Ю.Ф. Картавцев

МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА

Владивосток - 2008

Российская Академия Наук
Дальневосточное отделение
Институт биологии моря

Министерство образования РФ
Дальневосточный государственный
университет

Ю.Ф. Картавцев

МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА

Рекомендовано

Учебно-методическим объединением Министерства образования Российской Федерации

в качестве учебного пособия для студентов высших учебных заведений, изучающих курсы

«Популяционная генетика», «Общая генетика» и «Молекулярная биология»

УДК 576.1+551.7+575.17+575.224.4
ББК
К
Издание осуществлено при финансовой поддержке «Научно-образовательного центра морской биоты» ДВГУ

Рецензенты:
академик РАН, д.б.н., профессор Ю.Н. Журавлев;
к.б.н., с.н.с. К.А. Трувелер, к.б.н., с.н.с. В.В. Ефремов

Ответственные редакторы: д.б.н. И.В. Картавцева, к.б.н. О.Г. Корень

Ю.Ф. Картавцев
Молекулярная эволюция и популяционная генетика: Учебное пособие. – Владивосток: Издательство Дальневосточного государственного университета, 2-е изд. 2008. 25 п.л. ил. 182, библ. – 562.
ISBN …….

Учебное пособие предназначено для студентов старших курсов, обучающихся в университетах по биологическим специальностям. Книга может представлять также интерес для широкой аудитории специалистов, интересующихся генетикой и экологией популяций, проблемами вида и видообразования. В 16 главах и практических занятиях рассматривается широкий круг вопросов популяционной, эволюционной и экологической генетики с примерами из различных групп организмов, но преимущественно морских животных.

(C) Издательство Дальневосточного государственного университета, 2008
(C) Ю.Ф. Картавцев

Оглавление

Глава 1. ЭВОЛЮЦИОННАЯ ИСТОРИЯ ЖИЗНИ. *С. 4*

1.1. Главные этапы в геологической эволюции Земли.

1.2. Палеонтологические датировки эволюции.

1.3. Молекулярно-биологические датировки.

1.4. Задания для самообразования и практики.

Глава 2. МАТЕРИАЛЬНЫЕ ОСНОВЫ НАСЛЕДСТВЕННОСТИ. *С.13*

2.1. Структура хромосом.

2.2. Мейоз и формирование гамет.

2.3. Менделевская генетика.

2.4. Тонкая структура гена. Мутации и модификации.

2.5. Практическое занятие 2.

Глава 3. ТРАНСЛЯЦИЯ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ. ВВЕДЕНИЕ В ПРОТЕОМИКУ. *С.23*

3.1. Компоненты, критические для синтеза белков.

3.2. Процесс трансляции: от ДНК до полипептида.

3.3. Наследственность, белки и функция.

3.4. Введение в протеомику.

3.5. Практическое занятие 3.

Глава 4. ИММУНОГЕНЕТИЧЕСКИЙ АНАЛИЗ И ЭВОЛЮЦИОННЫЕ РАССТОЯНИЯ. *С.39*

4.1. Введение в иммуногенетику.

4.2. Наследование иммуногенетических признаков и популяционная изменчивость.

4.3. Иммуногенетическое датирование эволюции.

4.4. Датирование эволюции посредством анализа первичной последовательности аминокислот.

4.5. Практическое занятие 4.

Глава 5. БИОХИМИЧЕСКАЯ ГЕНЕТИКА. *С.54*

5.1. Введение в методы анализа изменчивости белков.

5.2. Генетическая интерпретация изменчивости белков.

5.3. Генетика изоферментов.

5.4. Эволюция регуляции генома.

5.5. Практическое занятие 5.

Глава 6. ГЕНЫ, МУТАЦИИ И ЭВОЛЮЦИЯ. *С.67*

6.1. Введение в проблематику исследования мутаций.

6.2. Структура и функции генов.

6.3. Типы мутаций на уровне ДНК.

6.4. Мутации и замены аминокислот.

6.5. Темп мутаций.

6.6. Практическое занятие 6.

Глава 7. ГЕНЕТИЧЕСКИЕ АСПЕКТЫ ВИДООБРАЗОВАНИЯ. *С.82*

7.1. Общий генетический подход: перспективы и ограничения.

7.2. Какие данные необходимы? Какова база данных?

7.3. Концепции вида. Обзор литературных данных по гетерозиготности и расстояниям.

7.4. Модели видообразования: популяционно-генетический взгляд.

7.5. Практическое занятие 7.

Глава 8. ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКАЯ СТРУКТУРА ВИДА. *С.103*

8.1. Популяционная структура бисексуальных видов.

8.2. Типы внутривидовой структуры.

8.3. Принципы исследования внутривидовой структуры.

8.4. Полевое популяционно-генетическое исследование.

8.5. Практическое занятие 8.

Глава 9. АНАЛИЗ ГЕНЕТИЧЕСКОГО РАЗНООБРАЗИЯ ВНУТРИ ВИДА: ТЕОРИЯ И ОПЫТ. *С.118*

9.1. Основные генетические параметры популяционной изменчивости.

9.2. Количественные меры дифференциации.

9.3. Модели популяционной структуры.

9.4. Эмпирические оценки параметров.

9.5. Практическое занятие 9.

Глава 10. ЕСТЕСТВЕННАЯ ГИБРИДИЗАЦИЯ И ИНТРОГРЕССИЯ В ПРИРОДЕ. *С.134*

10.1. Что такое гибриды и в чем суть гибридизации?

10.2. Методы определения гибридизации.

10.3. Генетические изменения в гибридных популяциях.

10.4. Эмпирические исследования гибридных зон.

10.5. Практическое занятие 10.

Глава 11. КОЛИЧЕСТВЕННЫЕ ПРИЗНАКИ: НАСЛЕДОВАНИЕ И ЭВОЛЮЦИЯ. *С.157*

11.1. Количественная изменчивость и ее законы.

11.2. Гипотеза множественных факторов.

11.3. Главные задачи генетического исследования количественных признаков: σ^2_{PH}, σ^2_G, Cv, h^2 и другие переменные.

11.4. Наследуемость.

11.5. Картирование локусов количественных признаков (ЛКП). Эволюция ЛКП.

11.6. Практическое занятие 11.

Глава 12. ГЕНЕТИЧЕСКИЙ ГОМЕОСТАЗ, ГЕТЕРОЗИС И ГЕТЕРОЗИГОТНОСТЬ. *С.176*

12.1. Концепция генетического гомеостаза.

12.2. Гетерозис и гетерозиготность.

12.3. Эмпирические результаты и интерпретация.

12.4. Практическое занятие 12.

Глава 13. ПОЛИМОРФИЗМ ДНК ВНУТРИ И МЕЖДУ ПОПУЛЯЦИЯМИ. *С.191*

13.1. Полиморфизм последовательности ДНК. Нуклеотидное разнообразие и *p*-расстояние.

13.2. Полиморфизм ДНК, оцененный по рестрикционным сайтам.

13.3. Полиморфизм длины ДНК. Теория и наблюдения.

13.4. Дивергенция популяций на уровне ДНК.

13.5. Практическое занятие 13.

Глава 14. ФИЛОГЕНЕТИЧЕСКИЕ ДЕРЕВЬЯ. *С.209*

14.1. Типы филогенетических деревьев.

14.2. Метод матрицы расстояний.

14.3. Методы парсимонии и максимального правдоподобия.

14.4. Роль популяционо-генетической теории в построении деревьев.

14.5. Практическое занятие 14.

Глава 15. ПРАКТИЧЕСКАЯ РАБОТА С ПОСЛЕДОВАТЕЛЬНОСТЯМИ НУКЛЕОТИДОВ: РЕДАКТИРОВАНИЕ, РЕГИСТРАЦИЯ, ВЫРАВНИВАНИЕ, КОНСТРУИРОВАНИЕ ДЕРЕВЬЕВ И АНАЛИЗ ФИЛОГЕНЕНИИ *227*

15.1. Редактирование последовательностей и их регистрация в генном банке.

15.2. Формат данных и доступные генные банки.

15.3. Выравнивание последовательностей.

15.4. Подбор оптимальной модели нуклеотидных замен.

15.5. Построение деревьев с помощью пакета программ MEGA-3 (MEGA-4). Аннотация возможностей программ PAUP, MrBayes и некоторых других.

15.6. Практическое занятие 15.

Глава 16. ЭВОЛЮЦИЯ ГЕНОМА. *С.246*

16.1. Возникновение и изменение генома.

16.2. Эволюционное изменение размера генома.

16.3. Формирование новых генов.

16.4. Повторяющиеся ДНК и мультигенные семейства.

16.5. Практическое занятие 16.

Оглавление на английском языке (Content in English) *С.265*

Предметный указатель *С.269*

Приложение I. Методическая инструкция для самостоятельной работы по курсу «Молекулярная эволюция и популяционная генетика». *С.270*

Приложение II. Вопросы для экзамена по курсу «Молекулярная эволюция и популяционная генетика». *С.274*

Приложение III. Основная литература и Интернет-поддержка. *С.278*

Приложение IV. Объяснение основных терминов. *С. 279*

Приложение V. Полный список литературы. *С. 283*

Предисловие

Данная книга является учебным пособием и возникла как результат объединения личного исследовательского опыта в популяционной, эволюционной и экологической генетике и более чем 27-летней практики преподавания спецкурса по этой проблематике для студентов-старшекурсников: преимущественно Дальневосточного государственного университета (Владивосток), а также студентов двух национальных университетов республики Корея - Корейского морского университета (Korea Maritime University, Busan) и Университета Чунгбук (Chungbuk University; Cheongju). В последние годы для удобства студентов материалы учебного пособия были адаптированы к современным компьютерным технологиям и к Интернету. В ходе обучения презентация материалов лекций базируется на программных средствах Микрософт (MS Power-Point), в ходе практических занятий осуществляется обращение к различным программным средствам поддержки, таким как POPULUS, BYOSIS, NTSYS, GENEPOP, ARLEQUINE, MEGA, STATISTICA и другим. Проводятся занятия по цифровому моделированию и статистическому анализу, включая обращение к многочисленным генетическим сайтам в Интернете и, наконец, с применением интерактивных обучающих программ. Каждая глава учебного пособия сопровождается практическим курсом, с разборкой алгоритмов расчетов и формул, знакомством с пакетами программ, рассмотрением конкретных примеров и решением задач. Книжное изложение, конечно, отличается от живого, интерактивного взаимодействия во время лекций и практических занятий. Например, каждая лекция сопровождается разделом, где суммируются и объясняются все новые термины. В книге все эти разделы сведены в специальное приложение. Однако, письменное изложение имеет свои неоспоримые преимущества. Оно позволяет неограниченно во времени прорабатывать сложные вопросы, дает возможность многократно возвращаться к необходимому материалу и др. Лекции базируются, как правило, на свободном изложении со слайдами-иллюстрациями в виде таблиц и рисунков. Каждая лекционная тема, включая практическое занятие, была дополнительно представлена письменной версией. На основе этой письменной версии и родилось данное учебное пособие. Весь материал изложен в 16 главах, составляющих основу курса. Он излагается по мере увеличения сложности, давая возможность слушателям (читателям) вспомнить основные понятия генетики и нарастить знания по специальным разделам, включая такие направления, как иммуногенетика, биохимическая генетика и молекулярная генетика в приложении к проблемам общей биологии, экологии и эволюции.

Исходно предполагается, что читатель (слушатель) владеет основами генетики. Поэтому для повторения представлены лишь два раздела по Общей генетике: Материальные основы наследственности (Глава 2) и Трансляция генетической информации (Глава 3). Эти два раздела являются принципиальными с точки зрения интерпретации фенотипической изменчивости и передачи признаков в поколениях потомства, что составляет основу генетического анализа. В кратком экскурсе по Общей генетике нет возможности повторить все темы. Поэтому, в настоящем курсе не дается систематического изложения репликации, транскрипции и многих других фундаментальных разделов Общей генетики. Хотя частично эти пробелы восполняются по мере раскрытия других тем курса. В частности, подробно рассматриваются генетический код и мутации, имеющие прямое отношение к данному курсу лекций, затрагиваются вопросы регуляции активности генов, сплайсинг и ряд других понятий.

Книга ориентирована не только на студентов, но может представлять интерес для широкой аудитории специалистов, интересующихся вопросами генетики и экологии популяций, вопросами вида и видообразования.

Хочется выразить свою благодарность вдохновителям данной книги, прослушавших лекции и рекомендовавших написание пособия, О.Г. Корень, А.Б. Холиной, а также первым читателям, внесшим много улучшений в стиль и смысл текста, И.В. Картавцевой, О.Г. Корень, А.В. Подлесных. Автор выражает искреннюю признательность также переводчику ИБМ Е.А. Коган за редактирование резюме на английском языке к главам книги.

Введение

Популяционная генетика и молекулярная эволюция – казалось бы, не столь близкие темы, чтобы объединять их в одной книге. Однако это не так. Более того, современная экспериментальная и теоретическая популяционная генетика получили сильный импульс после развития молекулярных методов анализа генетической изменчивости. С другой стороны, исследование дивергенции организмов во времени, то есть их эволюции, на основе молекулярных маркеров генов не мыслимо без понимания фундаментальных генетических свойств самих организмов и их групп, слагающих в природе такие единицы репродукции, как популяции и биологические виды. Популяционно-генетическую динамику во времени нельзя оторвать от пространственной динамики, от понимания основ внутривидовой генетической дифференциации. Увлеченные колоссальными возможностями филогенетического анализа на основе первичной последовательности дезоксирибонуклеиновой кислоты (ДНК) некоторые авторы вообще отвергают необходимость анализа пространственной дивергенции, противопоставляя филогенетическую концепцию вида биологической концепции вида. Однако многие генетики далеки от подобного экстремизма, понимая общность ряда механизмов внутри- и межвидовой дивергенции (Алтухов, 1983; 1989; Айала, 1984; Nei, 1987; Avise, Wollenberg, 1997). Эти вопросы, а также многие другие, и предполагается рассмотреть в предлагаемой книге в достаточно популярной форме в связи с ориентацией на студенческую аудиторию и специалистов широкого профиля. В тоже время, большинство глав книги являются краткими обзорами, что может представлять интерес и для более узких специалистов. Автор данной книги является морским биологом. Это обстоятельство наложило свой отпечаток на ее содержание, поставляя многие примеры именно для морского мира, – в основном, для рыб и беспозвоночных животных.

Истории органического мира и возникновению жизни посвящена Глава 1, а в заключительной части (Глава 16) рассмотрена эволюция генома. Основы генетического знания рассмотрены в Главах 2 и 3. Интенсивные исследования молекулярных основ эволюции организмов начались с разработкой и широким внедрением в биологию соответствующих методов, первоначально – иммуногенетических, затем генетико-биохимических и, наконец, молекулярно-генетических методов анализа различных макромолекул. Иммуногенетические работы получили развитие в середине пятидесятых годов минувшего века, с обоснованием количественно-генетических основ дивергенции по методике количественной фиксации микрокомплемента очищенного белка (Sarich, Wilson, 1966). Белок, обычно используемый для этой реакции, – это сывороточный альбумин (Champion et al., 1974). В связи с этим даны элементарные представления об иммунитете и иммуногенетике, а также представлены материалы по популяционно-генетическому и филогенетическому анализу ряда групп организмов (Глава 4).

Биохимическая генетика (Глава 5) в середине шестидесятых годов 20 века стала экспериментальной основой современной генетики природных популяций. Белки, чаще всего ферменты, как маркеры генов используются широко и в настоящее время, успешно конкурируя по дешевизне и экспрессивности с позднее появившимися ДНК-маркерами. Имеющаяся база данных в этой области, прежде всего по уровню гетерозиготности организмов и их генетическому сходству или различию, и сейчас представляет возможность

сравнительного анализа и обобщения, остро необходимых в популяционной, эволюционной и экологической генетике. В частности, в значительной мере на основе белковой изменчивости в лекциях рассмотрены: мутации (Глава 6), генетические аспекты видообразования (Глава 7), популяционно-генетическая структура вида (Глава 8) и анализ генетического разнообразия внутри вида (Глава 9), естественная гибридизация и интрогрессия в природе (Глава 10), гетерозис и его связь с гетерозиготностью, а также проанализирована сопряженность гетерозиготности и количественного признака в целом (Глава 12). Отдельно рассмотрены также количественные признаки, их наследование и эволюция (Глава 11).

С расшифровкой молекулярной сущности генов стало очевидным, что эволюционные взаимосвязи организмов могут быть исследованы посредством сравнения нуклеотидной последовательности ДНК или аминокислотных последовательностей в белках, кодируемых ДНК (Crick, 1958). Полиморфизм ДНК внутри и между популяциями, а также построение филогенетических деревьев рассмотрены в трех последовательных лекциях (Главы 13-15). Бум, порожденный возможностью анализа дивергенции различных линий или филумов организмов на основе первичных последовательностей нуклеотидов в генах, дал новое направление в эволюционной биологии – молекулярную филогенетику. Ее рождение и развитие дало не только решение многих злободневных вопросов эволюции органического мира, но потребовало также разработки новых методов математического анализа и программного обеспечения. В отечественной литературе имеется острый дефицит руководств по филогенетике. Хотя данная книга не специализирована на предмете молекулярной филогенетики как таковом, этим вопросам будет посвящена значительная доля излагаемого материала, что в какой то мере восполнит существующий пробел.

Во второе издание книги добавлена новая глава (Глава 15). Эта глава посвящена практической работе с первичными последовательностями нуклеотидов. рассмотрено их редактирование, регистрация, анализ и приемы построения молекулярно-филогенетических деревьев. Во втором издании переработаны и дополнены главы 13 и 14. Кроме того, все главы теперь имеют дополнительно к русскому заголовок и резюме на английском языке. Заголовки и пояснения к рисункам и таблицам также продублированы на английском языке. Введен предметный указатель и добавлено оглавление на английском языке. В конструкцию книги внесены композиционные изменения, обновлен список литературы и отредактированы практически все её разделы.

ГЛАВА 1. ЭВОЛЮЦИОННАЯ ИСТОРИЯ ЖИЗНИ

ГЛАВНЫЕ ВОПРОСЫ:

1.1. Главные этапы в геологической эволюции Земли.
1.2. Палеонтологические датировки эволюции.
1.3. Молекулярно-биологические датировки.
1.4. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1.

РЕЗЮМЕ

1. Земля, как часть Солнечной системы, проходит во времени через геохимические и геологические процессы. От этой геологической эволюции неотрывна и эволюция живого на планете, начиная с зарождения первых редуплицирующихся организмов.
2. Палеонтологические находки позволяют разбить историю живого мира на Земле на **эры, периоды и эпохи**. Неполнота палеонтологических датировок дает неточное представление о времени возникновения многих групп организмов.
3. **Молекулярно-биологические датировки** основываются на «молекулярных часах». Разные гены дают возможность исследовать различные в иерархии таксоны из-за неодинаковой скорости их эволюции. Как **молекулярные часы**, так и установленные филогении могут содержать ошибки. Но, как правило, молекулярные данные дают очень полезные и главное научно достоверные данные, позволяющие в совокупности с предыдущими знаниями уточнить как датировки времени, так и филогении любых ныне живущих организмов.

CHAPTER 1. EVOLUTIONARY HISTORY OF LIFE

SUMMARY

1. The earth, as a part of the Solar System, come in the course of time through geochemical and geological processes. Evolution of life at the planet inseparably linked with geological evolution from the origin of the first self-reduplicating organisms.
2. Paleontology records allow us to divide the history of living beings on the Earth into **Era, Periods and Epochs**. Incompleteness of paleontological records for many dating give us imprecise time for origin of many groups of organisms.
3. **Molecular biological dating** is based on "**molecular clocks**". Different genes because of their varied rates of evolution give an opportunity to investigate taxa of different hierarchical ranks. Both molecular clocks and reconstructed phylogenies can contain errors. However, as a rule molecular data give very useful and, which is especially important, a scientifically substantiated records that, togather with previous knowledge, sufficiently improve dating, as well as phylogeny for any recent living beings.

1.1. ГЛАВНЫЕ ЭТАПЫ В ГЕОЛОГИЧЕСКОЙ ЭВОЛЮЦИИ ЗЕМЛИ

Все, что происходит на Земле, как планете, является ее историей, слагающейся из геологической истории и эволюционной истории неорганической и органической материи. Земля, как часть Солнечной системы, возникла около 4,5 миллиардов (млд.) лет назад.

Возникновение Солнца датируется 5,0 млд. лет назад. С самого начала история Земли зависит от истории нашей звезды, Солнца, и излучаемой от него энергии. Другой источник энергии – это собственные химические реакции на Земле. Изначально это были исключительно геохимические реакции, а позднее, с появлением жизни, биохимические реакции и биогенно-зависимый оборот вещества на планете становится весьма существенным.

Набор химических реакций на планете меняется во времени и весь этот комплекс изменений можно именовать эволюцией Земли. На первых этапах на Земле осуществлялась только химическая эволюция, с превращением исходного межпланетного провещества в многообразие минералов, металлов и других соединений атомов и молекул. После образования воды в большом количестве и ее конденсации возникла жизнь. Фокс и Дозе (1975) в своей книге рассмотрели химическую эволюцию и возникновение жизни. Согласно этим авторам на Земле в период с 4,5 до 5,0 млд. лет назад преобладала молекулярная химическая эволюция, поскольку многообразие атомов в основном уже существовало. После возникновения жизни, около 4,0 млд. лет назад, появилась и начала интенсивно нарастать биохимическая эволюция, которая продолжается и по настоящее время. Геологи и палеонтологи делят всю временную шкалу на определенные эры и периоды. Основные из множества этих подразделений будут рассмотрены в следующем разделе 1.2.

Геологическая структура Земли также эволюционирует. Меняются очертания и свойства глубинных геологических пород и континентально-тектонических плит, что приводит к изменениям и движениям поверхности континентов, а также к изменению очертаний Мирового океана. Таким образом, без рассмотрения геологической эволюции трудно понять и биологическую эволюцию, эволюцию организмов, обитающих в этой меняющейся среде.

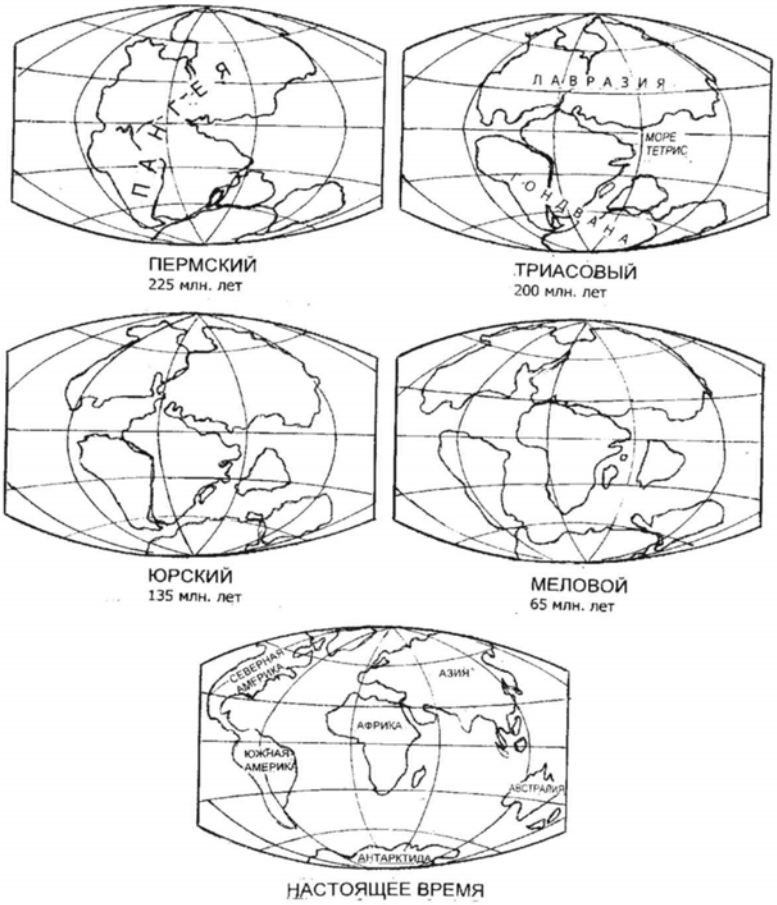


Рис. 1.1.1. Распределения континентальных масс в различные периоды Мезозойской эры (По On-Line Biology Book, 2002, с изменениями).
Fig. 1.1.1. Distribution of the land mass during the Mesozoic era (From On-Line Biology Book, 2002, with modifications).

Среди геологов сейчас широко признано, что континенты на Земле постоянно дрейфуют, в прошлом соединяясь и расходясь несколько раз (Calder, 1983). В Кембрийскую эру, около 550 млн лет назад, существовало 4 континента, но со временем, в Пермском периоде примерно 200 млн лет назад, они слились, образовав единый суперконтинент Пангею. Этот суперконтинент по прошествии 20 млн лет однако разделился на два суперконтинента – Лавразию и Гондвану (рис.1.1.1). Лавразия состояла из сегодняшних Северной Америки, Гренландии и Евразии, севернее Альп и Гималаев, тогда как Гондвана включала сегодняшнюю Южную Америку, Африку, Индию, Австралию и Антарктиду. Суперконтиненты позднее подразделились на более мелкие части, но рифты между ними не были эффективной преградой для миграции наземных животных вплоть до Мелового периода, когда расстояния между континентами стали велики. Примерно 75 млн лет назад Северная Америка полностью отделилась от Азии, а Гондвана разделилась на Южную Америку, Австралию, Антарктиду и т.д. Несколько важных аспектов настоящего географического распределения животных и растений могут быть объяснены континентальным дрейфом. Дрейф континентов продолжается, и даже в текущей Ценозойской эре, еще 60 млн лет назад, распределение наземных форм сильно отличалось от настоящего времени.

Информация о континентальном дрейфе использовалась многими авторами для датировки времени дивергенции между различными группами организмов (табл. 1.1.1). Так, Мэксон и Вильсон (Maxson, Wilson,1975) построили филогенетическое дерево лягушек из Северной Америки, Южной Африки, Австралии и Азии, используя молекулярные часы для гена альбумина. Они показали, что южно-американские и австралийские виды дивергировали около 75 млн лет назад, что находится в согласии с информацией о разделении двух континентов (см. табл. 1.1.1). Сибли и Элквист (Sibley, Ahlquist, 1984) использовали время разделения континентов, около 80 млн лет назад, для калибровки молекулярных часов по данным ДНК-гибридизации между страусами Африки и их общими предками в Южной Америке. Воутер с коллегами (Vawter et al., 1980) использовали время сформирования сухопутного моста между Северной Америкой и Южной Америкой для калибровки электрофоретических часов.

Таблица 1.1.1. Время различных геологических событий, использованных для калибровки молекулярных часов *(По Nei, 1987)*

Table 1.1.1. Times of various geological events that have been used for calibration of molecular clocks *(From Nei, 1987)*

Сравниваемые континенты или острова	***Время (млн. лет назад)***	***Автор***
I. Разделение земель		
Африка/Южная Америка	80	Sibley, Ahlquist, 1984
Новая Зеландия/Австралия	80	То же
Южная Америка/Австралия	75	Maxson, Wilson, 1975
Пацифика/Мексиканский залив	2-5	Vawter et al., 1980
II. Формирование островов		
Гавайи (Hawaii)	0.8	Hunt et al., 1981
Оаху (Oahu)	4	То же
Каваи (Kauai)	6	То же
Малые Антилы (Lesser Antilles)	3-5	Kim et al., 1976
Галапагосы (Galapagos)	0.5-4	Yang, Patton, 1981

Другой тип геологических событий, который важен для эволюционных исследований, – это оледенения. В истории Земли было несколько теплых и холодных периодов (рис. 1.1.2). Оледенение и таяние ледников вызывали резкие изменения климата планеты, которые в свою очередь изменяли распределение животных и растений. Оледенения изменяли также уровень воды в Мировом океане. Резкие колебания уровня моря (до 200 м) в свою очередь могли быть причиной быстрой смены ихтиофауны (Линдберг, 1948). В течение периода оледенения уровень моря понижался, так что острова или континенты, которые были ранее изолированы, могли соединяться. Это позволяло наземным животным и растениям мигрировать, преодолевая предыдущие барьеры. Аналогично, различные речные системы могли соединяться, а моря – разъединяться, устраняя или создавая различные изолирующие барьеры на пути миграции организмов и их генов. Длительные периоды оледенения приводили к изоляции морей, прекращению потока генов и раздельной дивергенции обитающих в них популяций, что могло стать основой для формирования новых видов. Аналогичные процессы происходили и на континентах.

Рис. 1.1.2. Возраст ледников, датированный по концентрации тяжелого изотопа кислорода (^{18}O) (По Hays et al., 1976).
Fig. 1.1.2. Ice ages recorded by heavy oxygen (^{18}O) (From Hays et al., 1976).

1.2. ПАЛЕОНТОЛОГИЧЕСКИЕ ДАТИРОВКИ ЭВОЛЮЦИИ

Когда возникли первые самореплицирующиеся организмы точно не известно. С того времени, когда Баркгоурн и Шопф (Barghoorn, Schopf, 1966) представили свои данные о вероятных ископаемых бактериях, появилось множество публикаций с заявлениями о находках микроископаемых с древними датировками Докембрийской эры (570 млн. лет назад и более). Хотя большинство из этих находок в ходе дальнейших перепроверок не подтвердились (Schopf, Walter, 1983; Hoffmann, Schopf, 1983), бактериоподобные микроископаемые, представленные Эурэмик с соавторами (Awramik et al., 1983), кажутся несомненными. Позднее Уэлш и Лоув (Walsh, Lowe, 1985) представили аналогичные данные о древних бактериоподобных микроископаемых. Обе находки датировали возрастом 3.5 млд. лет назад. Некоторые общепринятые датировки, возникновения различных групп организмов ранней истории Земли представлены на рис. 1.2.1. Обозначения и датировки для более недавних геологических периодов даны в Таблице 1.2.1.

Милли-оны лет	ГЕОЛОГИЧЕСКАЯ ВРЕМЕННАЯ ШКАЛА		Милли-оны лет	
	Эра	Период		
0	Ценозойская	Четвертичный	0-2	
		Третичный	65	Гибель динозавров
100	Мезозойская	Меловой	144	
200		Юрский	213	Первые динозавры, млекопитающие, птицы
300		Триассовый	248	
	Палеозойская	Пермский	286	
400		Угольный	360	Первые рептилии
		Девонский	408	Первые амфибии
500		Силурский	438	
		Ордовик	505	Первые наземные растения
600		Кембрий	590	Первые рыбы
700	Докембрийская	Докембрий	700	Первые беспозвоночные
…			…	
4600			4600	

Рис. 1.2.1. Геологическое время и ранняя история жизни (По On-Line Biology Book, 2002, с изменениями).

Fig. 1.2.1. Geological time and the early history of life (From On-Line Biology Book, 2002).

Когда возникли первые самореплицирующиеся организмы, точно не известно. С того времени, когда Баркгоурн и Шопф (Barghoorn, Schopf, 1966) представили свои данные о вероятных ископаемых бактериях, появилось множество публикаций с заявлениями о находках микроископаемых с древними датировками Докембрийской эры (590 млн лет назад и более). Хотя большинство из этих находок в ходе дальнейших перепроверок не подтвердились (Schopf, Walter, 1983; Hoffmann, Schopf, 1983), бактериоподобные микроископаемые, представленные Эурэмик с соавторами (Awramik et al., 1983), кажутся несомненными. Позднее Уэлш и Лоув (Walsh, Lowe, 1985) представили аналогичные данные о древних бактериоподобных микроископаемых. Обе находки датировали возрастом 3.5 млрд лет назад. Некоторые общепринятые датировки, возникновения различных групп организмов ранней истории Земли представлены на рис. 1.2.1. Обозначения и датировки для менее давних геологических периодов даны в таблице 1.2.1.

Рассмотрев эти микроископаемые и другие ископаемые находки, Шопф с соавторами (Schopf et al., 1983) выдвинули предположение, что жизнь возникла около 3.8 млрд лет назад. К периоду 3.5 млрд лет назад, по-видимому, уже возникли и анаэробные и фотосинтетические бактерии. Согласно этим авторам (Schopf et al., 1983) одноклеточные, митотически делящиеся эукариоты возникли 1.5 млрд лет назад, а дивергенция между животными и растениями произошла в районе 600 млн или 1 млрд лет назад, по-видимому, ближе ко второй дате.

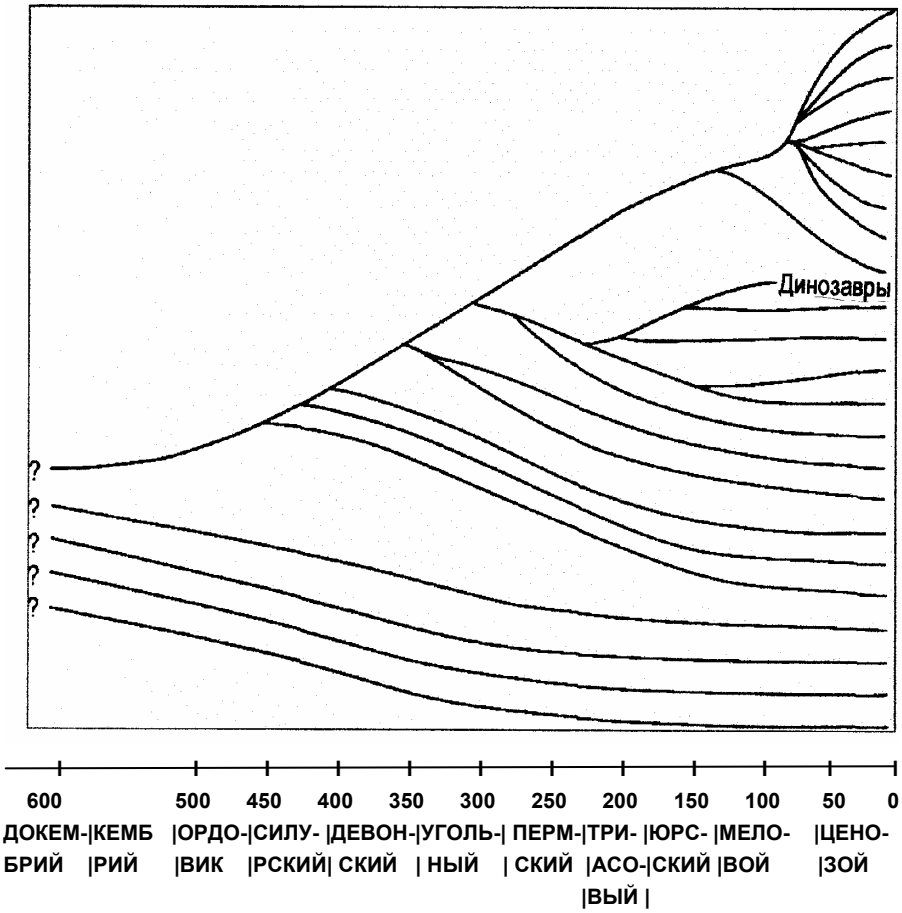
Общая картина филогенетических взаимосвязей среди живых существ представлена на рис. 1.2.2. Ископаемые палеонтологические находки позволяют датировать многие из отмеченных точек дивергенции (рис. 1.2.1). Множество дополнительных и увлекательных находок заинтересованный читатель найдет в Интернете (On-Line Biology Book, 2002).

Не стоит думать, что точные точки ветвления в эволюционном древе (рис. 1.2.1) обозначают точные знания о датах и родственных связях тех или иных филумов. Большинство из них неизвестны и являются продуктами различных реконструкций и экспертных оценок. Эти реконструкции в большинстве основаны на данных палеонтологии и сравнительной морфологии, и являются далекими от какой-либо точности. Поэтому молекулярно-генетическое датирование призвано сыграть здесь значительную роль с тем, чтобы появилась и большая определенность датировок, и прояснились генетико-генеалогические составляющие филогенетического дерева. Хотя молекулярно-генетические датировки сами по себе не являются абсолютно надежными, но в комплексе с другими данными позволили значительно продвинуть филогенетику. Молекулярная филогенетика – это сравнительно новая область биологии, которая сейчас быстро прогрессирует, успешно инкорпорируя бурно прогрессирующую молекулярную биологию и микроэлектронику с её новыми вычислительными возможностями.

Таблица 1.2.1. Шкала геологического времени и наименований градаций в Неогеновой (Ценозойской) эре

Table 1.2.1. Scale of a geological time for the Neozoic (Cenozoic) Era

Период	***Эпоха***	***Время, млн. лет***
Неогеновый (Четвертичный)	Плейстоцен	1.75
	Плиоцен	5.32
	Миоцен	23.8
Палеогеновый	Олигоцен	33.7
	Эоцен	54.5 (54.8)
	Палеоцен	65.0



- Приматы
- Грызуны
- Кроличьи
- Китообразные
- Хищные (собаки, коты)
- Парнокопытные (свиньи, бычьи)
- Непарнокопытные (лошади)
- Слоны
- Сумчатые
- Птицы
- Крокодилы
- Змеи
- Ящерицы
- Черепахи
- Лягушки
- Саламандры
- Костистые рыбы
- Акулы и скаты
- Миноги и миксины
- Насекомые
- Высшие растения
- Грибы
- Бактерии

(Миллионы лет)

Рис. 1.2.2. Реконструкция дивергенции позвоночных животных и некоторых других групп на основании палеонтологических и морфологических данных (По McLaughlin, Dayhoff, 1972; заимствовано из Nei, 1987).
Fig. 1.2.2. Divergence of the vertebrate groups and lineages for few other groups according to geological and morphological evidence (Modified from McLaughlin, Dayhoff, 1972).

1.3. МОЛЕКУЛЯРНО-БИОЛОГИЧЕСКИЕ ДАТИРОВКИ ЭВОЛЮЦИИ

С расшифровкой молекулярной сущности генов стало очевидным, что эволюционные взаимосвязи организмов могут быть исследованы на основе сравнения нуклеотидной последовательности дезоксирибонуклеиновой кислоты (ДНК) или аминокислотных последовательностей в белках, кодируемых на основе ДНК (Crick, 1958). Цукеркандль и Полинг (Zuckerkandl, Pauling, 1962; 1965) и затем Марголиш и Смит (Margoliash, Smith, 1965) показали, что скорость аминокислотных замещений в белках приблизительно постоянна в шкале времени, измеряемого годами. Это открытие дало новый метод для построения филогенетических деревьев.

Вскоре принцип константности мутационных замен в генах был перенесен на ДНК и РНК (рибонуклеиновая кислота). Многие авторы использовали этот метод для прояснения филогенетических взаимосвязей в различных группах организмов (Dayhoff, 1969; 1972; Ayala et al., 1974; Brown, 1983). Хотя эволюционные деревья, построенные этим методом, подвержены как большим выборочным ошибкам, так и систематическим ошибкам, полученные результаты оказываются очень часто вполне состоятельными. Недавние данные показывают, что молекулярные часы не являются очень точными, как первоначально представлялось, но это серьезно не влияет на пригодность молекулярных данных для филогенетических целей, да и сами датировки могут быть скорректированы (см. Глава 14).

Одно из преимуществ молекулярно-генетических методов состоит в том, что скорость (или темп) аминокислотных или нуклеотидных замен сильно варьирует для разных генов. Это позволяет исследовать эволюцию и на коротких временных отрезках, и длительную эволюцию во времени, используя различные гены. Можно привести здесь аналогию с датировками времени по различным радиоактивным элементам, с различающимся временем полураспада. Гены рибосомных РНК (рРНК) и транспортных РНК (тРНК) эволюционируют очень медленно в ядерном геноме, поэтому, их активно использовали для исследования ранних стадий биологической эволюции и радиации на Земле (McLaughlin, Dayhoff, 1970; Kimura, Ohta, 1973; Fox et al., 1977; Hori, Osawa, 1979). Хори и Осава (Hori, Osawa, 1979), например, исследовали нуклеотидные различия 5S рРНК у разных видов эукариот и прокариот и обнаружили естественную кластеризацию, как внутри этих групп, так и их самих (рис. 1.3.1).

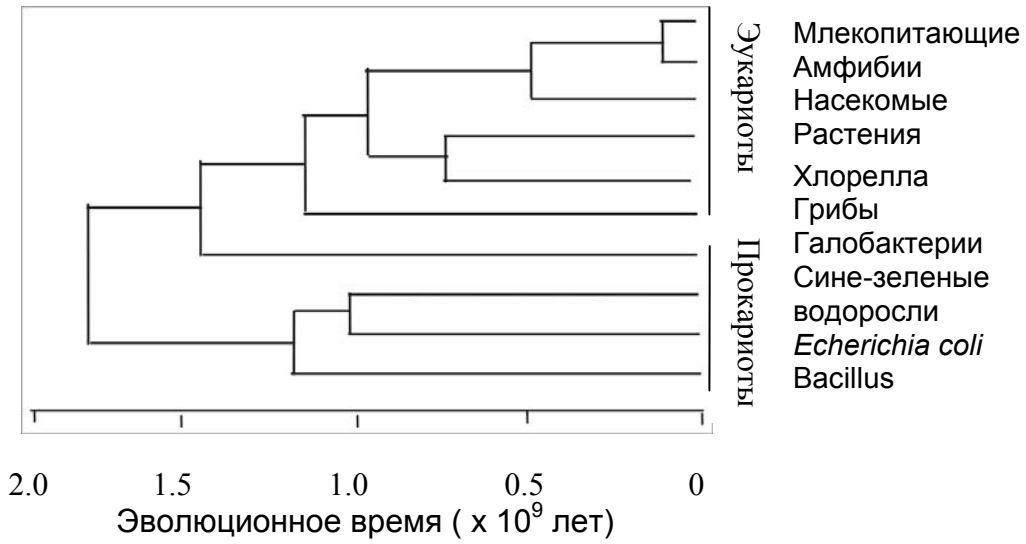


Рис. 1.3.1. Филогенетическое дерево различных видов эукариот и прокариот, полученное на основе анализа нуклеотидных различий 5S рРНК (По Hori, Osawa, 1979).

Fig. 1.3.1. Phylogenetic tree of various eukaryote and prokaryote species inferred from nucleotide sequence differences of 5S RNAs (From Hori, Osawa, 1979).

Временная шкала, приведенная на рис. 1.3.1, введена при допущении постоянной скорости нуклеотидных замен. Согласно этому дереву эукариоты дивергировали от прокариот 1,5 млрд лет назад, а растения и животные разошлись около 1,0 млрд лет назад. И хотя эти оценки строятся на многочисленных допущениях, они не являются бессмысленными, учитывая согласованность с палеонтологическими находками. Важно отметить также, что различные группы прокариот дивергировали очень давно и некоторые группы разошлись еще до дивергенции собственно прокариотических и эукариотических предков (Nei, 1987, p. 13). Так, галобактерии кластеризуются сначала с эукариотами и эта комплексная группа затем объединяется с тремя другими представителями прокариот (рис. 1.3.1). Однако, как мы увидим позже в Лекции 14, такие деревья имеют большие топологические ошибки, что вынуждает с осторожностью относиться к широкомасштабным обобщениям на основе только этих данных.

Еще в 50-е годы предыдущего столетия обычным было деление многообразия живых форм на две группы – животные и растения. Такая классификация впоследствии была признана упрощенной и дополнительно были введены такие таксоны как грибы, водоросли и бактерии (Whittaker, 1969). Позднее Воронцов (Vorontsov, 1989) продолжил эту логику и ввел для мегасистематики специальную империю Noncelulata, чтобы включить вирусы, и установил подимперию Mesokaryota для динофлагеллят. В империю Cellulata Воронцов ввел подимперию Mesokaryota, подимперию Prokaryota, с царствами Archeobacteria и Eubacteria, а также подимперию Eukaryota. Даже такая система была признана неудовлетворительной и введена многоцарственная система, учитывающая молекулярно-генетические данные и данные клеточной биологии (Кусакин, Дроздов, 1994; 1998). Такая система включает сейчас 26 царств, в которые входят 219 типов клеточных организмов (Drozdov, Kusakin, 2002).

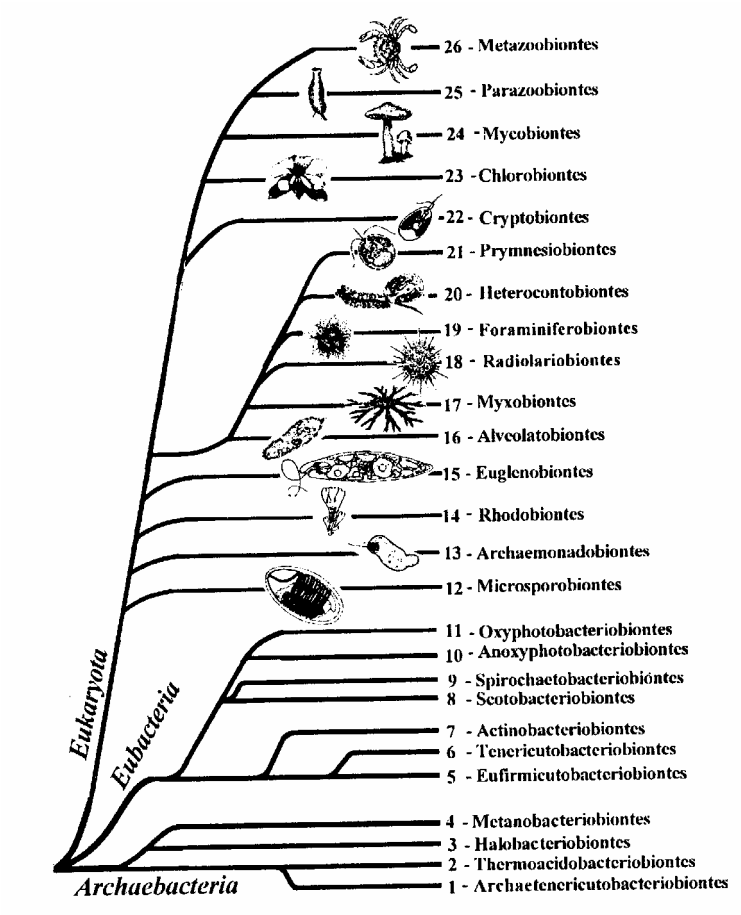


Рис. 1.3.2. Филогенетическая схема многоцарственной системы клеточных живых существ (По Drozdov, Kusakin, 2002).
Fig. 1.3.2. Phylogenetic scheme of multikingdom system of the cellular living beings (From Drozdov, Kusakin, 2002).

Для высших организмов, на основе анализа различных белков и генов, было предложено множество филогенетических деревьев (Fitch, Margoliash, 1967a; Sarich, Wilson 1967; Dayhoff, 1969; 1972; Lakovaara et al. 1972; Kim et al., 1976; Goodman et al., 1982; Sibley, Ahlquist, 1981; 1984; Brown et al., 1982; Avise et al., 1983; Картавцев и др., 1984; 2002, и многие другие). Эти деревья подвержены большим выборочным ошибкам и, вероятно, большинство из них не являются окончательными представлениями о филогении в той или иной группе. Тем не менее, многие из таких представлений оказались полезными для систематики или филогении изученных групп (см. подробности в Лекции 7). В частности, эти данные дали многое для понимания степени общности обезьян и человекоподобных (рис. 1.3.3).

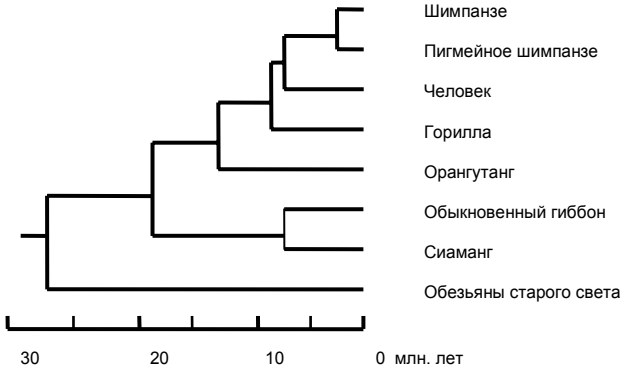


Рис. 1.3.3. Филогенетическое дерево для человекоподобных и обезьян старого света. Шкала эволюционного времени получена при допущении, что орангутанг дивергировал от линии человека 13 млн. лет назад (По Sibley, Ahlquist, 1984).
Fig. 1.3.3. Phylogenetic tree for hominoids and old world monkeys. The evolutionary time scale was obtained by assuming that the orangutan diverged from the human lineage 13 MY ago (From Sibley, Ahlquist, 1984).

Ранее, на основе палеонтологических находок, антропологи считали, что человек и шимпанзе дивергировали около 30 млн лет назад. Когда Сарич и Вильсон (Sarich, Wilson, 1967) обнаружили по данным иммуногенетического исследования, что время дивергенции составляет только 5 млн лет, антропологи критически восприняли эту информацию. Однако внимательно проанализировав палеонтологические данные, пришлм к заключению, что в действительности противоречия нет и возраст этой дивергенции на самом деле около 5 млн лет (Pilbeam, 1984), что согласуется и другими данными (см. рис. 1.3.3). Последние данные в целом не противоречат данным по аминокислотным последовательностям в белках (Goodman et al., 1982), данным дифференциального окрашивания хромосом (Yunis, Prakash, 1982) и данным по мтДНК (Глава 13).

1.4. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1

- Выберите тему реферата (Подберите по своему вкусу одну из тем, приведенных в Методической инструкции к курсу – см. Приложение I).
- Сообщите преподавателю, какая из тем вами выбрана и когда будет сделан доклад (дата защиты автореферата с устным выступлением).
- Принести калькуляторы на каждое последующее занятие для выполнения практических заданий.
- Уточнить остаточные знания по генетике (Проверить словарный запас. Напомнить некоторые термины).
- Выяснить, каковы остаточные знания по биостатистике и знанию программных средств.

ГЛАВА 2. МАТЕРИАЛЬНЫЕ ОСНОВЫ НАСЛЕДСТВЕННОСТИ

ГЛАВНЫЕ ВОПРОСЫ:

2.1. Структура хромосом.
2.2. Мейоз и формирование гамет.
2.3. Менделевская генетика.
2.4. Тонкая структура гена. Мутации и модификации.
2.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2.

РЕЗЮМЕ

1. Имеются четыре положения, или **закона**, **Менделя**:
   Наследуемые дискретные признаки контролируются единичными факторами, которые существуют парами в индивидуальном организме.
   Единичные факторы в первом поколении гибридов проявляют доминирование одного из элементов пары, что выражается в единообразии гибридов F_1.
   Единичные факторы расщепляются (сегрегируют) во втором поколении, в F_2, в определенной и устойчивой пропорции.
   Единичные факторы, контролирующие контрастные признаки, сочетаются независимо в поколении F_2 и в последующих поколениях сохраняют свою дискретность.

2. Понятия **локус, ген, аллель** и **цистрон** – это одни из ключевых терминов генетики, общность и различия между которыми надо понимать.

3. **Плейотропия** – это действие одного гена на несколько признаков. **Эпистаз** – это форма межлокусного взаимодействия, когда один из генов подавляет проявление другого. **Коплементация** выражается во взаимодополнении эффектов двух разных неаллельных генов. Эпистаз и комплементация приводят к отклонениям в пропорциях наследования, предсказываемых законами свободной сегрегации генов. **Полигения** – это контроль признака несколькими генами. Одним из важнейших понятий генетики является сцепление – совместное наследование генов, связанное с их расположением на одной и той же хромосоме. Внутри- и межгенные взаимодействия являются важнейшими свойствами генотипа.

4. Основные элементы тонкой структуры гена: кодирующие участки гена – **экзоны**, не кодирующие участки – **интроны, промотор** – регуляторный участок гена, с несколькими сайтами, и, наконец, **терминатор** – сайт, отмечающий окончание гена.

CHAPTER 2. THE MATERIAL BASIS OF HEREDITY

SUMMARY

1. There are four main conclusions or **Mendel's Laws.**

• *Genetic characters are controlled by* unit factors *that* exist in pair*s in individual organisms.*

• *One of unit factors appeared to be* dominant *to another unit factor, which resulted in uniformity of the first generation of cross, F1 hybrids.*

• *In the second generation of hybrids, in* F_2, segregation *of unit factors occurred in a certain and stable proportion.*

• *Unit factors that control contrasting pairs of traits give an* independent assortment *of one to another.*

2. Notions of a **gene, allele, locus**, and **cistrone** are the most important for genetics as a science. Students should precisely understand both similarity and difference between these terms.
3. **Pleiotropy** is a gene action mode when one gene has an impact on two or more phenotypic traits. **Epistasis** is an inter-locus interaction, when one gene inhibits expression of the other gene. **Complementation** is an inter-locus interaction, when one or more genes added to the expression of the original gene a new quality resulting in a new phenotype. **Polygene** is one of the series of genes that control one phenotypic character.
4. In accordance with the modern concept a fine gene structure is comprised by the following parts: *I, a site of initiation of transcription, R,* ***Regulator (CAAT)*** *and O,* ***Operator*** *sites of a* ***Promoter****,* ***Exon****, coding part of the gene,* ***Intron****, non-coding part of the gene, T, a site of termination of transcription. Other elements may also be pointed out such as an enhancing site of the Promoter (in some viruses) and complex pathways both for regulation of transcription and translation.*

2.1. СТРУКТУРА ХРОМОСОМ

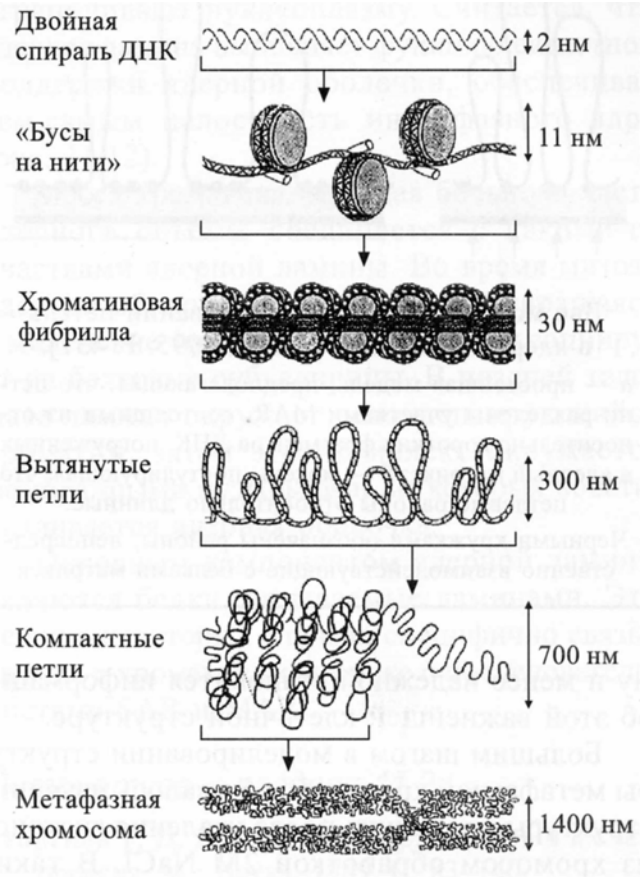
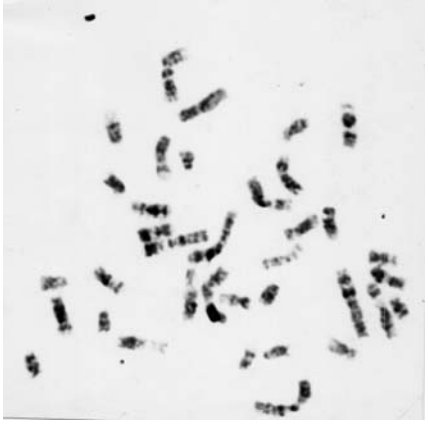


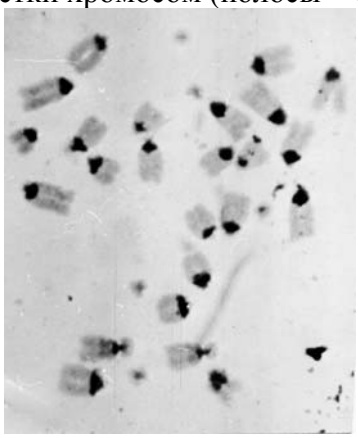
Рис. 2.1.1. Схематическое представление хромосомы и ее сложного строения.
«Бусы на нити» показывают нити ДНК, накрученные вокруг нуклеосом. Показаны также другие варианты упаковки ДНК в хромосоме (По Russel, 1998, p.333; Заимствовано из Жимулев, 2002, с.309).
Fig. 2.1.1. Schemmatic representation of a chromosome and its complex structure.
"Beads on a thread" show DNA that involute around the nucleosomes. Other modes of DNA packing are shown (From Russel, 1998, p.333; Adopted after Zhimulev, 2002, p. 309).

Хромосома – это органоид клетки, входящий в состав клеточного ядра и который отчетливо виден в световой микроскоп на стадии метафазы (рис. 2.1.1, показано внизу) и состоящий из ДНК и белков. В каждой метафазной хромосоме имеются по две хроматидные нити, которые представлены молекулой ДНК, упакованной вокруг нуклеосом, состоящих из комплекса белков (рис. 2.1.1, вторая сверху схема, нуклеосомы состоят из 8 субъединиц белков-гистонов). Репликация ДНК в процессе клеточных делений в митозе и мейозе является молекулярной основой точного копирования хромосом в клеточных линиях отдельного организма, а также в ряду родители - потомки. Нуклеосомы играют большую роль в процессах компактизации - декомпактизации хромосом в ходе клеточных циклов. Компактность хромосом в свою очередь является важной составляющей их цитологического поведения, точного удвоения и деления в

клеточных поколениях. В зависимости от положения центромеры выделяют три основных морфологических типа хромосом: метацентрические, акроцентрические и субметацентрические. Материал хромосомы называют хроматином. Его можно окрасить и, соответственно различиям окраски, выявить видоспецифичность и даже индивидуальность отдельных хромосом из их всего набора, называемого кариотипом или диплоидным набором, 2n (рис. 2.1.2). Окрашивание может быть как общим, так и выявляющим отдельные участки хромосом (полосы – G, C и др.; рис. 2.1.2).

A





B

Рис. 2.1.2. А. Хромосомы песчанки Виноградова (Meriones vinogradovi) в метафазе митоза, 2n=44; окрашивание по Гимза, G-полосы (Gimza G-banding). В. Хромосомы крысовидного хомяка (Tscherskia triton) в метафазе, 2n=28. С-полососы (Фотографии И.В. Картавцевой).
Fig. 2.1.2. A. Chromosomes of gerble (Meriones vinogradovi) in metaphase, 2n=44, Gimza G-banding. B. Chromosomes of hamster (Tscherskia triton) in metaphase, 2n=28. C-banding (Photographs presented by Dr. I.V. Kartavtseva).

2.2. МЕЙОЗ И ФОРМИРОВАНИЕ ГАМЕТ

***Мейоз** – это тип клеточного деления, при котором происходит редукция числа хромосом вдвое от их числа в домейотических клетках.* Когда клетка с диплоидным (двойным) набором хромосом (2n) проходит через мейоз, то в результате этого появляются четыре дочерние клетки, каждая генетически отлична одна от другой и каждая представлена гаплоидным набором хромосом (n) (рис. 2.2.1). Мейоз состоит из двух последовательных делений клеточного ядра. Сущность поведения хромосом в мейозе показана на рисунке 2.2.1. Схематически обозначены различные фазы и взаимодействие отдельных хромосом. При митозе число хромосом в клетке не меняется (2n=4). В противоположность этому, по завершении мейоза их число в каждой новой клетке уменьшается вдвое (n=2). Мейоз, как и митоз, начинается с покоящегося ядра (А), которое затем переходит в стадию профазы. Хромосомы в начале конденсации (Б) имеют вид одиночных нитей, на которых видны хромомеры. Затем соответствующие хромосомы попарно коньюгируют (В). В метафазе мейоза (Г) ядерная оболочка исчезает, а пары хромосом расположились в экваториальной плоскости клетки. Затем они разъединяются и направляются к разным полюсам клетки. Это приводит к тому, что каждая дочерняя клетка получает только две хромосомы, т.е. по одному представителю от каждой пары. Из этих клеток образуются впоследствии половые клетки (гаметы). На рисунке варианты (Г) и (Г’) иллюстрируют перекомбинацию родительских хромосом (показаны черным и белым цветом для отцов и матерей). В результате чего образуются четыре разные гаметы, в равном числе (Д и Д’). Видно также, что одна пара представлена метацентрическими хромосомами, а другая – акроцентрическими. Мейоз является физической (цитологической) основой расщепления аллельных генов и их случайной перекомбинации. В ходе

мейоза также осуществляется рекомбинация между неаллельными генами в результате кроссинговера.

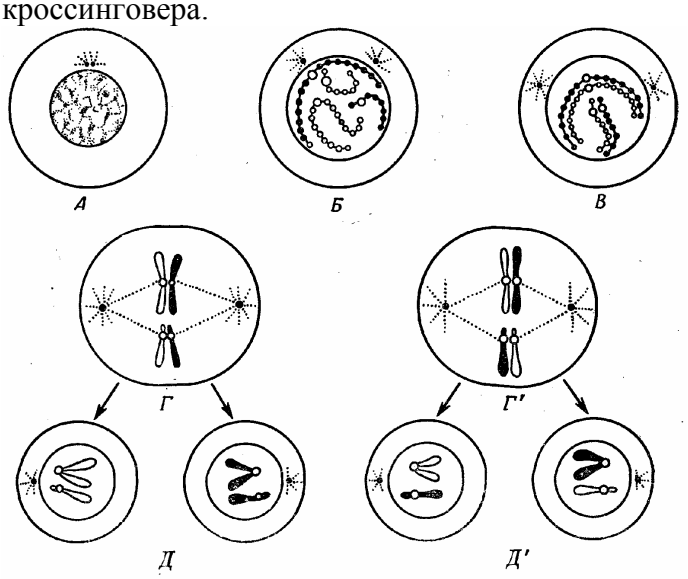


Рис. 2.2.1. Схематическое представление основных событий мейоза; вымышленный пример для двух пар морфологически различающихся хромосом.
Fig. 2.2.1. Overview of major events and outcomes of meiosis; an example for two chromosome pairs.

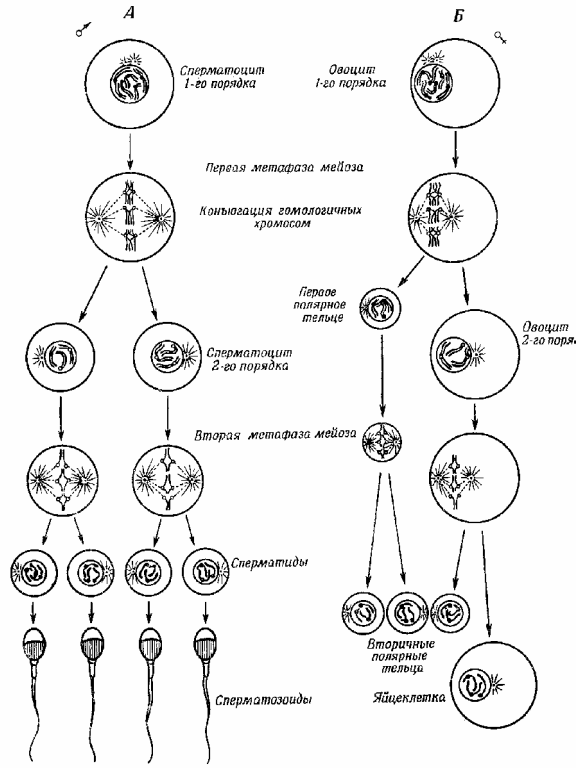


Рис. 2.2.2. Сперматогенез и овогенез клеток животных.
Fig. 2.2.2. A spermatogenesis and oogenesis in animal cells.

Хотя события, которые происходят в течение делений мейоза, сходны в клетках, которые участвуют в гаметогенезе различных видов животных, имеются определенные различия при формировании гамет самцов (сперматогенез) и при формировании гамет самок (овогенез). На рис. 2.2.2 дана общая схема этих процессов на примере позвоночных животных. **Сперматогенез** проходит в мужских половых органах (А). Процесс начинается с расширенного роста недифференцированных клеток зародышевого пути, называемых **сперматогониями.** Эти клетки увеличиваются и становятся **сперматоцитами 1-го порядка**, которые подвергаются первому мейотическому делению. Продукты этого деления называются **сперматоцитами 2-го порядка**, содержащими гаплоидное число диад. Вторичные сперматоциты в дальнейшем подвергаются второму мейотическому делению и каждая из этих клеток производит по два гаплоидных **сперматида**. Сперматиды проходят через серию последовательных стадий развития, **сперматогенез**, и становятся высокоспециализированными, мобильными **сперматозоидами** (спермой). Все клетки спермы, произведенные в течение сперматогенеза, получают равное количество ядерного генетического материала (равное число хромосом) и равную долю цитоплазмы (рис. 2.2.2).

В овогенезе животных (Б) формирование яйцеклетки или яйца происходит в яичниках – женских репродуктивных органах. Дочерние клетки возникают как результат двух мейотических делений, при которых генетический материал ядра (в виде хромосом) делится поровну, а цитоплазма распределяется неравномерно. Вместо этого, в каждом делении почти вся цитоплазма **овоцита 1-го порядка**, возникшего ранее из **овогония**, концентрируется только в одной из дочерних клеток. Такое неравное распределение диктуется потребностями будущего организма, во

многом осуществляющего начальные этапы эмбриогенеза за счет зрелого материнского яйца после оплодотворения. В течение первой мейотической анафазы в овогенезе тетрады первичного овоцита разделяются, а диады расходятся к противоположным полюсам (редукция числа хромосом). Далее, в течение первой телофазы, диады, отошедшие к одному из полюсов и получившие очень малую долю цитоплазмы, становятся так называемым **первым полярным телом**. Овогенез завершается формированием зрелого овоцита, который содержит гаплоидное число хромосом, n (рис. 2.2.2).

2.3. МЕНДЕЛЕВСКАЯ ГЕНЕТИКА

Передача по наследству простых дискретных признаков или качественных признаков, характеризующих внешний облик организмов и их другие фенотипические особенности, реализуется по законам, установленным еще Менделем. С современными цитологическими знаниями и обозначениями генов схема наследования отдельного признака очень проста (рис. 2.3.1). Также простым является представление скрещиваний и ожидаемых результатов с помощью решетки Пеннета (рис. 2.3.2).

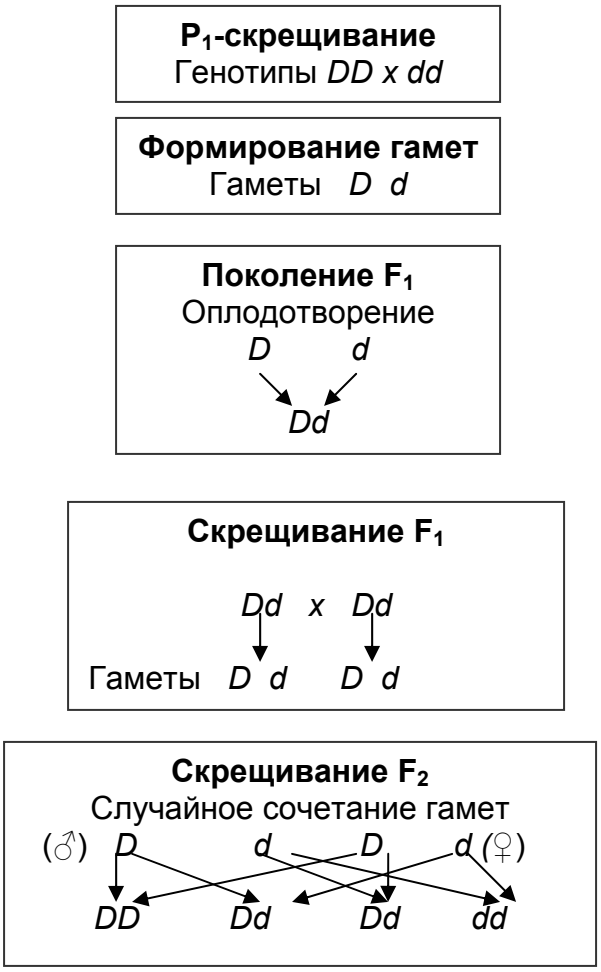


Рис. 2.3.1. Моногибридное скрещивание растений душистого горошка с длинным и коротким стеблем.
Fig. 2.3.1. The monohybrid cross between tall and dwarf pea plants.

Скрещивание F_1

Dd x Dd

Гаметы *D d D d*

Формирование решетки Пеннета для F_2

Гаметы (♀/♂)	*D*	*d*
D	*DD* (*длинные*)	*Dd* (*длинные*)
d	*Dd* (*длинные*)	*dd* (*короткие*)

Результаты F_2

Генотип	*Фенотип*
1 DD	*3/4 длинные*
2 Dd	
1 dd	*1/4 короткие*
1 : 2 : 1	*3 : 1*

Рис. 2.3.2. Формирование решетки Пеннета для F_2 потомств от скрещиваний F_1 x F_1, которые показаны на рис. 2.3.1.
Fig. 2.3.2. A Punnet square generates the F_2 of the F_1 x F_1 cross that shown in Fig. 2.3.1.

Простейший случай наследования – когда один ген контролирует один признак (фенотип). Такой контроль, как в рассмотренных примерах, называется моногенным, а скрещивание – моногибридным (рис. 2.3.1–2.3.2). Мендель рассматривал несколько признаков садового горошка и обнаружил четкие правила (законы) их передачи потомкам. На схеме (рис. 2.3.1) символ *d* и *D* обозначает единичные факторы (как их называл Мендель), детерминирующие растения с коротким и длинным стеблем, передающиеся потомкам через гаметы. То есть один признак определяется факторами в различных состояниях – доминантном (*D*) и рецессивном (*d*). Кроме прямых скрещиваний, в генетике обычны возвратные скрещивания и анализирующие скрещивания $\mathbf{F_b}$ и $\mathbf{F_a}$ с родительской линией. В случае коротко- и длинностебельных растений гороха имеем: $\mathbf{F_b}$ – $Dd \times DD$, $\mathbf{F_a}$ – $Dd \times dd$. Чуть более сложен случай, когда два гена контролируют формирование контрастных признаков, с доминированием одного из аллелей (рис. 2.3.3).

Скрещивание F_1: (♂) ***CcWw x*** (♀) ***CcWw***

Гаметы	CW	Cw	cW	cw
CW	CCWW Желтые, круглые	CCWw Желтые, круглые	CcWW Желтые, круглые	CcWw Желтые, круглые
Cw	CCWw Желтые, круглые	CCww Желтые, морщинистые	CcWw Желтые, круглые	Ccww Желтые, морщинистые
cW	CcWW Желтые, круглые	CcWw Желтые, круглые	ccWW Зеленые, круглые	ccWw Зеленые, круглые
Cw	CcWw Желтые, круглые	Ccww Желтые, морщинистые	ccWw Зеленые, круглые	ccww Зеленые, морщинистые

Рис. 2.3.3. Дигибридное скрещивание между двумя контрастными парами генов, контролирующих признак желтые – зеленые (C, c) и гладкие - морщинистые (W, w) плоды гороха. Скрещивания проиллюстрированы с помощью решетки Пеннета и вилочно-линейным методом.
Fig. 2.3.3. The dihybrid cross between yellow-green and round-wrinkle pea plants. Punnet square and forked-line methods are shown in the bottom.

Формирование пропорций по фенотипу исходя из частот

Пропорции генотипов, F_2		*Пропорции генотипов, F_2*
1/16 *CCWW*	→	9/16 *желтые, круглые*
2/16 *CCWw*	→	
2/16 *CcWW*	→	
4/16 *CcWw*	→	
1/16 *CCww*	→	3/16 *желтые, морщинистые*
2/16 *Ccww*	→	
1/16 *ccWW*	→	3/16 *зеленые, круглые*
2/16 *ccWw*	→	
1/16 *ccww*	→	1/16 *Зеленые, морщинистые*

Часты и более сложные случаи наследования. Когда три гена контролируют качественные признаки, мы имеем дело с тригибридным скрещиванием (рис. 2.3.4). Для таких скрещиваний наименее сложно использовать **вилочно-линейный метод (Forked-line method)**. Этот метод называют еще ветвистой диаграммой, и основывается он на простых вероятностных выкладках,

установленных для моногибридных скрещиваний. Предполагается, что каждая пара генов ведет себя независимо в течение формирования гамет. Последовательно рассматриваются три этапа.
Когда реализовано моногибридное скрещивание *AA x aa*, мы знаем, что:
1. Все особи F_1 имеют генотип *Aa* и проявляют (экспрессируют) фенотип, определяемый аллелем *A*, который называется **A** в последующих рассуждениях.
2. Поколение F_2 состоит из особей либо с фенотипом **A**, либо с фенотипом **a** в пропорции 3:1.
3. Такие же обобщения можно сделать для скрещиваний *BB x bb* и *CC x cc* (см. пропорции расщеплений, рис. 2.3.4 и Табл. 2.3.1).

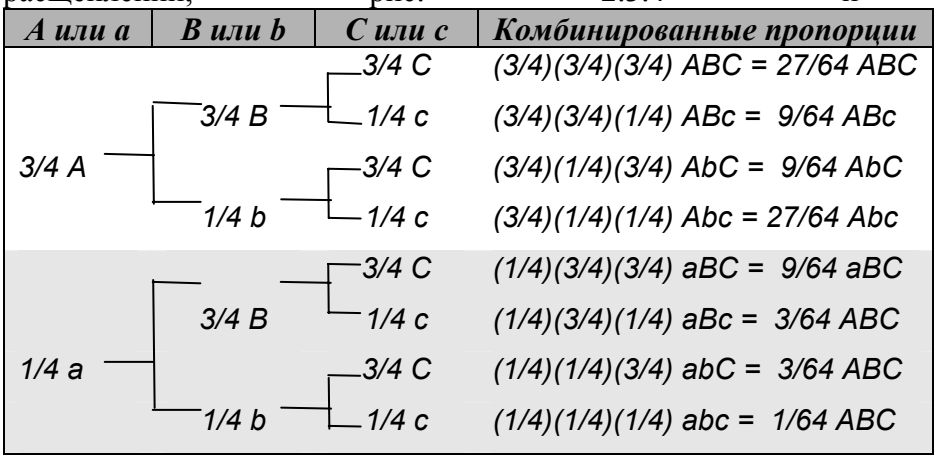


Рис. 2.3.4. Обобщение F_2 для пропорций тригибридного скрещивания, полученных вилочно-линейным методом (По Klug, Cummings, 2002).
Fig. 2.3.4. Generation of F_2 trihybrid genotypic ratio using the forked-line method (From Klug, Cummings, 2002).

Заканчивая данный раздел, давайте переформулируем основные правила или законы Г. Менделя, помня, что единичные факторы – есть гены в современном понимании.

Законы Менделя:

- Наследуемые дискретные признаки контролируются **единичными факторами**, которые **существуют парами** в индивидуальном организме.
- Единичные факторы в первом поколении гибридов проявляют **доминирование** одного из элементов пары, что выражается в единообразии гибридов F_1.
- Единичные факторы **расщепляются** (сегрегируют) во втором поколении, в F_2, в определенной и устойчивой пропорции.
- Единичные **факторы**, контролирующие контрастные признаки, **сочетаются независимо** в поколении F_2 и в последующих поколениях сохраняют свою дискретность.

Завершая повторение классической генетики, следует подчеркнуть, что независимое комбинирование гамет, отдельных хромосом и, даже без учета перекомбинации генов благодаря кроссинговеру (выяснившейся много позже первых Менделевских работ), дает колоссальные возможности для генетической изменчивости особей, популяций и видов (Табл. 2.3.1).

Таблица 2.3.1. Число классов гибридного потомства и расщепление в F_2 при различных комбинациях признаков при доминировании

Table 2.3.1. Number of classes in hybrid offspring and the mode of segregation in F_2 under different trait's combination

Скрещивание	**Число различающихся признаков, P**	**Число возможных типов гамет**	**Число возможных генотипов**	**Число классов**		**Пропорции фенотипов**
				фенотип	генотип	
Моногибридное	1	$2^1 = 2$	$4^1 = 4$	$2^1 = 2$	$3^1 = 3$	3:1
Дигибридное	2	$2^2 = 4$	$4^2 = 16$	$2^2 = 4$	$3^2 = 9$	9:3:3:1
Тригибридное	3	$2^3 = 8$	$4^3 = 64$	$2^3 = 8$	$3^3 = 27$	27:9:9:9:3:3:3:1

Тетрагибридное	4	$2^4 = 16$	$4^4 = 256$	$2^4 = 16$	$3^4 = 81$	81:27:27:27:27:9:9:9:9:9:9:3:3:3:3:1
Полигибридное	n	2^n	4^n	2^n	3^n	$(3:1)^n$

Некоторые другие генетические понятия

К важнейшим понятиям классической генетики относится понятие сцепления. Что такое **сцепление**? Сцепление - это совместное наследование пары или большего числа генов, кодирующих независимые признаки. Основные исследования по обоснованию сцепления выполнены в 20-м веке Морганом, Стертевантом и Бриджесом. Материальная основа сцепления сейчас абсолютно понятна и связана она с локализацией генов линейно вдоль ДНК и хромосомы. Частота кроссинговера, обмена участками хромосом, является, поэтому мерилом степени близости расположения генов; чем меньше частота, тем ближе располагаются гены. Если частота рекомбинации равна 50%, то гены считаются не сцепленными. Одна группа сцепления физически соответствует одной хромосоме.

К важнейшим понятиям генетики относится также понятие взаимодействия генов. Внутригенные взаимодействия – это доминирование, кодоминирование (неполное доминирование), сверхдоминирование, аддитивность и нейтральность. Межгенные взаимодействия – это плейотропия, эпистаз, коплементация и полигения. Все внутригенные взаимодействия и их количественное выражение будут рассмотрены подробнее в Главе 12, хотя понятие о доминировании сформулировано выше (см. постулаты Менделя), а механизм кодоминирования будет рассмотрен в Главе 5. На данном этапе для нас принципиально помнить, что аллели одного гена проявляют дифференциальную активность, которая может поглощаться, складываться, умножаться или быть индифферентной по отношению друг к другу. **Плейотропия** – это действие одного гена на несколько признаков. Частым плейотропным эффектом, например гена определяющего морфологический признак, является его влияние на жизнеспособность. **Эпистаз** – это форма межлокусного взаимодействия, когда один из генов подавляет проявление другого. **Коплементация** выражается во взаимодополнении эффектов двух разных неаллельных генов. Эпистаз и комплементация приводят к отклонениям в пропорциях наследования, предсказываемых законами свободной сегрегации генов. **Полигения** – это контроль признака, как правило, количественного признака, несколькими генами. Количественные признаки и их наследственную основу мы рассмотрим специально в Главе 11.

2.4. ТОНКАЯ СТРУКТУРА ГЕНА. МУТАЦИИ И МОДИФИКАЦИИ

Главные положения о тонкой структуре гена эукариот представлены на рис. 2.4.1. Основные элементы: кодирующие участки ДНК гена – экзоны, не кодирующие участки – интроны, промотор – регуляторный участок гена, с несколькими сайтами, и, наконец, терминатор – сайт, отмечающий окончание гена. Иногда гены могут перекрываться (показано стрелкой). Ген в современном представлении – это цистрон. **Цистрон –** это единица генетической функции, которая включает кодирующий участок молекулы ДНК и регуляторные элементы для синтеза макромолекул живых организмов (рис. 2.4.1). Таким образом, одним из важнейших свойств гена является детерминация и регуляция определенной элементарной функции. Не все гены активны одинаково. Регуляция их активности – очень сложный процесс. Можно подразделить регуляцию активности генов на две фазы, связанные с онтогенезом и нормальным функционированием отдельных клеток и организмов в целом. У высших организмов активация генов в онтогенезе происходит дифференцированно в ходе индивидуального развития, и их функционирование различается по тканям и органам (см. Глава 5). Активность генов в ходе нормального функционирования клеток регулируется, в основном, в ходе транскрипции, или синтеза мРНК на матрице ДНК, и в ходе трансляции, синтезе полипептидов на мРНК (Лекция 3). Известно несколько механизмов, с помощью которых один и тот же набор генов в неодинаковых условиях

жизнедеятельности организма и на разных стадиях развития детерминирует синтез белков. Регуляция экспрессии (выражения) генов может осуществляться на нескольких уровнях: генном, транскрипционном, трансляционном и функциональном (Алиханян и др., 1985). Первый из них связан с изменением количества или локализации генов, контролирующих данный признак. Второй определяет, какие и сколько мРНК должны синтезироваться в данный момент; это осуществляется сайтами оператора цистрона (рис. 2.4.1). Третий обеспечивает отбор мРНК, транслирующихся на рибосомах. Четвертый связан с аллостерической регуляцией активности ферментов. Контроль действия генов может осуществляться также путем посттрансляционной модификации полипептидов, посттранскрипционной модификации мРНК (сплайсинг) и другими путями.

*Рис. 2.4.1. Схема гена. I – сайт инициации транскрипции, R – регуляторный сайт (**CAAT**) и O – операторный сайт Промотора, Экзон – кодирующая часть гена, Интрон – не кодирующая часть гена, T – сайт терминации транскрипции. Могут быть указаны также усиливающий (enhancing) сайт Промотора (у некоторых вирусов) и комплексные пути как для регуляции транскрипции, так и для регуляции трансляции.*

*Fig. 2.4.1. A scheme of a gene. I, is a site of initiation of transcription, R, are a regulator (**CAAT**) and O, operator sites of a Promoter, Exon, is coding part of the gene, Intron, is non-coding part of the gene, T, is a site for termination of transcription. Other elements may also be pointed out, including enhancing site of the Promoter (in some viruses) and complex pathways both for the regulation of transcription and translation.*

Мутации

Гены определяют стабильные функции организмов. Однако иногда ген может неожиданно (спонтанно) измениться, или мутировать. Мутации генов в природных популяциях, как правило, редки. Хотя скорость мутаций (μ) широко варьирует у разных организмов и может сильно отличаться для разных генов: $\mu = 10^{-3} - 10^{-7}$, грубо. Генные или точечные мутации – это обычно изменения нуклеотидов: замены, вставки, утери. Имеются также различного типа хромосомные преобразования: инверсии, транслокации, дупликации и делеции. Эти хромосомные перестройки иногда также называют мутациями, но более правильно зарезервировать термин мутации для изменений на генном уровне. Отдельно о мутациях и их роли в эволюции мы еще поговорим в Главе 6.

Модификации

Модификации – это другой тип изменчивости генома, так называемые эпигенетические изменения, когда из принципиально подобного генотипа формируются фенотипически различные особи в зависимости от состояния среды. Для лососевых рыб, например, известны «жилые» и «проходные», речные и озерные экотипы или биотипы. Такая изменчивость играет большую роль в приспособленности популяций вида к меняющейся среде. Однако некоторые биологи (Б.

Медников, Россия) придают им значение, как одному из факторов видообразования.

2.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2

2.1. Генетический словарь. 1. Дайте определения для следующих обозначений - *AA, Aa, BbCc, ×, P, F1, Fa. 2.* Объяснить и разобрать в сравнении термины: *локус, ген, аллель, цистрон (см. ниже), инверсия, транслокация, дупликация и делеция (см. Приложение IV).*

- **Ген** – это кодирующая последовательность ДНК, которая определяет элементарную функцию организма, такую, например, как синтез фермента.
- **Аллель** – это вариант гена, с определенными особенностями функции (например, изменен заряд молекулы белка).
- **Локус –** это место на хромосоме где располагается данный ген. В тоже время термин локус употребляется как синоним термина ген.
- **Цистрон –** это единица генетической функции, которая включает кодирующий участок молекулы ДНК и регуляторные элементы для синтеза макромолекул живых организмов.

2.2. Независимое расщепление и статистика. Специфика качественных и количественных признаков.

2.3. Объяснить хи-квадрат критерий для оценки распределения частот качественных признаков. Разобрать алгоритм: $X^2 = \sum[(f - f')^2 / f']$, f' и f – это ожидаемые и наблюдаемые численности, d.f. = n - 1, где n – число фенотипических классов.

2.4. С помощью критерия хи-квадрат проверьте результаты опытов Менделя на рис. 2.5.1.

Моногибридное скрещивание				
Фактические численности (o)	Ожидаемые численности (e)	Отклонение (o-e)	(d^2)	$(d^2) / e$
740	3/4 (1000) = 750	-10^2	100	100/750 = 0.13
260	1/4 (1000) = 250	$+10^2$	100	100/250 = 0.40
Всего = 1000				χ^2= 0.53
				P = 0.48
Дигибридное скрещивание				
587	567	+20	400	0.71
197	189	+8	64	0.34
168	189	-21	441	2.33
56	63	-7	49	0.78
Всего = 1008				χ^2= 4.16
				P = 0.26

Рис. 2.5.1. Два примера расщеплений в скрещиваниях Менделя (По Klug, Cummings, 2002). Fig. 2.5.1. Two examples of segregation in Mendel's crosses (From Klug, Cummings, 2002).

2.5. Задание для самообразования с Интернет-поддержкой. Используя адреса сайтов Интернета, приведенные в Лекции 1, найдите три примера, иллюстрирующие законы Менделя. Покажите их преподавателю и поясните результаты.

ГЛАВА 3. ТРАНСЛЯЦИЯ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ. ВВЕДЕНИЕ В ПРОТЕОМИКУ

ГЛАВНЫЕ ВОПРОСЫ:

3.1. Компоненты, критические для синтеза белков.
3.2. Процесс трансляции: от ДНК до полипептида.
3.3. Наследственность, белки и функция.
3.4. Введение в протеомику.
3.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3.

РЕЗЮМЕ

1. **Трансляция** – это синтез полипептидных цепей, осуществляемый в клетках. Она реализуется на основе информации, хранимой в **мРНК** посредством рибосомного комплекса. Этот процесс, таким образом, конвертирует информацию, закодированную в последовательности нуклеотидов ДНК в последовательность аминокислот, представляющих полипептиды.
2. Трансляция – это комплексный энергоемкий процесс, зависящий от заряженных молекул **тРНК** и многочисленных белковых факторов, участвующих в синтезе. Транспортные РНК служат в качестве адапторных молекул между триплетами мРНК и соответствующей аминокислотой.
3. Исследование потребностей пищевых цепей у хлебных дрожжей *Neurospora* Бидлом с коллегами доказало, что мутации вызывают потерю активности соответствующих ферментов. Эта работа привела к концепции «**один ген – один фермент»**. Эта концепция была позже модифицирована в соответствии с более современным пониманием, что **один ген определяет синтез одного полипептида**.
4. Предположение, что нуклеотидная последовательность в гене определяет **колинеарно** последовательность аминокислот в полипептиде была доказана в экспериментах по синтезу триптофана у кишечной палочки *E. coli.*
5. **Белки**, являются конечными продуктами генов. Они обнаруживают четыре уровня структурной организации. **Первичная структура белка** определяется последовательностью аминокислот в цепи. **Вторичная и третичная структура** зависят от двух иерархий в трехмерной конформации полипептидной цепи. **Четвертичная структура** определяется комплексированием молекулы белка, имеющего несколько полипептидов (олигомерные белки).
6. Ферментативная каталитическая функция является одной из самых значимых среди многих сотен тысяч функций, выполняемых белками. Этот высокоспецифичный клеточный катализ играет центральную роль в синтезе многочисленных веществ в живых организмах. Белки состоят из одного или большего числа функциональных **доменов**, которые заимствуются различными молекулами. Происхождение этих доменов может быть связано с **перетасовкой экзонов** в процессе эволюции.
7. **Протеома** – это набор белков, экспрессирующихся с генов и модифицирущихся в течение жизни клетки. В узком смысле термин описывает также набор ферментов, имеющихся в клетке в данный момент времени. **Протеомика** – это наука, занимающаяся исследованием протеомы.

Chapter 3. TRANSLATION OF GENETIC INFO. AN INTRODUCTION TO PROTEOMICS

SUMMARY

1. **Translation** describes synthesis of polypeptide chains in cells, under the direction of mRNA and in association with ribosomes. This process ultimately converts the information stored in the genetic code of the DNA forming a gene into the corresponding sequence of amino acids making up the polypeptide.
2. Translation is a complex energy-requiring process that also depends on charged tRNA molecules and numerous protein factors. Transfer RNA (tRNA) serves as an adaptor molecule between an mRNA triplet and the appropriate amino acid.
3. Investigation of nutritional requirements in *Neurospora* by Beadle with colleagues made it clear that mutations cause the loss of enzyme activity. Their work led to the concept of **one-gene: one-enzyme**. The one-gene: one-enzyme hypothesis was later revised. Pauling and Ingram's investigations of hemoglobins from patients with sickle-cell anemia led to the discovery of the fact that one gene directs the synthesis of only one polypeptide chain.
4. The proposal suggesting that a gene nucleotide sequence specifies the sequence of amino acids in a polypeptide chain in a colinear manner was confirmed by experiments involving mutations in the tryptophan synthetase gene in *E. coli.*
5. Proteins, the end products of genes, demonstrate four levels of structural organization that collectively provide the chemical basis for their three-dimensional conformation, which is a basis for function of a molecule.
6. Of the myriad functions performed by proteins, the most influential role is assumed by enzymes. These highly specific, cellular catalysts play the central role in the production of all classes of molecules in living systems. Proteins consist of one or more functional domains, which are shared by different molecules. The origin of these domains may be the result of exon shuffling during evolution.
7. Proteome is a set of proteins expressed and modified during a cell entire lifetime. In a narrower sense, the term also describes a set of proteins expressed in a cell at a given time. **Proteomics** is the study of a proteome, and it uses technologies ranging from genetic analysis to mass spectrometry.

3.1. КОМПОНЕНТЫ, КРИТИЧЕСКИЕ ДЛЯ СИНТЕЗА БЕЛКОВ

Белки являются основными продуктами генов. Именно на основе белков строится все многообразие функций и черт фенотипа. Поэтому необходимо рассмотреть, какие процессы осуществляют синтез белков и как происходит передача генетической информации от гена (генотипический уровень) к белку (фенотипический уровень). Понимание этих процессов понадобится также, когда изменчивость белков будет использоваться для маркирования изменчивости генов.

Для синтеза белков наиболее существенными являются пять компонентов, информация о которых изложена ниже в соответствующих пунктах 1–5 Раздела 3.1, а суть трансляции дана в Разделах 3.2 и 3.3.

1. Матричная РНК

Матричная РНК (**мРНК**) играет роль в объединении вместе рибосомных субъединиц (объяснение ниже) и предоставления кодирующей информации, которая определяет последовательность аминокислот в производимой полипептидной цепи. Исходная мРНК подвергается у высших организмов созреванию, главным этапом которого является процессинг, то есть, вырезание интронов и регуляторных сайтов из первичной последовательности нуклеотидов. Роли мРНК в синтезе белков мы еще коснемся в последующих разделах.

2. Транспортная РНК

Транспортная РНК (тРНК) представлена классом молекул, предназначенных для преобразования специфической информации триплетного кода мРНК в соответствующие

аминокислоты. Адапторная или роль переносчика информации для тРНК была постулирована Криком (F. Crick) в 1957 г.

Структура тРНК. Вследствие их малого размера и стабильности тРНК были исследованы раньше и лучше других РНК. Различные тРНК состоят из 75–90 нуклеотидов, проявляя почти идентичную структуру у прокариот и эукариот. У обоих типов организмов тРНК транскрибируется как большего размера молекула-предшественник, которая посредством ферментативной обработки превращается в зрелую молекулу 4S тРНК. Например, у *E. coli,* тРНК состоит из 77 нуклеотидов, тогда как молекула-предшественник содержит 126 нуклеотидов. В 1965 г. Холи (R. Holley) с коллегами доложили о выяснении полной первичной последовательности нуклеотидов тРНК^{ала} (тРНК–переносчик аланина) у дрожжей. Оказалось, что число нуклеотидов является уникальным для тРНК. Каждая РНК является модификацией одного из азотистых оснований, используемых в РНК (G – гуанин, C – цитозин, A – аденин и U – урацил). тРНК включают, среди других, инозин, который содержит пурин гипоксантин, риботимидиловую кислоту и псевдоуридин.

Анализ первичной последовательности привел Холи к двумерной **модели** тРНК по типу **клеверного листа** (рис. 3.1.1). До этого было известно, что тРНК проявляет вторичную структуру, зависимую от сочетания пар оснований. Холи понял, что он может преобразовать линейную последовательность таким образом, что молекула РНК образует несколько вытягиваний или петель в результате смыкания пар оснований; возникающая структура похожа на лист клевера (рис. 3.1.1). Петли содержат модифицированные основания, которые в обычных парах оснований не встречаются. Поскольку модифицированный триплет GCU (GCC и GCA) определяет аланин, Холи провел поиск последовательности антикодона, комплементарного одному из этих кодонов в молекуле тРНК^{ала}. Он обнаружил его в виде CGI (в 3' → 5' направлении), в одной из петель «клеверного листа». Азотистое основание I (инозин) может формировать водородную связь с U, C или A, третьими членами триплета. Именно таким путем был распознана петля антикодона в тРНК.

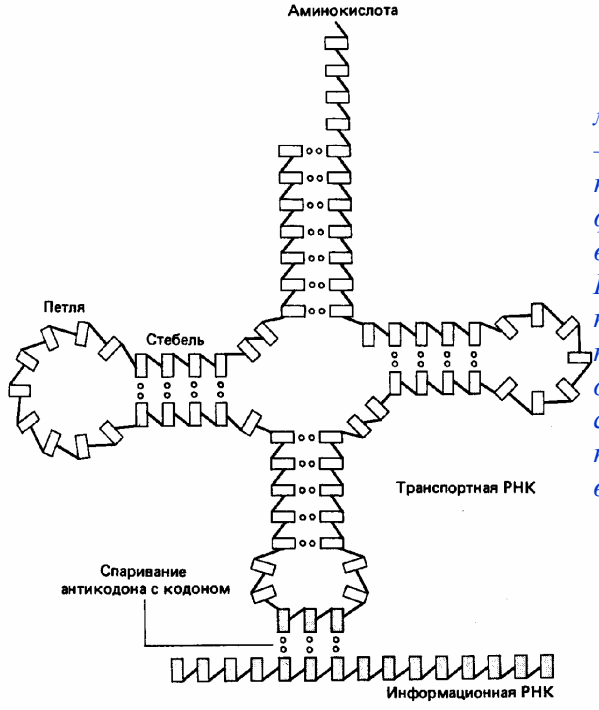


Рис. 3.1.1. Двумерная модель клеверного листа транспортной РНК (тРНК).

Представлена общая схема трансформации информации мРНК через сайт тРНК на компонент полипептида – аминокислоту. Остатки сахара изображены в виде прямоугольников, а соединяющие их линии обозначают фосфодиэфирные связи. Стебли шпилек образованы спариванием комплементарных оснований (показано кружками). На конце каждой шпильки находится одноцепочечная петля. В нижней петле расположен триплет, комплементарный кодону той аминокислоты, которую представляет данная тРНК. Этот антикодон узнает кодон путем спаривания оснований. Аминокислота присоединяется к последнему нуклеотиду 3'-конца, который изображен в верхней части рисунка (По Льюин, 1987, с.67).

Fig. 3.1.1. Two-dimensional cloverleaf model of transfer RNA.

A general scheme for adopting of mRNA information through binding of tRNA anticodon site and formation a polypeptide unit, aminoacid is outlined. Sugar residuals are shown as rectangles and lines that connect them denoted the phospho-di-ether bonds. Steam of pin is formed by pairing of complementary residuals (shown with dots). At the end of each pin there is a single-strained loop. A bottom loop contain a triplet that is complementary to the codon of an aminoacid, which represented by this tRNA. This anticodon recognize consequent codon by base pairing. Aminoacid is bonded on top

at 3'- end (From Lewin, 1987, p.67).

Рис. 3.1.2. Прокариотическая и эукариотическая рибосома.
Они состоят из двух субъединиц. Показаны три действующих сайта рибосомы – A, P, E (аминоацильный, пептидильный и выводящий).
Fig. 3.1.2. The comparison of components in procaryotic and eucariotic ribosome. Three sites of a ribosome are shown, A, P, E (Aminoacyl, Peptidyl, and Exit).

3. Рибосомы

Этот компонент представлен клеточными органоидами, на которых осуществляется синтез полипептидных молекул белков. Рибосомы двигаются вдоль молекулы мРНК, и получив соответствующую молекулу тРНК, присоединяют ее к полипептидной цепи посредством пептидной связи. Рибосомы состоят из двух отдельных рибопротеиновых частей, малой (40S, у эукариот) и большой субъединицы (60S), которые к моменту трансляции соединяются вместе, формируя зрелую рибосому (моносому), массой 80S у эукариот и 70S – у прокариот (рис. 3.1.2).

4. Аминоацил-тРНК синтетазы (А-тРНК)

Каждый фермент в этом наборе молекул катализирует присоединение определенной аминокислоты к соответствующей молекуле тРНК. А-тРНК, объединенная со своей аминокислотой, называется аминоацетиллированной тРНК или заряженной тРНК. Она поступает в специальный сайт рибосомы (рис. 3.1.2, А). После этого тРНК с аминокислотой при участии ряда других макромолекул перемещается в следующий, пептидильный сайт (рис. 3.1.2, Р), где удерживается растущий пептид. Освободившаяся после синтеза, не заряженная тРНК переходит в третий, выводящий (exit) сайт и затем покидает рибосому (рис. 3.1.2, Е).**5. Факторы инициации, элонгации и терминации**

Синтез полипептида можно подразделить на три стадии: инициация, элонгация и терминация. В каждой фазе участвуют специфические молекулы. Рассмотрим это более подробно.

3.2. ПРОЦЕСС ТРАНСЛЯЦИИ: ОТ РНК К ПОЛИПЕПТИДУ

Инициация

Инициация у эукариот начинается со сканирования мРНК и поиска инициирующего кодона. Инициация синтеза белка начинается (1) с образования стартовых комплексных форм на 5' конце мРНК. Здесь возникает кэп (5'-cap), или «шапка», и присоединение нескольких факторов инициации (IF). (2) Сформированный стартовый комплекс включает 40S рибосомную субъединицу, инициаторную $тРНК^{met}$ (met – метионин), и факторы инициации IF (IF1–IF3). (3) Стартовый комплекс присоединяет субъединицу 60S, в которой тРНК занимает Р (пептидильный) сайт. Этот комплекс двигается вдоль мРНК до прочтения первого кодона AUG, с которого начинается трансляция мРНК. Сам процесс трансляции начинается с присоединения молекулы $тРНК^{met}$ к рибосоме. Далее аминоацетилированные молекулы тРНК поступают последовательно одна за другой к рибосоме, которая транслирует мРНК.

Элонгация

Элонгация осуществляется последовательно кодон за кодоном посредством так называемого механизма трещетки (ratchet mechanism). Элонгация полипептида начинается с привлечения специального фактора EF-Tu-ГТФ (рис. 3.2.1). Элонгация состоит из трех процессов,

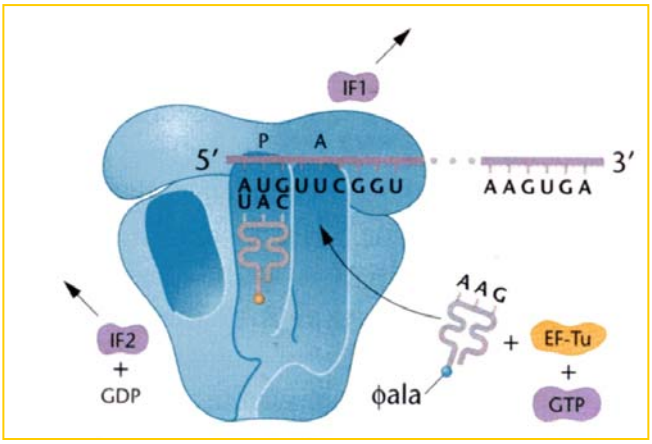


Рис. 3.2.1. Цикл элонгации белкового синтеза. Факторы инициации (IF1, IF2) освобождаются, давая возможность ассоциации двух субъединиц рибосомы. Заряженная инициаторная тРНК занимает сайт P. Пептидная связь формируется между первой и новой аминокислотой в этом же сайте A (φala). Энергетически перенос новой аминокислоты обеспечивает фактор элонгации EF-Tu. Функция фактора EF-Tu состоит также в переносе 40S субъединицы к следующему кодону. Функция другого фактора EF-G состоит в освобождении незаряженной тРНК (По Klug, Cummings, 2002, p.258, с сокращениями).
Fig. 3.2.1. Initiation of protein synthesis. (A) The initiation complex forms at the 5' end of the mRNA. (B) This consists of one 40S ribosomal subunit, the initiator tRNA^{met}, and the eIF initiation factors. (C) The initiation complex recruits a 60S ribosomal subunit in which the tRNA occupies the P (peptidyl) site of the ribosome. This complex travels along the mRNA until the first AUG is encountered, at which codon translation begins (From Klug, Cummings, 2002, p.258, with modifications).

осуществляемых повторно: 1. Привнесение новой аминоацил-тРНК в очередь. 2. Формирование новой пептидной связи для наращивания полипептида. 3. Передвижение рибосомы к следующему кодону вдоль мРНК. На растущем конце каждая поступающая тРНК с аминокислотой тестируется на соответствие и присоединяется пептидной связью посредством пептилтрансферазы к своей предшественнице: аминогруппа очередной аминокислоты соединяется с карбоксильной группой предшествующей аминокислоты. В конце, химическая связь между последней тРНК и присоединенной к ней аминокислотой разрушается и скомплектованный полипептид выводится из рибосомы.

Первую и следующие ступени элонгации иллюстрирует рис. 3.2.1.

Терминация

В сравнении с инициацией и элонгацией терминация полипептидного синтеза (фаза освобождения) является простой. Когда стоп-кодон прочтен, тРНК, находящаяся на конце полипептида, остается в P-сайте, а освобождающий фактор (*release factor, RF)* соединяется с рибосомой, гидролиз ГТФ (гуанозинтрифосфата) поставляет энергию и освобождает полипептид от тРНК, а также выводит RF и диссоциирует 80S субъединицу от мРНК. Таким образом, сигнал терминирующего кодона освобождает законченную полипептидную цепь.

Полирибосомы

Как только в процессе элонгации начальный фрагмент мРНК проходит через рибосому, эта мРНК получает возможность ассоциировать с другой малой субъединицей и формирует второй комплекс инициации. Этот процесс может повторяться несколько раз с той же самой мРНК, формируя структуру, называемую **полирибосома** или просто **полисома**. Полирибосомы можно выделить и исследовать, применяя мягкий лизис клеток. На рис. 3.2.2 полисомные комплексы видны под увеличением электронного микроскопа. На этом рисунке видна также мРНК (рис. 3.2.2, тонкая линия), располагающаяся между индивидуальными рибосомами. Видны также полипептидные цепи, выходящие из рибосомы в ходе трансляции (рис. 3.2.2, изогнутые цепи).

Формирование комплекса полисом является демонстрацией эффективного использования компонентов синтеза, имеющихся в наличии в данный момент времени.

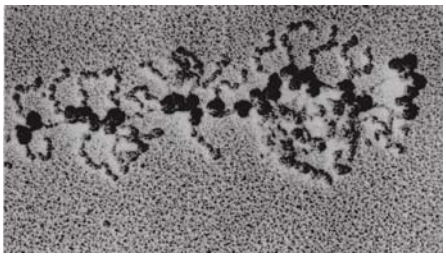


Рис. 3.2.2. Визуализация полирибосом, выделенных из гигантских клеток слюнных желез комара Chironomus thummi, в поле электронного микроскопа. Отчетливо видны полипептидные цепи, покидающие каждую рибосому. Их длина увеличивается по мере прохождения процесса трансляции слева (5') на право (3') вдоль мРНК (По Klug, Cummings, 2002, Fig. E.V. Kiseleva). Fig. 3.2.2. Polyribosomes visualized under the electron microscope from giant salivary gland cells of the midgefly, Chironomus thummi. The polypeptide chain is apparent as it emerges from each ribosome. Its length increases as translation proceeds from left (5') to right (3') along the mRNA (From Klug, Cummings, 2002, Fig.E.V. Kiseleva).

Прокариоты часто кодируют мульти-полипептидные цепи посредством одной мРНК, что давно известно как оперонный принцип работы генов. У прокариот мРНК не имеет специального посадочного кэпа (шапки) и отсутствует сканирующий механизм для поиска стартового кодона AUG. У *E. coli,* например, трансляция начинается, когда два инициирующих фактора (IF1 и IF3) взаимодействуют с 30S субъединицей в то же самое время, когда другой инициирующий фактор (IF2) соединяется со специальной инициирующей тРНК, заряженной формилметионином (тРНК^{fMet}). Эти компоненты подходят вместе и соединяются с мРНК. Присоединение происходит по водородной связи между 3' концом 16S рРНК, присутствующей в 30S субъединице, и в специальной последовательности, **рибосомно-связывающем сайте** в мРНК. Все вместе, 30S + тРНК^{fMet} + мРНК-комплекс, привлекают 50S субъединицу, в которой тРНК располагается в P-сайте и выравнивается по положению инициирующего кодона AUG. В ансамбле скомплектованной рибосомы факторы инициации диссоциируют и отходят от этого комплекса.

3.3. НАСЛЕДСТВЕННОСТЬ, БЕЛКИ И ФУНКЦИИ

Гипотеза один ген - один фермент

В двух независимых исследованиях, начатых в 1933 г., Бидл (G. Beadle) представил первые убедительные экспериментальные доказательства того, что гены непосредственно являются ответственными за синтез ферментов. Первое исследование, проведенное совместно с Эфрусси (B. Ephrussi), касалось пигментации глаза у *Drosophila*. Они выяснили, что мутантные гены, которые меняют цвет глаз у плодовой мушки, связаны с биохимическими ошибками, наиболее часто включающими потерю функции фермента. Бидл далее объединился с Татумом (E. Tatum) для исследования биохимических мутаций в пищевых цепях хлебной плесени *Neurospora crassa.* Это исследование позднее привело к гипотезе **«один ген – один фермент».**

Бидл и Татум: Мутации у нейроспоры. В начале 1940-х годов Бидл и Татум остановили свой выбор на исследовании нейроспоры, так как о ее биохимии уже многое было известно, а мутации можно было индуцировать и выделять относительно просто. Вызывая мутации, эти исследователи образовали линии, у которых биохимические реакции, критические для роста организма, были блокированы. Было известно, что нейроспора способна производить для собственной жизнедеятельности практически все необходимое. Например, она способна использовать источники атомарного углерода и азота, синтезировать витамины, 20 аминокислот, многочисленные каротиноидные пигменты, а также все главные пурины и пиримидины. Бидл и Татум облучали рентгеновскими лучами бесполые конидии (споры) для увеличения частоты мутаций и давали им расти на «полной» среде, содержащей все компоненты необходимые для роста (витамины, аминокислоты и т.д.). При таких условиях содержания мутантные линии, которые

были не в состоянии расти на минимальной среде, восстанавливали способность к росту. После этого все культуры переносились опять на минимальную среду. Если рост на минимальной среде обнаруживался, то исследователи делали вывод, что линия не содержала мутации и наоборот. Далее необходимо было определить, какого типа мутация возникла. Подобные опыты проводились и на других объектах. Принципиальная схема тестирования показана на рис. 3.3.1.

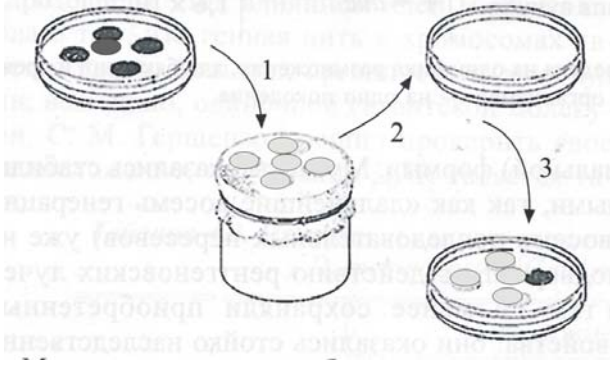


Рис. 3.3.1. Метод отпечатков для обнаружения мутантов у бактерий, устойчивых к фагу Т1. 1 – получение отпечатка колоний на материи, 2 – перенос отпечатка на среду, 3 – инкубация отпечатка. Колония бактерий с мутацией показана в нижней чашке Петри темным цветом.
Fig. 3.3.1. Stamp method applied for detection of mutations in bacteria that are resistant against T1 Fag. 1, getting "stamp" replicas from bacterial colonies. 2, stamps' transfer to medium. 3, incubation of a stamp. Bacterial colony with a mutation is shown in a bottom Petri plate with dark color.

Результаты, полученные при тестировании более 80 тыс. спор нейроспоры убедили Бидла и Татума в том, что биохимические синтетические процессы генетически детерминированы. С большой вероятностью можно было заключить, что каждая мутация, затрагивающая пищевую цепь, вызывает потерю активности фермента, который у дикого штамма нормально обеспечивал эту реакцию. Оказалось также, что мутации могут быть обнаружены практически для каждой ферментативно регулируемой реакции. На основании этого Бидл и Татум сделали заключение, что один ген детерминирует один фермент. С уточнением, сделанным в середине 1950-х годов Жакобом и Моно, о синтезе с гена одного полипептида как части работы оперона, эта концепция становится принятой в современной генетике.

Гены и ферменты: анализ биохимических цепей

Концепция «один ген – один фермент» и методы, на которых она базировалась, были использованы многократно в исследованиях метаболизма на нейроспоре, кишечной палочке и на многих других микроорганизмах. Один из первых метаболитических путей, исследованных в деталях, – это синтез аминокислоты аргинина у нейроспоры. Исследуя мутантные линии, каждая из которых требовала для роста аргинин *(arg^-),* Срб и Горовитц (A. Srb, N. Horowitz) установили точный биохимический путь, ведущий к синтезу этой молекулы. Их работа показывает, как генетический анализ может быть использован для получения биохимической информации. Срб и Горовитц протестировали способность каждой из семи мутантных линий восстанавливать рост, если цитруллин или/и орнитин, два близких химически к аргинину вещества, используются в минимальной среде для его замещения. Если любой из них был способен заместить аргинин, то предполагалось, что он должен быть вовлечен в биосинтетический путь аргинина. Эти авторы обнаружили, что обе молекулы могут быть замещены в одной или более линий нейроспоры.

Из семи мутантных линий, в четырех *(arg 4–7)* рост осуществлялся, если они получали либо цитруллин, либо орнитин, либо аргинин. В двух из линий *(arg 2* и *arg 3)* рост продолжался, если линии получали либо цитруллин, либо аргинин. Одна линия *(arg 1)* возобновляла рост только в случае получения аргинина; ни цитруллин, ни орнитин не могли заместить аргинин. На основании этих экспериментальных наблюдений был предложен следующий биохимический путь и установлены метаболитические сбои, вызываемые каждой мутацией (рис. 3.3.2):

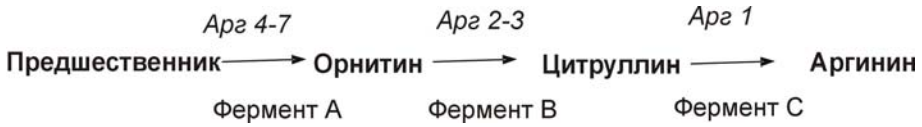


Рис. 3.3.2. Цепь химических реакций аргинина у нейроспоры (По Klug, Cummings, 2002).
Fig. 3.3.2. Chain of chemical reactions in Neurospora (From Klug, Cummings, 2002).

Один ген – одна полипептидная цепь

Описанная выше концепция «один ген – один фермент» не сразу была принята генетиками. Это не удивительно, так как не было понятно, как мутантный фермент может вызывать изменчивость фенотипических признаков. Так, у дрозофилы мутации изменяют размер и цвет глаз, жилкование крыльев и число щетинок на теле, а также многие другие признаки. Подобные фенотипические эффекты мутаций типичны для многих других видов животных и растений и они весьма разнообразны.

Однако вскоре ситуация прояснилась. Во-первых, стало ясно, что хотя все ферменты являются белками (или их комплексами с другими макромолекулами – липидами, гликозидами и т.п.), не все белки – ферменты. По мере развития биохимической генетики стало ясно, что все белки детерминируются информацией, хранящейся в генах, что привело к более точной терминологии: один ген – один белок. Во-вторых, белки часто обнаруживают четвертичную структуру, представленную конформационными объединениями двух или более полипептидных цепей. Это составляет основу четвертичной структуры белков, которая будет рассмотрена позже в данной лекции, а также в Лекции 5. Поскольку полипептидные цепи определенного белка кодируется отдельным геном, более точное определение появилось и для генетической концепции: «один ген – одна полипептидная цепь». Это стало очевидным после работ Жакоба и Моно, установивших оперонный принцип организации генов прокариот, а также исследования структуры гемоглобина у людей, пораженных анемией.

Серповидноклеточная анемия. Наиболее серьезные доказательства того, что гены определяют синтез белков, были получены при исследовании мутаций, затрагивающих молекулу гемоглобина у людей и вызывающих заболевание Серповидноклеточная анемия. Люди, пораженные этим недугом, обладают эритроцитами, которые при низкой концентрации кислорода становятся вытянутыми в дугу, наподобие серпа, из-за полимеризации гемоглобина. Эти эритроциты резко отличаются от нормальных и вызывают тяжелую анемию (рис. 3.3.3).

a. **b.**

Рис. 3.3.3. Эритроциты нормальных людей (a) и страдающих серповидноклеточной анемией (b) (По Klug, Cummings, 2002).
Fig. 3.3.3. A comparison of erythrocytes from normal individuals (a) and from individuals with sickle-cell anemia (b) (From Klug, Cummings, 2002).

Еще в 1949 г. Нил и Бит (J. Neel, E. Beet) показали, что эта аномалия наследуется в соответствии с Менделевскими законами. Проведенный анализ родословных выявил три фенотипа и три соответствующих генотипа, контролируемых парой аллелей Hb^A и Hb^S. Эритроциты гетерозиготных особей являются нормальными. Поэтому, являясь носителями дефектного гена, эти люди не страдают анемией, так как половина их гемоглобиновых молекул нормальна. Более того, оказалось, что у таких людей повышен иммунитет к малярии. Такие гетерозиготы, являясь обладателями дефектного гена, тем не менее, передают его согласно законам расщепления 50% их потомков. В этом же 1949 г. Полинг (L. Pauling) с соавторами представили серьезное

доказательство молекулярной основы анемии. Они показали, что аномальный и нормальный гемоглобин отличаются по подвижности при электрофорезе. При этой методике (см. подробнее в Лекции 5) заряженные молекулы мигрируют в электрическом поле в геле. Если заряд молекул различается, суммарная подвижность молекул также отличается. На основе этого было показано, что между нормальными и дефектными гемоглобиновыми молекулами имеются химические различия. Эти молекулы обозначают с той поры **Hb-A** и **Hb-S,** соответственно.

Электрофоретическое исследование, проведенное Полингом, оставило вопрос, к какой части молекулы гемоглобина, к железосодержащему гему или к глобину, состоящему из четырех-субъединиц, относятся найденные различия по подвижности у нормальных людей и людей с анемией. Этот вопрос удалось разрешить позже, в период между 1954-1957 гг. Ингрэму (V. Ingram), обнаружевшему пептидые различия и различия в аминокислотном составе этих гемоглобинов и показавшим, таким образом, что первичная структура белка ответственна за аномалию. Используя фингерпринт-анализ, Ингрэм показал, что Hb-S отличается по набору пептидов и по аминокислотному составу по сравнению с Hb-A. Гемоглобин человека содержит две идентичные полипептидные α цепи, состоящие из 141 аминокислоты, и две идентичные β цепи, представленные 146 аминокислотами, которые составляют его четвертичную конформационную структуру. Дальнейший анализ выявил единичную аминокислотную замену, а именно – валин был замещен глутаминовой кислотой в шестом положении β цепи, вызывая соответствующие различия пептидов.

Структура белков и биологическое разнообразие. Разнообразие белков впечатляет. Каким образом, и какие структурные особенности этих макромолекул обеспечивают это разнообразие и многочисленные потребности клеток, тканей и в целом организм? Для начала необходимо провести разграничение между полипептидами и белками.

Полипептиды – это предшественники белков. Прежде, чем функциональные молекулы начинают работать, реализуются несколько шагов их аггрегации, транспортировки и модификации. Полипептидные цепи белков – это линейные неветвящиеся полимеры (такими же полимерами являются РНК и ДНК молекулы). Имеется 20 аминокислот, которые выступают как мономеры в полипептидных цепях. Каждая аминокислота имеет карбоксильную группу, аминогруппу и боковые радикалы или **R** –группы, которые присоединены ковалентно к центральному атому углерода (рис. 3.3.4). **R** –группы дают каждой аминокислоте ее химическую уникальность. Последовательность аминокислот и комбинация **R** –групп определяют уникальность первичной структуры полипептида.

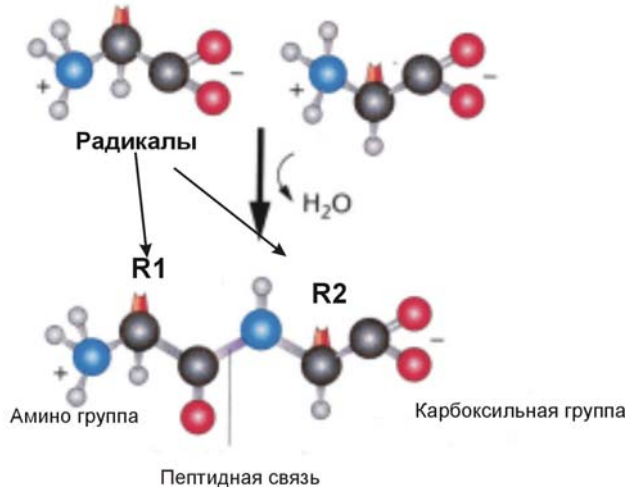


Рис. 3.3.4. Формирование пептидной связи между двумя аминокислотами.

Голубым цветом показаны атомы азота, серым – углерода (большие шарики) и водорода (малые шарики) и красным – кислорода (По Klug, Cummings, 2002, с добавлениями).

Fig. 3.3.4. Peptide bond formation between two amino acids.

In blue color the nitrogen, in gray – carbon (large balls) or hydrogen (small balls), and in red – oxygen atoms are shown (From Klug, Cummings, 2002, with adds).

С образованием пептидной связи, возникающей между атомами углерода и азота в цепи, начинается синтез полипептида (рис. 3.3.4). В дальнейшем на основе пептидов формируется большое разнообразие конфигураций, дающие

четыре основных свойства полипептидов: (1) неполярность (гидрофобность), (2) полярность (гидрофильность), (3) отрицательный заряд, и (4) положительный заряд. Поскольку полипептиды – это часто длинные полимеры и так как любое положение в цепи может быть занято какой-либо из 20 аминокислот, то число комбинаций при образовании этих полимеров очень велико. Разнообразие полимеров затем преобразуется в практически неограниченное разнообразие структур и функций. К примеру, если полипептид состоит из 250 аминокислот (средний белок), то возможное число комбинаций из 20 аминокислотных блоков составит $\mathbf{20^{250}}$.

В белковой молекуле можно распознать четыре структурных уровня: первичная, вторичная, третичная и четвертичная структура. **Первичная структура** формируется как последовательность аминокислот в цепи полипептида. Она определяется первичной последовательностью нуклеотидов в ДНК через мРНК-посредника. Такое *однозначное соответствие последовательностей ДНК и аминокислот называется* ***колинеарностью***. Первичная структура определяет остальные особенности конформации («упаковки») молекулы белка при его формировании, то есть, - его пространственную конфигурацию. **Вторичная структура** слагается при образовании химических связей между близко расположенными аминокислотами в полипептидной цепи – это так называемая α-спираль (рис.3.3.5, А), которая была предсказана исходя из теоретических соображений Л. Полингом и Р. Кори (R. Corey). Предсказана также конформация в виде «гармошки», называемая часто β-упаковкой (рис. 3.3.5,В). Большинство белков сочетают оба варианта упаковки, определяющие вторичную структуру. Согласно модели полипептидная цепь «замыкается» на себя в параллельные или антипараллельные структуры (в такой конформации могут участвовать и другие цепи). Каждая подобная структура стабилизируется водородными связями, формируемыми между соседними цепями (рис.3.3.5, В).

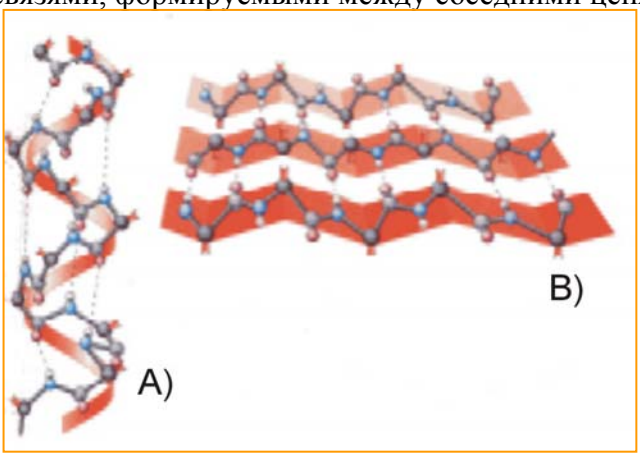


Рис. 3.3.5. (А) Правозакрученная α-спираль, которая представляет один из вариантов вторичной структуры, (В) β –листовая конформация или конформация типа «гармошки», - альтернативный вариант вторичной структуры полипептида. Для простоты показаны не все атомы. Основные группы атомов цепи показаны на предыдущем рисунке (По Klug, Cummings, 2002, с сокращениями).

Fig. 3.3.5. (a) The right-handed α-helix, which represents one form of secondary structure of a polypeptide chain, (b) The β-pleated-sheet configuration, an alternative form of secondary structure of polypeptide chains. For the sake of clarity, not all atoms are shown (From Klug, Cummings, 2002).

Третичная структура. Она определяется трехмерной конформацией всей полипептидной цепи в пространстве (рис. 3.3.6). Каждый белок поворачивается и формирует петли характерным только для него способом. Описанные три уровня имеют решающее значение в конформации и стабилизации молекулы белка. **Четвертичная структура.** Многие белки имеют в составе более чем одну полипептидную цепь или субъединицу. Упаковка этих субъединиц по отношению друг к другу и составляет четвертичную структуру. Белки, состоящие из нескольких субъединиц, называют олигомерными, а каждую отдельную субъединицу называют протомером. Субъединицы

белков образуют комплиментарные друг к другу структуры; у ферментов часто только после образования четвертичной структуры формируется полноценный активный (каталитический) центр. Формирование активной молекулы белка часто сопровождается включением атомов металла (гема), связывающих, например, кислород (рис. 3.3.6 – 3.3.7). Четвертичная структура гемоглобина слагается четырьмя олигомерами или единичными полипептидными цепями (рис. 3.3.8). Большинство ферментов – оксидоредуктазы, гидролазы, лигазы, трансферазы, включая такие полимеразы, как ДНК- и РНК-полимераза, являются олигомерами.

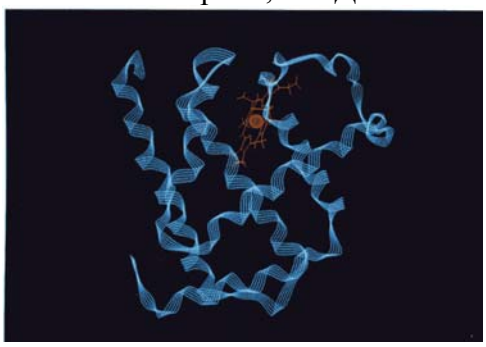


Рис. 3.3.6. Третичная структура белка на примере миоглобина. Присоединенный атом кислорода показан красным цветом.
Fig. 3.3.6. The tertiary level of protein structure in respiratory pigment myoglobin. The bound oxygen atom is shown in red.

Рис. 3.3.7. Четвертичная структура белка на примере гемоглобина. Четыре цепи (α-, β-цепи – светлые и темные) взаимодействуют с гемом (дисковидные структуры), формируя активную молекулу (Оба рисунка - из Klug, Cummings, 2002).
Fig. 3.3.7. The quaternary level of protein structure as seen in hemoglobin. Four chains (2α and 2 β) interact with four heme groups to form the functional molecule (Both pictures are from Klug, Cummings, 2002).

Посттрансляционные модификации белков

Важным достижение современной биохимии явилось понимание того, что большинство макромолекул подвергаются модификации после из синтеза. Это относится как к РНК-транскриптам, так и к полипептидным цепям. Эта дополнительная обработка белков называется **посттрансляционной модификацией**. Хотя детальное освещение этих модификаций находится вне целей данного курса лекций, представляется важным ознакомиться с принципиально важными событиями, сопровождающими конечную фазу формирования белка, как генного продукта. Ниже будут представлены несколько примеров посттрансляционных модификаций.

Посттрансляционная модификация очевидно важна в достижении функционального состояния тем или иным белком. Поскольку третичная и четвертичная пространственные структуры молекулы белка очень тонко связаны с его специфической функцией, вопрос, как полипептидные цепи образуют пространственную конфигурацию, является очень важным. Многие годы считалось, что конформация молекулы белка определяется спонтанно, по принципу максимизации ее термодинамической стабильности, базирующейся, в основном, на комбинации химических свойств аминокислот в их первичной последовательности в полипептидных цепях, составляющих данных белок. В противоположность этому, многочисленные недавние исследования показали, что конформация зависит от группы других повсеместно встречающихся белков, получивших название **шаперонов** (иногда также называемых шаперонинами или молекулярными шаперонами). Функция шаперонов – это способствование конформации других белков.

Функционирование белка

Для клеточных организмов, представляющих наиболее продвинутые формы жизни на Земле, одной из важнейших задач является синтез и функционирование белков. Можно даже сказать, что *самовоспроизводимая реализация информации, хранимой в ДНК (РНК), в*

многообразие белков, регулируемо функционирующих в клетке или организме – это и есть жизнь в основном своем проявлении. С другой стороны, многообразие функций тканей и органов, формируемых у высших организмов, зависит в подавляющем числе случаев от разнообразия функций белков и их свойств.

Белки – это наиболее многочисленные макромолекулы клетки, роль которых весьма многообразна. Так, обеспечение дыхания и транспорт кислорода осуществляют, как упоминалось, миоглобин и гемоглобин. Сократительные функции обеспечивают актин и миозин, присутствующие в больших количествах в мышечной ткани, но присутствующие также во всех делящихся клетках, обеспечивая внутриклеточные сократительные процессы, типа движения хромосом к полюсам в митозе и мейозе. Покровные функции, волосяной покров, соединительные ткани обеспечиваются кератином и коллагеном, специализированными структурными белками. Иммуноглобулины – это другая группа белков, обеспечивающих иммунитет (см. Глава 4). Структурную функцию в процессе компактизации - декомпактизации хромосом выполняют гистоны, консервативные белки с основными свойствами. Перечень этот может быть весьма обширным, обеспечивая практически все многообразие структур и функций живой материи.

Наибольшую группу белков, обеспечивающих многочисленные каталитические биохимические реакции, представляют **ферменты.** Они представлены четырьмя упомянутыми ранее группами. В Лекции 5 мы еще обратимся к ферментам с позиций биохимической генетики. Здесь представляется важным рассмотреть некоторые их особенности в более широком аспекте. Фермент увеличивает скорость, с которой биохимическая реакция достигает равновесного, асимптотического уровня. Однако они не изменяют конечную точку химического равновесия. Биологический катализ – это процесс, посредством которого, энергия активации для данной химической реакции снижается. Поскольку ферментативные реакции асимптотические, то очень важными являются соблюдение условий оптимума реакций – температурного, pH, концентрации субстрата. Превращение субстрата в продукт – это обратимая реакция; часто ее преобладающее направление зависит от их концентрации. Субстратами ферментативных реакций выступают многочисленные макромолекулы, такие как аминокислоты, витамины, сахара, спирты и др. Поэтому, во многих случаях за счет ферментов легко осуществляется авторегуляция внутриклеточного метаболизма.

Каталитические свойства и специфичность ферментативной реакции определяются химической конфигурацией **активного центра** молекулы. Он ассоциирован с впадиной, или складкой, на поверхности фермента. Активный центр связывает субстрат и облегчает его взаимодействие с другими реагентами. Все внутриклеточные метаболические активности так или иначе контролируются ферментами. Они подразделяются на две группы – анаболические и катаболические. Посредством анаболических реакций осуществляются реакции химического синтеза из простых компонентов в более сложные, такие как белки, нуклеиновые кислоты, липиды. Катаболические реакции осуществляют деградацию больших молекул в молекулы меньшего размера с высвобождением химической энергии.

Доменная структура белков и перетасовка экзонов

Выше мы отметили, что активный центр фермента формируется как некоторая складчатость поверхности, связанная со спецификой взаимодействия определенных последовательностей его молекулы на четырех структурных уровнях. Не вся молекула белка существенна для такой специфичной складчатости, обычно от 50 до 300 аминокислот влияют на конфигурацию молекулы, определяя ее активные центры, которые называют также **белковыми доменами**. У белков могут иметься один, два или несколько доменов, формируемых определенными модулями и определяющими уникальную, специфическую конфигурацию данного белка. Различные белковые домены обладают специфичными функциональными способностями.

Доменная и четвертичная структуры белков взаимосвязаны. Каждый домен – это смесь вторичных структур, включая α–спираль и β–гармошку. Уникальная, специфическая

конфигурация, которая складывается в отдельном домене, определяет его характерную функцию. Так, домен может служить основой для формирования каталитического или активного центра фермента, позволять специфическое соединение с мембраной или, наконец, соединение с другой макромолекулой (липидом, полисахаридом и т.п.). Поэтому типично употребление таких терминов как каталитические домены, ДНК-связывающие домены и другие. Главное в этом то, что белок должен рассматриваться как набор структурно и функционально связанных блоков. Ясно, что присутствие множественных доменов увеличивает универсальность белка и снижает энергетические затраты метаболизма. Гилберт (Gilbert, 1978) с позиций генетики предложил идею возникновения доменов за счет перетасовки экзонов, что дает основу для формирования новых доменов.

Отношения между экзонами могут быть достаточно сложными. Так, например, для низкоплотностного липопротеина имеется 18 экзонов, определяющих пять доменов и сигнальную последовательность (рис. 3.3.8). Первый экзон контролирует сигнальную последовательность, которая удаляется из белка прежде, чем низкоплотностный липопротеин станет частью мембраны. Следующие пять экзонов представляют домен, определяющий связывание холестерина. Этот домен состоит из повторяющейся 7 раз последовательности 40 аминокислот. Следующий домен детерминируется последовательностью из 400 аминокислот, обладающей сильной гомологией с пептидным гормоном мыши, епидермальным ростовым фактором (EGF, epidermal growth factor). Этот регион кодируется восьмью экзонами и содержит три повторяющихся последовательности по 40 аминокислот. Сходная последовательность обнаружена также у трех белков свертывания крови. Пятнадцатый экзон определяет домен, специализирующийся на пострансляционном добавлении карбо-гидрата. Остающиеся два экзона контролируют регионы белка, которые являются частью мембраны, привязывающей рецептор к специфическим сайтам, называемым «липкими» или «магнитными» точками клеточной поверхностной мембраны.

Экспериментальные данные, связанные с низкоплотностным липопротеином, весьма хорошо укладываются в рамки «**гипотезы перетасовки экзонов**».

Рис. 3.3.8. Сопоставление 18 экзонов, составляющих генетическую последовательность, кодирующую рецепторный белок LDL (низкоплотностный липопротеин; Low Density Lipoprotein). Экзоны собраны в пять функциональных доменов и сигнальную последовательность (По Klug, Cummings, 2002)
Fig. 3.3.8. A comparison of 18 exons making up the gene sequence encoding the LDL receptor protein. The exons are organized into five functional domains and one signal sequence (From Klug, Cummings, 2002).

3.4. ВВЕДЕНИЕ В ПРОТЕОМИКУ

Протеома – это набор белков, экспрессирующихся с генов и модифицирущихся в течение

жизни клетки. В узком смысле термин описывает также набор ферментов, имеющихся в клетке в данный момент времени. **Протеомика** – это наука, занимающаяся исследованием протеомы. Она использует множество методов молекулярной биологии, начиная от генетического анализа и заканчивая масс-спектрометрией полипептидов. После реализации геномного проекта основное внимание сконцентрировали на вопросах изучения структуры и функций белков. Геномный проект показал, что множество вновь открытых генов имеют неизвестные функции, а многие описанные функции являются предположительными, базирующимися на аналогии с другими организмами. Например, у двух подробно исследованных организмов, у дрожжей *S. cerevisiae* и кишечной палочки *E. coli,* для более чем половины генома неизвестно, какие функции, детерминируют открытые гены. Также неизвестна функция для более 40% белков, предсказанных в рамках геномного проекта человека (см. Глава 15, рис. 15.1.1).

Понимание функции требует не только идентификации продуктов генов, а гораздо большего. Из материала, изложенного выше видно, что многие полипептиды прежде, чем стать белком модифицируются добавлением метильных, ацетильных, фосфатных и других групп, из них вырезаются сигнальные и не функционирующие в данной ткани участки, а также - инициаторные метиониновые остатки, и, наконец, полипептиды комплексуются с сахарами, липидами и другими полипептидами. Белки саморегулируемо подвергаются конформационной «сборке», а также подвергаются внешнему регулированию шаперонами. Известен процесс процессинга внутренних аминокислот первичной последовательности в цепи полипептида. Известно также свыше тысячи механизмов пострансляционных модификаций белков вдобавок к высокому разнообразию, поставляемому альтернативным сплайсингом мРНК (см. Глава 5). Таким образом, геном человека, например, имеющий 35000-40000 генов, может производить более чем пол миллиона генных продуктов. Задачи протеомики – обширны. Это – получение информации о структурах, функциях и основных модификациях, происходящих с белками. Выяснение клеточной локализации белков, оценка их изменчивости и взаимосвязей с другими белками.

Методы протеомики

Протеомика использует большинство методов молекулярной биологии, развитых при реализации геномных проектов. Это – такие методы как ПЦР (полимеразной цепной реакции), применение различных векторов, автоматическое секвенирование ДНК и специализированные пакеты программ. Методы протеомики продолжают развиваться и усовершенствоваться, представляя одно из ведущих направлений генетики. Главные методы протеомики включают разделение и идентификацию белков, выделенных из клеток. Одним из наиболее распространенных методов разделения белков является двумерный электрофорез (рис. 3.4.1). По этой методике белки, выделенные из клеток различных тканей, помещаются в полиакриламидный гель (ПАГ) и подвергаются воздействию поля электрического тока. На первом этапе белки подвергаются разделению в соответствии с зарядом, коррелирующим с их молекулярным весом. На втором этапе гель поворачивают на 90 градусов и на этот раз белки разделяются по их молекулярному размеру, используя эффект ПАГ, как молекулярного сита.

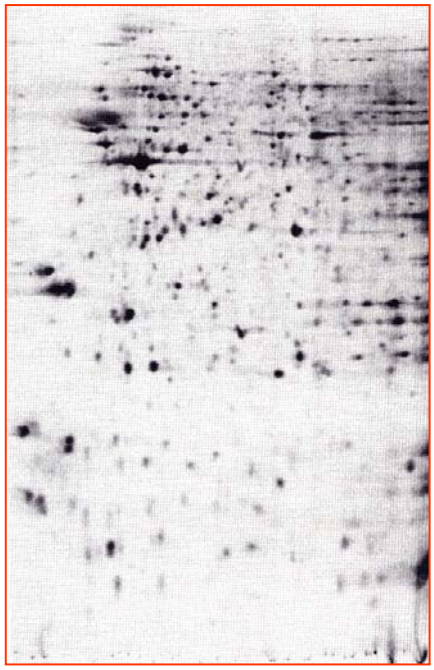


Рис. 3.4.1. Двумерный белковый гель электрофорез.

Полученные результаты показывают пятна-фракции расположения различающихся по молекулярной массе и размеру белков после специфического их окрашивания. На таком геле может быть выявлено до нескольких тысяч белков (По Klug, Cummings, 2002).

Fig. 3.4.1. A two-dimensional protein gel-electrophoresis results, showing the separated proteins as spots. Several thousand proteins can be displayed on such gels (From Klug, Cummings, 2002).

После окрашивания белков они выглядят как пятна. Обычно на геле можно насчитать от 200 до 10000 таких пятен или фракций (рис. 3.4.1). Для идентификации отдельных белков, фракции вырезаются из геля и белок обрабатывается и «переваривается», например, таким ферментом, как трипсин. Это дает серию характерных пептидных фрагментов. Фрагменты далее подвергаются масс-спектрометрическому анализу по методике, называемой пептидный масс-фингерпринтинг. Для идентификации белков, массы пептидов сравниваются с предсказанными массами и использованием белковых баз данных. Для увеличения скорости и точности анализа протеомы при ее исследовании широко применяются различные пакеты программ по ДНК-анализу и другие статистические средства.

Бактериальная протеома

Наиболее подходящими для анализа являются простые организмы, такие как *Mycoplasma genitalium,* с ее геномом порядка 480 нп (нуклеотидных пар; см. также данные в Главе 15). Несколько лабораторий изучали *M. genitalium* с целью установить минимальное число биохимических и метаболических реакций, необходимых для функционирования живого организма. В частности, Васинджер (V. Wasinger) с ее коллегами изучали экспоненциальную и пост-экспоненциальную фазы роста у *M. genitalium.*

Используя двумерный электрофорез они обнаружили всего 427 белковых фракций в первой фазе. Из них 201 были проанализированы и идентифицированы по методике пептидного масс-фингерпринта и сравнения с известными белками. Анализ позволил открыть 158 известных белков (33% протеомы) и 17 неизвестных белков. Остальные фракции фингерпринта включали фрагменты от больших белков, различные изоформы одних и тех же белков и посттрансляционные модификации белковых продуктов. Идентифицированные белки оказались ферментами, вовлеченными в энергетический метаболизм, репликацию ДНК, транскрипцию, трансляцию и транспорт различных молекул через клеточную мембрану. Переход от фазы экспоненциального роста к стационарной фазе привел к сокращению на 42% числа синтезируемых белков. В течение самого этого перехода появились некоторые новые белки, а другие подверглись драматическим изменениям в количестве. Эти изменения, очевидно, связаны с изменением питания, увеличением кислотности среды роста и другими адаптациями к изменениям среды.

Исследование Васинджер показало, что вряд ли могут быть найдены условия для реализации всего потенциала генома. В исследовании *M. genitalium* обнаружилось, что только 33% генов экспрессируются даже в условиях оптимизированых для максимального роста бактериальной культуры. С учетом технических ограничений эта величина представляет наиболее сильно экспрессируемые белки. Остающиеся 67 % белков очевидно экспрессируются в спектре

других средовых условий, а также находятся ниже порога чувствительности методики электрофореза, а также методов экстракции белков. Вопреки этим ограничениям анализ протеомы дает спектр информации недоступной при одном лишь секвенировании генома.

3.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3

1. Интернет-поддержка. Рассмотрите один пример синтеза белка на сайте Иллинойского университета США: http://www.gene.com.ac/AB/GG/protein_synthesis.html.

ГЛАВА 4. ИММУНОГЕНЕТИЧЕСКИЙ АНАЛИЗ И ЭВОЛЮЦИОННЫЕ РАССТОЯНИЯ

ГЛАВНЫЕ ВОПРОСЫ:

4.1. Введение в иммуногенетику.
4.2. Наследование иммуногенетических признаков и популяционная изменчивость.
4.3. Иммуногенетическое датирование эволюции.
4.4. Датирование эволюции посредством анализа первичной последовательности аминокислот.
4.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4.

РЕЗЮМЕ

1. Наилучшей методикой для обнаружения иммуногенетического полиморфизма является тестирование образцов крови антисывороткой.
2. Иммуногенетические признаки наследуются по законам Менделя. В природных популяциях по ним, как правило, выполняется закон Харди-Ванберга.
3. Полиморфизм иммуногенетических признаков был обнаружен у более, чем ста проанализированных видов. Поэтому использование иммуногенетических признаков сделало реальным генетические исследования на природных популяциях.
4. Популяции рыб и других организмов отличаются по частотам аллелей, кодирующих иммуногенетические признаки и иногда имеют даже уникальные аллели.
5. Иммуногенетическое датирование эволюции дало реалистичные результаты для многих групп организмов, как позвоночных, так и беспозвоночных животных и помогло улучшить филогению этих групп.
6. Оценка числа аминокислотных замещений с использованием закона Пуассона дает вполне точные эмпирические результаты, когда анализируемые отрезки времени невелики. В другом случае необходимо использовать метод Матрицы аминокислотных замещений. Полученные данные показывают реалистичность молекулярных часов.

SUMMARY

1. The best technique for immunogenetic polymorphism detection is a probation with antisera.
2. Inheritance of immunogenetic traits is simple Mendelian. Hardy-Weinberg (H-W) law is holding as a rule in natural populations.
3. Polymorphism of immunogenetic traits was found in more than 100 fish species. Thus, introduction of immunogenetic approach made it possible to conduct experimental population genetic research in the wild nature.
4. More than two alleles were found in many blood groups loci. So, the existence of multiple alleles and wide genetic variation were proved for natural populations. Fish populations differ in allele frequencies, and sometimes even have unique alleles.
5. Immunogenetic dating provides reliable results for many groups of animals, both vertebrate and invertebrate and helps to improve phylogeny in these groups.
6. Estimation of a number of amino acid substitutions through Poison theory gives quite accurate empirical results when the considered evolutionary time period is relatively short. In opposite case of long history the Amino Acid Substitution Matrix method is used. The obtained data provide the evidence on the reliability of the Molecular Clocks.

4.1. ВВЕДЕНИЕ В ИММУНОГЕНЕТИКУ

Все позвоночные животные имеют развитую иммунную систему. У рыб и тем более беспозвоночных животных она не такая совершенная, как у млекопитающих, но, тем не менее, достаточно развита, чтобы осуществлять реакции антиген-антитело, которые являются принципиальными для иммуногенетического анализа и для иммуногенетики в целом.

Иммунная реакция может осуществляться в ткани или в кровеносной системе. В последнем случае, так называемого гуморального ответа, вырабатываются антитела (представленные лабильными белками-иммуноглобулинами: α, β, γ-глобулинами), которые поступают в сыворотку, где в кровеносной системе и осуществляется иммунная реакция. Антитела развиваются против различных «чужеродных» источников или их частей, называемых антигенами. В качестве антигенов могут выступать любые макромолекулы (пептиды, липопротеины, гликопротеины), вирусы и бактерии. Иногда даже клетки собственного организма могут служить как антигены, что создает аутоиммунные заболевания, такие как ревматоидный артрит (в США от этой болезни страдают 6,5 млн. человек), юношеский диабет (страдает 0,5 % популяции человека) и, наконец, рассеянный склероз (аутоиммунная реакция мозга). В этом разделе мы рассмотрим два рисунка, которые призваны напомнить, как осуществляется реакция антиген-антитело. Для более детального ознакомления с этой проблематикой можно обратиться к монографии Поповой (1999) или другим источникам.

Иммунитет. Гуморальный иммунитет обеспечивается В-клетками и антителами, которые они производят (рис. 4.1.1). Тканевой или клеточный иммунитет обеспечивается Т-клетками. Реакции, запускаемые и осуществляемые антителами, защищают организм от бактерий и вирусов. Клеточный иммунитет обеспечивает функционирование тканей, борясь с клетками, инфицированными бактериями и вирусами, инвазированными грибами и простейшими, а также нейтрализуя клетки, пораженные раком. В ходе осуществления гуморального иммунитета происходит взаимодействие В-клеток с антигеном и последующая дифференцировка В-клеток в плазматические клетки, синтезирующие антитела. Секретируемые антитела связывают антигены и облегчают их удаление из организма. Система клеточного иммунитета включает различные субпопуляции Т-клеток, которые опознают антиген, связанный с белком МНС класса I. Тн-клетки отвечают на появление антигена продукцией цитокинов. Тс-клетки отвечают на антиген дифференцировкой в цитотоксические Т-лимфоциты (CTL), которые обеспечивают убийство измененной антигеном клетки, например, клетки, инфицированной вирусами.

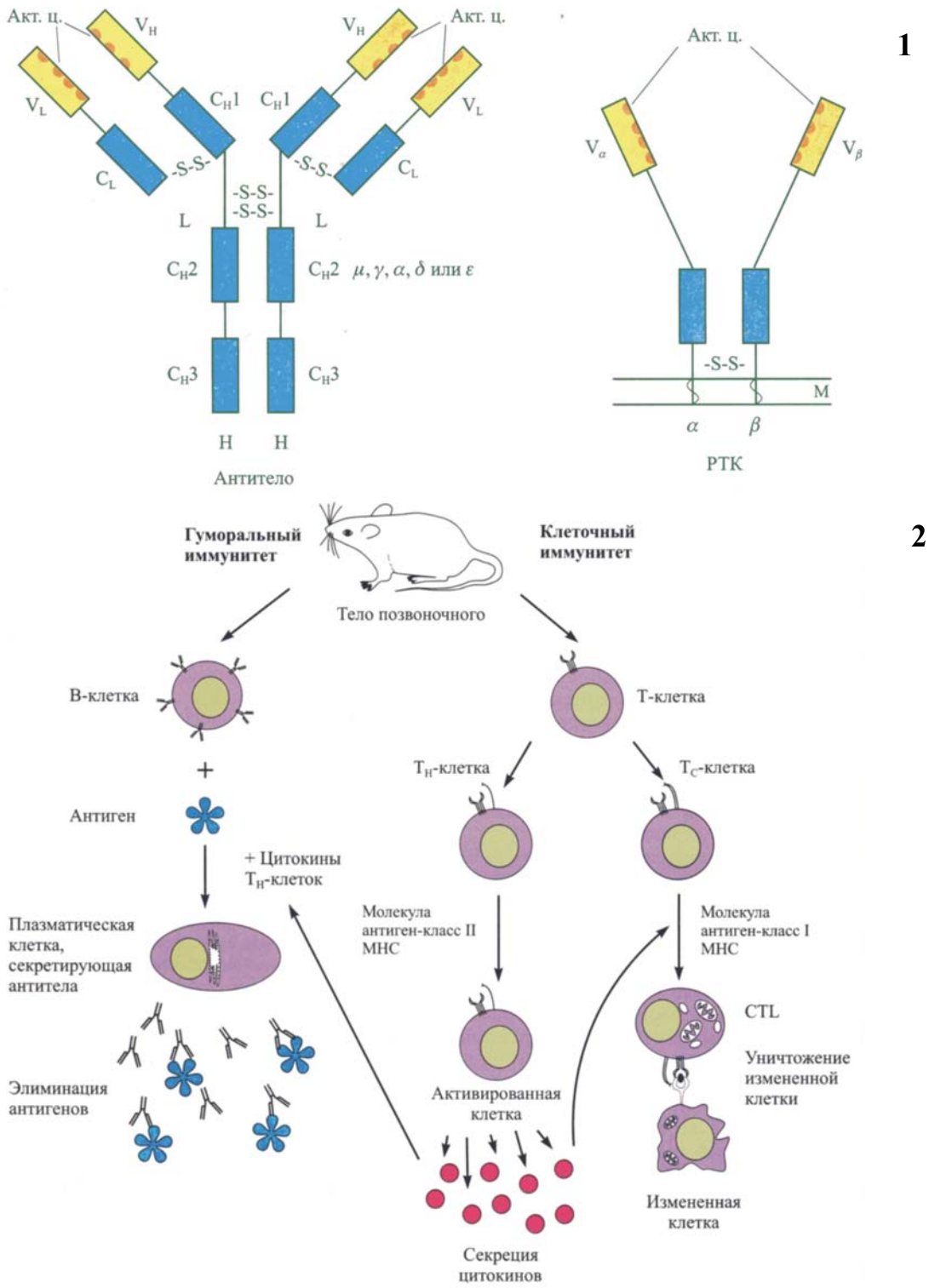
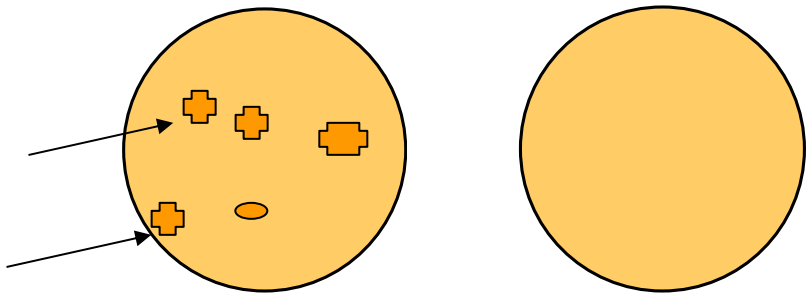


Рис. 4.1.1. Строение антитела (1) и схема гуморального и клеточно-тканевого иммунитета (2). Акт ц. – активный центр, МНС – главный комплекс гистосовместимости (По Kuby, 1994, p.13; Абелев, 1996; заимствовано из Жимулев, 2002, с.391, с изменениями).

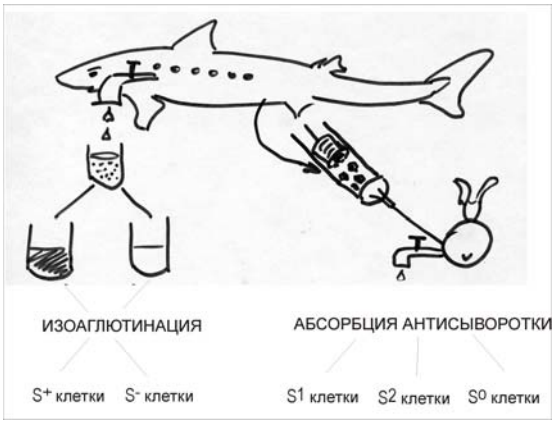
Fig. 4.1.1. Composition of an antibody (1) and the humoral and cell-mediated immune responses (2). MHC is Main Histo-compatibility Complex (From Kuby, 1994, p.13; Abelev, 1996; From Zhimulev, 2002, p.391, with modiifications).

Антитела связывают антигены, взаимодействуя посредством вариабельного участка иммуноглобулиновой молекулы (рис. 4.1.1, 1). Обе реакции и гуморальная (a), и клеточно-тканевая (b) могут реализоваться одновременно (рис. 4.1.1). В конце реакции, в эффекторной фазе, продукты взаимодействия антиген-антитело инактивируются (рис. 4.1.1, 2).

Если антигены локализуются на поверхности эритроцитов, то антитела реагируют с ними, приводя к их склеиванию и коагуляции крови. Такая кровь в виде сгустков хорошо отличается от нормальной, где иммунной реакции нет, как это показано на примере акулы (рис. 4.1.2).



A



B

Рис. 4.1.2. Определение S-системы групп крови у акулы Squalus acanthias. ***A*** *– Реакция агглютинации имеется, показывая реакцию антиген-антитело (агрегации эритроцитов показаны стрелками, слева), некоторые особи такой реакции не проявляют (справа),* ***B*** *– Показаны различные этапы анализа при простом и более сложном подходе (По Алтухов, 1974, с изменениями).*

Fig. 4.1.2. A detection of the S-system of the blood group in the shark Squalus acanthias. ***A,*** *There is an immune reaction of the agglutination (Blood cell aggregations are shown by arrows, left) and no reaction (right),* ***B,*** *Below different steps of the analysis are shown (From Altukhov, 1974, with modifications).*

Следующий шаг иммуногенетического анализа состоит в поиске источника (сыворотки), которая наиболее специфично реагирует с естественными антигенами эритроцитов. В частности, проведение реакции изоаглютинации (кровь тестировалась в сыворотке особей того же вида), позволяет обнаружить у акулы два типа реакции организмов (два фенотипа): S^+ и S^- (рис. 4.1.2, B, показано слева). Более сложная процедура, с предварительной иммунизацией кролика, повышает чувствительность сыворотки и позволяет выявить S^1-, S^2-, S^0-клетки или три фенотипа (рис. 4.1.2, B, справа). Как это обычно принято в генетике, теперь мы можем предложить (выдвинуть гипотезу) схему генного контроля S-системы группы крови акулы, и найти доказательства верности этого предположения. На первых этапах проводят тесты на воспроизводимость фенотипов, а затем проводят скрещивания и анализ расщепления в потомстве. Для диких популяций чаще всего используют данные естественно осуществляющихся в природе скрещиваний.

Чтобы подвести итоги данного раздела, необходимо определить, что такое **нормальная** и **иммунная сыворотка.** Нормальная сыворотка – это обычная сыворотка крови. *Иммунная сыворотка – это сыворотка, которая выработана организмом после введения в его кровяное русло источника антигена.* На рис. 4.1.2 были рассмотрены две реакции с нормальной и иммунной

сыворотками. Как правило, иммунные сыворотки показывают большую чувствительность (см. рис. 4.1.2, В). Иммунные сыворотки называют также **иммунными антисыворотками**. Чтобы еще более увеличить чувствительность антисыворотки, она может быть абсорбирована, однажды или повторно, с использованием крови одного и того же организма или различных организмов. Это дает возможность выявить антигены с небольшими иммунными эффектами. Агглютинация и наблюдение склеивания эритроцитов в микроскоп – не единственный путь выявления иммунной реакции. Преципитация антител – другая методика иммуногенетики. Антитела, выработанные в сыворотке какого-либо организма, могут быть осаждены или прецитипетированы, а затем подвергнуты исследованию, например, посредством иммуно-электрофореза. Заметим, что иногда в качестве заменителей сыворотки используют растворы лектинов, веществ, выделяемых из бобовых растений.

Группа крови АВ0. Простой случай сильной сывороточной реакции – это реакция на естественные антигены эритроцитов крови человека (рис. 4.1.3). АВ0 группа крови человека открыта К. Ландштейном (K. Landsteiner) в 1900 г. АВ0 система не единственная, дающая на поверхности эритроцитов антигены. У человека есть также MN-, C-, Rh-антигены (резус) и другие, всего известно более 20 групп крови. А- и В-антигены отличны от MN-антигенов и, соответственно, находятся под контролем различных локусов, локализованных в 9-й хромосоме. Наследование двух аллелей А и В кодоминантное, а аллель 0 проявляет рецессивность (ген не дает продукта). Когда отдельные особи тестируются с использованием антисыворотки против А и В антигенов, то удается обнаружить четыре фенотипические группы. Каждая особь в популяции человека имеет либо А антиген (А фенотип), В антиген (В фенотип), А и В антигены (АВ фенотип) и ноль антигенов (0 фенотип). Уже в 1924 г., на основе анализа наследования групп крови в семьях, было высказано предположение, что эти фенотипы контролируются тремя аллелями одного генетического локуса. Хотя возможны различные обозначения аллелей, мы используем символы I^A, I^B, и I^0 чтобы различать эти три аллеля АВ0 системы; *I* обозначает изоагглютиноген (другой термин для наименования антигена). В целом наследование АВ0 системы можно представить в виде следующей схемы (рис. 4.1.3):

Фенотип	**Антиген**	**Генотип**
О (I)	No	I^0I^0,
А (II)	А	I^AI^A, I^AI^0,
В (III)	В	I^BI^B, I^BI^0,
АВ (IV)	А и В	I^AI^B.

Рис. 4.1.3. АВ0 система группы крови человека.

Аллели I^A и I^B ведут себя доминантно по отношению к аллелю I^0, но кодоминантно по отношению друг к другу.

Fig. 4.1.3. The ABO human blood system.

The I^A and I^B alleles behave dominantly to the /° allele, but codominantly to each other.

Сходную с АВ0 системой группы крови имеют многие другие животные и, в частности, рыбы. Среди самых известных - можно отметить S- и А-систему антигенов у акулы (см. выше) и сельди. А-система сельди может быть представлена следующим образом:

Фенотип	**Антиген**	**Генотип**
А1 (I)	А1	A_1A_1, A_1A_2, A_1A_0,
А2 (II)	А2	A_2A_2, A_2A_0
А0 (III)	No	A_0A_0

$A_1 > A_2 > A_0$.

Как видно из приведенной схемы, не все генотипы сельди проявляются фенотипически. Чтобы увеличить точность генотипирования и дальнейшего анализа необходимо проводить истощение антисывороток. В заключение раздела можно сформулировать, что такое группа крови с точки зрения иммуногенетики. ***Группа крови*** *– это фенотип иммунной реакции, который экспрессируется одним или большим числом генотипов либо в гомозиготном, либо в гетерозиготном состоянии.*

4.2. НАСЛЕДОВАНИЕ ИММУНОГЕНЕТИЧЕСКИХ ПРИЗНАКОВ И ПОПУЛЯЦИОННАЯ ИЗМЕНЧИВОСТЬ

Иммуногенетические признаки наследуются обычно согласно простым Менделевским законам. Аллели одного локуса проявляют либо доминирование, либо кодоминирование. В первом случае доминирование осуществляется по отношению к нулевому аллелю, который не дает какого-либо продукта. Примеры с S-системой акулы и A-системой сельди это хорошо иллюстрируют. Сходное наследование обнаружено для нескольких групп крови полосатого тунца *Catsuvonus pelamis.*

Проведение скрещиваний и анализ наследования, которые необходимы для установления схемы генного контроля, довольно редко реализуются для природных видов из-за трудностей содержания и разведения животных из природы. Для рыб известно наследование иммуногенетических признаков у радужной форели *S. gairdneri* (Sanders, Wright, 1962). У нее обнаружено два антигена R-1 и R-2, на основе которых реализуются три фенотипа или три группы крови: R1 (генотип R_1R_1), R2 (R_2R_2) и R1-2 (R_1R_2). В скрещиваниях, проведенных цитированными авторами, обнаруживается: R1 X R1-2, Fa = 1:1; R1-2 X R1-2, F2 = 49 R1 + 91 R1-2 + 42 R2 или 1:2:1, как ожидается при моногибридном скрещивании с кодоминированием.

У карпа липопротеин *Lpt-1* наследуется как доминантный / рецессивный признак. У форели *S. trutta* обнаружена AB0-подобная группа крови, но детерминируемая двумя аллелями. В последнем случае было трудно провести анализ наследования из-за онтогенетической вариации.

Очень часто иммуногенетические признаки кодируются сериями множественных аллелей: сельдь имеет три аллеля A-системы, акула – три аллеля S-системы, полосатый тунец – до 12 аллелей в одной из нескольких определенных у него систем.

Общие особенности наследования иммуногенетических признаков можно суммировать следующим образом:

(1) Наличие доминирования, когда у гетерозигот проявляется только один из аллелей (антигенов);
(2) Наличие кодоминирования, когда у гетерозигот активны оба аллеля;
(3) Наличие новообразований: фенотип у гетерозигот не является простой суммой продуктов двух гомозигот (молекулярный механизм этого явления мы разберем в Лекции 5);
(4) Наличие эпистаза, когда один из генных продуктов ингибирует действие другого локуса;
(5) Наличие множественных аллелей: присутствие более чем двух аллелей в локусе – это норма для иммуногенетических локусов.

Закон Харди-Вайнберга

Из-за трудностей содержания большинства природных видов, включая морские организмы, наследование иммуногенетических признаков устанавливают по скрещиваниям, которые происходят естественным путем в природе. Если скрещивания происходят случайно, то между частотами гамет и генотипов существуют определенные пропорции. Эти пропорции впервые определили Харди (Hardy, 1908) и Вайнберг (Weinberg, 1908), по именам которых, и назван соответствующий закон (правило) Харди-Ванберга (Х-В). Прежде, чем сформулировать этот закон в словесном виде, необходимо определить, каковы условия для выполнения этого закона (модели).

Прежде всего, рассмотрим большую (в идеале - бесконечную) популяцию диплоидных бисексуальных организмов, спаривающихся случайно; то есть, гаметы в родительской популяции комбинируются в зиготах случайно. Далее, рассматриваются только отдельные аутосомные локусы. Пусть имеются три генотипа, а их численности равны:

Генотипы: R_1R_1 R_1R_2 R_2R_2

Численности: D H R;

D+H+R = N особей включенных в выборку (2N - генов). Аллельные (или равно гаметические) частоты тогда составляют для двух генов R_1 и R_2:

$p_{(R1)} = (2D + H) / 2N = (D + 1/2\ H) / N$ и $q_{(R2)} = (2R + H) / 2N = (R + 1/2\ H) / N$. Частота второго аллеля может быть найдена проще: $q_{(R2)} = 1 - p_{(R1)}$, так как $p_{(R1)} + q_{(R2)} = 1$. Частоты в таком определении представляют собой вероятности. Поэтому, для ситуации свободного (случайного) скрещивания можно записать частоты зигот как произведение частот гамет. Помня, что частота для R_1 = p, а для R_2 = q, имеем пропорции зигот в F_1, как они представлены в таблице

Самка (F)	Самец (M)	
	(R 1) p	(R 2) q
(R 1) p	p^2	pq
(R 2) q	pq	q^2

Другими словами: $(p + q)^2 = p^2$ (R_1R_1) + 2pq (R_1R_2) + q^2 (R_2R_2). Такие биномиальные или равновесные пропорции между гаметическими и генотипическими частотами сохраняются, пока действуют стохастические условия модели. Они могут нарушаться систематическими эволюционными факторами: естественным отбором, дрейфом, миграцией, а также другими причинами, нарушающими условия равновесия, например, ассортативностью скрещиваний или инбридингом. Выражение выше является аналитическим представлением закона Харди-Вайнберга. В словесном виде закон Х-В можно сформулировать следующим образом: *«В большой свободно скрещивающейся популяции, начиная с первого поколения, существуют биномиальные равновесные пропорции между частотами гамет и частотами генотипов, которые существуют неограниченно долго, если систематические эволюционные факторы – естественный отбор, дрейф и миграция, а также другие причины, не нарушают стохастические процессы».*

Правило Х-В дает очень легкий способ для проверки схемы генетического контроля того или иного полиморфного, качественного признака в природных популяциях. Можно сформулировать гипотезу генного контроля, вычислить аллельные частоты и по ним ожидаемые частоты генотипов исходя из правила Х-В, и затем сравнить их с наблюдаемыми (фактическими) численностями генотипов или фенотипов. Несколько примеров будет разобрано в рамках Практического занятия 4.

Сопоставление природных популяций

Популяции видов рыб были одними из первых природных объектов, проанализированных на основе иммуногенетических маркеров генов, когда с 50-х годов предыдущего столетия они были вовлечены в анализ. Кроме упомянутых акулы, сельди и тунца, были проанализированы такие виды, как европейский анчоус, треска, форель и другие. Открыты несколько групп крови: у трески *Gadus morhua* A- и E-, у полосатого тунца *Catsuvonus pelamis* - C-, Tg, G, Keyvamfar, у морского окуня *Sebastes mentella* и у европейского анчоуса *Engraulis encrasicholus* - A (Алтухов, 1974, Кирпичников, 1987).

Были описаны и сопоставлены частоты аллелей многих популяций. Полиморфизм иммуногенетических признаков был обнаружен у более, чем ста проанализированных видов. Поэтому использование иммуногенетических признаков сделало реальным генетические исследования на природных популяциях. Главный вывод на этом этапе заключался в том, что

генофонды различных популяций отличаются и популяции, как правило, находятся в равновесии Х-В. Однако, в тоже самое время было обнаружено, что популяции разных видов в природе непохожи одна на другую. Например, у полосатого тунца различия частот аллелей обнаружили, но только в популяциях отделенных очень большими расстояниями, такими как различные участки Тихого океана. Тогда как генетически отличающиеся популяции анчоуса были обнаружены не только в сравнительно близких, но географически разобщенных Азовском и Черном морях, но даже в пределах относительно небольшого Азовского моря выявлена значительная гетерогенность частот аллелей и даже был обнаружен уникальный аллель (рис. 4.3.1).

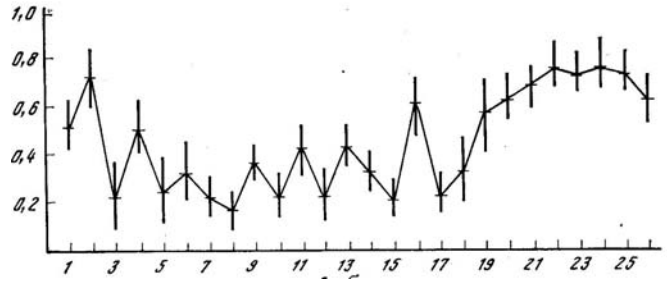


Рис. 4.3.1. Пространственная изменчивость группы крови А0 в нерестовых скоплениях анчоуса E. encrasicholus в Азовском море

По Y-оси – частоты аллелей, по X-оси - выборки. Различия частот значимы (По Алтухову, 1974).

Fig. 4.3.1. Space variation of blood group A0 in spanning gatherings of anchovy E. encrasicholus in the Azov Sea (From Altukhov, 1974).

On Y-axis is allele frequency, on X-axis are samples.

4.3. ИММУНОГЕНЕТИЧЕСКИЕ ДАТИРОВКИ ЭВОЛЮЦИИ

Данные секвенирования аминокислотных последовательностей, кроме прочего, используют для оценки эволюционного времени и прояснения генетических взаимосвязей организмов, хотя получение самих секвенс данных для полипептидов весьма трудоемко. Определение первичной последовательности нуклеотидов в ДНК много легче, но и это занимает много времени. Для таксономических целей необходимы и более простые методы, а самое главное важны различные методы (Nei, 1987; Avise, 1994; Avise, Wollenberg, 1997). Один из таких подходов – это использование интенсивности иммунологической реакции между антигеном и антисывороткой разных видов организмов.

Наиболее простым из нескольких методов оценки интенсивности иммунологической реакции является метод количественной фиксации микрокомплемента очищенного белка (Sarich, Wilson, 1966). Белок, обычно используемый для этой реакции, – это сывороточный альбумин (Champion et al., 1974). Вкратце процедура такова (рис. 4.3.1).

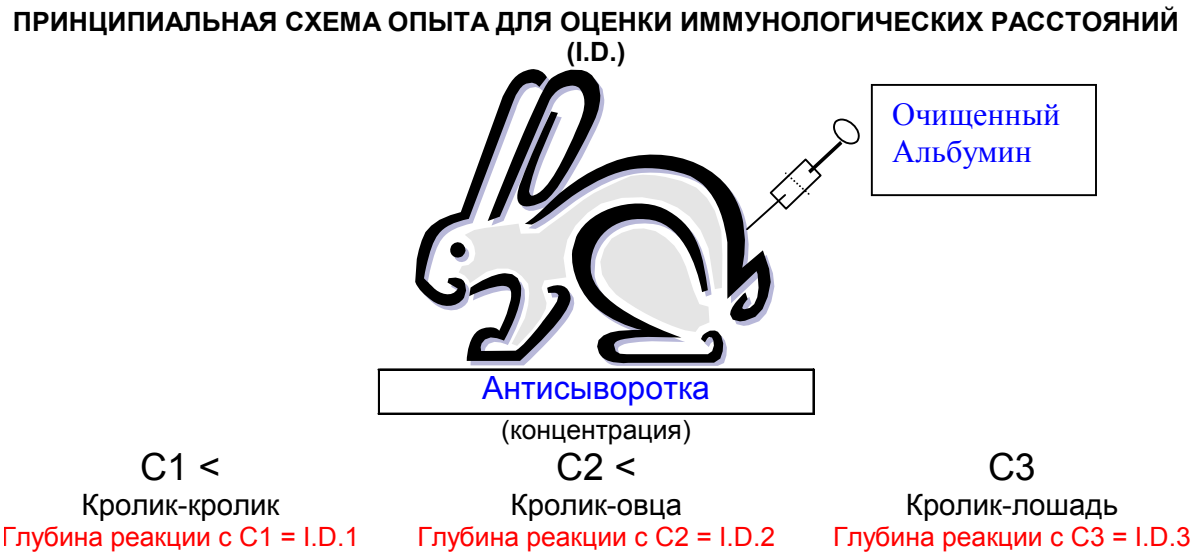


Рис. 4.3.1. Получение и использование антисыворотки для оценки I.D. Объяснения в тексте.
Fig. 4.3.1. Obtaining and usage an antisera for I.D. estimation. Explanation in the text.

Антисыворотку, которую предполагается использовать, получают иммунизацией кролика очищенным сывороточным альбумином организма, предназначенного для тестирования. Полученная антисыворотка реагирует максимально сильно с альбумином, взятым от того же организма (гомологичный антиген), но менее сильно - с альбумином организма, представляющего другой вид (гетерологичный антиген). Концентрация антисыворотки в опытах сохраняется одинаковой. Однако если концентрацию антисыворотки увеличить, то реакция с гетерологичным антигеном увеличивается, вплоть до уровня гомологичного антигена. Степень антигенных различий между двумя альбуминами измеряется величиной, на которую надо поднять концентрацию антисыворотки, чтобы гетерологичный альбумин производил с ней такую же активность, как гомологичный альбумин. Эта величина может быть названа индексом несходства (Index of dissimilarity, *I.D.).* Посредством этой реакции Сариш и Вильсон (Sarich, Wilson, 1967) показали, что логарифм *I.D.* приблизительно линейно связан со временем дивергенции между тестируемыми организмами. Они назвали величину $d_I = 100 \times \log_{10}$ (*I.D.*) иммунологическим расстоянием. Выяснилось, что для многих белков (альбумин, лизоцим, рибонуклеаза и др.) d_I линейно связано с долей различающихся аминокислот в их первичных последовательностях (Prager, Wilson 1971; Benjamin et al., 1984). Причины, почему d_I должно быть связано с долей различных аминокислот не вполне ясны. Однако эмпирические свойства d_I очень полезны для измерения генетических расстояний между видами, так как иммуногенетическая методика измерения много проще, чем анализ первичной аминокислотной последовательности. По Максону и Вильсону (Maxson, Wilson, 1974), одна единица d_I примерно соответствует различию в одну аминокислоту в альбумине. Связь между d_I и временем с момента дивергенции (t) можно выразить следующим образом:

$$T = cd_I. \qquad (4.1)$$

В 4.1 c – это константа пропорциональности, которая варьирует от белка к белку и частично от одной группы организмов к другой группе. Значение константы c, которое было использовано для млекопитающих, рептилий, амфибий, составило 5.5— 6.0×10^5, для птиц - 1.9×10^6 (Prager et al., 1974; Wilson et al., 1977; Collier, O'Brien, 1985). Эти значения показывают, что скорость изменения альбумина у птиц была в три раза меньше, чем у других организмов. Хотя оценка скорости эволюции у птиц может быть и неверной, поскольку ископаемые находки, использованные для

калибровки, не могут считаться очень надежными. В целом же если признать, что большинство отрядов птиц возникли относительно недавно (Wyles et al., 1983), то различие в скорости эволюции альбумина у птиц и других исследованных позвоночных может исчезнуть. Другая возможная причина этих различий будет обсуждена при рассмотрении базы данных по генетическим расстояниям (Глава 7).

Альбумин не может быть использован для реакции комплемент-фиксации у беспозвоночных. Поэтому Беверли и Вильсон (Beverley, Wilson, 1984; 1985), использовали белок гемолимфы для исследования филогенетических взаимосвязей различных видов Drosophilidae и в целом двукрылых. Используя информацию по ископаемым находкам в янтаре, по континентальному дрейфу и формированию островов, эти авторы установили, какова константа пропорциональности для белка гемолимфы, $c = 8 \times 10^5$. Как видим, эта величина такого же порядка, как обнаруженная для альбумина млекопитающих, рептилий и амфибий.

Оценка времени дивергенции посредством выражения 4.1 подвержена по меньшей мере четырем типам ошибок. 1. Экспериментальная ошибка. Она не превышает 2% оценки, даже если сама оценка относительно мала (Sarich, Wilson, 1966). 2. Вторая ошибка возникает в случае полиморфности антигена (белка) у исследуемого вида (ов). Для отдаленных видов величина этой ошибки относительно мала. 3. Третий тип ошибки связан с тем, что аминокислотные замены в белковом антигене возникают стохастично, а не детерминистично. Данная ошибка должна сделать дисперсию d_i больше величины средней. Поэтому она значительно более существенна, чем 1 и 2. 4. Этот тип ошибки связан с неполной пропорциональностью связи d_i с числом аминокислотных замен (Champion et al., 1974). Данный тип ошибки может быть не менее значимым, чем третий.

Часы иммунологических расстояний, по-видимому, несколько менее точные, чем часы, базирующиеся на аминокислотных последовательностях. Тем не менее, применение этого метода дало много важных находок для эволюционных исследований. Этот подход, в частности, был применен для уточнения филогенетических взаимосвязей обезьян и человека (Sarich, Wilson, 1966; 1967). Для своего времени результаты оказались неожиданными, показав близкое родство шимпанзе и гориллы к человеку, которого антропологи относят другому семейству (Hominidae), чем к орангутангу и гиббону, с которыми их объединяет одно номинальное семейство (Pongidae). Более того, как отмечалось в Лекции 1, дивергенция шимпанзе, гориллы и человека произошла только 5 млн. лет назад, а не ранее, как считалось в антропологии. Из примечательных результатов следует отметить, что виды лягушек, принадлежащие к одному и тому же роду, имеют часто иммунологические расстояния сравнимые с расстояниями между представителями семейств и даже отрядов млекопитающих (Maxson et al., 1975; Post, Uzzell, 1981; Maxson, 1984). К различиям в темпах эволюции в разных группах мы еще обратимся позже (Глава 7). Большие расстояния, обнаруженные у лягушек, а также у некоторых пар дрозофилид, предполагает их давнюю эволюционную дивергенцию. Беверли и Вильсон (Beverley, Wilson, 1984; 1985) оценили, что дивергенция различных подродов *Drosophila* (например, *D. virilis, D. mulleri)* от группы видов *D. melanogaster* произошла около 60 млн. лет назад, тогда как гавайские дрозофилиды отделились от рода *Drosophila* около 40 млн. лет назад (рис. 4.3.2). Возраст Гавайского архипелага датируется

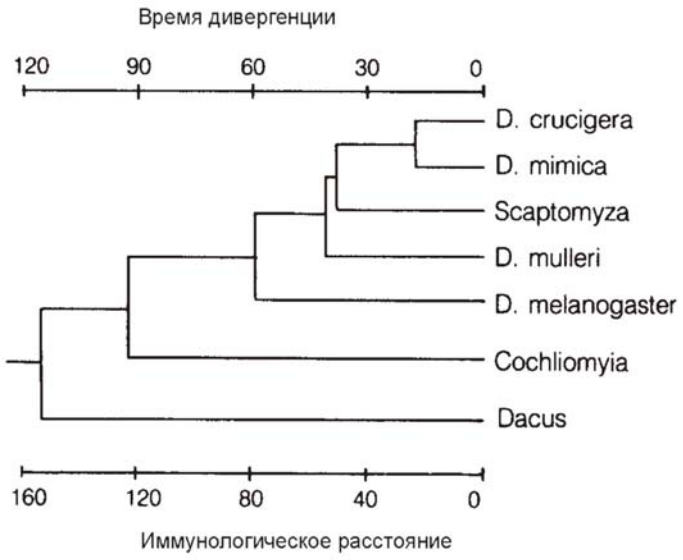


Рис. 4.3.2. Филогенетическое дерево реконструированное по иммуногенетическим расстояниям для семи групп дрозофилы.

D. crucigera и D. mimica являются Гавайскими дрозофилидами, тогда как D. mulleri и D. melanogaster представляют континентальных дрозофилид. Scaptomyza – это род, близкий к Drosophila. Drosophila, Cochliomyia и Dacus принадлежат к различным семействам (По Beverley, Wilson, 1984).

Fig. 4.3.2. Phylogenetic tree reconstructed from immunological distance data for seven fly lineages. D. crucigera and D. mimica are Hawaiian drosophilids, whereas D. mulleri and D. melanogaster are continental drosophilids. Scaptomyza is a fly genus closely related to Drosophila. Drosophila, Cochliomyia, and Dacus belong to different families (From Beverley, Wilson, 1984).

5-6 млн. лет и поэтому можно было бы ожидать, что дрозофилы с этих островов дивергировали относительно недавно. Однако иммуногенетические данные показывают, что такие виды как *D. crucigera* и *D. mimica* с Гавай дивергировали около 20 млн. лет назад. Некоторые близкие группы мух дивергировали здесь еще раньше, порядка 40 млн. лет назад, что в целом объяснимо историей этого архипелага, с несколькими последовательными поднятиями и опусканиями островов в последние 70 млн. лет (Beverley, Wilson, 1985).

4.4. ДАТИРОВКА ЭВОЛЮЦИИ ПОСРЕДСТВОМ АНАЛИЗА ПЕРВИЧНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ АМИНОКИСЛОТ

В предыдущих разделах мы имели дело с иммуногенетическими расстояниями, которые зависят от числа аминокислотных замещений в полипептидах и антителах. Эта методика имеет свои преимущества, но есть необходимость измерить число аминокислотных замен не только по иммуногенетическим маркерам, но и по другим локусам, кодирующим белки. Данные о первичной последовательности аминокислот в белках могут предоставить такую информацию. Секвенирование белков – это очень трудоемкое занятие и сейчас практически полностью замещено методами секвенирования ДНК. Однако рассмотреть эту идеологию полезно, так как она будет востребована при анализе дивергенции по ДНК.

Доля различных аминокислот и число аминокислотных замещений

Исследование изменений в полипептидных цепях начинается с сопоставления двух или большего числа аминокислотных последовательностей гомологичных белков различных организмов. Наиболее простая мера дивергенции между парой последовательностей аминокислот – это доля *(p)* различных аминокислот между двумя последовательностями, с оценкой равной:

$$\hat{p} = n_d / n, \qquad (4.2)$$

где n – это общее число сопоставленных аминокислот, а n_d – число различающихся аминокислот. Если аминокислотные сайты замещаются с равной вероятностью, то n_d следует биномиальному распределению. Поэтому дисперсия $\hat{p}$ задается выражением:

$$V(p\hat{}) = p(1-p) / n. \qquad (4.3)$$

Когда p мало, она равна среднему числу аминокислотных замещений. В других случаях отклонения от среднего могут быть велики. Для определения числа нуклеотидных замещений по их доле p, необходимо ввести определенную модель.

Пуассоновский процесс. Простая математическая модель, которая может быть использована для связывания с ожидаемым числом аминокислотных замещений на сайт – это Пуассоновский процесс в теории вероятностей (Nei, 1987). Пусть λ является скоростью аминокислотных замещений в год для некоторого аминокислотного сайта; допустим также для простоты, что λ одинаково для всех сайтов. Это допущение не обязательно выполняется в реальности, но ошибка от такого допущения невелика, если рассматриваемый временной отрезок не является очень большим. Среднее число аминокислотных замещений на сайт за период времени t тогда составит λt, а вероятность r появления аминокислотных замещений в определенном сайте задается следующим Пуассоновским распределением:

$$Pr\ (t) = e^{-\lambda t}\ (\lambda t)^r / r! \qquad (4.4)$$

Поэтому, вероятность того, что не случится ни одного замещения в определенном сайте равна $P_0(t)=e^{-\lambda t}$. Таким образом, если число аминокислот в полипептиде - есть n, то ожидаемое число неизмененных аминокислотных сайтов составит $ne^{-\lambda t}$. В действительности, у предкового вида число аминокислот в последовательности обычно неизвестно и поэтому выражение (4.4) неприменимо. Число аминокислотных замещений в обычной практике вычисляется на основе сравнения двух различных организмов, которые дивергировали t лет назад. Поскольку вероятность того, что ни одной аминокислотной замены не произойдет за t лет равна $e^{-\lambda t}$, вероятность (q), что ни один из гомологичных сайтов двух полипептидов не подвергнется замещению составит:

$$q = e^{-2\lambda t}. \qquad (4.5)$$

Эта вероятность на практике может быть оценена как $q\hat{} \equiv 1 - p\hat{} = n_i / n$, где n_i – это число идентичных аминокислот между двумя полипептидами. Равенство $q = e^{-2\lambda t}$ является примерным, поскольку обратные и параллельные мутации в расчет не принимаются. Однако эффект этих мутаций обычно весьма невелик, до той поры пока не рассматриваются очень длинные временные отрезки.

Если использовать формулу (4.5), общее число аминокислотных замещений на сайт для двух полипептидов $(d=2\ \lambda t)$ может быть оценено как:

$$d\hat{} = -\log_e q\hat{}. \qquad (4.6)$$

Поэтому, если известно t, то λ может быть оценено как $\lambda\hat{} = d\hat{} / (2t)$. Наоборот, если известно λ, t можно оценить из соотношения $t\hat{} = d\hat{} / (2\lambda)$.

Основная выборочная дисперсия d составляет:

$$V(d\hat{}) = [\mathrm{d}d / \mathrm{d}q]^2 * V(q) = (1 - q) / qn, \qquad (4.7)$$

Так как $V(q\hat{}) = q(1 — q)/n$ (Elandt-Johnson, 1970). Очевидно, что дисперсия $\lambda\hat{}$ и $t\hat{}$ задается $Vd\hat{} /(2t)^2$ и $V(d\hat{})/(2\lambda)^2$, соответственно.

Следует отметить, что если число аминокислотных замещений для всех сайтов аминокислот известно, дисперсия числа аминокислотных замещений на сайт должна быть при Пуассоновском процессе $2\lambda t / n$ (дисперсия равна среднему значению). На практике, однако, невозможно знать эти числа, поэтому приходится оценивать d через выражение (4.6). А поскольку (4.6) основывается на неполной информации об аминокислотных замещениях, (4.7) дает дисперсию большую, чем $2\lambda t / n$.

В описанных выше выкладках, предполагалось равенство аминокислотных замещений для всех сайтов. Однако это допущение не соответствует экспериментальным фактам, поскольку хорошо известно, что скорость замещений выше в функционально менее значимых сайтах (King, Jukes, 1969; Dickerson, 1971; Dayhoff, 1978; Nei, 1987). Было показано, что в действительности число аминокислотных замещений имеет большую дисперсию, чем это ожидается при Пуассоновском процессе (Fitch, Margoliash, 1967b; Uzzell, Corbin, 1971). Хотя, выражение (4.6)

является достаточно робастным и приблизительно выполняется, даже если скорость значительно варьирует от сайта к сайту. В этом можно убедиться, рассматривая предельный случай, когда доля *a* аминокислотных сайтов является неизменной, а доля 1 - *a* распределена по закону Пуассона. Ожидаемая доля идентичных аминокислот для этого случая задается как $q = a+(1 - a)\, e^{-2\lambda t}$. Исходя из этого выражения, очевидно, что $d \equiv -\log_e q$, строго говоря, не является линейным во времени. Однако результаты моделирования показывают, что *d* примерно линейно, когда $2\lambda t <= 1$ (рис. 4.4.1). Следует заметить, что когда $2\lambda t << 1$, то $e^{-2\lambda t} \approx 1 - 2\lambda t$. Поэтому, $d \approx -\log_e[1 - 2(1 - a)\lambda t]$. Так как $(1-a)\lambda$ является средней скоростью $(\bar{\lambda})$ аминокислотных замещений для всех аминокислотных сайтов, то *d* может быть записано как $2\bar{\lambda}t$. Именно, когда λ варьирует между аминокислотами, то все равно $d = 2\bar{\lambda}t$ выполняется до тех пор, пока $2\bar{\lambda}t$ велико.

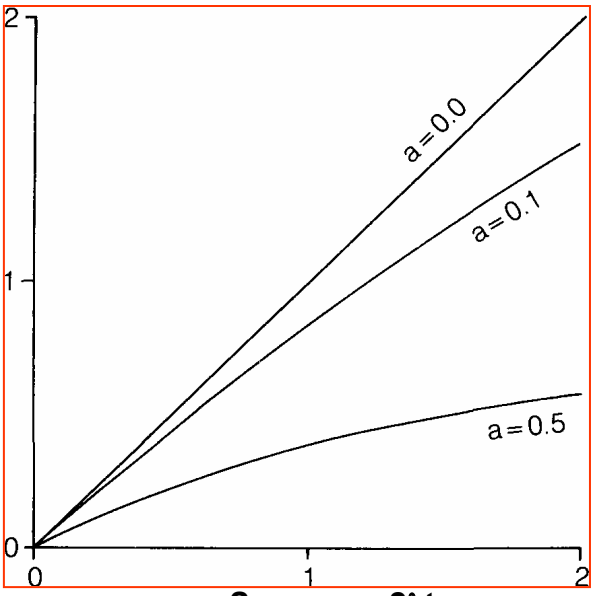


Рис. 4.4.1. Влияние неизменности аминокислот на оценку аминокислотных замещений (d); a – доля неизмененных аминокислот в полипептиде (По Nei, 1987).

Fig. 4.4.1. Effects of invariable amino acids on the estimate of amino acid substitutions (d); a, proportion of invariable amino acids in a polypeptide (From Nei, 1987).

Так как это было сформулировано, предполагалось, что сопоставленные аминокислотные последовательности имеют одинаковое число аминокислот и что дивергенция между ними осуществляется только путем замещений (замен). Однако когда сравниваются отдаленно связанные последовательности, вовлечены также вставки (инсерции) и делеции. В этом случае необходимо первоначально идентифицировать вставки и делеции. Когда число этих изменений невелико, как в примере, который будет рассмотрен в Практическом занятии 4 (п. 4), это может быть сделано относительно легко. Когда же число вставок и делеций относительно велико, выравнивание аминокислотных последовательностей довольно трудоемко. Поскольку эта проблема наиболее серьезна при сравнении последовательностей ДНК, а сама проблема почти идентична для двух типов данных, то ее обсуждение будет проведено при анализе первичных последовательностей ДНК.

Матрица аминокислотных замещений

Представленный выше Пуассоновский процесс, как модель для оценки числа аминокислотных замен, дает вполне точные оценки, если рассматриваемое эволюционное время относительно мало. На более протяженных отрезках времени нельзя исключить, как отмечалось, влияния возвратных и параллельных мутаций и поэтому распределение Пуассона будет давать недооценку числа замещений. Неравная скорость замен по разным аминокислотным сайтам также будет добавлять неточность получаемых оценок. Чтобы преодолеть эти трудности Дэйхофф с соавторами (Dayhoff et al., 1978) предложили так называемый метод **Матрицы аминокислотных замещений**, который основывается на числе ассимилированных точечных мутаций (PAM -*accepted point mutations*). Детали этого подхода можно найти в монографии Нея (Nei, 1987, pp.45-46).

4.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4

Предполагается рассмотреть четыре основных вопроса.

1. Закон Харди-Вайнберга (Х-В): Повторение формулировки и численные примеры.

2. Хи-квадрат тест для оценки соответствия наблюдаемых и ожидаемых из соотношения Х-В численностей генотипов. Оценивание стандартной ошибки (SE) частот из выражения (4.3) и гетерогенности частот аллелей тестом хи-квадрат.
3. Пакет программ: SPECSTAT. Разобрать численные примеры для тестирования закона Х-В и гетерогенности частот аллелей.
4. Провести расчеты величины *d*^ для четырех видов позвоночных по первичным последовательностям α гемоглобина.

Рисунок 4.4.2 показывает аминокислотные последовательности α цепи гемоглобина человека, лошади, быка и карпа. Цепь гемоглобина млекопитающих состоит из 141 аминокислоты, а у карпа гемоглобин представлен 142 аминокислотами. Сопоставление этих последовательностей предполагает, что делеции или вставки произошли в трех положениях после дивергенции млекопитающих от рыб. Если мы игнорируем эти делеции/вставки, то доля различных аминокислот (p^) и оценка числа аминокислотных замещений для каждой пары организмов могут быть представлены, как это дано в Таблице 4.4.1. Например, в случае человека и карпа p^ = 68/140=0.486, а $d^\wedge = -\log_e(1 - p^\wedge) = 0.666$. Дисперсия *d*^ равна $V(d^\wedge) = 0.486/(0.514 \times 140) = 0.006754$, а стандартная ошибка, соответственно - 0.082.

```
Человек VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGHG
Лошадь  VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHF-DLSHGSAQVKAHG
Бык     VLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHF-DLSHGSAQVKGHG
Карп    SLSDKDKAAVKIAWAKISPKADDIGAEALGRMLTVYPQTKTYFAHWADLSPGSGPVK-HG

Человек KKVA-DALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT
Лошадь  KKVA-DGLTLAVGHLDDLPGALSDLSNLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
Бык     AKVA-AALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT
Карп    KKVIMGAVGDAVSKiDDLVGGLASLSELHASKLRVDPANFKILANHIVVGIMFYLPGDFP

Человек PAVHASLDKFLASVSTVLTSKYR
Лошадь  PAVHASLDKFLSSVSTVLTSKYR
Бык     PAVHASLDKFLANVSTVLTSKYR
Карп    PEVHMSVDKFFQNLALALSEKYR
```

Рис. 4.4.2. Аминокислотные последовательности в α цепи гемоглобина четырех видов позвоночных. Аминокислоты представлены первыми буквами их английских наименований. Черточки отмечают позиции делеций или вставок (По Nei, 1987).
Fig. 4.4.2. Amino acid sequences in the α chains of hemoglobins from four vertebrate species. Amino acids are expressed in terms of one-letter codes. The hyphens indicate the positions of deletions or insertions (From Nei, 1987).

Таблица 4.4.1. Число аминокислотных различий (выше диагонали) между ***α*** цепями гемоглобина человека, лошади, быка и карпа. Делеции и вставки были исключены из подсчета; общее число использованных аминокислот составило 140. Цифры в скобках показывают доли различных аминокислот (p^). Значения, приведенные ниже диагонали – это оценки среднего числа аминокислотных замещений на сайт для пар видов *(d^) (По Nei, 1987).*

Table 4.4.1. Numbers of amino acid differences (above the diagonal) between hemoglobin ***α,*** chains from the human, horse, bovine, and carp. Deletions and insertions were excluded from the computation, the total number of amino acids used being 140. The figures in parentheses are the proportions of different amino acids. The values given below the diagonal are estimates of the average number of amino acid substitutions per site between two species *(d^) (From Nei, 1987).*

Организм	***Человек***	***Лошадь***	***Бык***	***Карп***
Человек		18(0.129)	17(0.121)	68(0.486)
Лошадь	0.138 ±0.032		18(0.129)	66(0.471)
Бык	0.129 ±0.031	0.138 ±0.032		65(0.464)
Карп	0.6б6 ±0.082	0.637 ±0.080	0.624 ±0.079	

Данные Таблицы 4.4.1 показывают, что *d*^ весьма близко для пар видов млекопитающих, тогда как для карпа и этих видов значения *d*^ значительно больше. Подобные различия согласуются с тем, что число аминокислотных замен приблизительно пропорционально времени прошедшего с момента дивергенции; для человека, лошади и быка – это порядка 75 млн. лет назад, а для карпа (костистые рыбы) и млекопитающих – это около 400 млн. лет назад (рис. 1.2.2, Глава 1). Средняя величина *d*^ для трех пар видов млекопитающих равна 0.135, тогда как для этих трех видов и карпа соответствующая величина составила 0.642. Соотношение этих величин близко к соотношению времени дивергенции.

ГЛАВА 5. БИОХИМИЧЕСКАЯ ГЕНЕТИКА

ГЛАВНЫЕ ВОПРОСЫ:

5.1. Введение в методы анализа изменчивости белков.
5.2. Генетическая интерпретация изменчивости белков.
5.3. Генетика изоферментов.
5.4. Эволюция регуляции генома.
5.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 5.

РЕЗЮМЕ

1. Изоферменты, а точнее аллозимы являются очень хорошими маркерами генотипической аллельной изменчивости.
2. Экспрессия ферментных локусов обычно кодоминантная.
3. По аллозимным локусам обычны множественные аллели.
4. Наследование аллозимной изменчивости простое Менделевское. По этим признакам, как правило, выполняется закон Харди-Вайнберга.
5. Изоферменты являются наилучшими маркерами для исследования тканевой активности генов.
6. Изоферменты позволяют исследовать время активации генов в онтогенезе и генетические последствия их работы.

Chapter 5. BIOCHEMICAL GENETICS

SUMMARY

1. Isozymes or rather allozymes are very good markers of intra loci or allelic variability.
2. Expression of enzyme loci is usually codominant.
3. Multiple alleles are common for allozyme loci.
4. Inheritance of allozymes is simple Mendelian. H-W law is normally held in nature for these traits.
5. Isozymes are the best markers to detect tissue activity of genes.
6. Isozymes permit to investigate time of gene activation and consequences for turn on the genes in ontogenesis.

5.1. ВВЕДЕНИЕ В МЕТОДЫ АНАЛИЗА ИЗМЕНЧИВОСТИ БЕЛКОВ

Биохимическая генетика родилась первоначально, как область биохимии применительно к микроорганизмам, включая и исследования дрожжей в середине 1940-х. Однако позже, с внедрением в нее методов гистохимии, бурное направление получила биохимическая генетика природных популяций. С середины 1960-х годов, в течение 30 лет наблюдался бум исследований в этой области. Сейчас акценты сместились на методики анализа ДНК, но анализ белков, как маркеров генов не утратил своей актуальности и дает очень много аналитических возможностей.

Имеется множество методов для исследования белков и оценки их изменчивости: хроматография, гель-фильтрация, осаждение ультрацентрифугированием и другие. Главным и наиболее чувствительным, однако, остается электрофорез белков в геле с последующим гистохимическим определением их локализации и активности.

Системы электрофореза

Пять из 20 наиболее распространенных аминокислот, составляющих белки, несут электрический заряд. Заряд таких аминокислот, как аргинин, гистидин и лизин является в норме положительным, тогда как у аспарагиновой и глутаминовой кислот – он отрицательный. Различные белки имеют отличающийся суммарный заряд в силу того, что представлены различными наборами аминокислот в своих последовательностях. Электрофорез использует это физико-химическое свойство белков для разделения их смесей на основе заряда. Если по локусу, кодирующему белок, обнаруживаются аллельные различия, то суммарный заряд полипептида или белка, детерминируемого ими, часто также отличается. Гелевый электрофорез белков позволяет обнаружить такие аллельные различия.

Методика гелевого электрофореза довольно проста (рис. 5.1.1). В гель (обычно крахмальный или полиакриламидный, ПАГ) вносятся образцы белков, и в конструкцию подается постоянный ток. Образцы белков представляют собой их водные растворы, полученные экстракцией из отдельных тканей (мышцы, печени, почки и т.д.) или из целого организма, если его размер мал. Могут, наконец, использоваться готовые организменные растворы, такие, например, как сыворотка крови. Экстракцию часто проводят с использованием буферных растворов с низкой ионной силой и pH близкой к нейтральной, с добавками стабилизаторов химических связей и не ионных детергентов. Образцы обычно вносят в гель, смочив белковым раствором специальную фильтровальную бумагу (ватман) или в случае ПАГ – пипеткой. Обычно в гель вносится 25-50 образцов, хотя на рис. 5.1.1 показано только восемь.

Методика электрофореза основывается на ионизации молекул в буферных растворах и их движении в зависимости от заряда к аноду (+) или катоду (-). В нейтральных или слабощелочных буферах, которые чаще всего используются для электрофореза, большинство белков, ионизируясь, приобретают положительный суммарный заряд и поэтому двигаются в сторону анода.

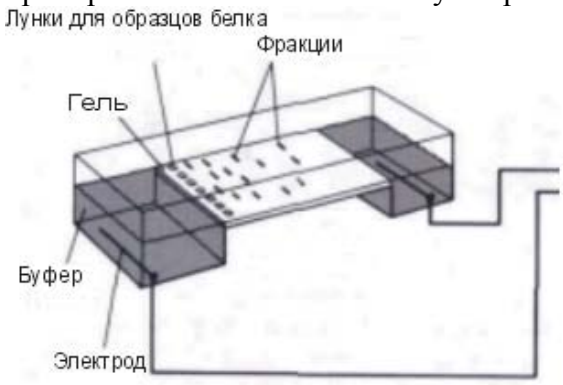


Рис. 5.1.1. Принципиальная схема электрофореза в геле.

Образцы белка после экстракции из тканей (мышца, печень) помещаются в лунки в геле. После подачи постоянного тока ("-", "+") молекулы белков двигаются, образуя фракции ввиду различий зарядов молекул. Фракции обычно видны только после завершения электрофореза и проведения специфического окрашивания.

Fig. 5.1.1. General scheme of electrophoresis.

Protein mixture (specimen) is extracted from tissue such as muscle or liver. Specimen from each individual is introduced one by one to gel by soaked filter paper inserted in slots. Different forms of a particular protein often move different distances from the point of applications when electric current ("-" and "+") is applied because they have different charges. The bands are identified later by a specific staining.

В биохимической генетике электрофорез проводят с использованием различных носителей. Чаще всего используются упомянутые крахмальный гель и ПАГ, а также агарозный гели. Реже используют карбоксиметилцеллюлозу, дезоксиэтиламиноэтилцеллюлозу и специальную бумагу.

Макромолекулы разделяют чаще всего в соответствии с их суммарным электростатическим зарядом, приобретаемом в буферном растворе (буфере), на котором изготовляется гель. Носитель также может служить в качестве молекулярного сита, разделяя макромолекулы в соответствии с их

молекулярной массой. Особенно эффективен в этом отношении ПАГ, так как это искусственный полимер и размеры его пор можно планировать, варьируя концентрацию мономерных компонентов. Существует два главных типа электрофореза: (1) непрерывный и (2) прерывистый. В первом типе, компоненты буфера являются одинаковыми для геля, а также для катодной и анодной частей электрофоретической системы. Во втором типе, состав буферных растворов и их концентрация могут быть различными в геле и электродных отсеках. Что такое буфер? Буфер – это специальный раствор (электролит), в состав которого входят два компонента, один – сильная кислота, другой – слабое основание или наоборот. Часто оба компонента – кислота и основание не являются сильными. Для примера, из множества буферов назовем три: буфер трис-HCl, боратный H_3BO_3- NaOH буфер и трис-цитратный буфер. Трис – это популярное в биохимии основание (трис-оксиметил аминометан).

Гистохимическое окрашивание макромолекул

После электрофореза стоит задача каким то путем визуализовать расположение в геле белков или других макромолекул, если они исследовались, поскольку подавляющее число макромолекул бесцветны или находятся в низких концентрациях, чтобы их можно было видеть. Методики гистохимического окрашивания позволяют это сделать. Имеются три типа таких реакций: (1) специфическое окрашивание на определенный тип макромолекул, (2) специфическое окрашивание на определенную группу радикала в белках и (3) оксидоредуктазная (дегидрогеназная) реакция фермента с использованием специфического субстрата.

(1) В реакциях первого типа используются специфические красители широкого спектра. Например, все белки окрашиваются Амидочерным – Б (Amido Black-B) или Кумаси голубым (Comassie Blue). ДНК окрашивается Цитоорсеином (Cytoorsein), Метиленовым синим (Methylene Blue) и др.

(2) Второй тип – это специфическое окрашивание, которое позволяет выявлять определенные химические группы в белках. С помощью этой реакции окрашивания можно выявлять 9 или более групп. Среди них такие как: 1) аминогруппы концевых альфа–аминокислот, 2) концевые и боковые карбоксильные группы, 3) сульфгидрильные группы (SH) цистеина и др.

(3) Третий тип – это дегидрогеназная реакция. Это наиболее многочисленный тип реакций, использующий специфичность ферментативной реакции со строго определенным субстратом. Схема типичной реакции такова:

$$\textbf{Субстрат} \xleftrightarrow{\textbf{ДГ}} \textbf{Продукт}$$

Некоторые другие компоненты также участвуют в реакции. Для примера рассмотрим реакцию окрашивания, проводимую для выявления лактатдегидрогеназы (ЛДГ):

РЕАКЦИЯ: **Лактат ↔ ЛДГ ↔ Пируват**

Компоненты реакции	**Лактат в буфере с pH- и температурным оптимумом**	**Пируват**	Продукты реакции
	NAD+, NaCN+, $MgCl_2$+	**NADH**	
	PMS+, NBT+	**PMS-, NBT- (Формазан в зоне реакции – голубые фракции)**	

В результате реакции краситель (NBT+) окисляется и превращается в формазан – ярко окрашенное вещество, находящееся непосредственно в месте прохождения ферментативной реакции. Большинство белков, которые исследуются посредством гелевого электрофореза, являются ферментами. Поэтому легко разработать специальные методики окрашивания, чтобы визуализовать активность специфического фермента. В биохимическую генетику подобная практика вошла с середины 50-х годов прошлого столетия (Hunter, Markert 1957). Множество источников дают детальные описания методик окрашивания ферментов на гелях после электрофореза (Shaw, Prasad, 1970; Harris, Hopkinson, 1976; Manchenko, 2002). Описано несколько

сот реакций, каждая из которых позволяет выявить активность отдельного фермента на геле из их смеси, подвергнутой электрофорезу.

5.2. ГЕНЕТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ИЗМЕНЧИВОСТИ БЕЛКОВ

Специфическое окрашивание с использованием ферментативной активности позволяет различить даже изоформы ферментов. На завершающем этапе электрофореза, после окрашивания, получают зимограмму (электрофореграмму) – полосу геля, на которой расположены выявленные для отдельных особей фракции изоферментов (рис. 5.2.1). Расположение фракций от отдельных особей (фенотипы) можно легко интерпретировать, как генотипическую изменчивость по конкретному ферментному локусу (рис. 5.2.1).

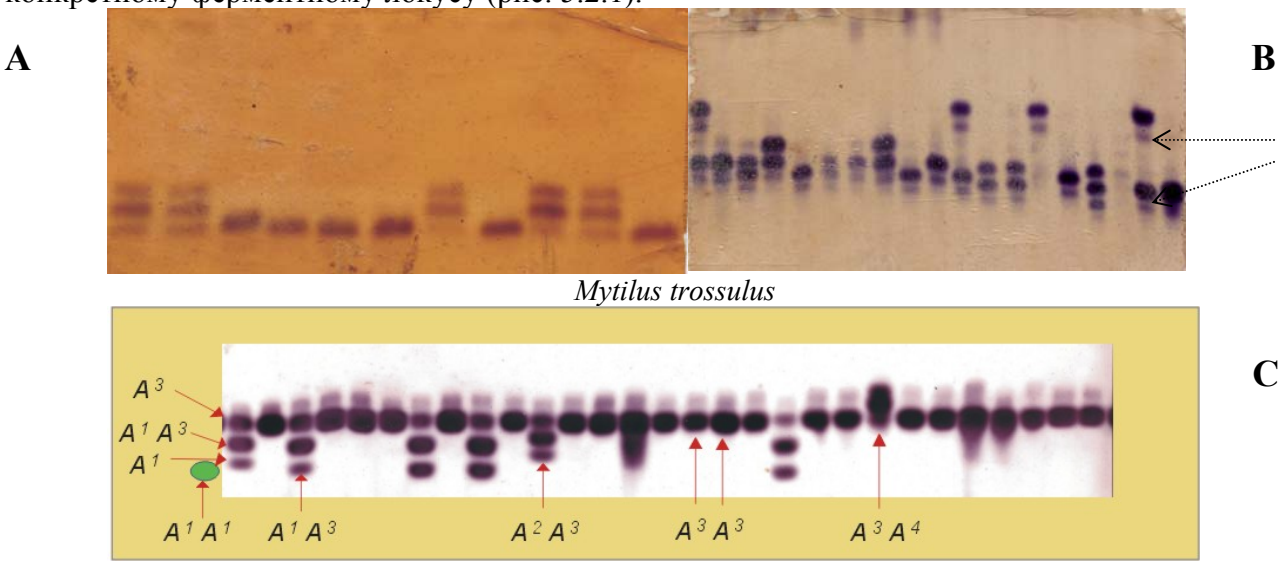


Рис. 5.2.1. Примеры зимограмм (полос геля с окрашенными ферментами), полученными в результате крахмально-гелевого электрофореза.

*А - аспартатаминотрансфераза (локус **AAT-1**), В - лецинаминопептидаза (**LAP-1**), С - формальдегиддегидрогеназа (**FDH**). Изоферменты у олигомерных ферментов могут быть комбинацией субъединиц разного типа (отличающихся подвижностью), как показано стрелками слева. Стрелками снизу показаны варианты фенотипов и их генетическая интерпретация. Пунктирные стрелки – не генетически определяемые ММФФ (см. ниже).*

*Fig. 5.2.1. Examples of starch-gel electrophoretic zimograms or enzymes stained in the gel. A, aspartate aminotranspherase (**AAT-1**), B, leucine aminopeptidase (**LAP-1**), and C, formaldehide dehydrogenase (**FDH**). Shown with arrows are the isozymes and their genetic interpreting. Dotted arrows show conformation changes.*

Генетическую интерпретацию можно представить в несколько ином виде, показав схематически фенотипические различия в зависимости от четвертичной структуры ферментов (рис. 5.2.2).

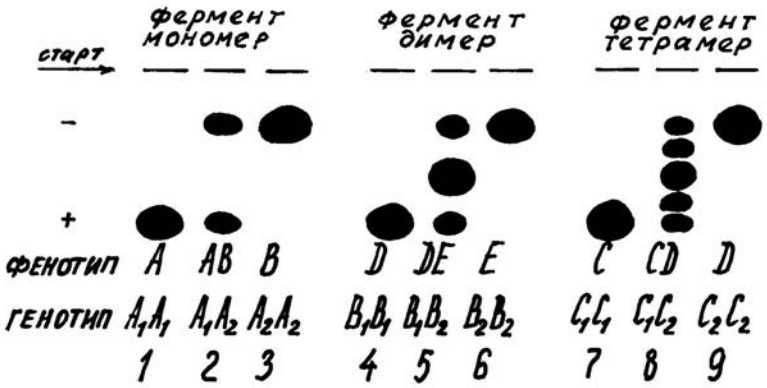


Рис. 5.2.2. Схема генотипической интерпретации фенотипической изменчивости по ферментам, отличающимся четвертичной структурой.

Локус кодирующий мономерный, димерный и тетрамерный фермент с двумя аллелями (A_1, A_2) и тремя вариантами фенотипа (полиморфизм) и генотипа (A_1A_1, A_1A_2, A_2A_2). Эти аллели производят субъединицы двух типов, медленные и быстрые. Комбинации субъединиц одного типа называются гомомерными, а различных типов – гетеромерными изоферментами.

Fig. 5.2.2. Interpretation of allozyme pattern of variability in the terms of genotype depending on their quaternary structure.

A locus is coded for a monomeric, dimeric and tetrameric enzyme having two alleles designated as A_1 and A_2 (i.e., a polymorphic locus). These alleles produce two types of subunits (sections within ellipses). The protein encoded by the A_1 allele migrates more slowly than the protein encoded by the A_2 allele. Combinations of same subunits are called homomeric, and different subunits called heteromeric isozymes. Three different genotypes are possible for an individual at these loci.

При генетической интерпретации полиморфизма изоферментов важно помнить (Глава 4):

1. Что один ген детерминирует один полипептид.

2. Что имеется колинеарность, - однозначное соответствие последовательности нуклеотидов в ДНК и последовательности аминокислот в белках.

Доказательства в поддержку первого утверждения были получены также в экспериментах по искусственному синтезу пептидов на матрице мРНК. Второе утверждение подтверждается прямым сопоставлением первичных последовательностей в генах и в полипептидах, синтезированных на них; мутации, приводящие к заменам нуклеотидов, ведут к замещениям аминокислот (рис. 5.2.3, верх). Данные, подтверждающие второе положение были получены также до появления методики секвенирования (рис. 5.2.4, низ).

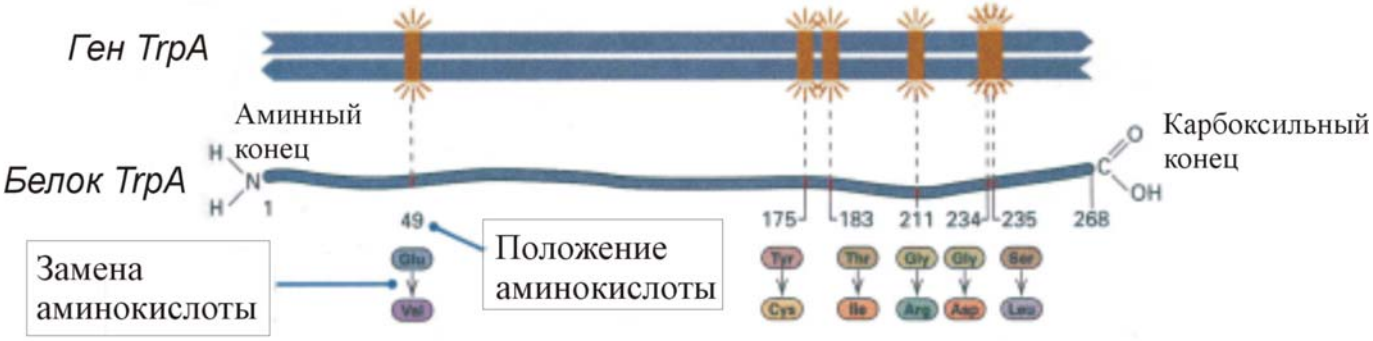


Рис. 5.2.3. Связь между положением мутации на генетической карте E. coli гена триптофана (TrpA) и положением замещенной аминокислоты в белке TrpA (По Hartl, Jones, 2002).

Fig. 5.2.3. Correlation of the positions of mutations in the genetic map of the E. Coli triptophan gene, trpA with positions of amino add replacements in the TrpA protein (From Hartl, Jones, 2002).

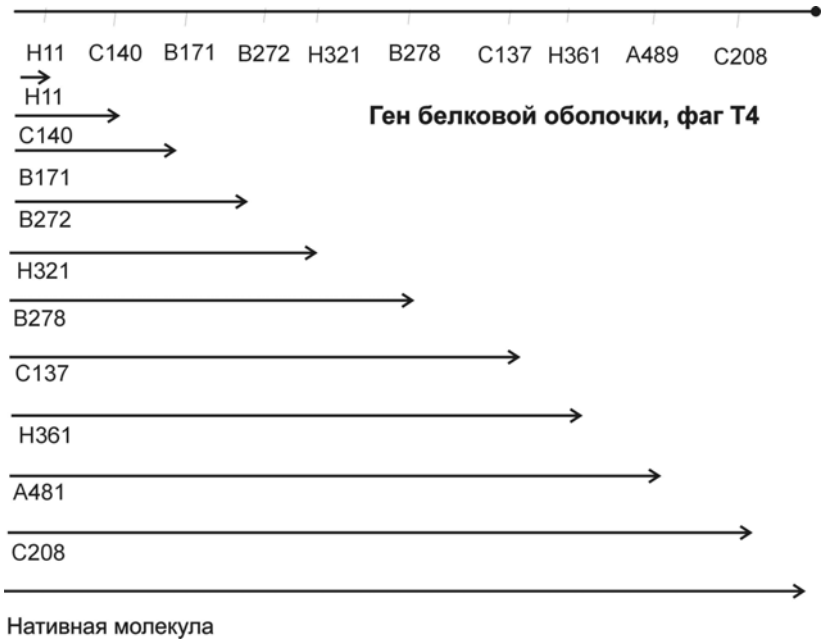


Рис. 5.2.4. Корреляция между положением мутации на генетической карте фага Т4 гена белковой оболочки и длиной синтезированного пептида (По Stent, 1974).
Fig. 5.2.4. Association of the positions of mutations in the genetic map of the Phage T4 gene of protein cover and the length of polypeptides synthesized (From Stent, 1974).

Таким образом, общая схема генетической интерпретации полиморфизма изоферментов, а также их аллельных вариантов, аллоферментов (аллозимов), весьма проста, особенно когда имеется четвертичная структура. Если два различных локуса кодируют изоферменты, генетическая интерпретация также может быть вполне простой.

5.3. ГЕНЕТИКА ИЗОФЕРМЕНТОВ

В предыдущей лекции мы провели разграничение между полипептидом и белком и отметили множество модификаций, происходящих с белками, в том числе с ферментами. Ферменты имеют ряд особенностей изменчивости, формируя изоформы различной природы. Ферменты и их изоформы – изоферменты (изозимы), как мы согласились ранее, являются наиболее важными маркерами в биохимической генетике. Давайте рассмотрим, каковы источники происхождения **множественных молекулярных форм ферментов** (ММФФ). ММФФ имеют сложную природу и только часть из них являются истинными изоферментами, имеющими наследственную детерминацию (Корочкин и др., 1977). Белки могут быть модифицированы *in vivo* и *in vitro*. Многие внутриклеточные модификации являются естественным процессом, ведущим к активной молекуле. Как упоминалось ранее (Глава 4), в отношении низкоплотностного липопротеина (LDL), пятнадцатый экзон определяет домен посттрансляционной добавки карбогидрата, а шестнадцатый и семнадцатый экзоны дают продукты связывания LDL с клеточной мембраной. Лишь после этого окончательно формируется активный LDL-белок, и сама мембрана приобретает новые свойства.

Можно разбить ММФФ на две группы: не генетически и генетически определяемые ММФФ. *Под генетическими ММФФ надо понимать такие вариации, которые детерминированы: (1) изменениями в первичной последовательности ДНК аллелей, вызывая соответствующие замены аминокислот в полипептидах (аллозимы) и/или (2) комбинированием продуктов (субъединиц) аллельных или неаллельных генов* (гомомерные и гетеромерные изоферменты; см. рис. 5.2.2). Могут быть и другие генетические источники ММФФ, обусловленные, например, альтернативным сплайсингом (см. ниже). Не генетические ММФФ - это изменения другого рода. Так, добавление карбо-гидрата к LDL-белку, меняет его молекулярную массу и конформацию и, хотя это наследственно детерминировано, но не создает генетически обусловленных вариантов. Таким образом, это случай не генетических ММФФ. В разделе 5.2 выше рассматривались только

генетически детерминированные ММФФ и интерпретация, в основном, относилась к **аллозимной изменчивости**. Реальная ситуация более сложна. Давайте обратимся еще раз к рис. 5.2.1. На рисунке, в подразделах B и C, можно видеть, кроме основных фракций, несколько минорных фракций (субфракций). Это не изоферменты, а конформеры, то есть, ММФФ, определяемые посттрансляционными модификациями молекул белков. В данном случае не ясно, как они возникли *in vivo* или *in vitro*. Но это для наших иллюстративных целей и не важно. Принципиально то, что они не являются наследуемой изменчивостью и должны быть исключены при генетической интерпретации. Другого рода осложнения возникают, когда два или более локусов экспрессируются в одной и той же ткани. В этих случаях субъединицы, синтезируемые с различных локусов, могут взаимодействовать, формируя изоферменты, в этом случае – межлокусные гетеромерные («гибридные») полипептидные продукты (рис. 5.3.1). Схематическое представление экспрессии полиморфных дуплицированных локусов для ферментов, с различной четвертичной структурой, дается в обзоре Аттера и соавторов (Аттер и др., 1987, рис. 2.5). Хороший пример взаимодействия представляет дуплицированный локус димерного фермента ГФИ (рис. 5.3.1).

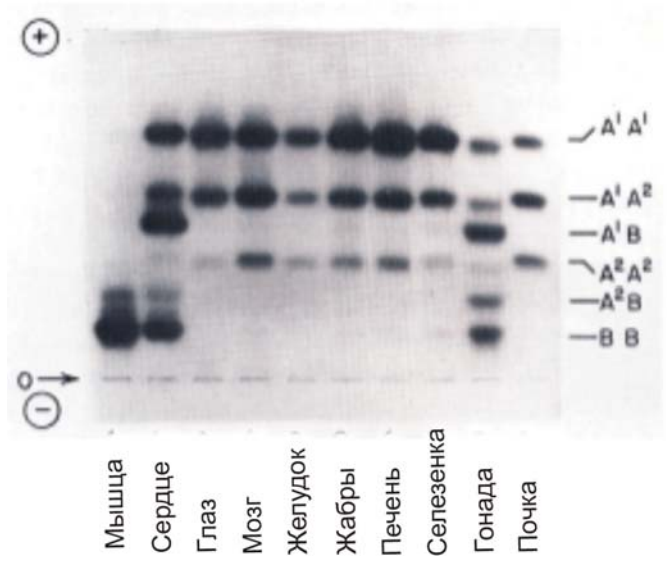


*Рис. 5.3.1. Дифференциальная экспрессия глюкозофосфатизомеразы **(ГФИ, локусы GPI-A*, GPI-B*)** в тканях рыбы Hypentelium nigricans. Между тканями наблюдается также интенсивная дивергенция экспрессии по дуплицированному локусу **GPI-A***. В мышце обнаружена экспрессия только одного локуса **GPI-B***, что обусловлено прошедшей диплоидизацией. Видно, что гетеромерные изоферменты могут образовываться между всеми тремя локусами (По Ferris, Whitt, 1979).*
*Fig. 5.3.1. Differential expression of glucosephosphate isomerase-A loci **(GPI*)** in tissues of fish Hypentelium nigricans. Extensive divergence in the expression of duplicate **GPI-A*** loci have occurred within and among tissues. The **GPI-B*** locus is singly expressed in tissues of this species due to gene diploidization (From Ferris, Whitt, 1979).*

Исключения в кодоминантном проявлении аллелей

Фенотипы ферментов, представленные на рисунках 5.2.1-5.2.2 и 5.3.1 обеспечиваются кодоминантным проявлением (экспрессией) аллелей. У гетерозигот оба различных аллеля проявляются в виде белковых продуктов, формируя в случае присутствия изоферментов гетеромерные молекулы или субъединицы. Именно за счет формирования гетеромерных молекул фенотип гетерозиготных особей не сводим к сумме фенотипов обоих родителей (см. рис. 5.2.2). Напомним, что впервые подобное свойство гетерозиготных особей было обнаружено при иммуногенетических исследованиях (Глава 4). Кодоминантное наследование является важной особенностью генетики изоферментов, поскольку от этого зависит генетическая интерпретация изменчивости. Кроме того, если кодоминантность, как правило, соблюдается, то это может быть основой для тестирования некоторых теоретических концепций (см. Глава 12). Однако отклонения от кодоминантного проявления имеются. Рассмотрим несколько подобных случаев.

Например, у некоторых лососевых рыб, обнаружено, что идентичные электрофоретически субъединицы синтезируются двумя различными локусами (рис. 5.3.2). Генетико-эволюционной основой для таких идентичных локусов (**изолокусов**), продуцирующих электрофоретически

идентичные ферменты, является дупликация генов, чаще всего связанная с полиплоидией, а у лососевых рыб - с тетраплоидией (Wright et al., 1983; Allendorf, Thorgaard, 1984). Могут быть полиморфными и один и оба локуса. В обоих случаях электрофоретические фенотипы, кодируемые изолокусами, осложняют и генетическую интерпретацию, и статистический анализ обнаруженной генотипической изменчивости. Одна из подобных сложностей состоит в том, что практически невозможно отличить аллельные продукты изолокусов, когда полиморфны два (или более) локусов. Данная проблема иллюстрируется на рис. 5.3.2, на котором изображена изменчивость фенотипов двух изолокусов, когда один из пары мономорфен, а другой полиморфен.

ФЕНОТИПЫ	**ГЕНОТИПЫ**			**Состав субъединиц**
	A_1A_1	**A_1A_2**	**A_2A_2**	
Мономер				a, b a'
Димер				aa, ab, bb aa', a'b a'a'
Тетрамер				aaaa, bbbb, aaab, aabb, abbb aaaa', a'bbb, aaa'b, aa'bb aaa'a', aa'a'b, a'a'bb, aa'a'a'. a'a'a'b a'a'a'a'

*Рис. 5.3.2. Вид ферментных фенотипов, если экспрессируются изолокусы. Особи являются гомозиготными и гетерозиготными по локусам, кодирующим мономерные, димерные и тетрамерные белки. Один из локусов является полиморфным (с аллелями A_1 и A_2, кодирующими субъединицы **a** и **a'**, соответственно), а второй локус – мономорфен, кодируя одну субъединицу (**b**), у которой электрофоретическая подвижность идентична подвижности субъединицы **a** (По Аттер и др., 1987).*
Fig. 5.3.2. Electrophoretic phenotypes when isoloci are expressed. Individuals are homozygous and heterozygous at loci coding for monomeric, dimeric, and tetrameric proteins: one locus is polymorphic (with alleles A and A' resulting from subunits a and a', respectively); and a second locus is monomorphic, coding for a subunit (b) with an electrophoretic mobility identical to that of subunit (a) (From Utter et al., 1987).

Дополнительное осложнение может предоставлять альтернативный сплайсинг, механизм, который открыт недавно (рис. 5.3.3). Процессинг, участвующий в таком варианте сплайсинга, потенциально представляет собой регуляторный этап в экспрессии гена. Так, к примеру, известно немало случаев, когда интроны, представленные в пре-мРНК, синтезированной с одного и того же гена, вырезались несколькими способами. Это ведет к формированию разных наборов экзонов в зрелой мРНК. Данный процесс получил название **альтернативный сплайсинг**. Альтернативный сплайсинг ведет к формированию группы мРНК, которые при трансляции, дают близкие белки, называемые в протеомике **изоформами**. По определению, данному выше для генетически детерминированных ММФФ, они могут называться также изоферментами. Случаи альтернативного сплайсинга описаны у многих живых форм, начиная от вирусов, дрозофилы и заканчивая человеком. Однако остается открытым вопрос, все ли описанные случаи изолокусов обусловлены дупликациями генов или часть из них - это продукты альтернативного сплайсинга и комбинирования близких экзонов. Чтобы разобраться в этом, необходимо провести секвенирование соответствующих генов и выяснить характер их расщепления или наследования в потомстве.

Приведенные примеры отклонений от кодоминантной экспрессии ферментных генов, по-видимому, не столь широко распространены среди высших животных, полиплоидия у которых ограничена хромосомным механизмом определения пола. Однако насколько широко

распространены дупликации ферментных локусов у беспозвоночных животных неизвестно. Из многолетней собственной практики работы с морскими беспозвоночными, можно заключить, что дуплицированные локусы у них встречаются не часто. Этому факту может быть два объяснения: 1) дупликации у них действительно генерируются редко или 2) дупликации обычны, но благодаря диплоидизации их частота невелика. Еще менее известно в отношении распространенности альтернативного сплайсинга в генерации изоферментов.

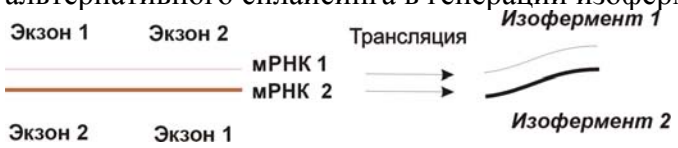


Рис. 5.3.3. Иллюстрация альтернативного сплайсинга - нового механизма создания разных мРНК продуктов (мРНК 1 и мРНК 2) и соответствующих изоформ (Изоформа 1 и Изоформа 2).
Fig. 5.3.3. Alternative splicing, the mechanism for the creation of new alleles (Gene 1 & Gene 2) and their products, the isoforms or isozymes (Isoform-1 & Isoform-2.

Перспективы и ограничения электрофоретического подхода для генетического исследования

Принципы, изложенные выше, для обнаружения генотипической изменчивости на основе электрофоретической изменчивости белков были в свое время признаны как прорыв для генетики природных популяций (Левонтин, 1978; Hartl, 1980).Превосходные свойства гелевого электрофореза в обнаружении генетической изменчивости дополняются относительной легкостью получения больших выборок генотипированных особей. Образцы белков посредством экстракции из тканей могут быть получены легко и в короткое время. В одном геле можно фракционировать одновременно много образцов и после электрофореза сам гель разрезается на продольные полосы. Эти срезы окрашиваются отдельно на различные ферменты или не ферментные белки, для того чтобы промаркировать различные локусы. Для опытного исследователя, имеющего одного технического помощника, вполне доступно проанализировать до 200-300 особей в сутки. При крахмально-гелевом электрофорезе по меньшей мере 6 локусов может быть выявлено в одном эксперименте от каждой особи, поскольку гель можно разрезать на 6 и более полос и каждый срез можно окрашивать на различные ферменты. Часто удается выявить больше, чем один локус с каждого среза, так как при окрашивании на отдельный фермент могут экспрессироваться более двух локусов.

Некоторые ограничения методики электрофореза белков также очевидны (Hubby, Lewontin, 1966, Lewontin, 1974). Так, наличие аминокислотных замен, оценивается не прямо, а косвенно по изменению суммарного заряда белка. Однако, (1) не каждая замена основания ведет к замене аминокислоты и (2) если замена аминокислоты произошла, это не обязательно ведет к изменению свойств белка, которые может обнаружить электрофорез. Было оценено, что примерно 25-30% аминокислотных замен обнаруживаются посредством методик электрофореза, применяемым в большинстве лабораторий (Shaw, 1965; Nei, Chakraborty, 1973; Lewontin, 1974), что согласуется с ожидаемой долей (25%), так как 5 из 20 аминокислот имеют заряд. Близкие цифры были получены для варьирующих условий электрофореза (Marshall, Brown, 1975). Однако иногда экспериментально обнаруживаемая доля мутаций по белковым генам достигает 50% (Ramshaw et al., 1979, Neel et al., 1986), а сам темп мутирования – высокой частоты (Алтухов, 1989: см. также Лекцию 6). Имеются некоторые другие ограничения. В частности, очевидно, что идентичность по электрофоретической подвижности не обязательно означает идентичность последовательности нуклеотидов в ДНК. Таким образом, степень сравнимости данных, например данных гетерозиготности, зависит от методики генотипирования (электрофоретические данные сравнимы, в основном, только между собой), что справедливо также в отношении любых генетических данных, полученных на основе фенотипов (Allendorf, 1977). В связи с пониманием возможной гетерогенности электрофоретически выявляемых аллелей, появился термин **электроморфы** (King,

Ohta, 1975; Allendorf, 1977). Экспериментальные данные подтвердили, что электрофоретически выявляемые аллели имеют много больше нуклеотидных замещений, чем аллели, выявленные методами анализа непосредственно ДНК (см. Глава 13).

Экспрессия генов в эмбриогенезе и онтогенезе

В результате дифференциальной активности генов формируются различные специализированные клеточные линии, а на их основе позже формируются ткани и органы. Исследование генетики индивидуального развития долгое время базировалось на исследовании морфологических мутаций в лабораторных условиях, в первую очередь в линиях дрозофилы (Корочкин, 1999; Жимулев, 2002). Морфологические признаки формируются поздно в онтогенезе, поэтому только с развитием методов биохимической и молекулярной генетики стал возможен анализ ранней экспрессии генов.

Давайте вспомним для начала несколько общих черт развития, в частности развития яйца. На начальной фазе развития женских репродуктивных органов происходит, так называемый, большой рост овоцитов, а потом мейоз (см. Лекцию 2). В фазе роста в овоците наблюдаются многочисленные «ламповые щетки», структуры указывающие на интенсивынй синтез мРНК, рРНК, рибонуклеопротеинов и других белков; мРНК при этом переходит в «замаскированную» форму – информасому. В зрелых овоцитах синтетические процессы практически прекращаются. В зиготе сразу после оплодотворения синтез также не происходит. Это обусловлено быстрыми клеточными делениями и короткими активными фазами – интеркинезами. В целом, до фазы гаструлы отцовские гены не активны. Другими словами, новый организм на первых этапах вынужден жить за счет материнских генных продуктов, накопленных про запас. Это знание биологии развития было получено в основном благодаря методикам биохимической генетики при анализе экспрессии ферментных генов в эмбриогенезе.

Экспрессия генов у самцов (M) и самок (F) в онтогенезе различна. Как мы заключили выше, эти различия запрограммированы спецификой онтогенеза у разных полов. Тем не менее, сам факт, что M-гены проявляются позже, был установлен экспериментально с применением методики электрофореза ферментов. При исследовании межвидовых гибридов различия экспрессии M - и F-родителей особенно показательны. В подавляющем большинстве многих проведенных скрещиваний M-гены экспрессируются позже F-генов и у гибридов (Нейфах, Лозовская, 1984). Например, в скрещивании между речным (M) и озерным (F) гольцами, два локуса ***LDH-B**** и ***GPDH****, которые отличаются подвижностями фиксированных аллелей, начинают первыми экспрессию F-изоферментов (Кирпичников, 1979).

Тканевая экспрессия ферментных локусов в развитии

Ферменты и другие белки дифференцированно экспрессируются в эмбриогенезе животных. Имеются три группы ферментов: (1) ферменты представлены во всех органах и тканях и присутствуют в течение всей жизни, (2) ферменты функционируют лишь в эмбриональных фазах, (3) ферменты появляются только на определенном этапе жизненного цикла и присутствуют лишь в некоторых тканях. (1) К первой группе относятся ДНК-, РНК-полимеразы, аденозинтрифосфатазы, АТФ-синтаза и другие ферментные и не ферментные белки, осуществляющие фундаментальные функции клетки. (2) Гемоглобин (Hb) – это наиболее известный представитель второй группы. Так, у человека в молекуле гемоглобина 2 субъединицы γ-Hb образуют тетрамер вместе с парой β-Hb субъединиц. После рождения субъединицы γ-Hb заменяются α-Hb полипептидными цепями. Аналогичный механизм работает, например, у лососевых рыб. Так, из 14-18 гемоглобиновых фракций, часть представлена только на стадии ранней личинки и после выклева отсутствуют. (3) Третья группа является самой представительной. Специфические спектры ГФИ присутствуют в нескольких тканях рыбы *Hypentelium nigricans*, как показано ранее (Рис 5.3.1). В мышечной ткани экспрессирутся практически только локус ***GPI-B*,*** тогда как в сердечной мышце и гонадах функционируют три гена, в остальных семи тканях - только два гена (Рис 5.3.1). Аналогично, у другого вида рыб *Lepomis cyanellus* ферменты ГФИ, МДГ (малатдегидрогеназа) и КК

(креатинкиназа), кодируются дуплицированными локусами, и один изофермент присутствует в овоцитах и далее в большинстве тканей, а второй появляется только после выклева и работает исключительно в мышечной ткани (Ferris, Whitt, 1979). Время начала экспрессии ферментных локусов совпадает у очень отдаленных таксонов: роды *Lepomis* и *Catostomus* представляют разные отряды рыб. Это заставляет думать, что генетическая регуляция онтогенеза очень сильно подвержена стабилизирующему отбору.

Функциональные различия изоферментов

Прежде, чем рассмотреть этот вопрос, необходимо сделать общее замечание о том, что дупликация и диверсификация функции генов – это взаимосвязанные события (Оно, 1973). Примеры имеются для лососевых рыб, и, в частности, для форели. Так, у них локус ***LDH**** дуплицирован и возникли последовательно три локуса: ***LDH-A*, LDH-B**** и ***LDH-C****. У этого тетрамерного фермента гомомерные изоферменты А4, В4, С4 значительно различаются по ряду свойств. А4 является наиболее термоустойчивым, наиболее активным в анаэробных условиях и наиболее устойчивым к ингибиторам. У С4 все эти свойства инвертированы, а В4 – обладает промежуточными свойствами. Изоферменты лейцинаминопептидазы (LAP) представляют другой хороший пример. Показано (Koehn, Innes, 1989), что его изоферменты обладают различной активностью при различной солености воды. Изоферменты ГФИ у *Lepomis* отличаются теплоустойчивостью. У различных фракций гемоглобинов угря были обнаружены отличия в сродстве с кислородом. Подробности на эту тему можно найти в обзорах (Салменкова, 1973, Хочачка, Сомеро, 1988).

Главные выводы из обсуждения материалов, представленных в разделах 5.2-5.3, таковы:
1. Изоферменты, а точнее аллоферменты, являются очень хорошими маркерами генетической изменчивости, несмотря на ряд ограничений.
2. Изоферменты являются наилучшими маркерами для выявления дифференциальной тканевой активности генов.
3. Изоферменты позволяют исследовать время активации генов в онтогенезе и особенности их экспрессии.

5.4. ЭВОЛЮЦИЯ РЕГУЛЯЦИИ ГЕНОВ

Взаимодействие генов

Имеется несколько форм взаимодействия генов. В предыдущих материалах мы встречали термин доминантность и кодоминантность, относящиеся к внутрилокусным взаимодействиям, эпистаз и плейотропию, которые объясняют межлокусные эффекты. В предыдущем разделе этой лекции была кратко подчеркнута роль, которую играет регуляция экспрессии и взаимодействие генов в индивидуальном развитии. Это очень большие разделы генетики индивидуального развития, как науки, и, в целом, они находятся за пределами основных задач курса. Тем не менее, представляется важным напомнить специально о большой сложности различных взаимодействий при формировании признаков, особенно таких сложных признаков, как морфологические или поведенческие признаки. Термин **взаимодействие генов** часто используется для обозначения идеи того, что несколько генов влияют на определенный признак. Это не значит, однако, что два или более генов, или их продуктов, обязательно взаимодействуют прямо, определяя конкретный фенотип. Скорее клеточные функции множества генных продуктов вносят совместный вклад для появления некоего суммарного (общего) фенотипа. Например, развитие органа, такого как многофасеточный глаз у насекомых, является предельно сложным и ведет к структуре с множественными фенотипическими проявлениями; наиболее просто они описываются характерным размером, формой, плотностью, цветом. Формирование глаза, например у дрозофилы, может быть представлено, как сложный каскад актов индивидуального развития. Этот процесс представляет пример **эпигенеза**, как онтогенетической концепции. Согласно этой

концепции каждый последующий шаг развития увеличивает сложность формируемого признака (органа) и находится под контролем и влиянием одного или более генов, действующих поэтапно.

С другой стороны, как упоминалось в Лекции 3, у многих высших организмов имеются различные жизненные формы или экотипы (биотипы), которые появляются на основе дифференциальной экспрессии одного и того же базисного генома в различных условиях внешней среды. Взаимодействие генотип – среда вкратце будет затронуто в Главах 11 и 12. Здесь мы схематично рассмотрим некоторые изменения, существующие в эпигенезе у простых и более сложных организмов. Прежде, чем рассматривать более сложные признаки, следует обратить внимание на простые биохимические признаки, такие как различные уровни активности ферментов. Это количественный признак и интересно понять, есть ли какая то генетическая регуляция на этом уровне вообще?

Генетические корреляции активности ферментов

Ферменты кодируются хорошо известными структурными генами и используя сбалансированные хромосомы и другие генетические средства, возможно проанализировать вклад структурной изменчивости и окружения, или модификаторов генов, на уровне активности ферментов. В серии работ на плодовой мушке, проанализирована генетическая изменчивость в активности 23 ферментов в серии хромосомнозамещенных линий для второй и третьей хромосомы (Laurie-Ahlberg et al. 1980; 1985; Laurie-Ahlberg, Vilton, 1982). В хромосомно-замещенных линиях поддерживалась генетическая идентичность всего набора хромосом, за исключением одной хромосомы. Схема эксперимента достаточно сложна и включала тестирование каждой линии в течение многих дней и их содержание в повтороностях. Сложная схема необходима для того, чтобы определить являются ли различия между линиями генетически обусловленными или связаны с эффектами содержания и другими экспериментальными ошибками. Существенные различия в активности были обнаружены для всех 23 ферментов, для второй и третьей хромосомнозамещающих серий. Исключение составила ТФИ (триозофосфатизомераза) для серии второй хромосомы. Принципиально сходные результаты были получены при представлении активностей как скорости реакции на мушку, на единицу веса или на белковую субъединицу. Была также проанализирована корреляция активностей для пар 23 ферментов и обнаружена тенденция к положительной генетической корреляции в активности ферментов (Wilton et al., 1982).

Сильная положительная корреляция активностей повторно проявлялась для двух ведущих ферментов пентозо-фосфатного шунта – для глюкозо-6-фосфатдегидрогеназы (Г6ФГД) и 6-фософоглюконатдегидрогеназы (6ФГД). Этот участок метаболизма регулирует поток углерода в цепи, начиная с глюкозо-6-фосфата через НАДФН, что является очень важным для многих биосинтетических процессов, включая липидный синтез. Оказалось, что плодовые мушки, у которых отсутствует активность либо Г6ФДГ, либо обоих этих ферментов, могут выживать. Но мушки, у которых нет активности только 6ФГД, не выживают. Это предполагает, что промежуточный продукт, создаваемый 6ФГД, может аккумулироваться до летального уровня, как это происходит со многими врожденными аномалиями метаболизма у человека. Очевидно, что в частном случае метаболизма у дрозофилы активности ферментов Г-6-ФДГ и 6ФГД пентозо-фосфатного шунта подвергаются координированному контролю и адаптивно (селективно) регулируются. В других случаях обнаруженная положительная корреляция активности может быть просто обусловлена эффектами использования общего субстрата. Например, ферменты Г-6-ФДГ, ГФИ, ФГМ (фосфоглюкомутаза) и ГК (гексокиназа) используют как субстрат глюкозо-6-фосфат и все обнаруживают положительную зависимость активностей. Направление и сила корреляций для ряда ферментов сходны в географически изолированных популяциях, и, кроме того, зависят от принадлежности к конкретному типу из двух хромосомных линий (вторая или третья хромосомы).

В нашем собственном исследовании были обнаружены различия активности ферментов, которые связаны с биотипической и таксономической дифференциацией (Картавцев, Мамонтов, 1983; Картавцев и др., 1983, 2002), а также с совместной регуляцией дуплицированных локусов

(Картавцев и др., 1985). Различия между жизненными формами (биотипами) и различия между молодыми видами по уровню активности многих ферментов велики и статистически несомненны, тогда как различия между «хорошими», давно сформированными видами, либо отсутствуют, либо доля различающихся ферментов невелика (Картавцев, Мамонтов, 1983; Картавцев и др., 1983, 2002). Данные результаты могут рассматриваться в поддержку возможной роли биотипической изменчивости в формировании новых видов; благодаря случаю тот или иной биотип может стать фиксированным и приобрести со временем репродуктивную изоляцию от родительского вида. Хотя, по-видимому, для такого видообразования решающую роль должна все равно играть изоляция и прекращение потока генов (см. Глава 7).

Не менее важными, чем эволюционные последствия, в этих данных являются синхронные изменения активностей по типу «включения - выключения», соответствующие резкому переходу в новое состояние особей, составляющих данный биотип. Этот включающий механизм приводит у рыб, как правило, к необратимым изменениям фенотипа. Жилые формы тихоокеанских лососей крайне редко становятся проходными и наоборот – проходные лососи ни когда не превращаются в озерные или речные формы (Коновалов, 1980). Можно привести много таких примеров для других групп животных, включая насекомых. Бактерии тоже имеют изменчивость, сходную с жизненными формами, но много проще. При пессимальных условиях они формируют споры, тогда как в норме – обычные клетки. Как отмечалось (Глава 4), в разных фазах развития бактериальных культур функционируют разные наборы генов. Можно заключить, что эпигенетическая регуляция эволюционировала, усложняясь с увеличением сложности организмов. Для всех организмов не известным остается, какой конкретный механизм запускает и переключает работу генов при формировании разных биотипов. По аналогии с онтогенетической регуляцией можно думать, что здесь также работает каскадный принцип, но возможно и что-то иное. Все эти захватывающие вопросы требуют научной проработки.

5.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 5

Предполагается рассмотреть три темы.

1. Гетерозиготность: наблюдаемая – Hobs и ожидаемая – Hexp.
2. Оценки генетических расстояний.
3. Программный пакет BIOSYS. С помощью пакета получить значения Hobs, Hexp и расстояний.

Средняя гетерозиготность (на локус/ на особь).

$H = \Sigma^{L}_{i=1} h_k / Lh_k = 1 - \Sigma^{m}_{i=1} p_i^2$, h_k – ожидаемая гетерозиготность (*Hexp=Hs*), p_i – частота *i*-того аллеля;

L – число локусов. $Hobs = h_j / n$, h_j – фактическое число гетерозигот, n – число особей.

Генетические расстояния. Стандартное расстояние Нея (D_N)

$D_N = - \ln I.I$ – нормализованная генетическая идентичнось по набору локусов. Для локуса j она равна: $I_j = (\Sigma^{k}_{i=1} x_i y_i) / \sqrt{\Sigma^{k}_{i=1} x_i^2 \Sigma^{k}_{i=1} y_i^2}$. Для всего набора локусов I равна:

$I = Jxy / \sqrt{JxJy}$; где Jxy, Jx, Jy – средние арифметические для сумм: $\Sigma^{k}_{i=1} x_i y_i$, $\Sigma^{k}_{i=1} x_i^2$, $\Sigma^{k}_{i=1} y_i^2$. Здесь, x_i и y_i – частоты аллелей в популяции (таксоне) *x* и *y*, k – число аллелей для всех попарных сравнений.

Минимальное расстояние Нея *(Dm).*

$Dm = 1 / k \, \Sigma^{k}_{i=1} d_m d_m = [(\Sigma^{k}_{i=1} x_i^2 + \Sigma^{k}_{i=1} y_i^2) / 2] - \Sigma^{k}_{i=1} x_i y_i$. d_m – расстояние по одному локусу, **k** – число локусов.

ГЛАВА 6. ГЕНЫ, МУТАЦИИ И ЭВОЛЮЦИЯ

ГЛАВНЫЕ ВОПРОСЫ:

6.1. Введение в проблематику исследования мутаций.
6.2. Структура и функции генов.
6.3. Типы мутаций на уровне ДНК.
6.4. Мутации и замены аминокислот.
6.5. Темп мутаций.
6.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 6.

РЕЗЮМЕ

1. Мутации ведут к резким изменениям нормального фенотипа или к изменчивости и, таким образом, поставляют материал для эволюции и, в частности, – для генетических исследований.
2. По определяемой ими функции гены можно разделить на две группы: кодирующие белки или структурные гены и РНК-кодирующие гены, осуществляющие регуляцию. Регуляторными бывают и структурные гены, поэтому между ними нет четкой границы.
3. Нуклеотидные замены могут быть подразделены на две различные группы: транзиции и трансверсии. Транзиции – это замены пуринового нуклеотида на другой пуриновый нуклеотид (A⇔G) или пиримидина на пиримидин (T⇔C). Трансверсия – это замены нуклеотидов разных типов (G⇔T, A⇔C, C⇔G or A⇔T)**.**
4. Классическая генетика установила, что **темп мутаций,** затрагивающих фенотипические признаки внешней морфологии или вызывающих летальные эффекты у таких эукариотических организмов, как человек, дрозофила, кукуруза, ***составляет порядка 10^{-5} в расчете на локус на поколение***. ***Средний темп аминокислотных замен в типичном белке оценивается, как 4×10^{-7} на локус на поколение.*** Учитывая, что электрофорез выявляет только одну-четвертую часть всех аминокислотных замен, ***темп мутаций для белковых локусов, определяемых методом электрофореза, составляет около 10^{-7} на локус на поколение.***
5. Мутации играют принципиальную роль в **молекулярных часах**. Темп мутаций и время, прошедшее с момента изоляции генофондов пары таксонов, прямо определяют величины накопленных за время изоляции генетических расстояний между ними. Таким образом, ***зная величину генетического расстояния и темп мутаций, можно определить время дивергенции таксонов***.

Chapter 6. GENES AND MUTATIONS

SUMMARY

1. Mutations lead to alteration of a current phenotype or to variability, and hence provide a subject for genetic research. From this their importance for genetics and evolution of life in general is obvious.
2. In terms of their function, genes can be classified into two groups: protein-coding or **structural genes** and RNA-coding, **regulatory genes**. Structural genes may participate in regulation too, and in this having no principal difference with the rest of the genes.
3. Nucleotide substitutions can be divided into two different classes: **transition** and **transversion**. **Transition** is a substitution of a purine for another purine (A⇔G) or the substitution of a pyrimidine for another pyrimidine (T⇔C). **Transversions** are inverse types of nucleotide substitutions, when a purine is substituted by a pyrimidine and vise versa (G⇔T, A⇔C, C⇔G or A⇔T).

4. Classical geneticists has determined that the **rate of mutations,** that alter phenotypic characters or induce lethal effects, ***is of the order of 10^{-5} per locus per generation*** in higher organisms such as man, *Drosophila,* and corn. ***The mean rate of amino acid substitution for average polypeptide is estimated to be 4×10^{-7} per locus per year.*** Due to the evidences that electrophoresis detects only about one-quarter of all amino acid changes, ***the rate of amino acid substitutions detectable by electrophoresis is estimated to be about 10^{-7} per locus per year.***
5. Mutations play the principal role in **molecular clocks**. Mutation rate and time period since gene pool separation of the taxonomic pair, directly determine the value of accumulated genetic distances between these taxa during the time of separation. Thus, ***when the value of the distance and mutation rate are known, the time of taxa separation can be calculated***.

6.1. ВВЕДЕНИЕ В ПРОБЛЕМАТИКУ ИССЛЕДОВАНИЯ МУТАЦИЙ

В этой лекции мы будем иметь дело с мутациями, их эффектами и близкой проблематикой. Главное внимание будет сосредоточено на природных видах. Мутации дают новые варианты обычного фенотипа, таким образом, формируя изменчивость, являющуюся основой генетического исследования. Из этого ясно, сколь велика их роль для генетики и эволюции в целом. У одних организмов мутации выявляются легче, чем у других. Это зависит от особенностей их биологической организации и, в случае экспериментальных исследований, от способности содержаться в искусственных условиях в лаборатории или других местах. Поэтому излюбленными объектами при исследовании мутаций являются вирусы и бактерии, или эукариоты – дрожжи, некоторые растения, дрозофила и мышь. Мутация – это скачкообразное изменение в генах, ведущее или не ведущее к изменению фенотипа.

Классификация мутаций

Спонтанные и индуцированные мутации. Мутации классифицируются по различным схемам, которые могут частично перекрываться. Все мутации можно разбить на две группы – спонтанные и индуцированные. **Спонтанные мутации** – это те, которые возникают периодически в природе, без какой-либо видимой причины. **Индуцированные мутации** вызывают искусственно, посредством воздействия определенного агента – химического вещества или излучения. Одним из первых, кто продемонстрировал индуцированные мутации в опытах по облучению дрозофилы рентгеновскими лучами, был Мёллер (H. Muller).

Гаметические и соматические мутации. При рассмотрении эффектов мутаций у эукариот важно различать, где происходят мутации в соматических клетках или в половых клетках – гаметах. Мутации в гаметах имеют большее значение для судьбы организма, так как передаются потомству. Доминантные аутосомные мутации будут проявляться фенотипически уже в первом поколении. Сцепленные с полом мутации могут проявляться в гемизиготном состоянии у самцов, таким образом, также проявятся уже в F_1. Аутосомные же мутации, при рецессивности, могут маскироваться у гетерозигот и не сразу проявятся в потомстве.

Другие категории мутаций. На основе эффектов, оказываемых на организмы, выделяют разнообразные типы мутаций. Одна и та же мутация может быть отнесена к разным категориям. В принципе, число различных категорий может соответствовать разнообразию существующих групп признаков. Так, можно выделить: (1) **морфологические мутации**, изменяющие, например, форму семян у душистого горошка – гладкие, морщинистые, (2) пищевые или **биохимические мутации**, которые вызывают изменения в метаболических цепях, (3) **поведенческие мутации**, меняющие поведение особей, (4) **регуляторные мутации**, влияющие на активацию и инактивацию генов, (5) мутации жизнеспособности можно классифицировать, как, **летальные, нейтральные, благоприятные и вредные мутации**, (6) наконец, большинство перечисленных мутаций может существовать как **условные мутации**. Допустим, что мутация присутствует в геноме организма,

но она может не проявляться в определенных условиях. Наиболее изученными из такого рода мутаций являются температурно-чувствительные мутации, обнаруженные у большого числа организмов. При определенной, «разрешенной» температуре мутантный генный продукт функционирует как нормальный, но теряет свою функциональность при другой, «запрещенной» температуре. В заключение назовем еще одну группу мутаций, - это так называемые **точковые** мутации, которые вызывают небольшие изменения в молекуле ДНК. Они противопоставляются хромосомным мутациям, которые, как отмечалось, лучше классифицировать как хромосомные перестройки.

Обнаружение мутаций и их причины

Проявление мутаций зависит от многих факторов, в том числе от самого типа организма. Соответственно, методы обнаружения мутаций зависят от специфики изучаемых объектов. Методы, которые пригодны для бактерий, совсем не обязательно хороши для человека или дрозофилы. Обнаружение мутаций у бактерий и грибов осуществить наиболее просто. Микроорганизмы из природы, «дикого» типа, при искусственном содержании требуют только минимальной среды, синтезируя для своей жизнедеятельности все остальные компоненты. Эти организмы называют **прототрофными** (прототрофами). Поэтому мутантными особями являются те, которые не могут расти на минимальной среде, и называются они **ауксотрофными** (ауксотрофами). Биохимические или пищевые мутанты могут быть определены и выделены методом исключения, на основании их способности расти в полной среде с необходимыми компонентами и отсутствии роста в минимальной среде. Мутанты могут быть обнаружены визуально по росту бактериальных колоний на агар-агаре. Далее применяются по необходимости различные другие методы молекулярной генетики, чтобы понять структуру исследуемого гена, характер самой мутации и точно картировать ее расположение.

В чем причины мутаций? Мутации – это сложный процесс и обсуждение разнообразия причин их обусловливающих лежит вне основных направлений этого курса. Тем не менее, необходимо привести хотя бы некоторые общеизвестные причины (Жимулев, 2002, с. 213-216; Klug, Cummings, 2002, p. 283-285). Отметим, что химической основой очень многих точковых спонтанных мутаций является замена обычных нуклеотидов их метилированными аналогами, которые плохо репарируются. Таутомерные переходы в энольную и иминоформы в молекулах нуклеотидов также являются причинами увеличения их нестабильности. Следствием таких переходов может быть образование атипичных A-C и G-T пар, ведущее к появлению спонтанных замен в ДНК. Очень велика роль механизмов репарации ДНК и РНК в возникновении мутаций. Так, ошибки в работе ДНК-полимеразы и химическая модификация цитозина дают спонтанные мутации. Ошибки в ходе полимеризации ДНК обычны и имеют частоту около 10^{-5}. Корректорная 3’-5’-экзонуклеазная активность ДНК полимераз снижает эту частоту до 10^{-10} в расчете на нуклеотид. Репаративные механизмы далее еще уменьшают эту частоту, но, тем не менее, определенная часть изменений, возникших в генах и популяциях особей, остаются, что и обеспечивает спонтанный мутагенез.

6.2. СТРУКТУРА И ФУНКЦИЯ ГЕНА

Мы рассмотрели концепцию гена и его действие в предыдущих лекциях. Хотя наше понимание функционирования генов в плане их участия в морфогенезе и регуляции остается весьма ограниченным. Прежде, чем мы рассмотрим темп появления мутаций и их распространение, необходимо напомнить, как устроен генетический код и с какими типами генов обычно имеет дело исследователь.

Типы генов и универсальность кода

Согласно своей функции гены можно разбить на две группы: структурные гены, кодирующие белки, и РНК-кодирующие гены. **Структурные гены** через процесс трансляции

обеспечивают синтез белков и их производных. РНК-кодирующие гены производят тРНК, рРНК и некоторые другие, типа мяРНК (малые ядерные РНК). Эти РНК являются конечными продуктами соответствующих генов. рРНК входят в состав рибосом и являются частью процесса трансляции и синтеза белка, а тРНК, хотя также участвуют в трансляции, являются посредниками, обеспечивающими превращение информации мРНК в аминокислотную последовательность в белке (см. Глава 3). Функции мяРНК не во всем понятны, но ясно, что они важны для транскрипции и процессинга. Часть структурных генов и большинство РНК-кодирующих генов участвуют в различных регуляторных функциях: регуляция репликации, транскрипции, трансляции и жизненного цикла, включая клеточное деление и онтогенез. Они называются **регуляторными генами**, хотя нет четкой границы между ними и структурными генами, как мы видим.

Генетический код, в целом, является универсальным для вирусов, прокариот и эукариот. Небольшие вариации имеются. Так, в митохондриях и хлоропластах некоторые кодоны отличаются от ядерных. Код для генов клеточного ядра приведен в Таблице 6.2.1. Каждому трехбуквенному обозначению кодона соответствует трехбуквенное обозначение для аминокислот (Табл. 6.2.1). Имеется 4^2=64 возможных кодонов, образующихся при комбинировании четырех различных нуклеотидов: аденина (A), гуанина (G), урацила (U) и цитозина (C) в трехчленные единицы-слова, называемые триплетами или кодонами. Три из них (UAA, UAG, UGA) являются только стоп кодонами или прерывателями синтеза (**нонсенс кодоны**), не обеспечивая никакого кода для аминокислот. Остальные кодоны (**смысловые кодоны**), всего таких 61, определяют ту или иную аминокислоту.

Таблица 6.2.1. Кодирующий словарь *(По Klug, Cummings, 2002)*
Table 6.2.1. The coding dictionary *(From Klug, Cummings, 2002)*.

Первая буква кодона	Вторая буква кодона				Третья буква кодона
	U	C	A	G	
U	Фен (Fen)	Сер (Ser)	Тир (Tyr)	Цис (Cys)	U
	Фен (Fen)	Сер (Ser)	Тир (Tyr)	Цис (Cys)	C
	Лей (Leu)	Сер (Ser)	Стоп (Stop)	Стоп (Stop)	A
	Лей (Leu)	Сер (Ser)	Стоп (Stop)	Трп (Trp)	G
C	Лей (Leu)	Про (Pro)	Гис (His)	Арг (Arg)	U
	Лей (Leu)	Про (Pro)	Гис (His)	Арг (Arg)	C
	Лей (Leu)	Про (Pro)	Гли (Gln)	Арг (Arg)	A
	Лей (Leu)	Про (Pro)	Гли (Gln)	Арг (Arg)	G
A	Иле (Ile)	Тре (Thr)	Асн (Asn)	Сер (Ser)	U
	Иле (Ile)	Тре (Thr)	Асн (Asn)	Сер (Ser)	C
	Иле (Ile)	Тре (Thr)	Лиз (Lys)	Арг (Arg)	A
	Мет (Met)	Тре (Thr)	Лиз (Lys)	Арг (Arg)	G
G	Вал (Val)	Ала (Ala)	Асп (Asp)	Гли (Gln)	U
	Вал (Val)	Ала (Ala)	Асп (Asp)	Гли (Gln)	C
	Вал (Val)	Ала (Ala)	Глу (Glu)	Гли (Gln)	A
	Вал (Val)	Ала (Ala)	Глу (Glu)	Гли (Gln)	G

Примечание. U – урацил, C – цитозин, A – аденин, G – гуанин. AUG кодирует метионин, инициаторную аминокислоту в большинстве полипептидов (выделен цветом). Все другие аминокислоты, за исключением триптофана, который кодируется только триплетом UGG, представлены от двух до шести кодонов. Триплеты UAA, UAG и UGA являются стоп кодонами.Поскольку имеется только 20 аминокислот, то ясно, что код является избыточным, то есть, более чем один кодон дает кодировку для одной и той

же аминокислоты. Кодоны, обеспечивающие образование одной и той же аминокислоты, называют **синонимичными кодонами**, а замены в них (мутации) – синонимическими.

Как отмечалось выше, код митохондриального генома слегка отличается, что проиллюстрировано для генов митохондрий млекопитающих (Табл. 6.2.2). В митохондриальном геноме кодон AUA, который кодирует изолейцин в ядерных генах, используется для шифрования метионина, UGA не является стоп кодоном, а кодирует триптофан. Наоборот, кодоны AGA и AGG – это стоп кодоны вместо кодирования аргинина. Генетический код дрожжей также несколько отличается от митохондриального кода (Табл. 6.2.2).

Таблица 6.2.2. Генетический код митохондриальных генов млекопитающих *(По Nei, 1987)*
Table 6.2.2. The genetic code for mammalian mitochondrial genes *(From Nei, 1987).*

Кодон	**Амино-кислота**	**Кодон**	**Амино-кислота**	**Кодон**	**Амино-кислота**	**Кодон**	**Амино-кислота**
UUU	*Phe*	*UCU*	*Ser*	*UAU*	*Tyr*	*UGU*	*Cys*
UUC	*Phe*	*UCC*	*Ser*	*UAC*	*Tyr*	*UGC*	*Cys*
UUA	*Leu*	*UCA*	*Ser*	*UAA*	*Ter*	*UGA*	*Trp*
UUG	*Leu*	*UCG*	*Ser*	*UAG*	*Ter*	*UGG*	*Trp*
CUU	*Leu(Thr)*	*CCU*	*Pro*	*CAU*	*His*	*CGU*	*Arg*
CUE	*Leu(Thr)*	*CCC*	*Pro*	*CAC*	*His*	*CGC*	*Arg*
CUA	*Leu(Thr)*	*CCA*	*Pro*	*CAA*	*Gin*	*CGA*	*Arg*
CUG	*Leu(Thr)*	*CCG*	*Pro*	*CAG*	*Gin*	*CGG*	*Arg*
AUU	*He*	*ACU*	*Thr*	*AAU*	*Asn*	*AGU*	*Ser*
AUC	*He*	*ACC*	*Thr*	*AAC*	*Asn*	*AGC*	*Ser*
AUA	*Met(Ile)*	*ACA*	*Thr*	*AAA*	*Lys*	*AGA*	*Ter(Arg)*
AUG	*Met*	*ACG*	*Thr*	*AAG*	*Lys*	*AGG*	*Ter(Arg)*
GUU	*Val*	*GCU*	*Ala*	*GAU*	*Asp*	*GGU*	*Gly*
GUC	*Val*	*GCC*	*Ala*	*GAC*	*Asp*	*GGC*	*Gly*
GUA	*Val*	*GCA*	*Ala*	*GAA*	*Glu*	*GGA*	*Gly*
GUG	*Val*	*GCG*	*Ala*	*GAG*	*Glu*	*GGG*	*Gly*

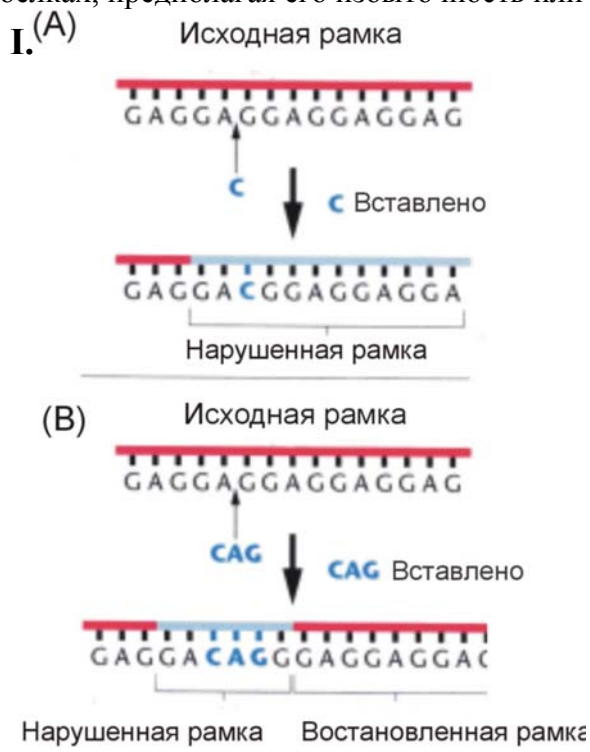
Примечание. Кодоны для дрожжей отличаются только немного. Отличия показаны в скобках.

Универсальность кода подтверждается многочисленными экспериментами, в том числе с применением рекомбинантных ДНК, в результате чего бактериальные гены успешно транскрибируются и транслируются в эукариотическом геноме и наоборот. Хотя, с увеличением новых экспериментальных данных, стало очевидным, что первоначальная уверенность в абсолютной универсальности кода уже не столь несомненна. В частности, у простейших *Paramecium* и *Tetrahymena,* кодоны UAA и UAG не являются терминирующими кодонами, как у всех эукариот в ядерных генах, а кодируют глутамин. Для одного из простейших прокариот, *Mycoplasma capricolum*, стоп кодон UGA используется для кодирования триптофана.

Триплетная природа кода

В 1950-е годы Гамов (Цит. по Ратнер, 1998), а позднее Брэннер с коллегами (Brenner et al., 1961) первыми предположили на основе теоретических соображений, что код должен быть трехбуквенным, так как три буквы смысловых «слов» представляют минимальное число знаков, которые необходимы для кодирования известных 20 аминокислот: $4^3 = 64$. Код по две буквы в слове, составленном из четырех нуклеотидов, представляет только $4^2=16$ уникальных кодирующих вариантов. Четырехбуквенный код может определять $4^4= 256$ слов, что является чрезмерной

избыточностью. Экспериментальные данные, обобщенные Криком и Уотсоном (F. Crick, T. Watson), подтверждали триплетную природу кода. Используя фаг Т4, они изучали **мутации рамки считывания** (**frameshift mutations**). ***Рамка считывания*** *– это начало и конец трансляции,* а интервал между стартовым и стоп кодоном называется открытой рамкой считывания, ОРС (Open Reading Frame, ORF). Мутации возникают как результат добавления или удаления одного или большего числа нуклеотидов в гене и соответственно в РНК, транскрибирумой с него. Приобретение или потеря одной или большего числа букв сдвигает рамку считывания в ходе трансляции. Однако если три замены вызывают мутацию, тогда рамка считывания восстанавливается (рис. 6.2.1, I и II). Такое поведение возможно, если код является триплетным. Проделанная работа также показала, что код не является пустым, а кодирует 20 аминокислот в белках, предполагая его избыточность или **вырожденность**.



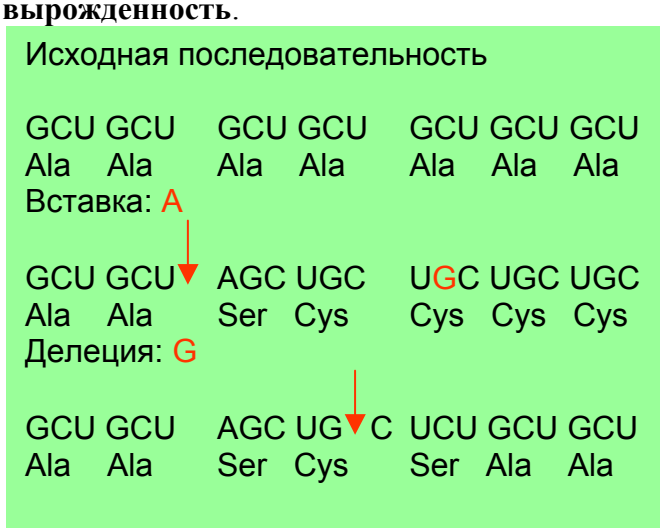


Рис. 6.2.1. Влияние эффекта мутации рамки считывания на последовательность ДНК, повторяющую триплетную последовательность GAG.

I. (A) Вставка единичного нуклеотида сдвигает соответствующую рамку считывания для триплета, (B) вставка трех нуклеотидов изменяет только два триплета, но рамка считывания затем восстанавливается (По Klug, Cummings, 2002).

II. Показан другой пример сдвига рамки считывания вследствие вставки и делеции.

Fig. 6.2.1. The effect of frameshift mutations on a DNA sequence repeating the triplet sequence GAG.

I. (A) The insertion of a single nucleotide shifts all subsequent triplet reading frames, (B) The insertion of three nucleotides changes only two triplets, but the frame of reading is then reestablished to the original sequence.

II. Another example in frame shift due to insertion and deletion (From Klug, Cummings, 2002 with adds).

Вырожденность кода и его неперекрываемость. Исходя из кодирующего словаря ясно, что код является вырожденным (Табл. 6.2.1), как и было предсказано в ранних работах. Почти все аминокислоты определяются двумя, тремя или большим числом кодонов. Три аминокислоты (серин, аргинин и лейцин) кодируются каждая шестью кодонами. Только триптофан и метионин детерминируются уникальными кодонами (Табл. 6.2.1). В генетическом коде кодон AUG кодирует метионин, но этот кодон также используется как инициирующий кодон. Поэтому каждая мРНК

имеет этот кодон в начале транскрипции и позднее он удаляется из цепи (см. Глава 3). Вырожденность кода имеет еще особенность в том, что значимость нуклеотидов в кодоне неодинакова. Наибольшую смысловую нагрузку несут первые два нуклеотида. Это ведет к некоторой свободе или «шатанию» кода. Эту особенность в виде гипотезы сформулировал Крик (Crick, 1966). Триплет или кодон определяет одну аминокислоту, следующий триплет, соответственно, кодирует последующую аминокислоту. То есть, код не имеет пробелов и не перекрывается; каждый кодон используется при прочтении ОРС последовательно один раз, начиная с фиксированной стартовой точки.

6.3. ТИПЫ МУТАЦИЙ НА УРОВНЕ ДНК

Замены, делеции, вставки и инверсии

Поскольку все морфологические, биохимические, физиологические и другие признаки контролируются генами, локализованными в ДНК, любые мутационные изменения этих признаков обусловлены изменениями в ДНК. Можно выделить четыре основных типа изменений в ДНК: (1) **замена** одного **нуклеотида** на другой в первичной последовательности, (2) **вставки** нуклеотидов в существующую последовательность (рис. 6.2.1, Ia-b), (3) **делеции** нуклеотидов из последовательности (рис. 6.2.1,II), (4) **инверсии** порядка последовательности нуклеотидов. Вставки, делеции и инверсии могут происходить для одного или нескольких нуклеотидов (см. примеры на рис. 6.2.1). Вставки и делеции могут сдвигать рамки считывания, ведя либо к частичному, либо к полному переосмысливанию кодирующей последовательности в гене (см. рис. 6.2.1).

Транзиции и трансверсии

Нуклеотидные замены (замещения) могут быть разделены на две категории: **транзиции** и **трансверсии**. Транзиция – это замена одного пурина другим пурином (A↔G) или пиримидина – другим пиримидином (T ↔ C). Замены обратного типа, то есть, пурина на пиримидин G↔T, A↔C, C↔G и A↔T называются трансверсиями (рис. 6.3.1).

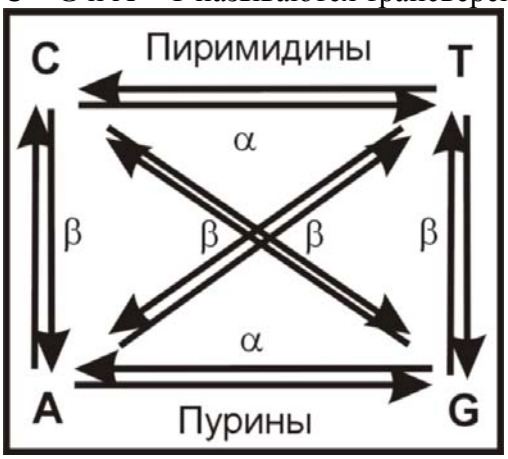


Рис. 6.3.1. Транзиционные, (A↔G и T↔C) и трансверсионные, (четыре других) категории нуклеотидных замен. Темп замен транзиций и трасверсий составляет, соответственно - α и β (По Nei, Kumar, 2000).

Fig. 6. 3.1. Transitional, α (A↔G and T↔C) and transversional, β (four others) nucleotide substitutions (From Nei, 1987).

Молекулярно-генетические данные показывают, что вставки и делеции встречаются довольно часто, особенно вне кодирующих участков ДНК. Число нуклеотидов во вставке, или вовлеченных в делецию, варьирует от одного или немногих до нескольких тысяч. Короткие вставки или делеции – это обычно ошибки репликации ДНК. Длинные вставки и делеции чаще всего являются результатом неравного кроссинговера или перемещения (транспозиции) фрагмента ДНК (см. Глава 16).

Преобладающие направления нуклеотидных замен

Имеется много различных генетических изменений, о которых говорилось ранее и те, что будут рассмотрены позже. Однако, если рассматривать только первичные последовательности ДНК, то большинством изменений являются нуклеотидные замены. При случайности возникновения мутаций для четырех нуклеотидов ДНК (A, T, C, и G), можно ожидать, что

трансверсии должны быть в два раза более частыми, чем транзиции (см. рис. 6.2.1, 4β / 2α). На практике, однако, транзиции случаются чаще, чем ожидается (Fitch, 1967; Vogel, 1972). Так, при исследовании псевдогенов, не функционирующих, «молчащих» генов, которые, как предполагается, возникли посредством дупликаций родственных генов (см. Глава 16), замены в них предполагаются нейтральными – то есть, случайными или в основном стохастическими, и зависят только от темпа мутаций (Nei, 1987, Chapter 13). В исследованиях Гойобори и Ли с соавторами (Gojobori et al., 1982; Li et al., 1984) выяснилось, что при сравнении псевдогенов с оригинальными генами частота транзиций (59.3%) много выше, чем ожидается (33.3%) (Табл. 6.3.1). Однако 4 возможных типа транзиций не одинаковы по частоте, так что доля замен C→ T выше, чем других типов. Большая частота C→ T замен, по-видимому, связана с метилированием цитозина в CG нуклеотидах в геноме, но возможны и другие причины. Наоборот, все трансверсии происходят с примерно равной частотой (Табл. 6.3.1). Биохимические причины неравных частот замен обсуждались в литературе (Topal, Fresco, 1976; Gojobori et al., 1982) и мы здесь на этом не останавливаемся подробнее. Для дальнейшего изложения важно лишь помнить, что темп мутационных замен для четырех нуклеотидов не равен.

Таблица 6.3.1. Относительный темп мутаций четырех нуклеотидов A, T, C и G в псевдогенах *(По Li et al., 1984)*

Table 6.3.1. Relative mutation rates among the four nucleotides A, T, C, and G in pseudogenes *(From Li et al., 1984)*

Мутантный нуклеотид	***Исходный нуклеотид***			
	A	**T**	**C**	**G**
A		4.4±1.1	6.5±1.1	20.7±2.2
T	4.7±1.3		21.0±2.1	7.2±1.1
C	5.0±0.7	8.2±1.3		5.3±1.0
G	9.4±1.3	3.3±1.2	4.2±0.5	

Примечание. Оценки темпа мутаций основываются на данных о нуклеотидных заменах в 16 псевдогенах млекопитающих.

6.4. МУТАЦИИ И ЗАМЕНЫ АМИНОКИСЛОТ

Поскольку аминокислотная последовательность полипептида определяется последовательностью нуклеотидов в структурном гене, то любая замена в аминокислотной последовательности вызывается мутацией, возникающей в ДНК. Однако мутационные изменения в ДНК не обязательно приводят к заменам аминокислот в силу вырожденности генетического кода и возможности синонимических замен. Экспериментальная методика обнаружения мутаций, как в случае электрофореза белков (см. Глава 5), также может вести к недооценке числа мутаций. Если замены нуклеотидов происходят случайно, то доля не синонимических мутаций может быть вычислена с использованием генетического кода. Если мы примем, что все кодоны имеют равную частоту в геноме, а темп замен приблизительно одинаков для всех четырех нуклеотидов, то доля не синонимических замен будет равна 71%, исключая нонсенс мутации (Nei, 1975). Таким образом, около 29% мутаций, возникающих на нуклеотидном уровне, не может быть обнаружено при анализе аминокислотных последовательностей. Эта цифра изменится не сильно, если мы сделаем более реалистичные предположения о вероятностях нуклеотидных замен и частотах кодонов.

Если воспользоваться генетическим словарем, приведенным в Таблицах 6.2.1 и 6.2.3, то можно обнаружить, что синонимические кодоны возникают из-за избыточности третьего положения в кодоне. Расчет показывает, что примерно 72% замен нуклеотидов в третьем положении дают синонимические кодоны. Аналогично можно обнаружить, что по первому

нуклеотиду синонимические замены происходят, но их доля меньше и составляет около 5%. Наконец, по второму нуклеотидному положению синонимичных замен не обнаруживается совсем.

Как подчеркивалось в предыдущих лекциях, в популяционной генетике и молекулярной таксономии белковые различия часто исследуются посредством электрофореза, но при этом выявляются лишь примерно 25-30% мутаций (Глава 5). Можно заметить, что чувствительность электрофореза ферментов в обнаружении генетической изменчивости можно заметно увеличить, применяя тепловую обработку (Ayala, 1975) и варьируя условия электрофореза (Singh et al., 1976; Ramshaw et al., 1979; Coyne, 1982; McLellan, 1984; Айала, 1984). Однако, это ведет к снижению скорости обработки выборок из популяций и, поэтому, эти усложненные методики не получили большого распространения. Потенциально подходит для анализа генетической изменчивости белков метод изоэлектрического фокусирования, который обладает большей разрешающей способностью (Whitney et al., 1985), но он не нашел применения при анализе ферментов, основных генных маркеров. Двумерный электрофорез – это другой подход, получивший развитие в генетике. При использовании этого метода, как пояснено в Лекции 3, суммарные экстракты, включающие многие белки, подвергаются электрофорезу поочередно в двух направлениях. Причем состав буферов может меняться в первой и второй фазе электрофореза, что потенциально позволяет выявить большое разнообразие белков, разложив их даже до пептидов, и также проанализировать значительное число особей. Так, Голдман с коллегами (Goldman et al.,1987), используя этот подход, выявили 383 различных полипептида у человека и других гоминид. В принципе, разрешающая способность варьирует от 200 до 10000 фракций (см. Глава 3, рис. 3.4.1). Однако доля полиморфных локусов, определяемая этим методом, ниже, чем при обычном электрофорезе (Aquadro, Avise, 1981; Ohnishi et al., 1983; McLellan et al., 1983; Айала, 1984). Данное обстоятельство связано, по-видимому, с тем, что при этом методе окрашиваются, в основном, мажорные компоненты образцов белков клеток, которые представлены структурными белками, такими, как актин, миозин, гистоны и т.п. Эти белки более сильно нагружены функционально и подвергаются нормализующему отбору, минимизирующему или устраняющему совсем их изменчивость (Kimura, 1983a; Nei, 1987).

6.5. ТЕМП МУТАЦИЙ

Прямые методы оценки

Темп спонтанных мутаций является одним из ключевых параметров в популяционной генетике, чем определяется важность получения его надежных оценок. Тем не менее, этот важный параметр не был оценен на основе молекулярных маркеров в достаточно большом количестве таксонов. Одна из трудностей получения представительных оценок состоит в низкой частоте мутаций, что требует анализа огромных выборок особей по нескольким маркерам. В классической генетике для нескольких эукариотических организмов (человек, мышь, дрозофила, кукуруза) спонтанный темп мутаций был оценен по морфологическим мутациям и мутациям жизнеспособности (летали) как величина порядка 10^{-5} на локус на поколение. Однако по этим данным ничего нельзя сказать о темпе мутаций на уровне ДНК. К тому же, большинство морфологических мутаций являются, как правило, вредными, не говоря о леталях, и, соответственно, мало влияют на эволюционную историю вида. Для сравнения, темп не вредных морфологических мутаций у ячменя был оценен, как не превышающий 8.6×10^{-9} на локус на поколение с 95%-ным доверительным уровнем (Kahler et al., 1984).

На основе методов биохимической и молекулярной генетики в последние годы появилось немало оценок темпа спонтанных мутаций в популяциях (Lewontin, 1974; Лимборская, 1981; Vogel, Motulsky, 1986; Nei, 1987; Алтухов, 1989). Масштабное исследование было проведено на дрозофиле Мукаи с соавторами (Mukai, Cockerham, 1977; Voelker et al., 1980). Эти авторы

анализировали спонтанные мутации на основе аллозимных данных для семи локусов у *D. melanogaster*, поддерживая 1000 линий в течение 211-224 поколений. Число гено-поколений (число проанализированных генов, умноженное на число поколений) достигло в этих опытах 3 111 598, а темп мутаций, по оценке других исследователей (Voelker et al., 1980), составил на локус на поколение 1.28×10^{-6} и 3.86×10^{-6}, соответственно, для мутаций затрагивающих подвижности и для нулевых аллелей. Позднее было обнаружено, что линии дрозофилы, с которыми работали эти авторы, несут гены-мутаторы, увеличивающие темп летальных мутаций почти в 10 раз. Осталось неясным, влияют ли эти гены на темп мутаций в аллозимных локусах или нет. В любом случае, полученные оценки показали, что темп мутаций на молекулярном уровне может быть ниже, чем темп некоторых морфологических и летальных мутаций.

Последнее заключение подтверждается данными исследования на ячмене (Kahler et al., 1984). Изучая электрофоретическую изменчивость пяти ферментных локусов, и получив, в целом, 841 260 гено-поколений и не обнаружив ни одной мутации, эти авторы сделали расчетное заключение, что темп мутаций должен быть не ниже 3.6×10^{-6} на локус на поколение, с 95%-ным доверительным уровнем. Масштабные исследования были проведены также на популяциях человека (Harris et al., 1974; Neel et al., 1980; Алтухов, 1989). Хотя в первых двух работах было исследовано суммарно 522 119 генов, ни одной мутации не было обнаружено. Исходя из этих данных Нил с коллегами (Neel et al., 1980) заключили, что темп мутаций должен быть $6 X 10^{-6}$ на локус на поколение с 95%-ным доверительным уровнем. По-видимому, эти оценки близки к селективно нейтральным точковым мутационным событиям. Некоторые молекулярные маркеры могут давать темп мутаций, заметно отличающийся от приведенных оценок - 1.9×10^{-4}, и близкий к частоте внутригенных рекомбинаций – 6.8×10^{-4} (Алтухов и др., 1983).

Непрямые методы оценки

Многие авторы применяли непрямые методы оценки темпа мутаций для аллелей, кодирующих ферменты (Neel, 1973; Nei, 1977a; Neel, Rothman, 1978; Bhatia et al., 1979; Алтухов, 1989). Эти оценки базируются на допущении равновесия в популяции между мутациями, дрейфом и естественным отбором. Неем (Nei, 1977a) получена оценка темпа мутаций для японских макак с использованием 29 белковых локусов, которая составила 2.3×10^{-6} на локус на поколение. Базируясь на работе Нея и расширяя ее Кимура (Kimura, 1983c) оценил долю вредных мутаций среди всего их набора как 80-96%. По его оценке (Kimura, 1983c) темп мутаций для оставшейся части нейтральных мутаций составил на локус на поколение 1.65×10^{-7}. Оригинальный подход предложен Алтуховым с соавторами (Алтухов и др., 1983; Алтухов, 1989). Авторами исследованы редкие аллельные варианты по ферментным маркерам отдельно для мономорфных и полиморфных локусов сосны обыкновенной на двух фазах жизненного цикла. «Истинный» темп мутаций был оценен как 1.9×10^{-4}, тогда как частота мутаций по полиморфным маркерам как 8.7×10^{-4} (Алтухов, 1989, Таблица 5.8). Как и в работе Кимура (Kimura, 1983c), подчеркивается также, что большинство новых мутаций по локусам, кодирующим белки, являются вредными. Коэффициент естественного отбора на двух фазах жизненного цикла оценен как примерно равный 70% (s=0.71; Алтухов, 1989, Таблица 5.8). Хотя полученные оценки вполне состоятельны, но они базируются на допущении генетического равновесия в популяции. Поскольку никаких данных об этом не имеется, то приведенные величины темпа мутаций можно рассматривать как предварительные.

Оценки темпа мутаций по скорости генных замен

Если допустить, что большинство замен в генах происходят стохастически и являются селективно нейтральными, то темп мутаций может быть оценен на основе данных о нуклеотидных заменах. Эта возможность связана с тем, что темп замен в генах (α) для нейтральных мутаций равняется скорости мутаций (μ) (Nei, 1987, Chapter 13). В реальности, исходя из предыдущего обсуждения, большинство новых мутаций не являются нейтральными, а обладают вредным эффектом и поэтому быстро удаляются из популяций нормализующим («очищающим») отбором.

Поэтому, если оценивать темп мутаций при допущении $\mu = \alpha$, то это должно привести к недооценке суммарного темпа мутаций. Исключение в этом отношении могут представлять псевдогены, в которых естественный отбор, очевидно, отсутствует. Однако, как отмечает Ней (Nei, 1987, p.32), подобная оценка может быть использована для проверки «нуль-гипотезы» нейтральных мутаций, поскольку в теории нейтральности молекулярной эволюции рассматриваются только те мутации, которые вносят вклад в генные замены или в полиморфизм (Kimura, 1983a,b).

В геномах многих таксонов эукариот псевдогены встречаются весьма часто. Поскольку псевдогены не имеют какой либо обнаруженной функции, общее число мутаций можно оценить по темпу нуклеотидных замен, если время потери функциональности является известным (Nei, 1987, p.147). Некоторые оценки темпа нуклеотидных замен для псевдогенов (b) приводятся вместе с данными для функционирующих генов в Табл. 6.5.1.

Таблица 6.5.1. Время с момента дупликации *(T_d)*, время с момента потери функции *(T_n)*, и темп нуклеотидных замен на сайт в год *(b)* для различных псевдогенов глобина; a_1, a_2 и a_3, обозначают темпы нуклеотидных замен для первого, второго и третьего положений в кодоне функциональных генов *(По Li et al., 1981 заимствовано из Nei, 1987)*

Table 6.5.1. Times since gene duplication *(T_d)*, times since non-functionalization *(T_n)*, and rates of nucleotide substitution per site per year *(b) for* various globin pseudogenes; a_1, a_2 and a_3, denote the rates of nucleotide substitution for the first, second, and third nucleotide positions of codons in functional genes, respectively *(From: Li et al., 1981 adopted from Nei, 1987)*

Ген	***Показатель***					
	Td (млн*)	***Tn (млн)***	***b*** $(\times 10^{-9})$	a_1 $(\times 10^{-9})$	a_2 $(\times 10^{-9})$	a_3 $(\times 10^{-9})$
Мышь *Ψα3*	27	23	5.0	0.75	0.68	2.65
Человек *Ψα1*	49	45	5.1	0.75	0.68	2.65
Кролик *Ψβ2*	44	43	4.1	0.94	0.71	2.02
Коза $\Psi\beta^x$ и Ψ^z	46	36	4.4	0.94	0.71	2.02
В среднем			4.7	0.85	0.70	2.34

Примечание. *млн – миллионов лет назад.

Из данных таблицы следует, что (1) темп замен нуклеотидов в функциональных генах миоглобина, проанализированных у тех же видов, что и псевдогены, наивысший по третьему положению в кодоне и (2) темп нуклеотидных замен в псевдогенах вдвое выше, чем соответствующая величина даже по третьему положению в кодоне (Табл. 6.5.1). В среднем темп нуклеотидных замен составляет 4.7×10^{-9} на нуклеотид в год. Учитывая сказанное ранее, эту величину можно рассматривать как темп мутаций для нейтральных генов. Глобиновые гены млекопитающих состоят из 423 нуклеотидов или, соответственно, из 141 кодона. Поэтому, суммарный темп мутаций на ген (в кодирующей части) можно оценить как 2.0×10^{-6} в год (Nei, 1987). Предполагая, что средняя протяженность поколения у человека в его эволюционной истории была порядка 20 лет, можно оценить суммарный темп мутаций для кодирующего региона гена α-глобина человека как $\mu = 4.0\times10^{-5}$ на поколение (Nei, 1987). Для других генов, как подчеркнуто выше, темп мутаций может быть еще выше: $\mu = 1.9\times10^{-4}$ или иногда даже достигает тысячных и сотых долей (Алтухов, 1989). Это высокие темпы мутаций. Однако следует помнить, что в функциональных генах эти мутации элиминируются очищающим отбором на ранних стадиях жизненного цикла и не вовлечены в дальнейшую эволюцию в популяциях. Темп остальных мутаций, соответственно, должен быть ниже.

Приведенные выше данные предполагают, что синонимичные замены нуклеотидов происходят более часто, чем не синонимичные, возможно даже, что темп здесь такой же, как по

псевдогенам. Темп синонимичных замен может быть оценен с помощью метода, который будет рассмотрен несколько позже. На основе подобных подходов было показано (Miyata et al., 1980; Li et al., 1985b), что темп синонимичных замен сходен для многих различных генов, таких как α- и β-глобина, гормона роста и других, и что средний темп мутаций для группы млекопитающих на нуклеотид в год равен 4.6×10^{-9} (Nei, 1987, Table 5.6). Этот темп почти совпадает с темпом мутаций для псевдогенов. Это предполагает, что синонимические мутации, по-видимому, не подвержены какому-либо очищающему отбору, хотя на эту тему остаются еще некоторые противоречивые данные, возможно, связанные со статистическими ошибками оценок. Если все-таки предположение справедливо, то темп нейтральных мутаций можно также протестировать на основе синонимичных замен нуклеотидов.

Выше рассматривались только ядерные гены. Однако известно, что темп мутаций митохондриальных генов у млекопитающих и других организмов выше, чем у ядерных генов (Brown et al., 1979; 1982; Avise, 1994). Весьма высокий темп точковых мутаций был обнаружен у бактерий и РНК вирусов (Air, 1981; Holland et al., 1982; Nei, 1983; Gojobori, Yokoyama, 1985; Алиханян и др., 1985; Льюин, 1987; Russel, 1998, p.629). У бактерий темп мутаций достигает 10^{-6} – 10^{-7} на локус на поколение (Льюин, 1987, с. 41). В исследованиях нуклеотидных замен у вируса гриппа А обнаружено, что темп мутаций должен составлять порядка 0.01 на нуклеотидный сайт в год, чтобы объяснить обнаруженную степень дивергенции штаммов. Это почти на шесть порядков выше, чем темп замен в ядерных генах эукариот. В горячих точках генома *E. coli* спонтанный темп транзиций и стоп мутаций составляет 2, 3, 10 и даже 100 на ген *lac* (Russel, 1998). Предполагается, что эти высокие темпы связаны с отсутствием или слабостью репарационных механизмов в ДНК и РНК у вирусов и бактерий (Holland et al., 1982; Льюин, 1987; Жимулев, 2002).

Для эволюционной генетики, как сказано ранее, важно знать темп мутаций в той части, которая вносит вклад в замещения аминокислот в белках, и, таким образом, определяет их полиморфизм в популяциях. При допущении нейтральности, темп мутаций может быть оценен по аминокислотным замещениям (Табл. 6.5.2). То есть, темп мутаций для определенного полипептида задается темпом замещений на аминокислотный сайт, помноженным на число аминокислот в полипептиде. Темп мутаций варьирует значительно для разных полипептидов, как это представлено в исследовании Чакраборти и соавторов (Chakraborty et al., 1978) (рис. 6.5.1). Полученное распределение приблизительно следует гамма-распределению со средним, равным 2.47×10^{-7}, и стандартным отклонением - 2.51×10^{-7}. Изменчивость в темпе замещений на полипептид может быть обусловлена либо различиями в числе аминокислот, либо темпом замещений на сайт.

Таблица 6.5.2. Темп аминокислотных замещений на аминокислотный сайт в 10^9 лет ($\times 10^9$) в различных белках *(По Dayhoff, 1978, заимствовано из Nei, 1987)*

Table 6.5.2. Rates of amino acid substitutions per amino acid site per 10^9 years ($\times 10^9$) in various proteins *(From: Dayhoff, 1978, adopted from Nei, 1987)*

Белок	**Темп**	**Белок**	**Темп**
Фибринопептиды	*9.0*	*Тиротропин, бета цепь*	*0.74*
Гормон роста	*3.7*	*Паратирин*	*0.73*
Ig каппа цепи С региона	*3.7*	*Парвальбумин*	*0.70*
Каппа казеин	*3.3*	*Ингибитор протеазы, тип BP1*	*0.62*
Ig гамма цепи С региона	*3.1*	*Трипсин*	*0.59*
Лютропин, бетта цепь	*3.0*	*Меланотропин бетта*	*0.56*
Ig лямбда цепи С региона	*2.7*	*Альфа кристаллин, А цепь*	*0.50*
Комплемент С3а анафилококсин	*2.7*	*Эндорфин*	*0.48*
Лакальбумин	*2.7*	*Цитохром b;*	*0.45*

Белок	Темп	Белок	Темп
Эпидермальный фактор роста	*2.6*	*Инсулин (из морской свинки и утрии)*	*0.44*
Соматотропин	*2.5*	*Кальцитонин*	*0.43*
Панкреатическая рибонуклеаза	*2.1*	*Нейрофизин 2*	*0.36*
Липотропин бетта	*2.1*	*Пластоцианин*	*0.35*
Гаптоглобин, альфа цепь	*2.0*	*Лактатдегидрогеназа*	*0.34*
Сывороточный альбумин	*1.9*	*Аденилат киназа*	*0.32*
Фосфолипаза Az	*1.9*	*Триозофосфат изомераза*	*0.28*
Ингибитор протеазы, типа PST1	*1.8*	*Вазоактивный кишечный пептид*	*0.26*
Пролактин	*1.7*	*Кортикотропин*	*0.25*
Панкреатический гормон	*1.7*	*Глицеральдегид-3-фосфатдегидрогеназа*	*0.22*
Карбоангидраза С	*1.6*	*Цитохром с*	*0.22*
Лютропин, альфа цепь	*1.6*	*Ферродоксин растений*	*0.19*
Гемоглобин, альфа цепь	*1.2*	*Коллаген (из неповторяющихся концов)*	*0.17*
Гемоглобин, бетта цепь	*1.2*	*Тропонин С, скелетной мускулатуры*	*0.15*
Липид-связывающий белок А-II	*1.0*	*Альфа кристаллин, В цепь*	*0.15*
Гастрин	*0.98*	*Глюкагон*	*0.12*
Лизоцим животных	*0.98*	*Глютаматдегидрогеназа*	*0.09*
Миоглобин	*0.89*	*Гистон Н2В*	*0.09*
Амилоид АА	*0.87*	*Гистон Н2А*	*0.05*
Ростовой фактор нервов	*0.85*	*Гистон Н3*	*0.014*
Кислые протеазы	*0.84*	*Убиквитин*	*0.01*
Миелин, основной белок	*0.74*	*Гистон Н4*	*0.010*

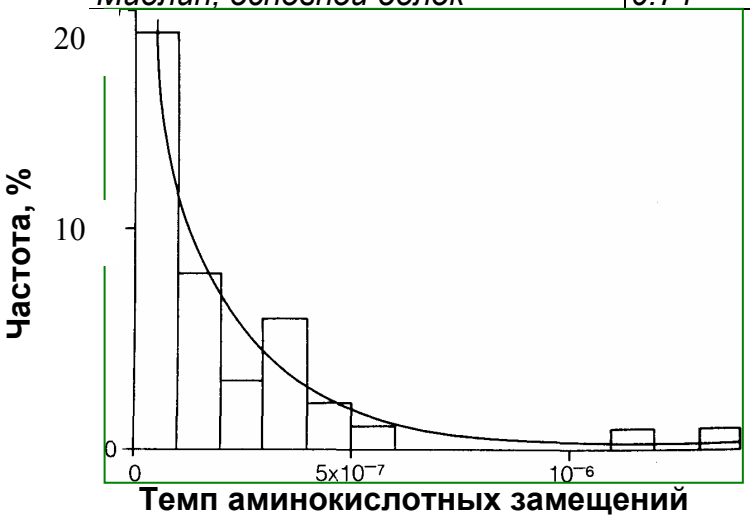


Рис. 6.5.1. Распределение темпа аминокислотных замещений на полипептид в год. Число проанализированных полипептидов равно 41 (По Chakraborty et al., 1978).
Fig. 6.5.1. Distribution of the rate of amino acid substitution per polypeptide per year. The number of polypeptides used is 41 (From Chakraborty et al., 1978).

Заметим, что средний темп замещений (мутаций) приблизительно равен произведению среднего темпа замещений из Табл. 6.5.2 (1×10^{-9}) и среднего числа аминокислот на полипептид (240). Различия в темпе замещений на сайт предположительно отражают изменчивость в интенсивности очищающего отбора.

Следует заметить, что белки, по которым был исследован темп замещений аминокислот выше, были в среднем небольшого размера; среднее число аминокислот на полипептид составило

240. Большинство белков, используемых при электрофоретических исследованиях, значительно больше в размере. Распределение молекулярного веса 119 белков показывает более точное соответствие гамма распределению (рис. 6.5.2). Средняя (M) и стандартное отклонение (s_x) для этого распределения равны M = 45 102, s_x = 24 531, соответственно. Поскольку средний молекулярный вес аминокислоты примерно равен 110, то средний полипептид должен иметь около 410 аминокислот. Если допустить, что темп аминокислотных замещений на сайт одинаков для обеих групп полипептидов, представленных на рис. 6.5.1-6.5.2, то ***средний темп аминокислотных замещений для среднего белка может быть оценен, как 4×10^{-7} на локус в год*** (Nei, 1987, с. 36). Ранее мы отмечали, что электрофорез выявляет только около одной-четвертой всех аминокислотных замещений. ***Поэтому, темп электрофоретически выявляемых аминокислотных замещений составляет около 10^{-7} на локус в год*** (Nei, 1987, с. 36)**.** При нуль-гипотезе нейтральных мутаций это равно темпу мутаций.

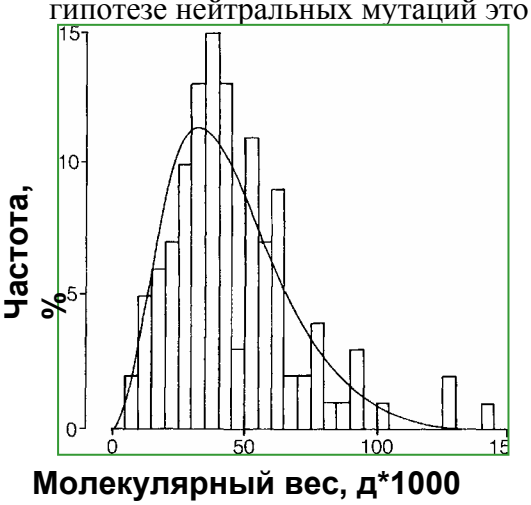


Рис. 6.5.2. Распределение молекулярного веса белковых субъединиц видов млекопитающих. Общее число использованных субъединиц составило 119. Гамма распределение очень хорошо соответствует эмпирическим данным: ($X^2(9)$ =6.43; P>0.65; По Darnell, Klotz, 1975, заимствовано из Nei, 1987).
Fig. 6.5.2. Distribution of molecular weights of protein subunits in mammalian species. The total number of subunits used is 119. The gamma distribution fits the data very well ($X^2(9)$ =6.43; P>0.65; From: Darnell, Klotz, 1975, adopted from Nei, 1987).

Темп мутаций и время дивергенции

Как упоминалось ранее, темп аминокислотных или нуклеотидных замен, по-видимому, является постоянным при расчете на год, но не на поколение. Если темп мутаций на уровне ДНК постоянен в годовом масштабе, то это легко понятно в рамках теории нейтральной эволюции (Kimura, Ohta, 1971; Nei, 1975). Однако исследования классической генетики показали, что темп мутаций в целом постоянен в масштабе на поколение, а не в год. В чем причина расхождений? Возможное объяснение состоит в том, что в классической генетике исследовали просто другие мутации, отличающиеся от замен нуклеотидов и аминокислот. Оценки темпа мутаций в классической генетике базировались, как отмечалось, на признаках внешней морфологии или жизнеспособности, которые либо частично вредны, либо вообще летальны. Возможно, что большинство этих мутаций связаны с крупными преобразованиями, типа больших делеций, вставок или на молекулярном уровне сдвигов рамки считывания, с утерей функции гена. Так, Магни (Magni, 1969) показал на дрожжах, что большинство вредных мутаций являются мутациями сдвига рамки считывания. Мёллером (Muller, 1959) на дрозофиле также было обнаружено, что летальные мутации происходят на стадии мейоза. Поэтому можно ожидать, темп вредных мутаций будет скорее связан с протяженностью поколения, чем в расчете на год. Наоборот, мелкие мутации или происходящие по другим, селективно менее значимым генам, должны быть небольшой частью всех мутаций и могут происходить в любое время, не завися от продолжительности поколения. Некоторые данные, показывающие что это именно так, были получены для фага T4 (Drake, 1966) и *E. coli* (Novick, Szilard, 1950). Не все в этом вопросе ясно. В целом частота мутаций должна зависеть и от циклов репликации ДНК, и от абсолютного времени. Например, Кубичек (Kubitschek, 1970) показал, что репликационно-зависимые мутации включают группу мутаций, вызываемых

кофеином и некоторыми другими химическими веществами, тогда как время-зависимые мутации, вызываются ультрафиолетом.

Тем не менее, приблизительная константность нуклеотидных и аминокислотных замен для прокариот и эукариот, как обнаружено для гена 5S РНК (Hori, Osawa, 1979), фибринопектина, гемоглобина и цитохрома-В (Dickerson, 1971), предполагает, что имеются мутации, которые возникают пропорционально абсолютному времени. Они могут возникать в результате действия ультрафиолета или по другим причинам. Важно отметить для себя, что суммарный темп мутаций (μ_T) складывается из нескольких компонент: $\mu_T = \mu_N + \mu_D + \mu_R$. Здесь: μ_N темп нейтральных мутаций, μ_D – темп вредных мутаций, μ_R – темп полезных мутаций. Последние мутации происходят весьма редко, но, возникнув, очень быстро становятся повсеместно распространенными, то есть, - вариация по ним теряется. Иными словами, молекулярные вариации и полиморфизм в популяциях на этом уровне сохраняются, в основном, за счет нейтральных мутаций (Kimura, 1983a; Nei, 1987).

Время дивергенции. Допущение о нейтральности молекулярной изменчивости лежит в основе реконструкции времени дивергенции на основе величины μ_N. Время в этом случае измеряется числом накопленных мутаций с момента изоляции генофондов пары таксонов. В случае белковых маркеров абсолютное время дивергенции (*T*) оценивается пропорционально величине стандартного генетического расстояния (Dn, Nei, 1972) и темпа мутаций для среднего белкового локуса μ_N ($\mu_N = 10^{-7}$, см. выше). То есть (Nei, 1975, 1987),

$$T = Dn\ 1/\mu_N. \qquad (6.1)$$

Оценка времени дивергенции на основе нуклеотидных и аминокислотных замен базируется на том же принципе – на аккумулированном числе мутаций для той или иной пары последовательностей. Подобный подход был кратко рассмотрен ранее в Лекции 4 для аминокислотных последовательностей и будет рассмотрен еще отдельно в Лекции 13 для последовательностей ДНК. В оценивании *T* остается немало вопросов. В частности, нередко необходимо специально масштабировать генетические расстояния с учетом различия темпов мутирования локусов, кодирующих разные по функциональной важности белки. Темп мутирования по ним может отличаться на порядок и более (см. текст выше и Табл. 6.5.2). Поэтому многие авторы использовали величину μ_N, отличную от приведенной выше (Carson, 1975, $\mu_N = 2*10^{-6}$; Avise et al., 1980, $\mu_N = 5*10^{-6}$, $\mu_N = 2*10^{-6}$, Картавцев и др., 1983; Картавцев, Мамонтов, 1983, $\mu_N = 5*10^{-6}$; Чичвархин и др., 2000, $\mu_N = 10^{-7}$; Картавцев и др., 2002, $\mu_N = 10^{-6}$). При оценке *T* по дивергенции нуклеотидных или аминокислотных последовательностей весьма важным является постоянство темпа замен (мутаций) в филетических линиях. Однако можно сказать, что постоянство темпа молекулярной эволюции остается нерешенной проблемой в настоящее время. Более того, постоянство является весьма приблизительным и имеется немало исключений, о чем еще пойдет речь. Мы практически не затронули в данной лекции транспозоны и вызываемые ими эффекты, которые близки к мутациям. О роли транспозонов будет отдельная дискуссия (Глава 16).

Реальных механизмов возникновения мутаций в последние годы открыто множество (Жимулев, 2002; Klug, Cummings, 2002). Надо понимать, что возникновение мутации – это сложный процесс, включающий изменения в цепи ДНК-РНК-белок и репарацию на нескольких уровнях. Поэтому предсказать темп мутаций только по одним молекулярным механизмам затруднительно (Kondo, 1977). Наконец, следует заметить, если даже темп мутаций зависит от продолжительности поколения, то можно объяснить линейную связь между молекулярной эволюцией и абсолютным временем, используя теорию частично вредных мутаций (Kimura, 1979), хотя это объяснение несколько запутанно. Выше было сказано, что полезные мутации редки и быстро закрепляются среди всех особей (фиксируются) в популяциях за счет механизма направленного естественного отбора. Однако надо сказать, что часть полезных мутаций остается многие поколения в популяциях, поддерживаемая балансирующим или варьирующим по

направлению отбором (Левонтин, 1978). Какая часть мутаций нейтральна, а какая поддерживается балансирующим отбором – это главный вопрос дискуссии, о которой уже упоминалось, и окончательного ответа на который пока не получено. Информация о влиянии естественного отбора на молекулярный полиморфизм будет дана еще в Главах 8-10 и 12.

Итоги для всего материала данной лекции можно сформулировать в трех пунктах.

1. Мутации имеют сложную природу. Появление даже относительно простых точечных мутаций зависит от комплекса химических процессов по модификации, замене нуклеотидов и их ферментативной репарации.

2. Из общего числа мутаций $\mu_T = \mu_N + \mu_D + \mu_R$ лишь часть, равная μ_N и части μ_R, сохраняется в виде вариации кодируемых ими молекул. Иными словами, молекулярный полиморфизм в популяциях сохраняется, в основном, за счет нейтральных мутаций и части, поддерживаемой сбалансированным естественным отбором.

3. Мутации играют принципиальную роль в молекулярных часах. Темп мутаций и время, прошедшее с момента изоляции генофондов пары таксонов, прямо определяют величины накопленных за время изоляции генетических расстояний между ними. Таким образом, зная величину генетического расстояния и темп мутаций можно определить время дивергенции таксонов.

6.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 6

1. Решите три примера в классе, чтобы углубить представление о мутациях.
(1) В Табл. 6.5.1 темп нуклеотидных замен на сайт в год равен в среднем $b = 4.7 \times 10^{-9}$ для изученных псевдогенов глобина. Согласуется ли эта величина с другими данными о мутациях для функциональных генов?
(2) Почему различаются значения темпов замен в зависимости от положения нуклеотида в кодоне: a_1, a_2, a_3 (Табл. 6.5.1).
(3) Исходя из кодирующего словаря определите, какие мутации, ведущие к заменам аминокислот, будут синонимическими?
2. В чем различие между хромосомной мутацией и генной мутацией? Каково различие между соматической и гаметической мутацией?
3. Почему мутации оказываются чаще вредными, чем нейтральными и полезными?
4. Большинство мутаций у диплоидных организмов являются рецессивными. Почему?
5. Посетите любую Интернет страницу или сайт, описывающий мутации и их последствия.

ГЛАВА 7. ГЕНЕТИЧЕСКИЕ АСПЕКТЫ ВИДООБРАЗОВАНИЯ

ГЛАВНЫЕ ВОПРОСЫ:

7.1. Общий генетический подход: перспективы и ограничения.
7.2. Какие данные необходимы? Какова база данных?
7.3. Концепции вида. Обзор литературных данных по гетерозиготности и расстояниям.
7.4. Модели видообразования: популяционно-генетический взгляд.
7.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 7.

РЕЗЮМЕ

1. Популяционно-генетический подход увеличивает значимость наших суждений о текущем статусе вида.
2. При использовании алгоритмического подхода, предлагаемого в лекции, возможно определить способы видообразования, реализовавшиеся в таксоне.
3. На основе молекулярно-генетических данных можно сделать филогенетические реконструкции, основанные на родстве, а не на сходстве, как ранее, а также оценить время дивергенции между таксонами.

Chapter 7. GENETIC ASPECTS OF SPECIATION

SUMMARY

1. Population genetic approach increases the validity of our conclusions on current species status.
2. Application of the approach, proposed in this chapter permit us to define modes of speciation in a taxon.
3. On the basis of molecular genetic data a phylogenetic reconstructions by kinship are possible, as well as dating time of divergence.

7.1. ОБЩИЙ ГЕНЕТИЧЕСКИЙ ПОДХОД: ПЕРСПЕКТИВЫ И ОГРАНИЧЕНИЯ

Главная цель данной лекции – это кратко осветить проблемы биологического вида и видообразования. Эти проблемы привлекают исследовательский интерес с момента становления биологии как науки. Данные проблемы являются одними из ключевых в общей биологии и общей генетике. Наиболее принята в настоящее время синтетическая теория эволюции (СТЭ), частью которой является биологическая концепция вида (БКВ). Обоснование и систематическое изложение СТЭ представлено в работах Холдейна (Haldane, 1932), Добжанского (Dobzhansky, 1937; 1943; 1951), Хаксли (Huxley, 1954), Майра и соавторов (Майр, 1947; Майр, 1968). В России популярным обобщением СТЭ стала книга Тимофеева-Ресовского и соавторов (1977). Конструктивное развитие положений СТЭ предложено Воронцовым (1980).

Одно из уязвимых мест СТЭ - отсутствие, как правило, возможности экспериментальной проверки ключевого в БКВ критерия не скрещиваемости видов в природе. Недавно, с развитием филогенетики, возникли новые противоречия между БКВ и филогенетической концепцией вида (Avise, Wollenberg, 1997; Harrison, 1998; Templeton, 1998). Не вполне разработана в СТЭ и теория видообразования. Точнее говоря, в количественном смысле теория как таковая вообще отсутствует. Имеется немало другой критики, которая была, например, суммирована Кингом

(King, 1993). Как понимание этого состояния появилась серия работ, теоретически и экспериментально развивающая данное направление (Avise, Wollenberg, 1997; Arnold, Emms, 1998; Harrison, 1998; Howard, 1998; Templeton, 1998; Wu, Hollocher, 1998; Картавцев и др., 2002).

Надо подчеркнуть, что, тем не менее, многие направления СТЭ и генетики видообразования развивались. Так, разносторонний анализ был проведен для понимания смысла и концептуальных основ видообразования (Bush, 1975; Grant, 1985; King, 1993; Harrison, 1998; Templeton, 1998; De Qeiros, 1998). Проанализированы также генетические основы создания репродуктивной изоляции (Templeton, 1981; Leslie, 1982; Nei et al., 1983; Coyne, 1992; Templeton, 1998; Butlin, 1998), возможность симпатрического видообразования (Bush, 1975; Lynch, 1989; McCune, Lovejoy, 1998), роль сальтаций и революций в эволюции (Алтухов, Рычков, 1972; Carson, 1975; Алтухов, 1974; 1989; 1997). Наконец, была подвергнута анализу генетическая дифференциация в ходе формирования разных форм и таксонов (Ayala et al., 1974; Avise, 1976; Avise, Aquadro, 1982; Nevo, Cleve, 1978; Thorpe, 1982; Nei, 1975; 1987). Каковы же в целом перспективы и ограничения генетического подхода?

Перспективы.

- Редукция данных до генотипических значений (кодов) дает возможность использовать генетическую теорию в анализе.
- Возможно сравнительное исследование изменчивости между таксонами по структурным и регуляторным элементам генома.
- Исследование дивергенции видов из природы по белкам и ДНК привело к открытию «молекулярных часов».
- Появилась возможность филогенетических реконструкций, основанных (1) не на сходстве, а (2) на родстве, за счет оценки хронологии и последовательности ветвлений в молекулярных филетических линиях.

Ограничения.

- Дедукция ограничена генотипическими описаниями и генетической теорией.
- Анализ связан с предварительным, трудоемким экспериментальным исследованием (со своими собственными ограничениями).
- Исследование видов из природы часто ограниченно уникальностью или редкой повторяемостью события (феномена).
- Генотипические эффекты маркерных локусов на фенотип являются слабыми.
- Теория по некоторым направлениям не достаточно развита.

7.2. КАКИЕ ДАННЫЕ НЕОБХОДИМЫ? КАКОВА БАЗА ДАННЫХ?

В целом необходимы данные такого рода:

- Данные, которые подтверждают (отвергают) центральную догму неодарвинизма – эволюция может происходить только на основе генетических изменений.
- Данные об изменчивости на различных уровнях биологической организации в генетических терминах (полокусные генотипические значения – AA, …) – одномерные таблицы данных (ТД).
- Данные о генотипических значениях для особей по набору локусов (комплексный генотип – AA Bb…) – многомерные ТД.
- Дополнительные информация: морфологические данные, данные об абиотической изменчивости (по меньшей мере, как экспертные оценки – группирующие переменные) и др.

Краеугольное видение СТЭ можно представить в виде схемы Добжанского (рис. 7.2.1), в которой разделение генофонда является ключевым для видообразования. Если кто-либо представит факт, что эволюция возможна без генетических изменений в филетических линиях, то тогда эволюционно-генетическая парадигма и, в частности, СТЭ могут быть отвергнуты.

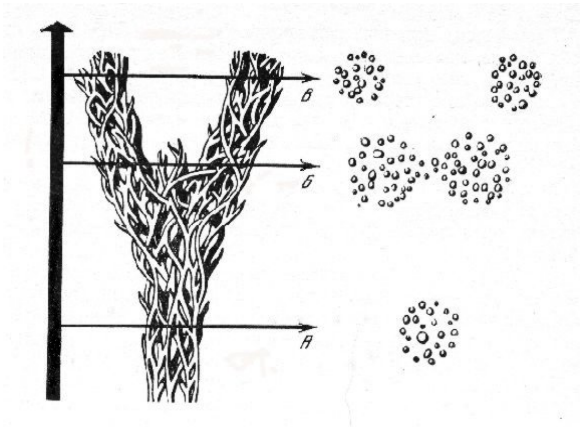


Рис. 7.2.1. Схематическое представление видовой дивергенции во времени.

А – единая видовая популяция, В – начальная фаза дивергенции (подвиды), С – различные виды. Кружки с правой стороны представляют объединения отдельных демов или популяций, между которыми имеется поток генов вплоть до фазы В (По Dobzhansky, 1955).

Fig. 7.2.1. A schematic representation of species divergence in time (From: Dobzhansky, 1955).

A – Single species population, B – Initial phase of divergence (subspecies), C – Different species. Circles on the right side represented the gatherings of separate demes or populations having exchange by gene flow with each other up to phase B.

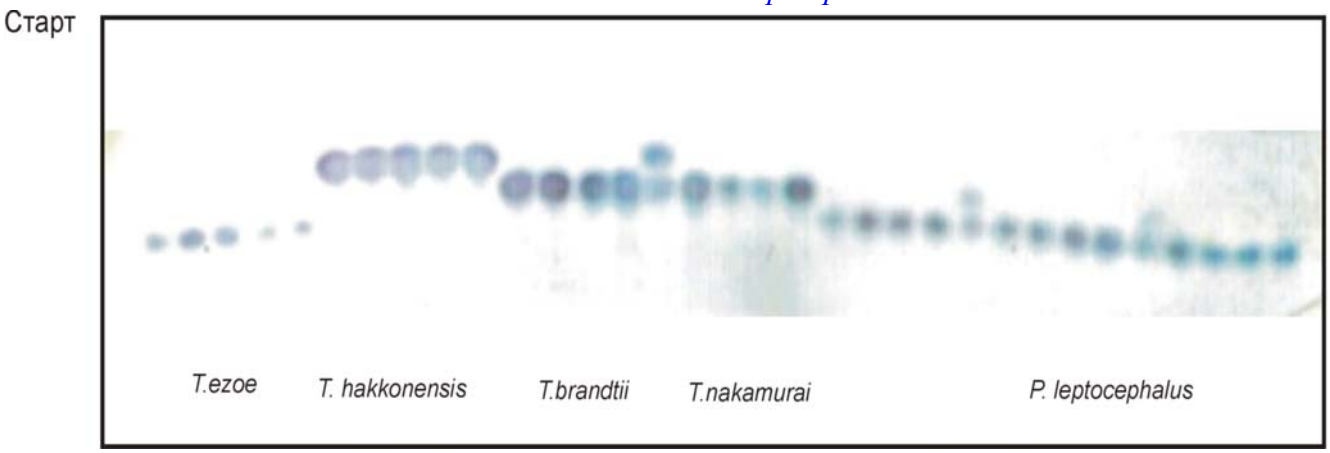


*Рис. 7.2.2. Электрофорез фермента фосфоглюкомутазы (локус **PGM***) у карповых рыб (Cyprinidae, Leuciscinae) из рода Дальневосточные красноперки (Tribolodon, 4 вида) и у жереха (Pseudaspius leptocephalus) (Картавцев и др., 2002, фото не публиковалось).*

При анализе геля одновременно можно оценить и генетическую изменчивость и сходство.

*Fig. 7.2.2. Enzyme electrophoresis of phosphoglucomutase (**PGM***) locus in the dace, the Cyprinid fishes (Pisces, Cyprinidae, Tribolodon) (Kartavtsev et al., 2002).*

Both the genetic variability and similarity can be estimated through the gel inspection.

Каковы имеющиеся базы данных, чтобы можно было с ними работать? Наиболее представительными в настоящее время являются данные по изменчивости белковых или точнее ферментных генов. Рассмотрим еще раз белковую изменчивость, теперь на конкретном примере внутри- и межвидового полиморфизма фосфоглюкомутазы у рыб красноперок (рис. 7.2.2). Посредством генетико-биохимического подхода и некоторых популяционно-генетических формул легко вычислить как уровень генетической изменчивости, так и степень сходства (или различия) генофондов особей из популяции одного вида или из различных видов. Для напомина-ния, ниже приведены некоторые выражения для расчета параметров изменчивости и сходства-различия, которые были разобраны ранее в Практическом занятии 5 соответствующей Главы 5.

Средняя гетерозиготность

$$H = \Sigma^{L}_{i=1} h_k / L.$$

H – это средняя ожидаемая гетерозиготность на локус или на особь.

$h_k = 1 - \Sigma^{m}_{i=1} p_i^2$, h_k – ожидаемая гетерозиготность по одному локусу (*Hexp=Hs*), p_i – частота *i*–

того аллеля; ***L*** – число локусов. Наблюдаемая гетерозиготность равна: $Hobs = hj / n$, hj – число гетерозигот в выборке, *n* – общее число особей в выборке.

Генетическое расстояние. Стандартное расстояние Нея (D_n)

$$D_n = - \ln I.$$

I – нормализованная генетическая идентичность для случайной выборки локусов. Для локуса ***j*** она равна: $I_j = (\Sigma\, k_{i=1}\, x_i\, y_i) / \sqrt{\Sigma\, k_{i=1}\, x_i^{\,2}\, \Sigma\, k_{i=1}\, y_i^{\,2}}$

Для всего набора локусов ***I*** равно: $I = Jxy / \sqrt{JxJy}$;

где Jxy, Jx, Jy – средние арифметические для сумм: $\Sigma\, k_{i=1}\, x_i\, y_i$, $\Sigma\, k_{i=1}\, x_i^2$, $\Sigma\, k_{i=1}\, y_i^2$.

Здесь, x_i и y_i – это частоты аллелей в популяциях (таксонах) *x* и *y*, ***k*** – число аллелей во всех попарных сравнениях.

Пример эволюционно-генетического исследования

Давайте рассмотрим пример конкретного исследования (Картавцев и др., 2002), в котором суммированы аллозимные данные для шести видов карповых рыб. В дискуссии будут также рассмотрены для сравнения и наши данные по мтДНК. Генетическая изменчивость среди изученных видов (Табл. 7.2.1, $\mathbf{N_A}$, **P**, **H**) типична в сравнении с другими представителями рыб (см. данные ниже). Оценки гетерозиготности для дальневосточных красноперок, например $H_{ож}$ = 0.034±0.013 (см. табл. 7.2.1, среднее для видов 1-4), хорошо согласуются с теми, что представлены для них ранее: H = 0.026 - 0.046. В целом оценки гетерозиготности для всей группы исследованных карповых рыб, $H_{ож}$ = 0.033±0.009 (см. табл. 7.2.1, среднее для видов 1-6), вполне типичны для рыб, для которые такие оценки в среднем составляют, H = 0.051±0.003 (Nevo et al., 1984, Ward et al., 1992).

Таблица 7.2.1. Генетическая изменчивость по 30 локусам у шести видов карповых рыб *(Pisces, Cyprinidae)*

Table 7.2.1. Genetic variability at 30 loci in 6 Cyprinid fish species *(Pisces, Cyprinidae)*

Вид	**N (SD)**	**N_A (SD)**	**P (%)***	**H (SE)**	
				$H_{наб}$	**$H_{ож}$ (X–B)****
(1) *Tribolodon hakonensis*	**24.1 (1.4)**	**1.3 (0.1)**	**23.3**	**.036 (.016)**	**.043 (.017)**
(2) *T. brandti*	**29.2 (1.5)**	**1.2 (0.1)**	**16.7**	**.018 (.010)**	**.021 (.013)**
(3) *T. ezoe*	**13.4 (0.4)**	**1.2 (0.1)**	**16.7**	**.045 (.020)**	**.067 (.029)**
(4) *T. nakamurai.*	**5.0 (0.0)**	**1.0 (0.0)**	**3.3**	**.007 (.007)**	**.007 (.007)**
(5) *Pseudoaspius leptocephalus*	**14.0 (0.0)**	**1.1 (0.1)**	**10.0**	**.017 (.010)**	**.016 (.009)**
(6) *Leuciscus waleckii*	**8.0 (0.0)**	**1.1 (0.1)**	**13.3**	**.029 (.014)**	**.042 (.022)**
Средняя (±SE), виды 1 - 4	**17.9±5.4**	**1.2±0.06**	**15.0±4.2**	**.026±.008**	**.034±.013**
Средняя (±SE), виды 1 - 6	**15.6±3.8**	**1.2±0.04**	**13.9±2.8**	**.025±.006**	**.033±.009**

Примечание. N – Средний размер выборки на локус, N_A – среднее число аллелей на локус, P – доля полиморфных локусов, H – средняя гетерозиготность на особь ($H_{наб}$, $H_{ож}$ –наблюдаемая и ожидаемая), SD – стандартное отклонение, SE – стандартная ошибка. * локус считается полиморфным, если частота преобладающего аллеля менее 0.99. ** X–B – гетерозиготность, которая ожидается при равновесии Харди-Вейнберга, не смещенная оценка (Nei, 1978).

На основе частот аллелей были получены оценки нескольких различных генетических расстояний между видами, два из них, индекс Роджерса, D_r и расстояние Нея, D_n приведены ниже (Таблица 7.2.2). Из данных таблицы следует, что виды одного рода генетически гораздо ближе друг к другу, чем представители двух других таксонов.

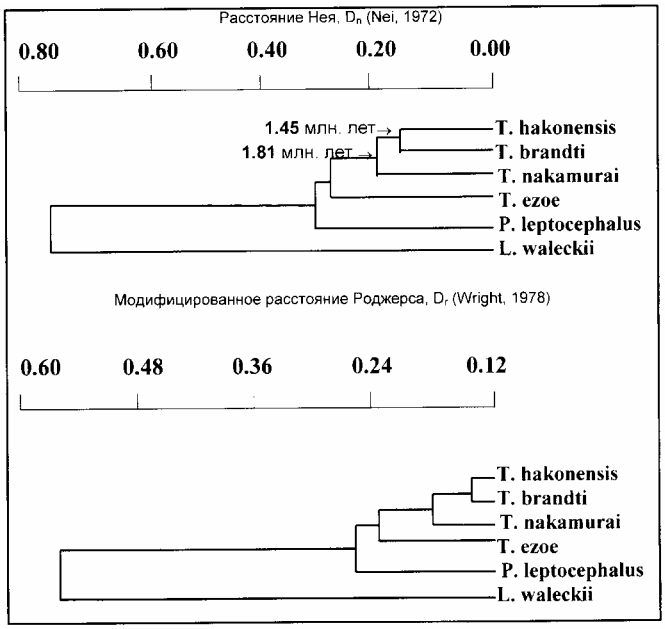
Таблица 7.2.2. Генетические расстояния между 6 видами карповых рыб *(Pisces, Cyprinidae)* по 30 белковым локусам

Table 7.2.2. Genetic distances among 6 Cyprinid fish species *(Pisces, Cyprinidae)* at 30 protein loci

Виды	**1**	**2**	**3**	**4**	**5**	**6**
(1) *T. hakonensis*	**–**	**.145**	**.243**	**.217**	**.300**	**.810**
(2) *T. brandti*	**.146**	**–**	**.177**	**.145**	**.266**	**.743**
(3) *T. ezoe*	**.242**	**.189**	**–**	**.284**	**.290**	**.762**
(4) *T. sp. (T. nakamurai)*	**.212**	**.146**	**.271**	**–**	**.223**	**.769**
(5) *P. leptocephalus*	**.272**	**.242**	**.276**	**.207**	**–**	**.760**
(6) *L. waleckii*	**.569**	**.538**	**.557**	**.547**	**.542**	**–**

Примечание. Ниже диагонали – расстояния Роджерса, D_R (Rogers, 1972), выше диагонали – стандартное расстояние Нея, D_n (Nei, 1972).

Главная выходная информация из треугольной матрицы сходства-различия – это чаще всего дендрограмма или дерево, которое иллюстрирует, как таксоны присоединяются друг к другу. Дендрограмма – это результат одного из вариантов кластерного анализа, призванного помочь свернуть цифровую информацию из треугольной матрицы в вид, легко воспринимаемый визуально. Когда единиц сравнения много, особенно это относится к внутривидовым сравнениям, то на самом деле более обоснованно визуализовать цифровую информацию из треугольной матрицы посредством других методов кластер-анализа (Pao, 1980), например, преобразовав расстояния в значения многомерных переменных и построив график рассеяния в двух-, трехмерном пространстве. В нашем случае построено дерево и наиболее близки друг к другу виды красноперок рода *Tribolodon*, а остальные два таксона попадают во внешние кластеры (рис. 7.2.3, слева и рис. 7.2.4, справа). Различия в топологии деревьев (в ветвлении) определяются несколькими причинами (Бейли, 1970; Pao, 1980; Nei, 1987): числом признаков (локусов), множественностью таксонов или единиц сравнения, методами формирования кластеров (см. рис. 7.2.3 и рис. 7.2.4) и, наконец, использованной мерой расстояния-сходства. Когда различия между таксонами большие, мера, как правило, мало влияет на кластеризацию. В частности, одинаковая кластеризация таксонов обнаруживается при использовании стандартного расстояния Нея, D_n и модифицированного расстояния Роджерса, D_r (рис. 7.2.3) Также как ранее представленные результаты, данные сопоставления полной последовательности нуклеотидов в митохондриальном гене цитохрома-В красноперок показывают их большую генетическую близость друг к другу и к жереху, *Pseudaspius leptocephalus* (рис. 7.2.5).



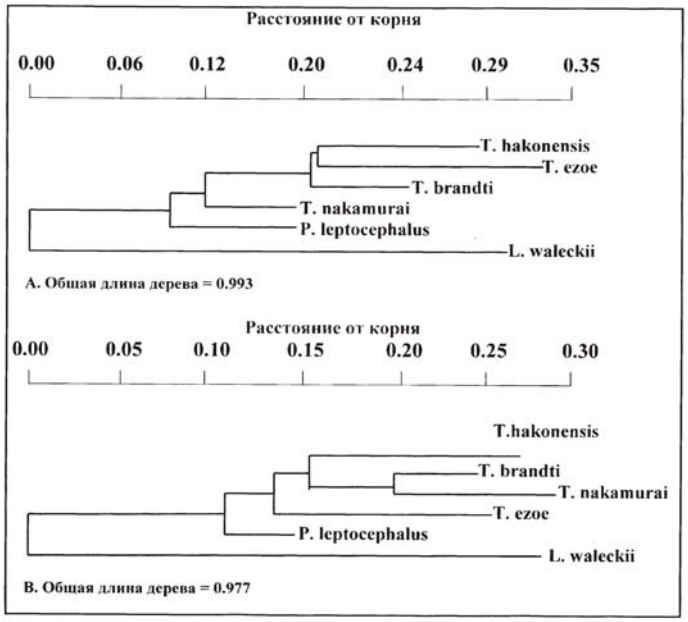


Рис. 7.2.3. Дендрограммы, иллюстрирующие кластеризацию 6 видов карповых рыб. Дендрограммы получены на основе метода НПГ по частотам аллелей 30 аллозимных локусов при использовании расстояний Нея – D_n (верхняя шкала, Nei, 1987) и модифицированного расстояния Рожжерса- D_r (нижняя шкала, Wright, 1978). Коэффициент кофенетической корреляции (r) равен: r = 0.99.
Fig. 7.2.3. Phenograms, which illustrating clustering via UPGM method among Cyprinid fishes on allele frequencies at 30 allozyme loci and using Nei's D_n (Nei, 1972) and modified Rogers' distance (D_r) (Wright, 1978). Same clustering obtained from Nei's D_m. Thus, clustering weakly depends of measure. Coefficient of cophenetic correlation: r = 0.99.

*Рис. 7.2.4. Варианты реконструкции филогенетических связей в подсемействе Leuciscinae (Pisces, Cyprinidae), полученные с использованием расстояния Роджерса (D_R) как исходной меры (Rogers, 1972). **A, B** – два возможных дерева (дендрограммы), которые построены по методу Вагнера. Дерево Вагнера укоренено с использованием внешнего таксона (L. waleckii). Длина дерева равна 0.993(A) и 0.977 (B). Значения r равны, соответственно: A – r = 0.979, B – r = 0.982.*
*Fig. 7.2.4. Variants of reconstruction of phylogenetic relations in Cyprinidae, which obtained by using Rogers' distance (D) as an initial metrics (Rogers, 1972). **A, B** are two possible phylograms, which constructed by Wagner's method. The r are equals, consequently: A, r = 0.979, B, r = 0.982.*

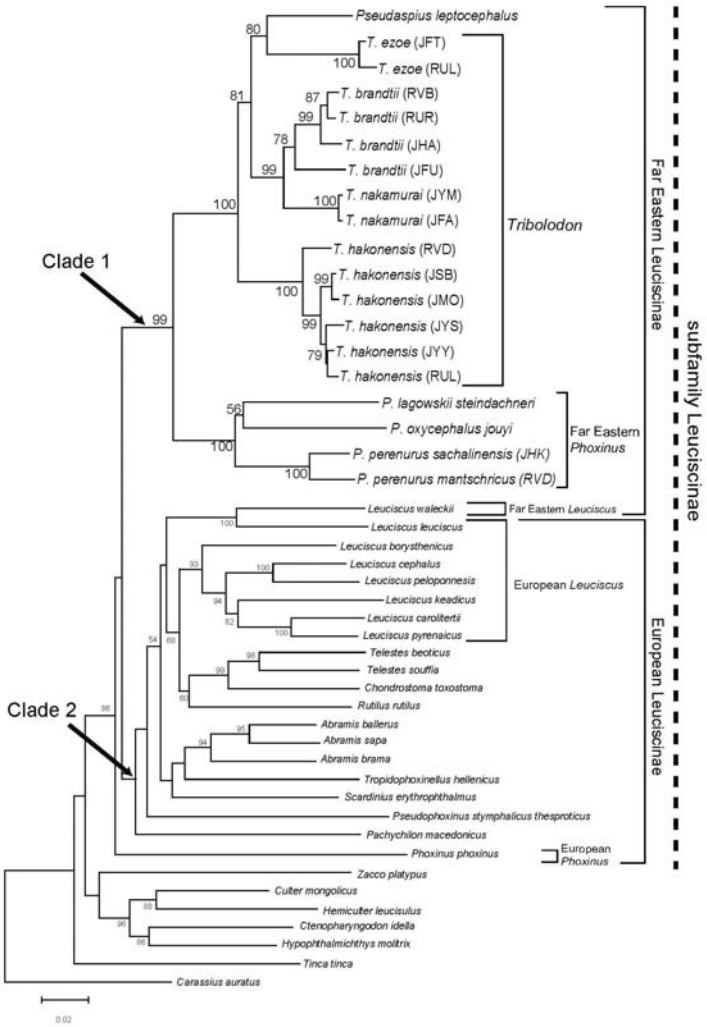


Рис. 7.2.5. Генетические расстояния и дерево, построенное по методу ближайшего соседства на основании данных секвенс-анализа по митохондриальному гену цитохрома-B.

Цифры в узлах дерева показывают величину поддержки бутстрепом в 1000 повторностей. Сокращения: T. ezoe, JFT – Япония, Фукушима, река Тадами, RUL – Россия, Сахалин, река Улике, T. brandti, RVB – Россия, залив Восток, RUR – Россия, Сахалин, река Улике, JHA - Япония, Хоккайдо, река Абира, JFU – Япония, Фукушима, река Укедо, T. nakamurai, JYM – Япония, Ямагата, река Могами, JFA – Япония, Фукушима, река Ага, T. hakonensis, RVD – Россия, залив Восток, JSB – Япония, Шига, озеро Бива, JMO – Япония, Миязаки, река Оидо, JYS – Япония, Ямагата, порт Саката, JYY – Япония, Ямагата, река Мамигасаки, RUL – Россия, Сахалин, река Улике, P. perenurus, JHK – Япония, Хоккайдо, RVD – Россия, район Владивостока (From: Sasaki et al., 2006). (По T. Sazaki, Y.P. Kartavtsev, T. Uematsu, V. Sviridov, N. Hanzawa, не опубликовано).

Fig. 7.2.5. Neighbor-joining tree based on mitochondrial cytochrome-b sequences' K2P distances (Kimura, 1980). Number at nodes represents bootstrap support in 1000 replicates The scale bar shows 0.02 base substitutions per site. Abbreviations: T. ezoe, JFT – Japan, Fukushima, Tadami River, RUL – Russia, Sakhalin, Uliker River, T. brandti, RVB – Russia, Vostok Bay, RUR – Russia, Sakhalin, Ulike River, JHA Japan, Hokkaido, Abira River, JFU – Japan, Fukushima, Ukedo River, T. nakamurai, JYM – Japan, Yamagata, Mogami River, JFA – Japan, Fukushima, Aga River, T. hakonensis, RVD – Russia, Vostok Bay, JSB – Japan, Shiga, Biva Lake, JMO – Japan, Miyazaki, Oydo River, JYS – Japan, Yamagata, Sakata Port, JYY – Japan, Yamagata, Mamigasaki River, RUL – Russia, Sakhalin, Ulike River, P. perenurus, JHK – Japan, Hokkaido, RVD – Russia, Vladivostok Area (From: Sasaki et al., 2006).

Ранее было отмечено, что топология дерева зависит от многих причин, а собственно «истинное» дерево обычно никогда не известно (Nei, 1987; Hillis et al., 1995; см. также Глава 14). Поэтому, чтобы построить правильное дерево зачастую приходится использовать различные подходы и комбинировать филогенетическую информацию для создания некоего обобщенного, или, как говорят, консенсус дерева или консенсус филогении. Для нашей группы видов карповых рыб такое дерево учитывало генетическую и экологическую информацию (рис. 7.2.6).

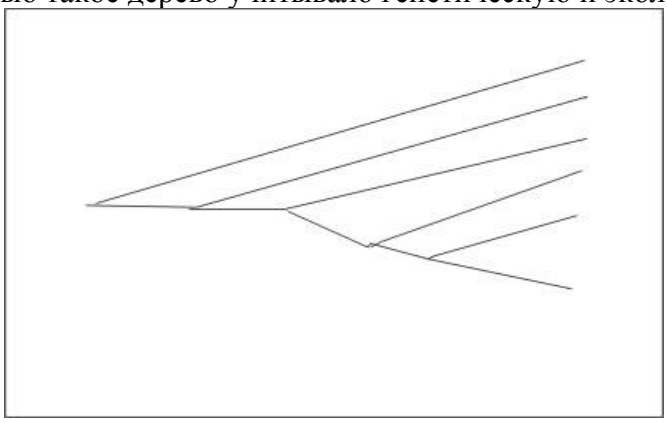


Рис. 7.2.6. Консенсус дерево для исследованных видов карповых рыб, подсемейства Leuciscinae (По Картавцев и др., 2002).
Fig. 7.2.6. Consensus dendrogramm for the studied species of Leuciscinae subfamily of the Cyprinid species.

7.3. КОНЦЕПЦИЯ ВИДА. ОБЗОР ЛИТЕРАТУРНЫХ ДАННЫХ О ГЕТЕРОЗИГОТНОСТИ И РАССТОЯНИЯХ

Прежде, чем сделать обзор, необходимо рассмотреть принципиальные схемы видообразования, с учетом генетических процессов, которые влияют на видообразование. В развитие основных идей СТЭ, и в частности схемы Добжанского (см. рис. 7.1.1), Буш (Bush, 1975) представил схему, которая отражает роль прерывания потока генов для видообразования (рис. 7.3.1).

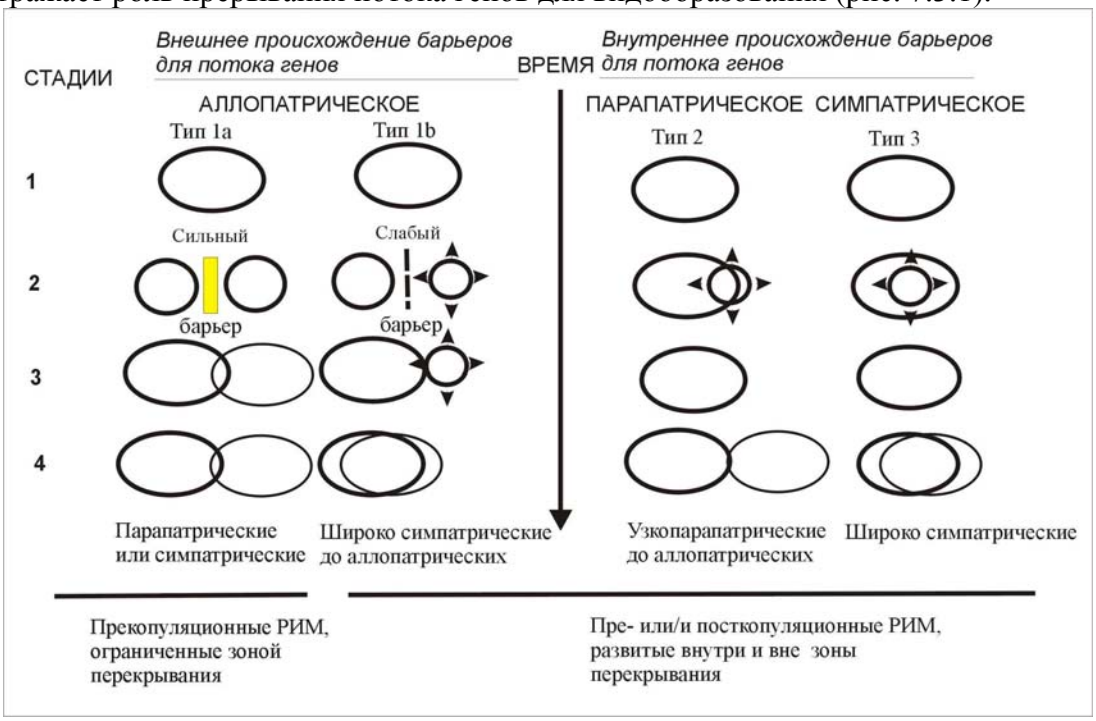


Рис. 7.3.1. Диаграмма главных способов видообразования. РИМ – репродуктивные изолирующие механизмы (По Bush, 1975).
Fig. 7.3.1. Diagram of main modes of species origin (From Bush, 1975).

Барьеры потоку генов могут создавать репродуктивные изолирующие механизмы (РИМ) или репродуктивно изолирующие барьеры (РИБ), которые в свою очередь ведут далее к формированию вида; при различных ситуациях в природе, реализуются различные способы видообразования (рис. 7.3.1). Ни схема, данная выше (рис. 7.3.1), ни сама работа Буша (Bush, 1975), не отвечают на многие фундаментальные вопросы видообразования. Например, неясно, какие способы видообразования наиболее часты в природе и действительно ли прерывание потока генов является единственным, принципиальным фактором при видообразовании, или действуют другие не менее важные факторы? Чтобы продолжить дальнейшее обсуждение также важно дать определение вида. Что такое вид?

Определение вида

Вид – это биологическая группировка, состоящая из одной или нескольких популяций перекрестно размножающихся особей, которые репродуктивно изолированы от других таких группировок и стабильны в природе, занимая определенный ареал (определение автора). В принципиальных позициях, это определение, типичное для БКВ. Так, одно из первоначальных определений БКВ формулируется следующим образом "***Вид – это репродуктивное сообщество популяций (репродуктивно изолированных от других), которое занимает определенную нишу в природе***" (Mayr, 1982, p. 273). Мы примем за основу в дальнейшей дискуссии БКВ, которая ограниченна, в основном, высшими бисексуальными организмами (Майр, 1968; Тимофеев-Ресовский и др., 1977; Templeton, 1998). БКВ наиболее близко подходит к теории популяционной генетики и, поэтому, я предпочитаю использовать ее в этой и в следующей лекции, как костяк дискуссии, не смотря на отмеченное выше ограничение. Имеются несколько других концепций вида со своими преимуществами и ограничениями, которые были критически проанализированы (Красилов, 1977; King, 1993). Список этих концепций дан ниже.

- Линнеевский вид.
- Биологическая концепция вида.
- Биологическая концепция вида в модификации Майра (БКВ) (Майр, 1947; 1968)
- БКВ, модификация II (Mayr, 1982).
- Концепция «распознания» (recognition) вида (Paterson, 1978; 1985).
- Концепция сплоченности (cohesion) вида (Templeton, 1989).
- Эволюционная концепция вида.
- Эволюционная концепция вида Симпсона (Simpson, 1961).
- Эволюционная концепция вида Вилея (Wiley, 1978).
- Экологическая концепция вида (Van Valen, 1976).
- Филогенетическая концепция вида (Cracraft, 1983).

Обзор данных о гетерозиготности и генетических расстояниях

Чтобы судить с генетической точки зрения о видообразовании, необходимо знать, как минимум, об уровнях генетической изменчивости среди внутривидовых группировок и групп таксономического ранга (требование парадигмы БКВ; Майр, 1968; Тимофеев-Ресовский и др., 1977), а также об относительной дивергенции генофондов форм, которые предположительно представляют различные стадии формирования вида (в соответствии с парадигмой БКВ и другими парадигмами; см. рис. 7.2.1, рис. 7.3.1, а также Майр, 1968; Тимофеев-Ресовский и др., 1977; King, 1993). В первую очередь, представляется полезным рассмотреть данные по генетической изменчивости, которую наиболее адекватно (Nei, 1975; 1987; Nei, Kumar, 2000) представляет средняя гетерозиготность на особь/локус, ***H*** (см. выражение выше). Репрезентативный обзор, хотя и довольно давно, дан Нево и его коллегами (Nevo et al., 1984; рис. 7.3.2).

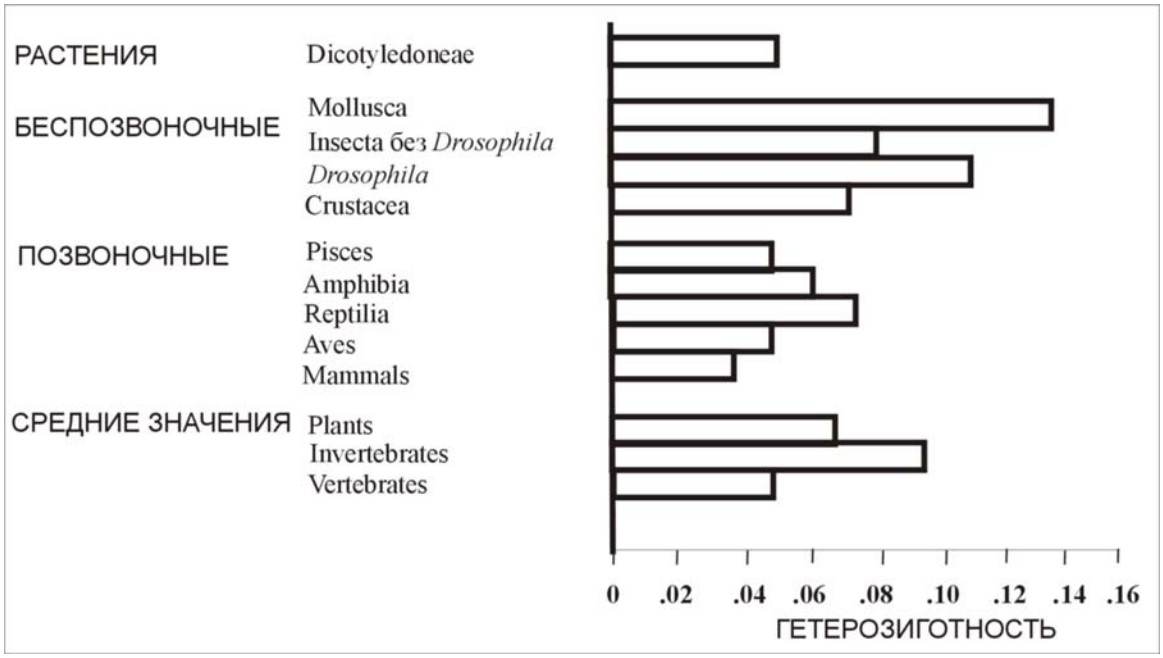


*Рис. 7.3.2. Уровни средней гетерозиготности (**H**) в различных таксонах (По Nevo et al., 1984).*
*Fig. 7.3.2. Levels of mean heterozygosity (**H**) in different taxa (From: Nevo et al., 1984).*

В среднем позвоночные животные имеют гетерозиготность ***H*** = 5%, беспозвоночные – ***H*** = 9%, а растения промежуточные значения между этими двумя величинами. Другие обзоры дают сходные значения, например, для рыб величина гетерозиготности составила ***H*** = 0,051± 0,003 (Ward et al., 1992). Принимая во внимание, что только около 25-50% изменчивости белковых маркеров генов обнаруживается посредством электрофореза (см. дискуссию в Лекции 5), можно сделать заключение о большой генетической изменчивости, поддерживаемой в природных популяциях – 10-20% и более локусов могут быть гетерозиготными у средней особи. Таким образом, этот факт поддерживает возможность, принимаемую СТЭ и БКВ, аккумулирования генетической изменчивости после разделения генофондов. На стадиях 1-4 в схеме, приведенной ранее (рис. 7.3.1), при способе видообразования 1а, это может вести к плейотропному возникновению репродуктивных изолирующих барьеров (РИБ или РИМ – см. рис. 7.3.1) и, соответственно, к возникновению вида.

Сопоставление некоторых других параметров изменчивости показывает, что генетическая изменчивость может зависеть от других источников и не так прямо связана с эволюцией и временем изоляции генофондов. Например, генетическая дифференциация в шкале Gst [Gst = (Ht-Hs) / Ht; см. более детальные объяснения о популяционной дифференциации, в том числе о дифференциации двух групп лососей в Главах 7-8) значительно различается в двух близких родах лососей (рис. 7.3.3), которые экологически отличны и имеют, в среднем, популяции различного размера (Алтухов и др., 1997, таблица 2.5). Приведенные данные поддерживают идею о том, что больший популяционный размер может определять большие значения ***H***, но не поддерживают временную концепцию разделения двух генофондов, так как *Salmo* более древняя группа, чем *Oncorhynchus* (Neave, 1958).

Другими словами размер, а точнее генетически эффективный размер популяции (Ne; см. Глава 8), может быть ключевым параметром, определяющим различия ***H*** (Soule, 1976; Nei, 1975, 1987). Литературные данные подтверждают, во всяком случае, что различия Ne у многих организмов существенны (рис. 7.3.4).

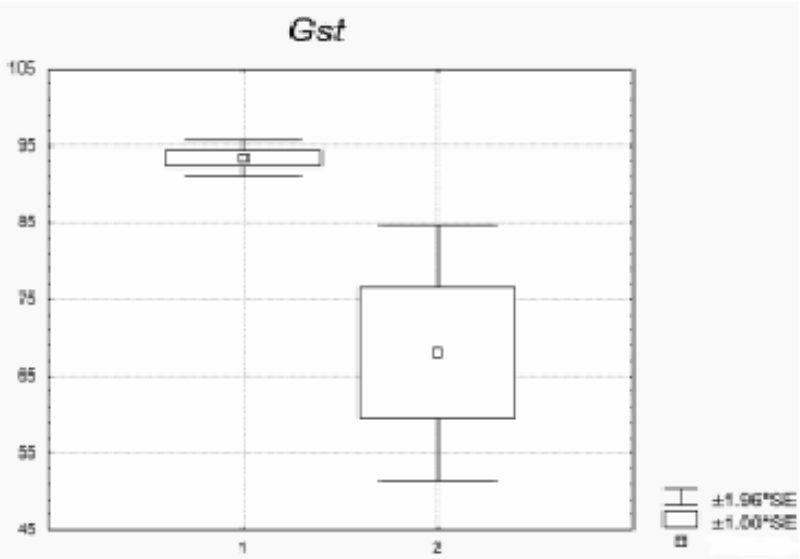


Рис. 7.3.3. Сопоставление генетической дифференциации в двух родах лососевых рыб в шкале стандартизированных частот аллелей (Gst = Fst). Судя по величине SE (стандартной ошибки), различия Gst являются значимыми статистически. Внутригрупповой компонент отражает генетическое разнообразие в каждом из двух таксонов.

Fig. 7.3.3. Comparison the genetic diversity in two salmon genera in the scale of standardize allele frequencies (Gst = Fst). Judging on the SE, the Gst differences are statistically significant.

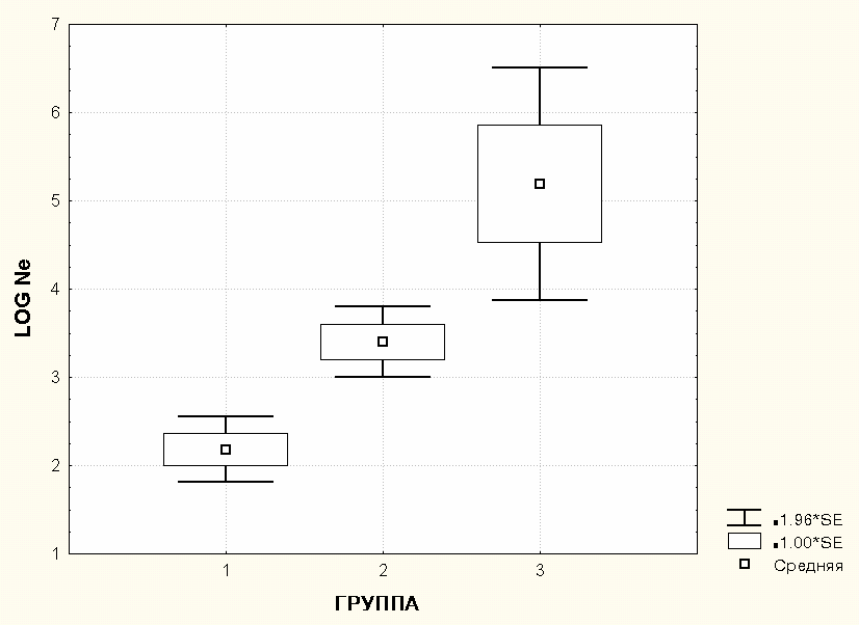


Рис. 7.3.4. Сопоставление средних величин эффективного размера популяций (Ne) в трех различных группах животных: 1 - позвоночные, 2 – наземные беспозвоночные, 3 – морские беспозвоночные.

Fig. 7.3.4. Plot of effective size (Ne) values across three different animal groups: 1, Vertebrates, 2, Land Invertebrates, 3, Marine Invertebrates.

В соответствии с последними данными можно интерпретировать различие ***H*** между позвоночными и беспозвоночными животными (см. рис. 7.3.2), как обусловленное, в основном, различиями Ne. Величины ***H*** и Ne являются функционально связанными согласно модели бесконечного числа мутаций (Infinite mutation model, Nei, 1987). Эта модель, которая является в принципе моделью нейтральных мутаций, не противоречит совсем не нейтральным плейотропным эффектам на РИБ в ходе процесса видообразования, которые возникают на основе накопленных

мутаций, когда каждая из них по отдельности нейтральна. Хотя многие ученые стоят в оппозиции к полностью нейтральной эволюции молекулярных маркеров (Lewontin, 1974, Алтухов, 1989), имеется немало других, во главе с разработчиком модели Кимура, которые считают, что нейтральные мутации наилучшим образом объясняют молекулярную эволюцию (Kimura, 1969; Nei, 1987). Мы не будем развернуто дискутировать на эту весьма сложную тему в данном курсе лекций, хотя будут представлены некоторые факты, которые не согласуются с экстремальными вариантами ни селекционистской, ни нейтралистской точек зрения (Лекции 8-10). Сейчас представляется возможным перейти к литературным данным о генетическом сходстве или, наоборот, различиям (расстояниям).

В иерархии таксонов животных подвиды имеют диапазон величины коэффициента сходства, или нормализованной идентичности Нея, ***I*** (Nei, 1972; ***Dn = -ln I***; см. выше Раздел 7.2): ***I*** = 0.6 – 1.0, с модой около 0.8, в сборной группе полувидов и близнецовых видов диапазон варьирования составил – ***I*** = 0.5 – 1.0, с модой около 0.7, среди видов диапазон варьирования равен – ***I*** = 0.5 – 1.0, с модой также около 0.7 и для родов – ***I*** = 0.0 – 1.0, с модой 0.2 (рис. 7.3.5).

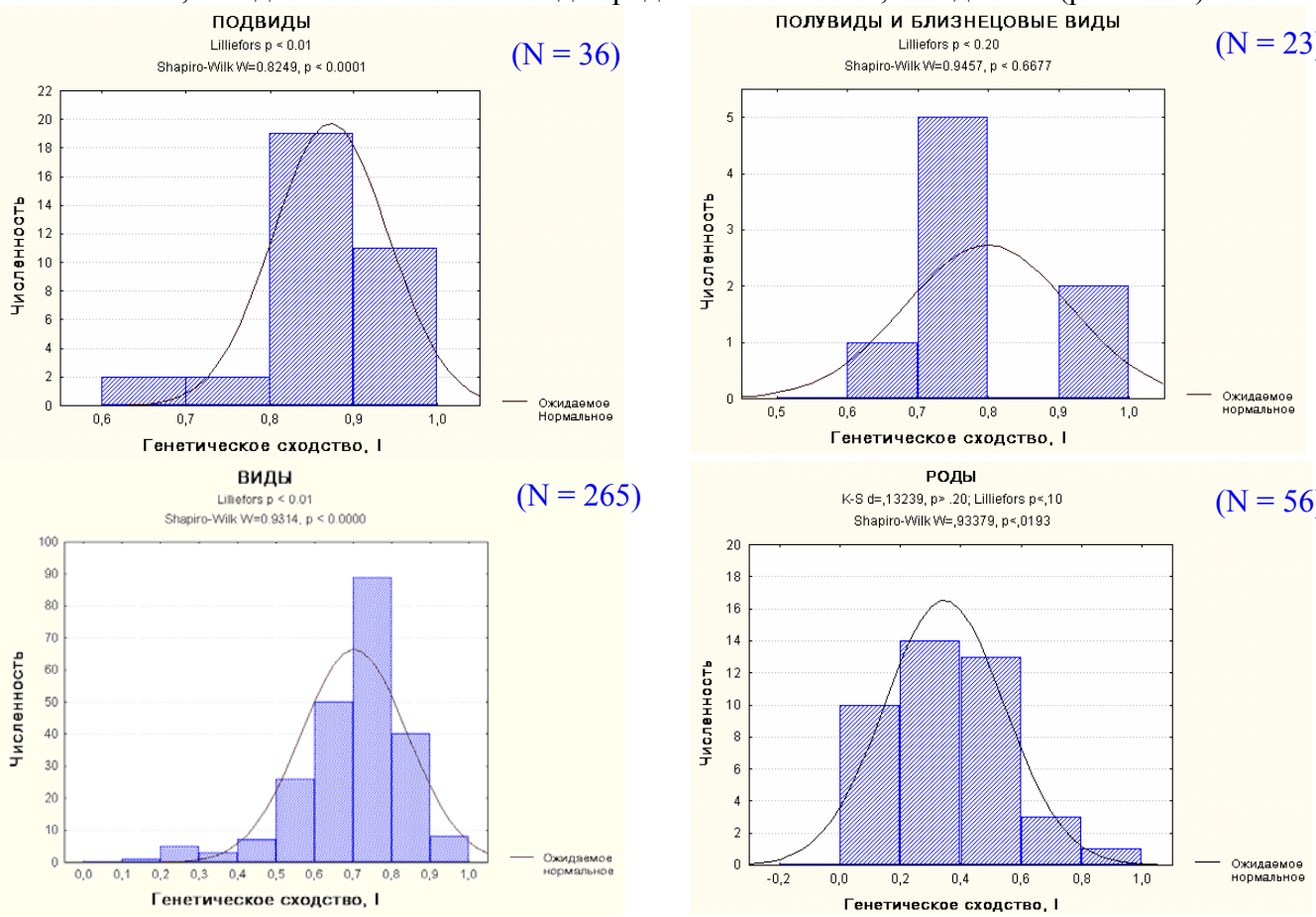


Рис. 7.3.5. Частотное распределение коэффициентов сходства у представителей различных таксонов животных.

*Данные автора (не опубликованы). Подавляющее большинство значений коэффициента сходства представлены **I,** коэффициентом сходства Нея (Nei, 1972). При отсутствии данных для I использованы значения и других близких коэффициентов. N – число сопоставленных таксонов.*
*Fig. 7.3.5. Frequency distribution the similarity coefficient among the representatives of different taxa in animals. Data are from Kartavtsev (in preparation). Bulk of **I**s are the normalized similarity indices (Nei, 1972). N is the number of taxa in comparisons.*

Три из четырех распределений являются одномодальными, хотя они и не соответствуют нормальному распределению, тогда как группа полувиды – близнецовые виды явно бимодальна (хотя статистическая значимость этого здесь отсутствует из-за малой выборки), что, очевидно, отражает искусственность их объединения в одну группу (рис. 7.3.5). То есть, генетическая дивергенция большинства близнецовых видов вполне сопоставима с таковой обычных видов, как это отмечали многие авторы. В среднем имеется очевидная тенденция уменьшения коэффициента сходства с увеличением ранга таксона (рис. 7.3.6).

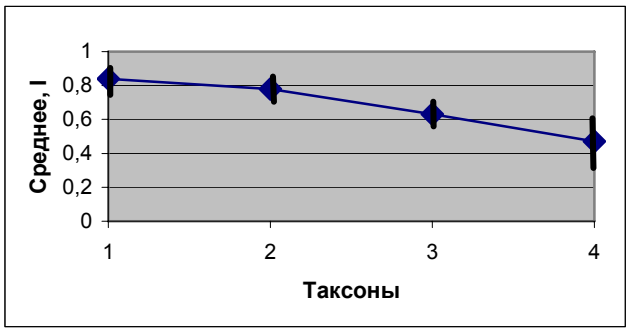


Рис. 7.3.6. Генетическое сходство в таксонах различного ранга: средние для групп.
1 – подвиды, 2 – полувиды и виды-двойники, 3 – виды, 4 - роды. Линии – доверительный интервал (95%).
Fig. 7.3.6. Genetic similarity in taxa of different rank: mean for the groups. 1, Subspecies, 2, Semispecies and sibling species, 3, Species, and 4, Genera.

Коэффициенты внутривидового генетического сходства или расстояния были измерены для многих групп организмов (Lewontin, 1974; Nei, 1987; Алтухов, 1989). Величина среднего генетического сходства на этом уровне составляет около ***I*** = 0.95 (см. детали в Лекции 8). Таким образом, имеющиеся генетические данные поддерживают (1) основную идею СТЭ о необходимости изоляции генофондов для формообразования и (2) географический способ видообразования (Тип 1а, см. рис. 7.3.1), как преобладающую в природе модель формирования вида с постепенным накоплением мелких генетических различий.

Предполагают ли представленные данные, что видообразование всегда следует Типу 1а? Я полагаю, что нет. Несколько примеров, приведенных ниже, подтверждают этот ответ. Так, в Швеции, в одном из горных озер, были известны две формы форели. Неизвестным было только, изолированы ли их генофонды или нет. После генетического исследования (Ryman et al., 1979), у этих форм были обнаружены два различных фиксированных аллеля, которые однозначно доказали полную изоляцию этих симпатрично обитающих форелей. Когда появился способ безошибочно различать всех особей, то для этих двух таксонов обнаружили и другие различия генофондов (Табл. 7.3.1).

Таблица 7.3.1. Генетические различия между симпатричными таксонами форели *Salmo trutta* в Швеции (По Ryman et al., 1979 с изменениями)
Table 7.3.1. Genetic differences among sympatric taxa of the trout *Salmo trutta* in Sweden (From: Ryman et al., 1979 with modification)

ЛОКУС (АЛЛЕЛЬ)	**ФОРМА I** ***LDH-1 (100/100)***	**ФОРМА II** ***LDH-1 (240/240)***	X^2
AGP-2 (100)	0.976	1.000	4.82*
CPK-1 (100)	0.562	0.293	26.04***
EST-2 (100)	0.967	0.886	3.37
LDH-5 (100)	0.872	1.000	12.63***
MDH-4 (100)	0.571	1.000	51.08***

SDH-1 (100)	0.891	1.000	10.58***
SOD (100)	0.989	1.000	0.01

Приведенные в таблице 7.3.1 данные иллюстрируют мысль о том, что иногда очень маленькие различия структурных генов могут создавать РИБ, и таким образом, репродуктивно изолированные биологические единицы. В конкретном случае, генетические расстояния, обнаруженные для двух форм форели равны, $D_n = 0.02$ (Ryman et al., 1979), что находится на уровне внутривидовой генетической дифференциации. Для лососевых рыб имеется много таких примеров (рис. 7.3.7).

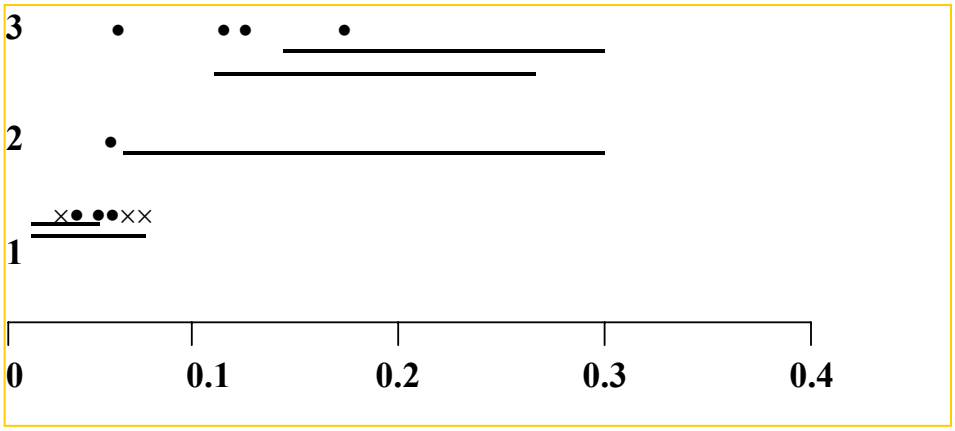


Рис. 7.3.7. Генетические расстояния в таксонах лососевых рыб

1 – популяции внутри вида, 2 – подвиды, 3 – виды. Точки – литературные данные, крестики – данные автора (По Картавцев, 1995).

Fig. 7.3.7. Genetic distances in taxa of salmon fishes (From Kartavtsev, 1995).

1, Populations within species, 2, Subspecies, 3, Species.

Этот график подтверждает мысль, что у лососевых небольшие изменения могут создавать биологические виды и за короткое время, предполагая реализацию альтернативной модели видообразования (об этом будет сказано ниже), хотя в целом (см. рис. 7.3.7), с очевидностью преобладает видообразование по модели 1а.

Другие факты также предупреждают о неправомочности упрощенных заключений о способах видообразования. Очевидно, что генетический «вес» вида, скажем в шкале D_n, может быть различным для различных таксонов животных. Например, у амфибий, в среднем $D_n = 1.1$, что на порядок величины больше, чем соответствующая величина у птиц, $D_n = 0.1$ (рис. 7.3.8). Напомним, что эти данные согласуются с иммуногенетическими данными (Глава 4, Раздел 4.3).

Скорость молекулярных изменений или темпы морфологической эволюции, или оба вместе, могут быть не одинаковыми в различных филетических линиях, что может вести к различиям средних значений признаков, например генетических расстояний, для таксонов одного ранга у птиц и лягушек, как это показано для уровня рода (рис. 7.3.8). Другой источник тоже может добавить неопределенности в этом вопросе, а именно – это популярность и большая изученность одних таксонов по сравнению с другими. Базы данных для млекопитающих и птиц много богаче, чем в других группах, что может вести к переоценке ранга таксона в одном случае и недооценки – в другом. В любом случае, нужно быть очень осмотрительным, определяя какой-нибудь простой, универсальный критерий ранга вида, рода и т.д., типа D_n.

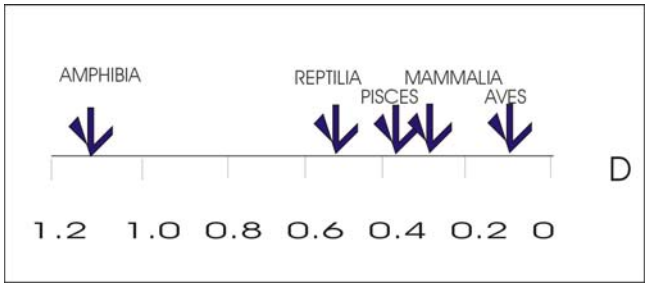


Рис. 7.3.8. Генетические расстояния (D_n) для видов одного рода, представляющих различные группы позвоночных животных (По Avise, Aquadro, 1982, с изменениями).
Fig. 7.3.8. Genetic distances (D_n) among species in separate animal genera (Avise, Aquadro, 1982).

Мы можем теперь принять, что видообразование не обязательно сопряжено с изменениями в структурных генах, которые могут быть очень незначительными (на уровне обычном для популяций видов). В противоположность этому, тогда можно ожидать в некоторых случаях видообразования значительных перестроек регуляторных генов (Wilson, 1976). Такие данные не часты в литературе, поскольку точное исследование регуляторных сдвигов, или изменений экспрессии генов, весьма трудоемко. Кроме того, как уже отмечалось, само деление на структурные и регуляторные гены достаточно условное (Wilson, 1976). Однако, если не ставить задачу очень точного измерения различий экспрессии, то можно получить очень полезную сравнительную информацию с точки зрения изучения процесса видообразования. В частности, были обнаружены достаточно большие регуляторные различия (в величине экспрессии ферментных генов) для двух близнецовых видов гольцов, у которых до 32% локусов дивергировали в этом отношении, тогда как расстояние $D_n = 0.08$, почти на уровне, типичном для популяций одного вида (Картавцев и др., 1983). Похожие результаты были получены для группы видов *in statu nascendy*, в семействах сиговых и хариусовых рыб в озере Байкал. Разброс генетических расстояний для них составил, $D_n = 0.01 – 0.03$ (Картавцев, Мамонтов, 1983), тогда как дивергенция по уровню экспрессии (*E*) достигает 9-27% (Табл. 7.3.2).

Таблица 7.3.2. Сопоставление активности ферментов у трех форм сиговых рыб (*Coregonus*) и двух форм хариусов (*Thymallus*) (По Картавцев, Мамонтов, 1983)
Table 7.3.2. Comparison of isozyme activity in three Whitefish forms (Coregonidae) and Grayling (Thymallidae)

ЛОКУС	РАЗЛИЧИЯ АКТИВНОСТИ ФЕРМЕНТОВ (ЭКСПРЕССИИ)			
	C. autumnalis* & *C. lavaretus pidschian	***C. autumnalis* & *C. lavaretus baikalensis***	***C. lavaretus pidschian* & *C. Lavaretus baikalensis***	***T. arcticus (Black form)* & *T. arcticus (White form)***
GPDH-1*	-	-	-	++
GPDH-2*	-	+	++	-
MDH-2*	-	-	-	+
ME*	-	-	+	-
6PGD*	++	-	+	+
IDH-1*	-	-	+	-
IDH-2*	+	-	-	-
PGM-2*	++	+++	++	-
FUM*	+	-	+	-
ACPH-1*	-	-	-	+++
Доля (*E*)	**18.2±8.2**	**9.1±6.1**	**27.3±9.5**	**17.4+7.9**

Примечание. $E = n / N$, %. N – общее число локусов, n – число локусов, для которых визуально было обнаружено различие активностей. N для сиговых – 22, для хариусов – 23. Знаки обозначают: “-” – активность значительно не различается, “+” воспроизводимые различия активности, “++” – двукратные

различия, “+++” – трехкратные различия или больше.

7.4. СПОСОБЫ ВИДООБРАЗОВАНИЯ: ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИЙ ВЗГЛЯД

В первом разделе лекции было отмечено, что теория видообразования в точном научном смысле, предполагающем формальную модель и элементы предсказания будущих событий этой моделью, в эволюционной генетике отсутствует. В конкретном случае, такая теория должна предсказывать образование вида или, по меньшей мере, быть способной отличать различные способы видообразования на базе некоторых, количественно определяемых параметров и их эмпирических оценок. Сделанные в этом направлении шаги (Avise, Wollenberg, 1997; Templeton, 1998) не удовлетворяют выше обозначенным требованиям.

В связи с этим была разработана схема и сформулирован алгоритмический подход (Картавцев и др., 2002), позволяющие различать способы (модели) видообразования на основе ряда важнейших популяционно-генетических параметров, имеющихся в литературе. Этот подход может заложить рамки будущей теории - генетической теории видообразования. В качестве основы для формулирования эволюционно-генетической концепции были использованы содержательные описания, сделанные Темплетоном (Templeton, 1981). В результате создана классификационная схема для семи известных способов видообразования (рис. 7.4.1). Этот путь ведет к достаточно простой экспериментальной схеме, которая позволяет: (1) организовать дальнейшее исследование видообразования в различных группах организмов и (2) вывести аналитические выражения (уравнения) для каждого способа видообразования (рис. 7.4.2). В первой схеме (рис. 7.4.1) можно определить, какие условия необходимы для видообразования, и какие достаточны для создания вида (рис. 7.4.1). Важно, что наряду с общим определением достаточных условий, вводится четыре (1-4) измеряемых экспериментально дескрипторов (их число на самом деле может быть при необходимости увеличено), которые позволяют прояснить, как или в виде чего эти достаточные условия проявляются в конкретном способе видообразования или в потенциальной модели (рис. 7.4.1). Например, дивергентный тип видообразования (ТВ) D1 объясняет классическое географическое (или аллопатрическое) видообразование. В согласии с БКВ большие популяции изолируются (разрыв потока генов) и эволюционируют раздельно, аккумулируя мутации. Чем больше время с момента изоляции, тем большие генетические расстояния приобретаются между соответствующими таксонами (Nei, 1987, см также рис. 7.2.1 и обсуждение выше в Разделе 7.3), и, соответственно, в нашей нотации вводится дескриптор: 1. $D_T > D_S$ (рис. 7.4.1). Аналогично, при реализации способа D1, не возникает ни каких различий в уровне генетической изменчивости, как в регуляторной части генома, так и по структурным генам; поэтому вводятся: 2. $H_D = H_P$ и 3. $E_D = E_P$ (рис. 7.4.1). Наконец, при некоторых способах видообразования, могут быть важны не только уровни изменчивости и генетические расстояния, но некоторые локусы, определяющие количественные признаки, неразличимые на молекулярном уровне, но ведущие к формированию РИБ, и соответственно, вводится: ТМ – экспериментальный тест на модификацию; последний позволит также отличить эпигенетическую изменчивость от таксономически фиксированных различий (рис. 7.4.1).

Очевидно, что возможно введение каких-то других дескрипторов, полученных, например, при исследовании ДНК и иммуногенетических маркеров. Самое важное, что схема, преобразованная в формальные уравнения (рис. 7.4.2) может быть операциональной, то есть, - позволять классифицировать типы видообразования. Она была протестирована на карповых рыбах (Картавцев и др., 2002) и хорошо объяснила наши ранние данные на лососях (Картавцев и др., 1983; Картавцев, Мамонтов, 1983), о которых уже упоминалось в этой лекции. Конечно, и сама предложенная схема и теоретические основы должны быть развиты далее.

7.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 7

1. Пакет программ: “POPULUS” – Случайный дрейф генов. Отбор и дрейф генов. Проведите цифровое моделирование для двух различных ситуаций дрейфа с устанавливаемыми параметрами (N = 10 и N = 999 и частотами аллелей p=0.5).
2. Интернет-поддержка. Посетите сайт и дополните свои знания о виде и видообразовании: //:http/www.biology.ucsc.edu/people/barrylab/public_html/classes/evolution/genetic.htm

Рис. 7.4.1. Типы видообразования (ТВ): популяционно-генетическая точка зрения
Fig. 7.5.1. Speciation modes (SM): population genetic view (From Kartavtsev et al., 2002)

ДЕСКРИПТОРЫ

D – генетические расстояния по структурным генам:
D_T – в предположительном родительском таксоне,
D_S – среди конспецифичных демов,
D_D – среди подвидов или близнецовых видов;
H_D - средняя гетерозиготность в предположительной дочерней популяции;
H_P - средняя гетерозиготность в предположительной родительской популяции;
E_P – дивергенция по регуляторным генам в предположительном родительском таксоне;
E_D – дивергенция по регуляторным генам в предположительном дочернем таксоне;
TM^+- тест на модификацию (положительный);
TM^- - тест на модификацию (отрицательный).

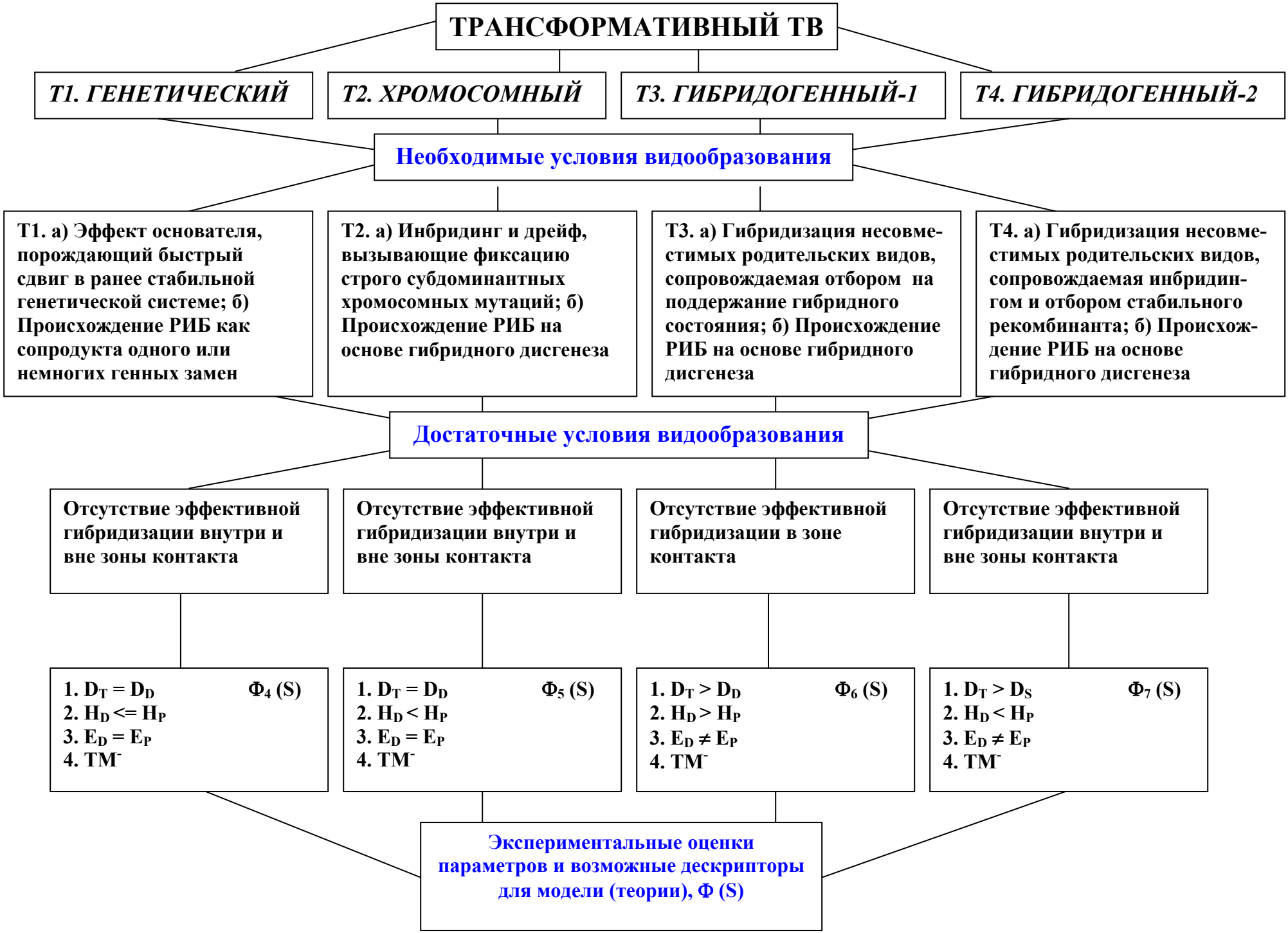

ТРАНСФОРМАТИВНЫЙ ТВ
Т1. ГЕНЕТИЧЕСКИЙ
Т2. ХРОМОСОМНЫЙ
Т3. ГИБРИДОГЕННЫЙ-1
Т4. ГИБРИДОГЕННЫЙ-2
Необходимые условия видообразования
Т1. а) Эффект основателя, порождающий быстрый сдвиг в ранее стабильной генетической системе; б) Происхождение РИБ как сопродукта одного или немногих генных замен
Т2. а) Инбридинг и дрейф, вызывающие фиксацию строго субдоминантных хромосомных мутаций; б) Происхождение РИБ на основе гибридного дисгенеза
Т3. а) Гибридизация несовместимых родительских видов, сопровождаемая отбором на поддержание гибридного состояния; б) Происхождение РИБ на основе гибридного дисгенеза
Т4. а) Гибридизация несовместимых родительских видов, сопровождаемая инбридингом и отбором стабильного рекомбинанта; б) Происхождение РИБ на основе гибридного дисгенеза
Достаточные условия видообразования
Отсутствие эффективной гибридизации внутри и вне зоны контакта
Отсутствие эффективной гибридизации внутри и вне зоны контакта
Отсутствие эффективной гибридизации в зоне контакта
Отсутствие эффективной гибридизации внутри и вне зоны контакта
1. $D_T = D_D$ Φ_4 (S)
2. $H_D <= H_P$
3. $E_D = E_P$
4. TM^-
1. $D_T = D_D$ Φ_5 (S)
2. $H_D < H_P$
3. $E_D = E_P$
4. TM^-
1. $D_T > D_D$ Φ_6 (S)
2. $H_D > H_P$
3. $E_D \neq E_P$
4. TM^-
1. $D_T > D_S$ Φ_7 (S)
2. $H_D < H_P$
3. $E_D \neq E_P$
4. TM^-
Экспериментальные оценки параметров и возможные дескрипторы для модели (теории), Ф (S)

Рис. 7.4.2. Аналитическое представление семи способов видообразования.

Описание дескрипторов в уравнения дано на рис. 7.4.1. D1 – D3 – это дивергентные способы видообразования, T1 – T4 – трансформативные способы видообразования.

Fig. 7.5.2. Analytical description of seven types of speciation modes.

Descriptors for equations are given in Fig. 6.4.1. D1 – D3 are divergence speciation modes, T1 – T4 are transilience speciation modes.

$$\Phi_1 (S) \in \{(D_T > D_S) \subset (E_D = E_P) \subset (H_D = H_P) \subset TM^-\} \qquad \textbf{(D1)}$$

$$\Phi_2 (S) \in \{(D_T > D_S) \subset (E_D \neq E_P) \subset (H_D = H_P) \subset TM^-\} \qquad \textbf{(D2)}$$

$$\Phi_3 (S) \in \{(D_T = D_S) \subset (E_D \neq E_P) \subset (H_D <= H_P) \subset TM^+\} \qquad \textbf{(D3)}$$

$$\Phi_4 (S) \in \{(D_T = D_D) \subset (E_D \neq E_P) \subset (H_D <= H_P) \subset TM^-\} \qquad \textbf{(T1)}$$

$$\Phi_5 (S) \in \{(D_T = D_D) \subset (E_D = E_P) \subset (H_D > H_P) \subset TM^-\} \qquad \textbf{(T2)}$$

$$\Phi_6 (S) \in \{(D_T > D_D) \subset (E_D \neq E_P) \subset (H_D > H_P) \subset TM^-\} \qquad \textbf{(T3)}$$

$$\Phi_7 (S) \in \{(D_T > D_S) \subset (E_D \neq E_P) \subset (H_D < H_P) \subset TM^-\} \qquad \textbf{(T4)}$$

ГЛАВА 8. ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКАЯ СТРУКТУРА ВИДА

ГЛАВНЫЕ ВОПРОСЫ:

8.1. Популяционная структура бисексуальных видов.
8.2. Типы внутривидовой структуры.
8.3. Принципы исследования внутривидовой структуры.
8.4. Полевое популяционно-генетическое исследование.
8.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 8.

РЕЗЮМЕ

1. Основными параметрами, определяющими генетическую динамику популяций являются: (1) ***p*** – **частота аллеля**, (2) ***Ne*** – **генетически эффективная численность популяции**, (3) ***m*** – **коэффициент миграции**, определяющий поток генов и (4) ***s*** – **коэффициент естественного отбора**.
2. Исследование **внутривидового генетического разнообразия** требует **комплексного подхода**, принимая во внимание комплексность механизмов, которые определяют внутривидовую изменчивость и дифференциацию.

Chapter 8. POPULATION GENETIC STRUCTURE OF A SPECIES

SUMMARY

1. The main parameters, which determined genetic dynamics of populations are (1) ***p***, **allele frequency**, (2) ***Ne***, **effective size of population**, (3) ***m***, **coefficient of migration** defining a gene flow, and (4) ***s***, **coefficient of natural selection**
2. The **intraspecies genetic diversity** research requires a **complex approach**, taking into consideration complexity of factors that cause variability and differentiation within a species.

8.1. ПОПУЛЯЦИОННАЯ СТРУКТУРА БИСЕКСУАЛЬНЫХ ВИДОВ

Давайте, определим для начала, о каких видах и о каких популяциях пойдет речь.

Вид.

В Главе 7 мы дали понятие вида согласно биологической концепции (БКВ). Это определение на основе БКВ наиболее органично соответствует популяционно-генетической теории, и именно поэтому мы его используем в дальнейшем, не смотря на ограничения, о которых упоминалось (Глава 7). Популяции, которые составляют вид по БКВ, могут быть легко определены в генетических терминах, и мы рассмотрим ряд свойств таких популяций в этой и последующих Главах именно в генетической плоскости. Естественно, наше рассмотрение не претендует на широту и объемность специальных руководств по другим аспектам популяционной биологии (Тимофеев-Ресовский др., 1973; Яблоков, 1987), к которым мы и отсылаем заинтересованных читателей.

Вид, о котором мы говорим, это также таксономическая единица. Виды и подвиды представляют наименьшие таксономические категории из нескольких, о которых мы здесь напомним (рис. 8.1.1).

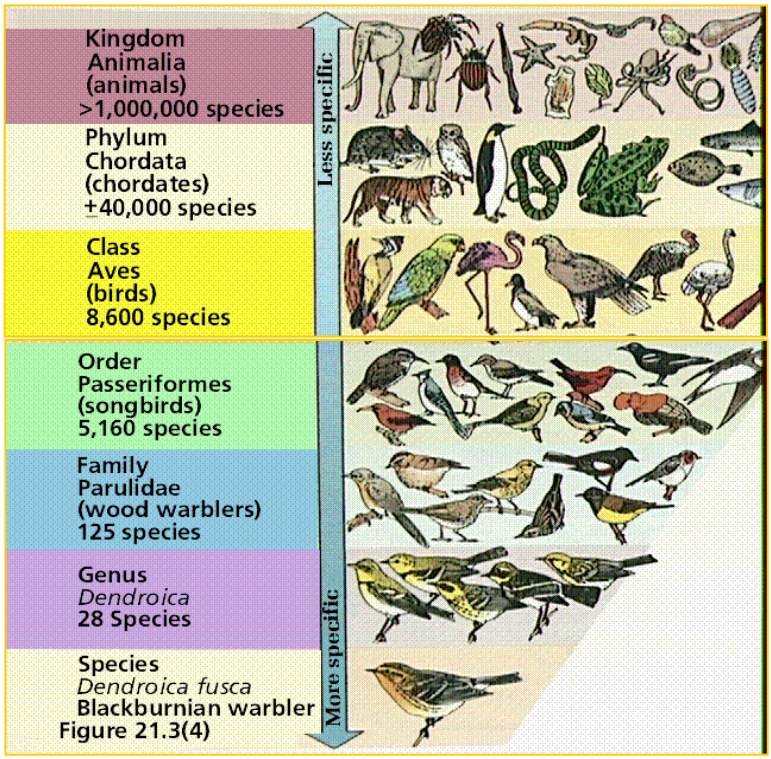


Рис. 8.1.1. Примеры основных таксономических категорий: царство (Kingdom), тип (Phylum), класс (Class), отряд (Order), семейство (Family), род (Genus) и вид (Species) (По On-Line Biology Book, 2002). Fig. 8.1.1. Examples of main taxonomic categories: Kingdom, Phylum, Class, Order, Family, Genus, and Species (From On-Line Biology Book, 2002).

Популяции

Популяции могут быть определены несколькими способами (Яблоков, 1987). Лебедев (1946) использовал понятие **элементарная популяция**, чтобы определить наименьшую единицу вида, а в целом им выделены такие подразделения вида: подвид (географическая раса), экологическая раса, стадо, элементарная популяция. Однако Лебедев не считал элементарные популяции наследственными группировками, то есть, самовоспроизводящимися репродуктивными группировками. Янулов (1962) использовал термин **стадо**, как самовоспроизводящуюся единицу, а Марр (Marr, 1957) полагал, что самовоспроизводящейся единицей является **субпопуляция**. Алтухов (1974) определил **локальное стадо** (**локальную популяцию**): *«Локальное стадо - это система субпопуляций, которая эквивалентна в теоретическом отношении **подразделенной популяции**, и является наименьшей, стабильной репродуктивной единицей вида»*. Изложение ряда популяционных понятий и многих аспектов внутривидовой организации и проблем вида дано Алтуховым (1974; 1983; 1989), Яблоковым (1987), Иванковым (1998).

Появились и более поздние определения популяции (Мина, 1986; Яблоков, 1987 и др.). По своей близости к генетике, кажется вполне удовлетворительным такое определение. «Популяция – это репродуктивная группировка организмов, которые обладают общим генофондом» (Dobzhansky, 1951; 1970). Такую популяцию часто определяют термином **дем** или **Менделевская популяция**. И локальное стадо и субпопуляция подпадают под это определение. Алтухов с коллегами (Алтухов, Рычков, 1970; Алтухов, 1974; 1989; Алтухов и др., 1997) представили веские доказательства, что подразделенные популяции, такие как локальные стада, являются более стабильными единицами, как в пространстве, так и во времени. Именно они являются единицами воспроизводства и эволюции, представляя также единицы запаса в промысле. Яблоков (1987, с.56) также считает, что единицами самовоспроизводства внутри вида являются группировки более высокого ранга, чем демы. Однако могут быть и исключения. В частности, у северной креветки единицами воспроизводства являются панмиктические Менделеевские популяции (см. Раздел. 8.4), а, в принципе, возможно, что и вид в целом будет единицей воспроизводства. Это зависит от особенностей биологии организмов. В нашем введении необходимо подчеркнуть, что имеется два

использования термина популяция: в смысле определения Добжанского и в общем смысле, как любой группы особей, не предполагая генетической нагрузки. В последнем случае временные скопления, типа стай и экологических групп могут именоваться популяциями. Лучше избегать такого второго смысла.

Чтобы подвести итог этой дискуссии, суммируем схематически представление о внутривидовой структуре (рис. 8.1.2).

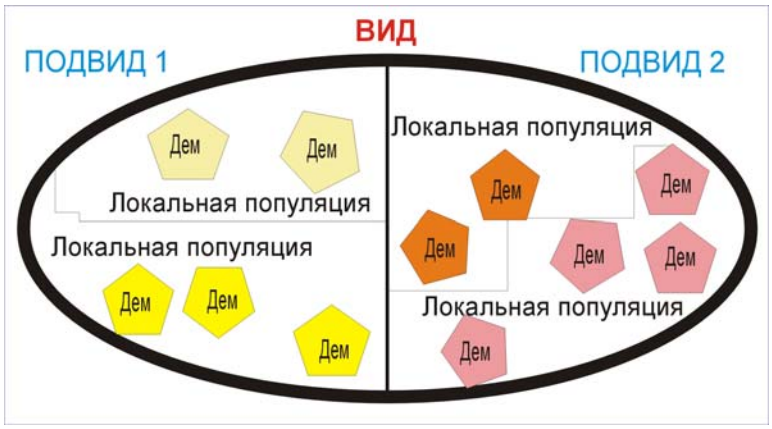


Рис. 8.1.2. Главные внутривидовые подразделения: 1) подвиды (географические расы), 2) локальные популяции (стада), 3) демы (субпопуляции, Менделевские популяции).
Fig. 8.1.2. Main intraspecies divisions: 1) Subspecies (geographic race), 2) Local population (stock), 3) Deme (subpopulation, Mendel population).

В данной классификации осознанно выпущены такие внутривидовые группировки, как экологическая раса и экотип, которые многие авторы (Никольский, 1968; Коновалов, 1980; Решетников, 1980; Иванков, 1998) включают в иерархический ряд внутривидовых категорий. Биологическая раса – это довольно сложное понятие (Коновалов, 1980; Иванков, 1998). В классическом понимании расы – это экологические, не наследственные группировки. В их формировании принципиальную роль играет модификационная изменчивость и механизм их детерминации эпигенетический. Хотя на этот счет имеются и иные мнения (Решетников, 1980). Таким образом, надо четко отделить внутривидовую генетическую изменчивость от эпигенетической изменчивости. С другой стороны, часто, по-видимому, за расы могут приниматься различные локальные стада, в местах их симпатрии, как это иногда имеет место у горбуши (Гриценко, 1981; Картавцев, 1986). Здесь требуется отдельная большая дискуссия, что, однако, выходит за рамки данного курса.

8.2. ТИПЫ ПОПУЛЯЦИОННОЙ СТРУКТУРЫ

Имеются два различных подхода в описании популяционной структуры.

Феноменологический подход

Пример такого подхода к популяции можно найти в книге Кирпичникова (1979). При подобном подходе выделено четыре типа популяционной структуры: 1) тип бельдюги, 2) тип сазана, 3) тип тихоокеанских лососей и 4) тип тунца.

На каком основании выделяется первый тип? Бельдюга – это малоподвижная, живородящая рыба с небольшой плодовитостью. Поэтому, видовой ареал населяют небольшие по численности, хорошо изолированные популяции. Следующий тип наиболее характерно представляет сазан, населяющий различные речные и/или озерные системы, хорошо изолированные друг от друга и населенные весьма многочисленными скоплениями рыб, которые мигрируют, но не очень далеко. Типичным представителем третьего типа являются виды тихоокеанских лососей (род *Oncorhynchus*). Каждая река, согласно концепции Кирпичникова, населена отдельным, изолированным локальным стадом, чья интегрированность поддерживается хомингом (инстинктом возврата в реку рождения). Наконец, четвертый тип представлен тунцами (и подобными организмами), пелагическими рыбами с очень высокой подвижностью и большой плодовитостью, которые дают огромные популяции, слабо дифференцированные генетически.

Что в этом подходе не верно. Давайте разберем в качестве примера тип 3. Даже среди наиболее многочисленных шести видов тихоокеанских лососей нет единства в смысле популяционной структурированности. Хоминг у горбуши довольно слабый, тогда как сима и нерка обладают весьма совершенным хомингом. Численность стад у горбуши может достигать миллионов экземпляров нерестовых особей, а у симы эти нерестовые группы представлены только сотнями или тысячами особей. Эти особенности биологии ведут к слабой у горбуши и сильной у симы и нерки генетической дифференциации. Даже еще более контрастна, как мы увидим позже, внутривидовая генетическая дифференциация лососей-форелей рода *Salmo* и тихоокеанских лососей рода *Oncorchynchus*.

Именно поэтому необходим другой подход для описания популяционной структуры. Можно назвать такой подход научным подходом.

Научный подход

Мы рассмотрим данный подход в Разделе 8.3. Главные идеи этого подхода базируются на теории популяционной генетики и, таким образом, имеют точные научные решения. Надо подчеркнуть, что иногда эти решения трудно достижимы на практике из-за малых баз данных, или, менее часто, из-за того, что сама теория в некоторых частях не достаточно развита.

8.3. ПРИНЦИПЫ ИССЛЕДОВАНИЯ ВНУТРИВИДОВОЙ СТРУКТУРЫ

Главные факторы популяционно-генетической динамики

Среди факторов, влияющих на динамику, имеются случайные, или стохастические, и систематические. Например, случайный дрейф генов – это один из наиболее важных стохастических факторов генетической дифференциации, особенно в малых популяциях. Естественный отбор и миграция – это систематические факторы. Основные популяционно-генетические параметры, которые определяют степень генетической дифференциации, это: (1) p – частота аллеля (мы определили ее ранее в Лекции 4, Раздел 4.2), (2) Ne – генетически эффективная (или просто эффективная) численность популяции, (3) m – коэффициент миграции, (4) s – коэффициент естественного отбора. Подробно эти параметры будут рассмотрены, наряду с другими аналитическими выкладками в следующей лекции.

Сейчас, мы лишь проиллюстрируем действие факторов, начав с первого и второго параметров. Давайте рассмотрим дрейф генов во времени для одной популяции небольшого размера. Дрейф генов возникает как результат стохастических ошибок при передаче генов из одного поколения в следующее поколение. Пусть численность популяции и частота аллеля (скажем, A) в ней равны Ne = N = 10 и $p_A = 0.5$, в исходном поколении (t_0). В этом случае можно рассчитать стандартную ошибку (припомним, как мы это делали в Практическом занятии 4 на основе выражения 4.3): $SE = \sqrt{[(0.5*0.5)/20]} = 0.16$. То есть, в следующем поколении p_A должна быть $0.5 \pm SE$. Другими словами, некоторые из выборок генофонда, дающие следующее поколение, могут иметь частоты: $p_A = 0.7$ или $p_A = 0.3$ (33% вероятности; см. Лекцию 11 для пояснений), или даже $0.5 \pm 2SE$, $p_A = 0.9$ или $p_A = 0.1$ (5% вероятности), с некоторыми упрощениями для большей ясности. Очевидно, что очень скоро в такой маленькой популяции десяти особей (N = 10) определенный аллель будет либо фиксирован ($p_A = 1$), либо утерян ($p_A = 0$). В пакете программ POPULUS на цифровых примерах мы наблюдали в динамике изменения частот в поколениях, для различных параметров p и Ne. Выше мы оперировали данными, полученными на основе теоретических соображений. Прямые экспериментальные данные подтверждают их (рис. 8.3.1).

Все сделанные заключения остаются верными для случаев, когда дрейф происходит в пространстве (Wright, 1943, Li, 1955).

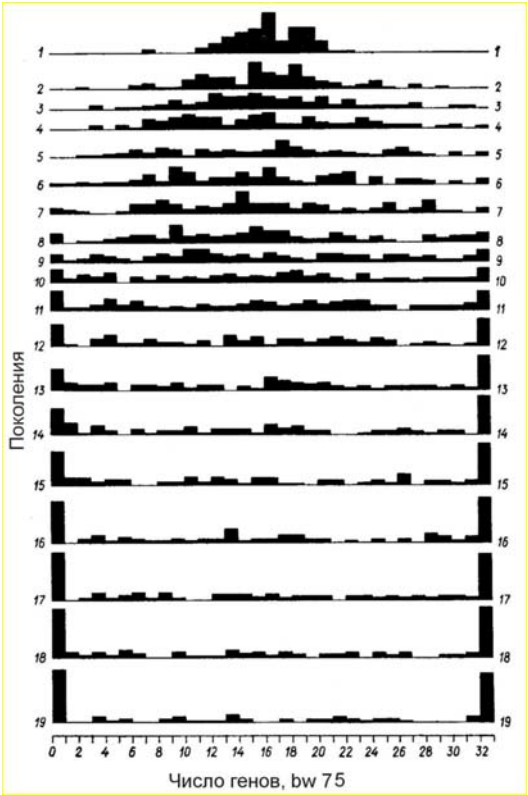


Рис. 8.3.1. Дрейф гена bw75 и частотное распределение в выборках из 100 экспериментальных популяций Drosophila melanogaster (По Buri, 1956; заимствовано из Алтухов, 1989).

Fig. 8.3.1. Drift and gene bw75 frequency distribution in the generations of 100 experimental populations of Drosophila melanogaster (From Buri, 1956; after Altukhov, 1989).

Как влияет частота гена p на изменчивость? Начальное значение частоты прямо влияет на число поколений, прошедших до фиксации или утери аллеля при рассмотрении динамики, определяемой только Ne (дрейфом). Очень наглядно это видно при моделировании с помощью POPULUS или другого подходящего пакета программ. В приведенном выше цифровом примере, если в t_0 частота аллеля $p_A = 0.9$, то при N = 10, SE = 0.07 и, соответственно, частота аллеля в следующем поколении может быть $p_A = 0.97$ или $p_A = 0.17$ (33% вероятности) или $p_A = 1.0$ или $p_A = 0.24$ (5% вероятности). То есть, при исходной частоте 0.9 фиксация аллеля A, или его переход в мономорфное состояние, может произойти уже в первом поколении.

Миграция и естественный отбор – это два следующих, из четырех важнейших факторов, влияющих на популяционную структуру. Мы рассмотрим подробнее влияние этих двух факторов в следующей лекции. Миграция влияет не только на интегрированность генофонда, но и на эффективную численность популяций, фактически объединяя отдельные субпопуляции при высоком потоке генов и создавая их слабые генетические различия (рис. 8.3.2). Фактические данные на рис. 8.3.2 для морских беспозвоночных соответствуют высоким расчетным (модельным) значениям миграции или потока генов. Поскольку аналитически m (коэффициент миграции) и Ne связаны пропорциональной зависимостью (см. Глава 9, Раздел 9.3, выражение 9.23), то это ведет к высоким Ne в популяциях этих организмов.

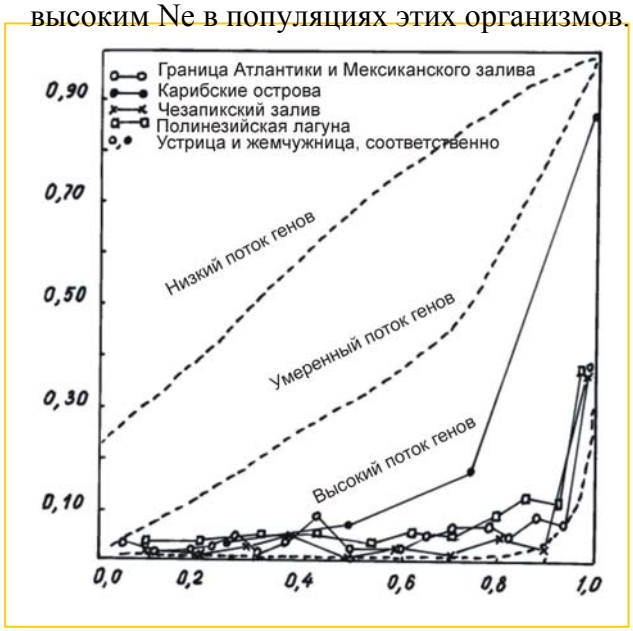


Рис. 8.3.2. График распределения условных средних частот аллелей (p_i) для некоторых популяций устрицы и жемчужницы, иллюстрирующий роль миграции в объединении популяций и в увеличении величины Ne.

По ординате - p_i, по абсциссе p_i, нормализованные на размер выборки. Прерывистые линии показывают расчетные уровни потока генов, которые соответствуют определенным значениям p_i (По Burroker, 1984 и Durand, Blank, 1986).

*Fig. 8.3.2. Graph of distribution of a conditional mean allele frequencies (**p_i**) within some species of oysters and pearl clam that illustrate the role of gene flow in creating high **Ne**.*

*On ordinate is **p_i**, on abscissa is **p_i** that normalized on sample number. Broken lines show gene flows, which correspond to certain **p_i** (From Burroker, 1984 and Durand, Blank, 1986).*

Влияние естественного отбора многогранно. Практически все процессы в популяции отбор может усиливать или тормозить. Влияние дрейфа генов при действии стабилизирующего отбора уменьшается, а при направленном отборе фиксация частоты происходит с увеличенной скоростью. Мы разберем эти вопросы, промоделировав на численных примерах влияние параметра s на динамику частот аллелей (Практическое занятие 8). Имеется множество разработанных моделей действия тех или иных факторов отбора. Часть из них будет рассмотрена в Главах 9 и 12, а желающим получить более подробную информацию на эту тему можно обратиться к другим источникам (Wright, 1968; Wright, 1969; Wright, 1978; Hartl, Clark, 1989).

Мера межпопуляционной дифференциации

В качестве меры степени генетической дифференциации между популяциями, или наоборот внутри них, полезными являются статистики Fst и Dm, первая – это стандартизированная дисперсия частот аллелей, а вторая – минимальное генетическое расстояние. Использование Fst (или Gst) при популяционном анализе является более предпочтительным, как мы увидим в Лекции 9 вместе с некоторыми деталями свойств статистик. Простейший способ понять, что такое Fst – это определить данную меру дифференциации в терминах гетерозиготности (разнообразия), как нормированную гетерозиготность: Fst = (Ht - Hs) / Ht. Здесь Hs – это ожидаемая гетерозиготность в субпопуляции (или внутри выборки), а Ht – это ожидаемая гетерозиготность во всей популяции (или в совокупности выборок). В этом определении Fst = Gst (Nei, 1977b).

Простейшая модель популяционной структуры

Мы уже имели дело с этой моделью. Это – соотношение Харди-Вайнберга: $(p + q)^2 = p^2 + 2pq + q^2$. Это равновесная модель для большой (бесконечной в теории) популяции случайно скрещивающихся двуполых особей. Она объясняет стабильность частот аллелей и в меньшей степени частот генотипов при отсутствии давления систематических факторов, основными из которых, в текущем поколении являются естественный отбор и миграция.

Комплексная природа генетической динамики и стационарности.

В природе ни один из перечисленных факторов не действует по отдельности. Обычно они взаимодействуют в различных комбинациях и варьируют сами в пространстве, в различных частях ареала, и во времени, в поколениях, от молодых возрастных групп - к взрослым, или в онтогенезе, и даже в пределах той же самой возрастной группы в течение сезонов года. Классическим примером изменчивости в природной популяции (никто еще не повторил с середины 70-х голов) является полиморфизм двух аллозимных локусов нерки, который подвержен влиянию двух факторов (рис. 8.3.3). По локусу *PGM** кривая, построенная с учетом некоторого давления отбора вместе с миграцией и вариацией

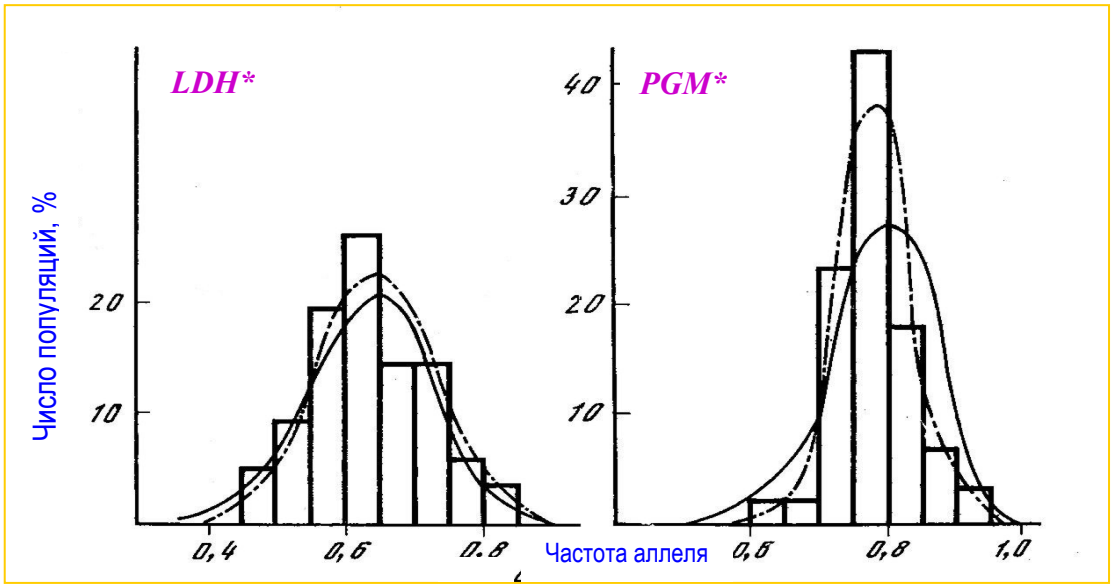


Рис. 8.3.3. Распределение частот аллелей по локусам LDH и PGM* в нерестовых скоплениях нерки (O. nerka) в озере Азабачье (Камчатка) (По Алтухов и др., 1975).*

*Гистограмма показывает фактическое распределение частот. Пунктирной линией показаны частоты, ожидаемые согласно островной модели Райта при действии дрейфа, миграции и естественного отбора, определяемых тремя параметрами – **Ne, m,** и **s.** Непрерывной линией показана вариация, зависимая только от **Ne** и **m** (дрейф и миграция).*

Fig. 8.3.3. Allele frequency distribution at two loci in spawning subpopulations of sockeye (O. nerka) in Azabachee Lake (Kamchatka) (From Altukhov et al., 1975).

*Histogram represents observed frequencies. With dotted line the island model expectations are shown for **Ne, m,** and **s** parameters in action (drift, migration and selection), with solid line - **Ne** and **m** (drift and migration).*

Ne, лучше объясняет фактические данные (рис. 8.3.3). Полиморфизм по *LDH** является селективно нейтральным и зависит только от двух параметров - ***Ne*** и ***m*** (рис. 8.3.3).

Классификация вариантов внутривидовой структуры, на основе использования популяционно-генетических средств

Если основываться на средствах популяционной генетики, то можно определить три главных варианта популяционной организации: (1) панмиктическая популяция, (2) подразделенная популяция и (3) непрерывная популяция. В данном подразделе мы рассмотрим кратко (а позже в Лекции 9 - подробнее) первый и второй варианты. Третий вариант несколько более труден в математическом смысле, а главное – достаточно редок в природе.

(1) Панмиктическая популяция. Этот вариант популяционной организации теоретически описывается моделью Харди-Вайнберга. Внутри такой популяции не существует генетической дифференциации. Поэтому, когда нет дифференцирующего действия отбора, то выполняется условие $Fst = 0$. Вид может быть представлен на ареале одной или несколькими популяциями, которые полностью, или почти полностью, изолированы одна от другой.

(2) Подразделенная популяция. Теоретически описана как **островная модель** популяционной структуры (Wright, 1943), или как **модель ступенчатой миграции** (Kimura, Ohta, 1971). Дифференциация в подразделенной популяции в равновесном состоянии, в отсутствие отбора, зависит, в основном, от двух факторов – миграции и эффективной численности (размера) приближенно через соотношение: $Fst = 1 / [4Nm + 1]$.

Особенности популяционной организации морских организмов

Имеются некоторые особенности внутривидовой структурированности в морской среде. Это связано с обычно весьма многочисленными популяциями, которые имеют огромный потенциал миграции или расселения (активного или пассивного). Однако все генетические

принципы также сохраняют свою верность и для этих популяций. Например, имеются в среднем большие и малые значения *Fst* для разных видов, что связано, главным образом, с большой и небольшой генетической дифференциацией, зависимой от обратно пропорционально связанных дрейфа и миграции (см. рис. 8.3,2, рис. 8.4.5). Давайте рассмотрим несколько общих данных о генетической изменчивости и дифференциации, а затем вернемся к особенностям популяционной структурированности морских организмов.

Генетическая изменчивость и дифференциация

Ранее (Глава 7, см. рис. 7.3.2) мы привели обобщение о генетической изменчивости. Средняя гетерозиготность на особь, ***H*** широко варьирует в таксонах растений и животных. Суммарная средняя составила: ***H***=0.076, для позвоночных - ***H***=0.054 и для беспозвоночных - ***H***=0.100 (Nevo et al., 1984).

Частоты коэффициентов генетического сходства по ферментным локусам, оцененные для разных видов, следуют U-образному распределению (рис. 8.3.4); при нейтральности же ожидается обратная связь с ожидаемой дифференциацией (Ayala, 1975), то есть, близкое к нормальному распределение.

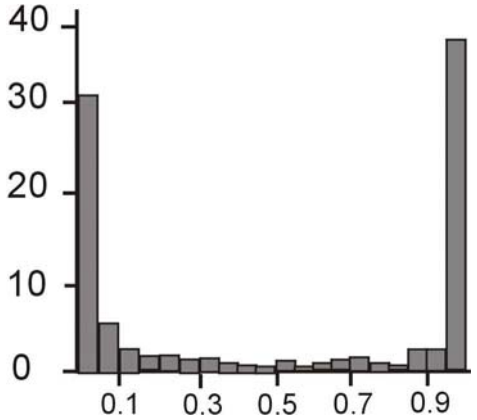


Рис. 8.3.4. Ось Y – частота (%), ось X – коэффициент сходства, I среди видов группы Drosophila willistoni. Большинство локусов обнаруживают либо близкое к единице, либо нулевое сходство, тогда как остальные варьируют в диапазоне от малых до больших коэффициентов сходства. U-образное распределение не согласуется с дрейфовой дифференциацией при нейтральности белковых маркеров. Распределение должно быть близким к нормальному (По Ayala, 1975).
Fig. 8.3.4. Y-axis, Frequency (%), X-axis, Similarity index, I in Drosophila willistoni species group. Most of loci show either fix or no similarity, while the rest exhibit from moderate to high levels of similarity. The U-shaped distribution is not agrees with drift or neutral theory differentiation in time. Rather it should be close to normal distribution (From Ayala, 1975).

Тем не менее, для некоторых белковых локусов можно обнаружить распределение коэффициентов сходства, которое близко к нормальному (рис. 8.3.5). Таким образом, по меньшей мере, по некоторым локусам обнаруживаемая дифференциация во времени согласуется с дрейфовым, нейтральным процессом.

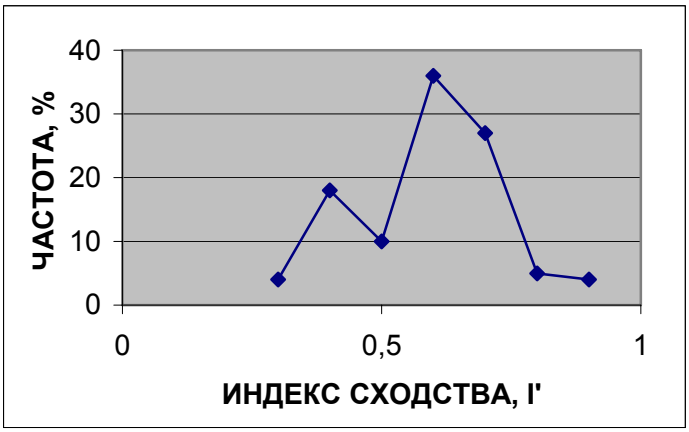


Рис. 8.3.5. Частотное распределение коэффициента сходства, рассчитанного по дуплицированным гемоглобиновым локусам лососевых рыб (По Картавцев, 1995).
Fig. 8.3.5. Frequency distribution of a similarity index at duplicated hemoglobin loci in different salmon species (From Kartavtsev, 1995).

Некоторые исследования показывают, что без привлечения концепции естественного отбора вряд ли возможно объяснить сопряженную изменчивость индивидуальной гетерозиготности по ферментным генам (H_o) и морфологическую изменчивость в популяционно-средовом градиенте (рис. 8.3.6).

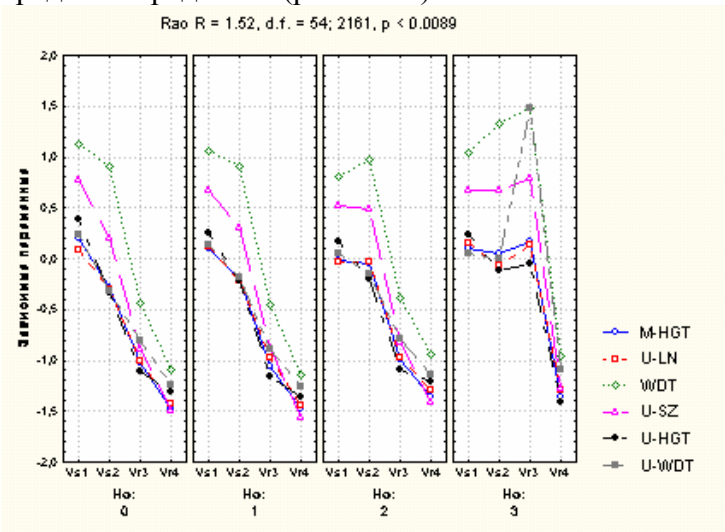


Рис. 8.3.6. График средних, который иллюстрирует результаты двухфакторного дисперсионного анализа: H_O и изменчивости морфологических признаков у брюхоногого моллюска (Nucella heyseana) в заливе Восток (Vs1-Vs2) и заливе Находка (Vr3-Vr4) в заливе Петра Великого Японского моря.

Увеличение размера тройных гетерозигот (H_O =3) в выборке Vr3 затруднительно объяснить не привлекая концепцию отбора в пользу гетерозигот (По Kartavtsev, Svinyna, 2003).
Fig. 8.3.6. Plot of mean values, which illustrate the results of two-factor variance analysis: the H_O and morphology traits variability in mollusk (Nucella heyseana) in Vostok and Nakhodka bays (Peter the Great Bay, Sea of Japan). Increase of size in the group of tri-heterozygotes (H_O=3) in two samples requires selection as an explanation of these results (From Kartavtsev, Svinyna, 2003).

Как продемонстрировано ранее, источники, поддерживающие генетическую изменчивость в популяциях, имеют комплексную природу (рис. 8.3.2, 8.3.4-8.3.6). Иногда совершенно очевидно, что ни естественный отбор, ни нейтральность не являются причиной того или иного типа обнаруженной изменчивости. Неоднородность самого материала и соответствующих данных могут вести к корреляциям или атипичным распределениям (рис. 8.3.7).

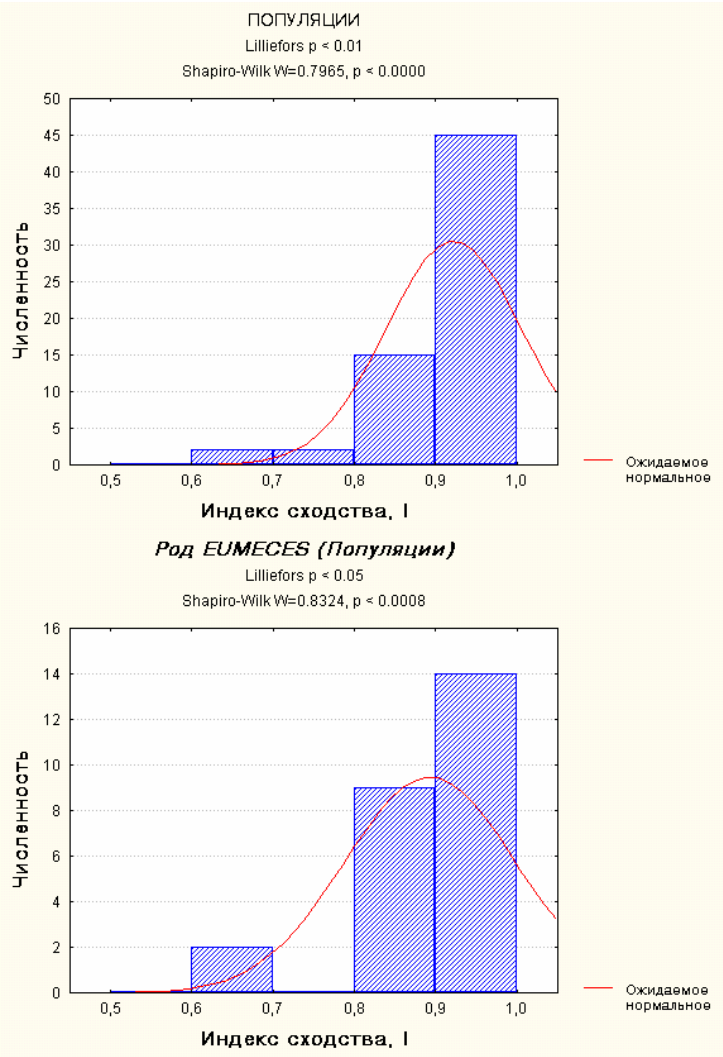


Рис 8.3.7. Частотное распределение коэффициента сходства (I) для внутривидового уровня.

Верхний график показывает распределение I для всей базы данных, тогда как нижний только для рептилий рода Eumeces (Картавцев, не опубликовано).

Fig 8.3.7. Frequency distribution of a similarity index (I) within species (Data of the author, unpublished).

Upper figure show I distribution for the whole database, while bottom for only one reptile genus Eumeces.

Очевидна ненормальность обоих распределений. Бимодальность на нижнем графике обусловлена гетерогенностью данных, представленных небольшим числом очень несходных островных популяций этих ящериц (I = 0.6-0.7 и менее; Kato et al., 1994) (рис. 8.3.7). Эти данные предполагают, что имеется гетерогенность среди популяционных групп по степени сходства и они дифференциально представлены в материале по внутривидовой изменчивости.

Если сравнивать более однородные группы, то нормальность можно обнаружить (рис.8.3.8).

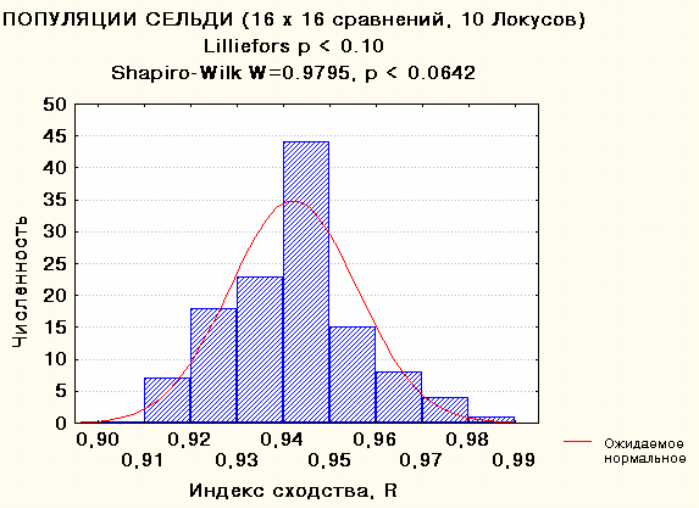
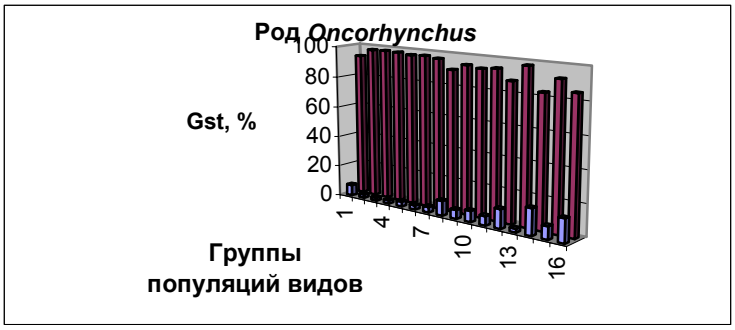


Рис. 8.3.8. Частотное распределение коэффициента сходства (R_C) внутри вида тихоокеанская сельдь, Clupea pallasi (Перерасчет по данным Картавцев, Рыбникова, 1999).

Fig. 8.3.8. Frequency distribution of a similarity index (R_C) within the species of Pacific herring, Clupea pallasi (Calculated and plotted from data in Kartavtsev, Rybnikova, 1999).

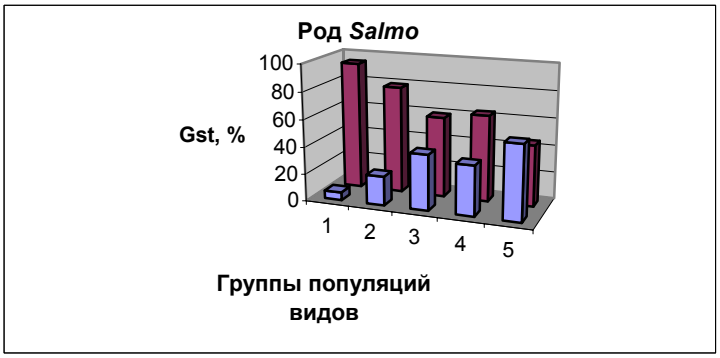
Данные о внутривидовой дифференциации показывают, что единообразия популяционной дифференциации не наблюдается ни у форелей рода *Salmo*, ни у тихоокеанских лососей рода *Oncorhynchus*. Хотя можно сделать вывод, что в обеих группах преобладает внутривидовой компонент (рис. 8.3.9).

*Рис. 8.3.9. Величина **Gst** (Nei, 1973, 1977b) как мера внутривидовой дифференциации лососевых рыб (Данные автора, не опубликованы).*
*Fig. 8.3.9. Values of **Gst** (Nei, 1973, 1977) as measurements of intraspecies differentiation in salmon (Data of author, unpublished).*
1 - Inter-, 2 – Intra-population components.

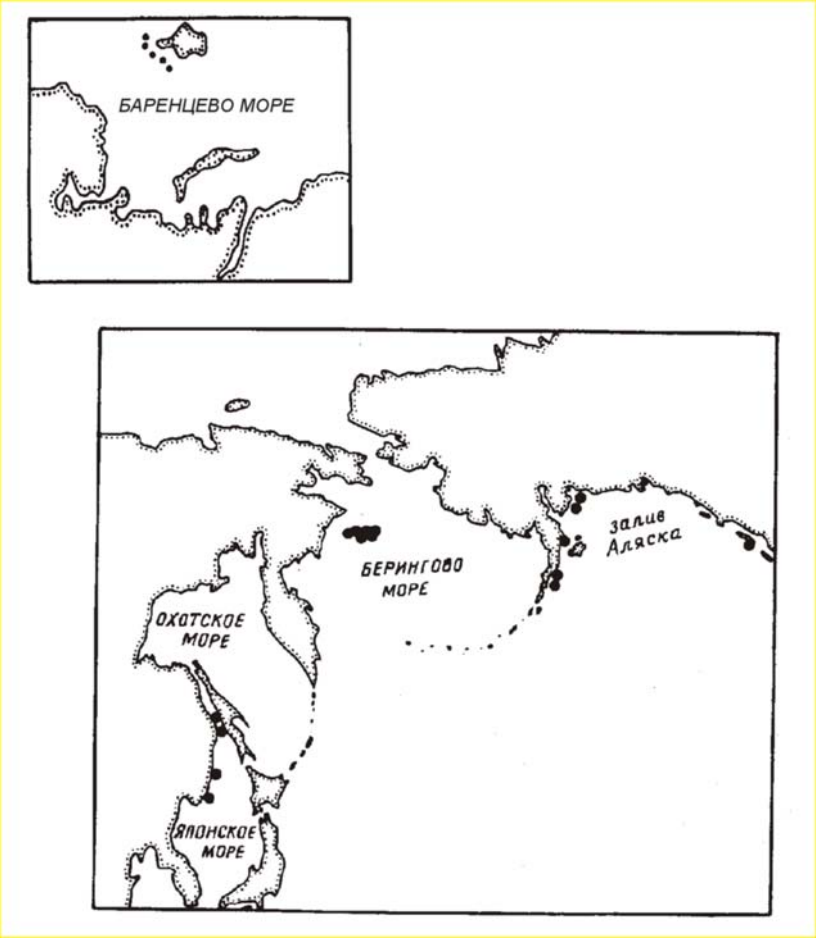


Данные на рис. 8.3.9 иллюстрируют: (1) относительно небольшую межпопуляционную генетическую дифференциацию (*Gst* = 7%, верхний рисунок)) по сравнению с внутрипопуляционным разнообразием (*Gst* = 93%) (верхний рисунок) и (2) различие величин *Gst* между родами *Salmo* и *Oncorchynchus*, с одной стороны, и между видами в каждом таксоне (средний и нижний рисунки). Виды и популяции у форелей более дифференцированы генетически, чем у тихоокеанских лососей, о чем свидетельствуют статистически значимые различия Gst (см. рис. 7.3.3, Глава 7).

Причины, обусловливающие различные уровни генетической дифференциации, могут быть различными. Как отмечалось, одной из самых важных причин может быть различие эффективной численности популяций этих двух групп (см. Глава 7).



Чем больше популяционный Ne, тем меньше генетический дрейф и, соответственно, меньше генетическая дифференциация популяций и вида в целом (см. детали в Лекции 9). По меньшей мере, для позвоночных и морских беспозвоночных, значимые различия Ne несомненны (рис. 7.3.4, Глава 7).



8.4. ПРИМЕР ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКОГО ИССЛЕДОВАНИЯ

Исследование внутривидовой популяционной структуры у северной креветки *Pandalus borealis* (Kartavtsev et al., 1991; 1993; Kartavtsev, 1994).

Исследование основывалось на совокупности выборок, охватывающих практически весь ареал вида (рис. 8.4.1).

Рис. 8.4.1. Карта исследованного ареала.

Точки – места взятия выборок.

Fig. 8.4.1. Map of the area studied (From Kartavtsev, 1994).

Первый наиболее важный результат проведенного исследования состоит в том, что в пределах морских бассейнов (и больших заливов) существует генетическая однородность, тогда как между морями и суммарно имеется отчетливая неоднородность (рис. 8.4.2).

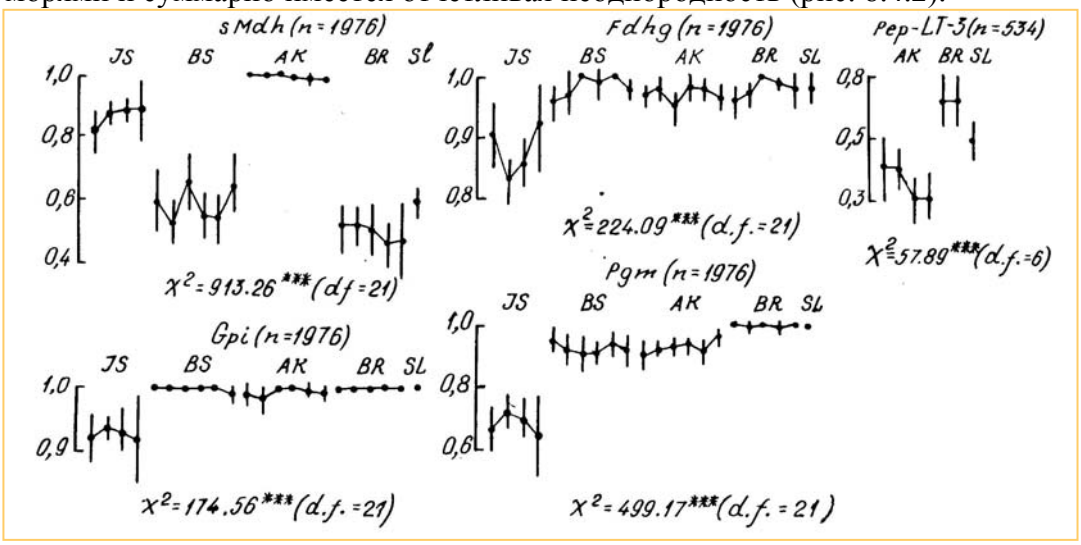


*Рис. 8.4.2. Изменчивость частот аллелей по 6 аллозимным локусам в выборках северной креветки, Pandalus borealis. Бассейны: JS – Японское море, BS – Берингово море, BR – Баренцево море, AK – залив Аляска, Sl – залив Св. Лоуренса. Над ломаными линиями показаны исследованные локусы и размер выборки (n). Неоднородность частот преобладающих аллелей: *** $p < 0.001$.*

*Fig. 8.4.2. Allele frequency variation at 6 allozyme loci in cohorts of pink shrimp, Pandalus borealis. Asterisks show statistically significant heterogeneity: *** $p < 0.001$.*

В соответствии с характером гетерогенности имеется маленькая, или отсутствующая вовсе, генетическая дифференциация в шкале *F'st* в пределах морского бассейна и много большая дифференциация между бассейнами и суммарно (рис. 8.4.3).

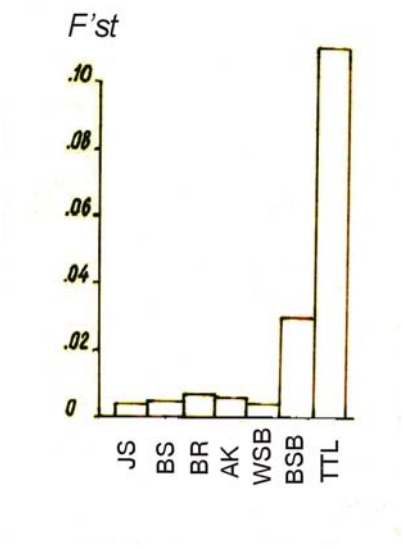


Рис. 8.4.3. График распределения F'st (Nei, 1987) для различных уровней внутривидовой дифференциации северной креветки, P. borealis.

JS – Японское море, BS – Берингово море, BR – Баренцево море, AK – залив Аляска. WSB – средние внутри бассейна, BSB – средние между бассейнами, TTL – суммарное значение для всего ареала.

Fig. 8.4.3. Plot of F'st distribution within and among basins investigated in pink shrimp.

JS, Sea of Japan, BS, Bering Sea, BR, Barents Sea, AK, Gulf of Alaska. WSB, mean within sea basin comparisons, BSB – mean for between sea basins comparisons, TTL, total value for the whole area.

Кластеризация выборок, основанная на сходстве частот аллелей четырех сопоставленных локусов, показывает, что выборки одного и того же бассейна составляют много более близкие группы, чем выборки, принадлежащие к различным морям (рис. 8.4.4).

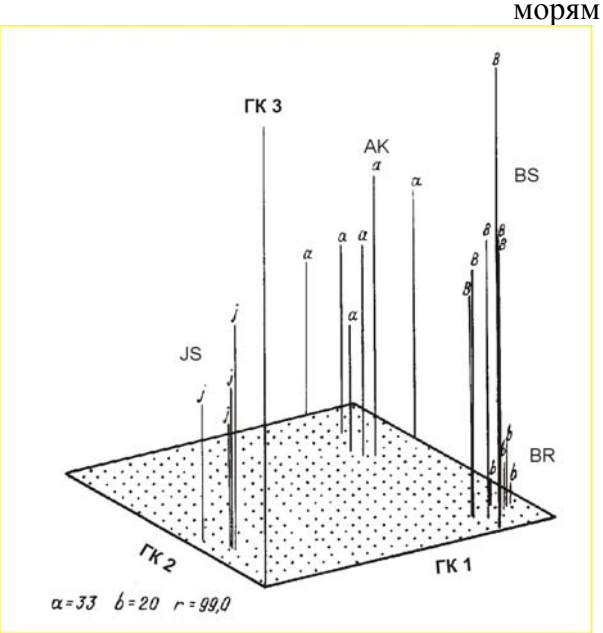


Рис. 8.4.4. Трехмерное распределение значений координат главных компонент (ГК), полученных по частотам аллелей аллозимных локусов северной креветки.

JS – Японское море, BS – Берингово море, BR – Баренцево море, AK – залив Аляска. Внизу даны углы вращения графика.

Fig. 8.4.4. Three-dimension distribution of principal components (PC) scores, took from allele frequencies at allozyme loci.

JS, Sea of Japan, BS, Bering Sea, BR, Barents Sea, AK, Gulf of Alaska.

Все популяции разных морей дискриминируются абсолютно точно (100%), на основе частот аллелей проанализированных локусов, что проиллюстрировано для наименее различающихся северных бассейнов (Табл. 8.4.1).

Таблица 8.4.1. Матрица классификации выборок северной креветки *Pandalus borealis* для трех северных морских бассейнов, полученная по результатам дискриминантного анализа частот аллелей четырех проанализированных аллозимных локусов

Группа	Процент точных классификаций	Число классифицированных выборок		
		Берингово море	Баренцево море	залив Аляска
Берингово море	100	6	0	0
Баренцево море	100	0	5	0
залив Аляска	100	0	0	6
Суммарно	100	6	5	6

В целом генетическая дифференциация у северной креветки хорошо согласуется с данными для других морских ракообразных и является маленькой в сравнении с популяциями пресноводных видов, имеющих ограниченный межпопуляционный поток генов (рис. 8.4.5).

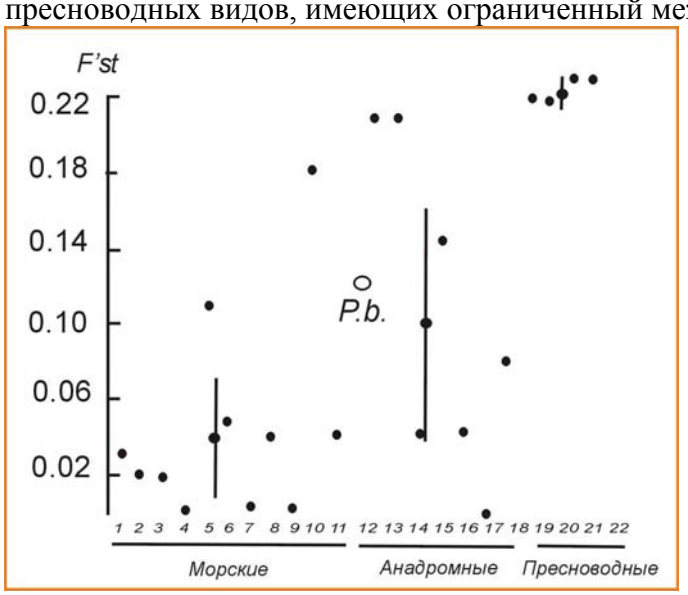


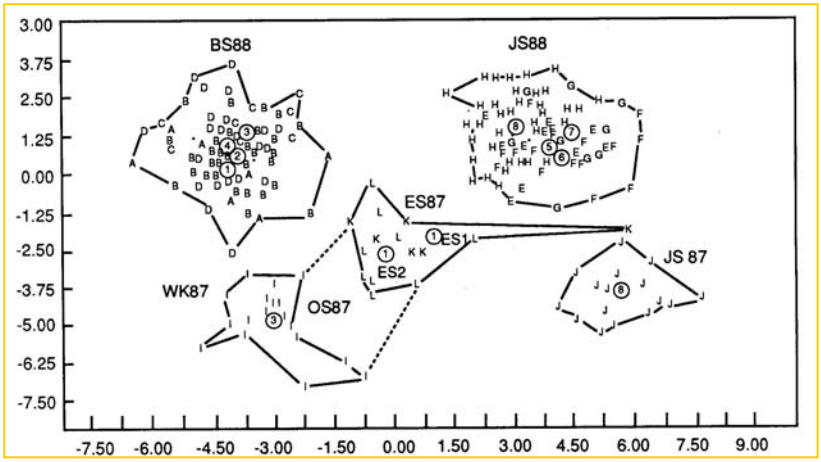
Рис. 8.4.5. График распределения F'$_{st}$ для высших ракообразных, представленных тремя группами: морскими (M), анадромными (A), и пресноводными (F).

Кружок и буквы P.b. отмечают наши данные для северной креветки. Точки с вертикальными линиями обозначают средние значения F'$_{st}$ с доверительными интервалами (95%) для трех групп M, A и F. Литературные данные взяты из Чоу и Фуджио (Chow, Fujio, 1987), но пересчитаны автором в величины F'$_{st}$.

Fig. 8.4.5. Plot of F'$_{st}$ distribution among the higher crustaceans for three groups: marine (M), anadromous (A), and freshwater (F).

The circle and the letters P.b. show the point for the pink shrimp. Dots with vertical lines show mean F'$_{st}$ values with confidence intervals (95%) for the three groups M, A, F. Literature data from Chow and Fujio (1987) were recalculated in the scale F'$_{st}$ by the author (From Kartavtsev, 1994).

Морфометрическое исследование, проведенное на северной креветке (Kartavtsev et al.,



1993), в общих чертах согласуется с генетическими данными, представленными выше (рис. 8.4.6).

Рис. 8.4.6. Распределение значений координат двух канонических переменных (CV), полученное на основе анализа комплекса морфологических признаков и индексов у северной креветки P. borealis.

Буквы показывают распределение значений для отдельных особей. Средние значения для выборок показаны кружками. Выборки: BS –Берингово море, JS – Японское море, OS – Охотское море, ES – восточный Сахалин, WK – западная Камчатка (По Kartavtsev et al., 1993).

Fig. 8.4.6. Female pink shrimp Pandalus borealis distribution on two canonical variables (CV) as exemplified by the for-sample, for-individual discriminative procedure based on the combine morphological traits and indices complex. The signs and abbreviations are the same as before (From Kartavtsev et al., 1992).

8.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 8

1. Оцените стандартное и минимальное генетические расстояния Нея (Nei, 1972) по данным для тихоокеанской мидии (SPECSTAT, File Edspec.dt1), а также рассчитаете Fst и F'st, используя пакет программ SPECSTAT. Рассчитайте величины Fst на основе пакета программ BIOSYS и GENEPOP. Сравните все эти значения. О чем говорят значения Fst (F'st)?
2. Постройте дендрограмму методом НПГ с использованием пакета программ BIOSYS по образцовому файлу из пакета на основе двух расстояний Нея (Dn, Dm). Есть ли между ними различие?
3. Проведите численное моделирование на основе программ POPULUS для иллюстрации влияния изменчивости по трем популяционно-генетическим параметрам - ***Ne, s,*** и ***p***.

ГЛАВА 9. АНАЛИЗ ВНУТРИВИДОВОЙ ГЕНЕТИЧЕСКОЙ ДИФФЕРЕНЦИАЦИИ: ТЕОРИЯ И ПРАКТИКА

ГЛАВНЫЕ ВОПРОСЫ:

9.1. Основные генетические параметры популяционной изменчивости.
9.2. Количественные меры дифференциации.
9.3. Модели популяционной структуры.
9.4. Эмпирические оценки параметров.
9.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 9.

РЕЗЮМЕ

1. Наиболее важными являются четыре популяционно-генетических параметра: ***(1) p* – частота аллеля*, (2) Ne* – генетически эффективная численность популяции*, (3) m* – коэффициент миграции*, (4) s* – коэффициент естественного отбора.**
2. ***Ne,* генетически эффективная численность популяции –** это число особей в идеальной популяции размножающихся особей, которая имеет такие же свойства в смысле дрейфа генов, как и реальная популяция организмов. Обычно выполняется соотношение $Ne < N_b < N$, где N_b и N – численности производителей в нерестовой популяции и общая численность всех особей в популяции.
3. В **подразделенной популяции**, если отбор отсутствует, ключевым параметром является ***Nm***, произведение ***Ne*** и коэффициента миграции, ***m.***
4. Популяционно-генетическая теория достаточно хорошо развита. Однако **эмпирические оценки** даже основных **параметров** имеются в очень ограниченном числе.

Chapter 9. ANALYSIS OF GENETIC DIVERSITY WITHIN A SPECIES: THEORY AND EXPERIENCE

SUMMARY

1. There are four the most important population genetic parameters: (1) ***p***, **allele frequency**, (2) ***Ne***, **effective size of population**, (3) ***m***, **coefficient of migration**, and (4) ***s***, **coefficient of natural selection.**
2. ***Ne,* effective size of population** is a number of individuals in an ideal reproducing population that has the same quality in the gene drift sense as a real bisexual population of organisms. In general $\mathbf{Ne < N_b < N}$ is hold, where N_b and N are the numbers of spanners and the total population number.
3. In a **subdivided population** if no selection occurred the key parameter is ***Nm.***
4. A population genetic theory is well developed. However, there are few **empirical estimates of** population genetic **parameters**.

9.1. ОСНОВНЫЕ ГЕНЕТИЧЕСКИЕ ПАРАМЕТРЫ ПОПУЛЯЦИОННОЙ ИЗМЕНЧИВОСТИ

В предыдущей лекции мы отметили наличие четырех основных параметров, определяющих генетическую дифференциацию популяций в конкретном поколении, то есть, в текущий момент времени. В целом, в совокупности поколений, в ходе эволюционной истории популяции или вида, большую роль могут играть также мутации. Их значение и эволюционная роль были рассмотрены

в Лекции 6. Напомним еще раз, о каких четырех параметрах текущей динамики идет речь: (1) ***p*** – частота аллеля**,** (2) ***Ne*** – генетически эффективная численность популяции, (3) ***m*** – коэффициент миграции, определяющий поток генов и (4) ***s*** – коэффициент естественного отбора. Давайте рассмотрим более подробно, что они собой представляют и какова их теоретическая основа.

Частота аллеля, *p*

Эмпирически при кодоминировании или неполном доминировании величину ***p*** можно оценить прямым подсчетом генотипов, как мы уже обсуждали (Глава 4):

$$p_{A1} = (2D\ A_1A_1 + H\ A_1A_2) / 2n = (D\ A_1A_1 + H/2\ A_1A_2) / n. \qquad (9.1)$$

Где p_{A1} – это частота аллеля A_1, ***D*** – численность в выборке генотипа A_1A_1, ***H*** – численность в выборке генотипа A_1A_2 и ***n*** – общее число особей в выборке. Частоты всех других аллелей могут быть найдены точно таким же способом. Частоту последнего, *k*-ого аллеля, так же как в диаллельном случае частоту второго аллеля, можно не рассчитывать, поскольку она определяется, как разница между единицей и суммой частот всех остальных аллелей: $q_{Ak} = 1 - \Sigma_i^k p_{Ak}$. Для локуса не сцепленного с полом, частоты аллелей и гамет равны. Поэтому в большой популяции двуполых организмов между частотами гамет и частотами генотипов существует биномиальная пропорция, известная как соотношение или закон Харди-Вайнберга:

$(p + q)^2 = p^2 + 2pq + q^2$. Или в общем виде для множественных аллелей соотношение может быть записано как:

$$(\Sigma p_k A_{ij})^2 = \Sigma p^2_k A_iA_i + \Sigma\, 2p_k(1-p_k)\, A_iA_j, \qquad (9.2)$$

где, k = 1,2,3..j – это порядковые номера аллелей. ***Равновесное соотношение между частотами гамет и частотами генотипов одного локуса в большой панмиктической популяции устанавливается уже в первом поколении скрещиваний и сохраняется неограниченно долго, до тех пор, пока не проявится действие фактора(ов), нарушающих условия равновесия, таких как естественный отбор, миграция, сокращение численности и ассортативное скрещивание.***

Для случая трех аллелей соотношение можно записать в виде: $(p + q + r)^2 = p^2+q^2+r^2+2pq+2pr+2qr$ и так далее.

В графическом виде соотношение между аллельными (гаметическими) и генотипическими частотами может быть представлено как парабола внутри треугольника (рис. 9.1.1).

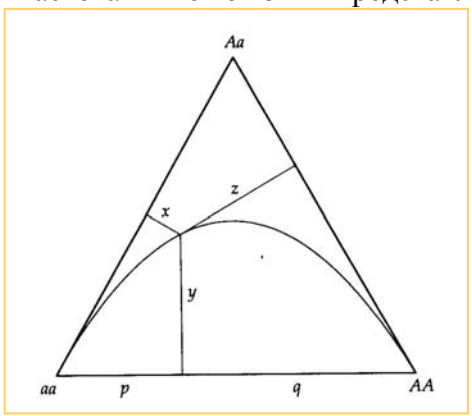


Рис. 9.1.1. Диаграмма Де-Финетти, показывающая частоты трех генотипов по расположению в треугольной системе координат.

*Длины отрезков x, y и z показывают частоты генотипов AA, Aa и aa, соответственно. Углы внизу отмечаают фиксацию аллеля **a** (слева) и аллеля **A** (справа). Вершина треугольника представляет популяцию, состоящую только из гетерозигот и не имеющую гомозигот. Парабола, которая проходит в треугольнике, отмечает набор точек, представляющих частоты генотипов при равновесии Харди-Вайнберга (По Hartl, Clark, 1989).*

Fig. 9.1.1. De Finetti diagram showing the frequencies of the three genotypes by location in a triangular coordinate system.

*The lengths of the line segments x, y, and z indicate the frequencies of genotypes AA, Aa, and aa, respectively. The vertices at the base represent fixation of the **a** allele (left) and the **A** allele (right). The top vertex represents a population with only heterozygotes and no homozygotes. The parabola that runs through the triangle is the set of points representing genotype frequencies in Hardy-Weinberg equilibrium (After Hartl, Clark, 1989).*

Частоты аллелей в принципе можно рассматривать, как вероятности появления того или иного аллеля. Давайте рассмотрим, как оценить такую частоту или вероятность теоретически. Пусть ***p,*** частота аллеля ***A***, а ***q=1-p*** частота для второго аллеля ***a*** из пары, являются случайными переменными, следующими биномиальному распределению.

$$[p(A) + q(a)]^{2N}. \qquad (9.3)$$

Здесь ***N*** – это число особей. В реальности ***N***, здесь и далее, представляет только размножающихся особей и оно равно *Ne* – эффективной численности популяции, как мы увидим позже.

Обозначим $p_A' = p_i - p$ (где p_i – частота аллеля в ***i***-той выборке) как случайное отклонение частоты аллеля от теоретически ожидаемого значения (в генеральной совокупности), отличающегося от отклонений, обусловленных действием систематических факторов (отбор, миграции, мутации). Таким образом, в большой популяции случайно скрещивающихся особей (в панмиктической популяции) дисперсия частоты за одно поколение составит:

$$\sigma^2_{\delta} pA' = [p(1-p)] / 2N. \qquad (9.4)$$

Поскольку выборочная оценка дисперсии равна

$$\sigma^2 pAi = [p_i(1-p_i)] / 2n, \qquad (9.5)$$

то равенство $\sigma^2 \delta pA' = \sigma^2 pAi$ может быть нуль-гипотезой, что тестируемая популяция соответствует большой панмиктической популяции с равновесием Х-В. В качестве критерия этого равенства дисперсий в случае многих выборок можно использовать хи-квадрат тест (X^2) неоднородности частот аллелей (Workman, Niswander, 1970) или другие его варианты. В случае природных популяций, большинство из которых являются подразделенными популяциями, как мы согласились в Лекции 8, оценка неоднородности должна быть более сложной, поскольку более сложна в этом случае дисперсия (Ли, 1978).

$$\sigma^2_{\delta}'' = p(1-p) F_{st}^{(t)} + [(p(1-p)) / 2N](1 - F''_{st}). \qquad (9.6)$$

В 9.6 $F_{st}^{(t)}$ и F''_{st} представляют собой межсубпопуляционную и межгрупповую стандартизированные дисперсии. Мы еще вернемся к этому выражение в Разделе 9.3, когда будем рассматривать модель подразделенной популяции.

Генетически эффективная численность популяции, *Ne*

Ne – это число размножающихся особей в некоторой идеальной популяции, которая имеет такие же свойства в смысле дрейфа генов, как реальная популяция вида. Представление о ***Ne***, как репродуктивной части популяции, вошло в популяционную генетику с работы Райта (Wright, 1931). В настоящее время имеются, по меньшей мере, шесть различных определений ***Ne*** (Kimura, Ohta, 1971; Nei, 1987; Алтухов, 1989).

Когда поколения дискретны (не перекрываются), а популяция является панмиктической, то определить величину ***Ne*** не трудно, если есть информация о числе самцов (m) и самок (f) (Wright, 1931). Дискретные поколения не так уж часто встречаются в животном мире. Тем не менее, предположение о дискретности является полезным и приемлемым, как для предварительной оценки ***Ne***, так и для пояснения сути самого понятия ***Ne***. В данном курсе лекций ***Ne*** для перекрывающихся поколений не рассматривается. Соответствующая литература и содержательное объяснение можно найти в обзоре Лэнде, Бэрроуклаф (1989). Разберем некоторые определения ***Ne***.

1. Ne при варьирующем соотношении полов. В цитированной выше статье Райта ***Ne*** определяется как:

$$Ne = (4N_m N_f) / (N_m + N_f), \qquad (9.7)$$

При неизменной численности родителей N, ***Ne*** будет максимальным, когда $N_m = N_f$. Если соотношение полов отклоняется существенно от соотношения 1:1, то ***Ne*** зависит главным образом от малочисленного пола.

2. Ne при варьирующем числе потомков. Пусть ***k*** – это число потомков особи за всю ее жизнь. Тогда ***Ne*** может зависеть от вклада самцов и самок в следующее поколение, определяемое

их плодовитостью. Запишем, соответственно, N_m, N_f, как средние численности для потомств особей каждого пола, с дисперсией, равной V_k, для каждого из k потомков (Ланде, Бэрроуклаф, 1989):

$$Ne_m = (N_m \bar{k}_m - 1) / [\bar{k}_m + (V_{km} / \bar{k}_m) - 1],$$
$$Ne_f = (N_f \bar{k}_f - 1) / [\bar{k}_f + (V_{kf} / \bar{k}_f) - 1].$$

Для стабильной популяции ($k = 2$)

$$Ne = (4N - 2) / (V_k + 2). \qquad (9.8)$$

Когда число потомков следует закону Пуассона $(k = V_k = 2)$, Ne приблизительно равняется репродуктивной части популяции (N_b). Обычно в природных популяциях $V_k > k$. Соответственно, в них $Ne < N_b$ и в целом можно записать неравенство $Ne < N_b < N$, где N – это общая численность популяции, включая неполовозрелых особей.

3. Ne при флуктуирующей численности популяции. Когда размер популяции флуктуирует, то эффективная численность в различных поколениях, *Ne (1), Ne (2),. . ., Ne (t),* может быть получена как гармоническая средняя для данного временного ряда по формуле

$$Ne = t / \sum[1 / Ne (i)]. \qquad (9.9)$$

Такая оценка через гармонические средние, однако, является состоятельной только для коротких рядов, в течение которых вариация средних для частот аллелей остается незначительной (Wright, 1968: p. 214; Crow, Kimura, 1970: p. 360). Поэтому представленная оценка Ne через выражение (9.9) не может использоваться в долговременных программах регулирования эффективной численности.

4. Дисперсионная Ne. Это определение Ne введено Кроу (Crow, 1954, Kimura, Crow, 1963). В идеальной популяции, которая состоит из N_b скрещивающихся особей, дисперсия частоты аллеля V_p, обусловленная случайностью выборки гамет от одного поколения к другому, задается выражением 9.4. Поэтому естественно определить Ne как

$$Ne = p (1 - p) / 2V \delta p. \qquad (9.10)$$

Имеется также определение Ne в терминах инбридинга. Мы его опустим по той причине, что не вводили здесь это понятие. Некоторые практические следствия из оценок будут рассмотрены в Разделе 9.4. Сейчас, завершая знакомство с эффективной численностью, полезно представить в наглядном виде соотношение между Ne, N_b и N (рис. 9.1.2).

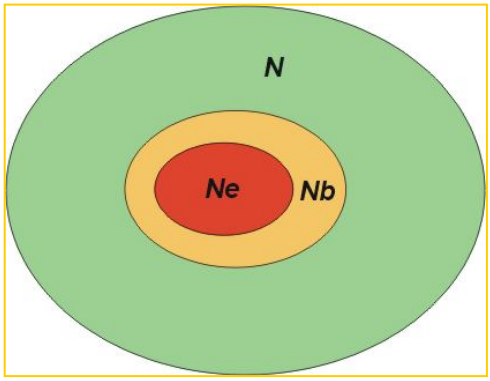


Рис. 9.1.2. Различия в понимании численности или размера популяции.

N – общая численность популяции: половозрелые и не половозрелые особи, N_b – половозрелые размножающиеся особи популяции, Ne – генетически эффективная численность или N_b, уменьшенное за счет не равного соотношения полов, дифференциальной плодовитости и колебаний численности поколений.

*Fig. 9.1.2. Differences among population sizes. N is total number of individuals in a population: mature and not mature, N_b is mature part of a population: **spawning** mates, Ne is genetically effective size or a part of spanning group. N_b affected by sex ratio, fecundity variance, and inter-generation abundance.*

О том, что Ne – это генетическая абстракция, как сформулировано в определении, данном в начале подраздела, не стоит забывать. В частности, исходя из выражения 9.7, легко подсчитать, что при равенстве полов Ne может достигать $4N$, что, конечно, не достижимо на практике.

То, как ***Ne*** может влиять на изменчивость частот аллелей в панмиктической популяции различного размера иллюстрирует Рисунок 9.1.3. При больших величинах ***Ne*** частоты аллелей от поколения к поколению изменяются очень мало, тогда как при малых ***Ne*** изменения велики.

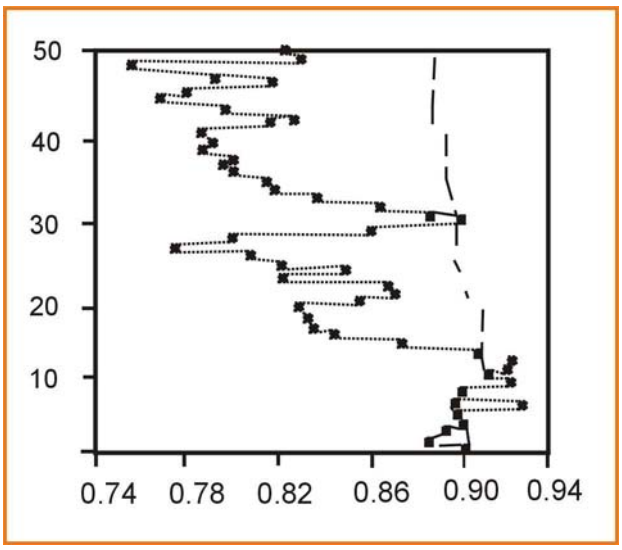


*Рис. 9.1.3. Влияние **Ne** на изменение частоты гена (q – ось X) в серии поколений (**T** – ось Y).*

Точками и сплошной линией показана изменчивость в популяции с Ne=118, а пунктирной линией с Ne=33182. Исходная частота гена в обоих случаях q = 0.9 (По Картавцев, 1995).

*Fig. 9.1.3. **Ne** influence on gene frequency change (**q, X-axis**) through generations (**T, Y-axis**). Changes in smaller population Ne=118 shown with dots and solid line, in larger Ne=33182 changes shown with broken line. Initial allele frequency in both cases is q = 0.9 (From Kartavtsev, 1995).*

Поток генов и коэффициент миграции, *m*

Новые аллели могут появиться в популяции посредством мутаций, но они также могут проникнуть в популяцию в качестве гамет, которые привнесут иммигранты из других популяций. Этот процесс проникновения чужих или «иностранных» генов может вызывать обмен генов, и если этот процесс постоянен, то мы будем называть его потоком генов. *Эффективный размер субпопуляций часто настолько мал, что повторные мутации и их закрепление в популяциях происходят пренебрежимо редко, даже на значительных отрезках времени, не оказывая никакого влияния на дифференциацию популяций в текущем поколении* (именно поэтому мы не рассматриваем роль мутаций в определении популяционной структуры). С другой стороны, иммиграция особей из других демов, сопровождаемая перекрестным скрещиванием, должна быть почти всегда очень важным фактором их текущей дифференциации, распространяясь на последующие поколения и становясь их историей. Пусть ***m*** является фактической долей популяции, которая привнесена со стороны, а *Q* является наблюдаемой частотой данного гена у иммигрантов, в отличие от ***q*** для этого же гена внутри аборигенной популяции (Wright, 1931). Частота гена ***a*** в исходном (нулевом) поколении равна q_0, а частота в первом поколении - q_1. Ее можно определить из соотношения: $q_1 = q_0 (1-m) + Qm = q_0 - q_0m + Qm = q_0 - m (q_0 - Q)$. Изменение частоты аллеля за одно поколение составит: $\Delta q = q_1 - q_0$. Заменяя q_1 в данном выражении, соответствующей величиной из предыдущего уравнения имеем:

$$\Delta q = q_1 - q_0 = q_0 - m (q_0 - Q) - q_0 = - m (q_0 - Q). \qquad (9.11)$$

Здесь ***m*** – это коэффициент миграции или фактическая доля иммигрантов, а Δq – изменение частоты гена за одно поколение.

Эффект миграции можно проиллюстрировать различными путями. Выше мы приняли, что миграция происходит только в одном направлении, например из большой континентальной популяции в маленькую частично изолированную островную популяцию. Предполагается также, что в обеих популяциях выполняется равновесие Х-В, хотя миграция действует. Чтобы лучше понять эффект потока генов, разберем численный пример (рис. 9.1.4).

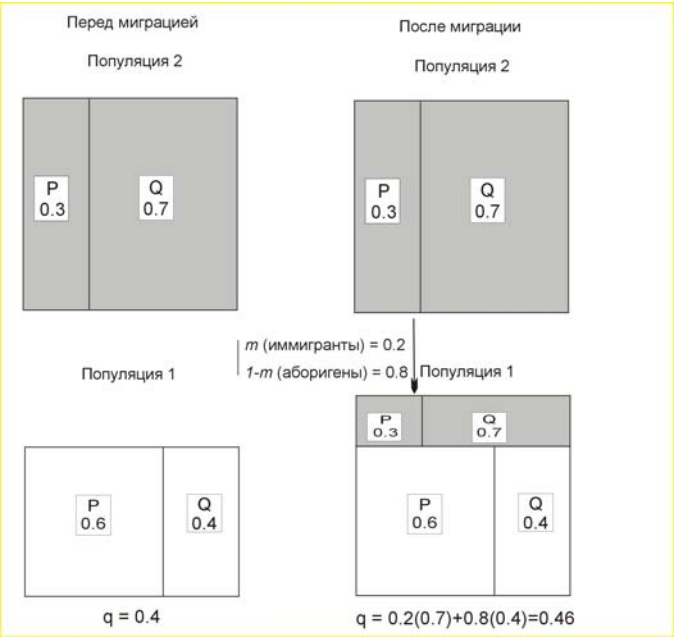


Рис. 9.1.4. Влияние миграции на изменение частоты гена (По Меттлер, Грегг, 1972).
Fig. 9.1.4. Migration influence on gene frequency change (From Mettler, Gregg, 1969).

Миграция влияет не только на интегрированность генофонда, но также и на эффективный размер популяций (Глава 8; рис. 8.3.2). Данные на рис. 8.3.2 соответствуют высокому потоку генов у морских беспозвоночных, что прямо связано с Ne (см. Раздел 9.3, формула 9.22). Это легко понять даже на интуитивном уровне. Так, если три популяции размером Ne_1, Ne_2 и Ne_3 интенсивно обмениваются генами, $m \approx 1$, то это значит, что их численность практически суммируется: $Ne = Ne_1 + Ne_2 + Ne_3$ и, естественно, много больше, чем каждая популяция по отдельности, если миграция отсутствует.

Коэффициент естественного отбора, *s*

Относительная приспособленность, скажем по жизнеспособности генотипов (W), может быть оценена по соотношению генотипов AA, Aa и aa среди потомков в скрещиваниях гетерозиготных родителей (Aa x Aa) исходя из выражения:

$$2n_{AA}/n_{Aa},\ 1,\ 2n_{aa}/n_{Aa}, \qquad (9.12)$$

где n – это число генотипов.

Если $W_{Aa} = 1$, тогда

$$W_{AA} = 1 - s_1, \text{ и } W_{aa} = 1 - s_2, \qquad (9.13)$$

где s_1 является коэффициентом отбора против AA, а s_2 – это коэффициент отбора против aa. После одного цикла отбора средние приспособленности (по данному локусу!) в популяции становятся (Табл. 9.1.1):

$$\underline{W} = W_0p^2 + W_1 2pq + W_2q^2. \qquad (9.14)$$

Таблица 9.1.1. Результаты одного цикла отбора (По Меттлер, Грегг, 1972)
Table 9.1.1. Results of one-circle of selection (From Mettler, Gregg, 1969)

Генотип	AA	Aa	aa	Суммарно
Частота до отбора (F_0)	p^2	2pq	q^2	1
Приспособленность	W_0	W_1	W_2	-
Пропорциональный вклад	W_0p^2	$W_1 2pq$	W_2q^2	$\underline{W}$
Частота после отбора (F_1)	$\frac{W_0p^2}{\underline{W}}$	$\frac{W_1 2pq}{\underline{W}}$	$\frac{W_2q^2}{\underline{W}}$	1

Другими словами, в нашем численном примере в результате отбора численности обеих гомозигот уменьшаются, а относительная частота гетерозигот увеличивается (Табл. 9.1.5). Эти изменения достигнут стабильного или стационарного состояния, когда частоты аллелей достигнут равновесных значений (p^ или q^), которые сохраняются долго, до той поры, пока не изменится W. В пакете POPULUS по умолчанию для t = 200 поколений, q^a = 0.3 (Populus>Evolutionary Simulations>Selection>Autosomal Selection: Defaults).

9.2. КОЛИЧЕСТВЕННАЯ МЕРА ДИФФЕРЕНЦИАЦИИ

Эффект Валунда

Для структурированной популяции необходима какая-то количественная мера дифференциации. В качестве основы для сравнения мы рассмотрим большую случайно скрещивающуюся популяцию и возьмем локус, по которому не имеется отбора. Если локус изменчив по двум аллелям, *A* и *a* с соответствующими частотами *p* и *q,* то ожидаемые пропорции генотипов *AA, Aa* и *aa* будут следовать распределению Харди-Вайнберга: p^2, $2pq$ и q^2. Если же популяция подразделена и частоты аллелей варьируют между субпопуляциями, то тогда, как показал Валунд (Wahlund, 1928), доли гомозигот в объединенной популяции будут больше, чем предсказывается соотношением Х-В (для гетерозигот, соответственно, справедливо обратное).

Пусть *p* это частота А в субпопуляции *i* ($i = 1, ..., m$), и положим, что в каждой субпопуляции генотипы находятся в равновесии Х-В, но отличаются, следуя локальным частотам аллелей. Частота *AA* в субпопуляции *i* тогда равна p^2_i и, если субпопуляции имеют равный размер, то пропорция *AA* во всей популяции составит:

$$(\sum p^2_i) / m = \bar{p}^2 - V_p. \qquad (9.15)$$

Здесь $\bar{p} = \sum p_i / m$ – это средняя частота аллеля *A* во всей популяции, а V_p – дисперсия p_i между субпопуляциями. Таким образом, V_p равняется избытку гомозигот и, соответственно, дефициту гетерозигот по отношению к пропорциям Х-В. Проиллюстрируем это заключение численным примером (рис. 9.2.1).

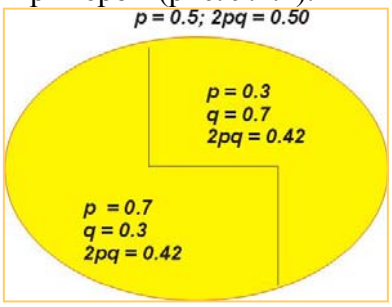


Рис. 9.2.1. Численный пример, иллюстрирующий эффект Валунда.

В подразделенной популяции имеется избыток гомозигот и дефицит гетерозигот по сравнению с ожиданием Х-В в не подразделенной панмиктической популяции. В панмиктической популяции при равенстве частот двух аллелей 2pq = 0.50, однако, в подразделенной популяции 2pq = 0.42, хотя суммарные частоты аллелей также равны 0.5.

Fig. 9.2.1. Numerical example of Wahlund effect.

In subdivided population there is an excess of homozygotes and deficit of heterozygotes in comparison with H-W expectations in a single panmictic population. In panmictic population 2pq = 0.50, however in subdivided population 2pq = 0.42.

Райт (Wright, 1921) ввел параметр F, названный индексом фиксации, для того, чтобы охарактеризовать изменчивость по диаллельному локусу. Для нашего случая частоты генотипов *AA, Aa* и *aa* в подразделенной популяции могут быть записаны, соответственно, как $p^2 + Fpq$, $2pq(1 - F)$ и $q^2 + Fpq$; сравнивая с (9.15) мы видим, что $F = V_p / \bar{p}\,\bar{q}$. Райт (Wright, 1943; 1951; 1965) развил свою концепцию и предложил три F –статистики: Fis, Fit и Fst, посредством которых можно оценить разные варианты отклонения от равновесия Х-В в объединенной популяции при наличии ее подразделенности на субпопуляции. Для оценки дифференциации между субпопуляциями применима статистика Fst, которая в интерпретации Райта является коэффициентом корреляции между гаметами, в выборках из одной популяции по отношению к объединенной популяции. Для конкретного текущего поколения это определение ведет к выражению $Fst = V_p / \bar{p}\,\bar{q}$, в котором *Fst* – это стандартизированная дисперсия частот аллелей. Отметим, однако, что $Fst = V_p / \bar{p}\,\bar{q}$ справедливо вне зависимости, выполняется ли равновесие Х-В или нет. Удобная формула для Fst может быть получена, если ввести $Hs = Hi/m$, где $Hi = 1 - (p_i^2 + q_i^2)$ – это ожидаемая из соотношения Х-В гетерозиготность в субпопуляции *i*, а $Ht = 1 -$

$(\bar{p}^2 + \bar{q}^2)$. Простые выкладки ведут к выражению (Nei, 1973):

$$Fst = V_p / \bar{p}\,\bar{q} = (Ht-Hs)/Ht = 1 - (Hs/Ht). \quad (9.16)$$

Анализ генетической дифференциации

Ней (Nei, 1973; 1977) расширил использование *Fst*, как меры генетической дифференциации, введя статистику ***Gst***, которая применима для множественных аллелей и нескольких локусов. Пусть p_{ik} для данного локуса является частотой аллеля ***k*** в субпопуляции $i = 1,2..m$. Обозначим ожидаемые гетерозиготности h_k, Hs и Ht как: $h_k = 1 - \sum^k_{i=1} p_i^2$, $Hs = \sum^m_{i=1} h_k / m$, $Ht = 1 - \sum^m_{i=1} \bar{p}_k^2$. Далее примем также, что наблюдаемая гетерозиготность равна $Ho=1/n\ h_{фак}$, где $h_{фак}$ – фактическая численность гетерозигот в выборке, а n – число особей в выборке. Теперь можно ввести F-статистику для случая множественных аллелей и многих локусов (Nei, 1973, 1977, 1987). Для одного локуса значения *F*–статистик равны:

$$F_{is} = (H_s - H_o) / H_s\ ,\ F_{it} = (H_t - H_o) / H_t,\ F_{st} = (H_t - H_s) / H_t\ . \quad (9.17)$$

Для нескольких локусов F-статистики рассчитываются усреднением полокусных значений со взвешиванием, процедура которого описана Неем (Nei, 1977b) и представлена также на русском языке (Картавцев, Заславская, 1983). Для того чтобы различать *Fst*-статистику Райта с последним вариантом *Fst*, было введено обозначение ***Gst*** (Nei, 1977b). В этом обозначении *Fst=Gst* является мерой межгрупповой генетической дифференциации (diversity) и всегда является положительной, даже если на самом деле нет реальной генетической гетерогенности. Это надо иметь ввиду при изучении реальных популяций, так как ошибки выборочности могут создавать кажущуюся дифференциацию. *Fis* – это индекс, измеряющий отклонения внутригрупповых (субпопуляционных) наблюдаемых гетерозиготностей от ожидаемых при равновесии Х-В (он может быть положительным – при дефиците гетерозигот по отношению к Х-В пропорции или отрицательным – при их избытке). *Fit* – это индекс, измеряющий отклонение суммарных ожидаемых гетерозиготностей в популяции от фактических гетерозиготностей. Он может быть, аналогично Fis, положительным или отрицательным, но всегда больше *Fis*.

Иерархическая структура

Можно показать, что $Ht = Hs + Dst$ (Nei, 1973). Здесь $Dst = D_{kl} / s^2$ – это средняя межгрупповая дифференциация по *l*–тому локусу; где D_{kl} – среднее минимальное генетическое расстояние среди *s* групп или субпопуляций. Используя эти обозначения можно ввести другую запись для Gst. Относительную величину генетической дифференциации между субпопуляциями можно измерить как:

$$Gst = Dst / Ht. \quad (9.18)$$

Величина Gst варьирует от 0 до 1 и называется коэффициентом генной дифференциации. Выражение (9.18) можно легко расширить до случая, когда каждая субпопуляция далее подразделена на несколько колоний. В этом случае *Hs* может быть разложена на генную дифференциацию внутри и между колониями *(Hc* и *Dcs,* соответственно). Поэтому,

$$Ht = Hc + Dcs + Dst. \quad (9.19)$$

Такого типа анализ может быть продолжен для любого числа иерархических подразделений. Относительная степень генной дифференциации, отнесенная к колониям внутри субпопуляций, может быть измерена как $Gcs(t) = Dcs/Ht$. Можно показать, что $(1 - Gcs)(1 - Gst)Ht = Hc$, где $Gcs = Dcs/Hs$. Детали этого анализа и обсуждение близких вопросов можно найти в литературе (Chakraborty et al., 1982; Chakraborty, Leimar, 1987).

9.3. ТИПЫ ПОПУЛЯЦИОННОЙ СТРУКТУРЫ

Модель панмиктической популяции

Как говорилось выше и в Лекции 8, простейший тип популяционной структуры описывает модель панмиктической популяции. Параметрами этой модели являются ***p*** и ***Ne***. В начале любого

популяционного исследования именно эту модель, с отсутствием какой либо генетической дифференциации (гетерогенности) и равновесием Х-В, необходимо тестировать в первую очередь. В том виде, как она сформулирована для случая множественных аллелей (9.2), эта модель такова: $(\Sigma p_k A_{ij})^2 = \Sigma p^2_k A_i A_i + \Sigma 2p_k(1-p_k) A_i A_j$, где A_{ij} обозначает аллель i или j, p_k его частоту, $k = 1,2,3..j$ – порядковые номера аллелей и $\Sigma p_k = 1$.

Модели подразделенной популяции

Прежде, чем рассмотреть детали давайте рассмотрим схему модели (рис. 9.3.1).

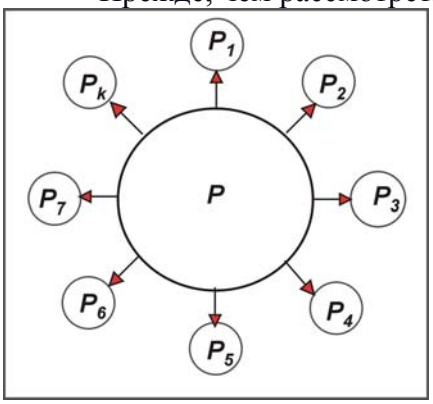


Рис. 9.3.1. Один из вариантов островной модели популяционной структуры.

P – это средняя частота аллеля в группе субпопуляций ***K (K → ∞)*** *(p_1 до p_k), которые составляют «острова архипелага» и частоты аллелей в которых равны частоте аллеля в «континентальной» популяции. Каждая субпопуляция получает гены с «континента» со скоростью, равной* ***m*** (стрелки).

Fig. 9.3.1. One of variants of the Island Model of population structure.

P is mean frequency of alleles among ***K (K → ∞)*** *demes (p_1 to p_k) that comprise "islands of archipelago" and which equals to allele frequency in a "main land" population. Each deme getting genes from "main land" with the rate* ***m*** (arrows).

Островная модель

Имеются два варианта этой модели. Схема одного из них дана выше. Во втором варианте «острова архипелага» равны в размере и обмениваются друг с другом m иммигрантами с равной вероятностью, а «континент» не рассматривается. Генофонд мигрантов формируется совместным вкладом из всех субпопуляций. Расстояния между субпопуляциями не влияют на поток генов. По генетическим последствиям эти два варианта подразделенной популяции идентичны.

Для простоты, мы не будем принимать в расчет возможное действие естественного отбора. То есть, мы рассмотрим действие только параметров p, Ne и m. В этом случае только дрейф генов будет определять дифференциацию субпопуляций, а равновесное состояние частот аллелей будет достигаться при влиянии потока генов (миграции). Дифференциация задается выражением (9.20), которое сходно с (9.6), но редуцировано, поскольку здесь мы рассматриваем субпопуляции равного размера и достаточно большие, чтобы поддерживать равновесие Х-В.

$$\sigma^2_{\delta''} = p\,(1 - p)\,F_{st}(t), \qquad \textbf{(9.20)}$$

где $F_{st}(t) = (1 - m)^2[1/2N+(1-1/2N)\,F_{st}(t-1)]$. **(9.21)**

Если выразить в (9.21) F через m и наоборот, то можно получить хорошо известные равенства (Ли, 1978, p. 474):

$$(1 - m)^2 = 2N\,F_{st} / ((2N - 1)\,F_{st} + 1), \qquad \textbf{(9.22)}$$

$$F_{st} = (1 - m)^2 / (2N - (2N - 1)(1 - m)^2). \qquad \textbf{(9.23)}$$

Эти соотношения свидетельствуют об отрицательной связи между дрейфом и миграцией в межгрупповой дифференциации, которая определяется величиной Ne и значением коэффициента миграции. При небольшой интенсивности миграции ($m << 1$) из (9.23) и (9.20) можно вывести более простое выражение для генетической дисперсии или дифференциации (Wright, 1931; Колмогоров, 1935):

$$\sigma^2_{\delta'} = p\,(1 - p) / (4Nm + 1). \qquad \textbf{(9.24)}$$

Из последнего соотношения теперь ясно, почему $Fst = 1 / (4Nm + 1)$, как было дано в предыдущей лекции. Здесь важно отметить, что для каждого текущего поколения важным является

не само по себе действие ***m*** или ***Ne,*** но их произведение ***Nm***; то есть, ключевым является совокупный эффект в виде числа мигрантов (причем речь идет о потоке генов, или иначе, о мигрантах, которые оставляют потомство!).

При взаимодействии дрейфа (изменения ***Ne***) и миграции изменение частоты аллеля во времени - это комплексная функция распределения плотности вероятности (Wright, 1938; 1969). Объяснение этого выражения находится вне задач этого курса. Соотношение между ***Nm*** и ***Fst*** в состоянии стационарности, при равновесии дрейфа и миграции, показано графически (рис. 9.3.2).

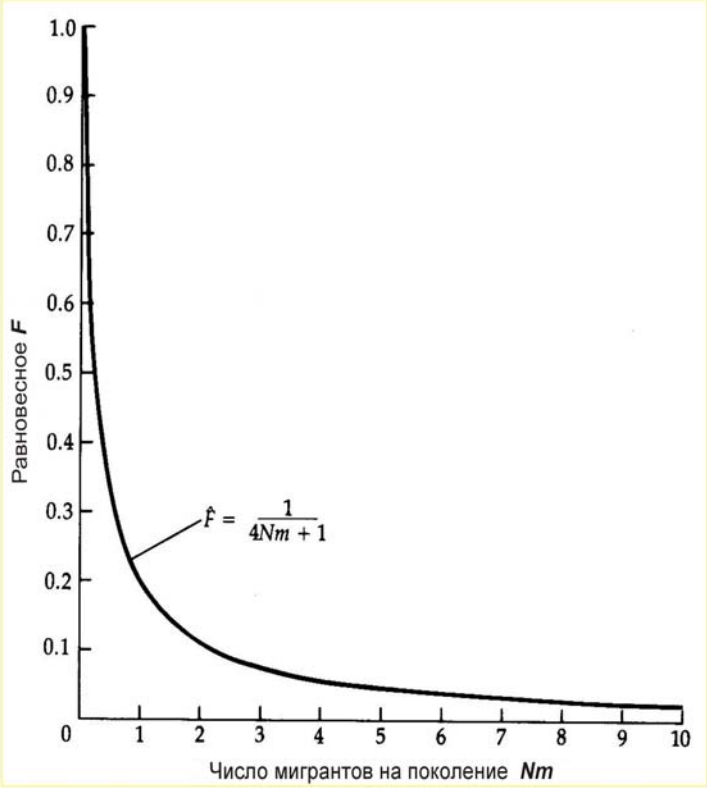


Рис. 9.3.2. Связь между равновесной величиной индекса фиксации (F^) и числом мигрантов на поколение (Nm).

Отметим, что лишь несколько мигрантов необходимо для значительного снижения F^ и, таким образом, снижения генетической дифференциации до весьма низкого уровня: Nm = 0.25 (один мигрант на каждое четвертое поколение), F = 0.50, Nm = 0.5 (один мигрант на каждое второе поколение), F = 0.33, Nm = 1 (один мигрант на поколение), F = 0.20, Nm = 2 (два мигранта в каждом поколении), F = 0.11 (По Hartl, Clark, 1989).

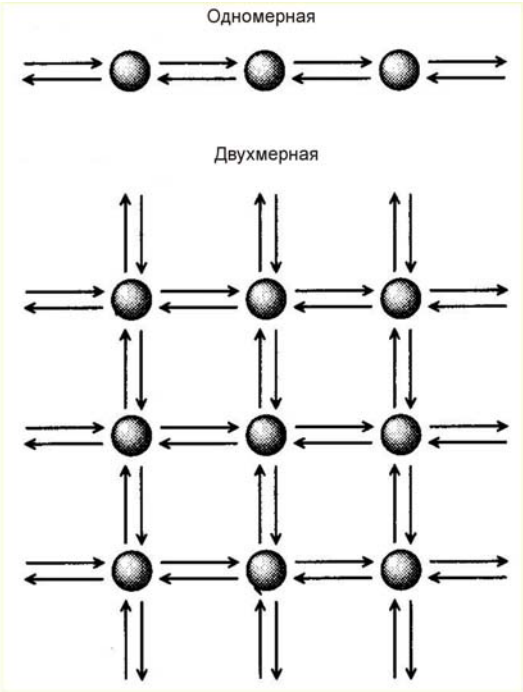
Fig. 9.3.2. Decrease of equilibrium fixation index (F^) against number of migrants per generation (Nm). Note that only a few migrants are necessary to reduce F^, and thus a population differentiation, to a very small level (From Hartl, Clark, 1989).

Nm = 0.25 (one migrant every fourth generation), F = 0.50, Nm = 0.5 (one migrant every second generation), F = 0.33, Nm = 1 (one migrant every generation), F = 0.20, Nm = 2 (two migrants every generation), F = 0.11.

Островная модель предполагает, что если вовлечена группа внутривидовых подразделений, то должно быть увеличение генетической дифференциации от более низких иерархических уровней, к более высоким.

Модель ступенчатой миграции

Одно-, двух- и трехмерные модели ступенчатой миграции были разработаны Кимурой с коллегами (Kimura, 1953; Kimura, Weiss, 1964; Weiss, Kimura, 1965). Мы не будем рассматривать детали этих моделей. Однако некоторое схематическое представление о них мы предлагаем (рис. 9.3.3). По меньшей мере, исходя из одномерной модели, даже интуитивно очевидно, что для этого случая между миграцией и генетическим расстоянием (генетической дифференциацией) существует обратно пропорциональная зависимость.



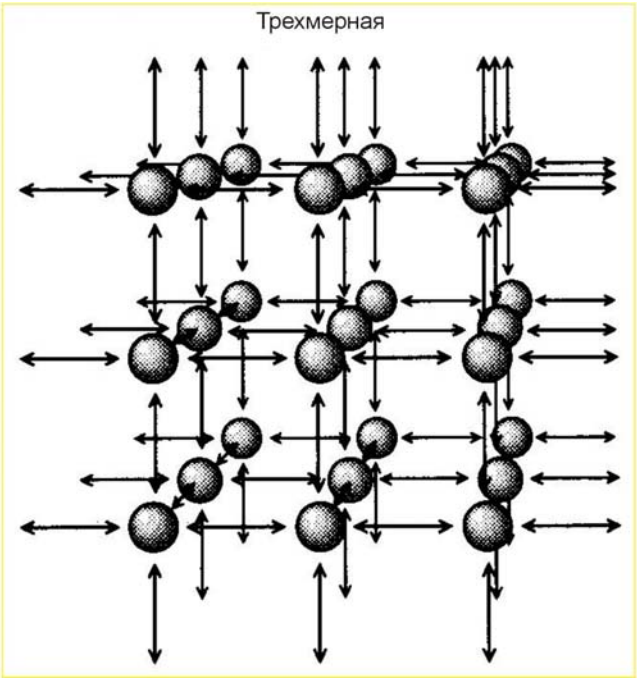
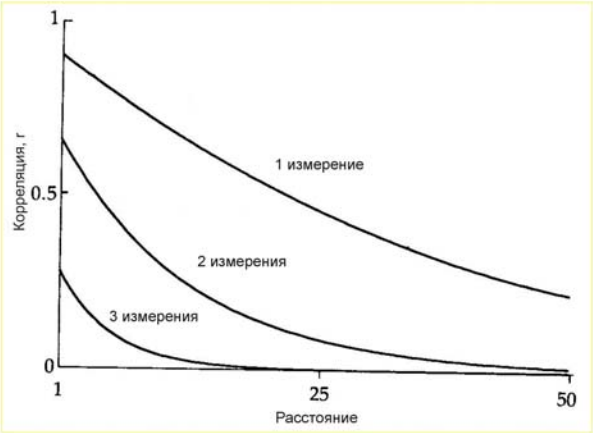


Рис. 9.3.3. Характер миграции при одно-, двух- и трехмерной моделях ступенчатой миграции.

Каждая стрелка представляет темп миграции m/2 в одном измерении, m/4 в двух измерениях и m/6 в каждом направлении в трех измерениях. Краевые эффекты могут быть элиминированы превращением одномерного массива в окружность, двумерного массива – в поверхность сферы, а трехмерного массива – в тор (По Hartl, Clark, 1989).

Fig. 9.3.3. Pattern of a migration in one-, two-, and three-dimensional stepping stone models Each arrow represents a migration rate of m/2 in one dimension, m/4 in two dimensions, and m/6 in each direction in three dimensions. Edge effects can be eliminated by wrapping the one-dimensional array into a circle, the two-dimensional array into the surface of a sphere, and the three-dimensional array into a torus (From Hartl, Clark, 1989).



В теории подразделенной популяции (последний вариант) имеется ясное предсказание взаимосвязи между степенью генетической дифференциации и расстояниями между субпопуляциями или демами (Рис 9.3.4), которое подтверждается эмпирическими исследованиями на индейцах Папаго (рис. 9.3.5). Также, сохраняется верным общее равновесное соотношение для случая нейтральности (модель без отбора):

Fst = 1 / [4Nm + 1].

Рис 9.3.4. Уменьшение корреляции частоты аллеля с расстоянием для одно-, двух- и трехмерных моделей ступенчатой миграции (По Kimura, Weiss, 1964).

Fig 9.3.4. The decline in the correlation of allele frequencies under the one-, two-, and three-dimensional stepping stone model (After Kimura, Weiss, 1964).

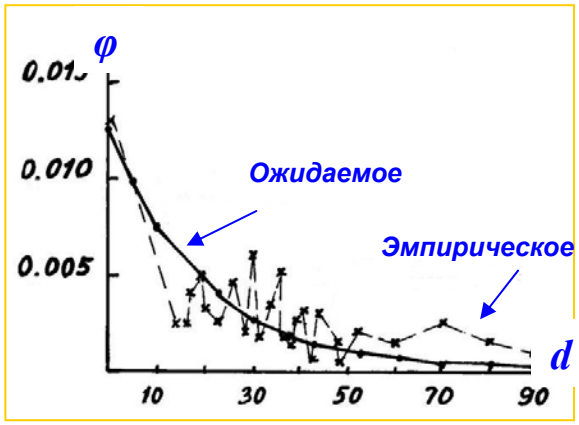


*Рис 9.3.5. Связь между географическим расстоянием (**d, км**), разделяющим племена индейцев Папаго, и генетической близостью между ними (φ) (По Workman et al., 1973 с добавлениями).*

***φ(d) = a*e - bd**, где a – коэффициент локального сходства, b – функция широкомасштабной миграции.*

*Fig 9.3.5. Association between geographic distances and (**d, km**) and genetic relatedness among them (φ) (After Workman et al., 1973 with adds).*

***φ(d) = a*e - bd**, where a - coefficient of local similarity, b - function of wide-scale migration.*

9.4. ЭМПИРИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ

Из необходимых параметров не многие оценены в реальных природных популяциях.

Рассмотрим имеющиеся данные оценивания эффективной численности, ***Ne*** (Табл. 9.4.1), генетической дифференциации, ***Gst*** (рис. 9.4.2), величины миграции, ***m*** (Табл. 9.4.2) и приспособленности, ***W*** (Табл. 9.4.3).

Таблица 9.4.1. Генетически эффективная численность (***Ne***) в природных популяциях животных

Table 9.4.1. Genetically effective size (***Ne***) in natural animal populations

Виды	*Ne*	Источник
Позвоночные		
Oncorhynchus nerka	200	Алтухов и др., 1975
То же	174	Marriott, 1964
O. gorbuscha	695	Картавцев, 1995

Porcelio scaber	19-180	Brererton, 1962
Chondrus bidens	50	Алтухов, Лившиц, 1978
Sceloporus olivaceus	250	Kerster, 1964
Uta stansburiana	14	Tinkle, 1965
Platycercus eximius	31-83	Brererton, 1962
Mus musculus	10	Anderson, 1970
Michigan populations	10-75	Rasmussen, 1964
Arizona populations	30-130	Rasmussen, 1964
Ovis canadensis	98	Geist, 1971
Homo sapiens, Parma Valley, аборигены	214-266	Cavalli-Sforza et al., 1964
Homo sapiens, азиатские монголоиды	45-218	Рычков, 1973, Рычков, Шереметьева, 1976
Homo sapiens, племена индейцев, Ю. Америка	288-14400	Neel, Rothman, 1978
Среднее (среднее по минимальным значениям)	**142**	
Среднее (среднее по максимальным значениям)	**1123**	
Наземные беспозвоночные		
Cepea nemoralis	236-8440	Lamotte, 1951
То же	190-12000	Greenwood, 1975, 1976
Aedes aegipti	500-1000	Tabachnik, Powell, 1978
Drosophila pseudoobscura	500-1000	Dobzhansky, 1972
D. subobscura	400	Begon, 1977
Euphydrias edita	10-3700	Ehrlich, 1965
Dacus olea	4000	Nei, Tajima, 1981
Среднее (среднее по минимальным значениям)	**834**	
Среднее (среднее по максимальным значениям)	**4362**	
Морские беспозвоночные		
Mytilus trossulus	33182	Картавцев, Никифоров, 1993
Pandalus kessleri	730800	Ситников и др., 1998
Среднее	**381991**	

Приведенные в Табл. 9.4.1 данные иллюстрируют: (1) относительно небольшую величину ***Ne*** по сравнению с общей численностью популяций животных, (2) наличие существенного различия величин ***Ne*** между позвоночными и беспозвоночными животными, особенно морскими беспозвоночными (рис. 9.4.1), (3) среди лососевых рыб наибольшее значение ***Ne*** обнаружено для горбуши и значительно меньшее для нерки (Табл. 9.4.1). Последние данные имеют важное значение в связи с дискуссией, представленной в Лекции 8, относительно причин популяционной дифференциации.

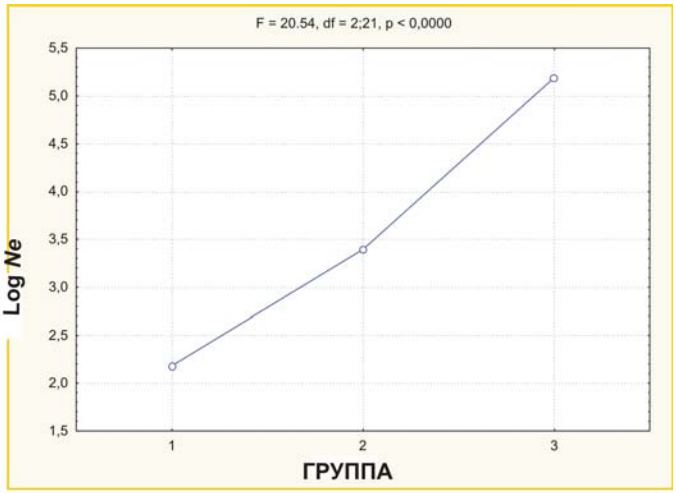


*Рис. 9.4.1. Дисперсионный анализ межгрупповой неоднородности значений **Ne** в таксонах животных: 1 - позвоночные, 2 – наземные беспозвоночные, 3 – морские беспозвоночные.*
*Fig. 9.4.1. Variance analysis among group heterogeneity of **Ne** values in animal taxa. 1 - Vertebrates, 2 - Land invertebrates, 3 - Marine invertebrates.*

В целом генетическая дифференциация, скажем в терминах Gst, среди исследованных видов различна, даже для весьма близких таксономически, как это проиллюстрировано на лососевых рыбах (см. Глава 8, рис. 8.3.9) и как это видно для брюхоногих моллюсков рода *Littorina* (рис. 9.4.2). Для *Littorina* в среднем для планктотрофов и лецитотрофов: Gst = 0.047±0.010 и Gst = 0.154±0.028, соответственно (Ward, 1990; расчет средних и SE автора).

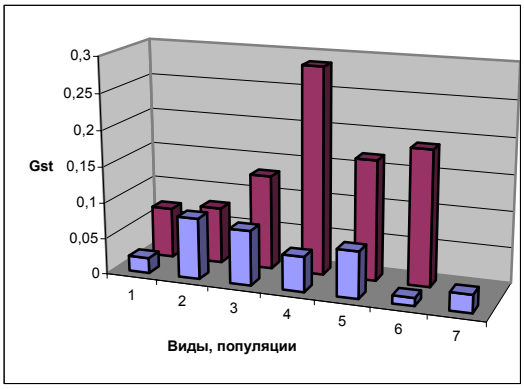


Рис. 9.4.2. Генетическая дифференциация среди локальных популяций различных видов брюхоногих моллюсков рода Littorina.

Передние столбцы – популяции видов, которые имеют планктонную личинку: 1 – L. littorea, 2 – L. scutulata, 3- L. plena, 4 – L. angustior, 5 - L. lineolata, 6 – L. ziczac, 7 – L. angulifera; задние столбцы – популяции видов, у которых отсутствует стадия планктонной личинки в жизненном цикле: 1 - L. saxatilis, 2-4 - L. saxatilis (другие популяции), 5 – L. arcane, 6 – L. nigrolineata (Данные из Ward, 1990).

Fig. 9.4.2. Gene diversity among local populations of various Littorina species.

Front columns are populations of species that have pelagic larvae, back columns are populations of species that have no pelagic stage in a life circle. Pelagic: 1 – L. littorea, 2 – L. scutulata, 3- L. plena, 4 – L. angustior, 5 - L. lineolata, 6 – L. ziczac, 7 – L. angulifera; Non-pelagic: 1 - L. saxatilis, 2-4 - L. saxatilis (other populations), 5 – L. arcane, 6 – L. nigrolineata (Data from Ward, 1990).

Различие средних значений Gst для этих двух групп статистически существенно: $t_d = 3.6$, d.f. = 15, $p < 0.005$. Обнаруженные различия для литторин *(Littorina)* с планктотрофной и лецитотрофной адаптивной стратегией, показывают, что биология видов, которая влияет на интенсивность миграции, может играть принципиальную роль в популяционно-генетической дифференциации.

Рассмотрим некоторые, немногочисленные данные, которые имеются по оценке темпа миграции. Темп миграции значительно варьирует среди изученных лососевых рыб (Табл. 9.4.2).

Таблица 9.4.2. Коэффициенты миграции (***m***) у лососевых рыб
Table 9.4.2. Coefficients of migration (***m***) in salmon

Виды	***m* (%)**	**Источник**
O. nerka	3	Vernon, 1957
То же	<=3	Hartmann, Releigh, 1964
То же	2	Алтухов, 1989

То же	2	Ильин и др., 1983
То же	0,001-0,4	Brannon, 1982
O. gorbuscha	5*	Bams, 1976
То же	9[2)]	Bams, 1976
То же	0-16*	Lane et al., 1990
То же	0,1-11[1)]	Рухлов, Любаева, 1980
O. kisutch	15-27[2)]	Donaldson, Allen, 1958
S. salar	5	Verspoor et al., 1991

Примечание. Коэффициенты ***m,*** отмеченные звездочкой, получены из косвенных источников. 1) Оценки для искусственно воспроизводимых стад. 2) Коэффициент возврата для трансплантированных рыб.

Сводка, представленная в Табл. 9.4.2, иллюстрирует относительно незначительный темп миграции ***m*** в природных популяциях лососей в сравнении с искусственно воспроизводимыми стадами. Эти данные также показывают, что для нерки характерны меньшие коэффициенты миграции ***m***, чем для горбуши, опять же согласующиеся с интерпретацией причин популяционной структурированности у лососевых, данной ранее (Глава 8).

Имеется немного данных по оценке коэффициентов приспособленности в природных популяциях. Оценки приспособленности, полученные для двух видов лососевых, показывают: (1) что имеются относительно небольшие различия ***Wi-***тых коэффициентов приспособленности и (2) что для разных белковых локусов также относительно невелики отличия значений относительных приспособленностей W_i (Табл. 9.4.3).

Таблица 9.4.3. Оценки относительной приспособленности (W_i) генотипов по белковым локусам у двух видов тихоокеанских лососей

Table 9.4.3. Estimates of relative fitness (W_i) of genotypes at allozyme loci in pacific salmon

Вид, локальность и др. источники	**Значение для локуса**			**Литература**
	Гомозигота 1	**Гетерозигота**	**Гомозигота 2**	
		LDH*		
O. nerka, чаши	0.90	1.00	0.86	1
ручьи	0.98	1.00	0.97	1
летняя раса	0.94	1.00	0.91	1
озеро	-	-	-	1
Стадо, суммарно	0.98	1.00	0.97	1
		PGM*		
чаши	0.94	1.00	0.80	1
ручьи	0.91	1.00	0.72	1
осенняя раса	0.92	1.00	0.74	1
озеро	0.99	1.00	0.97	1
стадо, суммарно	0.94	1.00	0.79	1
		MDH*		
O. gorbuscha, родители-личинки	0.63	1.00	-	2
личинки	0.66	-	-	3
-"-		***GPDH****		
	0.58	1.00	-	2
-"-	0.61	-	-	3

-"-	0.71	***6PGD**** 1.00	0.43	2
-"-	0.58	-	-	3
-"-	0.64	***PGM**** 0.78	1.00	2
-"-	0.86	-	-	3

Примечание. ***Wi*** – это значения, принадлежащие наиболее часто встречающимся генотипам. Источники: 1 – Алтухов, 1989, 2 – Салменкова, 1989, 3 – Животовский и др., 1987.

9.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 9

1. Различие Fst и F’st. Меры генетических расстояний, наиболее полезных при анализе внутривидовой структуры: индексы Dm и Rs.
2. Совместный эффект миграции и отбора. Пакет программ “POPULUS” -> Evolutionary Simulations -> Differentiation Models -> Selection, Gene Flow and Clines. Протестируйте две различные опции с различными характеристиками параметров.
3. Отбор в одной популяции. Пакет программ “POPULUS” -> Evolutionary Simulations -> Selection -> Autosomal Selection. Используйте опции по умолчанию. Затем поменяйте параметры.

ГЛАВА 10. ЕСТЕСТВЕННАЯ ГИБРИДИЗАЦИЯ И ИНТРОГРЕССИЯ ГЕНОВ У ВИДОВ В ПРИРОДЕ

ГЛАВНЫЕ ВОПРОСЫ:

10.1. Что такое гибриды и в чем суть гибридизации?
10.2. Методы определения гибридизации.
10.3. Генетическая интерпретация гибридизующих популяций.
10.4. Эмпирические исследования гибридных зон.
10.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 10.

РЕЗЮМЕ

1. Введение методов биохимической и молекулярной генетики в практику **исследования гибридизации** позволило дать точные сведения, как о самом феномене, так и о **частоте встречаемости гибридов** в природе.
2. **Гибридизация** – это динамический процесс. Она включает смешение и реорганизацию генома на индивидуальном уровне, а также смешение и реорганизацию генофондов на популяционном уровне. Гибридизация генерирует неравновесие по сцеплению, когда два ранее равновесных изолированных генофонда объединяются в один. Гибридизация приводит к разрушению сложившихся популяций, как единиц самовоспроизводства, а также хозяйственного воздействия человека.
3. **Генетическое смешение**, создание **гибридных зон** и автоматическое существование долговременного **гаметического неравновесия** могут иметь различные последствия – положительные и отрицательные. Положительными последствиями являются увеличение генетической изменчивости, а также «проверочная» роль **гибридных зон** (tension zone), которые являются «полем битвы» за видовую интегрированность. Наконец, **гибриды**, если иногда оказываются более приспособленными или просто обладающими нужными человеку свойствами, могут быть использованы прямо в аквакультуре или как источники для скрещиваний в селекционных программах. Отрицательные последствия весьма многочисленны: (1) неравновесие, ведущее к снижению общей приспособленности гибридной популяции или вида, (2) интрогрессия вредных (или просто неадаптированных) генотипов в аборигенную популяцию или вид, которая ведет к их деградации и замещению другим видом.

Chapter 10. NATURAL HYBRIDIZATION AND INTROGRESSION IN WILD SPECIES

SUMMARY

1. Introduction of biochemical genetic and molecular genetic methods in **hybridization research** provided a real opportunity to investigate a phenomenon of hybridization itself in nature as well as to estimate **hybrid frequencies** in populations and species.
2. **Hybridization** is a dynamic process. It includes mixing and reorganization of genomes at the individual level, as well as mixing and reorganization of gene pools at the population level. Hybridization is a process, which generates genetic disequilibrium from former genetic equilibrium by mixing and interbreeding two previously isolated gene pools. This process leads to a condition preventing management of fish and shellfish stocks as discrete population units or species.
3. **Genetic mixing** and automatic creation of the long-term **genetic disequilibrium** may have different

impacts, both positive and negative. Positive impact includes the increase of genetic variance and a special role of the **hybrid zone** (tension zone), which served as a "field of battle" for species integrity. **Hybrids**, if sometimes are advantageous, may be useful directly for aquaculture propagation or as a source for breeding programs. Numerous negative impacts are created by (i) disequilibria, decreasing fitness in hybrid deme (taxa), and (ii) introgression of deleterious (or just not adapted) genes into aboriginal populations or species and as a result causing their degradation and substitution by other species (as a rule they are worse, both for human being and ecosystems).

10.1. ЧТО ТАКОЕ ГИБРИДЫ И В ЧЕМ СУТЬ ГИБРИДИЗАЦИИ?

Для начала ответим на вопросы, поставленные в заголовке раздела. ***Гибрид*** *– это генетическая помесь или потомство от скрещивания между генетически различными организмами. Гибридом можно также считать особь со смешанной родословной.* В предельном случае гетерозигота по одному или нескольким локусам – это гибридная особь. Таким образом, скрещивание P_1: A_1A_1 B_2B_2 x A_2A_2 B_3B_3 дает гибрид $F_1 = F_H$: A_1A_2 B_2B_3. ***Гибридизация*** *– это процесс, посредством которого появляются гибриды.* Между тем, мы должны отметить разницу между простым внутрипопуляционным скрещиванием и скрещиванием между различными линиями, популяциями и видами. Истинные гибриды обычно представляются, как потомки более отдаленных скрещиваний. Причем отдаленность скрещивания довольно условное понятие, зависящее от конкретного вида организмов и сложившейся для него, нормальной системы скрещиваний. Кроме F_1, могут возникать другие типы гибридов: F_1x F_1= F_2, F_1x P_1= F_b и т.д. Гибридизация может быть искусственной или естественной природы. В данной лекции мы будем преимущественно иметь дело с естественной гибридизацией. Гибриды не обязательно должны быть промежуточными между родительскими формами, но в зависимости от сложности скрещивания могут быть более близкими к одному из родителей, так что гибридный индекс может быть далек от 0.5 (рис. 10.1.1). Несколько больше деталей мы рассмотрим позже.

Обычно принимается, что средняя приспособленность (***W***) ниже у гибридов в сравнении с родительскими формами (рис. 10.1.1). Это правило не всегда выполняется. Иногда даже обратное соотношение является обычным. Однако эти случаи находятся вне темы данной лекции и относятся к искусственным скрещиваниям между инбредными линиями или определенными типами пород и сортов.

Рис. 10.1.1. Диаграмма, показывающая происхождение гибридов и изменение генофонда, как смеси двух родительских видов (демов) – 1 и 2. Гибридный индекс отмечает степень промежуточности гибридов по отношению к родительским формам. Приспособленность гибридов в природных популяциях обычно снижена.
Fig. 10.1.1. A diagram showing the hybrids' origin and a gene pool change as a mixture between two parental species (demes) - 1 and 2.

Способность и склонность таксономически различных рыб скрещиваться и производить плодовитое потомство, сейчас хорошо документирована. Для американских видов такую сводку сделал Хаббс (Hubbs, 1955). Шварц (Schwartz, 1972; 1981) провел анализ почти четырех тысяч

работ с примерами искусственной и естественной гибридизации у рыб. Считается, что естественная гибридизация более обычна у рыб, чем в других группах позвоночных. Это заключение относится в большой степени и к морским беспозвоночным. На первое место среди причин более частой гибридизации в этих группах надо поставить не столь развитую, как, например, у млекопитающих, систему хромосомного определения пола. Известно, что у рыб пол определяется несколькими причинами, а не половыми хромосомами, которые могут отсутствовать вовсе (Кирпичников, 1979). Несколько особенностей биологии рыб и беспозвоночных могут лежать в основе таких различий: внешнее оплодотворение, слабые поведенческие изолирующие механизмы, неравная численность двух потенциальных родительских видов, конкуренция за ограниченные нерестовые биотопы и, наконец, восприимчивость к вторичному контакту недавно дивергировавших форм (Campton, 1987). В зависимости от локальных условий эти особенности могут сильно варьировать. Естественные и вызванные вмешательством человека изменения во внешней среде часто приводятся в качестве факторов, способствующих гибридизации у рыб (Алтухов и др., 1997). Например, гибридизация относительно обычна у пресноводных рыб из умеренных широт, где геологические и климатические условия, начиная с плейстоцена, резко менялись, меняя пресноводную среду, тогда как морская среда оставалась относительно стабильной. Порожденные цивилизацией изменения в среде в Северной Америке также коррелируют с повышенной гибридизацией, встречающейся между исходно аллопатрическими и естественно симпатрическими парами видов (Hubbs et al., 1953; Nelson, 1966; 1973; Stevenson, Buchanon, 1973). Для лососевых рыб такие примеры были суммированы отдельно (Simon, Nobble, 1968; Алтухов, Салменкова, 1991; Алтухов и др., 1997).

Перед тем как перейти к разделу, посвященному методам анализа, необходимо уточнить несколько терминов. Во-первых, мы будем называть ***потоком генов*** *процесс, маркируемый селективно нейтральными (или почти нейтральными) аллелями и примем, что* ***репродуктивная изоляция*** *между биологическими видами означает отсутствие какого бы то ни было потока генов* (в смысле, что F_1, и особенно F_2 или F_b, неплодовиты или не жизнеспособны; низко плодовиты, слабо жизнеспособны). Эта формулировка подходит под строгую, или ортодоксальную БКВ. Во-вторых, давайте примем, что ***гибридная зона*** *– это географическое пространство, где встречаются гибриды естественного происхождения между предполагаемыми родительскими формами*. Обычно в такой зоне возникает клина. В третьих, ***клина*** *– это постепенное (градиентное) или резкое изменение частот аллелей в прямом или обратном направлении: вид 1 (популяция 1) → гибриды → вид 2 (популяция 2), поддерживаемое балансом между расселением и отбором против гибридов* (модифицировано из Barton, Hewitt, 1985).

10.2. МЕТОДЫ ОПРЕДЕЛЕНИЯ ГИБРИДИЗАЦИИ

Выявление естественной гибридизации между видами из природы, в том числе между видами рыб, часто затруднено, поскольку существуют различного рода неопределенности. Например, два тесно связанных таксона или конспецифичных популяции, могут сосуществовать на ареале без скрещивания друг с другом и, все же, гибридизация может предполагаться из-за перекрывания фенотипических рядов. Или наоборот, гибридизация происходит, но сами гибриды никогда не скрещиваются по разным причинам, включая стерильность. Если гибриды дают обратные скрещивания с одним или обоими родительскими видами, то может возникнуть интрогрессия генов или возникнет гибридная популяция.

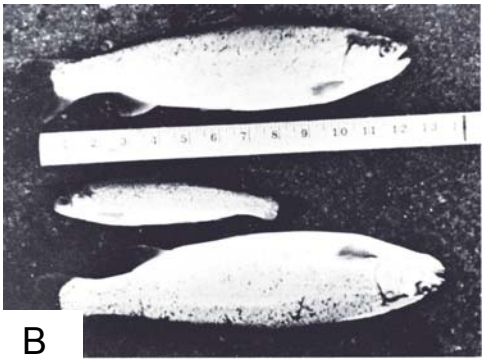
Если факт гибридизации установлен, то возникает потребность различить особей со смешанной родословной, исходящей от двух родительских видов. Однако, как будет отмечено позже, отличить истинных гибридов часто непросто, если гибридизация распространяется дальше

F_1. Ниже мы рассмотрим данные о гибридизации видов и популяций, полученные на основе различных методов.

Методы морфологии

В таких исследованиях основой являются мерные и счетные признаки, значения для которых получены на предполагаемых гибридах, а также на обеих потенциальных родительских формах. Имеются специфические процедуры сравнения морфологических признаков в целом и при изучении гибридизации, в частности (Бейли, 1970; Hubbs, Lagler, 1970). Особи рассматриваются как межвидовые гибриды, если значения их измеренных признаков являются в среднем промежуточными по сравнению с обоими родительскими таксонами.

Из морфологических признаков, используемых для различения таксонов, обычно не все являются промежуточными у гибридов (Simon, Noble, 1968; Ross, Cavender, 1981; Leary et al., 1983). Гибриды не редко проявляют смесь морфологических черт, которые могут быть в одном варианте более близки к одной родительской форме, а другом – к другой. В некоторых случаях гибриды могут быть морфологически промежуточными почти для всех признаков, но некоторые признаки остаются точно такими же, как у одной из родительских форм. Данные для форелей и гольцов помогут понять трудности различения гибридов по морфологии (рис. 10.2.1).



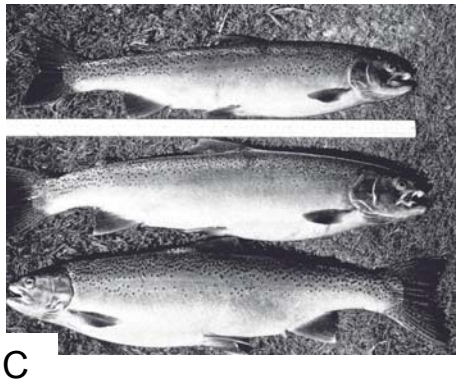


Рис. 10.2.1. Сопоставление форелей и стальноголового лосося.

A. *На снимке показана кумжа Salmo trutta в природе в речном биотопе (По Оммани, 1975).* ***B.*** *Показаны экземпляры радужной форели S. gairdneri из культивируемой линии (нижняя особь); в центре стальноголовый лосось S. gairdneri. Верхняя рыба на снимке - гибрид радужная форель × стальноголовый лосось (Donaldson). Гибрид более быстрорастущий, чем стальноголовый лосось, но обладает предпочтением к миграции в соленую воду, как стальноголовый лосось.* ***C.*** *Экземпляры гибридов радужная форель× стальноголо-вый лосось, вернувшиеся из моря (По Hines, 1976)*

Fig. 10.2.1. The rainbow-steelhead story.

A. *On the color slide is brown trout Salmo trutta in nature in the river.* ***B.*** *The fish at the bottom is a rainbow trout S. gairdneri from the cultivated strain; center, a steelhead S. gairdneri. At the top is a rainbow-steelhead hybrid. It is faster than the steelhead in growth, but with the steel-head's disposition to seek salt water.* ***C.*** *Speci-mens of rainbow-steelhead hybrids that have returned from the sea.*

Специально для морфологических признаков был предложен гибридный индекс, чтобы измерять среднее морфологическое сходство индивидуальной рыбы по отношению к обоим родительским таксонам или популяциям (Hubbs, Kuronuma, 1942; Hubbs et al., 1943).

Индекс (*I*) вычисляется отдельно для каждого признака как $I = 100 * [(u—X)/(Y-X)]$, где *u* – это значение признака у данной особи, а *X* и *Y* – это средние значения признака в популяции (таксоне) X и Y. Индивидуальная рыба со значением признака, равным X или Y будет иметь значение *I*, равное, соответственно, 0 или 100. Индекс со значением 50 обозначает точную промежуточность анализируемого признака. Более распространенным является подход на основе дискриминантного анализа, с использованием для расчетов особей исходных родительских форм, а затем вычисление значений дискриминантной функции для каждой

предполагаемой гибридной особи.

Недостатком гибридного индекса, и в целом одномерных подходов (базирующихся на единичных признаках), является невозможность безошибочно диагностировать по-индивидуально всех особей. Много более обнадеживающие результаты дают многомерные статистические методы, которые сейчас широко используются (Neff, Smith, 1979; Афифи, Эйзен, 1982). Основное их преимущество состоит в том, что по комплексу признаков можно с высокой степенью надежности классифицировать индивидуально каждую особь в таксонах, в том числе выявляя гибридов.

Дискриминантный анализ (ДА) – это один из многомерных статистических методов, который часто используется для анализа морфологических данных и в исследованиях гибридизации. Дискриминантный анализ позволяет вычислить линейные функции измеряемых признаков, таким образом, чтобы две или более групп особей могли быть по-индивидуально отличимы (дискриминированы) и классифицированы одна от другой на основе многомерных векторов значений. Факторный анализ (ФА), в вариантах главных компонент или принципиальных координат, это второй многомерный подход, который часто используется в морфометрических исследованиях. В отличие от ДА, ФА не требует априорной идентификации классифицируемых групп особей. По проекциям отдельных особей (методом наименьших квадратов или другими способами) в их многомерном пространстве значений, можно рассчитать и визуализовать на графике несколько главных компонент, или осей (Афифи, Эйзен, 1982). Набор таких проекций, полученных для особей из районов с наличием гибридизации и вне их, может представлять объективную информацию по морфологической изменчивости и свидетельствовать о наличии или отсутствии гибридов без каких-либо априорных допущений. Такого рода исследование карповых рыб на основе первой главной компоненты, спроектированной на гистограммах, позволило понять распределение форм рыб из нескольких локальностей (Dowling, Moore, 1984; рис. 10.2.2).

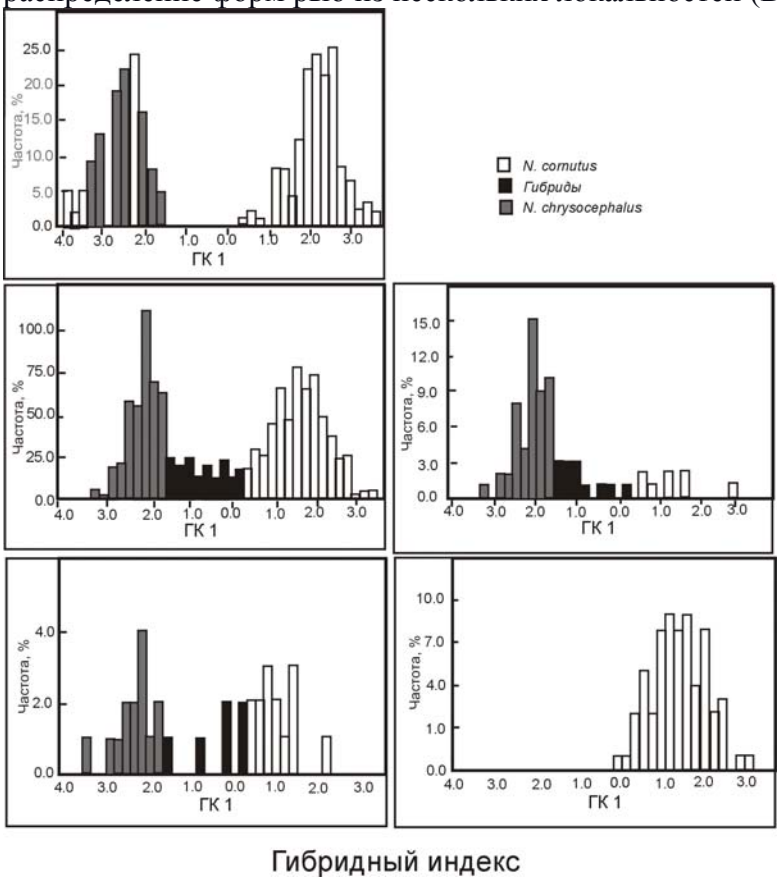


Рис. 10.2.2. Частотное распределение значений первой главной компоненты (ГК 1) для двух карповых рыб, Notropis cornutus и N. chrysocephalus, и их предполагаемых гибридов в пяти географических локальностях (По Dowling, Moore, 1984).
Fig. 10.2.2. Frequency distributions of first principal component scores for two cyprinid fishes, Notropis cornutus and N. chrysocephalus, and their supposed hybrids from each of five geographic locations (From Dowling, Moore, 1984).

Исследование естественной гибридизации на основе морфологических признаков сталкивается со многими проблемами. В общем случае морфологические данные могут дать лишь предварительные сведения о гибридизации в природе или интрогрессии поскольку гибриды априорно

предполагаются промежуточными по фенотипу к родительским формам. Однако это часто не так. Даже при многомерном подходе индивидуальная идентификация затруднена интерградацией признаков, так что гибридов F_1 нельзя отличить от F_2 и бэккроссов. Как отмечено ранее часто гибриды F_1 проявляют мозаику морфологических признаков, которые по отдельности представляют родительские формы. Определение же интрогрессии вообще предельно трудно, поскольку интрогрессирующая форма может оказаться неотличимой от одного из родителей (Greenfield, Greenfield, 1972; Busack, Gall, 1981).

Изложенные выше сведения, как мы видим, требуют использования более прямых генетических методов для установления гибридизации.

Кариотипирование

Использование кариологического анализа широко применяется в исследовании гибридов растений. Вспомним классический пример получения гибрида Карпеченко «Рафанобрассика» между редькой (*Raphanus sativa*) и капустой (*Brassica olearacia*), известный из школьного курса. Оба вида имеют 2n=18. Гибрид F_1 имеет 18 хромосом, по 9 от каждого вида. Мейоз у F_1 нарушен, так как хромосомы несут много не гомологичных участков. Это ведет к стерильности гибрида F_1. Однако, когда хромосомы F_1 были удвоены перед вступлением в стадию мейоза, то была получена зигота растения с 2n=36, которое вполне плодовито, поскольку каждая хромосома имеет гомолог.

Кариологические исследования при анализе гибридов рыб проводились не часто (Setzer, 1970; Greenfield, Greenfield, 1972; Busack et al., 1980; Sola et al., 1981). Большинство попыток дифференциально окрасить хромосомы рыб, и таким образом дифференцировать их индивидуальность по характерной «полосатости» (бэндингу) были не очень удачными. Тем не менее, некоторые успехи были получены для выявления C- и Q-бэндинга и окрашивания ядерного организатора (ЯОР) у лососевых рыб (Phillips, Zajicek, 1982; Delaney, Bloom, 1984; Phillips et al., 1985; Фролов, 2000). Получены даже G-бэндинги и Fish-бэндинги (Tontana et al., 1998), но пока это не вошло в широкую практику. Однако в принципе анализ кариотипов, при усовершенствовании методики, например G- и Fish-окрашивания, может дать объективную и независимую от морфологических данных информацию о гибридном происхождении особей (Greenfield et al., 1973; Gold, 1979; Васильев, 1985) и может быть комплементарным к другим генетическим методам.

Белковые маркеры генов

Обнаружение естественной гибридизации и интрогрессии методами анализа белковой изменчивости является объективным и точным, когда два родительских таксона или две формы имеют различные фиксированные аллели по двум или более локусам. В этом случае все особи двух родительских таксонов будут гомозиготами по различным аллелям, тогда как гибриды F_1 будут гетерозиготами по всем диагностическим локусам (Leary et al., 1983; Whitmore, 1983). Если гибридизация распространяется далее стадии F_1, то гибриды, однако, будут проявлять различный спектр рекомбинантных генотипов, включая два родительских типа (Avise, Van Den Avyle, 1984). Соответственно, особь с составным генотипом, которая идентична по генотипу с одной из родительских форм, может быть гибридным потомком F_1, полученным при самовоспроизводстве или при возвратном скрещивании с одним из родителей. Присутствие рекомбинантных генотипов иными словами может быть признано, как свидетельство плодовитости гибридов и наличия второго поколения гибридов (F_2 или F_b); это выполняется только, если у родительских форм имеется полная фиксация различных аллелей, дающая несомненные данные.

Если формы фиксированы только по аллелям одного локуса, то не всегда возможно отличить гибридов F_1 и F_2; когда доля F_1 велика, то и F_2 встречаются часто. Все особи в такой популяции будут проявлять либо гибридный фенотип F_1, либо один из двух родительских фенотипов. Однако, если частота F_1 невелика, то вероятность встретить F_2 вместо F_1 крайне мала даже по одному локусу (частота F_2 примерно на порядок меньше частоты F_1). Следующий пример иллюстрирует эту мысль. Пусть аллели A_1 и A_2 у родительских форм (P_1 и P_2) фиксированы, а их частоты и частота F_1 в выборке n = 100 из популяции равны: $P_1 = 0.01$, $P_2 = 0.99$, $F_1 = 0.01$. Частоты

гамет у гибрида равны р A_1 = рA_2 = 0.005. Соответственно, частоты гетерозигот в скрещиваниях, при допущении случайности оплодотворения, будут: F_1 x F_1 = 0.005*0.005*2=0.00004, F_1 x P_1 = 0.005 x 0.99 = 0.005 и F_1 x P_2 = 0.005 x 0.01 = 0.00005. То есть, частота самой частой гетерозиготы F_2 как минимум в 10-15 раз меньше частоты F_1. Когда частота гибридов F_1 высока, то присутствие гетерозиготных особей дает только доказательство, что два вида скрещиваются, но размах гибридизации оценить точно нельзя. Многие случаи естественной гибридизации были документированы по единичным экземплярам гетерозиготных особей (Soloman, Child, 1978; Beland et al., 1981). Поэтому важной составляющей исследований должна стать количественная оценка гибридизации. Один из принципиальных вопросов, на которые следует ответить, является ли доля особей со смешанным генотипом существенно выше, чем ожидаемая при свободном скрещивании с одним или обоими таксонами.

Этот вопрос был изучен на природных анадромных популяциях стальноголового лосося (*Salmo gairdneri)* и прибрежных популяциях лосося Кларка *(S. clarki clarki)* (Campton, Utter, 1985). Эти два таксона имеют различные наиболее частые аллели по четырем электрофоретически определяемым локусам, но один из родителей имеет полиморфизм по каждому из четырех локусов. В одном районе было обнаружено неожиданно большое число особей с промежуточными генотипическими комбинациями (Campton, Utter, 1985). Чтобы количественно оценить вероятность гибридизации был предложен индекс *(I_H),* оценивающий относительные вероятности того, что составной генотип для каждой особи возник как следствие случайного скрещивания в пределах каждого из двух таксонов (Campton, Utter, 1985).

$$I_H = 1.0 - [\log10(p_x) / (\log10(p_x) + \log10(p_y))], \qquad (10.1)$$

где $p_x = \Pi_{i=1}^{L} k_i \Pi_{j=1}^{Ai} (X_{ij})^{m_{ij}}$ и $p_y = \Pi_{i=1}^{L} k_i \Pi_{j=1}^{Ai} (Y_{ij})^{m_{ij}}$.

Здесь X_{ij} и Y_{ij} - средние частоты *j*-тых аллелей по *i*-м локусам для вида X и Y, соответственно; m_{ij} – это число аллелей *j*-того типа обнаруженных по *i*-му локусу для оцениваемой особи; A_i – общее число известных аллелей по *i*-му локусу для двух видов вместе; k_i – биномиальный выборочный коэффициент (например, $k_i = 2$ для Аа, $k_i = 1$ для АА или аа), связанный с генотипом особи по *i*-му локусу; и *L* – число диагностических локусов, используемых в различении двух видов. Выражения p_x и p_y являются условными вероятностями того, что составной генотип по всем локусам для особи может возникнуть в результате случайного скрещивания внутри вида X и вида Y, соответственно (предполагается, что выборочные частоты аллелей соответствуют средним для видов и существует равновесие по сцеплению для всех локусов). Этот индекс может принимать значения в диапазоне 0 – 1 и будет близок к одному из этих двух значений, когда особи имеют очень высокую относительную вероятность принадлежности к виду X или Y**,** соответственно.

В исследовании гибридизации между стальноголовым лососем и лососем Кларка (Campton, Utter, 1985), гистограммы значений гибридного индекса ясно показывают области значений, характерные для исходных видов в зоне симпатрии, а также отмечают существование третьей, промежуточной группы в одной из локальностей (рис. 10.2.3). Авторы исследования интерпретируют особей этой промежуточной группы, как естественно встречающихся гибридов.

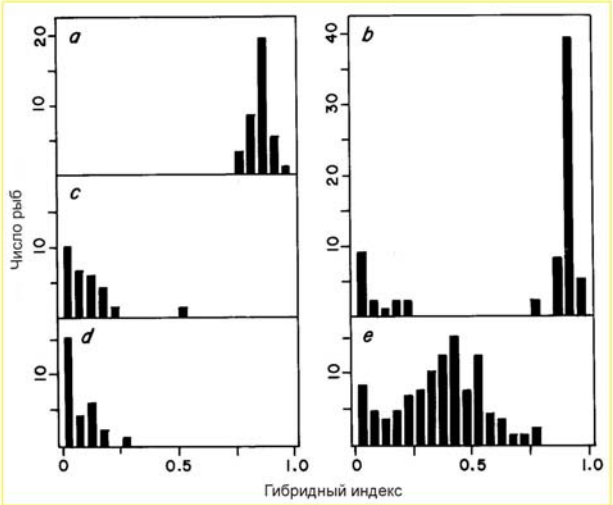


Рис. 10.2.3. Значения гибридного индекса (IH) для двух симпатричных видов форелей, анадромного стальноголового лосося (Salmo gairdneri) и прибрежного лосося Кларка (Salmo clarki clarki), а также их предполагаемых гибридов для трех выборочных участков небольшого ручья.

Особи со значениями индекса близкими к 1 или 0 проявляют составные генотипы по белковым локусам, которые имеют высокую относительную вероятность появления у S. gairdneri или S. clarki clarki, соответственно: (a) Участок 1, возраст рыб 0+ (только gairdneri); (b) Участок 2, возраст рыб 0 + (clarki и gairdneri, гибридов нет); (c) Участок 2, возраст рыб 1 + (clarki + 1 неизвестная или гибридная рыба); (d) Участок 3, возраст рыб 1+ (только clarki); (e) Участок 3, возраст рыб 0+ (clarki + большое число предположительных гибридов) (По Campton, Utter, 1985).

Fig. 10.2.3. Hybrid index scores for two sympatric species of trout, steelhead trout (anadromous Saimo gairdneri) and coastal cutthroat trout (Saimo clarki clarki), and their supposed hybrids from three sample sites within a small stream.

Individuals with values of the index close to 1.0 or 0.0 expressed composite electrophoretic phenotypes that have a high relative probability of occurring in S. gairdneri or S. clarki clarki, respectively: (a) Site 1, age 0+ fish (gairdneri only); (b) site 2, age 0 + fish (clarki and gairdneri, no hybrids); (c) site 2, age 1 + fish (clarki + 1 unknown or hybrid fish); (d) site 3, age 1+ fish (clarki only); (e) site 3, age 0+ fish (clarki + a large number of suspected hybrids) (From Campton, Utter, 1985).

Мощность метода для обнаружения гибридизации будет увеличиваться с увеличением количества локусов, используемых при анализе (Campton, 1987). Это легко продемонстрировать на численном примере (Табл. 10.2.1). Если частоты альтернативных аллелей (например A_1 и A_2) в локусе равны 0.8 и 0.2 в одной популяции и 0.2 и 0.8 – в другой, то вероятность, что особь будет гетерозиготной в каждой популяции просто равна $2pq = 0.32$. Если две популяции скрещиваются, то частота гетерозигот среди потомков F_1 будет 0.68, что только в 2.1 раза выше ожидаемой частоты гетерозигот в каждой из двух родительских популяций. С увеличением числа дискриминирующих локусов это соотношение растет. Например, для четырех локусов частота тетралокусных гетерозигот среди гибридов F_1 уже более чем в 20 раз выше ожидаемых значений для гетерозигот в родительских популяциях (Табл. 10.2.1). Эта пропорция возрастает до 92.1 для шести дискриминирующих локусов.

Таблица 10.2.1. Вероятность особи быть гетерозиготной одновременно по шести локусам при наличии 1-6 дискриминирующих локусов в каждой из двух родительских популяций (P) и среди F_1 гибридов (P_h), где частоты альтернативных аллелей равны 0.8 и 0.2 в одной популяции и 0.2 и 0.8, - в другой. Вероятности вычислялись как $P = [2(0.8)(0.2)]^L$ для родительской популяции и как $P_h = [(0.8)(0.8) + (0.2)(0.2)]^L$ – для гибридов F_1, где L - число локусов. Эти выражения предполагают наличие равновесия по сцеплению между всеми локусами и соблюдения соотношения Х-В в каждой родительской популяции (По Campton, 1987)

Table 10.2.1. Probability of an individual being heterozygous simultaneously at one to six distinguishing loci in each of two parental populations (P) and among F_1 hybrids (P_h), where the frequencies of alternate alleles at each locus are 0.8 and 0.2 in one population and 0.2 and 0.8, in the other. The probabilities were calculated as $P = [2(0.8)(0.2)]^L$ for the parental populations and as $P_h = [(0.8)(0.8) + (0.2)(0.2)]^L$ for the F_1 hybrids, where L = number of loci. These expressions assume gametic equilibrium between all loci and Hardy-Weinberg genotypic proportions in each parental population (From Campton, 1987)

Популяция	**Число дискриминирующих локусов**					
	1	**2**	**3**	**4**	**5**	**6**
Родители (P)	0.3200	0.1024	0.0328	0.0105	0.0034	0.0011
F_1 гибриды (P_h)	0.6800	0.4624	0.3144	0.2138	0.1454	0.0989
P_h/P	2.13	4.52	9.60	20.4	43.3	92.1

Однако даже с четырьмя дискриминирующими локусами ожидаемая частота тетралокусных гетерозигот внутри каждой из родительских популяций в Табл. 10.2.1 является выше, чем 1%. Соответственно, как отмечено выше, особи не могут быть несомненно различимыми как гибриды, если две родительские популяции или вида не имеют полной фиксации альтернативных аллелей по одному или более локусов.

Преимущество индексного метода состоит в том, что значения могут быть рассчитаны для каждой отдельной особи, даже если в данных имеются пропущенные значения по одному или большему числу локусов. Основной недостаток метода состоит в том, что он требует априорной оценки профиля частот аллелей для каждой из двух родительских популяций или таксонов. Чтобы преодолеть эти затруднения, можно оценивать вероятность гибридизации по серии двумерных графиков по главным компонентам. В данном случае, как и выше, многомерный подход имеет преимущество.

Митохондриальная и ядерная ДНК как маркеры генотипов

Использование рестриктаз для определения полиморфизма нуклеотидных сайтов митохондриальных ДНК (мтДНК) или ядерных ДНК (яДНК) дает ряд других методов маркирования генов и их аллельной изменчивости в природных популяциях (см. Глава 13). МтДНК наследуются по материнской линии и поэтому дают важное средство исследования гибридизации, посредством которого можно отличить в F_1 геномы родительских популяций или видов, если они были охарактеризованы по мтДНК.

Исследование генотипов по мтДНК, в комбинации с яДНК маркерами, или ферментными локусами, может позволить определить интрогрессию мтДНК от одного вида в ядерное окружение другого вида, если гибриды и их потомство фертильно. Такая интрогрессивная гибридизация требует успешности возвратных скрещиваний предковой гибридной самки с самцами родительских видов или других таксонов. Эта интрогрессия будет независимой от рекомбинационных и сегрегационных событий, реализующихся в ядерном геноме, если не вовлечен естественный отбор, направленный на поддержание ядерно-цитоплазматической совместимости (Takahata, Slatkin, 1984; Nei, 1987). Большое число примеров интрогрессии мтДНК (см. ниже) показывают, что такой отбор, если он и есть в принципе, то в природе встречается не часто. Таким образом, наличие примеров обладания «чужеродной» мтДНК у особей природных видов, идентифицированным другими методами, может являться доказательством гибридизации

близко родственных видов (таксонов). Такая межвидовая передача мтДНК обнаружена для видов родов *Drosophila, Mus* и *Rana* (Powell, 1983; Ferris et al., 1983; Spolsky, Uzzell, 1984; Yonekawa et al., 1984; 2000).

Ядерные ДНК также дают строгие доказательства наличия гибридов у организмов, обитающих в морской среде (рис. 10.2.4).

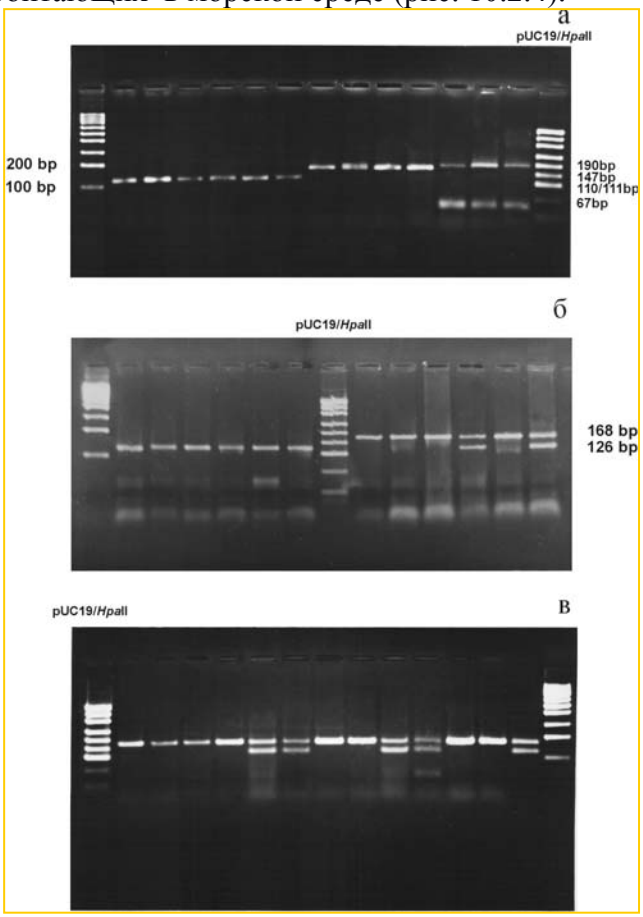


Рис. 10.2.4. Результаты анализа генотипов трех видов мидий рода Mytilus посредством ПЦР неповторяющегося региона гена клеящего белка биссуса. Ml - стандартный маркер, 100-тпн., M2, pUC19 ДНК обработанная рестриктазой Hpall, (a): 1-3, M. galloprovincialis (Франция); 4-6, M. galloprovincialis (островок Фальшивый, залив Петра Великого, Японское море); 7-10, M. trossulus (островок Фальшивый); 11-13, M. edulis (остров Гельголанд, Германия), (b): 1-6, M. galloprovincialis (островок Фальшивый); 7-12, M. trossulus (бухта Прибойная, залив Восток, залив Петра Великого, Японское море) (лунки 10 и 12, гибриды), (c): 1-6, M. trossulus (бухта Прибойная); 7-13, коллектор в заливе Восток (лунки 5, 6, 9, 10 и 13, гибриды) (По Скурихина, Картавцев и др., 2001).

Fig. 10.2.4. Results of analysis of genotypes of three species belonging to the genus Mytilus by means of PCR of the nonrepetitive region of the adhesive protein of bissus. Ml, a 100-kb standard marker, M2, pUC19 DNA digested by the Hpall restriction endonuclease, (a): 1-3, M. galloprovincialis (France); 4-6, M. galloprovincialis (Ostrovok Fal'shivyi, Peter the Great Bay); 7-10, M. trossulus (Ostrovok Fal'shivyi); 11-13, M. edulis (Helgoland Island, Germany), (b): 1-6, M. galloprovincialis (Ostrovok Fal'shivyi); 7-12, M. trossulus Priboinaya Bay, the Vostok Bay) (lanes 10 and 12, hybrid forms), (c): 1-6, M. trossulus (Priboinaya Bay); 7-13, the collector of the Vostok Biological Station (lanes 5, 6, 9, 10, and 13, hybrid forms) (From Skurikhina, Kartavtsev et al., 2001).

Использование мтДНК для исследования естественной гибридизации у рыб и морских беспозвоночных начато с середины 80-х годов предыдущего века (Avise, Saunders, 1984; Avise et al., 1983; 1984). Эвис и Сандерс (Avise, Saunders, 1984) использовали мтДНК в сочетании с аллозимами для исследования частоты гибридизации между 9 видами солнечных рыб (род *Lepomis)* из двух мест обитания на юго-востоке США. Основные итоги этого исследования можно суммировать в четырех пунктах.

1. Гибридизация происходит с относительно низкой частотой, но в нее вовлечены 5 из 9 исследованных видов.

2. Не обнаружено ни мтДНК, ни аллозимных доказательств существования интрогрессии генов между видами рода *Lepomis*; все гибриды оказались исключительно потомством F_1.

3. Каждый обнаруженный гибрид представляет скрещивание между наиболее распространенным и редким видом.

4. В шести из семи возможных гибридных комбинаций скрещиваний материнский родитель был представителем редкого вида, как было установлено по генотипу мтДНК. Эта особенность была объяснена как результат интенсивной конкуренции среди самцов за партнера по спариванию и общим промискуитетом самок.

Число примеров анализа мтДНК может быть легко увеличено. Мы рассмотрим некоторые из них в разделе 10.4.

10.3. ГЕНЕТИЧЕСКИЕ ИЗМЕНЕНИЯ В ГИБРИДНЫХ ПОПУЛЯЦИЯХ

Установление фактов естественной гибридизации требует проведения экспериментального и статистического анализа особей из природы. До проведения экспериментов и статистического анализа исследователи, как правило, не знают, имела место гибридизация или нет. Наиболее простым является случай исследования одной свободно скрещивающейся популяции, с отсутствием в прошлом гибридизации, или смешения представителей двух таксонов, и с небольшим числом встречающихся гибридов F_1. Другие и более сложные ситуации вполне возможны в природе. Если гибридизация и интрогрессия обнаружены, то необходимо оценить, каково их направление и насколько часто происходят эти события. В данном разделе мы ознакомимся с некоторыми теоретическими моделями популяционной генетики, которые могут оказаться полезными для различения случаев гибридизации.

Отклонения от соотношения Харди-Вайнберга

Первым этапом при анализе частот аллелей является оценка их однородности, которая позволяет сделать предварительное заключение, имеем ли мы дело с выборкой из одной популяции (см. Глава 9). Единая, свободно скрещивающаяся популяция должна проявлять по каждому локусу соотношения Х-В. С другой стороны, смесь двух не скрещивающихся популяций с различными частотами аллелей должна давать дефицит гетерозигот (и избыток гомозигот) по сравнению с соотношениями, ожидаемыми из равновесия Х-В. Таким образом, выявление эффекта Валунда является главным инструментом для определения наличия гибридизации в монолокусных случаях (см. Глава 9). Есть несколько специальных формул для ситуации, представляющей наличие двух популяций, которые формируют гибридов (Campton, 1987).

Здесь не все так просто, как хотелось бы. В частности, как было отмечено в двух предыдущих Главах, динамика частот аллелей зачастую имеет комплексную природу, с участием четырех факторов и естественного отбора, как одного их них. Соответственно, эффект Валунда может быть вызван не подразделенностью на две репродуктивные группы (или большее их число), а дифференцирующим влиянием отбора. Например, был описан случай гибридизации между двумя номинальными видами рыб рода *Notropis* (Cyprinidae) (Dowling, Moore, 1984). Авторы (Dowling, Moore, 1984) считают, что эти данные требуют признания *N. cornutus* и *N. chrysocephalus* как самостоятельных видов, а не подвидов, поскольку «гибридизация между подвидами должна давать равновесные Х-В генотипические частоты, тогда как гибридизация между видами даст значительный дефицит гетерозигот, либо за счет ассортативного скрещивания (прекопуляционная изоляция), либо отбора против гибридов (пост-копуляционная изоляция)». В последующем исследовании авторам удалось показать, что дефицит гетерозигот связан с естественным отборм против гибридов, а не с ассортативностью скрещиваний (Dowling, Moore, 1985). Мы знаем (Лекции 8-9), что не только подвиды, но и различные демы дают дефицит гетерозигот вследствие эффекта Валунда и, поэтому, ясно, что здесь необходимо дальнейшее усовершенствование интерпретации данных.

Неравновесие по сцеплению

Неравновесие по сцеплению (или гаметическое неравновесие) относится к не случайной ассоциации аллелей (гамет) между различными локусами диплоидного генома. Имеется обширная литература, в которой описываются условия, необходимые для создания и поддержания неравновесия по сцеплению (Hedrick et al., 1978; Животовский, 1984; Hartl, Clarke, 1989). Подразделенность популяции на несколько субпопуляций, также как скрещивание двух или большего числа ранее изолированных таксонов или популяций, является возможным механизмом создания неравновесия по сцеплению (Lewontin, 1974; Ohta, 1982; Животовский, 1984). Теоретическая оценка неравновесия по сцеплению поэтому является существенной для понимания динамики гибридизации и интрогрессии. Популяция находится в равновесии по сцеплению, если частоты гамет, несущих аллели двух или большего числа локусов являются равными произведению соответствующих частот аллелей.

Давайте рассмотрим два диаллельных локуса. Если p_1 и p_2 – это частоты двух аллелей (A1 и A2) по одному локусу, а q_1 и q_2 – частоты двух аллелей (B1 и B2) по второму локусу, то тогда гаметическое равновесие достигается в популяции, если $X_{11} = p_1q_1$, $X_{12} = p_1q_2$, $X_{21} = p_2q_1$ и $X_{22} = p_2q_2$, где X_{ij} – это частота гамет в популяции, которые несут аллели A_i и B_j. Если аллели между различными локусами ассоциированы не случайно (то есть, $X_{ij} \neq p_iq_j$), тогда можно показать, что

$$X_{11} = p_1q_1 + D \quad X_{22} = p_1q_2 - D \quad X_{21} = p_2q_1 - D \quad X_{22} = p_2q_2 + D. \qquad (10.2)$$

Здесь D измеряет величину межлокусной ассоциации между аллелями внутри гамет (Hedrick, 1983). D определяется неравновесием по сцеплению и задается как $D = X_{11} - p_1q_1$ для двухлокусной системы (Lewontin, Kojima, 1960). Подставляя $X_{11} + X_{12}$ вместо p_1 и $X_{11} + X_{21}$ вместо q_1, в предыдущем определении D, можно записать

$$D = (X_{11}\, X_{22}) - (X_{12}\, X_{21}). \qquad (10.3)$$

Таким образом, величина неравновесия по сцеплению между двумя локусами является просто разницей между произведением гаметических частот в цис-положении (т.е., частотами A_1B_1 и A_2B_2) и произведением гаметических частот в транс-положении (т.е., частотами A_1B_2 и A_2B_1). Иными словами, величина D измеряет отклонение частот неаллельных генов от независимого распределения. Неравновесие по сцеплению может быть определено, естественно, для большего числа локусов, но эти выражения более сложны (Bennett, 1954; Crow, Kimura, 1970, p. 50) и находятся вне задач данной лекции.

Величина возможных значений D зависит от частот аллелей по обоим локусам в рассматриваемой популяции. Исходя из выражения (10.3) можно рассчитать, что D достигает максимального значения 0.25, если $X_{11} = X_{22} = 0.5$ ($X_{12} = X_{21} = 0$), а минимального значения $D = -0.25$, если $X_{12} = X_{21} = 0.5$ ($X_{11} = X_{22} = 0$). Эти максимальные и минимальные значения D достигаются только, когда частоты аллелей по всем локусам равны, т.е., если $p_1 = q_1 = p_2 = q_2 = 0.5$. Поэтому абсолютное значение D достаточно трудно интерпретировать, особенно когда сравниваются несколько значений D. Поэтому для измерения неравновесия по сцеплению предложена более подходящая статистика $D' = D/D_{max}$ (Lewontin, 1964; 1974). Здесь D_{max} – это максимальное абсолютное значение D для конкретного набора частот аллелей (Hedrick, 1983). Значение D' варьирует в диапазоне от -1.0 до +1.0 и представляет собой относительную меру неслучайной ассоциации аллелей различных локусов. Определены и другие варианты оценки неравновесия (Weir, 1979). В качестве меры неравновесия по сцеплению можно ввести коэффициент корреляции R, нормировав D на ковариацию (произведение) частот аллелей:

$$R = \sqrt{[D / (p_1\, q_1 * p_2\, q_2)]}. \qquad (10.4)$$

Если принять значение 1 для A_1 и B_1, а также 0 для A_2 и B_2. R также может принимать значения в диапазоне от -1.0 до +1.0 при равенстве частот аллелей ($p_1 = p_2$ и $q_1 = q_2$). Для расчетов более приемлем вариант оценки D, предложенный Вейром (Weir, 1979), позволяющий использовать

фактические частоты генотипов и имеющий разработанный статистический критерий значимости отклонения от равновесия (Животовский, 1991, с.120).

Равновесие для двух локусов не устанавливается в первом поколении как в случае одного локуса. Оно достигается постепенно во времени, так что после *t* поколений

$$D_t = (1-r)^t * D_0. \qquad (10.5)$$

r - коэффициент рекомбинации, D_0 – неравновесие в нулевом поколении.

Для тесно сцепленных локусов (т.е., r << 1/2), неравновесие может существовать почти неограниченно долго, особенно в малых популяциях, где рекомбинация будет контрбалансироваться генетическим дрейфом.

Выше были рассмотрены теоретические аспекты оценки неравновесия по сцеплению. Однако оценка этого неравновесия на практике в природных популяциях базируется на выборках особей (не гамет). В этом случае гаметические генотипы обычно прямо не обнаруживаются, а определяются исходя из их диплоидных генотипов. Например, особей генотипа *A1A1B1B2* можно интерпретировать, как объединение гамет *A1B1* и *A1B2*. По этой схеме гаметические генотипы могут быть выведены по особям, которые гомозиготны по одному или обоим локусам. Однако гаметические генотипы нельзя вывести по двойным гетерозиготам, поскольку гаметическая фаза не известна. Так, различение *A1B1/A2B2* и *A1B2/A2B1* в норме возможно лишь в планируемых скрещиваниях. Вот почему неравновесие по сцеплению не может быть точно рассчитано для природных популяций диплоидных организмов. Тем не менее, предложено несколько методов оценки неравновесия по сцеплению для природных видов, когда двойные гетерозиготы двух типов не могут быть различимы. Наиболее известны из них метод максимального правдоподобия (ММП; Hill, 1974) и составной метод Барроу (Burrow) (Weir, 1979). Во втором методе *D* разбивается на два компонента ($D = D_W + D_B$) и обычно весьма хорошие результаты достигаются в пошаговой процедуре. Методика оценки *D* исходно была предложена несколько раньше (Cockerham, Weir, 1977). Пример, который иллюстрирует использование ММП и составного метода при оценке неравновесия по сцеплению приведен в Табл. 10.3.1.

Таблица 10.3.1. Оценки неравновесия по сцеплению на основе метода максимального правдоподобия (Hill, 1974) и составного метода Барроу (Weir, 1979) для потенциальных гибридов стальноголового лосося *(Saimo gairdneri)* и прибрежного лосося Кларка *(S. clarki clarki) (По Campton, 1987)*

Table 10.3.1. Estimates of linkage disequilibrium using Hill's (1974) maximum likelihood method and Burrows's composite measure (Weir, 1979) for suspected hybrids of steelhead trout *(Saimo gairdneri)* and coastal cutthroat trout *(S. clarki clarki).* The statistic *R* is the estimated correlation between alleles, and X^2 is the goodness of fit statistic ($X^2 = NR^2$) with d.f.=1 for testing *Ho: D = 0* (MLE) or $D + D_B = 0$ (Weir, 1979). The data and estimates are for the same fish as shown in Fig. 9.2.2,e in Campton and Utter (1985) *(From Campton, 1987)*

Локусы	**Максимальное правдоподобие**			**Составной метод**		
	D	***R***	X^2	$D + D_B$	***R***	X^2
GLD-1, ME-4	0.105	0.427	16.79***	0.026	0.239	5.26*
GLD-1, SDH-1	0.037	0.161	2.29	0.033	0.229	4.69*
ME-4, SDH-1	0.049	0.218	3.93*	0.062	0.248	5.09*

Примечание. *R* – это оценка корреляции между аллелями, X^2 – хи-квадрат оценка соответствия с ожидаемыми значениями ($X^2 = NR^2$) с d.f. = 1 для тестирования *Ho: D = 0* (ММП) или $D + D_B = 0$ (Weir, 1979). Приведены данные и оценки для тех же самых рыб, что на рис. 10.2.3,е. Значимость: * P < 0.05, *** P < 0.001

Оценивание пропорций в смеси

Гибридные популяции, которые прошли далее поколения F_1 с небольшим сокращением приспособленности (жизнеспособности или плодовитости) часто представляют собой

генетические или популяционные смеси. В понимание смеси здесь входит образование новых генотипических комбинаций через рекомбинацию. Впервые такое понимание было употреблено при описании скрещиваний между расами человека, последовавшее после установления вторичного контакта в западном полушарии (Glass, Li, 1953; Glass, 1955). Многие популяции человека в Северной и Южной Америке в действительности представляют собой генетические смеси двух или более рас, ранее находившихся в изоляции (Cavalli-Sforza, Bodmer, 1971). Если гибридизация была установлена, и если можно сделать заключение, что возникла единая свободно скрещивающаяся популяция, то тогда возникает необходимость оценить пропорциональный вклад родительских таксонов или популяций в генетическую смесь (Campton, Johnston, 1985).

Если гибридизация осуществляется в течение нескольких поколений в одном направлении, так, что гибридная популяция формируется каждое поколение за счет доли *m* одной из родительских популяций (скажем популяции 2) и доли, равной 1 - *m* гибридной популяции предыдущего поколения, то тогда ожидаемая частота аллеля в гибридной (смешанной) популяции в поколении $t[p_h(t)]$ задается выражением (Campton, 1987):

$P_h(t) = (1-m)^t p_1 + [1 - m)t] p_2$. Если *t* известно, то можно оценить *m,* среднюю скорость смешения на поколение, как долю

$$m = 1 - [(P_h(t) - p_2) / (p_1 - p_2)]^{1/t} . \qquad (10.7)$$

Эта доля может рассматриваться как простая однонаправленная модель миграции, где $P_h(0) = p_1$. Принципиальная схема изменения частоты аллеля вследствие миграции и численный пример были даны в Главах 8-9. Выражение (10.7) исходит из ранней оценки динамики смешения рас в США (Glass, Li, 1953), где преобладала интрогрессия генов от кавказской расы к американским черным поселенцам (Glass, 1955; Roberts, 1955; Saldana, 1957). Выражение (10.7) также подходит и для разрешения проблем гибридизации у любых других животных, включая морских организмов, поскольку обычная практика создания стад на рыбоводных заводах (например, для тихоокеанских лососей) состоит в первоначальном скрещивании двух или более локальных популяций, а потом продолжение скрещиваний с другой популяцией. Аналогично, постоянное введение экзотических видов в нативные биотопы других видов поднимает близкие проблемы, если межвидовые гибриды плодовиты (Busack, Gall, 1981).

Выражение (10.7) (а также 14 в Campton, 1987) дают оценки пропорций в смеси из двух популяций, основываясь на данных только по одному локусу. На самом деле, как правило, имеются данные для многих локусов и вклад в смесь могут вносить более чем две популяции. Элстон (Elston, 1971) описывает с помощью метода максимального правдоподобия и метода наименьших квадратов оценки пропорций в смесях для подобных генерализованных случаев. Детальное ознакомление с этими подходами лежит вне задач данной лекции, но желающие могут найти необходимые разъяснения в оригинальной работе (Elston, 1971) (рис. 10.3.1).

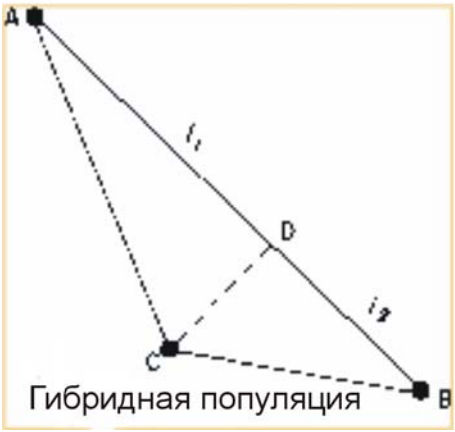


Рис. 10.3.1. Графическое представление относительного расположения двух гипотетических родительских популяций и гибридной популяции в n-мерном пространстве выборок и оцененных по частотам аллелей. Гибридная популяция не колинеарна с родительскими популяциями вследствие эффекта основателя, генетического дрейфа, ошибки выборки и, возможно, естественного отбора. В результате оценки пропорций в смеси, полученные из выражения 14 (Campton, 1987), будут варьировать в зависимости от используемых локусов. Проекции гибридной популяции по методу наименьших квадратов находятся в точке D. Если I_1 и I_2 являются длинами отрезков AD и BD, соответственно, то тогда оценки по наименьшим квадратам I_1 и I_2 (пропорции смесей популяции 1 и

2, соответственно) равны $I_2 / (I_1 + I_2)$ и $I_1 / (I_1 + I_2)$ (По Thompson, 1973).
Fig. 10.3.1. Graphical representation of the relative positions of two hypothetical parental populations and a hybrid population in the n-dimensional sample space of estimated allele frequencies. The hybrid population is not collinear with parental populations because of founder effects, genetic drift, sampling error, and possibly natural selection. As a result, estimates of admixture proportions obtained by equation 14 (Campton, 1987) will vary among loci. The least squares estimate projects the hybrid population to point D. If I_1 and I_2 are the lengths of line segments AD and BD, respectively, then the least squares estimates of I_1 and I_2 (the admixture proportions of populations 1 and 2, respectively) are $I_2 / (I_1 + I_2)$ and $I_1 / (I_1 + I_2)$ respectively (After Thompson, 1973).

10.4. ЭМПИРИЧЕСКИЕ ИССЛЕДОВАНИЯ ГИБРИДНЫХ ЗОН

При исследовании больших гибридных зон наиболее полезными являются молекулярные маркеры, включая аллозимные локусы. По аллозимам, в частности, описана широкая гибридная зона между двумя морфологически различными подвидами ушастого окуня *Lepomis macrochirus macrochirus* и *L. m. purpurescens* (Avise, Smith, 1974). Эти два подвида имеют различные и почти фиксированные аллели по двум аллозимным локусам в зоне их аллопатрии в юго-восточных районах США. Кроме того, существует зона генетического смешения в месте перекрывания ареалов двух подвидов (рис. 10.4.1).

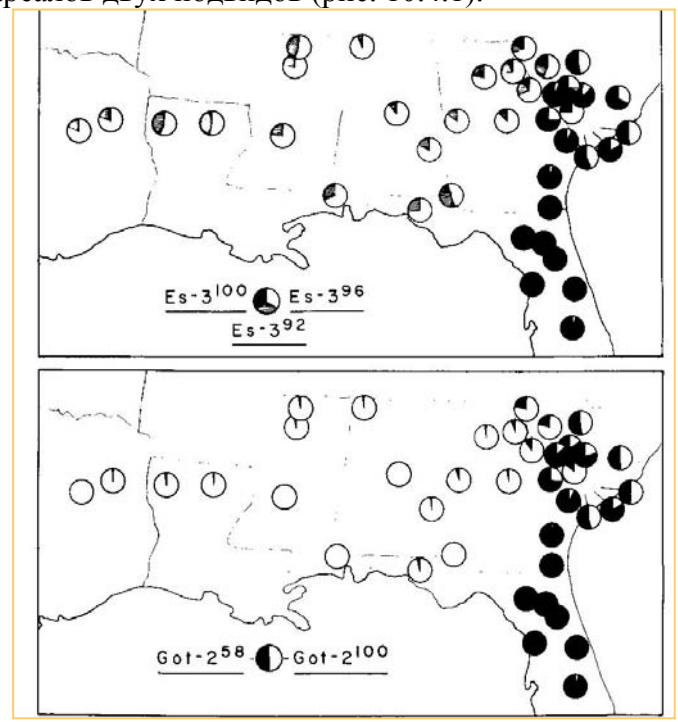


Рис. 10.4.1. Географическое распределение частот аллелей по двум аллозимным локусам (Es-3 и Got-2) для двух подвидов ушастого окуня, Lepomis macrochirus macrochirus и L. m. purpurescens, в юго-восточных участках США. На карте кружки со смешанными цветами (штриховка) обозначают наличие выявленных гибридов (По Avise, Smith, 1974).
Fig. 10.4.1. Geographic distributions of allele frequencies at two allozyme loci (Es-3 and Got-2) for two subspecies of bluegill sunflsh, Lepomis macrochirus macrochirus and L. m. purpurescens, in the southeastern United States. On the map circles with mixed colors denote hybrids occurrence (From Avise, Smith, 1974).

Сходная зона перекрывания и гибридизации обнаружена в юго-восточных районах США между северным и флоридским подвидом большеротого окуня, *Micropterus salmoides salmoides* и *M. s. floridanus* (рис. 10.4.2). Географическое распределение частот аллелей в зоне гибридизации для этих двух рыб имеет большое сходство (см. рис. 10.4.1 и рис. 10.4.2).

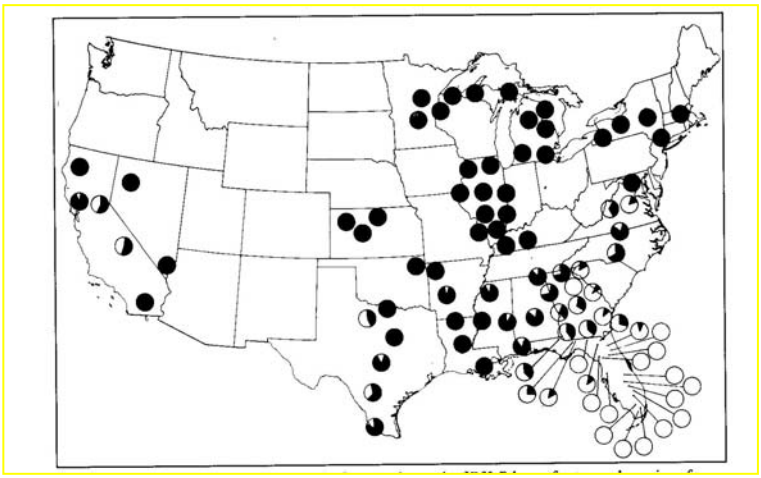


Рис. 10.4.2. Географическое распределение частот аллелей по локусу IDH-B для двух подвидов большеротого окуня, Microplerus salmoides salmoides и M. s. ftoridanus, в США (По Phillipp et al., 1983).*
Fig. 10.4.2. Geographic distribution of allele frequencies at the IDH-B locus for two subspecies of largemouth bass, Microplerus salmoides salmoides and M. s. ftoridanus, in the United States (From Phillip et al., 1983).*

На основе генетико-биохимических данных также обнаружена зона очевидной интрогрессии между нативной и интродуцированной радужной форелью в реке Якима (Yakima River) в штате Вашингтон (рис. 10.4.3). Эвис и Смит (Avise, Smith, 1974) выдвинули предположение, что острова земли или рефугии, сформировавшиеся в течение Плейстоценовых оледенений в юго-восточных США, изолировали пресноводную фауну в этом регионе вплоть до наступления вторичного контакта в исторически недавнее время. Наличие таких гибридных зон имеет большое значение для эволюционно-генетических работ, и итоги их исследований обсуждались неоднократно (Remington, 1968, Woodruff, 1973; Moore, 1977; Barton, 1979; 1983, Barton, Hewitt, 1985).

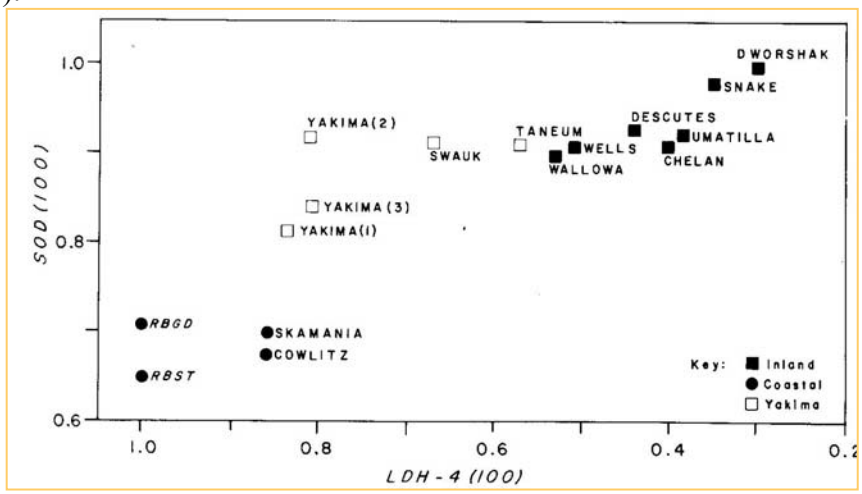


Рис. 10.4.3. Частоты основного аллеля (100) по локусам LDH-4 и SOD* в популяциях радужной форели (Salmo gairdneri) в Штате Вашингтон. Река Якима (Yakima River) является важным притоком реки Колумбия (Columbia River) восточнее горного каскада, но радужная форель, населяющая эту реку (точки 1, 2 и 3), а также два ручья-притока (Swauk и Tanеum), проявляет частоты аллелей, промежуточные с частотами в популяциях основной материковой части и интродуцентов в береговые участки. RBGD и RBST представляют две искусственно воспроизводимые популяции не анадромной радужной форели (По Campton, Johnston, 1985).*
Fig. 10.4.3. Frequencies of the common (100) allele at the LDH-4 and SOD* loci for populations of rainbow trout (Salrno gairdneri) in Washington state. The Yakima River is a major tributary of the Columbia River east of the Cascade Mountains, but rainbow trout inhabiting this river (sites 1, 2, and 3) and two tributary creeks (Swauk and Taneum) expressed allele frequencies intermediate to those for other inland populations and those for introduced coastal populations. RBGD and RBST represent two hatchery populations of nonanadromous rainbow trout (From Campton, Johnston, 1985).*

Впечатляющими являются результаты исследования сверчков *Allonemobius fasciatus* и *A. socius,* обнаружившим по многим локусам как эффект Валунда (положительные значения *Fis*, Табл. 10.4.1), так и гаметическое неравновесие (значимые *D*, Табл. 10.4.2).

Таблица 10.4.1. Внутрипопуляционное разнообразие *(Fis*) по маркерным локусам в смешанных популяциях сверчка (По Howard, Waring, 1991)
Table 10.4.1. The inbreeding coefficients *(Fis* values) of marker loci in mixed populations of ground crickets (From Howard, Waring, 1991)

Локус / Место	***Idh-1***	***Hk***	***Pep-3***	***Got***
ILL8	*0.399*	*0.763*	*0.573*	*0.385*
ILL9	*0.560*	*0.753*	*0.721*	*0.559*
ILL 13	*0.608*	*0.606*	*0.595*	*0.329*
ILL 14	*0.682*	*0.843*	*0.928*	*0.129*
MTN26	*0.646*	*0.812*	*0.546*	*0.189*
MTN27	*0.550*	*0.725*	*0.892*	*0.402*
MTN33	*0.332*	*0.714*	*0.514*	*0.260*
MTN37	*0.211*	*0.365*	*0.458*	*-0.113*
MTN40	*0.092*	*0.697*	*0.414*	*0.045*
EC57	*0.109*	*0.419*	*0.596*	*0.246*
EC58	*0.135*	*0.664*	*0.453*	*0.388*

Таблица 10.4.2. Составная мера неравновесия по сцеплению Барроу (Weir, Cockerham, 1989) для смешанных популяций сверчка *A. fasciatus* и *A. socius.* Цифры в рядах рядом с оценками неравновесия, которые выделены, представляют значения X^2 для гипотезы $D = 0$ (d.f. =1) (По Howard, Waring, 1991)

Table 10.4.2. Burrow's composite measure of linkage disequilibrium (Weir, Cockerham, 1989) for mixed populations of ground cricket *A. fasciatus* and *A. socius.* The numbers in the row beside the estimates of linkage disequilibrium (in bold) represent X^2 values and test the hypothesis $D = 0$ (d.f. = 1) (From Howard, Waring, 1991)

Место	***Idh-l/Hk***	***Idh-1/Pep-3***	***Idh- I/Got***	***Hk/Pep-3***	***Hk/Got***	***Pep-3/Got***
ILLS	***0.229*** *15.208***	***0.229*** *16.912***	***0.238*** *22.644***	***0.286*** *21.458***	***0.219*** *18.140***	***0.206*** *15.799***
ILL9	***0.323*** *100.968***	***0.270*** *56.571***	***0.271*** *62.075***	***0.317*** *65.853**	***0.309*** *65.407***	***0.295*** *68.521***
ILL 13	***0.239*** *18.080***	***0.265*** *22.207***	***0.198*** *24.600***	***0.289*** *26.421***	***0.218*** *19.493***	***0.223*** *37.843***
ILL 14	***0.308*** *15.204***	***0.341*** *17.374***	***0.182*** *9.778***	***0.320*** *14.872***	***0.185*** *9.396***	***0.211*** *9.903***
MTN26	***0.350*** *122.048***	***0.320*** *101.375***	***0.146*** *15.094***	***0.377*** *168.916***	***0.164*** *16.138***	***0.166*** *17.389***
MTN27	***0.324*** *132.606***	***0.334*** *112.741***	***0.235*** *68.255***	***0.350*** *108.934***	***0.223*** *59.123***	***0.232*** *46.449***
MTN33	***0.234*** *16.987***	***0.213*** *16.241***	***0.149*** *11.303***	***0.297*** *23.606***	***0.218*** *18.279***	***0.198*** *17.505***
MTN37	***0.164*** *21.811***	***0.145*** *17.513***	***0.038*** *2.481 NS*	***0.198*** *36.796***	***0.026*** *1.117NS*	***0.020*** *0.692 NS*
MTN40	***0.067*** *3.970**	***0.044*** *2.570 NS*	***0.016*** *0.535 NS*	***0.042*** *2.659 NS*	***0.012*** *0.317 NS*	***0.006*** *0.103NS*
EC57	***0.025*** *9.612***	***0.053*** *2.575 NS*	***0.092*** *5.991**	***0.051*** *2.695 NS*	***0.099*** *6.568**	***0.007*** *0.027 NS*
EC58	***0.178*** *33.994***	***0.035*** *3.270 NS*	***0.110*** *18.269***	***0.065*** *11.526***	***0.156*** *30.450***	***0.057*** *11.253***

Примечание. NS – не значимо, * $p < 0.05$, ** $p < 0.005$.

Было сделано заключение, что одно из теоретических предсказаний БКВ не подтверждается на этом материале. В противоположность ожиданию сила репродуктивной изоляции между двумя видами была одинаковой как в Иллинойсе, так и в горах Аппалачей. Этот факт предполагает, что если усиление изолирующего механизма и имеет место, то оно не зависит от ширины гибридной зоны.

Интересные результаты получены для обыкновенной мидии группы *Mytilus ex. group edulis* при исследовании большого материала из Нью-Фаундленда одновременно по аллозимным локусам и по морфометрии (рис. 10.4.4 – 10.4.5).

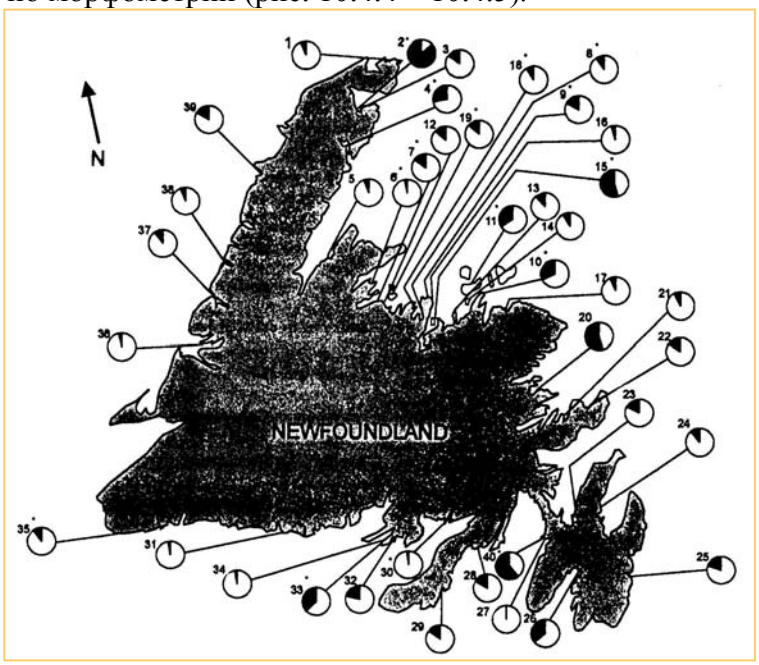


Рис. 10.4.4. Доли Mytilus edulis (белый цвет) и M. trossulus (черный цвет) в природных и культивируемых популяциях Нью-Фаундленда. Культивируемые популяции отмечены звездочкой (По Penney, Hart, 1999).
Fig. 10.4.4. Relative frequency of Mytilus edulis (white color) и M. trossulus (black color) in wild and cultivated populations in Newfoundland. Cultivated populations are denoted with asterisks (From Penney, Hart, 1999).

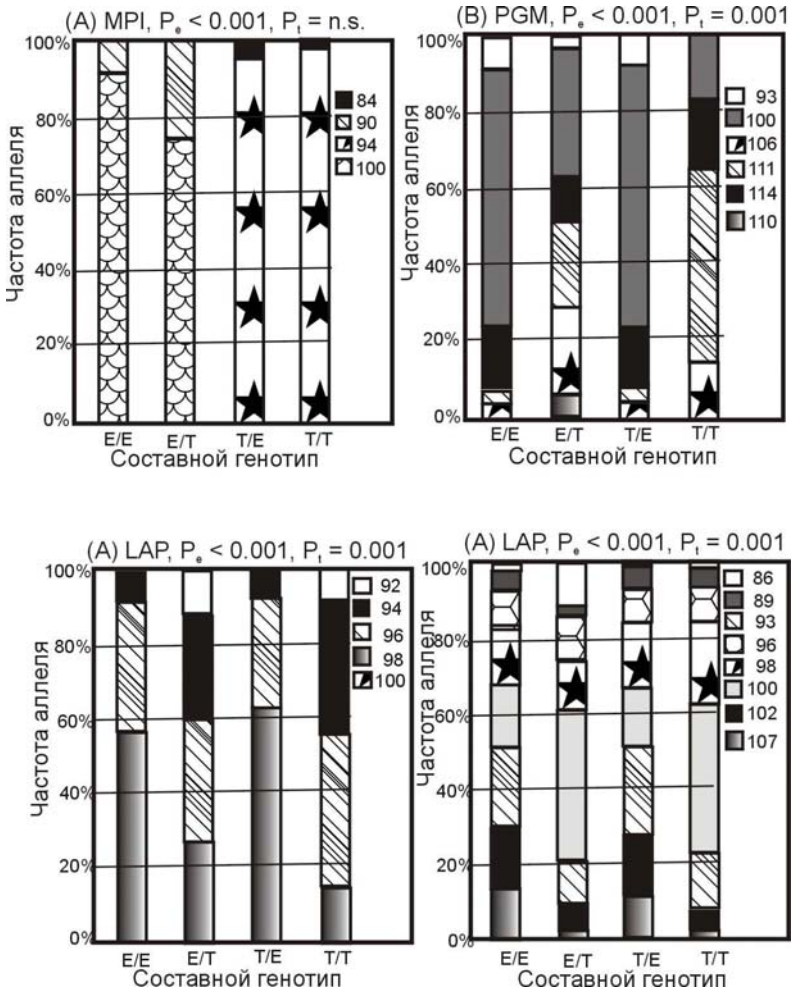


Рис. 10.4.5. Сопоставление частот аллелей для составных генотипических классов по четырем локусам (a) Mpi, (b) Pgm, (c) Lap и (d) Gpi. Вертикальные прямоугольники обозначают индивидуальные аллели пронумерованные, как дано в комментарии сбоку. Аллели с очень низкой частотой объединены с ближайшим аллелем по каждому из изученных локусов. Значения вероятностей, Pe и Pt даны для тестов X^2 по внутривидовым различиям между частотами аллелей (Pe = E/E против E/T; Pt = T/T против T/E) (По Penney, Hart, 1999).
Fig. 10.4.5. Character index profiles for ground cricket populations A. fasciatus and A. socius. collected along the Illinois transect (ILL1-ILL17). The height of the bar gives the proportion of individuals in a population having that score. The solid part of the bar represents the proportion of individuals with that character-index score having a genotype indicative of mixed ancestry. Each graph represents a different population. The N is sample size (From Howard, Waring, 1991).

Важно отметить: (1) присутствие гибридных особей как это видно по диагностическому локусу *MPI**, (2) увеличенную долю рекомбинантных генотипов по этому локусу в составном генотипе эдулис-подобного морфотипа (E/T) по сравнению троссулюс-подобным морфотипом (T/E; рис. 10.4.5, см. MPI – E/T и T/E), которая свидетельствует о преобладании интрогрессии в направлении T → E, (3) возможность, что гибриды могут быть не только F_1, учитывая результаты анализа ДНК из этого района (Comesana et al., 1999).

Оригинальные результаты для этой группы видов (*Mytilus spp.*) получены для района острова Ванкувер (рис. 10.4.6).

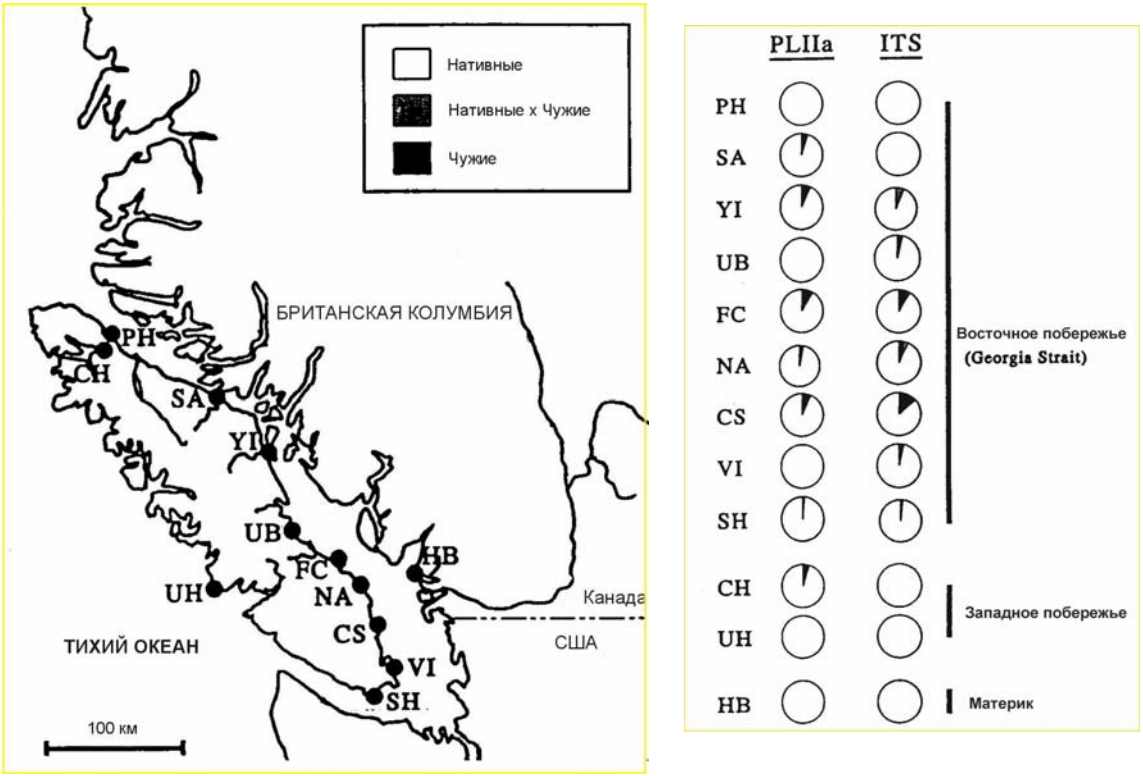


Рис. 10.4.6. Распределение 12 выборок обыкновенной мидии в юго-западной Британской Колумбии (слева) и присутствие «чужих» (M. edulis и/или M. galloprovincialis) и нативных (M. trossulus) генотипов по двум независимым ядерным маркерам ДНК (PLIIa и ITS). На карте показана локализация выборок, а справа обозначены частоты генотипов в этих выборках (По Heath et al., 1995).

Fig. 10.4.6. Southwestern British Columbia and Vancouver Island showing the 12 blue mussel sampling sites and the presence of the alien (M. edulis and/or M. galloprovincialis) and native (M. trossulus) genotypes as determined the two independent nuclear DNA markers (PLIIa and ITS) (From Heath et al., 1995).

Распределение чужих аллелей по двум маркерным локусам различается в различных выборочных точках, предполагая дифференциальную интрогрессию. Широкое распространение чужих аллелей, в сочетании с доказательствами интенсивной гибридизации между нативным и интродуцированным (чужим) видом указывает на то, что введенные в генофонд аллели, вероятно, существуют некоторое время в популяции мидии в Британской Колумбии (Heath et al., 1995).

Исследование, проведенное для обыкновенной мидии в другой части ареала, в Англии, также дало очень интересные результаты (рис. 10.4.7).

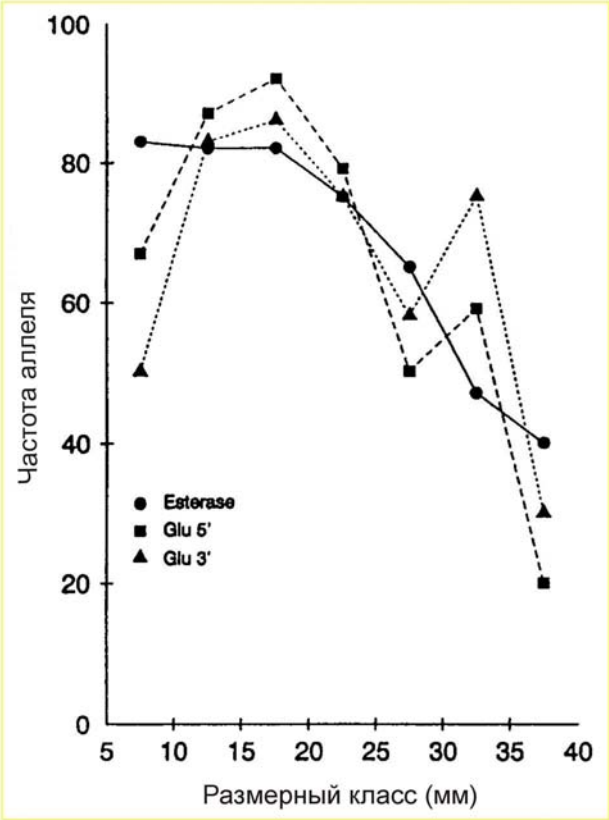


Рис. 10.4.7. Частота распределения у Mytilus edulis (e) аллелей эстеразы-D (Esterase), а также ДНК маркеров Glu-5' и Glu-3', как функция размера моллюсков в заливе Витсэнд (Whitsand Bay, UK) в зоне гибридизации популяций. Для оценки частот аллелей по маркеру Glu-5', M. edulis-специфичные 350 пн и 380 пн продукты рассматривались как различные аллели, тогда как для M. galloprovincialis 300 пн и 500 пн продукты считали единым аллелем. Таким образом, мидии с типом бэндинга 350 пн, 380 пн, или 350/380 пн были учтены как гомозиготы M. edulis, а с типом бэндинга 300 пн учтены как гомозиготы M. galloprovincialis, наконец, особи с типом бэндинга 300 пн в комбинации либо с 350 пн или 380 пн фракциями учитывались как гетерозиготы (По Rawson et al., 1996).

Fig. 10.4.7. Frequency of Mytilus edulis (e) alleles at the Esterase-D (Est), Glu-5' and Glu-3' DNA markers as a function of size for the Whitsand Bay, UK, hybrid zone population. For estimating allele frequencies at the Glu-5' marker, M. edulis-specific 350 bp and 380 bp products were treated as separate alleles whereas the M. galloprovincialis 300 bp and 500 bp products were assumed to be produced by a single allele. Thus, mussels with 350 bp, 380 bp, or 350/380 bp banding patterns were scored as M. edulis homozygotes, those with 300 bp patterns were scored as M. galloprovincialis homozygotes and those with a 300 bp band in combination with either the 350 bp or 380 bp band were scored as heterozygotes (From Rawson et al., 1996).

С использованием двух маркеров ДНК было показано, что гибридные мидии из залива Витсэнд (Whitsand Bay, UK) несут аллели генов, которые являются результатом внутригенной рекомбинации. Высокая частота (10%) этих рекомбинантных аллелей внутри гибридной популяции предполагает, что рекомбинация весьма часта в этом гене или что гибридизация между *M. edulis* и *M. galloprovincialis* существенна и происходит значительное время в эволюционной истории таксонов. Клина на рис. 10.4.9 может интерпретироваться как свидетельство естественного отбора. Хотя такие свидетельства обычно нелегко получить в нормальных условиях, найденная ассоциация, вместе с другими данными из гибридных зон группы *Mytilus ex. group edulis* (Skibinski et al., 1983; Gardner, Skibinsky, 1988; Wilhelm, 1993), заставляют нас сделать заключение об увеличении интенсивности естественного отбора в этих зонах, включая отбор против гибридов.

Данные настоящего раздела содержат не так много примеров конкретных гибридных зон. Главное то, что эти зоны существуют и не менее реальны, чем сами популяции и виды. Это надежно доказано с помощь нескольких генетических методов.

Ясно, что интрогрессия генов существует в природе, хотя в некоторых широких участках гибридных зон, например в группе *Mytilus ex. group edulis*, она может быть весьма ограниченной (Rawson et al., 1999) или быть асимметричной (см. выше, а также Heath et al., 1995; Rawson, Hilbish,

1998). Если принять, припоминая примеры из разделов 10.3-10.4, что многие перекрестно размножающиеся виды в морской и наземной среде широко интрогрессируют, то тогда мы должны признать, что ортодоксальная БКВ, как она определена в Главах 7-8, является неадекватной и требует уточнения. К счастью, виды ничего не знают о концепциях и продолжают, в большинстве, сохранять свою интегрированность вопреки провалу концепции.

Генетическое смешение, создание гибридных зон и автоматическое существование долговременного гаметического неравновесия могут иметь различные последствия – положительные и отрицательные. Положительными последствиями являются увеличение генетической изменчивости, а также «проверочная» роль гибридных зон (tension zone), которые являются «полем битвы» за видовую интегрированность. Наконец, гибриды, если иногда оказываются более приспособленными или просто обладающими нужными человеку свойствами, могут быть использованы прямо в аквакультуре или как источники для скрещиваний в селекционных программах. Отрицательные последствия весьма многочисленны: (1) неравновесие, ведущее к снижению общей приспособленности гибридной популяции или вида, (2) интрогрессия вредных (или просто неадаптированных) генотипов в аборигенную популяцию или вид, что ведет к их деградации и замещению другим видом (как правило, с худшими качествами и для человека и для экосистемы).

10.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 10

1. Программа SPECSTAT. Специфика использования модулей NEI-DIST и DISTANCE (см. файл Read.me).
2. Пакет программ NTSYS-pc: знакомство с основными программами и принципами управления.

ГЛАВА 11. КОЛИЧЕСТВЕННЫЕ ПРИЗНАКИ: НАСЛЕДОВАНИЕ И ЭВОЛЮЦИЯ

ГЛАВНЫЕ ВОПРОСЫ:

11.1. Количественная изменчивость и ее законы.
11.2. Гипотеза множественных факторов.
11.3. Основные задачи генетического исследования количественных признаков.
11.4. Наследуемость.
11.5. Картирование локусов количественных признаков (ЛКП). Эволюция ЛКП.
11.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 11.

РЕЗЮМЕ

1. Непрерывная изменчивость проявляется в скрещиваниях, вовлекающих количественные признаки, которые находятся под полигенным контролем. **Полигенное наследование** определяется, в основном, кумулятивным влиянием **аддитивных аллелей**.
2. **Полигенные признаки** могут анализироваться с использованием статистических методов, которые включают оценивание **средней**, **стандартного отклонения** и **стандартных ошибок** для них. Такой статистический анализ является описательным и может быть использован для разбивки общей фенотипической изменчивости на компоненты, выделяя, в том числе, и **генотипическую дисперсию**.
3. Для многих количественных признаков трудно установить точно, обусловлена ли их изменчивость генетическими факторами или факторами внешней среды. **Наследуемость** оценивается для КП как доля **генотипической изменчивости** по отношению к **фенотипической** изменчивости (общей изменчивости) признака: $h^2_B = \sigma^2_G / \sigma^2_{PH}$. **Коэффициент наследуемости** помогает установить соотношение наследственных и не наследственных факторов в детерминации хозяйственно важных количественных признаков и, поэтому, полезен в селекционной практике. Различают **наследуемость в широком смысле** и **наследуемость в узком смысле**.
4. Локусы, которые несут гены, контролирующие КП, называются **локусы количественных признаков** (**ЛКП**). Актуальной задачей генетики КП является **картирование ЛКП**, которое в последние годы активно осуществляется с применением молекулярно генетических маркеров генома.

Chapter 11. QUANTITATIVE TRAITS: INHERITANCE AND EVOLUTION

SUMMARY

1. **Continuous variation** is exhibited in crosses involving traits under **polygenic control**. Such traits are quantitative in nature and are inherited as a result of the cumulative impact of **additive alleles**.
2. **Polygenic characters** can be analyzed using statistical methods, which include the **mean**, the **variance**, the **standard deviation**, and the **standard error** of the mean. Such statistical analysis is descriptive, and can be used to make inferences about a population or to compare and contrast sets of data.
3. For many phenotypic characters, it is difficult to ascertain when variation depends on genetic or on environmental factors. **Heritability** is an estimate of the relative importance of **genetic factors** versus **nongenetic factors** and is determined correspondingly as a ration of **genotypic variance** and **phenotypic variance** in populations, $h^2_B = \sigma^2_G / \sigma^2_{PH}$. It can be calculated for many characters and is especially

useful in selective breeding of commercially valuable plants and animals. **Coefficient of heritability** is known in two modes as **narrow and wide sense heritability**.
4. Loci bearing genes that control a quantitative trait are called **quantitative trait loci** (**QTLs**). Location and distribution of QTLs within a genome can be ascertained by application of molecular genetic markers.

11.1. КОЛИЧЕСТВЕННАЯ ИЗМЕНЧИВОСТЬ И ЕЕ ЗАКОНЫ

Количественный признак (КП) это – признак, который имеет непрерывную шкалу изменчивости, в которой нельзя определить какие то дискретные классы. Однако мы можем измерить данный признак и определить пределы его изменчивости в значениях шкалы, которая задана тем или иным путем. Распределение КП можно объяснить теорией Гауса или **нормальным распределением**, следующего вида (рис. 11.1.1):

$$F(x) = j\,(x;\ \mu\sigma^2). \qquad (11.1)$$

Таким образом, нормальное распределение – это функция распределения плотности вероятности с двумя параметрами: μ и σ^2. μ и σ^2 – параметры функции $F(x)$, называются математическим ожиданием и дисперсией. Они могут быть оценены посредством вычисления средней и стандартного отклонения.

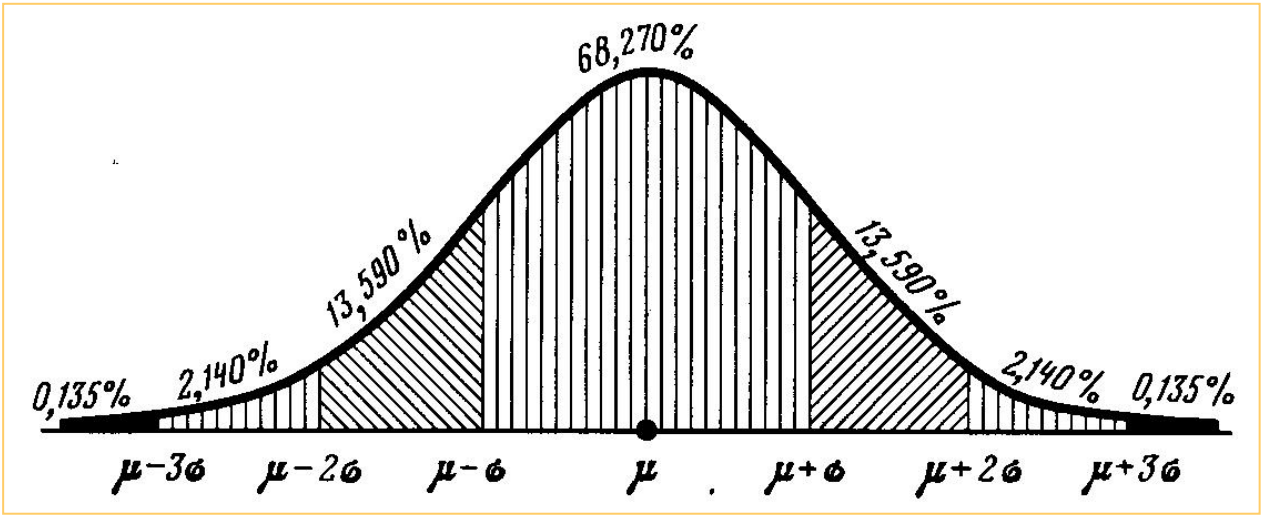


Рис. 11.1.1. Диаграмма, показывающая вид нормального распределения ***j (x; μs²)****. Цифры над кривой показывают площадь (в %) или вероятность попадания случайной непрерывной переменной в эти пределы или интервалы (По Животовский, 1991).*

Fig. 11.1.1. Diagram representation of the shape of the normal distribution ***j (x; μs²)****. Numbers denote a square (in %) or probability of putting a stochastic continuous variable in these limits or intervals (From Zhivotovsky, 1991).*

Нормальное распределение – классический закон математической статистики. Это распределение является важным по нескольким причинам. Одна из причин состоит в том, что это распределение является предельным случаем для многих других распределений. Например, с вычислительной точки зрения трудно задать биномиальное распределение для больших выборок. Аппроксимация нормальным распределением позволяет легко решить подобную задачу. Так, известно, что распределение оценки $p=k/N$ с хорошей точностью аппроксимируется нормальным распределением со средним значением, которое равно μ и с дисперсией (σ^2) - $\mu\,(1 - \mu\,) / N$. Вторая причина связана с выборочными оценками. Выборочные оценки средних значений для количественного признака, который может даже существенно отклоняться от нормального

распределения, описываются достаточно точно нормальным распределением, со средним, которое равно математическому ожиданию (генеральная средняя) и дисперсией σ^2 / N, где σ^2 – это генеральная дисперсия рассматриваемого признака. рис. 11.1.2 показывает распределение такого признака, как высота раковины (M_HGT) у брюхоногого моллюска из нескольких выборок в Японском море.

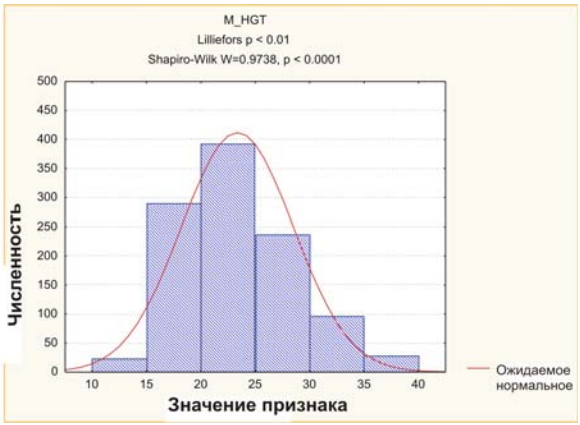


Рис. 11.1.2. Вариационный ряд для высоты раковины (M_HGT) в выборках брюхоногого моллюска Nucella heyseana из залива Петра Великого Японского моря.

Гистограмма показывает эмпирическую изменчивость, а кривая аппроксимацию нормальным распределением. Визуально аппроксимация вполне хорошая, хотя строгие критерии в данном случае показывают статистически значимые отклонения эмпирического и теоретического распределений (см. W=0.97, p < 0.0001; расчеты на основе данных Kartavtsev et al., 2004).

Fig. 11.1.2. Distribution of a gastropod mollusk Nucella heyseana shell height (M_HGT) in samples from Peter the Great Bay, Japan Sea.

Histogram shows the empirical raw variation and the curve presents normal distribution approximation. There is quite fine correspondence of two distributions, although statistically significant deviation observed for them (Based on data from Kartavtsev et al., 2003).

Истинные значения μ и σ^2 в генеральной совокупности не известны. О них можно судить только на основании случайных выборок из генеральной совокупности, для которых можно оценить значения средней ($\bar{x}$ или M) и среднего квадрата отклонений (среднего квадрата, для краткости, MS или s_x^2) или стандартного отклонения (s_x), которые являются оценивающими статистиками для двух параметров. Мы будем называть их оценками (estimates), хотя оценка в точном смысле – это только численное значение, скажем, 2.3 и 5.6, для $\bar{x}$ или s_x.

Средняя и стандартное отклонение

Средняя может быть определена следующим образом:

$$\bar{x} = (\sum x_i)/n = 1/n\ (\sum x_i). \qquad (11.2)$$

Здесь x_i - это значение индивидуального измерения у растения, животного или любого другого организма, а n – это число измерений. Средний квадрат и стандартное отклонение для серии таких измерений равны:

$$MS = \sum (x_i - \bar{x})^2/(n-1). \qquad s_x = \sqrt{[\sum (x_i - \bar{x})^2/(n-1)]}. \qquad (11.3)$$

Несколько других важных статистик также используются в генетических исследованиях КП. Одна из наиболее часто используемых статистик – это коэффициент вариации, Cv.

$$Cv = (s_x / \bar{x}) * 100\%. \qquad (11.4)$$

Этот коэффициент стандартизирует оценки дисперсии для различных признаков, делением на их средние значения и делает величины изменчивости более сравнимыми. Мы познакомимся с

некоторыми другими способами стандартизации и преобразованиями признаков в ходе Практического занятия №11. Очень важной также является другая статистика, нормализованное отклонение (t): $t = (x_i - \bar{x})/ s_x$. Во-первых, это выражение ведет к предсказанию значений x_i на основе равенства: $(x_i - \bar{x}) = t*s_x$ или $x_i = \bar{x} + t*s_x$. Во-вторых, все признаки после t-преобразования имеют вариационные ряды, в которых средняя равна нулю, а дисперсия равна единице: $\bar{x} = 0$ и $s_x = 1$. Это свойство широко используется при многомерном анализе, так как делает все переменные сравнимыми.

Доверительный интервал и доверительная вероятность

Оценка средней $\bar{x}$ называется точечной оценкой, потому что она имеет конкретное численное значение для определенной выборки, или может быть представлена точкой на числовой оси. Вследствие случайности выборки и ее ограниченного размера эта точка не будет совпадать с μ. Поэтому желательно определить такой интервал значений переменной, в который μ будет попадать. Как найти этот интервал? Мы можем знать функцию, скажем это $F(x)$, t, χ^2 или F. Хотя сложность здесь в том, что мы не знаем значение параметра μ для вычисления вероятностей и определения пределов изменчивости. Однако, чтобы избежать данного затруднения есть простой способ. Можно заменить μ выборочной оценкой - $\bar{x}$ и судить о генеральной совокупности по выборке. Соответствующее распределение будет подобно теоретическому распределению, хотя оно будет смещено от μ к $\bar{x}$ (рис. 11.1.3, А). Внутри этого распределения можно найти такой интервал значений I_1, I_2, что суммарная площадь под кривой и вне этого предела будет равной α ($\alpha/2$ с обеих сторон); площадь внутри интервала есть $P = 1 - \alpha$. А отрезок I_1, I_2 называется **доверительным интервалом** для оцениваемого параметра μ с **доверительной вероятностью** $P = 1 - \alpha$. Отрезок I_1, I_2 покрывает 100% доверительного интервала. Так, например, 95% доверительного интервала соответствует P = 0.95, 99% доверительного интервала соответствует P = 0.99, а 99.9% доверительного интервала соответствует P = 0.999.

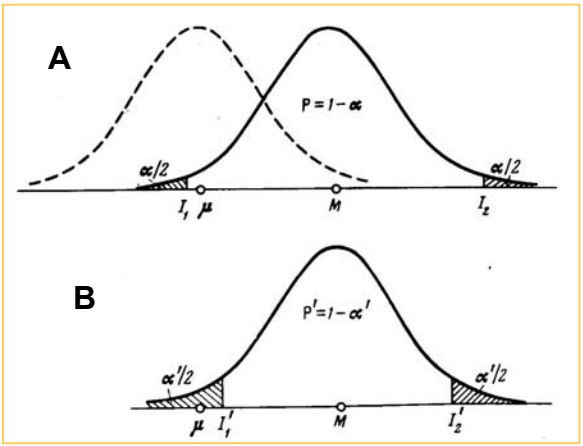


Рис. 11.1.3. Диаграмма, иллюстрирующая доверительный интервал и доверительную вероятность.

А. *Пунктирной линией показано теоретическое распределение оценочных значений. Сплошная линия – это тоже самое распределение, в котором «центр» представляет выборочную оценку средней* $\bar{x}$. I_1, I_2 *– это доверительный интервал, который с доверительной вероятностью* P *включает* μ, *генеральный параметр.* ***В.*** *Тоже самое распределение, но доверительный интервал* I'_1, I'_2 *показан с более низкими доверительными вероятностями,* $P' < P$ ($\alpha' > \alpha$) *и который уже не включает* μ, *значение генерального параметра.*

Fig. 11.1.3. Diagram for a confidence limit and a confidence probability. ***A.*** *With dotted line is shown a theoretical distribution estimates' values. Solid line is the same distribution, in which as a "center" taken sample estimates of* M *(or* $\bar{x}$*).* I_1, I_2 *is a confidence limit, which with a confidence probability* P *includes* μ, *a general parameter.* ***B.*** *The same distribution, but the confidence limit* I_1, I_2 *shown with lower confidence probability,* $P' < P$ ($\alpha' > \alpha$) *and no longer includes* μ, *a general parameter value.*

Стандартная ошибка статистик

Необходимо также знать стандартную ошибку (standard error, SE) средней ($s_{\underline{x}}$) и стандартного отклонения. ***SE*** для средней равна $s_{\underline{x}} = s_x / \sqrt{n}$. Здесь мы имеем дело со средней арифметической, а SE необходима для ее сравнения**.** ***SE*** для стандартного отклонения и ***Cv*** одинакова; она равна $SE = s_x / \sqrt{2n}$. В этом случае знание SE необходимо для сопоставления оценок s_x и ***Cv.*** Для других статистик аналогично - получение ошибок облегчает сравнение их оценок.

Анализ количественных признаков

Главная задача анализа – понять, имеется ли различие какого-либо измеренного признака между парой линий или популяций или между их совокупностями. Чтобы выполнить такое задание имеются несколько возможностей. Мы воспользуемся пакетом программ STATISTICA, чтобы оценить различия. С помощью современных программных средств можно сопоставить множество свойств распределений признаков. Поэтому важно понимать главное, как и для чего выполняются сравнения. Простейший случай – это когда сравниваются два средних значения или пара стандартных отклонений. Здесь нам в первую очередь необходимо получить их SE. Далее мы можем сопоставить две величины, скажем средние, на основе нуль-гипотезы случайности найденных различий. Если различия больше, чем ожидается, то нулевая гипотеза о случайности различий (или отсутствия различий) отвергается и можно говорить о двух (или более) различных линиях, популяциях и т.п. Главный инструмент для такого суждения – это t-статистика Стьюдента. Например, для выборок 1 и 2 фактическое значение ***t*** –статистики для средних арифметических есть:

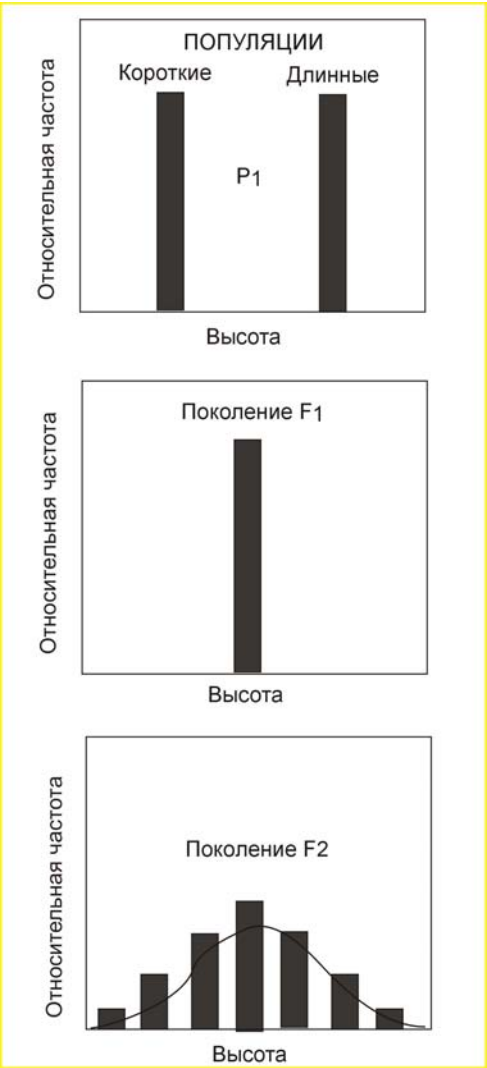
$$t = (M_1 - M_2) / \sqrt{(SE_1^2 + SE_2^2)}.$$

Также, как в случае хи-квадрат, фактическое значение ***t*** сравнивается с ожидаемым ***t*** (t-таблица): d.f. = $n_1 + n_2 - 2$; ***M*** - средняя, SE_1 и SE_2 – стандартные ошибки для выборок 1 и 2, d.f. – число степеней свободы (degree of freedom), а n_1, n_2 – размер каждой из двух выборок. Необходимо помнить, что представленная выше формула справедлива, когда размер выборок примерно равен и они достаточно велики по размеру. Если это не так, то необходимо использовать другие алгоритмы расчетов.

Принцип дисперсионного анализа, применяемого для выявления различий между несколькими выборками или линиями, также прост. Мы рассмотрим его ниже в Разделе 11.3.

11.2. ГИПОТЕЗА МНОЖЕСТВЕННЫХ ФАКТОРОВ

В начале 20-го столетия генетики заметили, что многие признаки у различных видов имеют сходный характер наследования. Это относится к таким признакам, как рост и вес человека, размер семян у различных бобовых растений, цвет зерен у пшеницы и другие. В каждом случае потомство в последовательных поколениях оказывалось как бы «смесью» значений родительских признаков. Вопрос, может ли изменчивость КП быть объяснена в терминах Менделевской генетики, породил противоречивую дискуссию в начале 19-го века. Бэтсон (Batson) и Юл (Yule), которые были сторонниками Менделевского объяснения наследования КП, предположили, что в их детерминации участвует большое число факторов (генов). Эта идея, названная **гипотезой множественных факторов**, предполагала, что в фенотип вносят вклад множество локусов и этот вклад носит кумулятивный или количественный характер. Однако другие генетики сомневались в том, что Менделевские единичные факторы могут играть роль в смешивании родительских фенотипов в КП и в таком характере наследования.



Мы можем наблюдать промежуточные значения по КП в F_1 и расщепление в F_2, которое ведет к большей изменчивости признака в этом поколении, чем в F_1 (рис. 11.2.1).

Рис. 11.2.1. Гистограмма, показывающая относительную частоту особей, проявляющих два различных фенотипа по высоте стебля и построенная по данным скрещивания Кёлрейтера (Kolreiter) между двумя сортами табака: «короткий» и «длинный» (По Klug, Cummings, 2002).
Fig. 11.2.1. Histogram showing the relative frequency of individuals expressing various height of phenotypes derived from Kolreiter's cross between two sorts of tobacco: dwarf and tall (From Klug, Cummings, 2002).

Результаты, представленные на рис. 11.2.1, являются несколько упрощенными, в том смысле, что не показывают вообще какой либо изменчивости в F_1. Более наглядными являются результаты, полученные Истом (East, 1936; рис. 11.2.2).

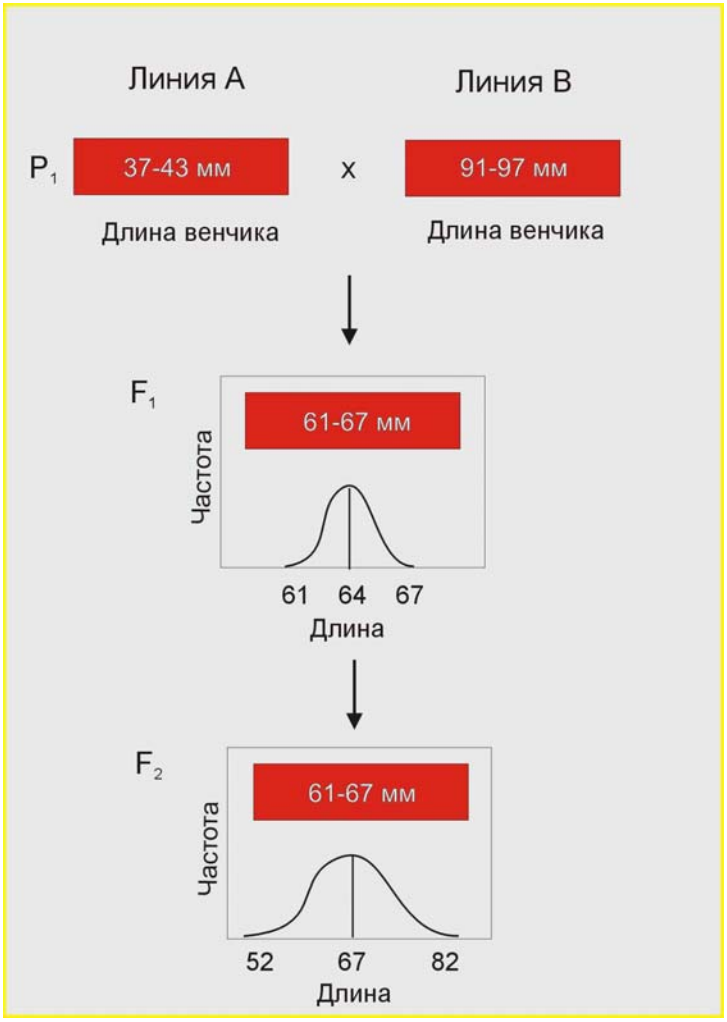


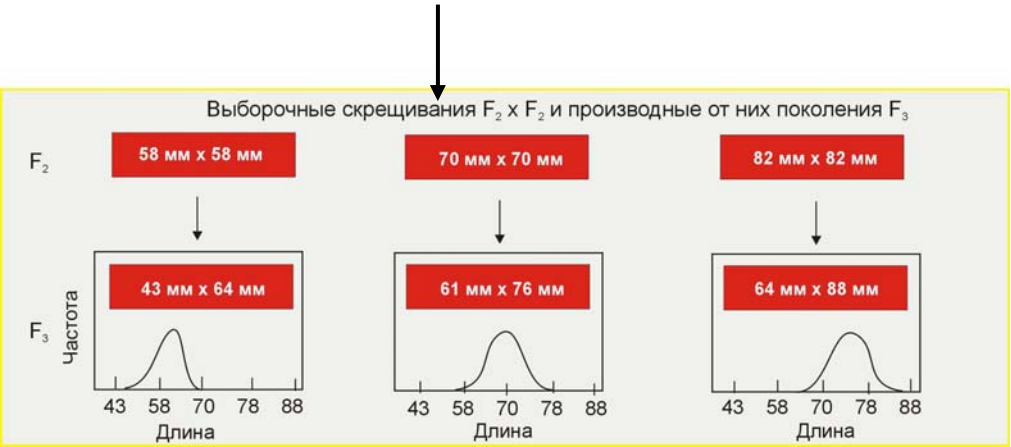
Рис. 11.2.2. Схема скрещивания двух линий и расщепления в потомстве F1, F2 и отобранном F3.

На схеме показан опыт Иста по скрещиванию двух линий табака с различной длиной венчика. Этот признак у растений линии А варьирует от 37 до 43 мм, тогда как у растений линии В варьирует в диапазоне 91-97 мм (По Klug, Cummings, 2002).

Fig. 11.2.2. The F1, F2 and selected F3 results of East's cross between two strains of Nicotiana with different corolla lengths. Plants of strain A vary from 37 to 43 mm, while plants of strain B vary from 91 to 97 mm (From Klug, Cummings, 2002).

Эксперименты Иста показывают, что хотя изменчивость в длине венчика является количественной, результаты экспериментального скрещивания демонстрируют расщепление отчетливых фенотипических классов, как обнаружено в трех группах F_3.

Это ключевое наблюдение стало основой гипотезы множественных факторов, которая объясняет, как признаки могут значительно различаться по экспрессии.



Аддитивные аллели: основа непрерывной изменчивости

Гипотеза множественных факторов, поддержанная наблюдениями Иста и других, включает следующие основные положения:

1. КП, который проявляет непрерывную изменчивость, обычно может быть определен посредством измерения, взвешивания, подсчета и т.д.

2. На наследование КП аддитивно оказывают влияние два и или более генов, локализованных в различных местах генома. Поскольку в детерминацию КП может быть вовлечено много генов, то наследование этого типа часто называют **полигенным**.
3. В каждом локусе может находится один из **аддитивных аллелей**, которые вносят свой вклад в суммарное фенотипическое значение, или **не аддитивный аллель**, который не вносит суммирующего вклада в фенотип.
4. Эффект каждого аддитивного аллеля на фенотип невелик и примерно эквивалентен вкладам отдельных аддитивных аллелей по другим генным локусам.
5. Все вместе гены, контролирующие отдельный КП определяют существенную фенотипическую изменчивость.
6. Анализ полигенных признаков требует исследования большого числа потомков из линии или популяции организма.

Сформулированные выше положения фокусируются вокруг концепции аддитивных аллелей по полигенным локусам, контролирующим КП. Чтобы проиллюстрировать это, давайте рассмотрим эксперимент Хермана-Нильсена-Элли (Herman-Nilsson-Ehle) по определению наследования цвета у семян пшеницы, который был проведен в начале 20-ого века. В одной из серий экспериментов пшеница с красными зернами была скрещена с линией, имеющей белые зерна (рис. 11.2.3).

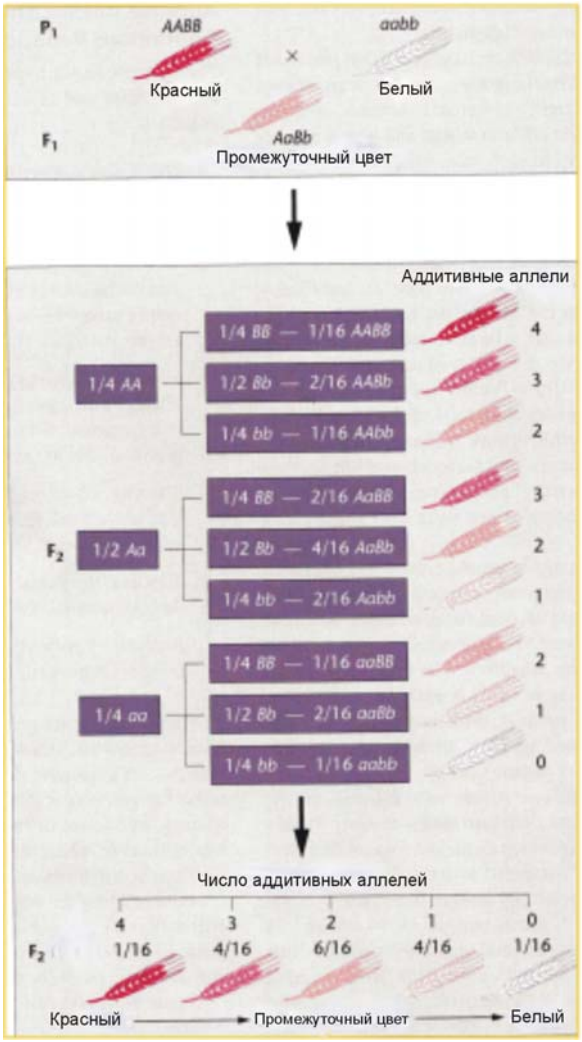


Рис. 11.2.3. Схема, иллюстрирующая, как гипотеза множественных факторов и аддитивные аллели обеспечивают расщепление по фенотипу в соотношении 1:4:6:4:1 для цвета зерен пшеницы.

Заглавные буквы обозначают аддитивные аллели, которые вносят равный вклад в пигментацию зерен или в фенотип растения (По Klug, Cummings, 2002).

Fig. 11.2.3. Scheme that shows how the multiple-factor hypothesis accounts for the 1:4:6:4:1 phenotypic ratio of grain color in wheat.

All alleles designated by uppercase letters are additive and contribute an equal amount of pigment to the grain or phenotype (From Klug, Cummings, 2002).

Наследование цвета происходит в зависимости от «дозы» аддитивных аллелей и следует соотношению 1:4:6:4:1. В исходном скрещивании предполагалось, что в контроле цвета зерен участвуют два гена, которые, расщепляясь в F_2, дают пять фенотипических групп, ожидаемые соотношения для которых равны 1:4:6:4:1. Поэтому, как мы видим, имеются веские основания принять, что количественная изменчивость может быть объяснена на основе Менделевских принципов. Нет ни каких причин, для того чтобы три, четыре или более генов не могли бы контролировать фенотип количественного признака. С увеличением числа вовлеченных генов, соответственно увеличивается сложность фенотипических расщеплений. Число фенотипов и ожидаемые их пропорции в F_2 при увеличении числа генов до пяти показаны на рис. 11.2.4.

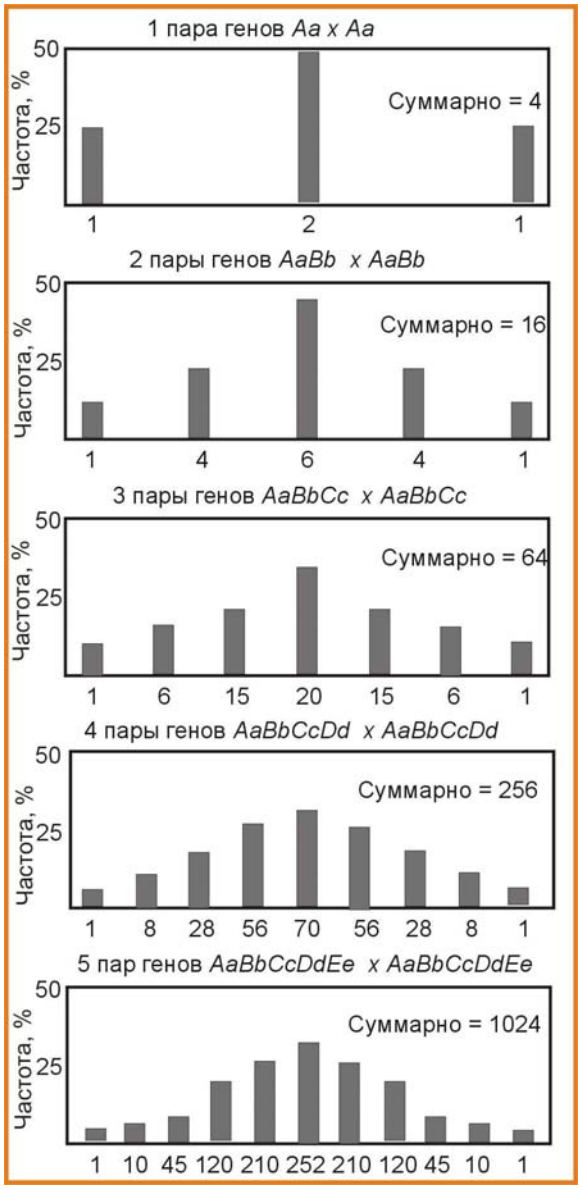


Рис. 11.2.4. Схематические результаты скрещивания двух гетерозигот при полигенном контроле КП и участии 1-5 локусов.

Каждый столбик гистограммы обозначает отчетливый фенотипический класс с одного края (левый конец) до другого края (правый конец). Каждый фенотип определяется различным числом аддитивных аллелей. Заметим: k = n + 1, k и n – число фенотипов и генов (По Klug, Cummings, 2002). Fig. 11.2.4. The results of two heterozygote cross when polygenic inheritance is operative with 1-5 gene pairs.

Each histogram bar indicates a distinct phenotypic class from one extreme (left end) to the other extreme (right end). Each phenotype results from a different number of additive alleles. Note: k = n + 1, k and n – number of phenotypes and genes (From Klug, Cummings, 2002)

В нашем предыдущем примере с пшеницей фенотипы родителей представляли два таких предельных класса – красные и белые зерна. На рис. 11.2.3 1/16 потомков F_2 являются либо красными, либо белыми, как их родители. В этом соотношении можно провести подстановку в правой части, так чтобы решить его по n: **$1/4^n = 1/16$, $4^n = 16$** или **n = Log416 = 2.** В общем случае **число генов, определяющих КП,** равно: ***n = Log4X***, где ***X*** – делитель в пропорции расщепления.

Значимость полигенного наследования

Полигенное наследование является важной концепцией. Оно служит генетической основой для огромного числа признаков, вовлеченных в селекцию животных и растений, в сельском хозяйстве и в аквакультуре. Такие признаки как длина, вес, форма тела у животных, размер и число семян у зерновых растений, молочность и выход мяса у сельскохозяйственных животных, яйценоскость у кур, плодовитость у рыб и беспозвоночных, - все они представляются, как находящиеся под полигенным контролем. Важно отметить, что наследуемый генотип устанавливает потенциальный спектр, в пределах которого реализуется конкретный фенотип. Однако факторы среды определяют, насколько данный генотип будет реализован. В скрещиваниях, описанных выше, мы предполагали, что организмы развивались в оптимальных условиях, которые минимизируют изменчивость, возникающую от различных источников. Разложить на компоненты данную изменчивость – это специальное задача генетики КП.

11.3. ГЛАВНЫЕ ЗАДАЧИ ГЕНЕТИЧЕСКОГО ИССЛЕДОВАНИЯ КОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ

Как наиболее важные, необходимо отметить три задачи: 1. Анализ характера распределения и изменчивости КП внутри и между выборками или селекционными линиями. 2. Оценивание

коэффициента наследуемости. 3. Определение числа генов, контролирующих КП. Последний вопрос мы не будем рассматривать здесь, так как рассмотрели его вкратце ранее, в конце предыдущего раздела 11.2.

1. Анализ характера распределения и изменчивости

Определение характера или вида распределения является важным, поскольку оно помогает применять соответствующее теоретическое распределение для того или иного исследуемого признака. Обычно распределение КП проверяется на его соответствие нормальному распределению, или более частным случаям, описываемым, например, распределениями Стьюдента и Фишера (***t-*** и ***F-***распределения). Так, на первом этапе анализа мы должны либо принять, либо отвергнуть нуль-гипотезу (H_n), что вариационный ряд и оценки его изменчивости статистически значимо отклоняются от нормального распределения: H_n >= P >= 0.95 (p <= 0.05) или выше (см. Раздел 11.1). Тестовые статистики для проверки H_n могут быть различными: хи-квадрат ($X^2_{набл.} < X^2_{ожид.}$), ***t-, F***-статистика и другие. На втором этапе анализа мы должны принять или отвергнуть H_n об однородности сравниваемых выборок. Это выполняется с использованием средств дисперсионного анализа, если сравниваются более чем две линии или выборки, или посредством ***t***-статистики, если анализируется пара выборок, как мы видели это ранее. Аналогично, мы решим несколько численных примеров в ходе Практического занятия 11, с использованием пакета STATISTICA. Кто-то может подумать, что понять дисперсионный анализ (ANOVA) трудно. В реальности же идея дисперсионного анализа проста для большинства примеров исследования, к примеру, природных популяций. Эта идея основывается на следующем равенстве:

$$\sigma^2_y = \sigma^2_A + \sigma^2_Z. \qquad \textbf{(11.5)}$$

Здесь σ^2_y – это суммарная дисперсия, σ^2_A – межгрупповая дисперсия и σ^2_Z – средняя внутригрупповая дисперсия. В генетическом анализе КП дисперсия может быть разложена далее на компоненты, в зависимости от сложности исследования и схемы скрещивания (мы увидим это ниже). На первом этапе генетики предпочитают представить суммарную дисперсию следующим образом:

$$\sigma^2_{PH} = \sigma^2_G + \sigma^2_E + \sigma^2_{GE}. \qquad \textbf{(11.6)}$$

Где σ^2_{PH} (=σ^2_y) – это суммарная или фенотипическая дисперсия, σ^2_G – генотипическая дисперсия, σ^2_E – дисперсия, обусловленная средой и s^2_{GE} – ковариационный член уравнения (обычно пренебрежимо мал и может быть опущен).

2. Определение коэффициента наследуемости

Получение оценки коэффициента наследуемости - это одна из самых важных задач в генетике КП. В целом несложно ввести коэффициент наследуемости (***h²***), после определения, того, что представляет собой суммарная дисперсия, а также ее компоненты. Это можно сделать исходя из 11.6. Таким образом, можно определить, что

Наследуемость в широком смысле (h^2_B) – это

$$h^2_B = \sigma^2_G / \sigma^2_{PH}. \qquad \textbf{(11.7)}$$

То есть, наследуемость КП – это доля генотипической дисперсии в общей изменчивости признака; h^2_B – это коэффициент наследуемости в широком смысле. Дисперсия σ^2_G является сложной по своей природе. Так, что $\sigma^2_G = \sigma^2_A + \sigma^2_D\ \sigma^2_I$, где σ^2_A - это **аддитивная дисперсия**, σ^2_D и σ^2_I – внутрилокусная и межлокусная дисперсии взаимодействия (первая - отклонение доминирования или отклонение от аддитивности аллельных эффектов, вторая – отклонения, обусловленные эпистатическими эффектами). Поэтому, можно сказать, что

Наследуемость в узком смысле (h^2_A) равна:

$$h^2_A = \sigma^2_A / \sigma^2_{PH} = \sigma^2_A / (\sigma^2_A + \sigma^2_E + \sigma^2_D\ \sigma^2_I). \qquad \textbf{(11.8)}$$

Влияние σ^2_I пренебрежимо мало и также может быть опущено. Как определить генотипическую дисперсию, σ^2_G и другие дисперсии? Это наиболее важный вопрос в генетическом анализе КП. Имеются специальные генетические приемы, которые позволяют разложить дисперсию на компоненты, и мы познакомимся с некоторыми из них ниже. До этого, приведем один простой пример, который полезен в понимании существа выявления σ^2_G. Представим себе, что имеется **клон**, - ***группа генетически идентичных организмов***. В пределах клона вся изменчивость определяется только средой, то есть, равна σ^2_E. Если мы проведем анализ многих клонов, то определим всю или суммарную изменчивость, σ^2_{PH}. Таким образом, мы можем вычислить генотипическую дисперсию, $\sigma^2_G = \sigma^2_{PH} - \sigma^2_E$. Такая логика может быть применима далее к различным генетическим уровням родственности: инбредным линиям, семьям (сибсы, полусибсы и т.п.). Полезным является также оценить результаты искусственного отбора.

11.4. НАСЛЕДУЕМОСТЬ

Искусственный отбор

Процесс подбора специальной группы организмов из исходно гетерогенной популяции для селекционных мероприятий называется **искусственным отбором**. Относительно высокие значения h^2 предполагают, что отбор по изменению популяции будет успешным. Как вы можете себе представить, основываясь на предыдущей информации, измерение компонент, необходимых для вычисления h^2 – это довольно сложное занятие. Простой подход включает оценку центральных тенденций (средних) признака на основе:

(1) родительской популяции *(M)*, проявляющей нормальное распределение,

(2) отбор части родительской популяции (*M1*), которая обладает наиболее подходящими значениями изучаемого количественного признака, и

(3) использование потомства *(M2),* возникшего от скрещивания отобранных *M1* групп.

Когда все это выполнено, то существуют следующие соотношения между средними и h^2: $M2 = M + h^2(M1 - M)$. Решение этого равенства для h^2 дает:

$h^2 = (M2 - M1) / (M1 - M)$. Это соотношение в свою очередь можно упростить, определяя *M2 – M1* как **ответ** *(R),* а *M1 – M* как **селекционный дифференциал** *(S)*. Здесь h^2 отражает долю фактического ответа на отбор по отношению к суммарному возможному ответу. То есть,

$$h^2 = R/S. \qquad (11.9)$$

Это так называемая **реализованная наследуемость**. Например: $h^2 = (13-20)/(10-20) = -7/-10 = 0.70$.

Самый длинный, из когда-либо проводимых селекционных экспериментов, это селекция в Государственной сельскохозяйственной лаборатории Иллинойса (США). Начиная с 1896 года, кукуруза отбиралась в двух направлениях, на увеличение и на снижение содержания масла. После 76 поколений искусственный отбор все еще дает увеличение содержания масла (рис. 11.4.1). Наследуемость некоторых хозяйственно важных КП у различных организмов суммирована в Табл. 11.4.1. Впечатляющие результаты ответа на отбор получены в лососеводстве (рис. 11.4.2).

Рис. 11.4.1. Ответ на искусственный отбор в линии кукурузы на высокое и низкое содержание масла, проводимый 76 поколений.

Числа в скобках над поколениями 9, 25, 52 и 76 для верхней кривой показывают вычисленные значения коэффициентов наследуемости в продолжающемся эксперименте. По прошествии 50 поколений коэффициенты наследуемости снизились в три раза, но остаются достаточно высокими, чтобы обеспечивать прирост выхода масла (По Klug, Cummings, 2002).

Fig. 11.4.1. Response of corn selected for high and low oil content over 76 generations. The numbers in parentheses at generations 9, 25, 52, and 76 for the "high oil" line indicate the calculation of heritability at these points in the continuing experiment (From Klug, Cummings, 2002).

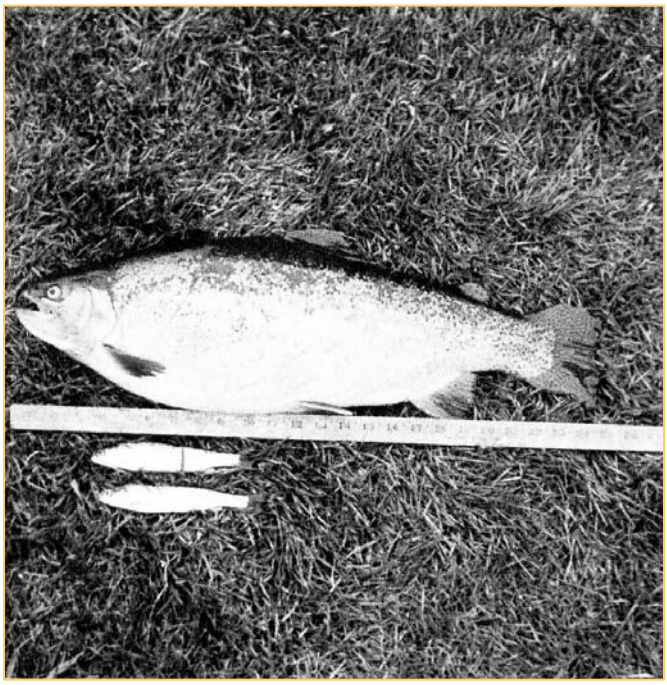


Рис. 11.4.2. Форель, отобранная на увеличение плодовитости Л. Дональдсоном (L. Donaldson).

Увеличение плодовитости составило от 3000 яиц для рыб в дикой популяции (внизу) до примерно 30000 яиц у самки “Сюзи” (вверху). Наследуемость не была определена (По Hines, 1976).

Fig. 11.4.2. Trout selected for fertility by L. Donaldson (Hines, 1976).

The increase is from 3000 eggs in wild fish (bottom) to some 30000 eggs in “Susie’s” body (top). Heritability was not detected.

Таблица 11.4.1. Оценки наследуемости у различных организмов (h^2, %)
(По Allendorf et al., 1987 и Klug, Cummings, 2002)

Table 11.4.2. Heritability estimates (h^2, %) *(From Allendorf et al., 1987, Klug, Cummings, 2002)*

Вид, признак	***h^2***
Человек	
Рост	65
Систолическое давление крови	48
Крупный рогатый скот	
Вес тела (взрослые)	65
Вес при рождении	51
Удойность	20-44
Жирность молока	40
Частота зачатия	3
Свинья	
Откладывание спинного сала	70
Прибавка веса в день	40
Домашняя птица	
Вес тела (в 32 недели)	50-55
Вес яиц (в 32 недели)	50
Яйценоскость	20
Выклевываемость из яиц	15
Овца	
Диаметр волоса	20-50
Вес тела	20-40
Мышь	
Длина хвоста (в 6 недель)	40-60
Вес тела (в 6 недель)	35-37
Численность помета	15
Drosophila melanogaster	
Вес тела	*40*
Длина торакса	*25-50*
Радужная форель	
Вес тела (молодь)	*12*
Вес тела (взрослые)	*17*
Длина тела (молодь)	*24*
Длина тела (взрослые)	*17*
Мясистость	*14*
Атлантический лосось	
Вес тела (молодь)	*8*
Вес тела (взрослые)	*36*
Длина тела (молодь)	*14*
Длина тела (взрослые)	*41*
Мясистость	*16*

Карп	
Вес тела (молодь)	*15*
Вес тела (взрослые)	*36*

Коэффициент наследуемости может варьировать в диапазоне: 0-1 (0-100%). В таблице, представленной выше, довольно много коэффициентов с величиной 0.4-0.5. Это умеренная наследуемость. Нередко наиболее важные в хозяйственном отношении признаки имеют низкие коэффициенты наследуемости, h^2 = 0.1-0.2 и менее.

Не следует думать, что все h^2, в приведенных таблицах, представляют собой реализованную наследуемость. Используются и многие другие методики. Очень частым является нахождение h^2 как корреляции (или регрессии) между родственниками: h^2 = r = b (известны данные для обоих родителей и потомства), h^2 = 2r = 2b (сибсы), h^2 = 4r = 4b (полусибсы). Один пример определения h^2 по этому способу дан ниже для бельдюги (рис. 11.4.3). Наиболее точным, но и наиболее сложным, является определение h^2 посредством разложения фенотипической дисперсии с помощью дисперсионного анализа, с предварительным скрещиванием производителей по специальной схеме. Мы разберем подобную схему в ходе Практического занятия 11.

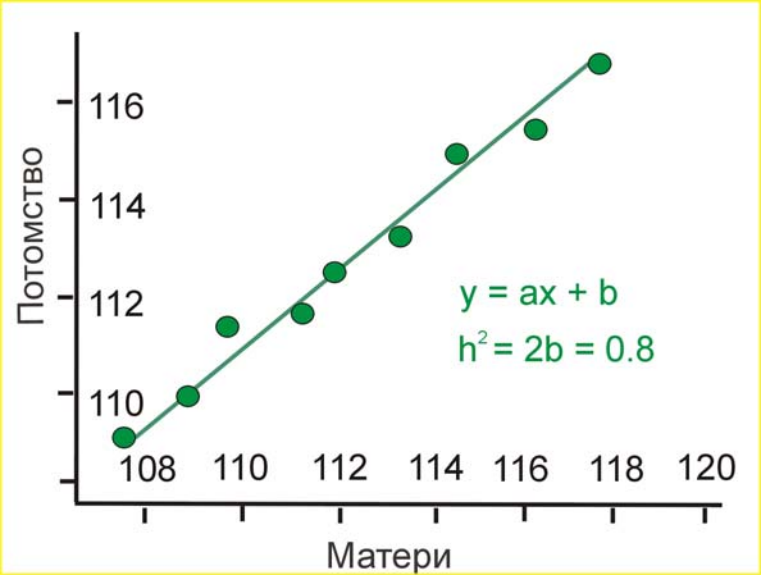


Рис. 11.4.3. Регрессия числа позвонков у бельдюги, Zoarces viviparus (По Smith, 1921).
Fig. 11.4.3. Regression of vertebrae number in eel-pout, Zoarces viviparus (After Smith, 1921).

11.5. КАРТИРОВАНИЕ ЛОКУСОВ КОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ (ЛКП). ЭВОЛЮЦИЯ ЛКП

Поскольку количественные признаки детерминируются множественными генами, перед генетиками стоит усложненная задача картирования этих генов в геноме. Расположены ли они на единственной хромосоме или «рассеяны» на многих хромосомах? Как мы знаем из курса общей генетики, локализация, или картирование генов в геноме – это один из важнейших шагов в направлении понимания генетической сущности организмов. Изначально необходимо локализовать гены, контролирующие количественные признаки, на конкретной хромосоме или хромосомах. В контексте данного раздела, эти гены будут называться **локусы количественных признаков** (ЛКП).

В качестве примера такого типа анализа, рассмотрим фенотипический признак устойчивости дрозофилы к ДДТ, который, как было установлено, находится под полигенным

контролем. Чтобы найти ответственные локусы, линии, отобранные на устойчивость к ДДТ, и линии, отобранные на увеличение чувствительности к нему, были скрещены с мушками, которые несут доминантные маркеры на каждой из четырех хромосом вида. В соответствии с различными скрещиваниями, было получено потомство, которое содержит много различных комбинаций маркерных хромосом и хромосом из обеих линий, устойчивых и чувствительных. Как показано на рис. 11.5.1, мушки с различными хромосомными комбинациями были далее протестированы на устойчивость (выживаемость) при экспозиции с ДДТ. Результаты показывают, что каждая хромосома дрозофилы содержит гены, которые вносят вклад в устойчивость к ДДТ. Другими словами, локусы несущие гены устойчивости к ДДТ рассеяны в геноме.

Рис. 11.5.1. Выживаемость дрозофил, несущих комбинации хромосом в ДДТ-устойчивых и ДДТ-чувствительных (контроль) линиях, при их обработке ДДТ.

Результаты показывают, что устойчивость к ДДТ является полигенной, с локализацией генов, обеспечивающих устойчивость, на каждой из основных хромосом. Хромосома 4 несет очень мало генов у дрозофилы и была опущена из этого анализа (По Klug, Cummings, 2002).

Fig. 11.5.1. Survival rates of Drosophila carrying combinations of chromosomes from DDT-resistant and DDT-sensitive (control) strains when exposed to DDT. The results indicate that DDT resistance is polygenic, with genes on each of the major chromosomes making a major contribution. Chromosome 4 carries only a few Drosophila genes and was omitted from this analysis (From Klug, Cummings, 2002).

Мы можем теперь картировать положение генов КП более специфично вдоль хромосом дрозофилы, поскольку появились молекулярные маркеры на каждой из этих хромосом. Расположение ЛКП определяется относительно известной локализации этих маркеров на хромосомах. Эти маркеры, к примеру, могут определяться сайтами узнавания различных специфических рестриктаз ДНК. В результате мутаций в таких сайтах возникает полиморфизм длины рестрикционных фрагментов (ПДРФ; restriction fragment length polymorphisms - RFLPs). Появляются также и другие маркеры (см. Лекцию 13).

Этот подход уточняет число и картирует сами локусы, ответственные за количественные признаки. ПДРФ маркеры в настоящее время разработаны для многих организмов, значимых для сельского хозяйства, что позволяет проводить планомерное картирование ЛКП. Например, сотни ПДРФ маркеров были локализованы у томатов. Они расположены вдоль 12 хромосом этого организма. Когда исследуемый фенотипический признак и маркер экспрессируются вместе, то говорят, что они **косегрегируют**. Существенная косегрегация устанавливает присутствие ЛКП в месте расположения ПДРФ или рядом на хромосоме. В случае, если ПДРФ маркер и ЛКП, ответственные за исследуемый признак, тесно сцеплены в одной хромосоме, то они будут с высокой вероятностью демонстрировать ассоциацию в родословных, нежели в случаях, когда они не сцеплены. Когда ЛКП локализованы, то далее строится карта идентифицированных генов и их местоположение в районах ПДРФ или других маркеров. Такие карты помогают лучше планировать селекционные программы.

Коррелированный ответ и генетические корреляции

Гены проявляют плейотропные эффекты на фенотип. То есть, каждый ген потенциально оказывает воздействие на несколько признаков организма, либо прямо, либо опосредованно. Поэтому, аллель, который благоприятен для одного КП, может иметь неблагоприятное воздействие на другой КП. По мере того, как эти аллели будут увеличиваться в частоте вследствие искусственного отбора, они будут вести, с одной стороны, к улучшению одного признака, но в тоже время, могут ухудшать другие признаки фенотипа. Плейотропия – это одна из причин **коррелированного ответа**, когда изменение фенотипического значения по одному признаку сопровождается также изменением по другому признаку(ам). Вторая возможная причина коррелированного ответа – это неравновесие по сцеплению. В результате чего пара (или более) близко расположенных генов вовлекаются в коррелированый ответ, хотя селекция ведется только по одному из них. В следующей лекции (12) мы также увидим, как различные внутрилокусные взаимодействия вызывают положительные и отрицательные корреляции с КП при объединении единичных локусов в мультилокусный генотип.

Коррелированный ответ вполне обычен при искусственном отборе и часто, но не всегда, приводит к ухудшению приспособленности. Например, в случае кур породы Леггорн (Leghorn) 12 поколений отбора на увеличение длины голени, снизили вылупляемость цыплят из яиц почти в два раза (Lerner, 1958). Выживаемость в линии форели Дональдсона, о которой упоминалось выше, вне строго регулируемого температурного режима, столь низка, что не приходится говорить о каком либо ее использовании, например, в поликультурных хозяйствах, с не очень жесткой регуляцией. Строго говоря, любая селекция ведет к созданию линии (породы), приспособленной к конкретным «селекционным» условиям. Вне этих условий высокие показатели данной породы или сорта не будут реализованы; они могут быть даже хуже, чем у особей, не подвергавшихся селекции (Жученко, 1980).

Сложность селекционного ответа хорошо иллюстрируется селекцией на прирост веса у крысы. Эксперимент по искусственному отбору проводился в 23 поколениях. В ходе эксперимента измерялся прирост веса между третьей и пятой неделями после рождения (Baker et al., 1975). Для того чтобы оценить коррелированный ответ, шесть линий крысы были охарактеризованы по 17 признакам черепа: длина черепа, ширина черепа, межорбитальное расстояние и т.д. (Atchley et al., 1982). Эксперимент включал две контрольные группы, две линии, в которых проводился отбор на прирост веса и две линии, отбираемые на потерю веса. Все линии происходили из одной общей группы. Прямой ответ был примечателен по величине асимметрии: самцы, отбираемые на больший вес, были на 3.46 стандартных отклонения больше, чем контроль, тогда как самцы, отбираемые на потерю, веса были на 1.30 стандартного отклонения меньше контроля. Аналогично, самки, отбираемые на увеличение веса, были на 1.46 стандартных отклонения больше, а самки, отбираемые на потерю веса, на 1.02 стандартных отклонения меньше, чем контрольные животные. Асимметричный ответ обычно относят на действие противоположно направленного естественного отбора, и каких-то специфических механизмов здесь не может быть предложено.

Коррелированные ответы показали сходные диапазоны асимметрии. Однако, тогда как линии-реплики показывали согласие величин прямых ответов, 35 из 51 сравнений между самцами из этих линий-реплик были различны в этом коррелированном ответе. Невоспроизводимость коррелированного ответа может быть обусловлена большей чувствительностью генетических коварианс к пертурбационным эффектам отбора и дрейфа. Проведенный анализ (Atchley et al., 1982) базировался на скоррелированных признаках и найденных канонических переменных (вычисляемых как линейные комбинации признаков по степени увеличения их дисперсий). Сравнивая наследуемость по индивидуальным признакам и наследуемость по каноническим переменным, цитированные авторы обнаружили, что для последних переменных коэффициенты наследуемости в узком смысле были более высокими. Эти факты подчеркивают произвольность измерения признаков и выведенного из этого генного действия, поскольку *характер генетической*

корреляции предполагает, что генетическая изменчивость затрагивает комплексы признаков в коррелированной манере. Полученные данные заставляют думать, что гены действуют не на отдельные признаки, такие как длина скулы и ширина челюсти, а на их комплекс или комплексный прообраз. Сходные идеи развивал Животовский (1984).

Отбор по многомерным признакам. Отбор по комплексам признаков, который может вызывать и эволюционные изменения, очевиден из данных по крысам, представленным выше, а также по голосеменным растениям и другим организмам (Животовский, 1984). В тоже время, очевидно, что разложение компонент дисперсии по комплексу признаков может быть безнадежно сложным. Действительно, как таковой не существует теории контроля количественного признака. Распределение Гауса – это статистическая аппроксимация и только. Однако теоретическая проблема становится более достижимой при принятии некоторых допущений. Наиболее важным является то, что многомерные распределения фенотипов и генетические эффекты предполагают выполнение многомерного нормального распределения. Если исходное распределение является многомерным нормальным, а отбор благоприятствует предпочтительному воспроизводству только определенных генотипов, то тогда фактическое распределение должно отклоняться от нормального как следствие отбора. Хотя это и верно, но если отбор достаточно слаб, то фенотипическое распределение будет оставаться примерно нормальным, по меньшей мере на коротких временных отрезках. Некоторые другие подходы к созданию генетической теории количественного признака обсуждаются Жученко (1980).

Эволюция количественных признаков: мутационно-селекционный баланс

Направленный отбор, как мы убедились в предыдущих разделах, сильно изменяет признаки и формы (породы) организмов. Со времен Дарвина в целом признается, что в дикой природе направленный естественный отбор также является доминирующим фактором эволюции. Однако в природе действует не только эта форма отбора. Стабилизирующий (очищающий) отбор представляет большой интерес, в связи с той ролью, которую он может играть в поддержании наследуемой изменчивости мультифакторных признаков. Интуитивно кажется резонным допустить, что обнаруживаемый уровень аддитивной генетической изменчивости может исходить от баланса между стабилизирующим отбором, который стремится уменьшить генетическую изменчивость, и новыми мутациями, которые стремятся увеличить ее. Очень простая, когда формулируется устно, модель становится сложной, если предпринимается попытка построить точную математическую модель или теорию. Осложнения возникают из-за необходимости оперировать точными знаниями о числе детерминирующих генов, знаниями о типе внутрилокусного действия аллелей и взаимодействия между локусами, эффектов плейотропии, сцепления и, наконец, направления и интенсивности отбора. Плейотропия выражается в широко распространенной тенденции затрагивать несколько признаков, включая, казалось бы, и не связанные признаки. Особенно это проявляется наглядно при реализации генетического контроля комплексных фенотипических признаков, которые возникают при последовательном взаимодействии многих генов в ходе индивидуального развития.

Число генов, определяющих признак, зависит от селекционно-мутационного баланса поэтому, как было показано (Crow, Kimura, 1970), для данной генетической дисперсии интенсивность отбора на локус уменьшается с увеличением числа вовлеченных локусов. Если суммарный темп мутаций постоянен, то скорость мутаций на локус при приближении к равновесию должна снижаться с увеличением числа локусов. Для популяции, находящейся в селекционно-мутационном равновесии, интенсивность отбора на локус оценивается как величина уже приблизительно независимая от числа локусов (Turelli, 1984). Сложная, как сама ситуация, гипотеза о том, что аддитивная генетическая изменчивость может поддерживаться стабилизирующим отбором, сбалансированным против мутаций, была поддержана математически новыми аппроксимациями в моделях, исследованных Ланде (Lande, 1975; 1977; 1979; 1980).

С другой стороны, аппроксимации оказываются достаточно чувствительными к конкретным генетическим и математическим допущениям, которые приняты при их разработке. Простейшие модели игнорируют доминирование и эпистаз, и они допускают большое число аллелей на локус и распределение эффектов, аппроксимирующихся нормальным распределением. При таких упрощенных допущениях решение может быть получено довольно легко (Kimura, 1965; Lande, 1975; Turelli, 1984). Развернутое изложение этой проблематики можно найти в другой литературе (Hartl, Clark, 1989, p.502).

11.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 11

1. Рассчитайте реализованную h^2 для рыбы тиляпии (см. Табл. 11.6.1, см. вопросы).
2. Решите следующие три задачи по определению наследуемости КП: примеры1-3 (Klug, Cummings, 2002, P. 117, Problems and Discussion Questions).
3. Ознакомление с пакетом программ STATISTICA. Решите один пример для расчета средней арифметической (***M***), стандартного отклонения и стандартной ошибки для ***M***, используя пакет программ STATISTICA.
4. Применение дисперсионного анализа (ANOVA) для оценки h^2. Рассмотрите основные схемы скрещивания: (1) иерархическую (2) полиаллельную (рис. 11.6.1). Ознакомление с нахождением варианс в диаллельной 2x2 факторной схеме ANOVA.
5. Рассмотрите модель направленного отбора по КП в пакете программ POPULUS: Evolutionary Simulations> Quantitative Genetics> Directional Selection at Quantitative Character.

Таблица 11.6.1. Определение h_B^2 у рыбы тиляпия в эксперименте Тхиеу (C.M. Thieu)
(По Кирпичников, 1979)
Table 11.6.1. Determination of h_b^2 in fish tilapia through selection by C.M. Thieu
(After Kirpitchnikov, 1979)

ПОЛ	*P*			*F1*			h^2 = R/ S
	Маленькие	Средние	Большие	Маленькие	Средние	Большие	
Самки	15.7	20.2	24.8	11.8	12.9	12.9	$h^2_{L/S}$= 0.12 $h^2_{L/M}$= ? $h^2_{M/S}$= ?
Самцы	16.8	23.8	28.4	19.3	20.3	20.7	$h^2_{L/S}$= 0.12 $h^2_{L/M}$= ? $h^2_{M/S}$= ?

Схемы скрещиваний. Чтобы выполнить ANOVA и вычислить компоненты дисперсий для определения коэффициента наследуемости необходимо предварительно поставить специальное скрещивание и получить потомство (рис. 11.6.1).

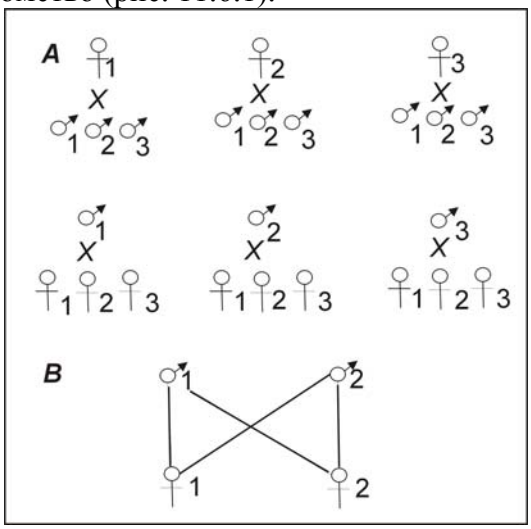


Рис. 11.6.1. Схема скрещивания: А – иерархический комплекс (два варианта) и В – диаллельное скрещивание.
Fig. 11.6.1. Scheme of crosses: A, Hierarchical complex (two variants) and B, Diallelic cross.

ГЛАВА 12. ГЕНЕТИЧЕСКИЙ ГОМЕОСТАЗ, ГЕТЕРОЗИС И ГЕТЕРОЗИГОТНОСТЬ

ГЛАВНЫЕ ВОПРОСЫ:

12.1. Концепция генетического гомеостаза.
12.2. Гетерозис и гетерозиготность.
12.3. Эмпирические результаты и их интерпретация.
12.4. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 12.

РЕЗЮМЕ

1. **Генетический гомеостаз**, – свойство популяции балансировать свой генетический состав. Оно пришло в генетику из физиологии. Понятие гомеостаза многогранно и относится, в том числе, к онтогенезу.
2. Сущность **гетерозиса** и **гетерозиготности** различна, хотя эти два понятия перекрываются. Между гетерозиготностью и гетерозисом может существовать положительная взаимосвязь, но первая не всегда является причиной второго.
3. Связь между гетерозиготностью и приспособительным признаком может быть объяснена как с позиций селективной нейтральности, так и с позиций селективной значимости маркерных локусов. Наиболее признанными сейчас являются **гипотезы инбредной депрессии**, предполагающей нейтральность маркеров, и **балансовая гипотеза**, которая в принципе является адаптивной. Обе эти гипотезы не вполне удовлетворяют эмпирическим данным. Нужна новая модель, которая, все же, должна разрабатываться на основе балансовой гипотезы.
4. Новая модель должна включать механизм балансирующего и релаксирующего естественного отбора, двояко действующего в ходе жизненного цикла: (1) как регулятор оптимальной гетерозиготности в популяции (поддержка стабильного сегрегационный груза) и (2) как регулятор фенотипического гомеостаза по количественным признакам.

Chapter 12. GENETIC HOMEOSTASIS, HETEROSIS AND HETEROZYGOSITY

SUMMARY

1. **Genetic homeostasis** is a feature of a population to balance its genetic composition; it came in genetics from physiology. A concept of homeostasis is manysided and can be attributed specifically to ontogenesis.
2. The essence of **heterosis** and **heterozygosity** is different; however, two notions are overlap. A positive relationship between heterozygosity and heterosis may exist but first is not always a reason for the second.
3. A relationship between heterozygosity and fitness trait can be explained both from the position of selective neutrality and from the position of selective significance of marker loci. The most recognized now are **hypothesis of inbred depression**, which suggests a neutrality of gene markers, and **balance hypothesis**, which is in principle an adaptive one. Both these hypotheses do not quite comply with empirical data. There is a need for a new model, which should be developed presumably on the basis of the balance hypothesis.
4. The new model should include a mechanism of balancing and relaxing selection, which acts bidirectionally during the life cycle, (1) as a regulator of the optimal heterozygosity in a population (support for a stable segregation load) and (2) as a regulator of a phenotypic homeostasis for **quantitative traits**.

12.1. КОНЦЕПЦИЯ ГЕНЕТИЧЕСКОГО ГОМЕОСТАЗА

Концепция гомеостаза появилась изначально в области физиологии (Canon, 1932). В целом можно сказать, что *«гомеостаз относится к свойству организма регулировать себя в зависимости от варьирующих условий или иными словами гомеостаз – это саморегулирующиеся механизмы организма, которые позволяют ему стабилизировать себя во флуктуирующей внутренней и внешней среде»* (Lerner, 1970). Гомеостаз может относится к различным системам: к человеческому обществу, к некоторым свойствам физических систем, химических соединений (буферов) и т.д. Поэтому, в определенном смысле можно говорить об экологическом гомеостазе, а генетический гомеостаз – это часть системных свойств живых организмов. ***Генетический гомеостаз*** *(ГГ) может быть определен, как свойство популяции балансировать свой генетический состав и противостоять неожиданным изменениям* (Lerner, 1970). ГГ включает не только групповые эффекты на популяционном уровне, но и регуляцию на индивидуальном уровне, включая онтогенетический гомеостаз.

При обсуждении ГГ необходимо учесть три главных допущения (Lerner, 1970, с небольшими добавлениями).

1. Эволюция перекрестно размножающихся организмов оказывает влияние на генотипы, которые определяют характер индивидуального развития со значительной степенью саморегуляции. В эмбриологических терминах процесс развития определенных форм становится канализированным, ведя к единообразию фенотипической экспрессии среди особей данной популяции вопреки генетической изменчивости между ними.
2. Менделевские популяции также обладают самобалансирующими свойствами, стремящимися поддерживать генетический состав, который дает максимальную среднюю приспособленность в определенной среде существования данной популяции. Это с очевидностью распространяется и на более высокие уровни, такие как подразделенные популяции и виды.
3. Представляется, что механизм обоих типов гомеостаза лежит в превосходстве гетерозиготных генотипов над гомозиготными: (а) онтогенетическая саморегуляция (гомеостаз развития) базируется на большей способности гетерозигот оставаться в пределах нормы канализированного развития; (б) саморегуляция популяций (генетический гомеостаз) основывается на естественном отборе, скорее благоприятствующего промежуточным, чем крайним фенотипам.

В этой лекции мы постараемся рассмотреть как теорию (или точнее некоторые модели), так и эмпирические результаты по ГГ в отношении популяционного уровня и связи гетерозиготности и приспособительных признаков (ПП) (fitness traits). ПП (=КП) относятся к признакам, которые являются важнейшими компонентами приспособленности, таким как высота, вес, плодовитость, выживаемость и другие количественные признаки, которые могут сами направлять генетический гомеостаз. Таким образом, анализ связи между маркерными локусами и ПП играет особую роль в понимании существа ГГ.

Важное место в понимании роли ПП в гомеостазе дают результаты искусственного отбора. Из данных рис. 12.1.1 следует, что, если отбор прекращается, то значение ПП возвращается к адаптивной норме ($\underline{W}_1$; нижняя прерывистая линия) или может сдвинуться к другой условной норме, если преодолен какой-то порог толерантности ($\underline{W}_2$; верхние кривые).

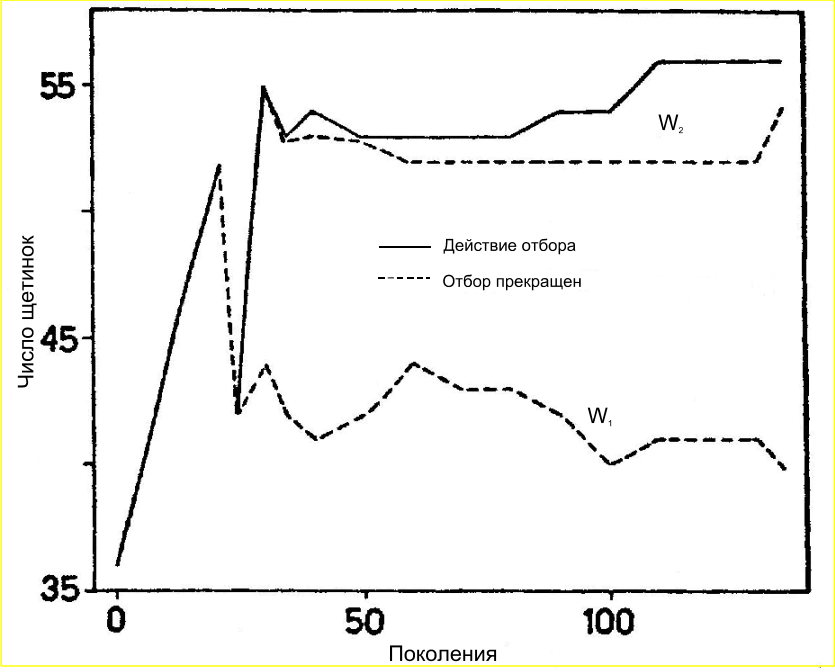


Рис. 12.1.1. Схематическое представление результатов искусственного отбора на увеличение числа абдоминальных щетинок (ПП) у Drosophila melanogaster в различных условиях.

$\underline{W_1}$ и $\underline{W_2}$– различные адаптивные нормы (По Mather, Harrison, 1949, заимствовано из Lerner, 1970 с добавлениями).
Fig. 12.1.1. Schematic representation of selection results for a high number of abdominal chaetae (FT) in Drosophila melanogaster (From Mather, Harrison, 1949, Op. Cit. Lerner, 1970 with adds).

Такого рода результаты приводят к мысли, что генотип организма имеет сбалансированный набор генов. Вследствие отбора он может быть изменен, но при релаксации отбора возвращается набор генотипов адаптивной нормы ($\underline{W_1}$), если изменения не преодолели порог толерантности. Если же отбор привел к преодолению порога толерантности, то устанавливается другая адаптивная нома ($\underline{W_2}$) (рис. 12.1.1, верхние кривые). Близкие идеи, в виде «концепции смещающегося равновесия», были предложены Райтом (Wright, 1935; 1969).

12.2. ГЕТЕРОЗИС И ГЕТЕРОЗИГОТНОСТЬ

Сущность гетерозиса и гетерозиготности различна, хотя эти два понятия перекрываются. Термин **гетерозис** был введен Шеллом (Shull, 1914), как *гибридная сила или увеличение «мощности» потомства F_1 по сравнению с обеими родительскими формами* без относительно механизма этого действия. Гибридная сила может проявляться по любому количественному признаку (КП): весу, длине, выживаемости и т.д. Формально это можно записать следующим образом: $P1 \times P2 = F_H$; так, что $P1 < F_H > P2$. Здесь: P1, P2 – это родительские линии, или популяции, а F_H – это гибридное потомство. Поскольку большинство КП связано с приспособленностью в Дарвиновском смысле, то гетерозис становится синонимом увеличенной приспособленности гибридов. А поскольку гибридизация в норме ведет к увеличенной гетерозиготности в F_1 (F_H), то гетерозис приравнивают к гетерозиготности. В настоящее время упоминаются, по меньшей мере, четыре различных варианта гетерозиса (Гершензон, 1983; Струнников, 1986) (рис. 12.2.1).

1. **Компенсация вредных эффектов рецессивных аллелей.**
2. **Аддитивные эффекты благоприятных генных комбинаций.**
3. **Сверхдоминирование или моногибридный гетерозис.**
4. **Сверхэпистаз модификаторов жизнеспособности** (терминология автора)**.**

1. *A, B, C, D* - норма, *a, b, c, d* – рецессивные маркеры, а КП контролируется генами x, y, z. Тогда: P_1 - *aaBBccDD* × P_2 - *AAbbCCdd*, $F_1 = F_H$ = *AaBbCcDd* и $P_1 < F_H > P_2$ или $x_1x_1y_1y_1z_1z_1 > x_0x_0y_0y_0z_0z_0$, где подстрочный индекс 0 и 1 обозначают P и $F_1 = F_H$.

2. $G_1W_1Y_1R_1$ и $G_2W_2Y_2R_2$ - норма, $G_1Y_1R_2$ – благоприятная комбинация, которая увеличивает значение КП. Тогда в скрещиваниях: P_1 - $G_1G_1W_1W_2Y_1Y_2R_1R_1$ × P_2 - $G_2G_2W_2W_2Y_1Y_2R_2R_2$, часть $F_1 = F_H = G_1G_2W_1W_2Y_1Y_2R_1R_2$ и $P_1 < F_H > P_2$.
3. Пусть $A_1A_2A_3$ – это норма. Тогда, если A_1A_2, A_1A_3, $A_2A_3 > A_1A_1$, A_2A_2, A_3A_3, $F_1 = F_H = P_1 < F_H > P_2$ или $x_1x_1y_1y_1z_1z_1 > x_0x_0y_0y_0z_0z_0$, где подстрочный индексы 0 и 1 обозначают P и F_1.
4. Пусть $x_1x_1y_1y_1z_1z_1$ и $x_1x_1y_1y_1z_2z_2$ – это субвитальные инбредные линии. В них рецессивные гены x, y контролируют КП, а z – является модификатором их жизнеспособности, который поддерживает их жизнеспособность на достаточном уровне. Кроме того, выполняется $z_1, z_2 > x_1, y_1$: (эпистаз). Тогда при скрещивании: P_1 - $x_1x_1y_1y_1z_1z_1$ (эпистаз) × P_2 - $x_1x_1y_1y_1z_2z_2$ (эпистаз), устанавливается сверхэпистаз: $F_1 = F_H = x_1x_1y_1y_1z_1z_2$ и $P_1 < F_H > P_2$.
Рис. 12. 2.1. Схематическое представление четырех вариантов гетерозиса: 1-4.
Fig. 12. 2.1. Four schemes of heterosis: 1-4.

Таким образом, мы должны ясно понять, что положительная взаимосвязь между гетерозиготностью и гетерозисом может существовать, но первая не всегда является причиной второго. Только в вариантах 1 и 3 гетерозиготность прямо влияет на гетерозис (рис. 12.2.1).

В символьном виде, взаимосвязь между полиморфным геном ***A*** и количественным выражением признака, находящимся под его контролем, может быть выражена такими соотношениями:

(1) $A_1A_1 = A_1A_2 = A_2A_2$ - нейтральность,
(2) $A_1A_1 > A_1A_2 > A_2A_2$ и $A_2A_2 > A_1A_2 > A_1A_1$ или $(A_1A_1 + A_2A_2) / 2 > A_1A_2$ - аддитивность,
(3) $A_1A_1 = A_1A_2 > A_2A_2$ и $A_2A_2 = A_1A_2 > A_1A_1$ или $(A_1A_1 + A_2A_2) / 2 < A_1A_2$ - доминантность,
(4) $A_1A_1 < A_1A_2 > A_2A_2$ - сверхдоминантность.

Связь между гетерозиготностью и ПП может быть объяснена как с позиций селективной нейтральности, так и с позиций селективной значимости маркеров. Наиболее признанными сейчас являются гипотезы инбредной депрессии (Mitton, Piers, 1980; Chakraborty, 1981) и балансовая (Bush et al., 1987). Анализ теоретической зависимости между гетерозиготностью и КП для трех из четырех описанных выше случаев провел Безруков (1989).

Гипотеза инбредной депрессии

Эта гипотеза предполагает, что сами по себе локусы нейтральны. Они являются только маркерами гетерозиготности генома (Mitton, Piers, 1980, Chakraborty, 1981). Чем ниже гетерозиготность маркеров, тем выше гомозиготность генома и инбредность особей; то есть, соответственно, в этом состоянии имеется множество вредных генов. Принципиальное значение здесь имеют отношения доминантности – рецессивности. Мы рассмотрим эту концепцию на примере из статьи Зоуроса (Zouros, 1987) (Табл. 12.2.1).

Пусть КП (скажем вес животного в возрасте X) находится под контролем двух локусов ***A*** и ***B***, каждый с двумя аллелями. Генотипические значения для локуса ***A*** составляют: $A_1A_1 = 2$, $A_2A_2 = 6$ и $A_1A_2 = 5$. Это средние значения для трех генотипов, взвешенные на их частоты в популяции. Генотипические значения для трех генотипов по локусу ***B*** равны, соответственно, 2, 8 и 7. Сверхдоминирования нет ни по локусу ***A*** (2<5<6)**,** ни по локусу ***B*** (2<7<8). Давайте для простоты получим двухлокусные генотипические значения как средние из однолокусных значений (Табл. 12.2.1). В каждой клетке таблицы косые линии разделяют значения для таких усреднений, в последовательности: средняя арифметическая, средняя геометрическая и средняя гармоническая.

Во второй части Табл. 12.2.1 (b) 9 двухлокусных генотипов были объединены в три группы (по 4, 4 и 1 значению в группе) в соответствии со степенью их индивидуальной гетерозиготности: $H_O = 0$, 1 и 2. После этого были рассчитаны средние значения для каждой группы. Мы видим, что чем выше гетерозиготность, тем больше генотипическое значение (Табл. 12.2.1, b). Это сохраняется при различных способах усреднения. Таким образом, без сверхдоминирования усреднение однолокусных значений в мультилокусные наборы дает корреляцию, в этом случае

положительную корреляцию. Во втором примере, значения даны в скобках, все эти вычисления были повторены с измененными значениями для однолокусных гетерозигот (A_1A_2 изменено с 5 до 3 и B_1B_2 - с 7 до 3) (Табл. 12.2.1). Теперь корреляция H_O и значения КП отрицательна (Табл. 12.2.1, b). В первом примере показан случай, который близок к ситуации внутрилокусных отношений (3) или к доминированию. Во втором примере, рассмотрен случай близкий к ситуации (2) или однолокусной аддитивности.

Численный пример для случая (2) аддитивности не поддерживается теоретическими предсказаниями для этого случая (Безруков, 1989, рис. 1), но поддерживаются численные примеры для случая (3) доминантности (Chakraborty, Ryman, 1983) и для случая (4) сверхдоминантности (Безруков, 1989, рис. 1) (см. обсуждение ниже). Однако, в соответствии с моими выкладками (Kartavtsev, 2003; Картавцев, 2004), теория предсказывает как отрицательную связь H_O и КП, когда преобладают аддитивные внутрилокусные эффекты, так и положительную ассоциацию H_O и КП, когда имеется внутрилокусное доминирование. Безруков (1989, рис. 1) показал отсутствие корреляции для случая аддитивности, но в действительности здесь предсказывается отрицательная корреляция (Картавцев, 2004).

Таким образом, *для общего суждения в поддержку гипотезы инбредной депрессии нет необходимости привлекать усреднение или мультилокусную сверхдоминантность* (см. выше и следующий абзац). *Необходимо только знать, преобладает ли доминирование в природе по сравнению с аддитивностью, или по отношению к другим типам действия гена*. Я не уверен, что факт превалирования доминантности в действительности существует. Наоборот, в основах генетики количественных признаков и в селекционной практике широко принята аддитивность действия генов и только в отношении жизнеспособности и плодовитости признается значительной роль не аддитивных эффектов (Кирпичников, 1987). Однако мнение Зоуроса (Zouros, 1987) противоположно и он поддерживает свое заключение резонными свидетельствами.

Таблица 12.2.1. Численный пример усредняющих эффектов генотипических значений отдельных локусов в мультилокусные генотипы в соответствии с определенным рангом гетерозиготности (По Zouros, 1987)

Table 12.2.1. Numerical example of the averaging effect of genotypic values at multi-loci genotypes in accordance to certain heterozygosity rank (From Zouros, 1987)

(a) Получение двухлокусных генотипических значений из однолокусных значений

ГЕНОТИПИ-ЧЕСКОЕ ЗНАЧЕНИЕ		**A_1A_1 2 (2)**	**A_2A_2 6 (6)**	**A_1A_2 5 (3)**
B_1B_1	**2 (2)**	2/2/2 (2/2/2)	4/3.46/3 (4/3.46/3)	3.5/3.16/2.86 (2.5/2.45/2.4)
B_2B_2	**8 (8)**	5/4/3.2 (5/4/3.2)	7/6.93/6.86 (7/6.93/6.86)	6.5/6.32/6.15 (5.5/4.9/4.36)
B_1B_2	**7 (3)**	4.5/3.74/3.11 (2.5/2.45/2.4)	6.5/6.48/6.46 (4.5/4.24/4)	6/5.91/5.83 (3/3/3)

(b) Средние значения для двухлокусных генотипов в соответствии с рангом гетерозиготности

Модель средней	Гетерозиготность, H_O		
	0	1	2
1. Арифметическая	4.5 (4.5)	5.25 (3.75)	6.0 (3)
2. Геометрическая	4.1 (4.1)	4.92 (3.51)	5.91 (3)
3. Гармоническая	3.76 (3.76)	4.64 (3.29)	5.83 (3)

Примечание. В скобках приведен второй набор значений (второй пример).

В связи с дискуссией, аналогичной представленной выше, были поставлены два вопроса (Mitton, Piers, 1980; Chakraborty, 1981; Zouros, Foltz, 1987). 1. Представляют ли исследованные

маркерные локусы подвыборку всех многих тысяч локусов генома? 2. И если это так, то представляют ли они случайную выборку в смысле связи H_O и КП, и в смысле изменчивости КП в соответствии с рангом H_O? Заключение последних цитированных авторов было отрицательным. То есть, если число исследованных локусов мало, скажем меньше, чем 20 (как в большинстве экспериментальных исследований), то ранг индивидуальной гетерозиготности практически не коррелирует с геномной гетерозиготностью и независим от изменчивости КП. По логике этих авторов, это является главной причиной, почему в большинстве исследований корреляции H_O и КП обнаруживались очень редко. Однако фактические данные, представленные ниже (см. Раздел 12.3), противоречат этому. Причина же может быть просто в отсутствии преобладания доминантности, как мы заметили ранее.

Балансовая гипотеза или гипотеза сверхдоминирования

В случае сверхдоминирования теория предсказывает прямо пропорциональную связь H_O и КП (рис. 12.2.2) и на эту тему имеются многочисленные формальные выкладки. Эта гипотеза предсказывает, что сами локусы или тесно сцепленные с ними маркеры влияют на связь H_O и КП.

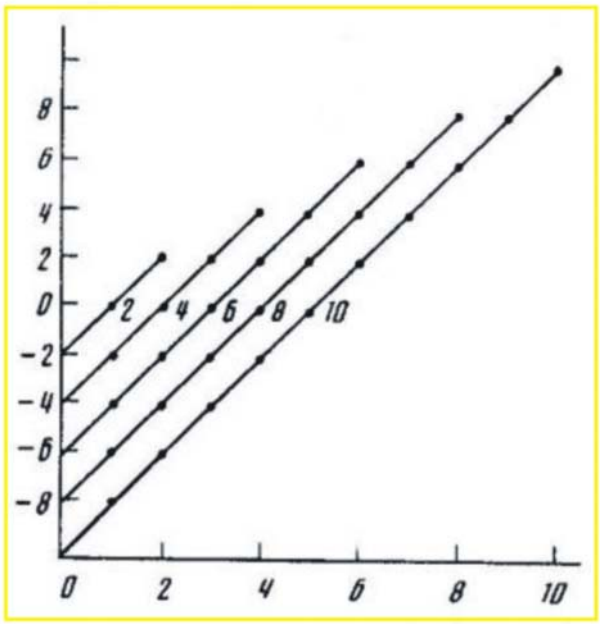


Рис. 12.2.2. Зависимость среднего значения КП от числа гетерозиготных локусов: случай сверхдоминирования.

*По оси X - H_O, число гетерозиготных локусов у особи. По оси Y - **Y-y**, где **Y** – это средний популяционный уровень КП, а **y** – значение КП у отдельной особи. Числа рядом с линиями обозначают число локусов в выборке (По Безруков, 1989).*

Fig. 12.2.2. Dependence of mean FT (fitness trait) score from the number of heterozygous loci: the over-dominance case (From Bezrukov, 1989).

*On the X-axis, H_O is the number of heterozygous loci in an individual. On the Y-axis is **Y-y**, where **Y** is mean population level of FT, and **y** is its score in an individual. Numbers beside lines denote number of loci in the sample under the simulation.*

Формально эта ситуация описывается случаем (4) выше. Турелли и Гинзбург (Turelli, Ginzburg, 1983) показали, что при любой из разнообразных моделей балансирующего отбора, с увеличением гетерозиготности, также ожидается увеличение приспособленности. Смоуз (Smouse, 1986) разработал мультипликативную модель сверхдоминирования, в которой модель инбредной депрессии является частным случаем. Автор называет ее моделью адаптивного расстояния. Идеологически эта модель связана с цитированной выше разработкой (Turelli, Ginzburg, 1983), где было показано, что маргинальное сверхдоминирование является наиболее частым следствием мультилокусного полиморфизма.

Рассмотрим модель адаптивного расстояния (AD), как это представлено в статье Буша и соавторов (Bush et al., 1987). Модель предназначена для описания связи между $\boldsymbol{Y = Log\ (W)}$, где W – это приспособленность, и $\boldsymbol{X}$, которое равно AD в виде частот аллелей: $\boldsymbol{P_A = f(A_1)}$ и $\boldsymbol{Q_A = f(A_2)}$.

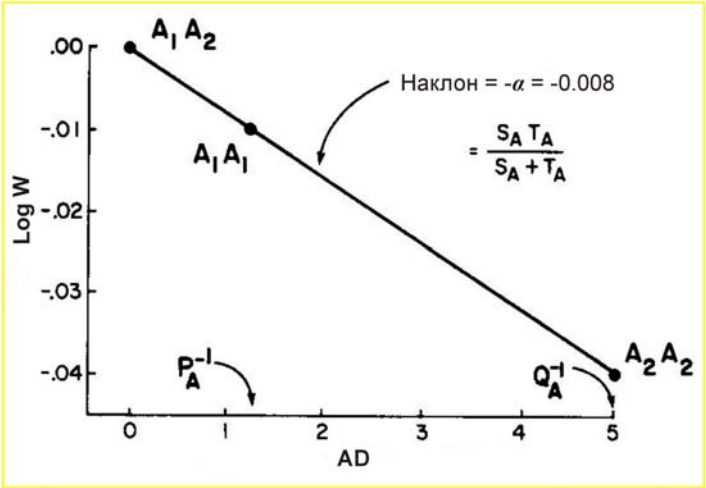


Рис. 12.2.3. Пример связи приспособленности (Log W) с адаптивным расстоянием (AD) для трех возможных генотипов по локусу A.

Значение AD гетерозиготы равно нулю. Значение AD для каждой гомозиготы равняется обратному значению равновесной частоты соответствующего аллеля. Здесь, P_A = 0.80 и Q_A = 0.20 (По Bush et al., 1987).

Fig. 12.2.3. Example plot of logW (fitness) against adaptive distance (AD) for the three possible genotypes at locus A.

The AD value of the heterozygote is zero. The AD value for each homozygote is equal to the inverse of the equilibrium frequency of the corresponding allele. Here, P_A = 0.80 and Q_A = 0.20 (From Bush et al., 1987).

Здесь три генотипа диаллельного локуса определяют значения AD через частоты аллелей: $P_A = f(A_1)$ и $Q_A = f(A_2)$. Было показано (Smouse, 1986), что если S_A и T_A являются селекционными дифференциалами (1-s и 1-t) для гомозигот A_1A_1 и A_2A_2, соответственно, то равновесные частоты аллелей являются функцией S_A и T_A. Таким образом, можно обозначить AD как:

Генотип	A_1A_1	A_1A_2	A_2A_2
AD	P_A^{-1}	0	Q_A^{-1}
W	$-S_A$	1	$-T_A$.

Тогда взаимосвязь между ***X*** и ***Y*** будет при равновесии линейной (рис. 12.2.3). Коэффициент регрессии X_A – это так называемый **сегрегационный генетический груз** для этого локуса (Morton et al., 1956):

$$\alpha = [S_A T_A / (S_A + T_A)]. \qquad \textbf{(12.1)}$$

В этом случае мы можем определить регрессионную модель для j-той особи следующим образом:

$$Y_j = 0 - aXA_j + e_j. \qquad \textbf{(12.2)}$$

Здесь e_j – это член, который определяет и ошибку оценки в Y_j и отклонение от модели. Для мультипликативной модели сверхдоминирования значения AD для различных локусов являются аддитивными. Поэтому (12.2) может быть расширено до выражения:

$$Y_j = 0 - aXA_j - bXB_j - \ldots - kXK_j + e_j. \qquad \textbf{(12.3)}$$

Нулевой первый член в уравнении предполагает использование k-локусной гетерозиготы, в качестве реперного генотипа. Обычные оценки P_A и P_Q определяются как:

$$P = Z / N \text{ и } Q = (N - Z) / N, \qquad \textbf{(12.4a)}$$

где ***Z*** – это число аллелей A_1 и ***(N - Z)*** - A_2. Выражения для P и Q, также как их обратных значений P_A^{-1} (1/P) и Q_A^{-1} (1/Q), являются смещенными. Чтобы избежать этого, возможна подстановка:

$$P = (Z + 1) / (N + 1) \text{ и } Q = (N - Z + 1) / (N + 1). \qquad \textbf{(12.4b)}$$

Обратные значения для этих последних оценок имеют ожидания (Smouse, Chakraborty, 1986):

$$E[1 / P] = 1 / P[1 - Q]^{N+1} \text{ и}$$

$$E[1 / Q] = 1 / Q[1 - P]^{N+1}. \qquad \textbf{(12.5)}$$

Последние выражения дают количественную основу для эмпирического тестирования гипотезы сверхдоминирования. В частности, в исследовании сосны *Pinus rigida* (Bush et al., 1987) эта модель лучше аппроксимирует эмпирические данные. Некоторые другие данные на эту тему будут рассмотрены в следующем разделе.

12.3. ЭМПИРИЧЕСКИЕ РЕЗУЛЬТАТЫ И ИХ ИНТЕРПРЕТАЦИЯ

В соответствие с нашим главным намерением мы будем рассматривать в основном данные для морских животных. Моллюски были исследованы наиболее подробно среди этих животных. Хотя надо отметить, что имеется множество данных для человека (Дуброва, Сусков, 1990; Дуброва и др., 1990), рыб (Mitton, Lewis, 1989) и хвойных растений (Политов и др., 1992). Из 17 проведенных исследований моллюсков, только в 5 не были обнаружены корреляции между гетерозиготностью и темпом роста (Табл. 12.3.1).

Таблица 12.3.1. Гетерозиготность и темп роста в природных популяциях моллюсков (По Zouros, 1987 с добавлениями)

Table 12.3.1. Heterozygosity and growth rate in natural population of mollusks (From Zouros, 1987 with adds)

Вид	**N**	**k**	***D***	**r_p**	**Источ-ник**	**Примечание**
Crassostrea virginica	400	5	-	+	1	молодь
То же	1500	7	-	+	2	То же
-"-	200	4	-	+	3	-"-
Mytilus edulis	650	5	-	+	4	-"-
То же	500	5	0	0/-	5	Смесь возрастных групп
-"-	300	7	-	+	6	Смесь возрастных групп
-"-	2739	5	-	0/+	7	Смесь возрастных групп, 6 выборок
-"-	106	?	?	+	8	Смесь возрастных групп, + - увеличенная плотность
Macoma baltica	150	6	-	+	9	?
Placopecten magellanicus	60	6	0/-	0	10	Одновозрастная группа
Pecten maximus	110	5	0	0	11	Смесь возрастных групп
Taias hemastoma	290	6	-	+	12	?
Mulinia lateralis	110	6	0	+	13	Смесь возрастных групп
То же	80	18	?	+	14	Молодь
Mytilus edulis	3726	17	?	+	15	Молодь
То же	>3000	5	-	+	16	Смесь возрастных групп
То же	551	5	-?	0	17	То же
Mytilus trossulus	1234	15	-	+/0*	18	То же

Примечание. N – число животных в исследовании, k – число проанализированных локусов, D – индекс избытка (+) или дефицита (-) гетерозигот (D = -Fis), r_p – это коэффициент корреляции между уровнем гетерозиготности и темпом роста (показаны только статистически существенные значения): "+" – положительная, "-" отрицательная, "0" – отсутствие корреляции. Источники: 1 - Singh, Zouros, 1978, 2 - Zouros et al., 1980, 3 - Singh, 1982, 4 - Koehn, Gaffney, 1984, 5 - Diehl, Koehn, 1985, 6 - Zouros, 1987, 7 - Gaffney, 1990, 8 - Gentini, Beaumont, 1988, 9 - Green et al., 1983, 10 - Zouros, Foltz, 1984, 11 - Beaumont et al., 1985, 12 - Garton, 1984, 13 – Garton et al., 1984, 14 - Diehl et al., 1986, 15 - Koehn et al., 1988, 16 - Gardner, Skibinski, 1988, 17 - Gosling, 1989, 18 – Картавцев, 1995 (*- дисперсионный анализ).

Корреляции, как правило, более рельефны среди молоди и ослаблены или вовсе отсутствуют на более поздних онтогенетических фазах. У моллюсков эта тенденция связана с индексом дефицита гетерозигот (D = - Fis), который у более взрослых животных уменьшается

(Levinton, Koehn, 1976; Картавцев, 1979; Koehn, Gaffney, 1984; Zouros, Foltz, 1987; Gaffney et al., 1990). У северной креветки корреляция H_O с двумя индексами обнаруживается только у самцов, самой маленькой размерной группы у этого переходного гермафродита: $r_P = -0.24$, $r_P = -0.26$ (n = 55, P < 0.05 и P < 0.025; Kartavtsev, 1996). Имеется ассоциация двух индексов W/L, H/L и H_O в двух выборках тихоокеанской мидии, *M. trossulus*: ***F*** = 3.126 (d.f. = 2; 67, P = 0.05) и ***F*** = 4.980 (d.f. = 2; 36, P = 0.012), но не в 15 других. Гетерогенность в этом отношении среди 15 выборок статистически состоятельна: ***F*** = 1.535 (d.f. = 33; 1234, P = 0.027; Картавцев, 1995). Близкие результаты получены для обыкновенной мидии, *M. edulis* (Gaffney, 1990). У лососей известны многие КП и H_O связи или влияние градации в зависимости от гетерозиготности (Leary et al., 1985; 1987; Danzman et al., 1989; Mitton, Lewis, 1989; Картавцев, 1990; 1998). Например, ранг гетерозиготности родителей влияет на смертность личинок у горбуши, *Oncorhynchus gorbuscha* (рис. 12.3.1).

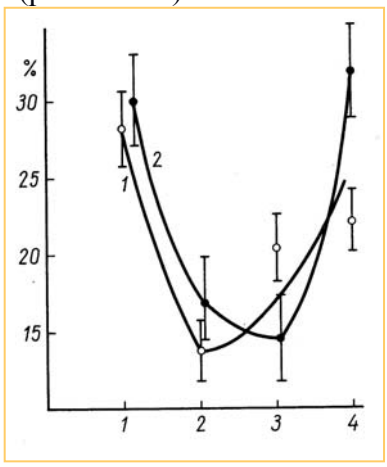


Рис. 12.3.1. Связь ранга гетерозиготности родителей (P) и уровня смертности личинок горбуши,Oncorhynchus gorbuscha (По Картавцев и др., 1990).

По оси X – ранг гетерозиготности (число гетерозиготных локусов P1+P2), по оси Y – доли погибших (%). Ранги определены как сумма гетерозиготных локусов у двух родителей: 1 - 0 (♀) + 0(♂); 2 -2 (♀) + 0 (♂); 3 - 0 (♀) + 4 (♂); 4 - 2 (♀) + 4 (♂). Значимость, кривая 1: t_d (1-2) = 8.7, P<0.001; t_d (2-4) =5.9, P<0.001, n=5184; кривая 2: t_d (1-2) =8.2, P<0.001; t_d (3-4) = 8.1, P<0.001, n = 2732.

Fig. 12.3.1. Association of heterozygosity rank of parents and the death rate of larvae in pink salmon,Oncorhynchus gorbuscha (From Kartavtsev et al., 1990).

On the X-axis is heterozygosity rank (number heterozygous loci in P1+P2), on Y-axis is death rate. Ranks defined as the sum of heterozygous loci in two parents: 1 - 0 (♀) + 0(♂); 2 -2 (♀) + 0 (♂); 3 - 0 (♀) + 4 (♂); 4 - 2 (♀) + 4 (♂). Significance, curve 1: t_d (1-2) = 8.7, P<0.001; t_d (2-4) =5.9, P<0.001, n=5184; curve 2: t_d (1-2) =8.2, P<0.001; t_d (3-4) = 8.1, P<0.001, n = 2732.

Две другие связи КП и H_O показаны на рис. 12.3.2 и 12.3.3. Очевидно, что у молоди горбуши обнаруживается отрицательная связь между отклонением значения КП у особи от групповой средней и уровнем индивидуальной гетерозиготности, H_O (рис. 12.3.2). То есть, в согласии концепцией ГГ, можно говорить о том, что более гетерозиготные по аллозимным маркерам особи, являются лучше гомеостазированными морфологически.

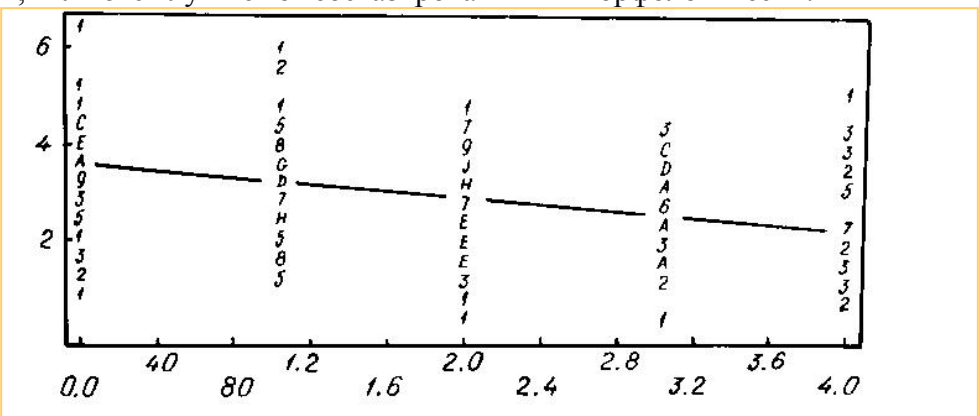


Рис. 12.3.2. Корреляция изменчивости H_O с отклонением от группового центроида по

комплексу из 10 морфологических признаков у молоди горбуши, полученной в диаллельных скрещиваниях.

По оси X - H_O, по оси Y – апостериорная вероятность (P) исключения особи из определенной H_O-группы в ходе дискриминантного анализа (По Картавцев, 1990).

Fig. 12.3.2. Correlation of H_O variability with 10 morphology traits in pink salmon fry, taken in di-allele crosses.

On X-axis, H_O, on Y-axis, aposterior probability (P) of an individual exclusion from the certain H_O-group during discriminant analysis (From Kartavtsev, 1990).

В данном случае с горбушей, очевидно, что в этом участвует естественный отбор против слишком мелких особей, поскольку они наименее гетерозиготны по сравнению с группой средних, а также крупных и диспропорциональных особей (рис.12.3.3).

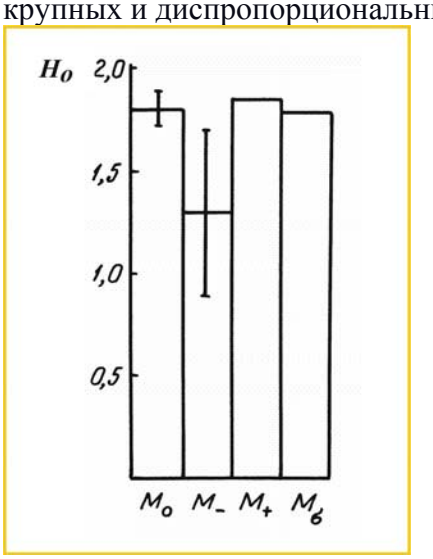


Рис. 12.3.3. Сопоставление H_O в четырех морфологически различных группах молоди горбуши в экспериментальном блоке 2.

M_O, M-, M+ - группы особей со средними, малыми, и большими значениями по комплексу морфологических признаков, M_σ – это диспропорциональные особи. Линии отмечают доверительный интервал (95%) (По Картавцев, 1990).

Fig. 12.3.3. H_O comparison in four morphologically distinct groups of pink salmon fry in experimental unit 2: M_O, M-, M+ are individuals with average, small, and large morphology scores, M_σ is disproportional individuals. Lines are confidence limits (95%) (From Kartavtsev, 1990).

Мультилокусная гетерозиготность коррелирует с КП у многих видов рыб (Mitton, Lewis, 1989). У лососей, которые лучше изучены, чем многие другие виды рыб, ассоциации КП и H_O, обычно, обнаруживаются на ранних стадиях жизненного цикла, или в группах, ранжированных в соответствии с качеством окружающей среды (Алтухов, Варнавская, 1983; Алтухов, 1989; Leary et al., 1985: 1987; Danzman et al., 1989; Картавцев, 1990; 1998; Картавцев и др., 1990; Kartavtsev, 1992), но не обнаруживаются у взрослых особей из природы (Beacham, Withler, 1985).

Взаимосвязь КП с гетерозиготностью на практике не всегда линейна, но нередко и не линейна (рис. 12.3.4; см. также рис. 12.3.1 выше). Такие результаты представлены во многих работах (Zouros, Foltz, 1987; Mitton, Lewis, 1989; Дуброва, Сусков, 1990; Дуброва и др., 1990; 1995).

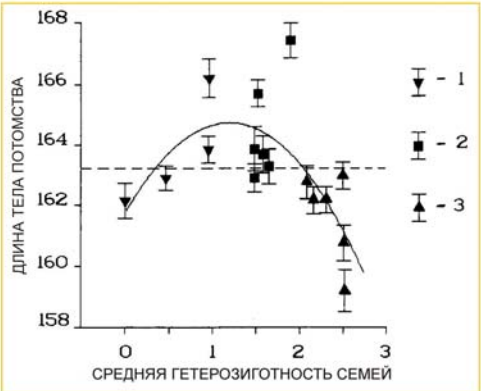


Рис. 12.3.4. Взаимосвязь между средней внутрисемейной гетерозиготностью и длиной особей в потомствах: 1 - семьи с низкой гетерозиготностью; 2 - семьи с промежуточной гетерозиготностью; 3 – семьи с высокой гетерозиготностью. Пунктирная линия обозначает средний вес тела для всего исследованного потомства. Линейная аппроксимация (не показана): $y = 164.9 - 0.87x$; $F (1; 2563) = 66.79$, $P< 0.001$; $R^2 = 0.0253$; квадратическая аппроксимация (сплошная кривая): $y=161.7+5.51x -2.27x^2$; $F (2; 2562) = 189.23$, $P< 0.0001$, $R^2 = 0.1290$ (По Dubrova et al., 1995).

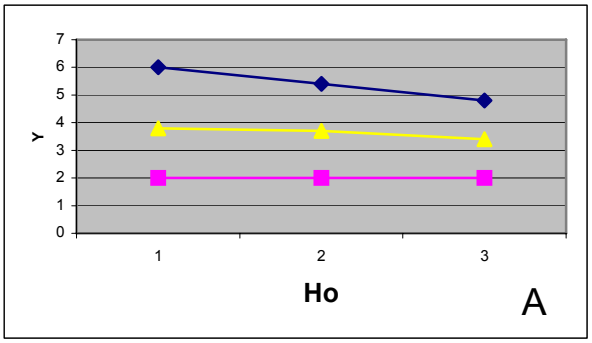
Fig. 12.3.4. The relationship between mean family heterozygosity and progeny body length of pink salmon: 1, families with low

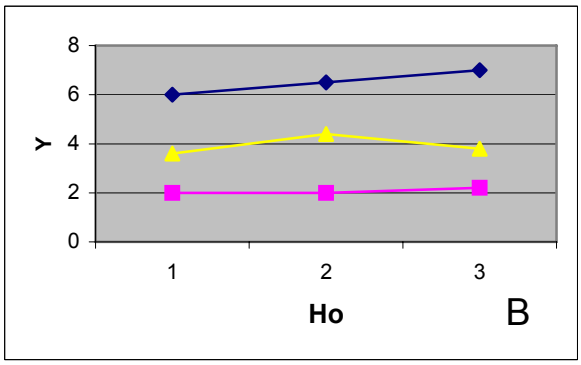
heterozygosity; 2, families with intermediate heterozygosity; 3, families with high heterozygosity. The line indicates the mean body length value for all the progeny examined. Linear fit (not shown) $y = 164.9 - 0.87x$; $F(1; 2563) = 66.79$, $P < 0.001$; $R^2 = 0.0253$; quadratic fit $y=161.7+5.51x -2.27x^2$; $F(2; 2562) = 189.23$, $P < 0.0001$, $R^2 = 0.1290$ (From Dubrova et al., 1995).

Какая из двух гипотез, которые мы рассмотрели в Разделе 12.2, удовлетворяет лучше эмпирическим результатам? На первый взгляд гипотеза инбредной депрессии является более привлекательной, вследствие общности и хорошо известным вредным эффектам при различных уровнях инбредности. Однако, эмпирические результаты, кажется, не очень хорошо согласуются с этим. Несогласие видится в следующих 7 пунктах.

1. Обнаруженные корреляции в большинстве случаев принадлежат к выборкам небольшого числа маркерных генов из генома.
2. Сила корреляций меняется в зависимости от фаз жизненного цикла.
3. Ранжирование в соответствие с рангом пресса среды усиливает корреляции.
4. Знак корреляций между гетерозиготностью и КП может меняться в различных популяциях и видах. То есть, между двумя переменными обнаруживаются не только положительные, но и отрицательные связи.
5. Дисперсия по КП может коррелировать с гетерозиготностью положительно, а не отрицательно, как предсказывает гипотеза инбредной депрессии.
6. Истинная сверхдоминантность по маркерным локусам обнаруживается в некоторых случаях.
7. Связь КП и гетерозиготности во многих случаях имеет не линейный характер.

Значит ли это, что балансовая гипотеза лучше согласуется с эмпирическими результатами? Я в этом сомневаюсь, хотя большинство из 7 пунктов, рассмотренных выше, могут быть объяснены в рамках этой гипотезы. В чем причины такого скептицизма? Прежде, чем ответить на этот вопрос, давайте рассмотрим теоретические предсказания о связи гетерозиготности при различных типах аллельного взаимодействия внутри локусов (Kartavtsev, 2003; Картавцев, 2004) (рис. 12.3.5).





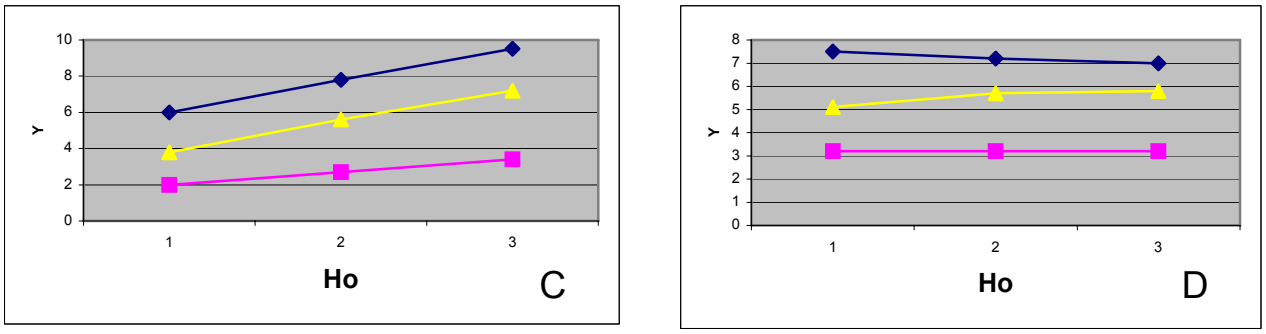


Рис. 12.3.5. Связь между уровнем индивидуальной гетерозиготности (H_o) и средним генотипическим значением в диаллельных генотипических классах (Y).

1, 2, 3 – уровни гетерозиготности, H_o, которые представляют число гетерозиготных локусов у особи: $H_o = 0$, $H_o = 1$ и $H_o = 2$. Кривые 1, 2 и 3 сверху вниз (ромбы, треугольники и квадраты) – это различные искусственные примеры из таблицы данных. A – аддитивность, B - доминантность, C – сверхдоминантность, D – нейтральность (По Картавцев, 2004).

Fig. 12.3.5. Association of individual heterozygosity level (H_o) and mean genotypic value at di-loci genotype classes (Y).

1, 2, 3 are the groups of H_o-classes, which represented the number of heterozygous loci in an individual, $H_o = 0$, $H_o = 1$ and $H_o = 2$. Curve 1, 2 and 3 from top to bottom are examples (rows 1-3) from the source table. A – Additivity, B - Dominance, C – Over-dominance, D – Neutrality (From Kartavtsev, 2004).

Для четырех типов внутрилокусного действия (аддитивности, доминантности, сверхдоминантности и нейтральности) выражение генотипического значения (***Y***) согласно дозе гена (***X***) рассмотрено на основе линейной модели. Аналитически и с помощью искусственных численных примеров показано, что при аддитивности, доминантности и сверхдоминантности существует положительная взаимосвязь между ***X*** и ***Y***, тогда как при нейтральности связь отсутствует. Усреднение однолокусных генотипических значений в мультилокусные генотипы (как в примере выше) дает не одинаковые результаты для различных типов внутрилокусного действия (рис. 12.3.5). При доминировании и сверхдоминировании генотипические значения и индивидуальная мультилокусная гетерозиготность коррелируют положительно, а при аддитивности – обнаруживается отрицательная корреляция этих двух переменных. При нейтральности не обнаруживается связи между значениями двух переменных, также как и при действии только внутрилокусных эффектов.

Объединение всех этих эффектов на одном графике дает очень реалистичную взаимосвязь H_o и КП (рис. 12.3.6).

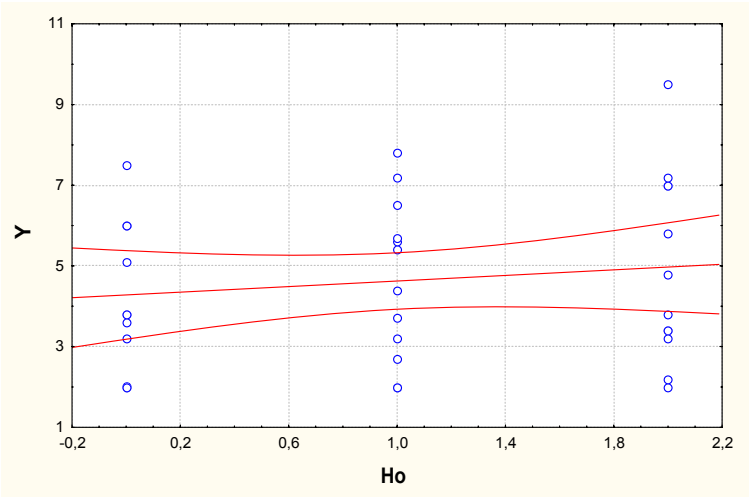


*Рис. 12.3.6. Усредненная корреляция между индивидуальной гетерозиготностью (**H_o**) и генотипическим значением (**Y**) для всех четырех типов внутрилокусного действия генов, полученная на основе искусственных примеров.*

*Значения корреля-ций равны: r_p = 0.256, r_s = 0.311, n =36, p > 0.5. Линейная аппроксимация: Y = 4,2806 + 0,34500 * H_o. Показана линия регрессии и ее доверительный интервал (95%) (По Картавцев, 2004).*

*Fig. 12.3.6. Averaged correlation between individual heterozygosity (**H_o**) and genotypic value (Y) for all four modes of intra-loci gene action as exemplified by artificial examples in source table (Kartavtsev, 2004).*

Correlation value is: r_p = 0.256, r_s = 0.311, n =36, p > 0.5.

Как мы могли убедиться, имеются различия между предсказаниями связей при различных моделях внутрилокусного действия для маркерных генов. Нейтральность предсказывает отсутствие связи. Аддитивность предсказывает отрицательную связь КП и H_O; аддитивность также требует различия дисперсий в композитных группах гомозигот и гетерозигот по мультилокусным генотипам, то есть, предсказывается отрицательная связь дисперсии гетерозиготности и КП (см. Chakraborty, Ryman, 1983; Безруков, 1989). Доминирование дает предсказание положительной связи КП и H_O, но также как при аддитивности, связь дисперсии гетерозиготности и КП должна быть отрицательной. Сверхдоминирование предсказывает положительную связь КП и H_O (Безруков, 1989; Картавцев, 2004) и отсутствие различия дисперсий у мультилокусных гомозигот и гетерозигот (Безруков, 1989).

Принимая все изложенное во внимание, мы понимаем, что ни одна из моделей не верна в приложении к эмпирическим данным. Более того, в реальном геноме возможно встретить все четыре перечисленные типа генного действия в комбинации, с результирующим эффектом подобным нашему примеру (рис. 12.3.6). При численном моделирования таких смешанных ситуаций генного действия сразу обнаруживаются криволинейные связи КП и H_O (Безруков, 1989, рис. 3). Межлокусные взаимодействия будут создавать еще большую сложность взаимосвязи КП и H_O и они также могут быть источником нелинейности связи, как предполагалось ранее (Дуброва, Сусков, 1990; Dubrova et al., 1995).

Необходимо отметить еще три важных обстоятельства. Во-первых, ожидается, что каждый индивидуальный маркерный локус (аллозимный локус, сайт ДНК) оказывает только небольшое воздействие на приспособленность особи на фенотипическом уровне (Lewontin, 1974). В нашем случае (см. рис. 12.3.4) факториальное влияние по аллозимным локусам было в диапазоне величин: R^2 = 0.025 - 0.129 или около 2-13% (Kartavtsev, 1996; Dubrova et al., 1995; Картавцев, 1990; 1998; 2004). Во-вторых, даже когда факториальное влияние велико, 0.4-0.6 или 40-60% (Zouros, Mallet, 1989; Алтухов, Шереметьева, 2000), оно не обусловлено влиянием H_O, а отражает влияние групповой гетерозиготности. В-третьих, когда проводится комплексный анализ, обнаруживаются межлокусные эффекты, и нет простого взаимодействия H_O и КП (или среды) (Dubrova et al., 1995). Последнее обстоятельство очень примечательно, поскольку гетерозиготность и гетерозис могут

коррелировать, но первая не всегда является причиной второго, как мы отметили в начале этой лекции (см. также Струнников, 1986).

Вопреки всем этим противоречиям мы, все же, должны принять балансовую гипотезу как рабочий инструмент в дальнейшем поиске истины. Главное, что побуждает к этому – это обобщенность предложенных селективных моделей (Turelli, Ginzburg, 1983; Smouse, 1986; Bush et al., 1987). Когда более простая модель, используемая как нуль-гипотеза, отвергается, как не адекватная, то допустимо принять за основу более сложную.

Еще один вопрос необходимо заострить в связи с завершением этой дискуссии. Это концептуальная необходимость разработки модели, позволяющей осуществлять авторегуляцию или ГГ на популяционном уровне. Имеющаяся модель частотно-зависимой регуляции не решает многих проблем, в том числе – генетического груза и изменчивости КП. В качестве наброска к такой модели можно предложить балансировку различных векторов отбора, действующего в пользу гетерозигот, определяя темп роста (размер) и наоборот, действующего против увеличения гетерозиготности, определяя выживаемость. На определенных фазах жизненного цикла при этом происходит релаксация интенсивности отбора вплоть до нуля. Три рода фактов подталкивают к созданию такой модели. 1. Во многих случаях (см. Табл. 12.3.1; Алтухов, 1989; Алтухов и др., 1991) мультилокусные гетерозиготы имеют преимущества в отношении влияния на КП, в виде увеличенного темпа роста или веса, как, например, обнаружено у молоди горбуши (рис. 12.3.7). 2. В тоже время выживаемость на групповом уровне оказывается наилучшей среди промежуточных по гетерозиготности особей (см. рис. 12.3.2, 12.3.4). 3. Учтем также, что факториальное влияние гетерозиготности на КП может снижаться за счет снижения интенсивности отбора в онтогенезе, а также комбинирования типов внутрилокусного и межлокусного взаимодействий (см. выше). Эти факты, в развитие изложенных ранее идей (Livshits, Kobyliansky, 1985; Алтухов, 1989; Алтухов и др., 1991), позволяют предложить *механизм балансирующего и релаксирующего в ходе жизненного цикла естественного отбора, (1) как регулятора оптимальной гетерозиготности в популяции (стабильный сегрегационный груз) и (2) как регулятора фенотипического гомеостаза по КП*. Если такая модель будет создана, то это будет лучшим развитием идей ГГ на популяционном уровне, хотя некоторые из допущений Лернера, сделанные в 1950-х годах, безусловно требуют корректировки с учетом современных генетических знаний.

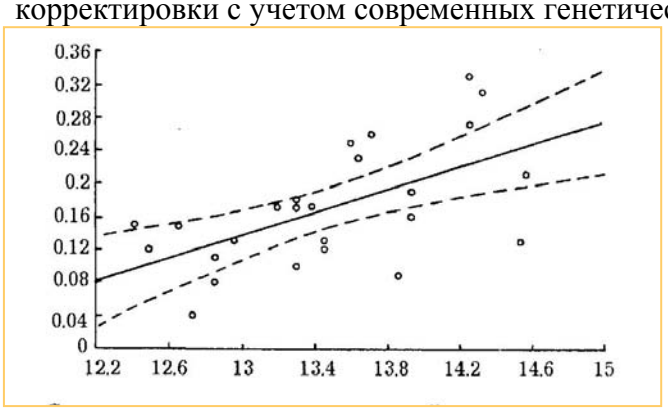


Рис. 12.3.7. Сопряженная изменчивость ожидаемой гетерозиготности (Hs) и веса тела у молоди горбуши в группах, содержавшихся в различных условиях плотности.

По оси X – квадратный корень веса тела (m), по оси Y - Hs для двух аллозимных локусов. Связь статистически значима: $r_P = 0.59$ ($n = 25$, $P < 0.025$). Показаны линия регрессии: $y = -0.7537 + 0.06855x$ и ее доверительный интервал (95%) (По Картавцев, 1998).

Fig. 12.3.7. Associated variation in expected heterozygosity (Hs) and body weight in pink salmon fry among groups that subjected to rearing in conditions of different density.

On the X-axis, square root of weight (m), on the Y-axis, Hs for two allozyme loci. Association is statistically significant: $r_P = 0.59$ ($n = 25$, $p < 0.025$). The regression line $y= -0.7537 + 0.06855x$ and its confidence limit (95%) are shown (From Kartavtsev,

1998).

12.4. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 12

1. Решите два примера связи H_o и КП используя пакет STATISTICA.
2. Принципы дисперсионного анализа (ANOVA) для определения изменчивости КП. Нахождение компонент изменчивости при реализации полной факторной схемы скрещивания и использовании статистических пакетов программ.
3. Принципы дискриминантного и факторного анализа при исследовании изменчивости КП.

ГЛАВА 13. ПОЛИМОРФИЗМ ДНК ВНУТРИ И МЕЖДУ ПОПУЛЯЦИЯМИ

ГЛАВНЫЕ ВОПРОСЫ:

13.1. Полиморфизм последовательностей ДНК. Нуклеотидное разнообразие и *p*-расстояние.
13.2. Полиморфизм ДНК, оцененный по данным рестриктазного анализа.
13.3. Полиморфизм длины ДНК. Теория и факты.
13.4. Дивергенция популяций на уровне ДНК.
13.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 13.

РЕЗЮМЕ

1. Нуклеон и гаплотип (нуклеоморфа) соответствуют гену (локусу) и аллелю в обычном генетическом понимании.
2. Гаплотипическое разнообразие *(h)* – это мера изменчивости, которая эквивалентна ожидаемой гетерозиготности (генному разнообразию) на локус.
3. Нуклеотидное разнообразие *(π)* – представляет собой среднее число нуклеотидных различий на сайт между двумя последовательностями.
4. Среднее число замен нуклеотидов для случайно выбранной пары гаплотипов из двух популяций или таксонов – это разница между средним числом замен нуклеотидов на сайт между *i*–тым и *j*–тым гаплотипами (d_{ij}) и средним взвешенным числом замен нуклеотидов внутри популяций (d_x и d_y). Эта величина эквивалентна мере генетического расстояния.
5. Филогения, установленная для видов и генов, не является идентичной. Генные линии обычно разделяются раньше, чем произойдет дивергенция популяций или видов.
6. Кодирующие последовательности структурных генов эволюционируют медленнее, чем не кодирующие участки ДНК, которые маркируют молчащий полиморфизм. Аллели, обнаруживаемые ферментными маркерами, могут иметь более высокий уровень замен нуклеотидов, чем другие аллели, выявляемые по ДНК-маркерам. По гену *Adh* у *D. melanogaster*, к примеру, это обусловлено большей длительностью дивергенции первых.

Chapter 13. DNA PLYMORPHISM WITHIN AND AMONG POPULATIONS

SUMMARY

1. A nucleon and haplotype (necleomorph) correspond to a gene (locus) and allele in a common genetic sense.
2. Haplotypic diversity *(h)* is a measure of variability, which is equivalent to the expected heterozygosity (gene diversity) per locus.
3. Nucleotide diversity *(π)* is a mean number of nucleotide differences per site between two sequences.
4. A mean number of nucleotide substitutions for a randomly sampled pair of haplotypes from two populations or taxa is a difference between a mean number of nucleotide substitutions per site between *i* and *j* haplotypes (d_{ij}) and a mean weighted number of nucleotide substitutions within populations (d_x and d_y). This measure is equivalent to a genetic distance.
5. Phylogeny established for species and genes is not identical. Gene lineages usually have more ancient divergence then divergence of populations or species themselves.

6. Coding sequences of structural genes evolve slower than non coding DNA sequences, which mark silent polymorphism. Alleles detected by enzyme markers may have higher level of nucleotide substitutions than the other DNA markers' alleles. For example, for *Adh* gene in *D. melanogaster* it is caused by a longer history of divergence for the former.

13.1. ПОЛИМОРФИЗМ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК. НУКЛЕОТИДНОЕ РАЗНООБРАЗИЕ

В предыдущих лекциях мы поняли, что белковые маркеры генов обнаружили большой размах генетической изменчивости в природных популяциях. Однако точная оценка изменчивости не может быть выполнена без исследования, проведенного на уровне ДНК.

Для генетики, прежде всего, представляет интерес понимание полиморфизма последовательности ДНК (DNA sequence polymorphism), как результата нуклеотидных замен. Что и как необходимо измерять? Давайте примем, что нуклеотидная последовательность для определенного набора локусов или аллелей в выборке из популяции известна. В этом случае имеются несколько путей оценки степени полиморфизма ДНК.

Число полиморфных последовательностей

Начать измерения можно с простой меры – доли различных или полиморфных последовательностей *(k)* во всей их выборке. Эта мера имеет такие же статистические свойства, как число аллелей в выборке и прямо зависит от размера выборки. Поэтому, для точного измерения уровня полиморфизма, скажем между двумя популяциями или локусами, используя эту меру необходимо иметь выборки одинакового или сравнимого размера. При определении первичной последовательности ДНК (секвенирование, sequence analysis), когда исследуются длинные последовательности ДНК, все сопоставляемые последовательности могут отличаться одна от другой. В этом случае доля полиморфных последовательностей не является вообще подходящей мерой, поскольку все выборки будут одинаково полиморфны или иметь 100% полиморфизма.

Число полиморфных нуклеотидных сайтов

Более подходящая мера полиморфизма ДНК – это доля полиморфных нуклеотидных сайтов на нуклеотидную последовательность (p) (Nei, 1987): $p_n = s_n/m_T$. Здесь s_n и m_T – это число полиморфных сайтов на последовательность и общее число проанализированных нуклеотидов, соответственно. Эта мера лучше, чем доля различающихся последовательностей, поскольку она применима всегда, даже если все проанализированные последовательности различны. Однако эта мера также зависит от размера выборки. В случае модели бесконечного числа нейтральных мутаций (Kimura, 1971), ожидаемое значение p_n в независимых популяциях (с учетом только стохастического процесса) равно (Nei, 1987):

$$E(p_n) = L[1 + 1/2 + 1/3 + \cdots + (n - 1)^{-1}], \qquad (13.1)$$

где n – это число проанализированных последовательностей, а $L=4N\mu$ (Watterson, 1975), N и μ – эффективный размер популяции и темп мутаций на нуклеотидный сайт на поколение, соответственно. Если n велико, $n>20$, то выражение (13.1) приблизительно задается (Nei, 1987):

$$E(p_n) = L[0.577+\log_e(n-1)]. \qquad (13.2)$$

Если рекомбинация отсутствует, то дисперсия p_n для стохастического процесса задается выражением (Nei, 1987):

$$V(p_n) = E(p_n)/m_T + \Sigma_{i=1}^{n-1} L^2\, 1/i^2. \qquad (13.3)$$

Поэтому, возможно оценить $L=4N\mu$, через соотношение (Nei, 1987):

$$L\hat{} = p_n / A, \qquad (13.4)$$

где $A = 1 + 1/2 + \cdots + (n - 1)^{-1}$. Это L^ может быть использовано в качестве меры полиморфизма. Однако следует помнить, что выражение (13.4) правомочно только в случае нейтральности мутаций и когда популяции, находятся в равновесии между дрейфом и мутациями.

Нуклеотидное разнообразие

Лучшей мерой полиморфизма ДНК, чем две предыдущие является среднее число нуклеотидных различий на сайт между двумя последовательностями или **нуклеотидное разнообразие**. Эта мера, также как ее оценка и дисперсия, дается ниже в варианте, предложенном Неем (Nei, 1987):

$$\pi = \Sigma_{ij}\, x_i x_j\, \pi_{ij}\,, \qquad \textbf{(13.5)}$$

где x_i, x_j – это популяционные частоты i-того и j-того типов последовательностей ДНК, а π_{ij} - доля различных нуклеотидов между i-тым и j-тым типами последовательностей ДНК. В панмиктической популяции π является обычной гетерозиготностью на нуклеотидном уровне. Оценка нуклеотидного разнообразия может быть найдена из уравнения:

$$\pi\hat{} = [n/(n-1)]\, \Sigma_{ij}\, x_i\hat{}\, x_j\hat{}\, \pi_{ij} \qquad \textbf{(13.6)}$$

или из

$$\pi\hat{} = \Sigma_{i<j}\, \pi_{ij} / n_c. \qquad \textbf{(13.7)}$$

Здесь n, x_i^, x_j^ и n_c – это число проанализированных последовательностей ДНК, частота i-того и j-того типов последовательности ДНК в выборке и общее число сопоставлений последовательностей, $[n(n - 1)/2]$, соответственно. В выражении (13.7), i и j относятся i-тым и j-тым последовательностям, а не i-тым и j-тым типам последовательностей. Предполагая, что $\pi\hat{}_{ij}$ являются постоянными, дисперсия π^, полученная из (13.6) определяется через соотношение:

$$V(\pi\hat{}) = 4 / n(n - 1)\, [(6 - 4n)\, (\Sigma_{i<j}\, x_i x_j\, \pi_{ij})^2 + (n - 2)\, \Sigma\, x_i x_j\, x_k\, \pi_{ij}\, \pi_{ik} + \Sigma_{i<j}\, x_i x_j\, \pi_{ij}^2] \qquad \textbf{(13.8)}$$

(Nei, Tajima 1981). Эта дисперсия возникает в момент определения частот аллелей и не включает стохастическую компоненту.

Если все мутации нейтральны, то ожидание π^ при стохастическом процессе равно:

$$E(\pi\hat{}) = L/[1 + (4/3)L] \approx L \text{ для } L << 1 \qquad \textbf{(13.9)}$$

(Kimura, 1968). Также, дисперсия π^ при взятии случайных выборок или при стохастическом процессе есть:

$$V(\pi\hat{}) = [(n + 1) / 3n(n - 1)\, m_T]\, L + [2(n2+n+3) / 9(n(n - 1)]\, L^2 \qquad \textbf{(13.10)}$$

(Tajima, 1983a; модифицировано Nei, 1987). По мере увеличения n, это уравнение принимает вид

$$V(\pi\hat{}) = (1 / 3\, m_T)\, L + 2/9L^2. \qquad \textbf{(13.11)}$$

Рассмотрим пример (Nei, 1987). Крейтман (Kreitman, 1983) секвенировал регион гена алкогольдегидрогеназы *(Adh)* для 11 аллелей *Drosophila melanogaster.* Исследованный регион состоял из 2579 нуклеотидов (m_T), исключая делеции и вставки. Девять из одиннадцати последовательностей отличались одна от другой, отмечая высокую степень полиморфизма последовательностей. Обнаружено 43 полиморфных нуклеотидных сайта, так что p_n = 43/2379=0.0181. При допущении нейтральности, поэтому, мы имеем $L\hat{}$ = 0.0062 из уравнения (13.4), так как A = 2.929 при n = 11. Из данных Крейтмана можно вычислить долю различных нуклеотидов для всех пар аллелей. Результаты вычислений представлены в Табл. 13.1.1. Нуклеотидное разнообразие для этого региона гена, поэтому может быть оценено как: $\pi\hat{}$ = 0.0065 ± 0.0017 по любому из уравнений (13.6) или (13.7). Заметим, что x_i = 1/11 для всех, кроме одного гаплотипа (тип последовательности). Этот гаплотип имел выборочную частоту 3/11. Стандартная ошибка оценки была получена с использованием равенства (13.8). Можно отметить, что оценка нуклеотидного разнообразия является близкой к оценке, полученной по числу полиморфных сайтов, при допущении нейтральности.

Таблица 13.1.1. Доля нуклеотидных различий (%) 11 аллелей локуса *Adh* у *Drosophila melanogaster* (*По Kreitman, 1983*)

Table 13.1.1. Percent nucleotide differences between 11 alleles of the *Adh* locus in *Drosophila melanogaster*. The total number of nucleotide sites compared is 2379 (*From Kreitman, 1983*)

Аллель	***(1)***	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***	***(9)***	***(10)***
(1) Wa-S										
(2) F1-1S	0.13									
(3) Af-S	0.59	0.55								
(4) Fr-S	0.67	0.63	0.25							
(5) F1-2S	0.80	0.84	0.55	0.46						
(6) Ja-S	0.80	0.67	0.38	0.46	0.59					
(7) F1-F	**0.84**	**0.71**	**0.50**	**0.59**	**0.63**	**0.21**				
(8) Fr-F	**1.13**	**1.10**	**0.88**	**0.97**	**0.59**	**0.59**	0.38			
(9) Wa-F	**1.13**	**1.10**	**0.88**	**0.97**	**0.59**	**0.59**	0.38	0.00		
(10) Af-F	**1.13**	**1.10**	**0.88**	**0.97**	**0.59**	**0.59**	0.38	0.00	0.00	
(11) Ja-F	**1.22**	**1.18**	**0.97**	**1.05**	**0.84**	**0.67**	0.46	0.42	0.42	0.42

Примечание. Суммарное число сопоставленных нуклеотидных сайтов равно 2379.

По локусу *Adh* у *D. melanogaster* имеются также два электрофоретически выявляемых аллеля (электроморфы), *S* и *F*. Аллели, исследованные Крейтманом, входят в шесть *S* и пять *F* электроморф. Можно оценить нуклеотидное разнообразие отдельно для этих двух электроморф. Оно оказывается: $\boldsymbol{\pi}$^ = 0.0056 для *S* и $\boldsymbol{\pi}$^ = 0.0029 - для *F*. Таким образом, нуклеотидное разнообразие для *S* вдвое выше этой величины для *F*. Таблица 13.1.1 показывает, что нуклеотидное разнообразие между аллелями электроморф *S* и *F* (выделено) значительно больше чем внутри электрофоретических аллелей. Это предполагает, что электрофоретические аллели дивергировали во времени много раньше аллелей, определяющих изменчивость последовательности ДНК в гене.

Доля нуклеотидных различий на сайт или *p*-расстояние

Очень простой и удобной мерой полиморфизма ДНК может являться доля (*p*) нуклеотидных различий на сайт между двумя последовательностями или ***p*-расстояние**. Эта доля может быть оценена как (Nei, Kumar, 2000, p. 33):

$$\boldsymbol{p\hat{} = n_d / n}\ , \qquad \textbf{(13.12)}$$

где n_d - доля различных нуклеотидов между последовательностями ДНК *X* и *Y*, а *n* – это общее число проанализированных нуклеотидов. Дисперсия ***p***^, полученная из (3.12) определяется через соотношение:

$$\boldsymbol{V(p\hat{}) = p\ (p - 1) / n.}\ \textbf{(13.13)}$$

Это обычная биномиальная дисперсия (см. уравнения 9.4 – 9.5).

Оценка числа нуклеотидных замен

Величина расстояния *p* дает удовлетворительную оценку числа нуклеотидных замен на сайт, когда последовательности ДНК близки. Но если *p* велико, то оно дает недооценку числа нуклеотидных замен на сайт, поскольку не учитывает обратные и параллельные мутации. Однако при величинах $p < 0.2$ (Nei, Kumar, 2000, p. 41, Figure. 3.1) эта величина дает весьма близкие значения с более строгими оценками дивергенции.

Для понимания существа процесса замен необходимо использовать определенную математическую модель. Мы проанализировали обоснования самих мер, аналитические выражения для них (модели) и их дисперсии в разной степени подробности по четырем источникам (Nei, 1987; Li, 19997; Nei, Kumar, 2000; Felsenstein, 2004). Одними из часто используемых моделей является модель Джукса-Кантора и двухпараметрическая модель Кимура, которые, соответственно, предполагают равенство темпов замен всех нуклеотидов и различные соотношения для транзиционных (α) и трансверсионных (β) замен. Всего известно не менее восьми различных моделей: 1) Джукса-Кантора (Jukes-Cantor), 2) Двухпараметрическая Кимура (Kimura, *K2P*), 3) Равного ввода (Equal-input), 4) Тамуры (Tamura), 5) Хасигава-Кишино-Яно (Hasigawa-

Kishino-Yano, *HKY*), 6) Тамуры-Нея (Tamura-Nei, *TrN*), 7) Общая реверсионная (General time reversible, *GTR*), 8) Неограниченная модель (Unrestricted).

В модели Кимура (*K2P*) равновесные частоты всех четырех нуклеотидов равны 0.25. Однако предлагаемые алгоритмы применимы вне зависимости от исходных частот (Nei, Kumar, 2000, p.38) и в данном отношении она сходна с моделью Джукса-Кантора и, соответственно, как и первая также приложима к более широкому спектру эмпирических данных по сравнению с шестью другими моделями. Надо подчеркнуть, что в модели Кимура $R = \alpha / 2\beta$, в то время как многие авторы и многие пакеты программ используют иную пропорцию $k = \alpha / \beta$. К этому надо внимательно относится, чтобы избежать нелепых ошибок в сравнениях. Кроме того, проведенный анализ алгоритмов моделей убеждает в необходимости вдумчивого выбора модели при анализе собственных данных. Иногда стоит потратить время и может быть выбрать более сложную модель оценки числа нуклеотидных замен (№ 3-8), вопреки рутинной опции в программе, ведущей к *K2P*, чтобы получить более корректные данные. Хотя надо также заметить, что в более сложных моделях большее число параметров ведет и к относительно большим стандартным ошибкам при их оценивании (Nei, Kumar, 2000). Отметим, что в проанализированном массиве работ наиболее часто (29% по данным таблицы, Приложение) для оценки *p*-расстояния авторы используют модель Кимура, но нередко применяют просто нескорректированную величину *p*^. Имеются примеры использования Модели HKY (Williams et al., 2001; Kontula et al., 2003; Martinez-Navarro et al., 2005), Модели Тамуры-Нея (Kontula et al., 2003; Bertsch et al., 2005) и Общей реверсионной модели (Kontula et al., 2003; Wuster et al., 2005; Kartavtsev et al., 2007 и др.). Численное моделирование, выполненное для бесконечного числа нуклеотидов, показывает, что когда число замен невелико (<20%), то все меры дают весьма близкие значения (Nei, Kumar, 2000, Figure 3.1). Однако по мере увеличения числа замен и нарастания гомоплазии быстрее других дает искажение величина *p*^. Важной корректировкой мер, в связи с гетерогенностью темпа замен в разных участках последовательностей, является коррекция посредством гамма-распределения (Li , Zarkhih, 1995; Nei, Kumar, 2000; Creer et al., 2003; Felsenstein, 2004). Популярной для выбора наиболее подходящей к конкретным эмпирическим данным модели является программа MODELTEST 3.06 (Posada, Grandal, 1998). Очень полезную информацию о свойствах моделей и их применимости в зависимости от специфики данных можно найти в работах (Nei, Kumar, 2000; Hall, 2001; Sanderson, Shaffer, 2003; Felsenstein, 2004). Разные опции (модели) для расчета филогенетических деревьев предоставляются пакетами программ PAUP* (Swofford, 2000), MEGA2, MEGA3 (Kumar et al., 1993; 2000) и другими. Удобный интерфейс и разнообразные статистические возможности, в том числе для анализа последовательностей ДНК и ДНК-маркеров, разработаны в пакете ARLEQUINE (Schneider et al., 2000). Очень хорошее руководство для филогенетического анализа написано Хэллом (Hall, 2001). Это руководство ориентированно в основном на PAUP*, но в нем популярно излагаются и общие принципы филогенетического анализа изменчивости по ДНК.

13.2. ПОЛИМОРФИЗМ ДНК, ОЦЕНЕННЫЙ ПО ДАННЫМ РЕСТРИКТАЗНОГО АНАЛИЗА И СЛУЧАЙНЫМ ПРАЙМЕРАМ

Методики анализа полиморфизма ДНК

Кроме трудоемкого секвенирования, имеются несколько других методик, позволяющих обнаружить изменчивость ДНК. Распространенной методикой исследования уровня полиморфизма ДНК является анализ сайтов рестрикции, по месту действия специфических ферментов эндонуклеаз или рестриктаз. Обычно эта методика сочетается с другой – полимеразной цепной реакцией (ПЦР; polymerase chain reaction – PCR). Полиморфизм обычно изучается для определенного фрагмента последовательности ДНК. Этот фрагмент может включать любой регион ДНК, включая структурные гены, или кодирующие последовательности, интроны, фланкирующие

и межгенные участки. Исследовать также можно короткие фрагменты, включающие только один тип последовательностей ДНК, а также длинные фрагменты ДНК, которые могут включать много типов последовательностей.

Ней и Таджима (Nei, Tajima, 1981) предложили называть весь участок, подвергающийся исследованию **нуклеоном** (nucleon), а конкретную последовательность ДНК изучаемого участка (последовательность рестриционного сайта) – **нуклеоморфой** (nucleomorph). Нуклеон и нуклеморфа соответствуют гену (локусу) и аллелю в обычном генетическом смысле. В литературе, однако, нуклеоморфу чаще называют **гаплотипом**, когда речь идет о мтДНК. По мере необходимости будут использоваться все эти термины.

Рассмотрим три наиболее часто встречающихся варианта использования рестриктазного анализа для определения полиморфизма ДНК: полиморфизм длины рестрикционных фрагментов (ПДРФ; Restriction Fragment Length Polymorphism – RFLP), рэндемично амплифицируемые полиморфные ДНК (Randomly Amplified Polymorphic DNAs – RAPD) и полиморфизм длины амплифицирующихся фрагментов (ПДАФ; Amplified Fragment Length Polymorphism – AFLP).

ПДРФ – полиморфизм длины рестрикционных фрагментов

ПДРФ – это методика, как уже упоминалось, основанная на использовании ферментов эндонуклеаз или рестриктаз, способных разрезать ДНК в определенных местах (сайтах) узнавания. Возникающие при ферментной обработке фрагменты ДНК могут иметь различную длину, если в сайтах узнавания изучаемой ДНК происходили мутации (рис. 13.2.1). Встречаются эндонуклеазы, специфичные к последовательностям из четырех, шести, восьми и десяти пар нуклеотидов. При использовании определенного фермента, можно определить наличие полиморфизма по конкретному сайту узнавания, анализируя длину рестрикционных фрагментов ДНК. Поскольку концентрация таких фрагментов невелика, то обычно используют ПЦР-реакцию для амплификации имеющихся фрагментов и увеличения их числа. После этой процедуры проводится электрофорез фрагментов и устанавливается, есть ли изменчивость в подвижности фрагментов по данному сайту рестрикции. Учитывая наличие многих сотен рестриктаз, число определяемых полиморфных сайтов с помощью ПДРФ потенциально достигает тысяч **молекулярных маркеров изменчивости ДНК**. Наиболее часто в анализе коротких ДНК, таких как мтДНК, используются эндонуклеазы, специфичные к четырем, шести и восьми парам нуклеотидов.

Применение различных комбинаций рестриктаз позволяет идентифицировать более чем два различающихся по размеру фрагментов ДНК. Всякий раз, когда распознаваемые фрагменты не одинаковы по длине, можно говорить, что длина полиморфна. Образцы повторяющейся мтДНК, например, обычно обнаруживают ПДРФ (рис. 13.2.1). Изменчивость мтДНК относится к гаплоидному геному и, поэтому, в результате электрофореза у всех особей обнаруживаются единичные фракции или полосы (рис. 13.2.1). По ядерной ДНК при обнаружении полиморфизма тип изменчивости при электрофорезе может указывать на кодоминантное проявление аллелей (рис. 13.2.2). В представленном примере обнаружено три варианта ДНК (генотипа), хотя по подвижности имеются только два типа аллелей (рис. 13.2.2; см А, В и АВ фенотипы). ПДРФ используются для анализа как внутривидовой, так и межвидовой генетической изменчивости.

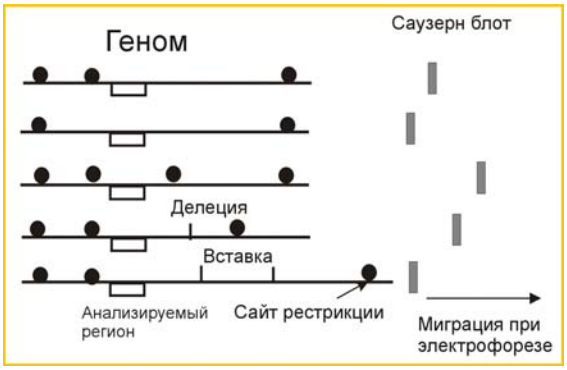


Рис. 13.2.1. Схема эксперимента, позволяющего обнаружить ПДРФ.
Fig. 13.2.1. The schemes of experimental design which permit to obtain RFLP's (left).

Рис. 13.2.2 Иллюстрация изменчивости по единичному сайту ядерной ДНК. Две полосы (фенотип АВ) показывают присутствие гетерозигот в данной выборке особей и кодоминантное проявление аллелей. А и В – гомозиготные особи по разным аллелям.
Fig. 13.2.2. Illustration of single site variation and codominant type of inheritance. Two bands are seen in heterozygotes (right).

РАПД – рэндемично амплифицируемые полиморфные ДНК

РАПД – это молекулярный полиморфизм, идентифицируемый случайно, и используемый для маркирования изменчивости ДНК и картирования генов. По этой методике, после выделения ДНК, не идентифицированные фрагменты тотальной ДНК не обрабатываются рестриктазами, а подвергаются многократным циклам денатурации и синтезу ДНК со случайными праймерами (синтезированные последовательности из 8 нуклеотидов). Сложность ядерных ДНК у эукариот достаточно велика. В результате чего случайные пары сайтов могут оказаться комплементарными, например, по единичным окта- или декануклеотидным последовательностям и быть близко расположенными на ДНК чтобы осуществлялась их амплификации посредством ПЦР (рис. 13.2.3).

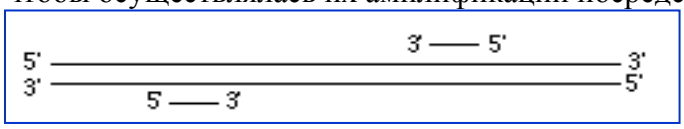


Рис.13.2.3. ДНК и ее фрагмент, который может быть случайно разрезан и амплифицирован.
Fig.13.2.3. DNA and fragments that can randomly be restricted in an individual (top).

Некоторые случайно выбранные декануклеотиды не дают амплифицируемых изменчивых фрагментов (мономорфные полосы). Тогда как другие, обнаруживают такие изменчивые фрагменты ДНК у некоторых особей, но не у всех (рис. 13.2.4). Эти изменчивые полосы и называют РАПД полосами. Три группы полос показаны на рис. 13.2.4. РАПД используются как молекулярные маркеры генетической изменчивости внутри видов и между видами. Однако, ввиду того, что РАПД проявляют доминантность, в том смысле, что невозможно провести различие между гомозиготами и гетерозиготами, их использование при анализе внутривидовой изменчивости ограничено. Ввиду обилия РАПД маркеры могут быть использованы в селекционной практике, частности для картирования ЛКП. РАПД маркеры при сцеплении с другими исследуемыми генами могут служить стартовой точкой в методике «геномной ходьбы».

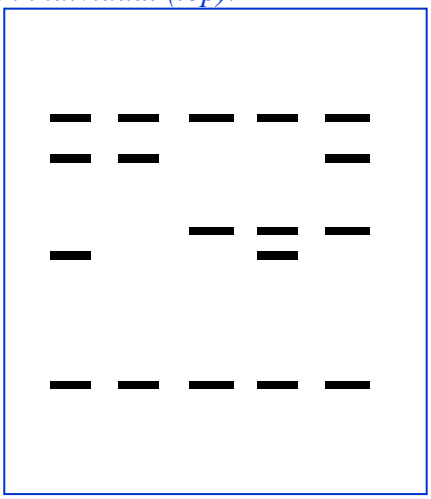


Рис. 13.2.4. Схема электрофореза РАПД фрагментов. Fig. 13.2.4. Scheme of electrophoresis of RAPD fragments (bottom).

ПДАФ – полиморфизм длины амплифицированных фрагментов

ПДАФ – это комбинация ПДРФ и РАПД. На первом этапе ПДАФ анализа геномную ДНК денатурируют рестриктазами, причем такими, которые разрезают ее часто (как, например,

рестриктаза *Mse*I, распознающая 4 пары оснований), и менее часто (например, *Eco*RI, распознающая 6 пар оснований) (рис. 13.2.5). Получающиеся фрагменты лигирут к конеце-специфическим адапторным молекулам. Далее осуществляется пилотная, преселективная (preselective) ПЦР-амплификация с использованием праймеров, комплементарных к каждой из двух адапторных последовательностей, исключая присутствие одного добавочного основания по 3'-концу. Это добавочное основание выбирается экспериментатором. Происходит амплификация только 1/16-той части фрагментов *Eco*RI-*Mse*I. При второй, «селективной» ПЦР, с использованием предыдущих продуктов как затравки, используются праймеры, содержащие два последующих добавочных основания, выбранных экспериментатором. Используемый *Eco*RI адапторный специфичный праймер несет метку (флуоресцентную или радиоактивную). На последнем этапе проводится электрофоретический анализ, который обнаруживает типы фрагментов (фингерпринты), представляющие около 1/4000 часть *Eco*RI-*Mse*I фрагментов (рис. 13.2.5).

ПДАФ могут быть кодоминантными маркерами, как ПДРФ. Кодоминантность проявляется, когда полиморфизм обусловлен последовательностями внутри амплифицируемого участка. И все же, из-за того, что при этой методике выявляется множество полос, требуются дополнительные доказательства принадлежности изменчивости аллелей к отдельным локусам. Кроме того, если полиморфизм обусловлен присутствием-отсутствием основания в изучаемом посредством данного праймера сайте, то изменчивость становится доминантного типа. Аллель вне праймера не будет определяться как полоса. По сравнению с РАПД в ПДАФ методике для обнаружения изменчивых сайтов необходимо меньше праймеров.

Рис. 13.2.5. Пример ПДАФ – полиморфизма длины амплифицированных фрагментов.
Fig. 13.2.5. AFLP, Amplified Fragment Length Polymorphism example.

Число гаплотипов и гаплотипическое разнообразие

Рассмотрим уровень полиморфизма ДНК для определенного региона ядерной ДНК, или целиком мтДНК какого-либо организма в его популяции. Определенное число особей было взято в виде выборки и их ДНК проанализирована с использованием рестриктаз. Эти ДНК далее классифицированы в различные гаплотипы в соответствии с характером рестрикционной изменчивости. Когда различные гаплотипы идентифицированы, то их относительные частоты в выборке могут быть подсчитаны так же, как частоты аллелей по ферментному локусу. Можно обозначить популяционную частоту *i*–того гаплотипа, как x_i, а выборочную частоту – как x_i^.
Полиморфизм ДНК может быть оценен двумя различными способами – на уровне нуклеона и на нуклеотидном уровне. Если рассматривать уровень нуклеона, то полиморфизм ДНК можно изучать без различения нуклеотидных замен и делеций - вставок. Если рассматривать нуклеотидный уровень, то тогда в расчет берут только нуклеотидные замены. Как ранее для анализа последовательностей ДНК, наиболее простой мерой полиморфизма ДНК на уровне нуклеона является число гаплотипов *(k)*, обнаруженное в выборке. Также как в случае с числом различных последовательностей ДНК, это число зависит от размера выборки. Однако более приемлемым, в качестве меры полиморфизма нуклеона, является нуклеонное или **гаплотипическое разнообразие** *(h)* (Nei, Tajima, 1981). Эта мера эквивалентна ожидаемой гетерозиготности (или генному разнообразию), используемой в биохимической популяционной генетике. Она определяется такими же выражениями, 5.1а и 5.1, как даны ранее (Лекция 5,

Практическое занятие 5) и их оценки могут быть получены несколькими путями (Nei, 1975; 1987). Статистические свойства и математическое ожидание для h при модели бесконечного числа аллелей являются точно такими же, как для ожидаемой гетерозиготности.

Среднее число различий по рестрикционным сайтам

Уровень полиморфизма ДНК может быть измерен также средним числом различия сайтов рестрикции между двумя случайно выбранными гаплотипами (Nei, Tajima, 1981). Это число определяется как:

$$v = \Sigma_{ij}\, x_i x_j\, v_{ij}. \qquad (13.14)$$

Здесь v_{ij} – это число различий по сайтам рестрикции между i-тым и j-тым гаплотипами и суммирование проводится по всем комбинациям гаплотипов. Оценка $\boldsymbol{v}$ и ее выборочная дисперсия задаются уравнениями (13.6) и (13.8), соответственно, если заменить $\boldsymbol{\pi_{ij}}$ на $\boldsymbol{v_{ij}}$. Для нейтральных гаплотипов вследствие нуклеотидных замен, ожидание для $\boldsymbol{v}$ при модели бесконечного числа мутаций есть:

$$E(v) = M, \qquad (13.15)$$

где $M = 4Nv$ и v – есть темп мутаций на нуклеон на уровне рестрикционных сайтов. С другой стороны, стохастическая дисперсия v^ задается примерно через соотношение (13.10), если заменить L на M и удалить m_T. Удаление m_T является необходимым, поскольку сейчас исследуется изменчивость на нуклеон, а не на нуклеотидный сайт (Tajima, 1983a).

Число полиморфных сайтов рестрикции

Величина близко связанная с $\boldsymbol{v}$ – это число полиморфных сайтов рестрикции (s_r), то есть, - это число сайтов рестрикции, которые полиморфны в выборке n нуклеонов. Эта мера очень сильно зависит от размера выборки и поэтому не вполне удобна для сравнений.

Нуклеотидное разнообразие

Основная проблема с использованием двух мер, представленных выше, связана с их зависимостью от размера исследованного фрагмента ДНК; значения мер увеличиваются с увеличением размера фрагментов. На практике, размер фрагментов сильно варьирует в зависимости от исследуемого гена или региона ДНК, поэтому они не могут быть использованы, как общие меры полиморфизма.

Чтобы обойти эту проблему, лучше измерять полиморфизм на нуклеотидном уровне. Когда основу полиморфизма по сайтам рестрикции составляют нуклеотидные замены, то такой мерой может быть нуклеотидное разнообразие согласно выражению (13.5). При использовании рестриктазной методики, π_{ij} в (13.5) может быть оценено посредством p^ = 1 – S^1/r или другим выражением (5.50; Nei, 1987), где S^=m_{XY}/m^, m^=$(m_X+m_Y)/2$ и m_X, m_Y, - число сайтов рестрикции для последовательностей ДНК X и Y, m_{XY} – число общих сайтов рестрикции в двух последовательностях. Таким образом, если выборочные частоты гаплотипов получены, то $\boldsymbol{\pi}$ может быть оценено через (13.6) или (13.7).

Как упоминалось ранее, ожидаемое значение $\boldsymbol{\pi}$ для модели бесконечного числа нейтральных мутаций, равно $L = 4N\mu$, тогда как ожидаемое значение v равно $M = 4Nv$. Поэтому, $\boldsymbol{\pi}$ может быть оценено через v, если известна взаимосвязь между μ и v. Было показано (Nei, Tajima, 1981), что

$$v = 2\Sigma_{i=1}^{s}\, m_i r_i\, \mu, \qquad (13.16)$$

где m_i и r_i – это число сайтов рестрикции и число нуклеотидов в распознаваемой последовательности i-той рестриктазы, соответственно, а s – число исследованных ферментов. Поэтому, $\boldsymbol{\pi}$ может быть оценено через соотношение (Nei, 1987):

$$\pi = v / R, \qquad (13.17)$$

где $R = v = 2\Sigma_{i=1}^{s}\, m_i r_i\, \mu$. Если использовать (13.4), то $\boldsymbol{\pi}$ может быть также оценено из s_r (Nei, 1987):

$$\pi\text{^} = s_r / (AR), \qquad (13.18)$$

где $A = 1 + 1/2 + \cdots + (n - 1)^{-1}$. Имеется также другая формула для оценки π (Engels, 1981):

$$\pi = (nc - \Sigma c_i^2) / [rc (n - i)]. \quad (13.19)$$

Где c_i – это число членов выборки, по сайту рестрикции i (i = 1,…, s_r), а $c = \Sigma c_i$. Эта формула не требует допущения нейтральности, как подчеркнуто автором (Engels, 1981), но зависит от допущения равновесия по сцеплению между сайтами рестрикции. В контрасте с этим, π^, получаемое из (13.6) или (13.15), не требует допущения ни нейтральности, ни равновесия по сцеплению (Nei, 1987). Однако, когда число проанализированных последовательностей ДНК велико, скажем, $n > 100$, то выражение (13.17) видимо будет более легким для вычисления, чем (13.6) или (13.15). В цитированной работе (Engels, 1981) была представлена также формула для оценки дисперсии **π^,** получаемой исходя из (13.17), но вследствие зависимости от равновесия по сцеплению, она должна давать недооценку.

Рассмотрим пример (Nei, 1987). У трех видов дрозофил *(D. melanogaster, D. simulans* и *D. virilis)* была исследована изменчивость мтДНК с использованием рестриктаз *Hae*III (GGCC), *Hpa*ll (CCGG), *Eco*RI (AATTC), and *Hin*dlll (AAGCTT) (Shah, Langley, 1979). Исследователи идентифицировали семь гаплотипов и данные по этим гаплотипам для каждого вида приведены в Табл. 13.2.1. У *D. simulans,* были проанализированы только пять нуклеонов (последовательностей ДНК) и полиморфизм не был обнаружен вообще. У *D. melanogaster* было обнаружено четыре гаплотипа и оценка нуклеотидного разнообразия составила: h = 0.71±0.04 (Table 13.2.2). Из этого значения, можно оценить $M=4N\,v$ по $M=h/(1—h)$ при допущении отсутствия отбора. M равно 2.46. С другой стороны, оценка M, полученная исходя из числа гаплотипов *(k)* [при использовании выражения (8.16; Nei, 1987)] составляет 1.95, что значительно меньше, чем оценка, полученная на основе нуклеотидного разнообразия. Это различие, очевидно, возникает вследствие того, что нуклеонное разнообразие (гетерозиготность) имеет тенденцию давать переоценку M, когда оно основывается на единичном локусе (Zouros, 1979).

Таблица 13.2.1. Частоты гаплотипов в выборках мтДНК при исследовании трех видов рода *Drosophila (Данные из Shah, Langley, 1979)*

Table 13.2.1. Haplotype frequencies in samples of mtDNAs from three species of *Drosophila; n* is number of mtDNAs sampled *(Data from Shah, Langley, 1979)*

Гаплотип	***m***	m_a	m_b	m_c	***s***	***v***	v_d	***(n)***
D. melanogaster	0.1	0.3	0.5	0.1				10
D. simulans					1.0			5
D. virilis						0.6	0.4	10

Примечание. n – это число исследованных мтДНК.

Таблица 13.2.2. Гаплотипическое разнообразие *(h^),* число гаплотипов *(k),* среднее число различий сайтов рестрикции *(v^),* число полиморфных сайтов *(s_r)* и нуклеотидное разнообразие *(π)* у *Drosophila melanogaster* и *D. virilis (По Nei, 1987)*

Table 13.2.2. Haplotypic diversity *(h^),* number of haplotypes *(k),* average number of restriction-site differences *(v^),* number of segregating sites *(s_r),* and nucleotide diversity *(π)* in *Drosophila melanogaster* and *D. virilis (From Nei, 1987)*

Вид, значение	***h^***	***k***	***v^***	s_r	π
D. melanogaster					
Оценка	0.71±0.12	4	1.22±0.27	3	0.008±0.002
M=4Nv	2.46	1.95	1.22	1.06	
L= 4Nμ	0.017	0.013	0.008	0.007	0.008
D. virilis					
Оценка	0.53±0.09	2	0.53±0.09	1	0.004±0.001
M=4Nv	1.144	0.43	0.53	0.35	
L=4Nμ	0.0088	0.003	0.004	0.002	0.004

Чтобы вычислить среднее число различий по сайтам рестрикции, удобно представить данные в виде таблицы числа различий сайтов рестрикции *(v_{ij})* для всех пар гаплотипов (Табл. 13.2.3). Эти числа могут быть получены на основании рисунка 1 в опубликованной статье (Shah, Langley, 1979). Из значений v_{ij} для *D. melanogaster* в Табл. 13.2.3, находим v^ =1.22 при использовании (10.5) и π_{ij} замененными на v_{ij}. При допущении нейтральности гаплотипов, эта величина является оценкой *4Nv.* Она все еще меньше соответствующей оценки, полученной на основе числа гаплотипов. Изменчивость мтДНК у *D. melanogaster* вызвана полиморфизмом по трем сайтам рестрикции. То есть, у этого вида s_r=3. Если мы приравняем эту величину к ожиданию в (13.2), с *L* замененным на *M,* то получим 3=2.83M, так как *n*=10. Поэтому, имеем еще одну оценку M= 1.06, которая близка к оценке исходя из v^.

Таблица 13.2 3. Различия сайтов рестрикции *(v_{ij})* и значения *S^$_{ij}$* для пар гаплотипов *(По Nei, 1987)*

Table 13.2 3. Restriction-site differences *(v_{ij})* and *S^$_{ij}$* values for pairs of haplotypes. The figures above the diagonal are *v_{ij}'s* and those below the diagonal are *S^$_{ij}$*'s. The upper *S^$_{ij}$* value for each pair of haplotypes is for *Hae*III and *Hpa*II *(r = 4),* whereas the lower *S^$_{ij}$* value is for *Eco*RI and *Hin*dIII (r=6) *(From Nei, 1987)*

Гаплотип	***m***	***m_a***	***m_b***	***m_c***	***s***	***v***	***v_d***
M		1	1	1	11	13	14
m_a	0.93 1.00		2	2	10	12	13
m_b	1.00 0.94	0.93 0.94		2	12	14	15
m_c	1.00 0.94	0.93 0.94	1.00 0.89		12	14	15
m_c	0.43 0.84	0.46 0.84	0.43 0.80	0.43 0.80		14	15
V	0.18 0.71	0.20 0.71	0.18 0.67	0.18 0.67	0.22 0.59		1
v_d	0.18 0.67	0.20 0.67	0.18 0.63	0.18 0.63	0.22 0.56	1.00 0.92	

Примечание. Цифры выше диагонали представляют v_{ij}-тые, а цифры ниже диагонали - *S*^-тые. Верхние *S*^ значения для каждой пары гаплотипов принадлежат *Hae*III и *Hpa*II (r = 4), тогда как нижние *S*^ значения принадлежат *Eco*RI и *Hin*dIII (r=6).

Давайте теперь свяжем *4Nv* с *4Nμ* при допущении, что все эволюционные изменения в этих двух видах возникли за счет замен нуклеотидов, и все они были нейтральными. Среднее число сайтов рестрикции для *Hae*II, *Hpa*II, *Hin*dIII и *Eco*RI для четырех гаплотипов *D. melanogaster* составляют 3.7, 4, 4.6 и 4, соответственно. Поэтому, имеем взаимосвязь: $v=2\Sigma\, m_i r_i\, \mu = 2 * [(3.7 + 4) * 4 + (4.6 + 4) * 6]\, \mu = 164.8\, \mu$. Сходно, находим $v=180\, \mu$ для *D. simulans* и v=100.8 μ для *D. virilis.* Среднее этих оценок представляет собой 149 μ. Надо заметить, что содержание G+C пар в мтДНК *Drosophila* составляет около 0.22 даже в случае, если A-T богатый регион исключен (Kaplan, Langley, 1979). Поэтому *Hae*III (GGCC) и *Hpa*II (CCGG) дают не много сайтов рестрикции. В любом случае, если использовать взаимосвязь v=149μ, то оценка *4Nμ* может быть получена исходя из значения *4Nv*. Эта величина варьирует от 0.007 до 0.017, но последнее значение, которое было получено из *h*, является, вероятно, переоценкой по причинам, отмеченным ранее.

Для оценки нуклеотидного разнообразия $\boldsymbol{\pi}$ необходимо сначала вычислить долю общих сайтов рестрикции для каждой пары гаплотипов, используя формулу $S^{\wedge}_{ij} = 2m_{ij} / (m_i + m_j)$, где m_i и

m_j – это числа сайтов рестрикции для i-того и j-того гаплотипов, соответственно, а m_{ij} – число общих сайтов рестрикции (см. выше). При этом, оценки π_{ij} задаются соотношением $\pi_{ij} = (-\log_e S\hat{}_{ij})/r$. Когда используются два или более ферментов с одинаковыми значениями r, тогда $S\hat{}_{ij}$ –тые должны вычисляться с объединением m_i, m_j и m_{ij} по всем ферментам. Если r не являются одинаковыми, то $S\hat{}_{ij}$ –тые должны вычисляться раздельно. В Табл. 13.2.3, $S\hat{}_{ij}$ –тые вычислены раздельно для *Hae*III и *Hpa*II (r=4) и для *Eco*RI и *Hin*dIII (r=6). Из этих значений, $\boldsymbol{\pi}\hat{}$ можно вычислить таким же способом, как вычислялись $v\hat{}$. У *D. melanogaster*, $\boldsymbol{\pi}\hat{}$ составляет 0.0080 для ферментов с r=4 и 0.0076 для ферментов с r=6, со средним, равным 0.008. Это значение является другой оценкой $4N\mu$ и близко к оценке, полученной из $v\hat{}$ и s_r. У *D. virilis*, были проделаны те же самые вычисления, а полученные значения представлены в Табл. 13.2.2. Все оценки генетической изменчивости у этого вида меньше, чем аналогичные величины у *D. melanogaster.*

13.3. ПОЛИМОРФИЗМ ДЛИНЫ ДНК. ТЕОРИЯ И НАБЛЮДЕНИЯ

Данные анализа ДНК показывают, что существенная часть ее полиморфизма обусловлена делециями и вставками (Langley et al., 1982; Bell et al., 1982; Chapman et al., 1986; Льюин, 1987; Жимулев, 2002; Klug, Cummings, 2002). Полиморфизм ДНК по длине фрагментов также во многом вызывается неравным кроссинговером (Coen et al., 1982; Жимулев, 2002; Klug, Cummings, 2002). Полиморфизм такого типа обычно обнаруживается по присутствию интервалов или «пропусков» (gaps) между двумя сравниваемыми ДНК или рестрикционными последовательностями. Один из методов, приемлемых для оценки этого полиморфизма – это оценка гаплотипического разнообразия. Очевидно, что эта мера зависит от длины исследованной ДНК. Более подходящей мерой является среднее число нуклеотидов в пропусках на нуклеотидный сайт между двумя случайно выбранными гаплотипами.

Пусть m_{gij} будет число нуклеотидов в пропусках между i-тым и j–тым гаплотипами, а m_{Tij} – общее число сопоставляемых нуклеотидов, включая пропуски. Число нуклеотидов в пропусках на нуклеотидный сайт (g_{ij}) между двумя гаплотипами тогда задается соотношением $g_{ij} = m_{gij} / m_{Tij}$, тогда как среднее для g_{ij}, взвешенная на m_{Tij}, для всех комбинаций гаплотипов составит (Nei, 1987):

$$\boldsymbol{G = \bar{m}_g / \bar{m}_T}, \qquad \textbf{(13.20)}$$

где $\bar{m}_g$ и $\bar{m}_T$ – средние для m_{gij} и m_{Tij}, соответственно. Отметим, что $\bar{m}_g$ в (13.20) эквивалентно (13.6), так что дисперсия $\bar{m}_g$ может быть получена заменой $\boldsymbol{\pi_{ij}}$ на m_{gij} в (13.8). Дисперсия для g может быть найдена приблизительно, допуская, что $\bar{m}_T$ является константой, так как $\bar{m}_T$ является обычно много большей, чем $\bar{m}_g$.

Рассмотрим еще один пример (Nei, 1987). У *D. melanogaster* был проанализирован рестрикционный полиморфизм в одном из регионов гена алкогольдегидрогеназы (Langley et al., 1982) и идентифицировано шесть гаплотипов ДНК (Табл. 13.3.1). Эти гаплотипы очевидно обусловлены делециями, произошедшими по наиболее частому гаплотипу (N1), длина которого составляет около 12000 пн. Все делеции были обнаружены в некодирующем регионе ДНК. Гаплотипическое разнообразие для этого региона ДНК составляетs 0.682 ±0.111, что немного меньше, чем гаплотипическое разнообразие (0.853), полученное для полиморфизма по сайтам рестрикции для этого же региона гена без делеций (Langley et al., 1982). Чтобы оценить g, необходимо сначала вычислить g_{ij} для всех пар гаплотипов. Например, g_{12} задается как m_{g12}/m_{T12} = 20/12000 = 0.002. В данном случае, все делеции - вставки не перекрываются, так что m_T = 12000 для всех g_{ij}. После того как g_{ij} вычислены, g можно получить исходя из (13.18). Оно становится

равным, g =0.0175 ± 0.009. Эта величина почти в три раза больше, чем нуклеотидное разнообразие (0.006) для того же самого региона гена (см. Табл. 13.3.2).

Таблица 13.3.1. Полиморфизм по делециям и вставкам в неструктурном регионе гена алкогольдегидрогеназы у *Drosophila melanogaster (По Nei, 1987; данные из Langley et al., 1982)*

Table 13.3.1. Deletion/insertion polymorphisms in the alcohol dehydrogenase gene region of *Drosophila melanogaster (From Nei, 1987; data from Langley et al., 1982)*

Гаплотип		***Делеции - вставки***				
Наименование	***Наблюда-емое число***	***Δa 20 bp***	***Δb 550 bp***	***Δc 900 bp***	***Δd 180 bp***	***Δf 30 bp***
N1 (Стандартный)	**10**	-	-	-	-	-
N2	**2**	+	-	-	-	-
N3	**1**	-	+	-	-	-
N4	**1**	-	-	+	-	-
N5	**2**	-	-	-	+	-
N6	**2**	-	-	-	-	+

Некоторые комментарии к полиморфизму ДНК

Нуклеотидное разнообразие. В предыдущих разделах мы встретили несколько способов измерения полиморфизма ДНК. Если не рассматривать эффект делеций и вставок, то наиболее подходящей мерой полиморфизма является нуклеотидное разнообразие. Эта мера не зависит от длины фрагментов ДНК и размера выборки и поэтому может использоваться для сравнения изменчивости ДНК в различных генах и популяциях.

Нуклеотидное разнообразие было изучено для многих генов (Табл. 13.3.2). Большинство оценок были получены с использованием рестрикционного анализа. В Табл. 13.3.2 также включены данные о доле нуклеотидных различий (π_{ij}) между выбранными парами аллелей в последовательностях ДНК. Нуклеотидное разнообразие варьирует от 0.002 до 0.019 для эукариот и мало отличается между генами митохондриальных и ядерных ДНК. Если вспомнить, что обычный ген содержит несколько тысяч пар нуклеотидов, то можно предположить, что две нуклеотидные последовательности двух генов, случайно выбранные из популяции, очень редко могут быть сходными. Большинство мтДНК имеют материнское наследование и существуют в форме гаплотипов. Вследствие этого *Ne* для мтДНК ожидаются как 1/4величины, полученной на основании изменчивости ядерных генов.

Однако, темп мутаций по мтДНК очевидно значительно выше, чем для ядерного генома (Лекция 6). Эти два компенсирующих фактора, вероятно, ведут к тому, что π для митохондриальных и ядерных генов почти одинаков. В некоторых случаях, однако, нуклеотидное разнообразие по-видимому весьма низко. Например, мтДНК индейцев из Венесуэлы полностью мономорфна (Johnson et al., 1983). Очевидно, что эта популяция претерпела относительно недавно эффект бутылочного горлышка. В отличие от генов эукариот, ген гемагглютинина вируса гриппа А обнаруживает очень высокое нуклеотидное разнообразие. Это высокое нуклеотидное разнообразие трудно объяснить чем-либо иным, чем необычно высоким темпом мутаций у этого вируса.

Таблица 13.3.2. Оценки нуклеотидного разнообразия *(π)* или доли нуклеотидных различий между выбранными парами последовательностей ДНК *(π_{ij}) (По Nei, 1987)*
Table 13.3.2. Estimates of nucleotide diversity *(π_i)* or the proportion of nucleotide differences between a selected pair of DNA sequences *(π_{ijj}) (From Nei, 1987)*

ДНК или регион гена	***Организм***	***Метод***	***n***	***пн***	***π или π_{ij}***
Нуклеотидное разнообразие (π)					
мтДНК	*Человек* [d]	*R*	100	16500	0.004
мтДНК	*Шимпанзе* [d]	*R*	10	16500	0.013
мтДНК	*Горилла* [d]	*R*	4	16500	0.006
мтДНК	*Полевая мышь* [d]	*R*	19	16500	0.004
мтДНК	*Дрозофила* [d]	*R*	10	11000	0.008
β-глобин	*Человек* [d]	*R*	50	35000	0.002
Гормон роста	*Человек* [a]	*R*	52	50000	0.002
Регион гена Adh	*Дрозофила* [d]	*R*	18	12000	0.006
Кодирующий регион Adh	*Дрозофила* [b]	*S*	11	765	0.006
Ген гистона H4	*Морской еж* [c]	*S*	5	1300	0.019
Гемагглютинин	Вирус гриппа А [d]	*S*	12	320	0.510
Выбранные пары последовательностей ДНК					
Инсулин	*Человек* [d]	*S*	2	1431	0.003
Ген иммуноглобулина $C_к$	*Крыса* [d]	*S*	2	1172	0.018
Ген иммуноглобулина IgG2a	*Мышь* [d]	*S*	2	1114	0.100

Примечание. *n* – размер выборки. пн*:* пар оснований. *R* – методика рестрикционного анализа. *S* – секвенирование ДНК. Источник: [a] Chakravarti et al., 1984; [b] Kreitman, 1983; [c] Yager et al., 1984; [d] Nei, 1983.

Молчащий полиморфизм. Интересно проанализировать полиморфизм, который не проявляется на фенотипическом уровне, даже на простейшем уровне – аминокислотной последовательности в белках. В теории нейтральности ожидается, что этот, так называемый «молчащий» полиморфизм, будет более частым, чем полиморфизм, реализующийся на уровне аминокислотных последовательностей. Это должно происходить вследствие того, что «молчащие» мутации будут с меньшей интенсивностью подвергаться очищающему отбору, чем не молчащие. Наоборот, если полиморфизм в основном поддерживается отбором (в случае, когда эффект дрейфа пренебрежимо мал), то можно ожидать, что молчащий полиморфизм будет меньшим, чем не молчащий полиморфизм. Один из способов обосновать эти предположения – это проанализировать полиморфизм по первому, второму и третьему положению в кодоне в структурных генах. Дискуссия на эту тему представлена ранее (Лекция 6).

13.4. ДИВЕРГЕНЦИЯ ПОПУЛЯЦИЙ НА УРОВНЕ ДНК

Дивергенция ДНК между популяциями

Для того, чтобы оценить уровень дивергенции ДНК между двумя близко родственными популяциями или видами необходимо рассматривать влияние полиморфизма на ДНК. Это может быть сделано аналогично тому, как мы поступали при оценке генетических расстояний.

Нуклеотидные различия между популяциями. Предположим, что для определенного региона ДНК имеется *m* различных гаплотипов в популяциях *X* и *Y*. Пусть также x_i^ и y_i^ – являются выборочными частотами *i*–того гаплотипа в популяциях *X* и *Y*. Среднее число замен

нуклеотидов для случайно выбранной пары гаплотипов (d_X) в популяции X может быть оценено из (Nei, 1987):

$$d_X\hat{} = [n_X / (n_X - 1)] \sum_{ij} x_i\hat{}x_j\hat{}\, d_{ij}\,, \qquad (13.21)$$

где n_X – это число последовательностей в выборке, а d_{ij} - это число нуклеотидных замен на сайт между i–тым и j–тым гаплотипами. Величину d_{ij} можно оценить исходя из равенства $d\hat{}_{ij}=(-\log_e S\hat{}_{ij})/r$ или другим способом [например, Nei,1987; выражение (5.3)]. Когда последовательности ДНК различны, то $x_i\hat{} = 1/n_X$ (другие обозначения даны в Разделе 13.2). Среднее число замен нуклеотидов (d_Y) для Y можно оценить таким же способом. С другой стороны, среднее число замен нуклеотидов (d_{XY}) между гаплотипами ДНК из X и Y может быть оценено из равенства (Nei, 1987):

$$d_{XY}\hat{} = S_{ij}\, x_i\hat{}x_j\hat{}\, d_{ij}\,, \qquad (13.22)$$

где d_{ij} – это замены нуклеотидов между i–тым гаплотипом из X и j–тым гаплотипом из Y. Число всех нуклеотидных замен между двумя популяциями (d_A) тогда можно оценить из соотношения (Nei, 1987):

$$d_A = d_{XY}\hat{} - (d_X\hat{} + d_Y\hat{}) / 2. \qquad (13.23)$$

Если темп замен нуклеотидов постоянен и равен 1 на сайт в год, а время с момента дивергенции между двумя популяциями есть T, то тогда ожидаемое значение (d_A) из $d_A\hat{}$ равно (Nei, 1987):

$$d_A = 2\lambda T. \qquad (13.24)$$

Смысл уравнений (13.23) и (13.24) состоит в том, что в присутствие полиморфных последовательностей среднее время с момента дивергенции между аллелями из X и из Y больше, чем время расхождения популяций, как это следует из данных рис. 13.4.1. Поэтому, для оценки времени (T) расхождения популяций необходимо вычесть из d_{XY} средние нуклеотидные различия для полиморфных аллелей на момент расхождения популяций. Среднее нуклеотидное различие оценивается исходя из $(d_X\hat{} + d_Y\hat{}) / 2$ при допущении, что ожидаемые значения $d_X\hat{}$ или $d_Y\hat{}$ являются одинаковыми для всего эволюционного процесса (Nei, 1987; Chapter 13).

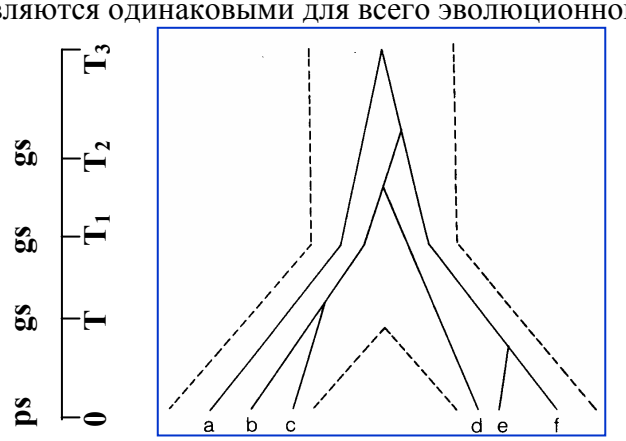


Рис. 13.4.1. Диаграмма, показывающая, что при существовании полиморфизма время дивергенции генов (gs) является обычно большим, чем время дивергенции популяций (ps). a – f – это различные генные линии, часть из которых (например, a, b и f) дивергировали задолго до расхождения популяций (По Nei, 1987).
Fig. 13.4.1. Diagram showing that the time of gene splitting (gs) is usually earlier than the time of population splitting (ps) when polymorphism exists. a – f are different gene lineages, part of which (e.g., a, b and f) diverged long before population splitting (From Nei, 1987).

Рассмотрим еще один пример (Nei, 1987). Ген алкогольдегидрогеназы был секвенирован у двух видов дрозофил *D. simulans* и *D. mauritiana*, являющимися близнецовыми к *D. melanogaster* (Bodmer, Ashburner, 1984; Cohn et al., 1984). Состав последовательностей, представленных двумя группами исследователей немного отличается и, очевидно, что эти различия обусловлены наличием полиморфизма внутри видов. Стифенс и Ней (Stephens, Nei, 1985) исследовали межвидовые и внутривидовые различия этих последовательностей, совместно с данными для *D. melanogaster* (Kreitman, 1983), рассматривая общий регион в 822 нуклеотида. Полученные

результаты представлены в Табл. 13.4.1. Значения d_A предполагают, что *D. simulans* и *D. mauritiana* являются более близкими друг к другу видами, чем к *D. melanogaster* и, соответственно, что видообразование произошло сначала между *D. melanogaster* и группой «*simulans-mauritiana*».

Если допустить постоянство темпа нуклеотидных замен, то можно оценить время дивергенции между этими видами. Средний темп замен нуклеотидов для кодирующих участков генов составляет около 2×10^{-9} на сайт в год ($0.71 \times 0.88 \times 10^{-9} + 0.29 \times 4.65 \times 10^{-9} \approx 2 \times 10^{-9}$) (см. Лекция 6). Если использовать это значение, то оценка времени дивергенции между видами *D.simulans* и *D. mauritiana* становится: $T = 0.0086 / (2 \times 2 \times 10^{-9}) = 2.1 \times 10^6$ лет согласно выражению (13.22). Это время дивергенции только немного больше времени (1.8 млн. лет) для расхождения двух полиморфных аллелей ($d_X = 0.0073$) у *D. simulans.* С другой стороны, время дивергенции между *D. melanogaster* и *D. simulans* и между *D. melanogaster* и *D. mauritiana* составляет 4.4×10^6 и 5.8×10^6 лет, соответственно. Среднее равно 5.1 млн. лет. Поэтому, дивергенция между *D. melanogaster* и группой «*simulans-mauritiana»,* очевидно, возникла в три раза раньше, чем дивергенция между *D. simulans* и *D. mauritiana.*

Таблица 13.4.1. Оценки межпопуляционных (d_{XY}), внутрипопуляционных (d_X или d_Y), и всех (d_A) нуклеотидных различий между тремя видами *Drosophila* для гена алкогольдегидрогеназы (*По Stephens, Nei, 1985)*

Table 13.4.1. Estimates of interpopulational (**d_{XY}**), intrapopulational (**d_X** or **d_Y**), and net (d_A) nucleotide differences among three species of *Drosophila* for the alcohol dehydrogenase gene (822 nucleotides examined). The figures on the diagonal refer to **d_X** (or **d_Y**), and those below the diagonal **d_{XY}**. The figures above the diagonal represent the values of **$d_A = d_{XY} - (d_X + d_Y) / 2$**. All values should be divided by 100 *(From Stephens, Nei, 1985)*

Вид	*D. melanogaster*	*D. simulans*	*D. mauritiana*
D. melanogaster	0.70±0.4l	1.73±0.28	2.36±0.24
D. simulans	2.45±0.49	0.73±0.30	0.86±0.22
D. mauritiana	2.96±0.54	1.47±0.38	0.49±0.24

Примечание. Проанализировано 822 нуклеотида. Цифры по диагонали относятся к d_X (или d_Y), а цифры ниже диагонали – к d_{XY}. Цифры выше диагонали представляют значения $d_A = d_{XY} - (d_X + d_Y) / 2$. Все значения должны быть разделены на 100.

Генеалогические взаимосвязи генов внутри и между популяциями

Имеется одно очень важное преимущество анализа полиморфизма ДНК по первичным последовательностям или сайтам рестрикции по сравнению с генетико-биохимическим ферментным полиморфизмом. А именно, филогенетические взаимосвязи можно оценивать по отдельному гену или фрагменту ДНК.

Выше в этой лекции были представлены оценки числа замен нуклеотидов для 11 аллелей по локусу *Adh* у *D. melanogaster.* Используя эти и некоторые дополнительные данные, был проведен филогенетический анализ этих аллелей (Kreitman, 1983; Stephens, Nei, 1985). Эти результаты показали, что в среднем *S* аллели являются более древними, чем *F* аллели. При оценке времени получается, что *F* аллель возник из *S* аллеля около миллиона лет назад. То есть, *S-F* полиморфизм по гену *Adh* является достаточно древним. Этот древний полиморфизм, по мнению Нея (Nei, 1987; Chapter 13) может быть объяснен без участия балансирующего отбора. Ввиду очень большого эффективного размера популяций у этих видов, составляющего порядка 2 миллионов, вся изменчивость может быть обусловлена накопленными мутациями, которые инкорпорированы дрейфом.

Если филогенетическое дерево построено на основе генов, выборки которых сделаны из нескольких связанных миграцией популяций, то теоретически мы будем иметь дело со смешанными генеалогиями (Malecote, 1970). Реальный пример такого смешения генеалогий можно наблюдать по данным мтДНК на расах человека, активно перемешивающихся в настоящее время

(рис. 13.4.2). Характер кластеризации представителей разных рас в филогенетическом дереве однозначно убеждает в «смешанности» ветвей (рис. 13.4.2). Близкие результаты были получены в других исследованиях человека (Cann et al., 1982; Cann, 1982) и такую неоднозначную кластеризацию интерпретировали, как свидетельство миграции между расами (Cann, 1982). Надо заметить, однако, что генеалогическое смешение мтДНК для различных популяций является ожидаемым, даже при отсутствии миграция, если предковая популяция была полиморфна и время, прошедшее после дивергенции относительно мало. Причины для этого были объяснены ранее (см. рис.13.4.1). Самая древняя дивергенция мтДНК, около 300 тыс. лет назад, принадлежит представителям монголоидной расы и негроидов, если исключить одну особь (рис. 13.4.2). Очевидно, что дивергенция мтДНК происходила ранее расхождения самих рас, как это следует из их дивергенции, оцененной по другим генам (Nei, 1987).

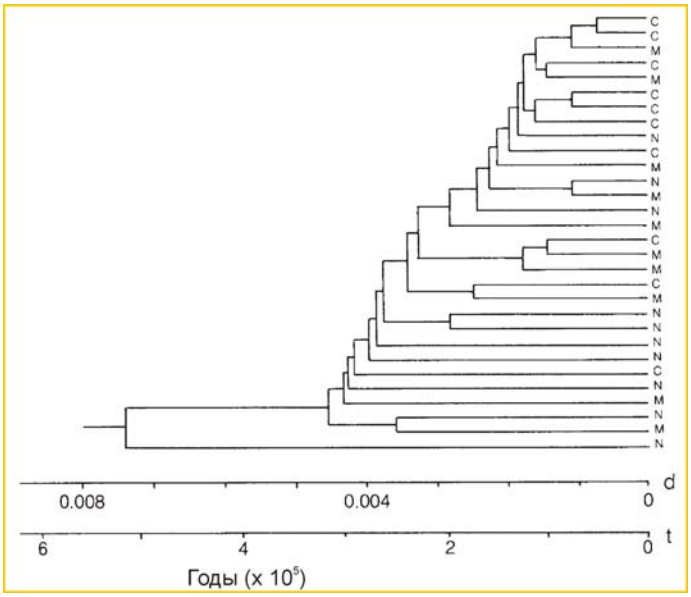


Рис. 13.4.2. Филогенетическое дерево мтДНК 10 особей, представленных кавказоидами (C), негроидами (N) и монголоидами (M); d, t – шкала расстояний и шкала времени, соответственно (По Nei, 1985).
Fig. 13.4.2. Phylogenetic tree of mtDNAs from 10 individuals from each of Caucasoid (C), Negroid (N), and Mongoloid (M) (From Nei, 1985).

Интрогрессия митохондриальной ДНК

Сформированные виды являются репродуктивно изолированными и, как правило, между ними поток генов отсутствует. Недавно, однако, многие авторы представили данные о том, что мтДНК пересекает границы видов. Так, в южной Дании имеется гибридная зона между двумя видами домовых мышей, *Mus musculus* и *M. domesticus* (ранее рассматривались как два подвида). Северная Дания населена в основном *M. musculus.* При анализе мтДНК этих видов было обнаружено (Ferris et al., 1983), что в отличие от формы *M. musculus* в Восточной Европе, в Дании *M. musculus* к северу от гибридной зоны имеют также мтДНК формы *M. domesticus*, которая показывает 5% дивергенции первичной последовательности мтДНК от *M. musculus*. Эта часть *M. domesticus* ограничена северной Данией и некоторыми частями Швеции. Поэтому интрогрессия мтДНК из *M. domesticus* в *M. musculus*, по-видимому, возникла относительно недавно. Интересно, что ядерные гены не обнаруживают свидетельств интрогрессии. По-видимому, интрогрессия по ядерным генам предотвращается за счет стерильности или нежизнеспособности гибридов, которая обусловлена самими ядерными генами, тогда как мтДНК, не влияя, или слабо влияя, на приспособленность, может наследоваться и передаваться независимо. Эти данные уже обсуждались специально в связи с анализом гибридизации и интрогрессии (Лекция 10), а также в связи с построением и филогеографическим анализом деревьев.

13.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 13

1. Выполните оценивание p_n и p, используя пакеты программ GENEPOP или ARLEQUINE и примеры, приводимые в них.

2. Выберите через Интернет в Генном банке данных две последовательности нуклеотидов для видов и проведите их анализ, используя GENEPOP или ARLEQUINE.

ЛЕКЦИЯ 14. ФИЛОГЕНЕТИЧЕСКИЕ ДЕРЕВЬЯ

ГЛАВНЫЕ ВОПРОСЫ:

14.1. Типы филогенетических деревьев.
14.2. Методы матриц расстояний.
14.3. Методы парсимонии и максимального правдоподобия.
14.4. Популяционно-генетическая теория и построение деревьев.
14.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 14.

РЕЗЮМЕ

1. Имеются два типа деревьев, называемых **видовыми деревьями** и **генными деревьями.** Они могут совпадать, но не обязательно идентичны.
2. Видовое дерево – это всегда ожидаемое дерево расстояний. Генное дерево может быть либо ожидаемым деревом расстояний, либо реализованным деревом расстояний. В ожидаемом дереве длина ветвей, ведущих к двум новым видам от общего предкового вида должна быть равной как для видового, так и для генного дерева. Соответственно, время после дивергенции между парой генов должно быть идентичным. Когда скорость замен в генах постоянна, то ожидаемое эволюционное расстояние должно также быть одинаковым вдоль каждой ветви. Этот тип дерева принято называть **ожидаемым деревом расстояний**. Когда подсчитывается число мутационных изменений или замен в гене, которые реально произошли (реализовались), то в этом случае длина каждой ветви может быть различной. Этот тип дерева называется **реализованным деревом расстояний**. Реализованные деревья получаются, например, при использовании методов Фитча и Марголиаша, а также методов парсимонии.
3. Взаимосвязи в филогенетических линиях организмов или генов могут быть визуализованы в деревоподобной форме с корнем. Этот тип деревьев называется **укорененным деревом**. Взаимосвязи могут быть выражены также в виде **сети** или **графа.** Их называют **не укорененным деревом.**
4. Имеются две группы наиболее популярных методов реконструкции филогенетического дерева: **Методы матриц расстояний** и **Методы максимальной парсимонии**.
5. Простейший из методов матриц расстояний – это не взвешенный парно-групповой метод (НПГМ) анализа или метод средних расстояний. Часто весьма эффективным является NJ метод.
6. Дерево, которое требует минимального числа мутационных изменений при построении, называется **деревом максимальной парсимонии** (максимально экономное дерево).
7. Построить правильное филогенетическое дерево – это очень трудно выполнимое дело, особенно, когда число ОТЕ является большим. Однако возможность устанавливать генные генеалогии открывает новую эру в филогенетических исследованиях.

Chapter 14. PHYLOGENETIC TREES

SUMMARY

1. There are two types of trees, which called **species trees** and **gene trees**. They can coincide but not necessarily are identical.
2. A species tree is always an expected distance tree. A gene tree can be either an expected distance tree or a realized distance tree. In the expected distance tree the length of branches, leading to two new species from a common ancestor species is equal as for species as for gene tree. Correspondingly, time after divergence between a pair of genes must be identical. When a rate of substitutions in genes is constant then the expected evolutionary distance is also equal along each branch. This type of a tree is conveniently called an **expected distance tree**. When a number of mutations or substitutions, which actually have occurred (realized) in a gene is calculated, then in this case length of each branch may be different. This type of a tree is called a **realized distance tree.** Realized trees can be obtained, for instance, when Fitch and Margoliash methods as well as a parsimony one are used.

3. Interrelationships in phylogenetic lineages of organisms or genes can be visualized in the tree-like mode with a root. This type of tree is called a **rooted tree**. Interrelationships can also be shown in a form of a **net** or **graph.** It called unrooted tree.
4. There are two groups of the most popular methods of phylogenetic tree reconstructions, **Distance Matrix Methods** and **Maximum Parsimony Methods**.
5. The simplest of the Distance Matrix Methods is the Unweighted Pair-Group Method of Analysis (UPGMA) or the method of mean distances. Frequently the NJ method is very efficient.
6. A tree requiring the minimal number of substitutions being built is called a **Maximum Parsimony Tree (Maximum Economy Tree).**
7. Building a correct phylogenetic tree is a very difficult task, especially when OTU number is large. Nevertheless, a possibility to obtain gene genealogies reveals a new era in the phylogenetic research.

14.1. ТИПЫ ФИЛОГЕНЕТИЧЕСКИХ ДЕРЕВЬЕВ

Одно из наиболее важных достижений молекулярной эволюционной генетики – это открытие относительного постоянства темпов аминокислотных и нуклеотидных замен. Это открытие дало исследователям новое средство для реконструкции филогенетических деревьев. Как было подчеркнуто в Лекции 4, постоянство темпа аминокислотных или нуклеотидных замен выполняется только приблизительно, с множеством допущений и возникающими осложнениями при его оценивании. Поэтому реконструкция филогении и оценка времени дивергенции требует профессионального подхода и знания этих допущений и возможных ошибок. Однако в сравнении с морфологическими или палеонтологическими признаками молекулярные данные имеют лучшую экспериментальную обоснованность, а также собственную теоретическую базу для оценки эволюционных изменений. Молекулярная дивергенция может быть измерена в единой, сопоставимой шкале, как это представлено в Лекции 7, и, таким образом, дает сравнительные ряды, позволяющие на количественной основе строить генеалогию в отличие от морфологических или палеонтологических данных. Поэтому они могут дать более ясное представление об эволюционных взаимосвязях организмов, чем, к примеру, морфологические признаки. Для филогенетических деревьев, построенных по морфологическим признакам, очень трудно дать эволюционную временную шкалу, тогда как молекулярно-генетическую дивергенцию легко перевести в оценку времени дивергенции. Все это не значит, что другие подходы должны быть отвергнуты как устаревшие. Сложность жизни, разнообразие видов и путей их эволюции не дает гарантии на абсолютный успех лишь какого-либо одного подхода, даже будучи сверхсовременным сегодня. Это понимание появилось (Hillis et al., 1996; Avise, Wollenberg, 1997), после того, как не столь модные аллозимные и морфологические данные и деревья, основанные на них, предсказали взаимоотношения в таксонах в некоторых случаях лучше (более согласованно со всей другой информацией), чем филогенетический анализ ДНК.

Имеется давнее противоречие между представителями различных направлений нумерической таксономии, в частности, между сторонниками фенетического и кладистического подходов (Бейли, 1970; Sneath, Sokal, 1973; Eldredge, Cracraft, 1980). Существенная часть этого противоречия связана с идеологией и методологией классификации организмов. В этой лекции нас интересует в первую очередь построение филогенетических деревьев, чем таксономия. Поэтому мы не будем вдаваться в детали этих противоречий. Хотя мы познакомимся с различными методами построения деревьев (дендрограмм), часть из которых предпочитается сторонниками одного направления, а часть – сторонниками другого направления нумерической таксономии. Мы рассмотрим методы только по степени их применимости к молекулярным данным. Хотя многие подходы применимы также к морфологическим признакам.

Видовые и генные филогенетические деревья

Встречаются два типа деревьев – **видовые деревья** и **генные деревья** (Tateno et al., 1982). Один из вариантов видового дерева мы видели в Лекции 7 (рис. 7.1.1, схема дивергенции во

времени). Два этих дерева не обязательно совпадают. Одна причина этого несовпадения была пояснена в предыдущей лекции (рис. 13.4.1). Более того, даже характер ветвления, или топология дерева, которое построено по генам, может быть отличающимся от видового дерева.

Для трех видов возможны три различных взаимосвязи между видовым и генным деревом. (рис. 14.1.1). В варианте (a) и (b) топология видового и генного дерева одинакова, но во взаимоотношениях (c) они различны. Вероятность появления взаимоотношений по типу (c) достаточно высока, когда временной интервал $(t_1 - t_0)$ между первым и вторым ветвлениями у вида мал.

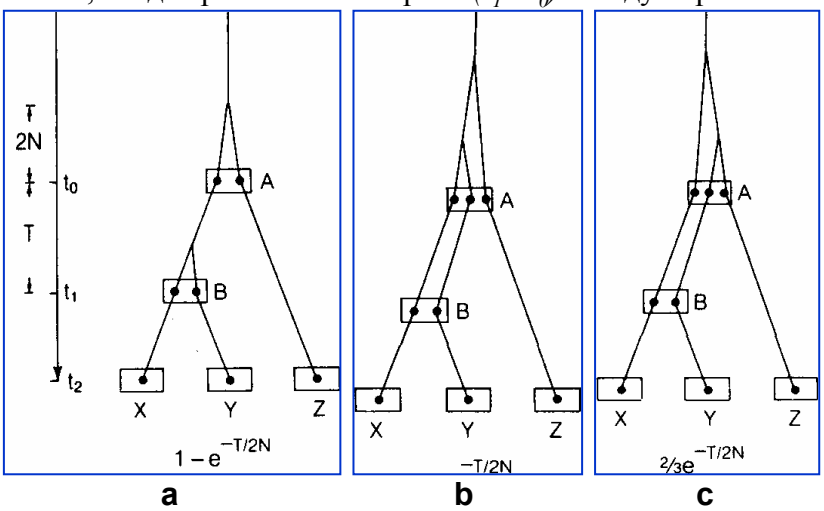


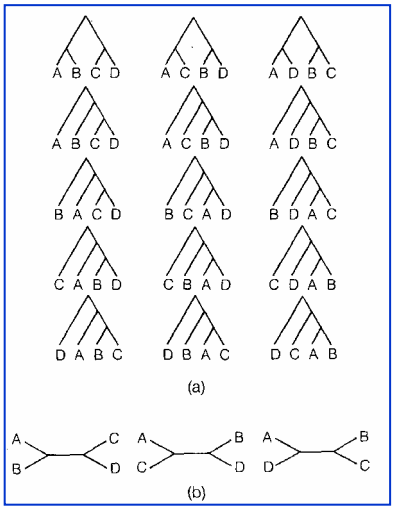
Рис. 14.1.1. Три возможных типа отношений между видовым и генным деревом для случая трех видов (X, Y и Z) в присутствии полиморфизма.

A и B точки генного дерева, в которых оценивается близость видов. t_0 и t_1 – это время первого и второго ветвления видов, соответственно. Вероятность появления каждого дерева дана внизу. $T=t_1 - t_0$, а N – это эффективная численность популяции (По Nei, 1987).

Fig. 14.1.1. Three possible relationships between the species and gene trees for the case of three species (X, Y, and Z) in the presence of polymorphism, t_0 and t_n are the times of the first and second species splitting, respectively. The probability of occurrence of each tree is given underneath the tree. $T=t_1 — t_0$, and N is the effective population size (From Nei, 1987).

Ожидаемые и реализованные деревья расстояний

Видовое дерево – это всегда ожидаемое дерево расстояний. Генное дерево может быть либо ожидаемым деревом расстояний, либо реализованным деревом расстояний. В чем различие между этими деревьями? Когда исследуются видовые деревья, то мы обычно интересуемся временем, прошедшим после дивергенции между каждой паров видов. В таком типе дерева длина ветвей, ведущих к двум новым видам от общего предкового вида должна быть равной. Аналогично для генного дерева, время после дивергенции между парой генов должно быть идентичным. Когда скорость замен в генах постоянна, то ожидаемое эволюционное расстояние должно также быть одинаковым вдоль каждой ветви. Поэтому необходимо построить дерево, каждая ветвь которого пропорциональна эволюционному времени или ожидаемому эволюционному расстоянию. Этот тип дерева принято называть **ожидаемым деревом расстояний**. Наоборот, число мутационных изменений или генных замен, которые фактически произошли (реализовались) в каждой из двух эволюционных линий, может быть не одинаковым из-за стохастических ошибок или варьирующего темпа генных замен в этих линиях. Оценка фактического числа мутационных



изменений, реализованных в каждой ветви дерева, будет различной. Соответственно, длина каждой ветви может быть различной в этом случае. Этот тип дерева называют **реализованным деревом расстояний**.

Укорененные и не укорененные деревья

Взаимосвязи в эволюционных линиях организмов или генов визуально могут быть представлены в деревоподобном виде с корнем, что является наиболее употребительным (рис. 14.1.2, a). Дерево этого типа называется **укорененным деревом**. Однако также возможно построить дерево, не имеющее корня (рис. 14.1.2, b).

Рис. 14.1.2. Пятнадцать возможных укорененных деревьев (a) и три возможных не укорененных дерева (b) для двух четырех (По Nei,1987).
Fig. 14.1.2. Fifteen possible rooted trees (a) and three possible unrooted trees (b) for four species (From Nei, 1987).

Имеется много возможных укорененных и не укорененных деревьев для данного числа видов *(n)*. Так, для случая *n*=4, имеются 15 возможных укорененных деревьев и 3 не укорененных дерева (рис. 14.1.2). Число возможных бифуркационных деревьев увеличивается с увеличением *n* согласно следующему выражению (Nei, 1987):

$$1 * 3 * 5 \ldots (2n - 3) = [(2n - 3)!] / [2^{n-2} (n - 2)!]. \qquad (14.1)$$

Фенетические и кладистические деревья

В нумерической таксономии для построения деревьев имеются два различных подхода (Sokal, Sneath, 1970; Бейли, 1970). Один из них – это фенетический подход, при котором дерево строится на основе фенотипического сходства видов без понимания их эволюционной истории. Этот подход разумен, когда законы эволюционных изменений признаков не достаточно ясны. Данный подход также обоснован, если исследователь в первую очередь заинтересован в таксономии, которая строится в основном на фенотипических признаках. Поскольку дерево, построенное этим методом, может не представлять эволюционную взаимосвязь, то оно часто называется **фенограммой**. В этом случае дерево является, главным образом, средством группировки организмов. Хотя, некоторые статистические методы, разработанные в фенетике, могут быть использованы для построения филогенетического дерева. Тем не менее, критицизм по отношению к фенетическому подходу имеется (Farris, 1977).

При втором подходе, который называется **кладистическим** (кладистским) **подходом**, филогенетическое дерево строится на основе рассмотрения разнообразия возможных способов эволюции признаков, следуя определенным правилам и выбору наилучшего возможного дерева. Дерево, реконструированное в согласии с этим подходом, чаще всего называют **кладограммой** (или дендрограммой), в контрасте с фенограммой. На первый взгляд можно подумать, что этот подход для конструирования филогенетических деревьев лучше, чем фенетический. Однако фактически это не так. Допущения, выдвигаемые кладистическим подходом, не всегда удовлетворяются, как молекулярными (Nei, 1987; Nei, Kumar, 2002), так и морфологическими данными (Бейли, 1970). Поэтому превосходство данного подхода не гарантированно.

Реконструкция филогенетического дерева

Как правило, очень трудно реконструировать истинное филогенетическое дерево, которое показывает, каким образом виды и другие таксоны возникали и дивергировали. Наиболее распространенными являются два подхода к реконструкции филогенетических деревьев: 1. **Методы матриц расстояний** (ММР) и 2. **Методы максимальной парсимонии** (ММП). В ММР для всех пар видов или популяций вычисляют эволюционное или генетическое расстояние (сходство) и группируют его в матрицы (ОТЕ или ОПЕ, см. объяснения ниже в Разделе 14.2), а филогенетическое дерево впоследствии строится на основе рассмотрения взаимосвязей среди всех

измеренных попарных расстояний в этой матрице. После того, как матрица получена, необходимо найти способ наглядного представления информации, содержащейся в ней. Имеется несколько способов сделать это, и построение дерева – это один из таких способов. В ММП нуклеотидная или аминокислотная последовательность предков выводится на основании данных, полученных для видов потомков, а дерево строится с минимизацией числа изменений в ветвях во всем филогенетическом дереве. Такое дерево максимальной парсимонии или экономии (Maximum parsimony tree) называют еще минимальным деревом, а также – максимально экономным деревом. При этом имеются также несколько алгоритмов конструирования дерева максимальной парсимонии. Дополнительно к двум описанным методам, имеются методы максимального правдоподобия и, так называемые, методы сопоставимости, которые тоже используются при построении деревьев. Мы затронем данные методы только вскользь.

14.2. МЕТОДЫ МАТРИЦ РАССТОЯНИЙ

Не взвешенный парно-групповой метод

Простейший из методов матриц расстояний – это не взвешенный парно-групповой метод (НПГМ; unweighted pair-group method, UPGM) или метод средних расстояний. Этот метод исходно был разработан для построения фенограмм (Sokal, Michener, 1958; Бейли, 1970), но он может также использоваться для построения филогенетического дерева. Метод применим в этом отношении до тех пор, пока ожидаемые значения расстояний не очень сильно различаются в исследуемых филогенетических линиях. Например, цифровое моделирование показало, что когда оценки расстояний подвергаются большим стохастическим ошибкам, НПГМ часто превосходит другие методы матриц расстояний в нахождении правильного дерева (Tateno et al., 1982; Nei et al., 1983; Sourdis, Krimbas, 1987). НПГМ предназначен для оценки видового дерева или ожидаемого генного дерева. Для достижения этой цели лучше использовать линейные меры расстояний, такие как число аминокислотных замен или стандартное генетическое расстояние Нея (D_n).

Рассмотрим алгоритм НПГМ. На первом этапе НПГМ вычисляется какая-либо мера расстояния (сходства) для всех пар опрерационных таксономических единиц (ОТЕ) или опрерационных популяционных единиц (ОПЕ), то есть, для пар видов или популяций. На основе матрицы всех попарных расстояний далее строится дерево, первый кластер которого получают объединением двух ОТЕ с наименьшим расстоянием. Затем рассчитывается новая матрица расстояний, с объединенными видами, представляющими теперь одну ОТЕ. Если объединяемые виды были индексированы как i и j, то тогда для всех расстояний $n \neq i,j$ из n формируются кластеры $r_{n(i,j)} = (r_{in} + r_{jn})/2$. Другими словами, расстояние от каждой n–й ОТЕ до кластера - это среднее арифметическое расстояний n–й ОТЕ до каждого другого члена ОТЕ в конкретном кластере. Остальные расстояния остаются такими же, как в исходной матрице. Новая рассчитанная матрица расстояний анализируется для выявления наименьшего расстояния между ОТЕ и объединения очередной пары в кластер. Эта процедура повторяется до тех пор, пока все ОТЕ не кластеризуются в дерево.

Разберем простой пример работы алгоритма НПГМ. Пусть полученные расстояния представлены в виде следующей матрицы ОТЕ (ОПЕ).

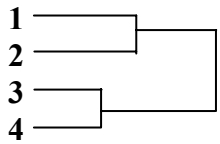
ОТЕ (ОПЕ)	**1**	**2**	**3**
2	r_{12}		
3	r_{13}	r_{23}	
4	r_{14}	r_{24}	r_{34}

Здесь, r_{ij} обозначает расстояние между i-той и j-той ОТЕ. Пусть расстояние между ОТЕ 3 и 4 (r_{34}) является наименьшим среди всех 6 значений в данной матрице. Тогда эти две ОТЕ объединяются в одну (ОТЕ 3-4), с точкой ветвления, находящейся на расстоянии $r_{34}/2$. Предполагается, что длина

ветвей, ведущих от точки ветвления к ОТЕ 3 и ОТЕ 4 одинаковы. Новые расстояние между этой объединенной ОТЕ и другими ОТЕ будут:

ОТЕ (ОПЕ)	**1**	**2**
2	r_{12}	
(34)	$r_{1(34)}$	$r_{2(34)}$

Здесь, $r_{1(34)}$ и $r_{2(34)}$ получают как $(r_{13} + r_{14}) / 2$ и $(r_{23} + r_{24}) / 2$, соответственно. Наконец, пусть r_{12} – это самое маленькое расстояние. Включаем в дерево $r_{12}/2$ и определяем последний кластер как: $r_{(12)(34)} = (r_{1(34)} + r_{2(34)})/2$ и присоединяем его к дереву также. Мы получили следующее дерево для четырех ОТЕ:



Теоретически при использовании НПГМ для n=4 ОТЕ возможно 15 типов деревьев (см. рис. 14.1.2). Выражение 14.1 позволяет представить, сколько типов деревьев теоретически возможно, скажем, для рода животных или семейства с n=25 ОТЕ, включающего двадцать пять видов (небольшой таксон). Оно равно 3^{49}! Это только одна из причин, почему трудно определить истинную топологию дерева, но имеются и другие причины. Об этом – ниже.

НПГМ - это простой метод и при этом он имеет несколько хороших статистических свойств. Основное допущение НПГМ, при использовании молекулярных данных, состоит в принятии постоянства ожидаемого темпа замен в генах. Если используемая мера расстояния точно линейно и без ошибки связана с временем эволюции, то НПГМ дает точную топологию и правильные длины ветвей. Если мера расстояния подвержена статистической ошибке, то и топология и длина ветвей могут быть неправильными, даже при постоянстве ожидаемого темпа замен. В эмпирических данных истинная топология обычно не известна. Поэтому трудно оценить объем топологических ошибок и ошибок длин ветвей. Известно, однако, что если точная топология найдена, то НПГМ дает наилучшие оценки для длин ветвей по методу наименьших квадратов (Chakraborty, 1977). Один из способов, как оценить правильность топологии филогенетического дерева, вычислить стандартные ошибки (SE) каждой точки ветвления в дереве. В случае дерева НПГМ, эти SE можно вычислить легко, если мы допустим постоянство темпа замен в генах. Очевидно, что SE точки ветвления, ведущей к двум ОТЕ (например, А на рис. 14.2.1), равняется половине SE расстояния между двумя ОТЕ. Мы можем также использовать SE для измеренных расстояний в исходной матрице и применить их к точкам ветвления. Однако не все меры имеют SE. Имеется способ оценки дисперсий и SE исходя только из расстояний и числа ОТЕ (Nei et al., 1985; Nei, 1987, p. 295-296).

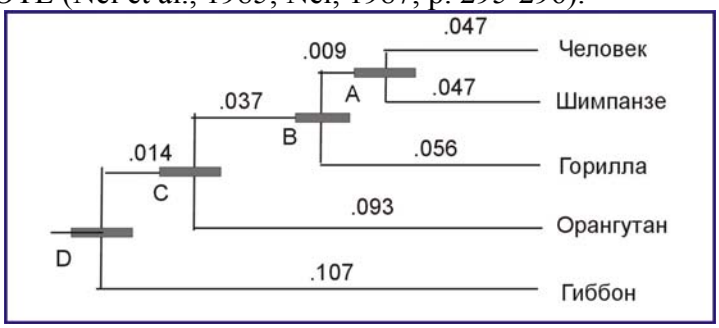


Рис. 14.2.1. Филогенетическое дерево, реконструированное на основе НПГМ по матрице расстояний, представленной в Таблице 14.2.1.

Цифрами показаны длины ветвей. Заштрихованные полосы представляют стандартные ошибки с каждой стороны точек ветвления, A-D (По Nei, 1987).

Fig. 14.2.1. Phylogenetic tree reconstructed by UPGM from the distance matrix in Table 14.2.1. The hatched bar represents one standard error on each side of the branching point (From Nei, 1987).

Рассмотрим численный пример построения дерева НПГМ для гоминид (Nei, 1987). Для шести видов был секвенирован фрагмент мтДНК (Brown et al., 1982). Этот фрагмент длиной 895 нуклеотидов сопоставлен у человека, шимпанзе, гориллы, орангутана и гиббона. Исходя из этих данных, можно оценить число нуклеотидных замен *(d)* и их стандартные ошибки, используя соответствующие уравнения (Nei, 1987; 5.3 и 5.4) или 13.5 и 13.7 из Лекции 13, если базироваться на нуклеотидном разнообразии. Полученные результаты представлены в Табл. 14.2.1. Очевидно, что значение *d* между человеком и шимпанзе является наименьшим, поэтому человек и шимпанзе должны образовать первый кластер с точкой ветвления *(A)* в d_{HC}= 0.094/2 =0.047 (рис. 14.2.1). Человек и шимпанзе теперь объединяются в одну ОТЕ *(HC)*. Значения *d* для этой ОТЕ и гориллой, орангутаном и гиббоном становятся следующими: (0.111+0.115)/2 = 0.113, (0.180+0.194)/2 = 0.187, и (0.207+ 0.218)/2 =0.212, соответственно. Другие значения расстояний остаются неизменными. Наименьшее значение *d* новой матрицы теперь (0.113) между *(HC)* и гориллой. Таким образом, горилла соединяется на дереве с ОТЕ *(HC)* с точкой ветвления $d_{G(HC)}$ = 0.113/2=0.056. Если эти расчеты повторять, то в итоге получим дерево (рис. 14.2.1).

Таблица 14.2.1. Доля различающихся нуклеотидов *(p)* (выше диагонали) и оценки числа нуклеотидных замен на сайт *(d)* и их стандартные ошибки (ниже диагонали), полученные по данным секвенирования нуклеотидов для пяти видов приматов (По *Brown et al., 1982*)

Table 14.2.1. Proportion of different nucleotides *(p)* (above the diagonal) and estimates of the number of nucleotide substitutions *(d)* per site and their standard errors (below the diagonal) obtained from nucleotide sequence data for five primate species (*From Brown et al., 1982*)

	Человек	***Шимпанзе***	***Горилла***	***Орангутан***	***Гиббон***
Человек		.088	.103	.160	.181
Шимпанзе	.094 ±.011		.106	.170	.189
Горилла	.111 ±.012	.115 ±.012		.166	.189
Орангутан	.180 ±.016	.194 ±.016	.188 ±.016		.188
Гиббон	.207 ±.017	.218 ±.017	.218 ±.017	.216 ±.017	

Метод ближайшего соседства

Метод ближайшего соседства (**Neighbor Joining,** NJ) (Saitou, Nei, 1987) является близким к НПГМ, поскольку он также базируется на матрице расстояний (или наоборот – сходства). Редуцируя исходную матрицу в размерности на каждом шаге, NJ строит дерево, базирующееся на серии матриц меньшего размера. NJ отличается от НПГМ тем, что не формирует кластеры на каждом шаге, а прямо вычисляет расстояния для внутренних узлов. Начиная с исходной матрицы, NJ сначала вычисляет для каждой ОТЕ её дивергенцию в не укорененном дереве или сети (графе) от всех других ОТЕ, как сумму индивидуальных расстояний. Затем NJ использует эту «сетевую» дивергенцию для вычисления скорректированной матрицы расстояний. Далее NJ находит пары ОТЕ, с наименьшими скорректированными расстояниями и вычисляет расстояние от каждой из этих ОТЕ к узлу, который их соединяет. При этом, расстояние от каждой ОТЕ в паре к узлу не обязательно должно быть одинаковым. Затем создается новая матрица, в которой новый узел замещает старый для данной пары ОТЕ. NJ не предполагает, что все ОТЕ равноудалены от корня.

NJ, аналогично методу парсимонии, является методом минимальных длин ветвей или эволюционный изменений, но он не гарантирует нахождения дерева с наименьшим общим расстоянием. Действительно, имеются случаи, для которых существуют многие более короткие, чем NJ деревья (Hillis et al., 1996). Некоторые авторы считают (Hillis et al., 1996; Swofford et al., 1996), что наилучшим использованием NJ дерева является базирование на нем, как на стартовой точке других методов, основанных на предопределенных моделях нуклеотидных замен, например, таких как метод максимального правдоподобия (см. Раздел 14.3 и Глава 15). Однако опыт показывает (Kartavtsev et al., 2007b; Sazaki et al., 2007), что и сами NJ-деревья часто дают хорошую

реконструкцию филогенетических взаимосвязей. Учитывая очень эффективный вычислительный алгоритм NJ, это имеет большое значение при построении больших деревьев. По-видимому, именно это обстоятельство, а также весьма привлекательное свойство «минимальности», послужили толчком к весьма высокой популярности NJ-деревьев в молекулярной филогенетике.

Метод Фитча и Марголиаша (МФМ)

Один из методов реализованных расстояний предложен Фитчем и Марголиашем (Fitch, Margoliash, 1967a). Предположим, что число замен, различающих последовательности *i* и *j*, составляет d_{ij}. Если дерево, связывающее видовые последовательности 1, 2 и 3, имеет длины ветвей A, B и C (рис. 14.2.2), то тогда длина всех ветвей может быть оценена из равенств:

$$A = (d_{12} + d_{13} - d_{23}) / 2$$
$$B = (d_{12} + d_{23} - d_{13}) / 2 \qquad (14.2)$$
$$C = (d_{13} + d_{23} - d_{12}) / 2.$$

Эти равенства получены решением уравнений $d_{12} = A + B$, $d_{13} = A + C$ и $d_{23} = B + C$. Если сравниваются более, чем три последовательности, то дерево строится из трех ОТЕ на каждом этапе, начиная с наиболее близко связанных последовательностей и присоединяя остальные последовательности (C теперь представляет сборную группу). Если последовательности 1 и 2 являются наиболее сходными, то тогда расстояние от последовательности 1 до остальных групп представляет собой среднее расстояний от последовательности 1 к каждому члену группы. При таком подходе только три расстояния рассматриваются в конкретный момент, а выражения 14.2 позволяют оценить длины ветвей. Процедура повторяется до тех пор, пока все ОТЕ не объединятся в одно семейство. В отличие от НПГМ, этот метод не позволяет найти корень дерева, как среднее значение. Чтобы найти этот корень необходимо ввести «внешнюю» ОТЕ или вычислить корень как (A+B+2C)/4 на завершающем этапе построения дерева. Топология дерева МФМ всегда такая же, как при использовании НПГМ, но длины ветвей отличаются.

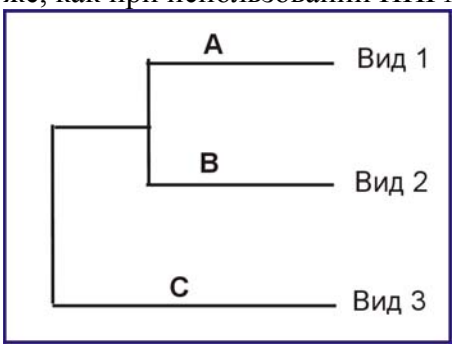


Рис. 14.2.2. Простое филогенетическое дерево.

A, B и C представляют длины ветвей по отношению к ближайшему общему предку (По Hartl, Clark, 1989).

Fig. 14.2.2. A simple phylogenetic tree. A, B, and C represent branch lengths from the most recent common ancestor (From Hartl, Clark, 1989).

Процедура, которая описана выше, это только часть МФМ. Рассматривая вероятность того, что топология полученного дерева не верна, необходимо проанализировать другие возможные топологии, прежде чем принять окончательное решение. Поскольку число деревьев при большом числе ОТЕ очень велико (см. выше), то при МФМ рассматривается применимость только небольшого числа первых деревьев (обычно 2-3). Решение, какое дерево является наилучшим, принимается с использованием дополнительных критериев, типа: (1) значение коэффициента кофенетической корреляции, r (r>0.95-0.99) или/и (2) так называемого, процента стандартных отклонений, s_{FM} (Nei, 1987; выражение 11.9), и других. Одно из деревьев считается более правильным, если r велико, а s_{FM} мало; s_{FM} – это отклонение между длиной ветвей в построенном дереве и реконструированном патристическом (предковом) дереве. Больше деталей на эту тему можно найти в книге Нея (Nei, 1987, P.299-301; см. также Практическое занятие 14).

Метод трансформированных расстояний

Этот метод близок к МФМ. Фэррис (Farris, 1977) обнаружил, что если темп замен варьирует между эволюционными линиями, то трансформированные расстояния могут улучшить топологию дерева для метода средних расстояний. Алгоритм преобразования таков:

$$d'_{ij} = (d_{ij} - d_{ir} - d_{jr})/2, \qquad (14.3)$$

где *r* относится к внешней группе (внешняя ОТЕ). Вследствие стохастических ошибок, различий в обратных и параллельных мутациях вдоль различных ветвей, могут возникать ошибки топологии и ошибки длины ветвей. Вот почему рекомендуется использовать две или более внешних ОТЕ. После того, как построено несколько деревьев, выбирается истинное дерево, как наиболее состоятельное или консенсусное (consensus) дерево. Как найти это консенсусное дерево? Чтобы сделать это имеются несколько способов. Самый простой – это выбрать дерево, которое удовлетворяет всем или большинству известных фактов (эмпирический подход). Начинают разрабатываться также и некоторые теоретические средства (Swofford et al., 1996).

Фаррисовский метод расстояний Вагнера (ФМРВ)

Этот метод был предложен для реализованных деревьев. Однако при ФМРВ для оценки расстояний должны использоваться меры, которые удовлетворяют правилу треугольника (гипотенуза короче суммы двух катетов). Однако, также как при МФМ, с помощью ФМРВ можно построить несколько деревьев, но только одно выбирается в качестве правильного. Рассмотрим ту же матрицу расстояний, которая представлена в Таблице 14.2.1 для 5 ОТЕ. Пусть d_{ij} – это расстояние между *i*-той и *j*-той ОТЕ. Предположим далее, что расстояние d_{12} является наименьшим в матрице ОТЕ; так что ОТЕ 1 и 2 объединяются первыми. Расстояния между этой парой ОТЕ (1-2) и каждой из остальных ОТЕ вычисляется как средние между ОТЕ (1) и ОТЕ (3), а также ОТЕ (2) и ОТЕ (3). Предположим, что теперь расстояние между ОТЕ (1-2) и ОТЕ (3) является минимальным из всех расстояний в матрице, полученной этим способом. Тогда ОТЕ (3) соединяется с ОТЕ (1-2), как это показано на рис. 14.2.3, а. Это дерево представляет собой сеть или не укорененное дерево (граф), поскольку сам подход не предполагает определения корня.

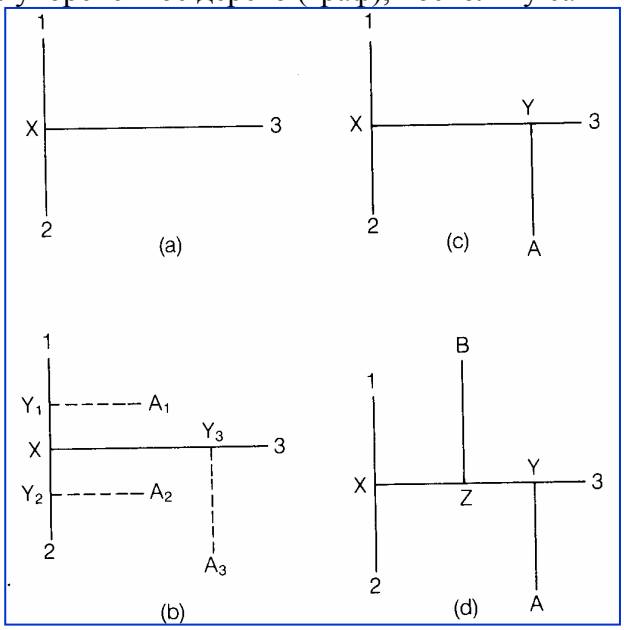


Рис. 14.2.3. Методика построения дерева расстояний Вагнера.

Объяснения – в тексте (По Nei, 1987).

Fig. 14.2.3. Procedure of constructing a distance-Wagner tree.

See text for explanations (From Nei, 1987).

Корень может быть найден позже, введением внешней группы или каким либо иным путем. На рис. 14.2.3 *X* – это точка ветвления. Длина каждой ветви вычисляется при использовании следующих формул:

$$L(3,X) = d_{13} + d_{23} - d_{12})/2, \qquad (14.4a)$$

$$L(1,X) = d_{13} - L(3,X), \qquad (14.4b)$$

$$L(2,X) = d_{23} - L(3,X). \qquad (14.4c)$$

Здесь *L(a,b)* – это длины отрезков между точками *a* и *b*. В реальности эти длины вычисляются отдельно для каждой ветви от пары ОТЕ (1-2) и остальными ОТЕ, а ОТЕ, которые показывают наименьшее расстояние к *X*, выбирается следующим по порядку. Теперь переходим ко второму шагу, при котором еще одна ОТЕ добавляется к сети. Имеются три возможности для одной ОТЕ, скажем А, из всех остальных ОТЕ (4 и 5) быть присоединенными к сети (рис. 14.2.3, b). Необходимо вычислить длину ветвей $L(A_1,Y_1)$, $L(A_2,Y_2)$, $L(A_3Y_3)$. Эти вычисления осуществляются для всех остальных ОТЕ (4 и 5), а ОТЕ, которые показывают наименьшие расстояния, далее выбираются для присоединения к сети. Как и ранее, в действительности $L(A_iY_i)$ вычисляются по следующим формулам:

$$L(A_1, Y_1) = \{L(A_1, 1) + L(A_1, X) - L(1, X)\}/2, \qquad (14.5a)$$

$$L(A_2, Y_2) = \{L(A_2, 2) + L(A_2, X) - L(2, X)\}/2, \qquad (14.5b)$$

$$L(A_3, Y_3) = \{L(A_3, 3) + L(A_3, X) - L(3, X)\}/2. \quad \textbf{(14.5c)}$$

В этих формулах $L(A_1, 1)$ получают прямо из матрицы расстояний, однако $L(1, X)$ были уже найдены из выражений (14.4). С другой стороны, эти значения могут быть оценены исходя из следующих формул:

$$L(A_1, X) = L(A_1, 2) - L(2, X) = L_1 \text{ или } L(A_1, 3) - L(3, X) = L_2, \quad \textbf{(14.6a)}$$

$$L(A_2, X) = L(A_2, 1) - L(1, X) = L_5 \text{ или } L_2, \quad \textbf{(14.6b)}$$

$$L(A_3, X) = L_3 \text{ или } L_1. \quad \textbf{(14.6c)}$$

Из L_1, L_2 или L_3 алгоритм Фэрриса выбирает наибольшие значения и использует их для всех значений из $L(A_i, X)$ согласно уравнениям 14.5. Предположим, что $L(A_3, Y_3)$ является наименьшим из $L(A_1, Y_2)$, $L(A_2, Y_2)$ и $L(A_3, Y_3)$. Тогда ОТЕ *A* должно быть соединено с ветвью 3-X, как показано на рис. 14.2.3, c. Последнее ОТЕ (B, которое либо ОТЕ 4 либо 5) далее добавляется к сети, как это показано на рис. 14.2.3, d. Процедура присоединения такова же, как ранее, за исключением того, что теперь должны рассматриваться пять возможных путей присоединения B. Этот процесс повторяется до тех пор, пока все ОТЕ не будут кластеризованы.

Сейчас для построения деревьев имеются модификации метода Фэрриса (Tateno et al., 1982; Faith, 1985).

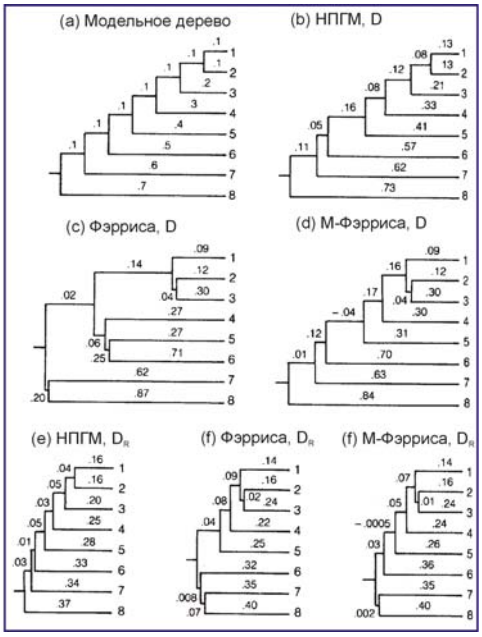
Сравнительные свойства методов матриц расстояний

1. В случаях, когда темп замен в генах постоянен, наилучшие результаты в отношении топологии дерева, как показывают результаты цифрового моделирования (Tateno et al., 1982), дает метод Фэрриса и его модификации. Это особенно верно, когда *d* невелико ($d <= 0.09$). Когда *d* больше, $d >= 0.37$, наилучшим является НПГМ (Nei, 1987, p.309).

2. Для оценки ожидаемых длин ветвей и построения правильной топологии при постоянстве темпа замен почти всегда наилучшим является НПГМ (Blanken et al., 1982; Nei, 1987, p.309).

3. При неоднородности темпа замен в ветвях наилучшие результаты кластеризации дает МФМ в модификации Фэрриса (Farris, 1977) (Blanken et al., 1982).

Ней с соавторами (Nei et al., 1983) показали три наиболее важных свойства методов матриц расстояний, которые частично иллюстрирует рис. 14.2.4. 1. Во всех методиках построения деревьев точность отражения топологии дерева и длин ветвей в построенном дереве (укорененное дерево) является очень малой, когда число локусов меньше 20, но постепенно увеличивается с увеличением их числа. 2. В случаях, когда ожидаемое число замен на локус, b = 0.1 или выше, и используется 30 локусов или более, то тогда топологические ошибки, которые оцениваются



индексом искажения d_T (Robinson, Foulds, 1981), невелики. Однако, вероятность обнаружения правильной топологии остается меньше, чем 0.5, даже когда размер выборки достигает 60 локусов. При небольших расстояниях, $d \approx 0.004$, эта вероятность много меньше. 3. НПГМ и модифицированный метод Фэрриса дают лучшую топологию по сравнению с методом Фэрриса.

Рис. 14.2.4. Модельное дерево (a) и реконструированные деревья (b, c, d), полученные при использовании расстояния Нея, D_n, а также (e, f, g), полученные на основе расстояния Роджерса, D_r. Реконструкции основываются на одинаковых данных для частот аллелей. Представлен результат одного компьютерного моделирования по 50 локусам (По Nei et al., 1983).

Fig. 14.2.4. Model tree (a) and reconstructed trees (b, c, d) obtained by using Nei's distance, D_n, and those (e, f, g) obtained by using Rogers' distance, D_r. These are based on the same set of gene frequency data from one replication of computer simulation with 50 loci (From Nei et al., 1983a).

14.3. МЕТОДЫ ПАРСИМОНИИ И МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Имеется несколько вариантов дерева парсимонии (Felsenstein, 1982; Swofford et al., 1996). Впервые идея подобного подхода появилась в середине 60-х годов (Eck, Dayhoff, 1966). Смысл метода состоит в получении последовательности аминокислот или нуклеотидов предкового вида и затем выбрать дерево, которое имеет минимальное число мутационных изменений. Дерево, которое построено таким методом, как мы говорили в начале, называется деревом максимальной парсимонии. Метод максимальной парсимонии исходно был предложен для выяснения топологии и длина ветвей при этом, как правило, не вычисляется.

Алгоритм

Рассмотрим алгоритм в изложении Нея (Nei, 1987). Процесс начинается с рассмотрения топологии конкретной группы ОТЕ и выводится предковая последовательность для такого дерева. Далее подсчитывается минимальное число замен, которое требуется для объяснения эволюционных изменений в дереве. Когда это число получено, то проверяется другая топология и вычисляется минимальное число замен для этого дерева. Процедура повторятся до тех пор, пока все разумные топологии не будут проверены. В конце топология, которая требует минимального числа замен, выбирается как правильная, а дерево, как дерево максимальной парсимонии.

Минимальное число замен, которое требуется для данного дерева, вычисляется для каждого нуклеотидного (или аминокислотного) сайта, а после этого определяется общее число замен для всех сайтов. Начнем с того, что определим первую топологию, получаемую методом Фэрриса, модифицированным методом Фэрриса или методом трансформированных расстояний. Далее проанализируем шесть ОТЕ, которые предполагаются связанными укорененным деревом (рис. 14.3.1, а). Предположим, что в определенном положении нуклеотида каждая ОТЕ имеет нуклеотиды, выписанные на ветвях дерева внизу. Из этих нуклеотидов мы можем вывести нуклеотиды для пяти предковых последовательностей (узлы): 1, 2, 3, 4 и 5. Нуклеотид в узле 1 должен быть либо C, либо T, если мы рассматриваем минимальное число замен. Аналогично, нуклеотид в узле 2 должен быть либо G, либо T, тогда как нуклеотид в узле 3 должен быть A, G или T. В узле 4 ожидается нуклеотид T, потому, что он ближайший предок узлов 1 и 3 (оба имеют T). Наконец, мы выводим нуклеотид в узле 5, который должен быть или A, или T. Теперь ясно, что минимальное число нуклеотидных замен для этого набора из 6 ОТЕ получается, при допущении,

что все предковые узлы имели нуклеотид T. Минимальное число, соответственно, равно четыре замены. Однако, это не единственный путь объяснения эволюционных изменений ОТЕ с четырьмя заменами.

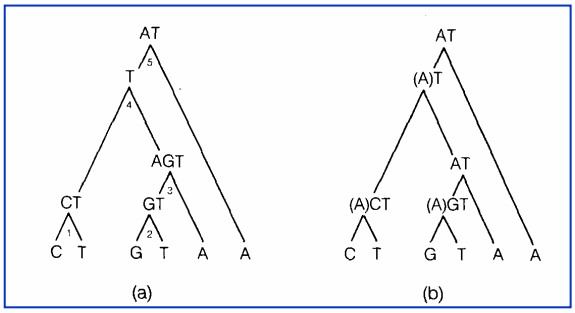


Рис. 14.3.1. Нуклеотиды у шести рецентных видов и возможные нуклеотиды у пяти предковых видов (По Fitch, 1971a).
Fig. 14.3.1. Nucleotides in six extant species and the possible nucleotides in five ancestral species (From Fitch, 1971a).

Если мы будем полагать, что все средние узлы имеют A, то число требующихся замен также равно 4 (рис. 14.3.1, b). Имеются еще девять других возможностей с тем же минимальным числом замен. Четыре из них таковы: (5-A, 4-T, 3-T, 2-T, 1-T), (5-A, 4-A, 3-A, 2-A, 1-T), (5-A, 4-A, 3-A, 2-A, 1-C) и (5-A, 4-A, 3-A, 2-G, 1-A). Из изложенных выше положений, ясно, что нуклеотиды в предковом узле не могут всегда быть определены уникальным образом и что все нуклеотиды, приведенные на рис. 14.3.1 удовлетворяют принципу парсимонии. Однако, возможно подсчитать минимальное число требующихся замен.

Альтернативные топологии для одного и того же набора данных могут возникать при использовании различных методов. Можно для этого использовать метод Фитча и Марголиаша (Fitch, Margoliash, 1967a), упомянутый ранее. Другой подходящий метод (Dayhoff, Park, 1969), позволяет разрезать ветви первого дерева, а каждая полученная часть затем трансплантируется на все ветви другой части. Минимальное число замен определяется для всех топологий, производимых этим способом, и затем выбирается лучшая. Этот процесс повторяется несколько раз и наилучшая из всех проанализированных топологий выбирается как дерево максимальной парсимонии.

Информативные сайты для определения топологии

Можно предположить, что любой полиморфный (политипический) сайт, в котором существует два или более различных типов нуклеотидов (или аминокислот), подходит для определения топологии дерева. Однако, для того, чтобы построить дерево максимальной парсимонии, нуклеотиды, существующие уникально в ОТЕ, так называемые единичные нуклеотиды (singular nucleotides; Fitch, 1977) являются **не информативными**. Для таких нуклеотидов можно всегда предположить, что они возникли в результате единичной мутации, в ближайшей ветви, ведущей к ОТЕ, в которой данный нуклеотид существует и, поэтому мутационные изменения совместимы с любой топологией. Нуклеотидный сайт является **информативным**, только тогда, когда имеются, по меньшей мере, два различных типа нуклеотидов, каждый представленный, по крайней мере, два раза. Например, первый нуклеотидный сайт на рис. 14.3.2 является не информативным, тогда как четвертый нуклеотидный сайт – информативный.

Разберем пример (Brown et al., 1982, из Nei, 1987). Рис. 14.3.2 показывает нуклеотидную последовательность, определенную для мтДНК человека (H), шимпанзе (C), гориллы (G), орангутана (O) и гиббона (B). На этом рисунке, показаны только те сайты, в которых встречается не менее двух различных типов нуклеотидов. Имеются 282 таких полиморфных сайтов, но только 90 из них являются информативными для определения топологии. Они отмечены точками на рис. 14.3.2.

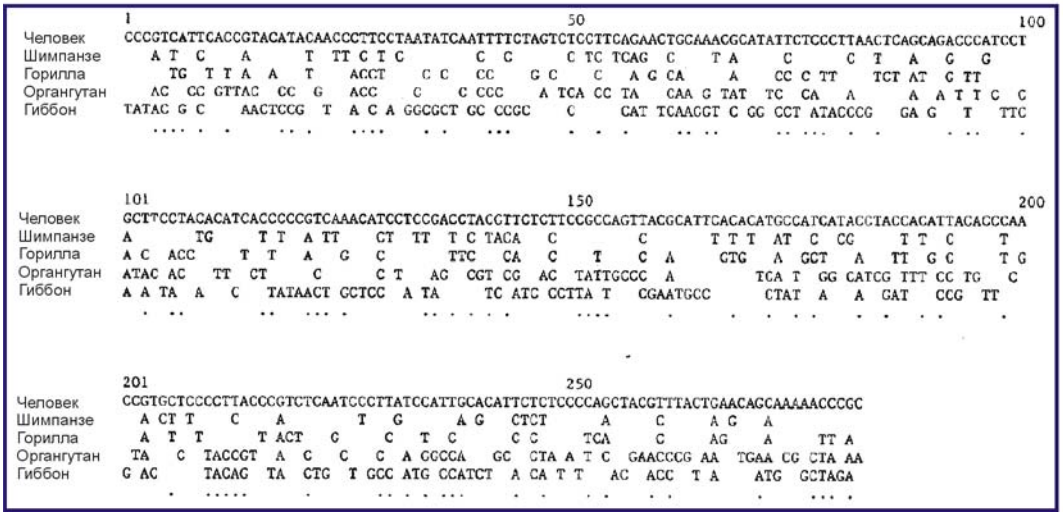


Рис. 14.3.2. Нуклеотидные последовательности мтДНК человека и обезьян. Представлены только полиморфные сайты последовательности. У шимпанзе, гориллы, орангутана и гиббона нуклеотиды, которые идентичны с нуклеотидами человека, не показаны. Нуклеотидные сайты, помеченные точками – это информативные сайты (По Nei, 1987; данные из Brown et al., 1982).
Fig. 14.3.2. Nucleotide sequences of mitochondrial DNAs from humans and apes. Here, only polytypic sites are presented. In chimpanzees, gorillas, orangutans, and gibbons, the nucleotides that are identical with those of humans are not shown. Nucleotide sites with dots are informative sites (From Nei, 1987; data from Brown et al., 1982).

Браун с коллегами рассматривают шесть различных топологий, которые были ранее предложены различными авторами для сопоставления этой группы организмов. Четыре наилучших дерева представлены на рис. 14.3.3.

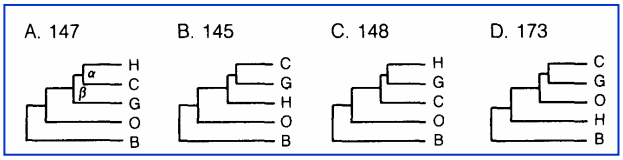


Рис. 14.3.3. Четыре (A-D) возможных филогенетических дерева (топологии) для человека (H), шимпанзе (C), гориллы (G), орангутана (O) и гиббона (B). Цифра, над каждой топологией – это минимальное число нуклеотидных замен, требуемое для объяснения различий последовательностей; α и β - узлы ветвления, где произошла единичная нуклеотидная замена (По Nei, 1987).
Fig. 14.3.3. Four possible phylogenetic trees (topologies) for the human (H), chimpanzee (C), gorilla (G), orangutan (O), and gibbon (B). The number given to each topology is the minimum number of nucleotide substitutions required for explaining the sequence differences (From Nei, 1987).

В конструкции максимального дерева парсимонии приемлемо классифицировать информативные сайты согласно наличию различных политипических или полиморфных типов. В данном случае имеются 15 различных полиморфных типов, как это показано в таблице 14.3.1.

Таблица 14.3.1. Полиморфные типы и их частоты по информативным сайтам для нахождения дерева парсимонии человека (H), шимпанзе (C), гориллы (G), орангутана (O) и гиббона (B) *(По Nei, 1987)*

Table 14.3.1. Polytypic patterns and their frequencies in informative sites for determining the parsimonious tree for the human (H), chimpanzee (C), gorilla (G), orangutan (O), and gibbon (B). Singular mutations are listed separately *(From Nei, 1987)*

Полиморфный тип	***Частота***	***Полиморфный тип***	***Частота***
I. Информативные сайты			
a. (HC)-(GOB)	10	i. (GB)-(HCO)	8
b. (HG)-(COB)	5	j. (OB)-(HCG)	29
c. (HO)-(CGB)	2	k. G-(HC)-(OB)	3
d. (HB)-(CGO)	4	l. O-(HG)-(CB)	2
e. (CG)-(HOB)	10	m. O-(HB)-(CG)	1
f. (CO)-(HGB)	2	n. B-(HG)-(CO)	2
g. (CB)-(HGO)	4	o. B-(HO)-(CG)	1
h. (GO)-(HCB)	7	Total	90
II. Единичные мутации			
p. Человек	20	s. Орангутан	59
q. Шимпанзе	26	t. Гиббон	79
r. Горилла	26	Суммарно	210

Примечание. Единичные мутации показаны отдельно.

Полиморфный тип a, (HC)-(GOB) показывает, что OTE H и C обладают одним общим нуклеотидами, тогда как G, O и B - обладают другим общим. Символ G-(HC)-(OB) показывает, что OTE G имеет уникальный нуклеотид, тогда как H и C обладают одним общим нуклеотидом, также как O и B обладают общим нуклеотидом. Мы отметили, что некоторые полиморфные типы совместимы с топологией (A) на рис. 14.3.3 при допущении единичной нуклеотидной замены, тогда как другие нет. Например, полиморфный тип (a) совместим с топологией (A), при допущении единственной нуклеотидной замены, поскольку предполагается, что единственная вовлеченная замена произошла между узлами α и β на рис. 14.3.3, A. Однако полиморфный тип (b) требует, по меньшей мере, двух замен, чтобы быть совместимым с топологией (A), тогда как требуется только одна, чтобы быть совместимым с топологией (C). Поэтому тип (b) совместим с топологией (C), но не с топологией (A). Список совместимых полиморфных типов для каждой из четырех топологий сведен в Таблицу 14.3.2. Чтобы вычислить минимальное число замен, требуемых для топологии, мы должны рассмотреть минимальные числа для всех совместимых и не совместимых полиморфных типов и просуммировать их. Полученные результаты также представлены в Таблице 14.3.2. Это вычисление позволило определить, что топология (B) требует 145 замен и что это число наименьшее для четырех проанализированных топологий. В данном случае имеется еще 11 не укорененных деревьев, но ни одно из них не имеет такого минимального числа замен, равного 145. Поэтому, топология (B) выбирается как максимальное дерево парсимонии.

Таблица 14.3.2. Сопоставимые полиморфные типы с четырьмя топологиями (рис. 14.3.3) и минимальным числом нуклеотидных замен, требуемым для каждой топологии *(По Nei, 1987)*

Table 14.3.2. Compatible polytypic patterns with the four topologies (Fig. 14.3.3) and the minimum number of nucleotide substitutions required for each topology *(From Nei, 1987)*

Топология	*Сопоставимые полиморфные типы*	*Число совместимых сайтов*	*Минимальное число замен*
(A) HCGOB	a, j, k	42	147
(B) CGHOB	e, j, k, m, o	44	145
(C) HGCOB	b, j, k, l, n	41	148

(D) CGOHB	d, e, m, o	16	173

Есть мнение, что для методов парсимонии не требуется допущения о приблизительном постоянстве темпа нуклеотидных или аминокислотных замен. Когда число замен на сайт мало, - это действительно так. Однако, если это число велико настолько, что параллельные и обратные мутации появляются достаточно часто, то методы парсимонии могут давать большие ошибки, в случаях, когда темп замен не постоянен или не совсем постоянен. Это правило выполняется, даже если проанализирована очень большая последовательность нуклеотидов или аминокислот (Felsenstein, 1978b). Более того, если число исследованных нуклеотидов или аминокислот мало и имеется много параллельных и обратных мутаций, методы парсимонии имеют высокую вероятность давать ошибочное дерево, даже если темп мутаций постоянен (Peacock, Boulter, 1975). Поскольку деревья парсимонии основываются на минимальных оценках реализованного числа замен, то трудно предложить статистический подход для проверки числа замен в различных таких деревьях. Ошибки зависят от структуры дерева, числа замен, неоднородности темпа замен между различными ветвями и т.д. Для решения этой проблемы был предложен (Templeron, 1983) не параметрический подход, базирующийся на рестрикционных данных. Однако было показано (Nei, Tajima, 1985; Li, 1986), что оценки парсимонии по числу изменений в рестрикционных сайтах, для полиморфных сайтов, подвержены систематической ошибке, зависимой от структуры дерева и числа изменений. Когда вовлечено четыре или большее число ОТЕ, то вообще даже неясно, какая нуль-гипотеза должна тестироваться (Cavender, 1981; Felsenstein, 1985). Очевидно, что этот вопрос нуждается в теоретической проработке.

Методы парсимонии обычно не дают оценок длины ветвей, как уже было упомянуто. Однако возможно ввести все предковые нуклеотиды по каждому информативному сайту и вычислить вероятности обладания ветвью каждым нуклеотидом, реализуя специальный подход (Dayhoff, Park, 1969; Fitch, 1971a; Goodman et al., 1974). При таком подходе, среднее число замен для каждой ветви в дереве может быть вычислено для всех информативных сайтов. После этого можно добавить единичные замены и получить общее число замен для каждой ветви. После того, как дерево максимальной парсимонии найдено, для оценки длины ветвей можно использовать метод Фитча и Марголиаша (Fitch, Margoliash, 1967a). Также как для методов средних расстояний, для деревьев парсимонии тоже трудно определить правильную топологию. Так, для группы, включающей человека и обезьян, чтобы достигнуть вероятности правильной топологии (>0.95) для линии человек – шимпанзе – горилла по мтДНК, необходимо как минимум сопоставление последовательности в 2600 пар нуклеотидов. Для ядерной ДНК это число должно быть еще больше, поскольку темп замен по ним меньше.

Методы максимального правдоподобия

Метод максимального правдоподобия для конструирования дерева был впервые применен Кавали-Сфорца и Эдвардсом (Cavalli-Sforza, Edwards, 1967). Эти авторы построили дерево, используя модель Броуновского движения, то есть, следуя Гаусовскому процессу. Однако в случае нуклеотидных или аминокислотных последовательностей эволюционные изменения скорее следуют Пуассоновскому распределению или другим вероятностным законам. Позднее этот алгоритм был использован другими авторами (Felsenstein, 1973b; 1981; Langley, Fitch, 1974). Последние авторы (Langley, Fitch, 1974) также использовали метод максимального правдоподобия для оценки ожидаемой длины ветвей и соответствующего эволюционного времени для данной топологии, когда фактическое число замен для каждой ветви известно. Мы вкратце рассмотрим оба подхода.

Метод Лэнгли и Фитча. Рассмотрим абстрактное филогенетическое дерево (рис. 14.3.4). Здесь x обозначает фактическое число замен для i–той ветви, а t_i –тое – это временной интервал. Мы задаемся целью оценить t_1, t_2 и t_3 используя метод максимального правдоподобия. Чтобы применить этот метод необходимо иметь определенную математическую модель изменения генов

или белков. Мы примем, что темп замен (λ) является постоянным в течение лет или поколений в ходе эволюционного процесса и что число замен в рассматриваемом гене или белке следует распределению Пуассона для любого временного интервала. То есть, вероятность того, что *x* замен произойдет в течение эволюционного времени *t* в одной линии есть: $exp(-\lambda t)\ (\lambda t)^x / x!$. Поэтому, вероятность обнаружения $x_1, x_2, \ldots x_6$ для дерева, представленного на рис. 14.3.4, равна (Nei, 1987):

$$L = (e^{-v1}\ v^{-x1} / x_1!) * (e^{-v1}\ v^{-x2} / x_2!)\ [e^{-(v1+v2)}\ (v_1 + v_2)^{x3} / x_3!)]\ [e^{-(v1+v2)}\ (v_1 + v_2)^{x4} / x_4!)], \quad (14.7)$$

где $v_1 = \lambda t_1$. Поскольку λ обычно является неизвестным, мы ищем максимально правдоподобные оценки v_i'-тых скорее, чем сами t_i'-тые. Они задаются следующей системой уравнений.

$$dLog_eL / dv_1 = -4 + (x_1 + x_2)/v_1 + (x_3 + x_4)/(v_1 + v_2) = 0, \quad (14.8a)$$

$$dLog_eL / dv_2 = -3 + (x_3 + x_4)/(v_1 + v_2) + x_5/(v_2 + v_3) = 0, \quad (14.8b)$$

$$dLog_eL / dv_3 = -2 + x_5/(v_2 + v_3) + x_6/v_3 = 0. \quad (14.8c)$$

Для уравнений можно найти численное решение. Мы можем использовать такой подход для любого числа ОТЕ, поскольку число параметров, необходимых для оценки, всегда меньше чем число фактических значений. Мы предполагаем, что λ является постоянным во всех филогенетических линиях. Однако оно может быть определено прямо из v_i'-тых тестом хи-квадрат (Langley, Fitch, 1974). По оценке этих авторов для 18 видов позвоночных, была обнаружена высокая гетерогенность, которая показала, что темп замен аминокислот в белках не является строго постоянным. И все же, число изменений оказывается линейной функцией эволюционного времени, когда *t* велико.

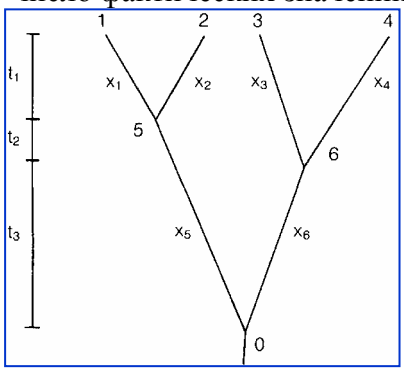


Рис. 14.3.4. Абстрактное филогенетическое дерево для четырех ОТЕ (1, 2, 3 и 4).

Здесь: t – представляет эволюционное время, тогда как x – обозначает фактическое число замен (По Nei, 1987).

Fig. 14.3.4. Artificial phylogenetic tree for four OTUs (1, 2, 3, and 4). t, represents the evolutionary time, whereas x, stands for the observed number of substitutions (From Nei, 1987).

Метод Фелзенштейна. Этот метод (Felsenstein, 1981) позволяет найти максимально правдоподобные оценки для фактически наблюдаемых нуклеотидных последовательностей в группе ОТЕ. Они вычисляются для многих различных топологий и топологии, которые показывают наибольшие значения правдоподобия, выбираются, как истинное дерево. Этот метод более сложен, чем предыдущий. Фелзенштейн обосновывает свой метод, рассматривая предельный случай бесконечного числа нуклеотидов. Фактически же используемое число обычно достаточно невелико. Возможно, что метод является не чувствительным к этим осложняющим факторам, как нередко бывает для многих статистических методов, но это надо проверять. Больше подробностей на эту тему можно найти в других источниках (Nei, 1987; Swofford et al., 1996).

В текущем разделе 14.3 мы опустили из рассмотрения некоторые методы парсимонии, которые стали популярны последние годы (Saitou, Nei, 1987), или только упомянули другие, как методы максимального правдоподобия. Эта область достаточно трудна для усвоения. Специально интересующимся этими вопросами рекомендуем обратиться к другим источникам (Saitou, Nei, 1987, Swafford et al., 1996). Мы еще обратимся к некоторым методам в ходе практического занятия. Мы также ознакомимся с некоторыми программными средствами, которые сейчас имеются в изобилии для конструирования деревьев и популяционно-генетического анализа. В своем обзоре Своффорд с соавторами (Swofford et al., 1996) цитируют 42 программы, имеющиеся для построения деревьев и популяционного-генетического анализа, как для операционной системы Windows, так и Macintosh. Наша цель – познакомится хотя бы с одной или двумя простейшими. В будущем желающие могут легко нарастить эти знания на основе освоенных программ.

14.4. ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКАЯ ТЕОРИЯ И ПОСТРОЕНИЕ ДЕРЕВЬЕВ

Молекулярная филогения основывается на числе нуклеотидных замен в ДНК или числе генных замен в геноме. Нуклеотидные или генные замены – это результат сложных процессов мутации, естественного отбора и генетического дрейфа. Поэтому важно рассмотреть влияние этих процессов на построение дерева. В обычной практике большинство молекулярных деревьев строится при использовании одного гена, исследованного у всех видов. Эти деревья часто рассматриваются как видовые деревья. Однако, как упоминалось ранее, генное дерево иногда очень отличается от видового дерева. Чтобы понять степень различия между этими деревьями, необходимо оценить специфику генных замен в популяциях. Такие вопросы, как нейтральность или селективная значимость, величина *Ne*, темп мутирования и величина миграции – все это влияет на точность оценок, как топологии дерева, так и времени с момента изоляции, прошедшего в различных линиях.

Многие авторы предпринимали попытки построить дерево на основании данных электрофореза белков, рассматривая каждый аллель, как частотно независимый кладистский признак. При этом почти никакого внимания не обращалось на процессы динамики частот аллелей и генных замен в популяциях, а также на причины их определяющие. Однако чтобы построить корректное филогенетическое дерево, необходимо знать эти причины. Рекомендации по использованию НПГМ и стандартного генетического расстояния Нея для аллозимных данных базируются именно на анализе таких причин (Nei et al., 1983; Nei, 1987). Тем не менее, в присутствие значительного эффекта бутылочного горлышка или влияния инбридинга на генетические расстояния, обнаруживается, что, НПГМ, например, для человеческих популяций, а также любой другой метод, дают бессмысленное дерево (Nei, Roychoudhury, 1982). Необходимо подойти к разрешению этих трудностей на основе понимания популяционно-генетических процессов, ведущих к филогенетическим изменениям. Это весьма сложная область. Недавние, кажущиеся и явные, противоречия между филогенетической и биологической концепциями видов являются безусловным доказательством этого (Avise, Wollenberg, 1997, Harrison, 1998). И все же, принимая во внимание множество трудностей, упомянутых и не упомянутых выше из-за краткости раздела, удивительно хорошее согласие деревьев, основанных на ферментных локусах и данных анализа ДНК, были получены для ряда групп животных (Nei, 1987; King, 1993; Kartavtsev, Hanzawa, 2007).

14.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 14

14.1. Выполните оценку процента стандартных отклонений (s_{FM}), индекса искажения (***Dt***) или кофенетической корреляции (***r***), а также времени дивергенции с момента изоляции (***t***) используя программы GENEPOP или ARLEQUINE и примеры в них.
14.2. Посетите Интернет и используйте три любые последовательности из генетического банка данных для применения программ MEGA или ARLEQUINE для построения филогенетического дерева.
14.3. Сопоставьте дерево для человека и обезьян на рис. 14.5.1 с двумя предыдущими (рис. 14.2.1 и рис. 14.3.3, А-В). Каким должно быть консенсус-дерево для линии HU-CH-GO? Нарисуйте его.
14.4. Меняется ли скорость эволюции вируса гриппа во времени (рис. 14.5.2) и как это согласуется с теорией молекулярных часов?

Процент стандартных отклонений (s_{FM}) для метода Фитча-Марголиаша и других. Статистика определяется следующим образом:

$$s_{FM} = [1/\{n(n-1)\} \{2\Sigma ij \{(dij - eij)/dij\}^2\}]^{1/2} * 100$$

для ***I<j,*** где ***n*** - это число исследуемых ОТЕ, а ***dij*** и ***eij*** – это фактическое и патристическое расстояния между ОТЕ ***i*** и ***j***, соответственно. Патристические расстояния между ОТЕ ***i*** и ***j*** равны сумме длин всех ветвей, связывающих две ОТЕ. Например, патристическое расстояние между человеком и гориллой в дереве на рис. 14.2.1 равно a+c+d=0.108. Для какого дерева s_{FM} является наименьшим, то и принимается как наилучшее; s_{FM} = 1.3 для дерева человека и обезьян (рис. 14.2.1, Табл. 14.2.1). Поскольку для двух других вариантов, когда рассматриваются изменения в линии GO-HU-CH, s_{FM} = 3.11 (HU->CH-GO) и s_{FM} = 1.88 (CH->GO-HU), то, наилучшим является дерево с величиной s_{FM} = 1.3, которое представлено на рис. 14.2.1. По мнению Нея (Nei, 1987, p.302) s_{FM}, или s_O (Farris, 1972) имеют смысл, когда эти статистики достаточно велики, а большие ***r*** - вообще не гарантируют, что дерево является правильным.

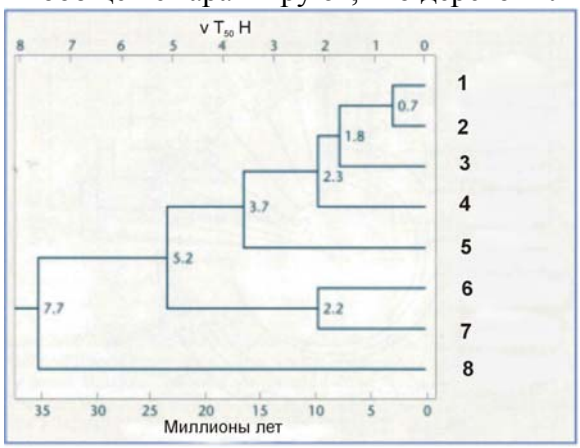


Рис. 14.5.1. Филогенетическая последовательность приматов и обезьян Старого света, полученная на основе анализа гибридизации ДНК. Числа в точках ветвления – это измерения ∇T50H. Эволюционные точки ветвления датированы с использованием ископаемых остатков и дивергенции нуклеотидов. Виды: 1 – шимпанзе, Pan troglodytes,2 - пигмейное шимпанзе, P. poniscus, 3 – человек, Homo sapiens, 4 – горилла, Gorilla gorilla, 5 – орангутан, Pongo pigmeus, 6 – обыкновенный гиббон, Hilobates lar, 7 – симанг, H. sindactilus, 8 – обезьяны старого света, Cerocopithecidae (Данные Sibley; заимствовано из Klug, Cummings, 2002).

Fig. 14.5.1. Phylogenetic sequence of hominoid primates and Old World monkeys as estimated by DNA hybridization. The numbers at the branch points are DT50H measurements. The evolutionary branch points are dated using the fossil record and nucleotide divergence (Sibley's data; adopted from Klug, Cummings, 2000).

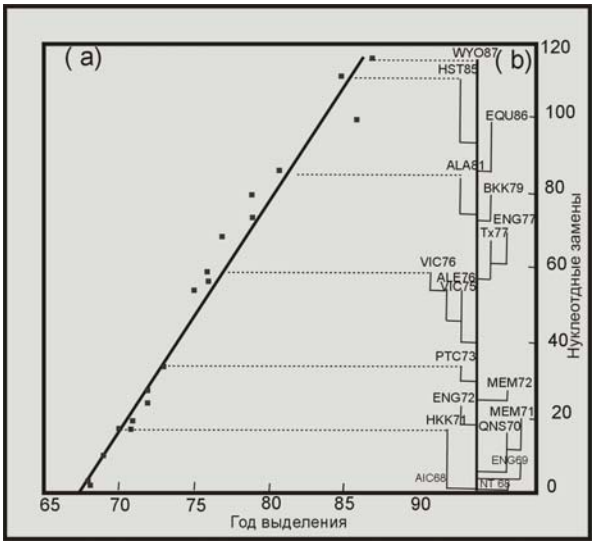


Рис. 14.5.2. Молекулярные часы для гена гемагглютинина вируса гриппа А.

(a) Число нуклеотидных различий между первым и каждым последующим выделением показано для соответствующего года выделения вирусного штамма. (b) Оценка филогении изолятов (По Fitch, et al. 1991, с упрощениями).

Fig. 14.6.2. Molecular clock in the influenza A hemagglutinin gene. (a) Number of nucleotide differences between the first isolate and each subsequent isolate as a function of year of isolation. (b) Estimate of the phylogeny of the isolates (From Fitch, et al. 1991. Positive Darwinian evolution in human influenza A viruses. PNAS 88:4270-73).

ГЛАВА 15. ПРАКТИЧЕСКАЯ РАБОТА С ПОСЛЕДОВАТЕЛЬНОСТЯМИ НУКЛЕОТИДОВ: РЕДАКТИРОВАНИЕ, РЕГИСТРАЦИЯ, ВЫРАВНИВАНИЕ И КОНСТРУИРОВАНИЕ ДЕРЕВЬЕВ

ГЛАВНЫЕ ВОПРОСЫ:

15.1. Редактирование последовательностей и их регистрация в генном банке.
15.2. Формат данных и доступные генные банки.
15.3. Выравнивание последовательностей.
15.4. Подбор оптимальной модели нуклеотидных замен.
15.5. Построение деревьев с помощью пакета программ MEGA-3 (MEGA-4). Аннотация возможностей программ PAUP, MrBayes и некоторых других.
15.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 15.

РЕЗЮМЕ

1. Получив с помощью молекулярно-генетических методов данные о последовательности нуклеотидов какого-либо гена или его фрагмента, необходимо провести их редактирование и регистрацию в генном банке. Одним из наиболее удобных программных пакетов (ПП) для первичного редактирования последовательностей является «**Хромас**» (**Chromas-pro**, http://www.flu.org.cn/en или http://www.technelysium.com.au/chromas.html).
2. Легкий доступ и большие возможности для регистрации, извлечения и первичного анализа массивов последовательностей предоставляет **генный банк, GenBank** (http://www.ncbi.nlm.nih.gov). Если последовательности небольшие по длине и представлено немного особей, линий и т.п., то подходит интерактивный режим регистрации через сеть Интернет, с использованием программной утилиты банка **Bankit**.
3. Перед началом молекулярно-филогенетического анализа надо провести выравнивание последовательностей. Затем необходимо осуществить подбор оптимальной модели нуклеотидных замен для набора исследуемых последовательностей. Удобным средством для этого является программа MODELTEST.
4. Для построения молекулярно-филогенетических деревьев имеется много ПП. Ресурсы для построения деревьев, которые просты в использовании и легко доступны через Интернет, дают ПП **MEGA-3** и **MEGA-4** (http://www.megasoftware.net/). ПП MEGA рассмотрен в той мере подробности, чтобы неискушенный пользователь мог провести простой анализ и построить филогенетическое дерево. Дана также аннотация возможностей программ PAUP, MrBayes и PHYLIP.

Chapter 15. PRACTICAL TRAINING WITH NUCLEOTIDE SEQUENCES: EDITION, SUBMISSION, ALIGNMENT, TREE BUILDING AND ANALYSIS OF PHYLOGENY

SUMMARY

1. When nucleotides sequence for a gene or its section is obtained using methods of molecular genetics, it is necessary to make the next step, which includes the sequence edition and its submission to a gene

bank. One of most convenient program packages (PP) for a primary edition of sequences is **Chromas-pro**, http://www.flu.org.cn/en or http://www.technelysium.com.au/chromas.html).
2. Easy access, big opportunities for submission, retrieving and preliminary investigation of sequences are provided by the GenBank (http://www.ncbi.nlm.nih.gov). If sequences are not big in size and few individuals, lines, etc. are submitted to GenBank, they can be submitted in an interactive mode through a Web Net and usage of the **Bankit** program utility.
3. Before the beginning of the molecular genetic analysis, alignment of sequences should be done. After that it is necessary to choose the optimal model of nucleotide substitutions for a set of analyzed sequences. MODELTEST program is convenient tool for this.
4. There are many PP that can be used for molecular phylogenetic tree building. Resources for tree building that are simple in usage and easily accessed via Web provides PP **MEGA-3** and/or **MEGA-4** (http://www.megasoftware.net/). PP MEGA is examined in detail in order that an inexperienced user can make a simple analysis and build a phylogenetic tree. An annotation of PAUP, MrBayes, and PHYLIP software was also provided.

15.1. РЕДАКТИРОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ИХ РЕГИСТРАЦИЯ В ГЕННОМ БАНКЕ

После выяснения последовательности нуклеотидов в ДНК с помощью молекулярно-генетических методов и использования на последнем этапе секвенатора возникает необходимость ее редактирования и регистрации в одном из существующих генных банков. Секвенаторы таких производителей, как Applied Biosystems и Amersham MegaBace предоставляют на выходе файлы с графической информацией в виде пиков (хроматограмм) и их интерпретацией для каждого из четырех основных типов нуклеотидов C, T, A, G или замещающей их буквой N, если интерпретация затруднена или другими буквами, если она неоднозначна. Различные редакторы позволяют перевести эту информацию в другой файл, представляющий буквенную последовательность соответствующих нуклеотидов в гене или его фрагменте. Однако эта информация еще не вполне подходит для содержательного анализа степени сходства или различия последовательностей с её последующим филогенетическим или иным отображением.

Для того, чтобы последовательность приобрела необходимую законченность её надо отредактировать. Многим требованиям для редактирования отвечают программные пакеты (ПП) MEGA-3 или MEGA-4 (http://www.megasoftware.net/), GeneDOC (http://www.nrbsc.org/) и др. Одним из наиболее удобных ПП для первичного редактирования последовательностей является «Хромас» (Chromas-pro, http://www.flu.org.cn/en или http://www.technelysium.com.au/chromas.html). Предлагаемая сейчас версия (Chromas-pro 2.31) позволяет осуществить множество опций редактирования.

- Открывает файлы хроматограмм разных форматов двух выше перечисленных производителей секвенаторов.
- Открывает файлы хроматограмм формата SCF, создаваемых программами ряда других секвенаторов ALF, Li-Cor, Visible Genetics OpenGene, Beckman CEQ 2000XL, CEQ 8000 и др.
- Просматривает генотипические файлы ПП Genescan.
- Позволяет сохранить файлы в форматах SCF или Applied Biosystems.
- Позволяет печать хроматограмм в увеличенном или уменьшенном виде, а также со сжатием до формата одной страницы.
- Экспортирует последовательности в текстовом формате или в формате с нумерацией оснований, а также в форматах FASTA, EMBL, GenBank и GCG.
- Копирует последовательности в оперативное запоминающее устройство, ОЗУ (clipboard) в виде простого текста или в формате FASTA для переноса в другие ПП и редакторы.

- Экспортирует последовательности из текстовых записей файлов хроматограмм автоматически удаляя векторные последовательности.
- Инвертирует последовательности и хроматограммы и выполняет их комплементацию.
- Ищет последовательности по точному спариванию или оптимальному выравниванию.
- Показывает трансляции в трех окнах, сопровождая их нуклеотидными последовательностями.
- Копирует графическую картинку хроматограммы для вставок в документы или презентации.
- Позволяет создавать консенсусные последовательности.

Таким образом, главное что позволяет «Хромас» – это провести сопоставление последовательностей, удалить векторные вставки в начале и конце цепей, инвертировать антипараллельные последовательности (цепи), создать консесную последовательность и записать информацию в удобном для последующих расчетов виде. Важно проверить на нескольких повторах точность описываемой последовательности. Обычный прием – это прогон или копирование ДНК с применением ПЦР для антипараллельных цепей в прямом 5’-3’ и обратном 3’-5’ направлениях с использованием соответствующей пары праймеров, типа F1 (Forward) и R1 (Reverse) и т.п. Опыт показывает, что точнее всего удается установить последовательность с использованием четырех праймеров. Обычно только при таком подходе консенсусная последовательность не содержит ошибок. Вид последовательностей в графическом и буквенном виде в редакторе последовательностей ПП «Хромас» представлен на рис. 15.1.1.

Рис. 15.1.1. Графическое и символьное представление фрагмента последовательности гена цитохромоксидазы 1 (Co-1) камбалы, Liopsetta pinifasciata.

Секвенирование проведено с использованием прибора ABI-3100 (Applied Biosistems, USA). Четыре повтора последовательности (1K_F2 и т.д., слева) представлены пиками и в виде их буквенной интерпретации. После инвертирования антипараллельных цепей (1KR1_L_p и 1K_R2) и проведения их комплементации они автоматически выравниваются. Консенсусная последовательность, находящаяся в процессе редактирования отображена сверху. В оригинале кривые хроматограмм и буквы для разных нуклеотидов изображены для наглядности разным цветом.

Fig. 15.1.1. Graphical and symbolical representation of a partial cytochrome oxidase 1 (Co-1) sequence in flounder, Liopsetta pinifasciata.

Sequencing made on the machine ABI-3100 (Applied Biosistems, USA). Four repeats of single individual sequence (1K_F2, etc. left column) are represented with peaks and their letter's interpretation. After inversion of ant- parallel chains (1KR1_L_p and 1K_R2) and making complementation their have been automatically aligned. Consensus sequence that is being under edition is shown on top. Originally lines on chromatogram and four different letters of nucleotides are given in different color.

После редактирования в «Хромас» или ином редакторе полученная последовательность нуклеотидов может быть зарегистрирована в одном из трех признанных генных банков. Все они связаны в сеть и обмениваются данными. Поэтому, зарегистрировав последовательность, например в GenBank (http://www.ncbi.nlm.nih.gov), в согласованный с авторами срок гарантируется доступ к последовательности любому пользователю Интернета. В генном банке проводится тестирование качества последовательности. При необходимости ведется работа с авторами для уточнения тех или иных характеристик записи. На представленном выше сайте GenBank, имеются различные утилиты для регистрации последовательностей, а также множество средств извлечения последовательностей, их тестирования и анализа. Для простых последовательностей отдельных генов или их фрагментов удобна утилита **Bankit**, которая позволяет зарегистрировать в интерактивном режиме одну или серию последовательностей, с присвоением им сначала первичных номеров, а после проверки – номеров доступа генного банка. Часть извлекаемой информации для уже зарегистрированной нуклеотидной последовательности на сайте GenBank отражена на рис. 15.1.2.

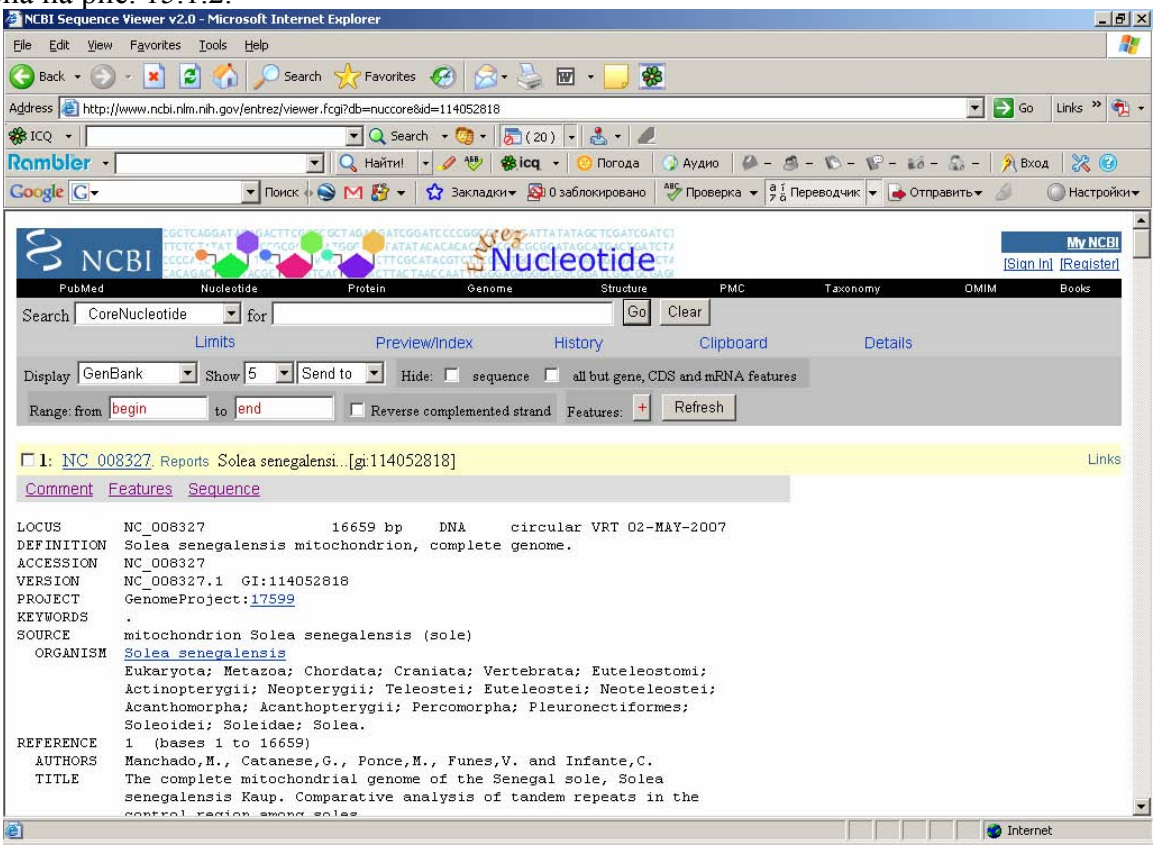


Рис. 15.1.2. Фрагмент окна генного банка GenBank. Представлены данные для полного митогенома одного из видов рыб отряда Камбалообразные.

Fig. 15.1.2. Sample of the GenBank window. Data for complete mitogenome one of Pleuronectiformes order are represented.

15.2. ФОРМАТ ДАННЫХ И ДОСТУПНЫЕ БАНКИ ДАННЫХ

Зарегистрированные последовательности будут доступны для всеобщего использования в оговоренный срок, после опубликования данных в открытой печати, обычно через 6-12 месяцев. Отдельная последовательность доступна для вывода в формате GenBank, в FASTA-формате и др. В первом случае она выглядит следующим образом (рис. 15.2.1).

```
  1 gtgcctgagc cggaatagtc ggggacaggc ctaagtctgc tcattcgagc agagctaagc
 61 caacctgggt gctctcctgg gagacgacca aatttataac gtaatcgtca ccgcacacgc
121 ctttgtaata atcttcttta tagtaatacc aattatgatn cggagggttc ggaaactgac
181 ttattccatt aataattggg gccccccgnat atggccttcc ctcgaataaa taacatgagt
241 ttctgacttc tacccccatc ctttctcctc cttctagcct cttcaggncg tcgaagctgg
301 ggcagggaca ggatgaaccg tgtatccccc actagctgga aatctagcac acgccggagc
361 atcggtagac ctcaccattt tctctcttca ccttgccgga atttcatcaa ttctaggggc
421 aatcaacttt attactacta tcatcaacat gaaaccaaca gcagtcacta tgtaccaaat
481 cccactattt gtctgagccg tactaatcac cgcacgtcct tcttcttctt tcacactacc
541 acgtcactgg ccgctggcat tacaatgcta ctgactagac cgcaacacta aacacaaaca
601 cttctttgac cctgcyg
```

Рис. 15.2.1. Фрагмент записи неполной последовательности гена Co-1 камбалы, Pseudopleuronectes obscurus.

В левом столбце приведены порядковые номера первых нуклеотидов в ряду. Нуклеотиды сгруппированы по 10, с общей численностью по 60 на один ряд.

Fig. 15.2.1. Partial sequence of Co-1 gene obtained in flatfish, Pseudopleuronectes obscurus.

In the left column ordinal numbers for first nucleotides in a row are given. Nucleotides are grouped by 10 with a total number of 60 in a row.

Полная запись содержит также дополнительную информацию, начиная с локуса, номера доступа, кодируемого продукта и далее включая источник выделения ДНК, таксономическую информацию, данные об авторах и др. (см. рис. 15.1.2). В FASTA формате информации меньше, но зато такой файл пригоден для прямого импортирования многими ПП, для выполнения последующих расчетов и построения филогенетических деревьев.

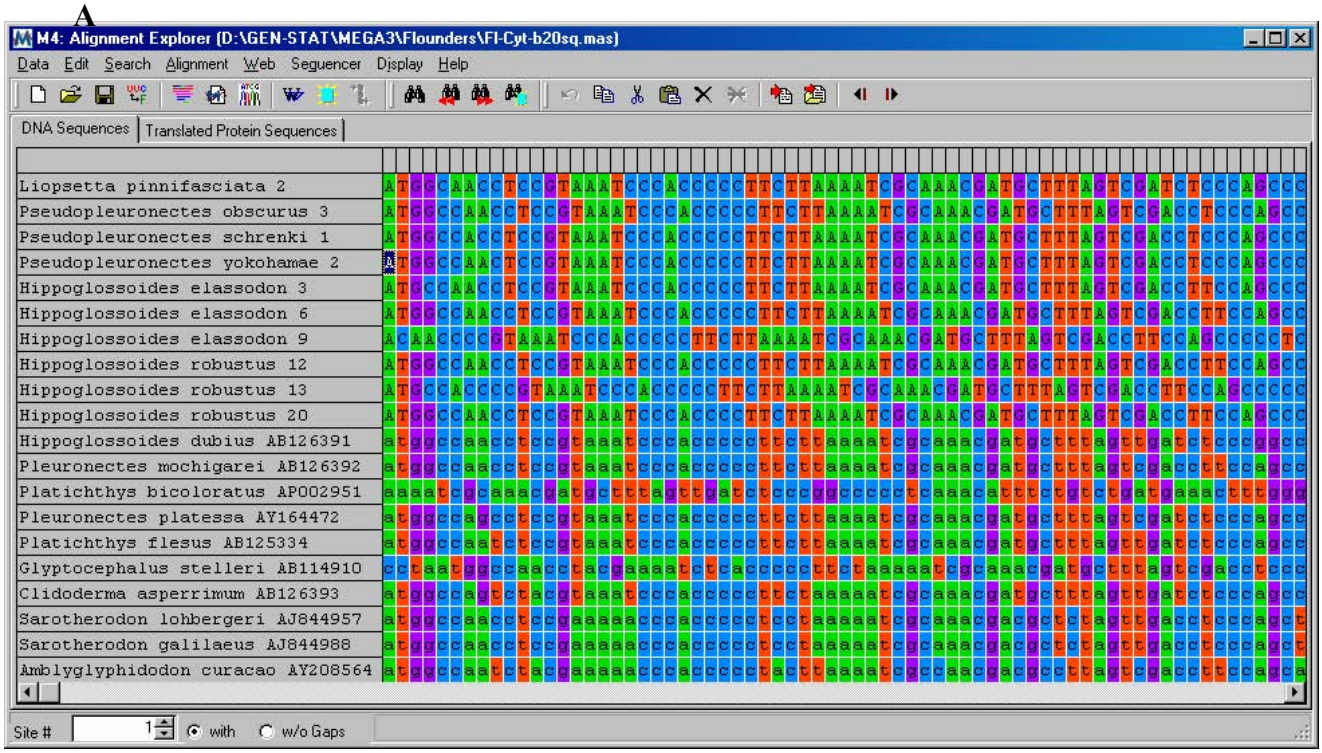
Для регистрации последовательности через Интернет в интерактивном режиме или по электронной почте общедоступны в настоящее время 3 генных банка, соединенных в глобальную систему. Кроме упомянутого ранее банка GenBank на сайте NCBI (США), можно зарегистрировать свои данные в европейском банке последовательностей ДНК, EMBL (http://www.ebi.ac.uk/embl/), в ДНК банке данных Японии, DDBJ (DNA Data Bank of Japan, http://www.ddbj.nig.ac.jp/searches-e.html). Имеются также локальные банки ДНК, например японского центра биоресурсов, RIKEN (http://www.brc.riken.jp/lab/dna/en/), «северный» банк, NGB (http://www.ngb.se) и др. Кроме нуклеотидных последовательностей к регистрации принимаются и полипептидные последовательности. Однако большинство регистрируемых последовательностей из-за относительной простоты их расшифровки – это ДНК-последовательности, которые затем транслируют в белковые (ферментные) полипептидные последовательности.

15.3. ВЫРАВНИВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Выравнивание (alignment) последовательностей – очень важная процедура, предваряющая их сравнительный количественный анализ, с расчетом мер сходства-различия, оценки гомологии и, наконец, построением различных филогенетических деревьев. Существует несколько алгоритмов выравнивания, осуществляемых различными редакторами

последовательностей. В этом разделе разобрано выравнивание последовательностей посредством редактора CLUSTAL W, адаптированного для ОС Windows. Для осуществления выравнивания последовательностей относительно друг друга надо провести сначала загрузку выбранных последовательностей в редактор. Загрузить последовательность можно тремя способами: 1) непосредственно записав последовательность нуклеотидов в соответствующей строке окна редактора, 2) импортировав ее из предварительно подготовленного файла и 3) скопировав через ОЗУ вектор символов нуклеотидной последовательности. На рис. 15.3.1 показан интерфейс редактора CLUSTAL W (Thompson et al. 1994), интегрированного в ПП MEGA, с фрагментом серии последовательностей гена *Cyt-b* камбалообразных рыб и двух видов окунеобразных рыб до и после выравнивания.

A



B

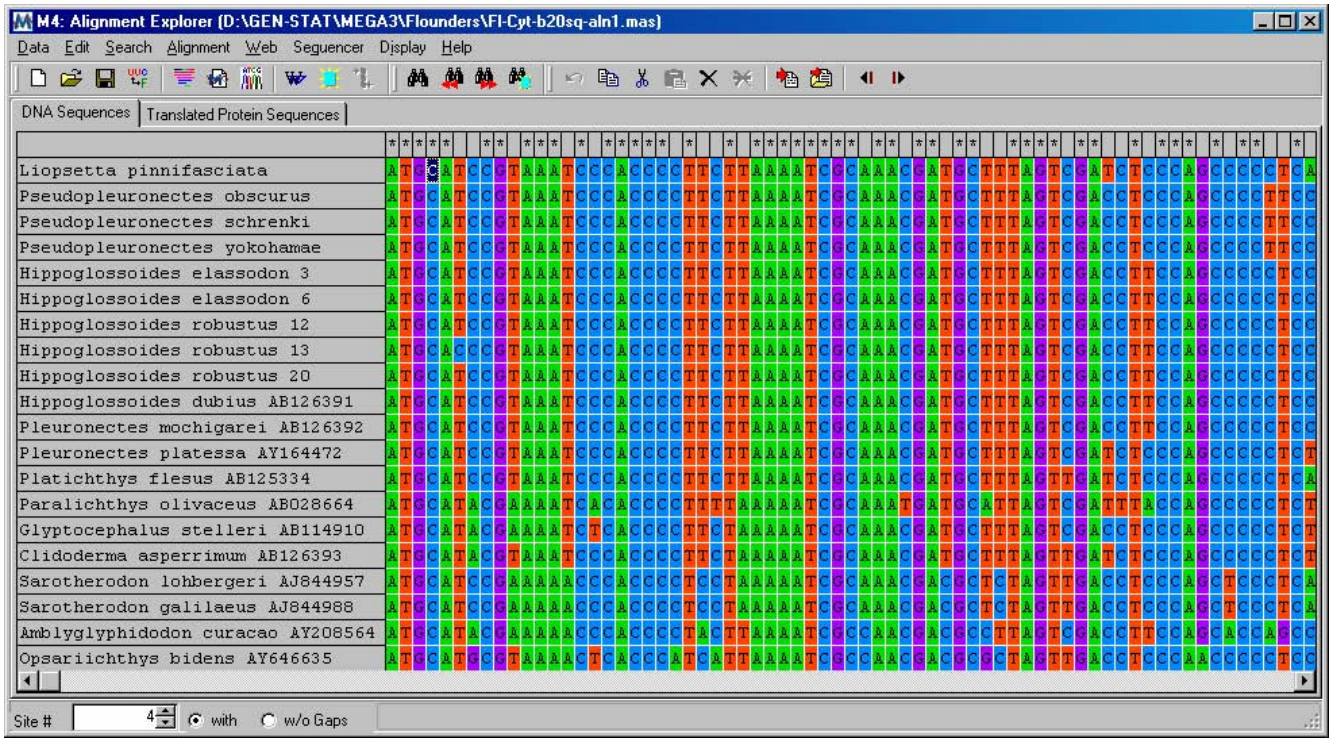


Рис. 15.3.1. Окна редактора выравнивания CLUSTAL W (Alignment explorer) в MEGA, с фрагментами последовательностей нуклеотидов гена Cyt-b до выравнивания (A) и после завершения выравнивания (B).

Оттенками серого цвета (в оригинальных файлах – разным цветом) показаны группы общих сайтов. Звездочкой отмечены сайты со 100%-ной гомологией нуклеотидов, то есть, – идентичные во всех последовательностях.

Fig. 15.3.1. A window of the Alignment explorer in CLUSTAL W of MEGA with a sample of nucleotide sequences at Cyt-b gene before alignment (A) and after alignment (B).

Different gray colors (Different color in original software) show sets of identical sites. With an asterisk sites are indicated that have 100% homology, i.e. identical in all sequences.

Загрузка последовательностей в данном случае (рис. 15.3.1) проводилась через ОЗУ. Запустив ПП MEGA-3 (MEGA-4), выбираем из главного меню: «**выравнивание**» → «**редактор выравнивания/Clustal**»→ «**создать новое выравнивание**» (**Alignment → Alignment explorer/Clustal → Create a new alignment**). В последней опции меню на самом деле имеются три возможности: «**создать новое выравнивание**» (**Create a new alignment**), «**открыть сохраненную сессию выравнивания**» (**Open a saved alignment session**), «**вывести последовательность из файла**» (**Retrieve sequence from a file**). Загрузив последовательности, автор сталкивается, как правило, с их неодинаковой размерностью: последовательности имеют различную длину, начало и концы последовательностей не совпадают и, кроме того, отдельные последовательности имеют пропуски (Gaps), которые не совпадают у различных индивидуумов и видов. Все эти проблемы и позволяет разрешить выравнивание. Технически, чтобы начать исполнение программы необходимо выделить все последовательности и запустить опцию «**выравнивание**» из главного меню**.** В итоге появляется диалоговое окно (рис. 15.3.2). На рис. 15.3.2 показано два диалоговых окна для реализации выравнивания в два этапа.

А. Первый этап. В. Второй этап.

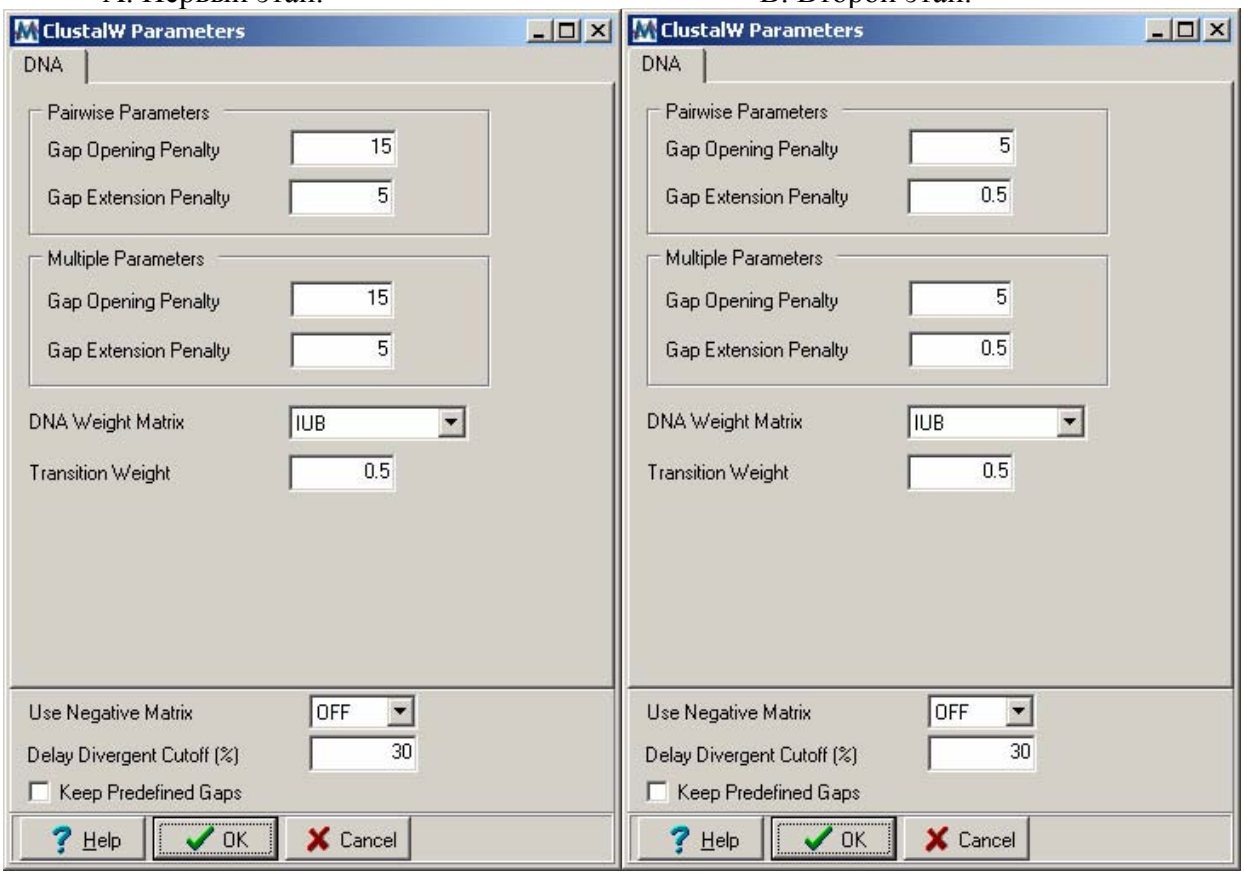


*Рис. 15.3.2. Диалоговые окна редактора выравнивания последовательностей CLUSTAL W в MEG , показывающие различные опции для штрафов (Penalties) при попарном выравнивание (**Pairwise Parameters**) и множественном выравнивании (**Multiple Parameters**).*
*Fig. 15.3.1. A dialog window of the Alignment explorer in CLUSTAL W of MEGA exemplified different options for penalties under pair wisealignment (**Pairwise Parameters**) and multiple alignment (**Multiple Parameters**).*

Нажав на опцию «**выполнить**» (**ОК**) запускаем выравнивание. Выравнивание – это деликатная процедура и требует терпения. Рекомендуется поэкспериментировать с величиной штрафов для каждого своего набора последовательностей. Алгоритм выравнивания устроен таким образом, что при увеличении штрафа получаем большее число выявленных пропусков (Gaps), обусловленных делециями или вставками, и относительно высокую гомологизацию нуклеотидов (или аминокислот, если выравниваются полипептиды) последовательностей. Однако слишком высокие штрафы могут привести к потере некоторой доли нуклеотидов, которые на самом деле гомологичны, но представлены только в части соответствующих сайтов последовательностей. Опыт работы автора с последовательностями различных митохондриальных генов показывает, что штрафы в диапазоне 15-30 за открытие пропуска и 0.5-8 за удлинение пробелов вполне достаточны для первого этапа выравнивания. После завершения работы программы CLUSTAL W, с установками в нашем примере (рис. 15.3.2, А): «**штрафы за открытие пропусков**» (**Gap Opening Penalties**) – 15 единиц и «**штрафы за удлинение пропусков**» (**Gap Extension Penalties**) – 5 единиц, как для попарных, так и для множественных процедур, появляется окно с обозначенными пропусками и гомологично сориентированными (выровненными) последовательностями (рис. 15.3.3). При этом выявляются самые большие (основные) пробелы и последовательности выглядят следующим образом (рис. 15.3.3).

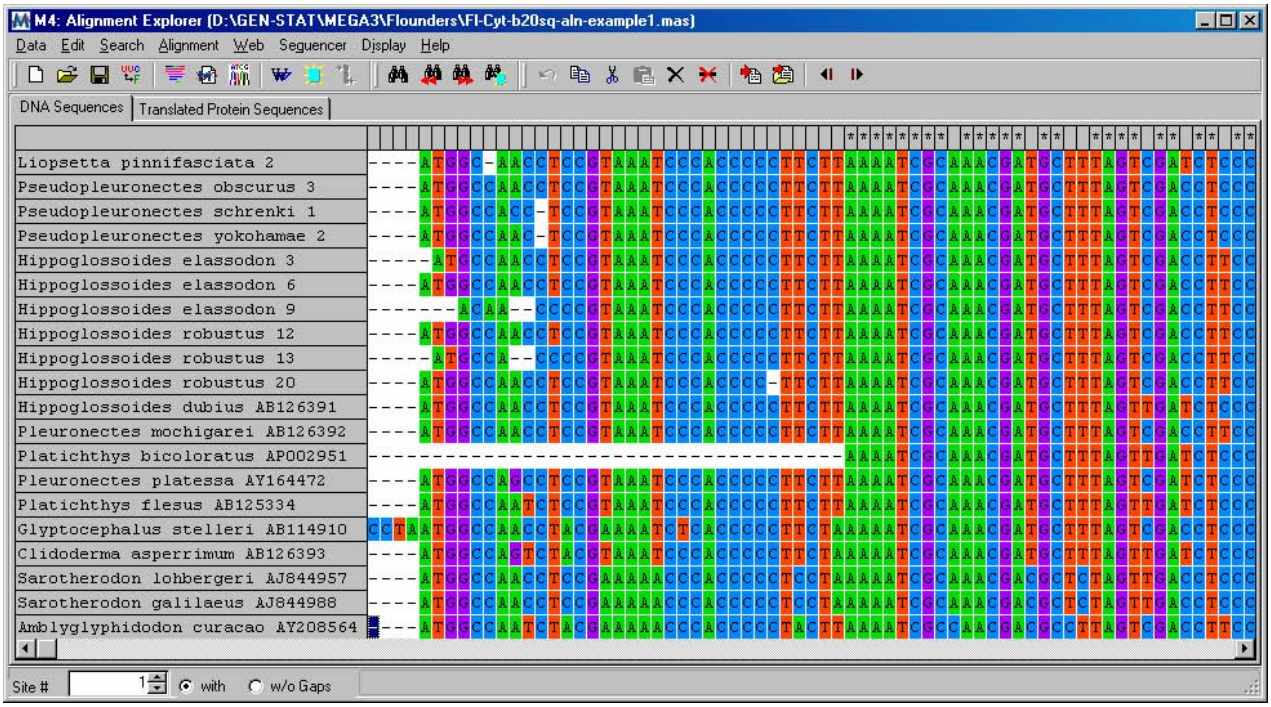


*Рис. 15.3.3. Окно редактора выравнивания CLUSTAL W в MEGA, с фрагментами последовательностей нуклеотидов гена Cyt-b после исполнения опции «**выравнивание**» и реализации первого этапа процесса выравнивания.*

Видны пропуски (Gaps) выравниваемых последовательностей, после удаления которых последовательности приобретают окончательный вид, как изображено на рис. 15.3.1, B.

Fig. 15.3.3. A window of the Alignment explorer in CLUSTAL W of MEGA with a sample of nucleotide sequences at Cyt-b gene after running the option "Alignment" and before gaps removal.

Gap are seen as blank sectors (white color).They will be removed and after that sequences take the mode as in Fig. 15.3.1, B.

Последовательности просматривают, и затем большие пропуски удаляют вручную. Можно удалить все пробелы, используя программные средства. Затем вызывают снова диалоговое окно CLUSTAL W и запускают выравнивание с уменьшенными величинами штрафов (рис. 15.3.2, B). Теперь после завершения работы программы удаляют все пропуски, записывают полученный файл в формате, необходимом для последующего исследования, и проводят дальнейший анализ: сопоставление нуклеотидного состава, встречаемости нуклеотидов по позициям в кодоне или суммарно, частот использования кодонов, величину попарных генетических расстояний между последовательностями, построение филогенетических деревьев и т.п.

15.4. ПОДБОР ОПТИМАЛЬНОЙ МОДЕЛИ НУКЛЕОТИДНЫХ ЗАМЕН

Для выбора наиболее подходящей к конкретным эмпирическим данным модели удобна программа MODELTEST 3.06 (Posada, Grandal, 1998) и более поздние версии 3.6 - 3.7. Полезную информацию о свойствах моделей и их применимости в зависимости от специфики данных можно также найти в других источниках (Nei, Kumar, 2000; Hall, 2001; Sanderson, Shaffer, 2003; Felsenstein, 2004).

Для того, чтобы использовать MODELTEST придется освоить предварительно ПП PAUP, так как эта программа использует некоторые модули PAUP. Принципиально работа с программой достаточно проста и включает 5 этапов.

1. Подготовить рабочий файл в формате **Nexus** (.nex), с последовательностями нуклеотидов и необходимыми идентификаторами параметров, в соответствии с требованиями для ПП PAUP;

2. Войти в сайт программы MODELTEST, доступный через Интернет, скачать все рекомендуемые модули и скопировать в подготовленный ранее nex-файл отдельный файл «modelblockPAUPb10.txt», распространяемый с программой (это относится к версии PAUP 4b10 для Windows);

3. Запустить инсталлированный ранее ПП PAUP 4b10 (исходный файл с данными лучше переименовать) и запустить исполнение подготовленного рабочего файла;

4. При нормальном завершении программы в той же директории, из которой запускался рабочий файл, появится новый файл с именем «model.scores»;

5. Теперь нужно запустить из окна ОС DOS собственно саму программу, лучше MODELTEST 3.7, из директории, содержащей исполнительный программный файл – «modeltest3.7.win.exe». Соответствующие идентификаторы в командной строке будут таковы: «modeltest3.7.win <model.scores> test.out» (последний выходной файл может иметь произвольное имя).

В выходном файле будет представлена необходимая информация и даны параметры двух наилучших моделей при оценке из 67 вариантов моделей по методу максимального правдоподобия и по информационному критерию Акайки (Acaike).

15.5. ПОСТРОЕНИЕ ДЕРЕВЬЕВ С ПОМОЩЬЮ ПАКЕТА ПРОГРАММ MEGA-3 (MEGA-4). АННОТАЦИЯ ВОЗМОЖНОСТЕЙ ПРОГРАММ PAUP, MRBAYES И НЕКОТОРЫХ ДРУГИХ

Опции или параметры моделей и сами модели для расчета филогенетических деревьев предоставляются различными пакетами программ PAUP* (Swofford, 2000), MEGA-3 (MEGA-4) (Kumar et al., 1993; 2000) и другими. Удобный интерфейс и разнообразные статистические возможности, в том числе для анализа последовательностей ДНК и ДНК-маркеров, разработаны в пакете ARLEQUINE (Schneider et al., 2000). Хорошим руководством для филогенетического анализа является книга Хэлла (Hall, 2001). Это руководство ориентированно в основном на PAUP*, но в нем излагаются и общие принципы филогенетического анализа изменчивости по ДНК и полипептидам. Кроме того, в книге приведены конкретные примеры и даны указания по использованию ПП CLUSTAL X, MrBayes и др.

Начало основной аналитической работы в ПП MEGA-3 и MEGA-4 может быть осуществлено сразу по завершении выравнивания (см. Раздел 15.3). Закрываем сохраненный файл редактора выравнивания (Alignment Explorer; он имеет расширение **.mas**). При этом появляется окно с предупреждением: «**сохранить файл для MEGA**», с опциями: «**да**», «**нет**», «**сброс**» (**Save data to MEGA file: Yes, No, Cancel**). Выбор опции «**да**» открывает следующее окно с именем файла, готового к сохранению на диске. Имя файла по умолчанию предполагается, таким же, что дано ранее для файла выравнивания, но имеет расширение **.meg**. Выбрав опцию «**сохранить**», запускаем рабочий процесс ПП MEGA. Перед открытием meg-файла для исполнения надо в появившемся окне указать, какая последовательность обрабатывается: «**данные с белок-кодирующей нуклеотидной последовательностью**» (**Protein-coding nucleotide sequence data**), с альтернативой «**да**», «**нет**». Наконец появляется диалоговое окно с вопросом: «**открыть файл с данными в MEGA**», «**да**», «**нет**». Выполнив «**да**», получаем рабочий файл MEGA, с открытием специального «**редактора последовательностей**» (**Sequence Data Explorer**) (рис. 15.5.1). В ПП MEGA-3 и MEGA-4 рабочие файлы в этом редакторе выглядят одинаково. В ПП MEGA-4 добавлено несколько новых опций расчетов и расширен объем динамической памяти, что увеличивает скорость расчетов, объем анализируемых массивов данных и длину векторов-

последовательностей. Подробности функционирования ПП и руководство пользователя можно найти на сайте MEGA (http://www.megasoftware.net/).

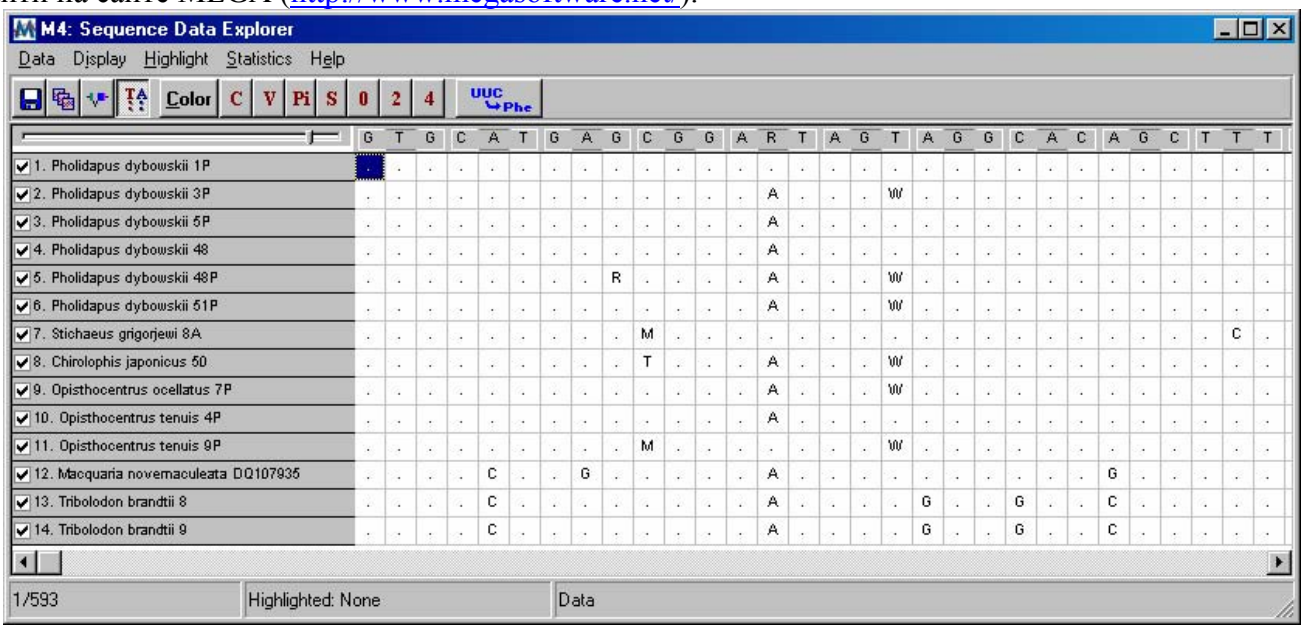


*Рис. 15.5.1. Вид рабочего файла MEGA-3 (MEGA-4), с открытым специальным «**редактором последовательностей**» (**Sequence Data Explorer**).*

Нуклеотиды, одинаковые для всех последовательностей, заменены точками. Для неоднозначно интерпретируемых нуклеотидов применяются дополнительные буквы, в примере – R,T,M,W.

*Fig. 15.5.1. A window of the MEGA-3 (MEGA-4) with open sequence editor **Sequence Data Explorer**. Identical nucleotides are replaced with dots. Ambiguous nucleotides are denoted with additional letters, in the example that are R,T,M,W.*

В открытом редакторе можно выполнить, если необходимо, дополнительную работу с набором последовательностей: выключить часть последовательностей из анализа, провести их трансляцию в аминокислотную последовательность, проанализировать состав нуклеотидов и др. Последняя функция реализуется через одну из опций меню редактора последовательностей «**статистика**» (**Statistics**) и далее выбором опции «**нуклеотидный состав**» (**Nucleotide Composition**). В результате расчетов появляется таблица, удобная для прямого использования в статье и требующая лишь небольшого редактирования.

Закрыв окно редактора последовательностей, получаем доступ к главному меню MEGA. Рассмотрим основные опции главного меню MEGA: «**файл**» (**File**), «**данные**» (**Data**), «**расстояния**» (**Distances**), «**филогения**» (**Phylogeny**), «**тип**» (**Pattern**), «**отбор**» (**Selection**), «**выравнивание**» (**Alignment**). Опция «**выравнивание**» была рассмотрена в общих чертах ранее (см. Раздел 15.3). В главном меню имеются еще две опции – «**окна**» и «**помощь**» (**Windows, Help**), функции которых очевидны и которые мы здесь не рассматриваем, поскольку они типичны для всех приложений ОС Windows.

Начинается главное меню с опции «**файл**» (**File**), которая позволяет провести серию операций с файлом (рис.15.5.2, открыто диалоговое окно с этой опцией). Поскольку рабочий файл был уже открыт ранее, то первая строка «**открыть данные**» (**Open Data**) не активна (серый цвет). Из открывшихся опций наиболее часто используемой является опция «**экспорт данных**» (**Export Data**). Она запускает новое диалоговое окно, в котором можно выбрать формат для экспортирования рабочего файла MEGA в другие форматы ПП (окно рабочего файла, будучи свернутым, вновь раскрывается одновременно с запуском последней опции): Nexus для PAUP 4.0,

Nexus для PAUP 3.0/McClade и PHYLIP. Используя диалоговое окно «**файл**», можно закрыть текущий рабочий файл, исполняя директиву «**закрыть данные**» (**Close Data**), затем можно открыть новый рабочий файл, «**открыть данные**» (**Open Data**), переоткрыть ранее использовавшиеся файлы (которые сохраняются с использованными координатами директорий), вызвать «**текстовый редактор**» (**Text Editor**), опции печати и, наконец, закрыть ПП MEGA («**выйти**», **Exit**).

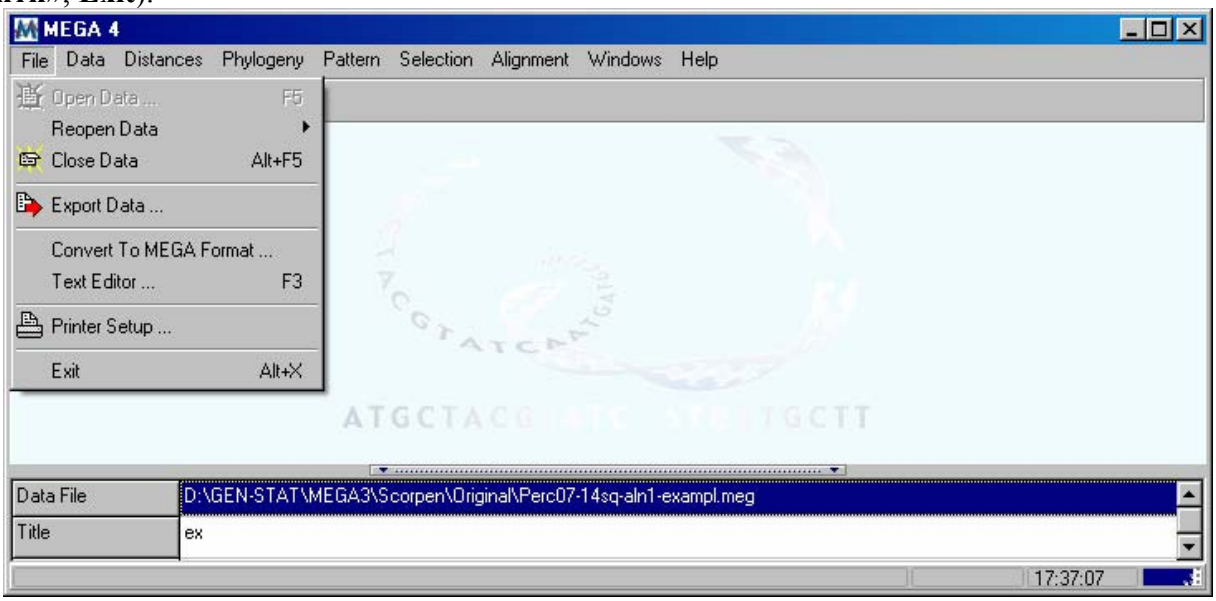


Рис.15.5.2. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.

*Открыто диалоговое окно опции «**файл**» (**File**) с рядом функций. Ниже показана локализация на диске рабочего файла (Data File) и заголовок задания (Title).*

Fig. 15.5.2. A window of the MEGA-3 (MEGA-4) software with the main menu options.

*A dialog window for file (**File**) option with several functions is shown. In two bottom lines there are records for working file location (DataFile) and its name (Title).*

Следующая опция в главном меню – это «**данные**» (**Data**) (рис.15.5.3). Можно вновь активировать рабочий файл запустив «**исследователь данных**» (**Data Explorer**), с включением из подменю «**редактора последовательностей**» (**Sequence Data Explorer),** который очень коротко был охарактеризован ранее, но имеет и множество других полезных функций. В разных опциях редактора можно: показать (выделить) определенные последовательности, выделить нужные группы видов или последовательностей, показать свойства организации генов (или доменов в полипептидах), выделить нуклеотиды цветом, показать консервативные сайты, вариабельные сайты, единичные сайты, провести трансляцию, вызвать опцию «**статистика**», о чем говорилось ранее, и др. В опции «**данные**» часть этих функций для удобства повторена: «**установки/выделить таксоны и группы**» (**Setup/Select Taxa&Groups**). «**установки/выделить гены и домены**» (**Setup/Select Genes&Domains**). Кроме этого, можно «**выбрать предпочтения**» (**Select Preferences**), «**выбрать таблицу генетического кода**» (**Select genetic Code Table**).

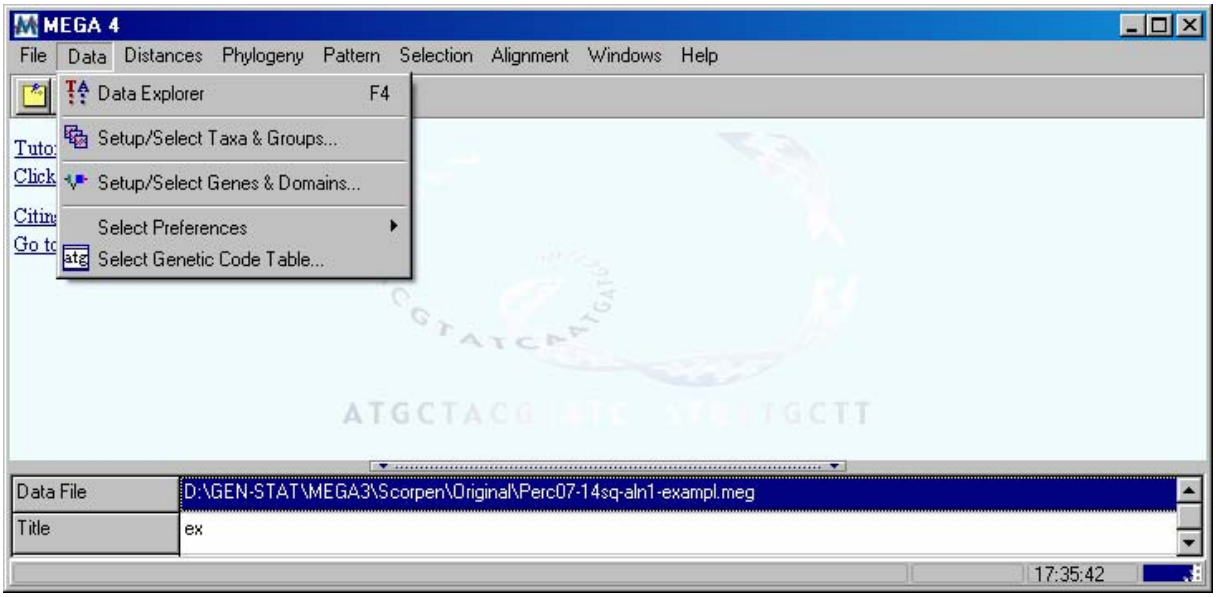


Рис. 15.5.3. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.
*Открыто диалоговое окно опции «**данные**» (**Data**) с рядом функций, которые пояснены в тексте.*
Fig. 15.5.3. A window of the MEGA-3 (MEGA-4) software with the main menu options.
*A dialog window for data (**Data**) option with several functions that explained in the text is shown.*

Следующая опция в главном меню – это «**расстояния**» (**Distances**) (рис.15.5.4). Выбрав в диалоговом окне первую строку с опцией «**выбрать модель**» (**Chose Model**), запускаем следующее диалоговое окно, которое позволяет использовать 9 различных расстояний (моделей). О том какую из моделей выбрать и как сказано ранее (см. Раздел 15.4). На предварительных этапах анализа уместно использовать просто *p*-расстояние или K2P (двухпараметрическая модель Кимура; см. Раздел 13.1). В открытом окне можно выбрать также две другие опции, важные при расчетах. Первая – это скорость замен нуклеотидов в филетических линиях, «**тип между линиями**» (**Pattern among Lineages)**. ПП MEGA допускет 2 возможных варианта типов замен – «**одинаковые**» (Same (Homogeneous)) или «**различные**» (Different (Heterogeneous)). Вторая – это вариация темпа замен между сайтами, «**скорость между сайтами**» (**Rates Among Sites**). Закрытие этого окна позволяет продолжить работу с другими опциями из меню «**расстояния**». Следующие две опции дают возможность рассчитать матрицу попарных расстояний, «**вычислить попарные**» (**Compute Pairwise**) и определить среднее расстояние для набора анализируемых последовательностей, «**вычислить суммарную среднюю**» (**Compute Overall Mean**). Три следующие опции из меню «расстояния» доступны, если среди набора последовательностей выделены какие либо группы. В этом случае возвожен рассчет внутри- и межгрупповых расстояний, рассчитать сеть между групповыми средними и наконец – определить разнообразие последовательностей.

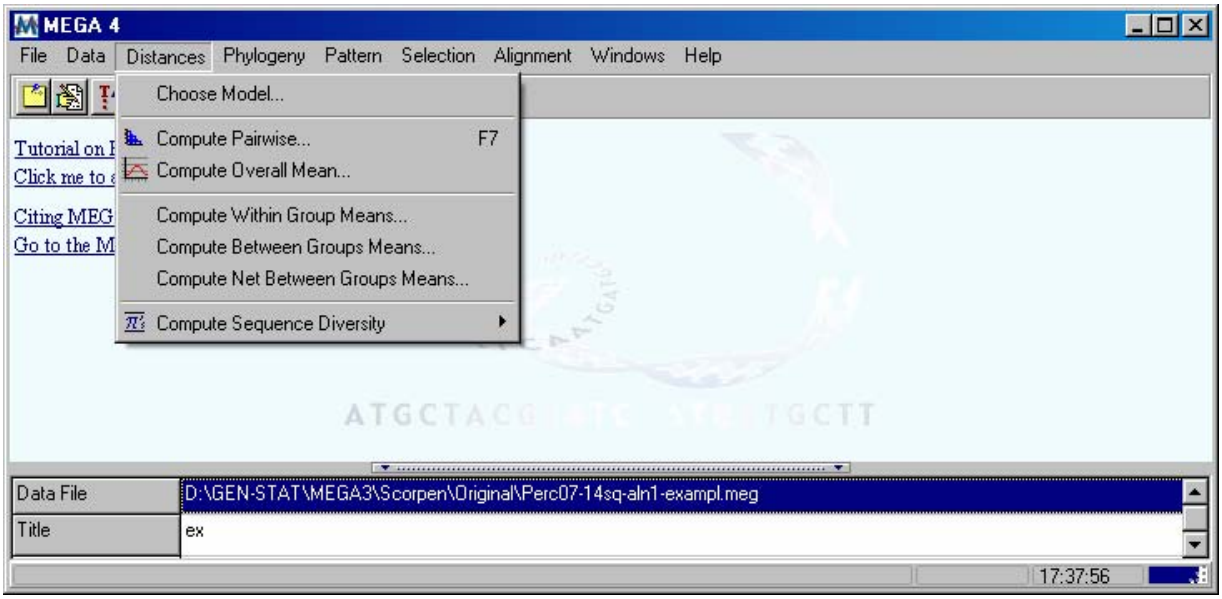


Рис. 15.5.4. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.
*Открыто диалоговое окно опции «**расстояния**» (**Distances**) с функциями, которые пояснены в тексте.*
Fig. 15.5.4. A window of the MEGA-3 (MEGA-4) software with the main menu options.
*A dialog window for distances (**Distances**) option with several functions that explained in the text is shown.*

Следующая опция в главном меню – это «**филогения**» (**Phylogeny**) (рис.15.5.5). Выбрав в диалоговом окне первую строку с опцией «**построить филогению**» (**Construct Phylogeny**), или вторую строку «**бутстреп тест филогении**» (**Bootstrap Test of Phylogeny**), получаем доступ к четырем программам для расчета молекулярно филогенетических деревьев. Сверху вниз в подменю это программы: (1) «**ближайшего соседства**» (**Neighbor Joining**; NJ), (2) «**минимальной эволюции**» (**Minimal Evolution**), (3) «**максимальной парсимонии**» (**Maximum Parsimony**) и (4) **НПГМ** (**UPGMA**). При конструировании небольших деревьев предпочтение следует отдать второй опции «**бутстреп тест филогении**». При этом при построении молекулярно-филогенетического дерева одновременно проводится оценка его устойчивости (надежности), с использованием бутстреп теста. Если дерево большое, включающее более 50 последовательностей с 1000-1500 пн или длиннее, то можно для прикидочного анализа взять первую опцию и запустить программы, использующие относительно простые алгоритмы – NJ и НПГ. На Практическом занятии 15 будет выполнен расчет дерева для набора последовательностей, показанных частично на рис. 15.5.1, с использованием программы «**ближайшего соседства**» (рис. 15.6.1-15.6.2). Сейчас же продолжим рассмотрение следующей опции главного меню.

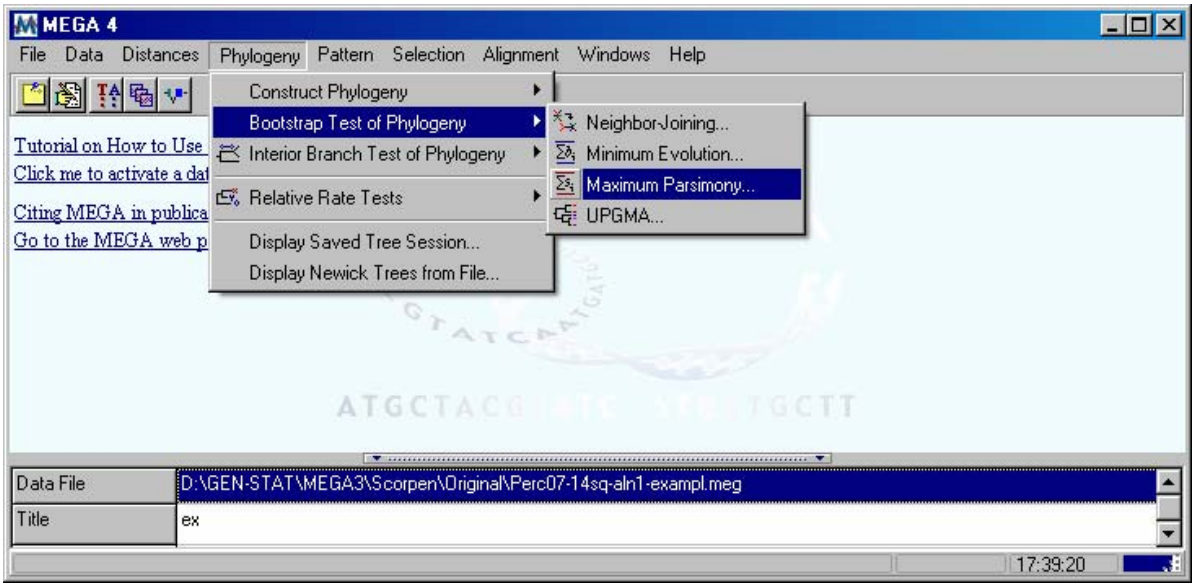


Рис. 15.5.5. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.

*Открыто диалоговое окно опции «**филогения**» (**Phylogeny**) с раскрытой опцией «**бутстреп тест филогении**» (**Bootstrap Test of Phylogeny**) и подменю четырех основных программ построения молекулярно-филогенетических деревьев: (1) «**ближайшего соседства**» (**Neighbor Joining; NJ**), (2) «**минимальной эволюции**» (**Minimal Evolution**), (3) «**максимальной парсимонии**» (**Maximum Parsimony**) и (4) **НПГ** (**UPGMA**), а также с другими функциями, которые пояснены в тексте.*

Fig. 15.5.5. A window of the MEGA-3 (MEGA-4) software with the main menu options.

*A dialog window for phylogeny (**Phylogeny**) option with a sub-option **Bootstrap Test of Phylogeny** is shown. Also opened a sub-menu for four main programs of tree building (1) **Neighbor Joining; NJ**), (2) **Minimal Evolution**, (3) **Maximum Parsimony,** and (4) **UPGM;** several other functions that explained in the text are available in this menu as well.*

Опция «**тип**» (**Pattern**) – это опция, вызывающая несколько важных функций MEGA. Эти функции позволяют получить оценки для параметров моделей, используемых при построении деревьев, или являются важными для иных суждений о молекулярной эволюции (рис.15.5.6). Эту опцию стоит запустить до анализа филогении, то есть, до использования предыдущей опции главного меню «**филогения**». В частности, при анализе дивергенции последовательностей важно рассмотреть, есть ли различия в соотношении между разными типами замен или типами оснований. Нередки случаи, когда доля пиримидинов выше, чем доля пуринов или/и их пропорции меняются в филетических линиях (Kartavtsev et al., 2007a,b). Последняя опция показанного меню, как раз позволяет оценить это смещение: «**вычислить смещение транзиций/трансверсий**» (**Compute Transition/Transversion Bias**).

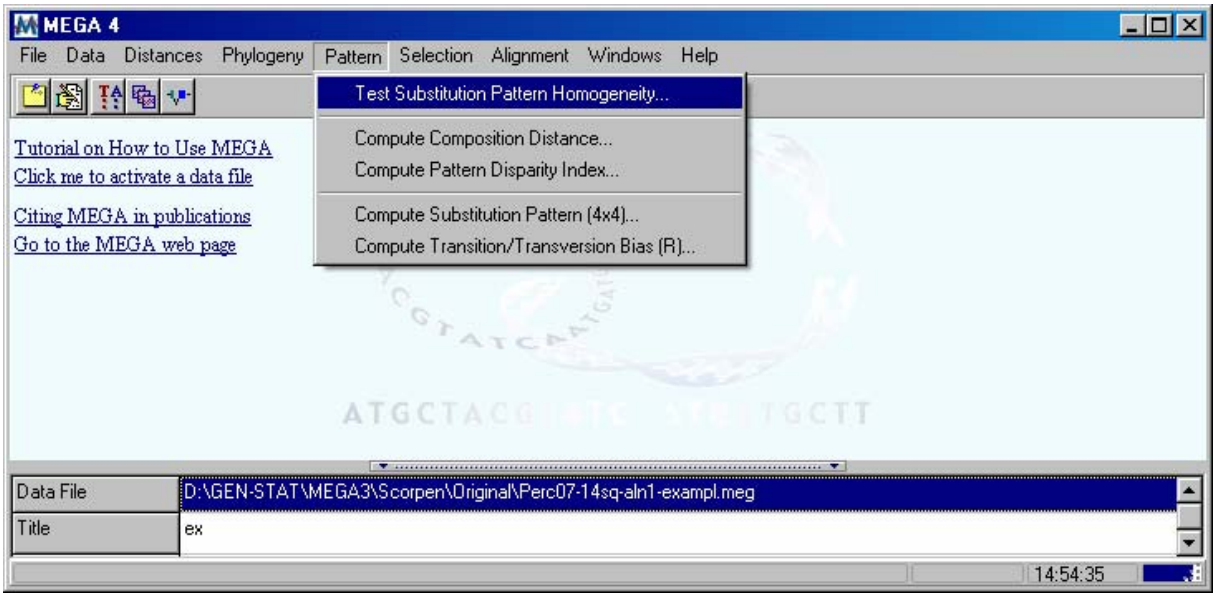


Рис. 15.5.6. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.

*Открыто диалоговое окно опции «**mun**» (**Pattern**) с раскрытым подменю и пятью его опциями.*

Fig. 15.5.6. A window of the MEGA-3 (MEGA-4) software with the main menu options.

*A dialog window for **Pattern** option with a sub-menu and its 5 options is shown.*

Последняя опция в главном меню, которую мы разберем, – это «**отбор**» (**Selection**) (рис.15.5.7). Программа позволяет вызвать три опции, с тремя различными тестами отбора. В меню выделен «**тест Таджима на нейтральность**» (**Tajima's Test of Neutrality**) (рис.15.5.7). Оценка эффектов, связанных с возможным действием естественного отбора, занимает особое место в эволюционной биологии со времен Дарвина. Эти работы получили особенный импульс после работ Кинга и Джукса (King, Jukes, 1968) и Кимура (Kimura, 1969; Кимура, 1985) о большой роли дрейфа в эволюции или наличии нейтральной молекулярной эволюции. Однако, выполняя эти тесты, не стоит забывать об огромном числе неучтенных факторов, которые могут влиять на значимость оценок. Авторы MEGA это вполне понимают и, соответственно, многие функции выдают сообщение о том, что за результаты расчетов авторы сами несут ответственность.

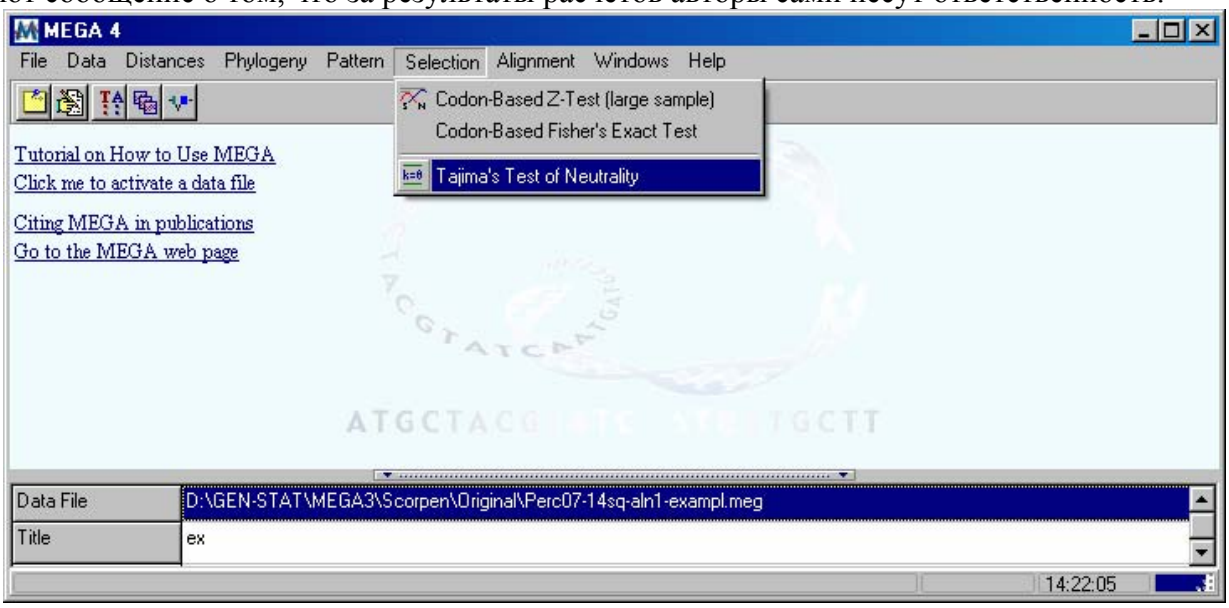


Рис. 15.5.7. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.

*Открыто диалоговое окно опции «**отбор**» (**Selection**) с раскрытым подменю и тремя его опциями.*
Fig. 15.5.6. A window of the MEGA-3 (MEGA-4) software with the main menu options.
*A dialog window for **Selection** option with a sub-menu opened and its 3 options is shown.*

Кроме рассмотренного выше ПП MEGA, широко используемыми ПП являются PAUP 4.0, MrBayes, PHYLIP и многие другие.

PAUP 4.0 (Swofford, 2002) является коммерческим ПП и в этой версии распространяется под ОС «Макинтош» (Macintosh). Популярное изложение ПП PAUP 4.0, как отмечалось, представил Хэлл. В версии для ОС Windows (PAUP 4.0 10b) он также доступен, но его интерфейс несовершенен и от пользователей требуется большой труд, чтобы освоить многочисленные операторы управления программой. Тем не менее, PAUP 4.0 стоит этих усилий. Как упоминалось, без этого ПП нельзя запустить программу MODELTEST. Кроме того, среди предоставляемых программ пакета PAUP есть практически все, необходимые для детального анализа молекулярных филогений. Например, есть программа для конструирования дерева максимального правдоподобия (Maximum Likelihood, ML), NJ-дерева и дерева максимальной парсимонии. Имеются различные опции для оценки состоятельности деревьев, то есть их надежности, включая бутстреп поддержку. Построение ML-дерева требует очень больших вычислительных ресурсов. К примеру, один из расчетов дерева, состоящего из 67 последовательностей по гену *Cyt-b* (Kartavtsev et al., 2007a), длился у автора 3 недели на хорошем ПК. Учитывая данное обстоятельство, а также требования ряда ведущих журналов представлять в статьях ML-деревья, создана версия PAUP под ОС Linux/Unix, которая допускает работу на многопроцессорных компьютерах, позволяющих параллельную работу нескольких пользователей, и являющихся более быстрыми.

MrBayes (Hulsenbeck, Rondquist, 2001; Ronquist, Huelsenbeck, 2003) – это относительно небольшой ПП. Однако вычислительный алгоритм, используемый им, является исключительно эффективным. Упоминавшийся выше набор из 67 последовательностей, был обработан этим ПП уже в течение двух дней. Построенные деревья также как ML-деревья, имеют вероятностную основу и высоко котируются в молекулярно-филогенетических журналах. Кроме самих деревьев, ПП MrBayes предоставляет и ряд других вычислительных возможностей для моделирования эволюции, в том числе и по морфологическим признакам. Сами деревья ПП MrBayes не строит, но специальные выходные файлы могут быть использованы для этих целей. Например, ПП TreeView (Page 1996) позволяет легко построить дерево по такому файлу ПП MrBayes, в том числе вывести и консенсусное дерево.

ПП PHYLIP (Felsenstein, 1995) – это очень хороший инструмент для филогенетического анализа. Его огромным достоинством, как и в случае с MEGA, является то, что программы базируются на теоретических выкладках и рекомендациях, изложенных в известных книгах (Nei, Kumar, 2000; Felsenstein, 2004). PHYLIP предоставляет возможность рассчитать все основные типы деревьев. Однако интерфейс ПП рассчитан под ОС DOS и в этом, конечно, менее удобен для использования.

15.6. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 15

1. Регистрация последовательностей в банке GenBank с помощью утилиты Bankit (http://www.ncbi.nlm.nih.gov/BankIt/help.html#getting). Откройте окно Bankit, введите любую последовательность нуклеотидов и другую требуемую информацию.
2. Выравнивание последовательностей посредством редактора CLUSTAL W.
1) Откройте ПП MEGA.

2) Откройте «**редактор выравнивания**» (**Alignment Explorer**) и проведите выравнивание любых трех последовательностей, извлекши их предварительно их генного банка GenBank (http://www.ncbi.nlm.nih.gov).

3. Расчеты и построение молекулярно-филогенетических деревьев с использованием ПП MEGA.

1) Откройте ПП MEGA. Далее запустив опцию «**файл**» (**File**) и открыв директорию «примеры» (Examples), используйте файл с именем «Drosophila_Adh.meg».

2) В соответствии с рекомендациями в Разделе 15.5 запустите опцию «**бутстреп тест филогении**» (**Bootstrap Test of Phylogeny**) и подменю программы построения молекулярно-филогенетического дерева «**ближайшего соседства**» (**Neighbor Joining**). Для этого в открывшемся диалоговом окне «**анализ предпочтений**» (**Analysis Preferences**) выберете из подменю: **Phylogeny Test of Evolution** → опции **Bootstrap**, **Replications** = 1000 и **Random Seed** = 20044 (или другое случайное число). Далее в опции «**модель**» (**Model**) выберете модель K2P, как указано в открытых диалоговых окнах (рис. 15.6.1). Запустите опцию «**вычислить**» (**Compute**). Построенное дерево в отрывшемся окне «**исследователя деревьев**» (**TreeExplorer**) должно иметь следующий вид (рис. 15.6.2).

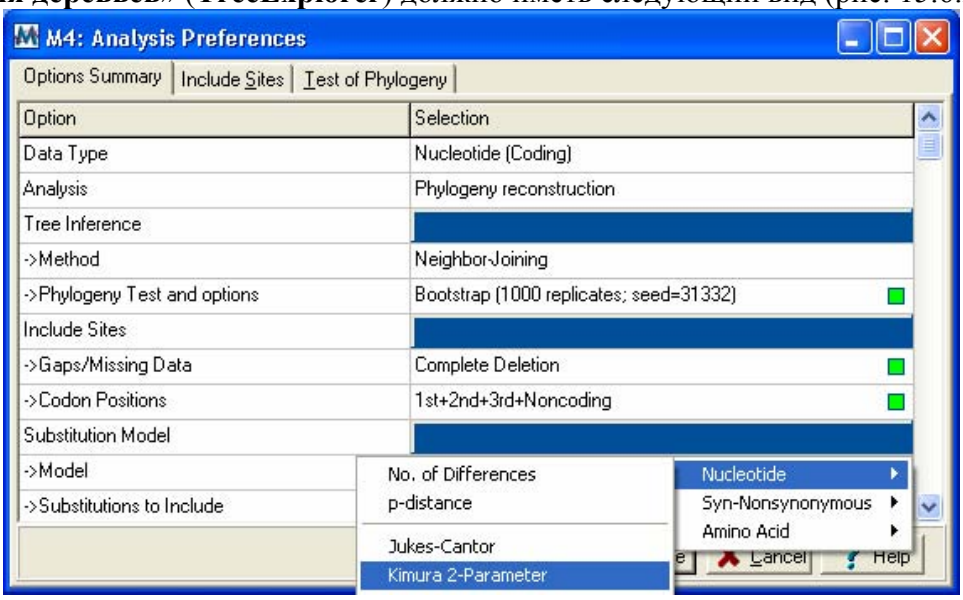


Рис. 15.6.1. Вид открытого ПП MEGA-3 (MEGA-4) с опциями главного меню.

*Открыто диалоговое окно опции «**филогения**» (**Phylogeny**) с раскрытой опцией «**бутстреп тест филогении**»: **Bootstrap Test of Phylogeny → Neighbor Joining → Phylogeny Test of Evolution.***

Fig. 15.6.1. A window of the MEGA-3 (MEGA-4) software with the main menu options.

*A dialog window for phylogeny (**Phylogeny**) option with a sub-option **Bootstrap Test of Phylogeny → Neighbor Joining → Phylogeny Test of Evolution.***

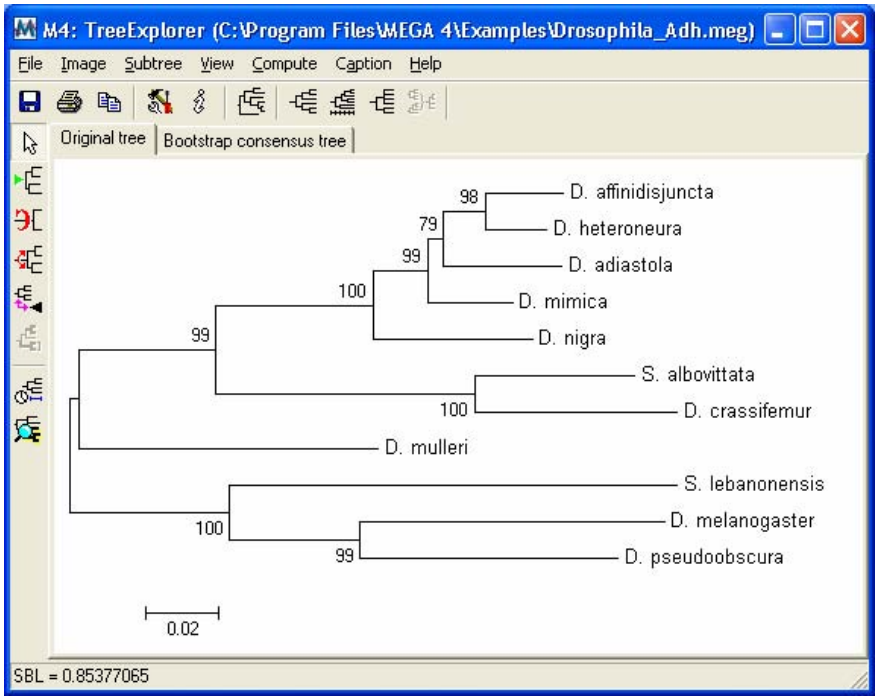


*Рис. 15.6.2. Вид открытого в ПП MEGA-3 (MEGA-4) «**исследователя деревьев**» (**TreeExplorer**).*

*Открыт файл с деревом, рассчитанным посредством программы NJ. На концах ветвей даны наименования изученных видов по нуклеотидной последовательности гена **Mdh** из примера, описанного в тексте и распространяемого вместе с ПП MEGA (Examples). В нижнем левом углу дан масштаб для длин ветвей.*

*Fig. 15.6.2. A view of the MEGA-3 (MEGA-4) software in option of the **TreeExplorer**.*

A file with the tree built by NJ opened. On the tips of branches names for species are given that were analyzed at MDH gene sequences. The example is from MEGA distributive folder (Examples). In left bottom corner a line represents scale for branch length.

ГЛАВА 16. ЭВОЛЮЦИЯ ГЕНОМА

ГЛАВНЫЕ ВОПРОСЫ:

16.1. Происхождение и изменение генома.
16.2. Эволюционные изменения размера генома.
16.3. Формирование новых генов.
16.4. Повторяющиеся ДНК и мультигенные семейства.
16.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 16.

РЕЗЮМЕ

1. Минимальное число генов, которое необходимо для существования самовоспроизводящегося организма у простейших форм *M. genitalium* и *M. pneumonia*, составляет 250–350 генов.
2. Тандемные дупликации и геномные дупликации (полиплоидия) являются главными источниками возникновения новых генов.
3. Геном эукариот содержит различные типы повторяющихся ДНК. Наиболее известными из них являются: (1) сателлитные ДНК, (2) варьирующие в числе тандемные повторы, ВЧТП (VNTRs), (3) микросаттелиты, (4) короткие чередующиеся элементы, КЧЭ (SINEs), (5) длинные чередующиеся последовательности, ДЧЭ (LINEs), (6) мультигенные семейства (рРНК, тРНК, глобин, иммуноглобулин и др.).
4. Геном с момента возникновения в ходе эволюции развивался, в основном, по линии усложнения, что в итоге привело к иерархическим, каскадным, гомеостазисным и сете-подобным принципам регуляции на клеточном, индивидуальном и популяционном уровнях у эукариот. Эволюция многих генов и их частей порождается случайными или стохастическими процессами.

Chapter 16. GENOMIC EVOLUTION

SUMMARY

1. The minimal number of genes necessary for organism existance as an independent, self-reproducing system was established on comparisons of *M genitalium* and *M. pneumonia* genomes, and on mutation experiments in *M. genitalium.* Now it is estimated that living cellular organisms require a minimum of 250-350 genes. **2.** Tandem duplications and genome multiplications (polyploidy) are main sources of a new gene origin.
3. Genome of eukaryotes contains various types of repetitive DNA. The most important are (1) satellite DNA, (2) VNTRs, minisatellites, (3) microsatellites, (4) SINEs, (5) LINEs, (6) gene families (rRNA, tRNA, globin, immunoglobulin, etc.).
4. Genome since its origin and during a long history of life evolved from quite simple ring linear molecules to highly organized chromosomes and several orders of hierarchical and network-like relationships which permitted it to be realized in a complexity of all living higher forms.

16.1. ПРОИСХОЖДЕНИЕ И ИЗМЕНЕНИЕ ГЕНОМА

Первые простейшие клеточные организмы, как свидетельствуют ископаемые остатки, возникли на Земле около 3.5 млд. лет назад (Лекция 1). Вероятно, что геном этих организмов уже состоял из двух нитей ДНК. Задолго до этого времени химическая эволюция должна была развить более простые, самовоспроизводящиеся создания, наподобие вирусов или вирусоподобных форм. Возможно, что вирусоподобные и протобактериальные формы были заселены на Землю в

результате падения на неё кометы или метеорита, во времена, когда атмосфера еще не была существенным барьером для проникновения инопланетных объектов. Поэтому, возникновение самой жизни могло происходить в каком-то другом месте вселенной. Первые ископаемые находки эукариотоподобных организмов, похожих на одноклеточные сине-зеленые водоросли, датируются возрастом порядка 1.5 млд лет назад. В течение своей жизни и прокариоты, и эукариоты постоянно сталкиваются с меняющейся средой обитания и должны уметь изменять себя, чтобы выжить. Проще говоря, изменения организмов в ответ на меняющуюся среду обитания и есть эволюция, когда процесс происходит во времени. Среда обитания сложна и представляет собой не только абиотические компоненты (вода, минералы и т.д.), но и биотические компоненты (различные организмы и продукты их жизнедеятельности). **Геном**, как *взаимосвязанная совокупность всех генов организма у живых существ*, также должен меняться динамически. Поэтому, причины эволюции многогранны. Первичные причины эволюционных изменений могут исходить как из неорганических составляющих среды, так и из изменения самих организмов и их генома. Как в прошлом, так и в настоящее время, изменения генома включают: мутации, рекомбинации, транспозиции, делеции, дупликации и другие. Изучая и сравнивая геномы филогенетически различных организмов, которые существуют в данное время, можно ретроспективно представить себе, какие механизмы и как определяют изменение генома, или понять его эволюцию.

Минимальный геном живой клетки

Сколько генов необходимо для существования живого клеточного организма? Мы не будем говорить о вирусах, жизненный цикл которых сейчас, в большинстве случаев, связан с бактериями или другими клеточными организмами, и они вообще могут рассматриваться, как упростившиеся, примитивные паразиты. Точно ответить на поставленный вопрос пока невозможно, поскольку не известны функции каждого гена в геноме, даже если известна его полная первичная последовательность. Все же, основываясь на генах с известной функцией, по аналогии с человеком или другими организмами (рис. 16.1.1), можно предположить, каково должно быть минимальное число генов, чтобы обеспечить основные функции клетки одноклеточного организма. Для этого мы будем использовать информацию о первичных последовательностях нуклеотидов в генах для двух самых маленьких бактериальных геномов: *Mycoplasma genitalium* и *M. pneumoniae.* Эти организмы являются одними из простейших прокариот (см. также Лекцию 3). Геном *M. genitalium* равен 0.6 мпн. (мега пар нуклеотидов), тогда как у *M. pneumoniae* размер генома равен 0.8 мпн. Оба вида принадлежат к группе бактерий, которые не имеют клеточной стенки, и, проникая в другие организмы, часто вызывают инфекции, в частности, - у растений, насекомых и человека. У животных заражение этими бактериями вызывает генитальные и респираторные заболевания.

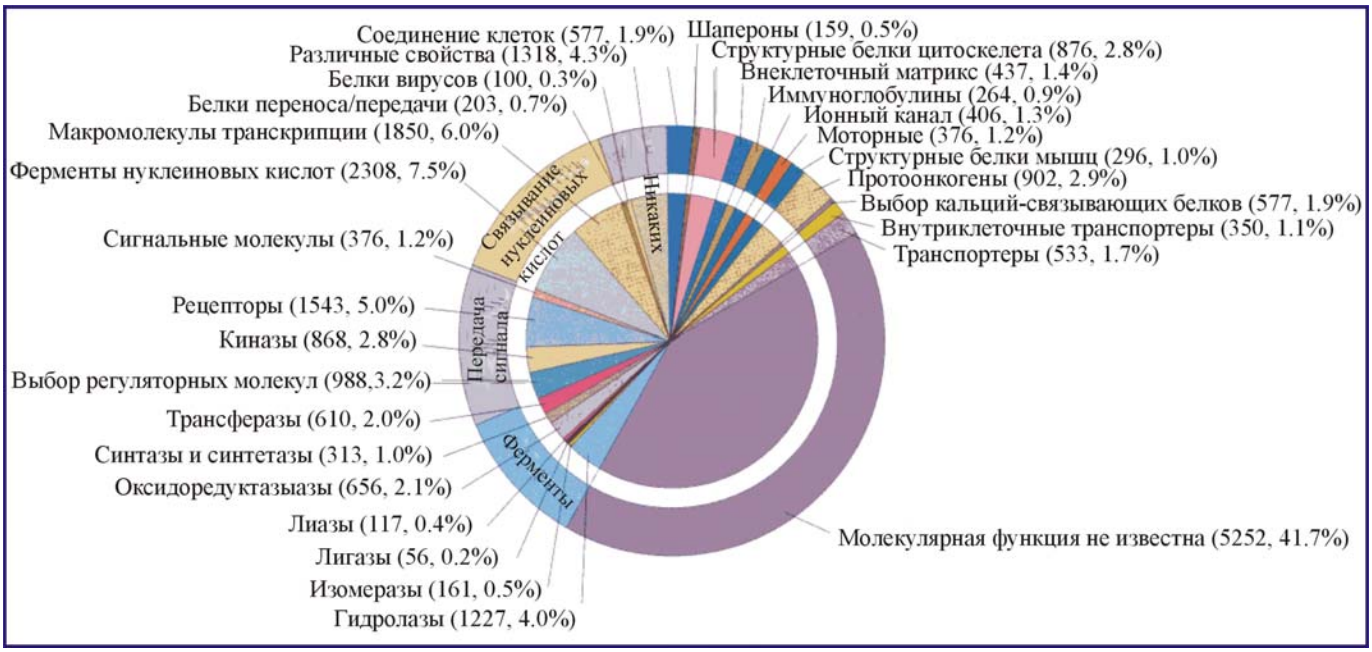


Рис. 16.1.1. Предварительный список выясненных функций для 26588 генов генома человека (По Venter et al. 2001).

Fig. 16.1.1. A preliminary list of assigned functions for 26588 genes in the human genome (From Venter et al. 2001; adopted from Klug, Cummings, 2002).

Оценка функций генов, включенных в список расшифрованных генов человека, основывается на сходстве белков известной функции у разных организмов (рис. 16.1.1). Наиболее обычны в геноме гены, которые вовлечены в метаболизм нуклеиновых кислот (7.5% всех идентифицированных генов), рецепторы (5%), протеинкиназы (2.8%) и структурные белки цитоскелета (2.8%). В целом, из 12809 предсказанных на основе секвенс анализа белков, 5252 (41%) не имеют известной функции, показывая, насколько много еще впереди работы для понимания функций лишь отдельных генов и только генома человека. В геноме *M. genitalium* имеются 467 открытых рамок считывания (ОРС), тогда как *M. pneumoniae* имеет 677 ОРС. Все 467 генов *M. genitalium* обнаружены среди 677 генов *M. pneumoniae.* Для сравнения, другие бактерии имеют значительно больший размер генома: *E. coli* – 4.6 мпн с 4288 ОРС, а *Haemophilus influenzae* – 1.8 мпн с 1727 ОРС. Некоторые гены, которые функционируют у упомянутых бактерий и их число приводятся ниже (Табл. 16.1.1). Очевидно, что среди основополагающих генов должны быть те, которые кодируют продукты, требуемые для репликации и репарации ДНК, транскрипции и трансляции, поддержания общих клеточных функций (метаболизм, катаболизм, иммунитет), и, наконец, транспорта белков.

Таблица 16.1.1. Функциональные классы генов у трех видов бактерий *(По Klug, Cummings, 2002)*

Table 15.1.1. Functional classes of genes in three bacterial species *(From Klug, Cummings, 2002)*

Функциональный класс	***E. coli***	***H. influenzae***	***M. genitalum***
Гены кодирующие белки	4288	1727	470
Репликация ДНК, репарация	115	87	32
Транскрипция	55	27	12
Трансляция	182	141	101
Регуляторные белки	178	64	7
Биосинтез аминокислот	131	68	1
Биосинтез нуклеиновых кислот	58	53	19
Метаболизм липидов	48	25	6
Энергетический метаболизм	243	112	31
Распознающие и транспортные белки	427	123	34

Давайте сравним гены, требуемые для определенной функции у разных видов бактерий. Так, *Escherichia coli* имеет для метаболизма 131 аминокислоту, *H. influenzae* – 68, а *M. genitalium* – только 1 (Табл. 16.1.1). Хотя *Haemophilus* имеет почти в 4 раза больше генов, его физиологические возможности остаются очень сходными с *Mycoplasma.* Очевидно, что много больше генов участвуют в реальности в кодировании ферментов в каждой функциональной группе. Продвинутость *Haemophilus* включает его исключительную пластичность в биосинтетической способности. Эта более сложная бактерия имеет 68 генов, вовлеченных в биосинтез аминокислот, тогда как *Mycoplasma*, как отмечено выше, имеет только один такой ген. В отсутствие такой пластичности *M. genitalium* вынуждена зависеть от многочисленных метаболических продуктов своего хозяина. Проведенные сравнения двух видов и мутационные эксперименты на *M. genitalium* позволяют сделать заключение о минимальном числе генов, необходимых для функционирования самовоспроизводящегося прокариотического организма. Это число составляет 250-350 генов.

Возникновение и эволюция эукариотического генома

Эукариот традиционно отличают от прокариот по нескольким принципиальным признакам, таким как отделенное мембраной ядро, система мембран цитоплазмы (цитоплазматический ретикулум) и цитоскелет. Новые исследования, и в частности Геномный проект, дали дополнительную информацию о структуре генома эукариот в сравнении с прокариотами. Это, соответственно, продвинуло понимание эволюции генома. К новым данным относятся: аминокислотный анализ белков у эукариот, анализ первичных последовательностей генов и расшифровка ряда неизвестных метаболических путей. Благодаря новым данным выяснилось, что эукариотический геном – это сложная конструкция и является мозаикой, составленной из нескольких элементов представителей двух групп прокариот Археобактерий и Настоящих бактерий (Archaea и Eubacteria). Эукариотический ядерный геном, как правило, не имеет оперонов, а его гены включают интроны. Обе эти особенности также отличают Археобактерий. В свою очередь, митохондриальный геном эукариот очень похож на геном альфа-протеобактерий (alpha-proteobacteria). Чтобы объяснить все эти данные, было выдвинуто предположение, что эукариоты возникли, как последовательность симбиотических связей между древним представителем анаэробной Археобактерии хозяина и аэробной альфа-протеобактерии, подобной риккеции (*Rickettsia)*, которые эволюировали далее в митохондрии. Все эукариоты имеют подобный тип клеточной организации, что, объясняет их монофилетическое происхождение и, кроме того, предполагает, что это успешное событие произошло лишь однажды в истории жизни на Земле. Однако, в ходе эволюции генома возникали и другие, хотя и менее важные симбиотические ассоциации. На самой ранней стадии произошло включение в клетки секс-фактора, который эволюционировал в полифакторно определяемую половую дифференциацию. Предполагается, что геном высших растений возник, как результат введения в растительную прото-клетку хлоропласт-подобного прокариотического организма.

16.2. ЭВОЛЮЦИОННЫЕ ИЗМЕНЕНИЯ РАЗМЕРА ГЕНОМА

Геномные дупликации

Как показывают имеющиеся данные, наиболее частыми в эволюции генома прокариот являются мутации, ассимиляции малых мобильных элементов, наподобие вирусов или плазмид, и ***тандемные дупликации*** *– участки, повторяющиеся один за другим*. На последних изменениях и дупликации целых хромосом мы и сосредоточим свое внимание в данном разделе. Увеличение всего набора хромосом или всего генома – это особенность эукариот; не известно ни одной бактерии, имеющей более чем одну настоящую хромосому. Долгое время большая часть внимания генетиков была сфокусирована на дупликации отдельных генов. Колоссальное значение полиплоидии в эволюции растений обычно не подвергается сомнению. С животными дело обстоит сложнее. Оно (1973) одним из первых обосновал, что геномные дупликации являются также важным эволюционным механизмом и для животных. Анализ данных по первичным

последовательностям нуклеотидов подтвердил мысль о том, что различия в числе генов, которые различают прокариот и эукариот, возникли как следствие наращивания исходного прото-генома. Сейчас принимается, что большинство событий этого наращивания размера генома эукариот возникают за счет дупликаций генома. В частности, появление позвоночных совпадает с массивным наращиванием генома, хотя дупликации возникали в эволюции многократно (Оно, 1973; Ohno, 1984; Ferris, Whitt, 1979).

Анализ генома дрожжей показывает древность наращивания генома за счет дупликаций. Вольф и Шильдс (Wolfe, Shields, 1997) исследовали первичную последовательность каждого гена у дрожжей и сопоставили весь геном. Они проанализировали 55 дуплицированных регионов, содержащих 376 генов, которые покрывают 50% длины генома. Из 55 регионов, 50 находятся в одинаковом относительном положении на хромосомах (рис. 16.2.1). Например, хромосомы XI и XIII содержат дуплицированный блок генов (блок 43), в котором порядок генов и ориентация по отношению к центромере сохранены неизменными. Другие хромосомы содержат внутренние дупликации, такие как блок 53, который расположен по разные стороны от центромеры на хромосоме XII. Филогенетический анализ показывает, что это событие геномной дупликации произошло около 100 млн. лет назад. Анализ генома человека также обнаруживает свидетельства древней масштабной дупликации, сопровождаемой перестановками и утратами генов в некоторых дуплицированных регионах (Venter et al., 2001). Эти дупликации в геноме человека включали небольшие сегменты, затрагивающие только несколько генов, а также большие участки, которые распространяются почти на всю хромосому. Большие дупликации, очевидно, датируются моментом возникновения позвоночных, то есть, около 500 млн. лет назад. В целом в геноме человека имеется 1077 блоков дуплицированных регионов, содержащих свыше 10000 генов; одна такая дупликация имеется в хромосомах 18 и 20 (Klug, Cummings, 2002).

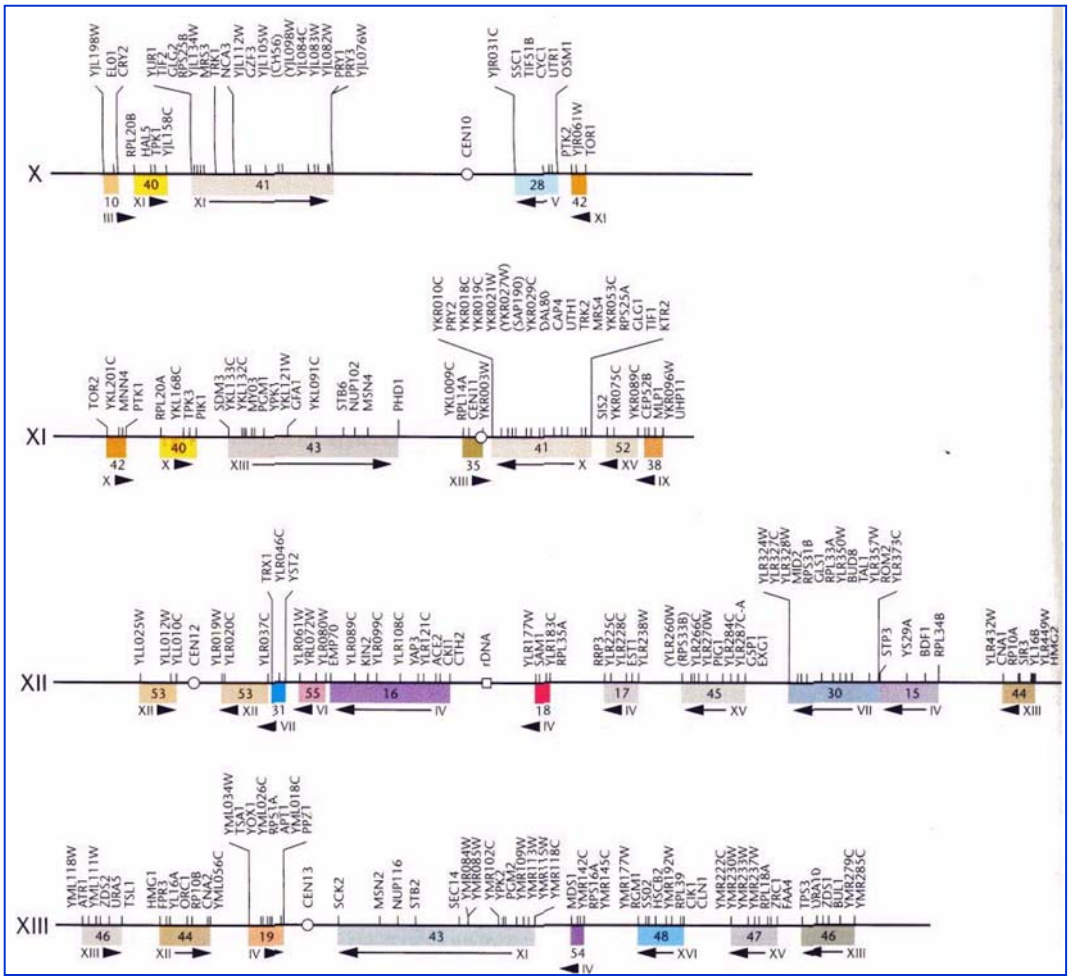


Рис. 16.2.1. Расположение дуплицированных регионов у дрожжей на хромосомах X, XI, XII и XIII. Дуплицированные блоки выделены, а их расположение показано римскими цифрами ниже самих блоков. Ориентация дуплицированных блоков в геноме консервативна по отношению к центромере. Исключение составляют пять блоков (По Wolfe, Shields, 1997; заимствовано из Klug, Cummings, 2002).

Fig. 15.2.1. Location of duplicated regions on yeast chromosomes X, XI, XII, and XIII. The duplicated blocks are numbered, and roman numerals below the blocks give the location of the duplicate block. Arrows show the relative orientation of the blocks. Genomewide orientation relative to the centromere is conserved except for five blocks (From Wolfe, Shields, 1997; adopted from Klug, Cummings, 2002).

Более высокое содержание ДНК у высших организмов, по-видимому, возникло за счет генных дупликаций. На самом деле имеется два типа дупликаций генов. Один – это дупликация всей хромосомы однократно, а второй – многократная дупликация небольшого участка хромосомы, так называемые тандемные дупликации. Первый тип хромосомной дупликации – происходит во время дупликации всего генома при полиплоидии. Как показано в Табл. 16.2.1, ДНК млекопитающих почти в 1000 раз больше, чем ДНК *E. coli*. Однако далеко не всегда увеличение содержания ДНК ведет к более совершенному строению организма.

В эволюции растений дупликации всего генома через полиплоидию происходили во многих группах и играли важную роль при видообразовании. В эволюции животных роль полиплоидии также важна. И цитологические, и биохимические данные показали, что дупликации генома происходили несколько раз в ходе эволюции рыб и амфибий (Оно, 1973; Ferris, Whitt, 1979). Полиплоидия была довольно частым явлением у животных до той поры, пока не появились дифференцированные половые хромосомы, и её важность сохраняется и по настоящее время в некоторых группах животных (Боркин, Даревский, 1980; Алтухов, 1989). Когда половая дифференциация за счет хромосомной детерминации пола становится совершенной, как у млекопитающих, птиц и, в меньшей степени, у рептилий, то полиплоидия приобретает более ограниченное значение. У большинства рыб и амфибий половые хромосомы развиты слабо. Поэтому, например, у рыб тетраплоидные группы хорошо известны (Оно, 1973; Фролов, 2000).

Таблица 16.2.1. Содержание ДНК у различных организмов *(По Nei, 1987)*
Table 15.2.1. DNA content in various organisms *(From Nei, 1987)*

Организм	***Пар нуклеотидов на геном***	***Организм***	***Пар нуклеотидов на геном***
Млекопитающие	3.2×10^9	*Дрозофила*	0.1×10^9
Птицы	1.2×10^9	*Кукуруза*	7×10^9
Ящерицы	1.9×10^9	*Нейроспора*	4×10^7
Лягушки	6.2×10^9	*E. coli*	4×10^6
Костистые рыбы	0.9×10^9	*Фаг T4*	2×10^5
Двоякодышащие рыбы	111.7×10^9	*Фаг λ*	1×10^5
Иглокожие	0.8×10^9	*Фаг φX174*	6×10^3

Оно (1973) убеждает, что дупликации всего генома обычно более важны, чем тандемные дупликации, поскольку последние могут увеличить только часть единой генетической системы, нарушая отношения между структурными и регуляторными элементами (генами). При полиплоидии, наоборот, происходит умножение всех элементов генома с минимальным вредным эффектом, особенно когда происходит автополиплоидия. Однако молекулярные исследования организации генома эукариот показали, что большинство генов существуют не в единственном экземпляре, но чаще всего в качестве кластеров из нескольких копий. Поэтому можно сделать вывод, что тандемные дупликации также играют значительную роль в эволюции. Хорошо известными примерами кластеров генов у многих организмов являются гены рРНК и тРНК (Табл. 16.2.2). Митохондриальный геном млекопитающих, также как бактерий, имеет только один набор генов рРНК - 12S и 16S (Табл. 16.2.2).

Таблица 16.2.2. Число генов рРНК и тРНК на гаплоидный геном у различных организмов *(По Nei, 1987)*
Table 15.2.2. Numbers of rRNA genes and tRNA genes per haploid genome in various organisms *(From Nei, 1987)*

Ген	***Организм***	***Число***	***Размер генома (пн)***
рРНК	Митохондриальный геном человека	1	16600
	Mycoplasma capricolum	1	1×10^6
	Escherichia coil	1	4×10^6
	Saccharomyces cerevisiae	140	5×10^7
	Drosophila melanogaster	130-250	1×10^8
	Xenopus laevis	400-600	8×10^9
	Человек	300	3×10^9
тРНК	Митохондриальный геном человека	22	16600
	E. coli	100	4×10^6
	S. cerevisiae	320-400	5×10^7
	D. melanogaster	750	1×10^8
	X. laevis	7,800	8×10^9
	Человек	1,300	3×10^9

Тандемные дупликации важны не только как источник увеличения числа генов с однотипной биохимической функцией, но также для генерирования генов с новыми функциями. Рассмотрим в качестве примера ген β –глобина. Семейство генов β –глобина человека имеет пять функциональных генов (рис. 16.2.2): ε функционирует только на ранних фаза эмбриогенеза, G_γ и A_γ производят глобиновые полипептиды только на фетальной стадии, а δ и β являются

ответственными за производство полипептида после рождения. Однако нуклеотидная последовательность этих генов имеет высокую степень сходства. Очевидно, что все эти гены возникли вследствие тандемных дупликаций.

*Рис. 16.2.2. Комплекс локусов **β**-глобина млекопитающих (По Edgell et al., 1983; заимствовано из Nei, 1987).*
*Fig. 15.2.2. Mammalian **β** globin complex loci (From Edgell et al., 1983; adopted from Nei, 1987).*

16.3. ФОРМИРОВАНИЕ НОВЫХ ГЕНОВ

1. Дупликации генов

Тандемные дупликации генов с последующей диверсификацией являются самым простым способом возникновения новых генов. Давайте уточним, каковы механизмы возникновения тандемных дупликаций. Тандемные дупликации могут возникать в результате действия трех механизмов: (1) **неравный кроссинговер** – рекомбинация между гомологичными парами хромосом, при которой на одном из гомологов участок ДНК дуплицируется, а на другом утрачивается; (2) **неравный обмен сестринских хроматид** – рекомбинационное событие сходное с неравным кроссинговером, которое возникает между двумя хроматидами одной и той же хромосомы; (3) **ошибки репликации** – при репликации матричная нить молекулы ДНК может сжиматься, вызывая вставку короткого фрагмента во вторую синтезируемую нить, что может поменять ОРС, то есть – дать новый ген. Такой ген, как правило, будет не функциональным, но в исторической перспективе возможно и приобретение им функции.

Если ген дуплицировался и появились две идентичные копии, то один из них имеет возможность мутировать не вызывая вредных воздействий на приспособленность особей и может постепенно приобрести новую функцию (Оно, 1973). Чтобы установить, является ли пара генов потомками общего предка, обычно исследуют сходство их первичных последовательностей на нуклеотидном или аминокислотном уровнях. Ингрэм (Ingram, 1963) был одним из первых, кто показал, что действительно три цепи гемоглобина человека возникли в результате дупликации первоначально единственного гена. Сопоставление этих трех цепей показывает, что доля общих аминокислот в α и β цепях равна 41%, а в β и γ даже больше – 73% (Табл. 16.3.1). Сходство настолько велико, что вероятность его случайного возникновения крайне мала. Ингрэм позже также обосновал, что миоглобин тоже возник из одного общего с гемоглобинами предкового гена. Позже появилось много других примеров возникновения новых генов посредством генных дупликаций (Табл. 16.3.1).

Эволюционная судьба единичной дупликации гена – это, как правило, потеря его функции и превращение в псевдоген, поскольку вредные мутации происходят чаще, чем полезные (Лекция 6). Реальные экспериментальные факты, подтверждающие данную информацию, имеются для эукариот (Lynch, Conery, 2000) и в изобилии – для прокариот (Kondrashov et al., 2002). Тем не менее, например, у позвоночных темп дупликаций и их фиксации высок, он оценен как одно событие на 100 млн. лет (Lynch, Conery, 2000). Благодаря тому, что темп дупликаций высок, и равен приблизительно темпу замен нуклеотидов или точковых мутаций, их роль в эволюции

остается очень значительной. Интересно, что некоторые гены, например ген, кодирующий ксантиндегидрогеназу (*Xdh*), дуплицируются повторно и конвергентно эволюционируют к одинаковой функции, обеспечиваемой геном альдегид оксидазы, *Ao* (Rodriguez-Trelles et al., 2003; Pereira, 2004), хотя теоретически это должно быть очень редким событием (Pereira, 2004).

2. Элонгация генов

Иной способ изменения гена – это его удлинение или элонгация. Элонгация гена может происходить за счет частичной или полной его дупликации. Многие белки живых организмов имеют внутренние повторы аминокислотных последовательностей, и эти повторы соответствуют различным структурно-функциональным доменам белков (см. Лекция 3). Можно предположить, что гены, кодирующие такие белки, возникли за счет внутригенных дупликаций. Сейчас становится популярной точка зрения о том, что гены эукариот возникли за счет дупликаций и элонгаций примордиальных генов или мини генов, которые могли существовать на ранних фазах эволюции жизни (Darnell, 1978; Doolittle, 1978; Ohno, 1981; Blake, 1985; Nei, 1987; Nei, Kumar, 2000).

Таблица 16.3.1. Дивергенция и функциональные различия между белками, возникшими за счет генных дупликаций *(По Dayhoff, Barker, 1972; заимствовано из Nei, 1987).*

Table. 15.3.1. Extents of divergence and functional differences between proteins derived from gene duplications. Chemical activities include differences in catalytic action and in binding to substrates, inhibitors, antigens, etc. *(From Dayhoff, Barker, 1972; adopted from Nei, 1987)*

Белки	***Различия в амино-кислотах (%)***	***Время Диверген-ции (млн. лет)***	***Химические активности***	***Агрегаци-онные свойства***	***Активные центры***
Гемоглобин-миоглобин	77	1100	—	++	+
Гормон роста - пролактин	75	200	+	—	+
Иммуноглобулин, Тяжелые и легкие цепи	75	400	++	+	—
Иммуноглобулин, μ и γ цепи С регионов	70	350	+	+	+
Трипсин-тромбин	65	1500	+	—	+
Лактальбумин - лизоцим	63	350	++	—	+
Иммуноглобулин, κ и λ цепи С регионов	62	300	—	—	—
Гемоглобин, α и β цепи	59	600	—	+	—
Гемоглобин, β и γ цепи	27	130	—	—	—
Протамины, сальмин AI и AII	22	100	—	—	—
Хемотрипсин A и B	21	270	—	—	—
Гормон роста - лактоген	15	23	+	—	+
Гемоглобин, β и δ цепи	8	40	—	—	—
Алкогольдегидрогеназа, F и S цепи	2	10	+	—	—

Примечание. + + Очень разные. + Разные. — Сходные. Химические активности включают различие в каталитическом действии и связывании с субстратом, ингибитором, антигенами и т.д.

В частности, данная точка зрения поддерживается данными о первичной последовательности овомукоидного гена у птиц. Овомукоид – это белок, присутствующий в яйцах и являющийся ответственным за трипсино-ингибиторную активность в яичном «белке». Овомукоидная белковая

цепь состоит из трех функциональных доменов (рис. 16.3.1). Каждый из этих доменов ответственен за связывание одной молекулы трипсина или другой сериновой протеиназы. Сходство аминокислотных последовательностей между доменами I и II, I и III, а также II и III составляет 46%, 33% и 30%, соответственно. Такая высокая степень сходства предполагает, что овомукоидный ген возник за счет трипликации домена примордиального гена (Kato et al., 1978).

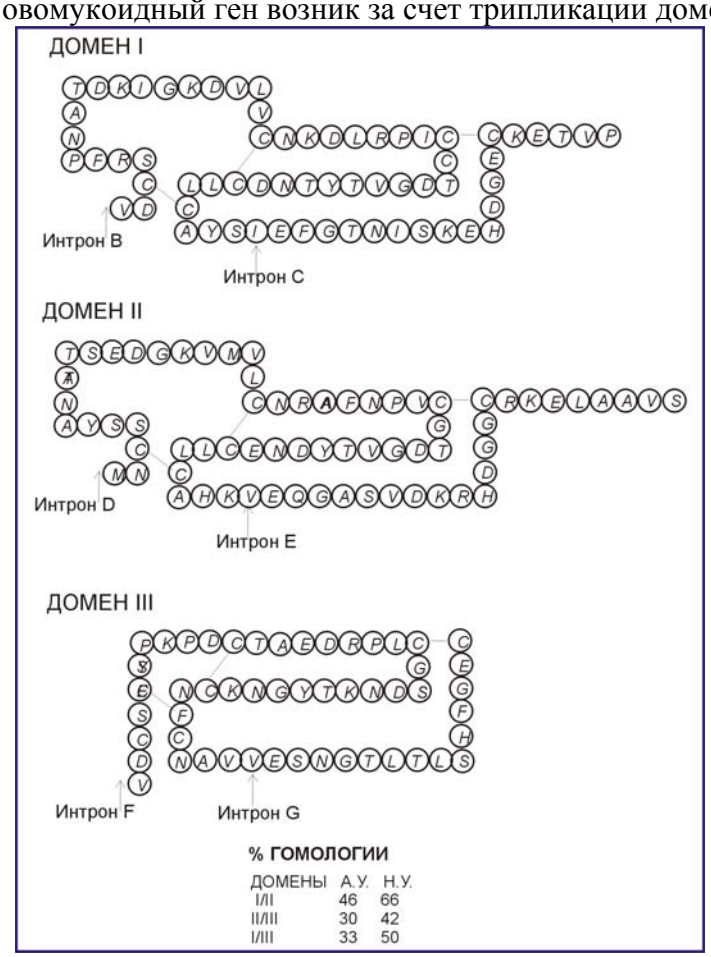


Рис. 16.3.1. Три функциональных домена секретируемого овомукоидного белка у кур. Гомологии между доменами на аминокислотном уровне (А.У.) и на уровне мРНК (Н.У.) показаны в нижней части рисунка (По Stein et al., 1980; заимствовано из Nei, 1987).

Fig. 15.3.1. Three functional domains of the secreted ovomucoid in chicken. The homologies between domains at the amino acid (A.A.) level and the mRNA (N.A.) level are shown at the bottom of the figure (From Stein et al., 1980; adopted from Nei, 1987).

Число примеров элонгации генов за счет дупликаций может быть увеличено (Li, 1983; Doolittle, 1985; Klug, Cummings, 2002). Многие случаи включают один или более повторов доменов. Таким образом, дупликационные события могут быть твердо обоснованы на основе данных о первичных последовательностях. Если идентифицировать функциональные домены трудно, то на помощь для открытия древних функциональных доменов могут прийти особенности упаковки молекулы белка (Go, 1981), его третичная и четвертичная структуры. Вполне возможно, что модули, определяющие эти уровни организации, представляют собой продукты примордиальных генов древних организмов. Предполагается, что примордиальные гены были очень короткими и что большинство современных генов возникли за счет нескольких раундов дупликаций этих предковых генов (Ohno, 1984).

3. Гибридные гены

Неравный кроссинговер может произойти в участке ДНК, включающем два гена. Это может дать новый ген, содержащий части этих двух генов. Гибридные гены, как правило, должны давать вредный эффект. Если же исходный ген был дуплицирован и сохранился, то гибридный ген может эволюционировать в новый ген. Возможный пример подобного возникновения гена представляет Z ген клупеина сельди, который, вероятно, возник вследствие кроссинговера между генами клупеина VI и VII (Fitch, 1971b). Вероятно, что эти три гена возникли вследствие простой дупликации и последовавшие аминокислотные замены редки здесь (Fitch, 1971b). Гилберт (Gilbert, 1978) предложил гипотезу перетасовки экзонов, а позже Дулиттл (Doolittle, 1985) проанализировал в этом аспекте последовательности аминокислот многих белков у высших организмов. Они обнаружили, что, казалось бы, несвязанные белки часто имеют частичную гомологию последовательностей и пришли к выводу, что это обусловлено перетасовкой одних и тех же

экзонов в разных генах. Эти авторы рассматривают такой механизм, как один из весьма важных для формирования новых генов. Новые данные, например для низкоплотностного липопротеина человека, очень хорошо поддерживают такую интерпретацию (Klug, Cummings, 2002, P. 272; Лекция 3, рис. 3.3.9).

16.4. ПОВТОРЯЮЩИЕСЯ ДНК И МУЛЬТИГЕННЫЕ СЕМЕЙСТВА

Геном эукариот содержит различные типы повторяющихся ДНК (рис. 16.4.1), первоначально обобщенных Бриттеном и Конэ (Britten, Kohne 1968). Эти ДНК включают последовательности нуклеотидов, которые повторяются в геноме от нескольких до миллионов раз. Количество повторяющейся ДНК сильно варьирует у различных организмов. Так, высокоповторяющаяся ДНК составляет 5% генома человека и 10% генома мыши. Функции многих повторяющихся ДНК не известны. Некоторые ее классы, по-видимому, являются не функциональными («ДНК хлам», Ohno, 1972). Однако некоторые группы повторяющейся ДНК имеют ясно очерченные функции: кодирование тРНК, рРНК, множественные семейства структурных генов (мультигенные семейства), центромерные повторы и т.д. (рис. 16.4.1).

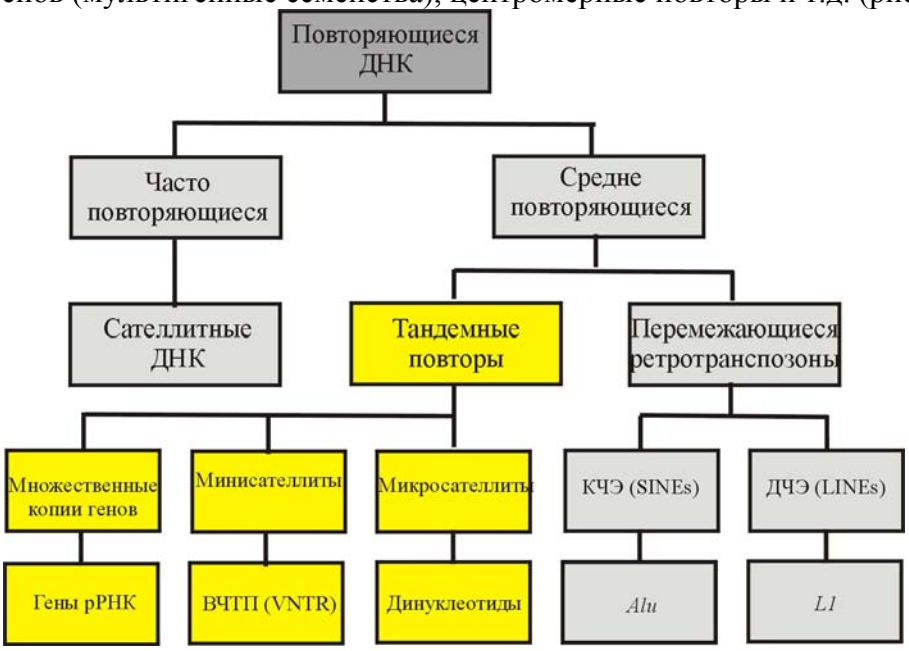


Рис. 16.4.1. Различные категории повторяющихся ДНК.
Объяснение сокращений и роли повторов ДНК даны в тексте (По Klug, Cummings, 2002).
Fig. 15.4.1. An overview of the various categories of repetitive DNA.
Explanation of abbreviations and the role of DNA repeats are in the text (From Klug, Cummings, 2002).

Нуклеотидный состав ДНК (например, доля пар G ≡ C по отношению к A = T парам) для определенного вида отражает ее плотность, что может быть измерено посредством центрифугирования в градиенте плотности. Когда таким способом анализируется ДНК эукариот, то большинство ее молекул представлено одним главным пиком или фракцией, с достаточно однообразной плотностью. Однако обычно также наблюдаются добавочные пики, которые немного отличаются по своей плотности. Эта часть молекул, называемая **сателлитной ДНК** (**satellite** DNA), составляет варьирующую по количеству от вида к виду долю суммарной ДНК. Для примера, соотношение основной фракции и сателлитной ДНК у мыши показано на рис. 16.4.2. В контрасте с этим, ДНК прокариот представлена только одним пиком. Значение сателлитной ДНК оставалось загадкой природы до середины 1960-х годов, когда Бриттеном и Кони (Britten, Kohne, 1967) была разработана методика для измерения кинетики реассоциации ДНК, которая была

предварительно диссоциирована в единичные цепи. Эти авторы продемонстрировали, что определенная доля молекул ДНК повторно ассоциирует более быстро, чем другие. Они сделали вывод, что быстро реассоциирующие молекулы представлены множественными копиями ДНК.

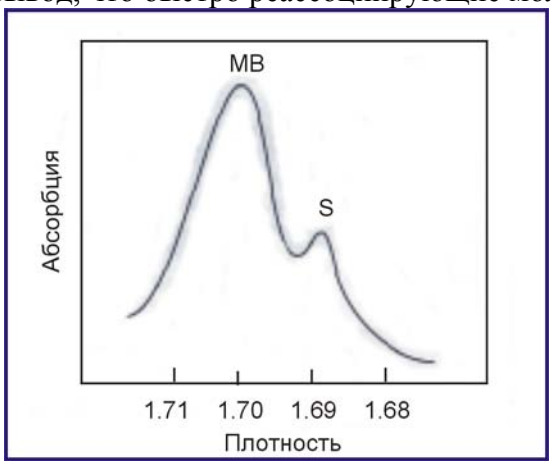


Рис. 16.4.2. Распределение основного пика (MB) и добавочного пика, сателлитной ДНК (S) у мыши с использованием ультрацентрифугирования в градиенте плотности CsCl (По Klug, Cummings, 2002). Fig. 15.4.2. Separation of main-band (MB) and satellite (S) DNA from the mouse using ultracentrifugation in a CsCl gradient (From Klug, Cummings, 2002).

Центромерные и теломерные последовательности ДНК

Разделение хроматид является существенным для точности деления хромосом в ходе митоза и мейоза. Определение числа ошибок при делении показывает, что они весьма редки: 1×10^{-5} до 1×10^{-6}, или происходит только 1 ошибка на 100000 - 1000000 клеточных делений. Поэтому, было выдвинуто предположение, что исследование последовательностей ДНК в центромерных регионах будет перспективно для понимания тонкого строения хромосом и генома в целом. Этот участок ДНК обозначается как ЦЕН (**CEN**; от слова центромера**).** Его функции и структура сейчас достаточно понятны. Структура особого участка внутри гетерохроматинового региона такова, что дает, во-первых, основу для связывания белков, формирующих центромеру, и, во-вторых, присоединяет кинетохор (дающий точки прикрепления нитей веретена при делении клетки). Основы для понимания ЦЕН региона заложили исследования хромосом дрожжей *S. cerevisiae*, которые дали принципиальную модель, впервые изложенную Карбоном и Кларком (Carbon and Clarke; цитировано по Klug, Cummings, 2002). Поскольку каждая центромера имеет идентичную функцию, то не удивительно, что все ЦЕН-участки имеют очень большое сходство в организации. ЦЕН-регион дрожжей состоит примерно из 225 пн., которые могут быть подразделены на три участка (рис. 16.4.3). Первый и третий участки (I и III) являются относительно короткими и высоко консервативными, составляя только 8 и 26 пн., соответственно. Участок II, который больше (80-85 пн) и исключительно А-Т обогащен (до 95% нуклеотидов) изменчив по своей последовательности в разных хромосомах.

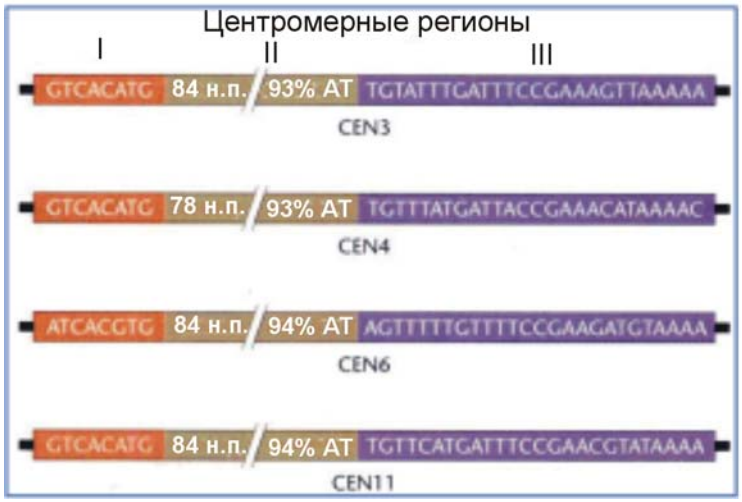


Рис. 16.4.3. Информация о последовательности нуклеотидов, полученная для ДНК трех основных центромерных участков (I, II, III) хромосом 3, 4, 6 и 11 дрожжей (По Klug, Cummings, 2002).
Fig. 15.4.3. Nucleotide sequence information derived from DNA of the three major centromere regions of chromosomes 3,4, 6, and 11 of east (From Klug, Cummings, 2002).

Перейдем теперь к теломерным участкам хромосом. Имеются два типа теломерных последовательностей (ТП; ТП являются существенными для стабильности и целостности хромосом). Один тип ТП называется теломерные последовательности ДНК. Они состоят из коротких тандемных повторов. У реснитчатого организма, *Tetrahymena* обнаружено свыше 50 тандемно повторенных гексануклеотидных последовательностей GGGGTT. У человека похожая GGGATT последовательность также повторяется много раз. Анализ ТП в ДНК показал, что они строго консервативны в ходе эволюции и отражают ту критическую роль, которую они играют в поддержании целостности хромосом. Другой тип ТП – это теломерно-ассоциированные последовательности. Они также являются повторяющимися. Этот тип ТП обнаружен как по соседству с теломерой, так и внутри ее. Данные о ТП варьируют в размере и повторяемости у разных организмов, и их функция пока остается не ясной.

Умеренно повторяющиеся последовательности ДНК

Имеется другая известная категория повторяющихся ДНК, которые важны для понимания структуры и эволюции генома эукариот. Это средне или умеренно повторяющиеся ДНК (**middle** or **moderately repetitive DNA**). Они исследованы достаточно подробно. У человека умеренно повторяющиеся ДНК чаще всего состоят либо из тандемных повторов или чередующихся (interspersed) последовательностей. Какой либо функции, выполняемой в геноме этими ДНК не известно. Те, которые охарактеризованы, включают ДНК с **варьирующим числом тандемных повторов**, ВЧТП (variable number tandem repeats**,** VNTR). Эти ВЧТП обычно длиной 15-100 пн и обнаруживаются как между генами, так и внутри них. ВЧТП часто называются **мини сателлитами,** и эти ДНК рассеяны по всему геному. Другая группа тандемно повторяющихся последовательностей ДНК представлена динуклеотидами, чаще всего их называют **микро сателлитами**. Также как ВЧТП, они рассеяны по всему геному и варьируют между особями по числу повторов, представленных в сайтах. У человека наиболее обычным микро сателлитом является динуклеотид СА, с числом повторений - 5-50 копий. Микро сателлиты сейчас широко используются как молекулярные маркеры для популяционно-генетического анализа.

Повторяющиеся перемещающиеся последовательности

Другая категория повторяющихся ДНК состоит из последовательностей, которые рассеяны по геному, но не являющихся тандемными. Они могут быть длинными или короткими, но главной их отличительной чертой является способность к перемещению в геноме. Это, так называемые, **перемещаемые или мобильные последовательности, транспозоны** (transposable or mobile sequences). Большая часть генома эукариот состоит из таких последовательностей, впервые

открытых Мак-Клинток (B. Mc-Clintock) у кукурузы. Один из типов **коротких чередующихся элементов**, КЧЭ (SINE, short interspersed elements), по размеру достигает 500 пн. Он может быть представлен, например, в геноме человека 500000 раз или даже больше. Одна из хорошо известных КЧЭ - это набор близко родственных последовательностей у человека, из ***Alu* -семейства** (названных по имени распознающей их рестриктазы *AluI). Alu* –семейство обнаружено также и у других млекопитающих. Обычно они 200-300 пн длиной и распределены достаточно однообразно в разных частях генома, и в генах, и между ними. Повторы *Alu* –семейства иногда транскрибируются. Хотя роль соответствующей РНК не очень понятна, но она должна быть определенным образом связана с их подвижностью в геноме. Предполагается, что *Alu* – последовательности возникли из РНК-элемента, чей ДНК комплемент был диспергирован в геноме как следствие активности обратной транскриптазы.

Группа **длинных чередующихся элементов**, ДЧЭ (long interspersed elements, LINE) – это еще одна категория перемещающихся последовательностей ДНК. У человека наиболее известным является семейство, обозначаемое как ***L1***. Члены этого семейства последовательностей достигают длины 6400 пн. Они могут быть представлены в геноме 100000 копий. У ***L1*** 5'–конец является высоко вариабельным. Роль этих последовательностей, также как механизм их транспозиции до конца не выяснены. Последовательность ***L1*** ДНК сначала транскрибируется в молекулу РНК. После этого РНК служит как матрица для синтеза ДНК комплемента с помощью обратной транскриптазы. Этот фермент кодируется среди других самой последовательностью ***L1***. Новая копия ***L1*** далее встраивается в ДНК хромосомы в новом сайте. Из-за сходства этого механизма перемещения с тем, который используется ретровирусами, ДЧЭ также называют **ретротранспозонами**. КЧЭ и ДЧЭ (SINEs и LINEs) составляют существенную часть ДНК человека. Оба перечисленных представителя этих последовательностей ДНК имеют сходные черты организации и состоят из смеси примерно 70% уникальных и 30% повторяющихся последовательностей. Вместе КЧЭ и ДЧЭ составляют около 10% генома.

Средне повторяющиеся множественные копии генов.

В некоторых случаях средне повторяющаяся ДНК включает функциональные гены, представленные тандемно множественными копиями – это **мультигенные семейства** (рис. 16.4.4). Например, существуют много копий генов, кодирующих не только рРНК, но и тРНК, как уже упоминалось в Разделе 16.2 (см. Табл. 16.2.2).

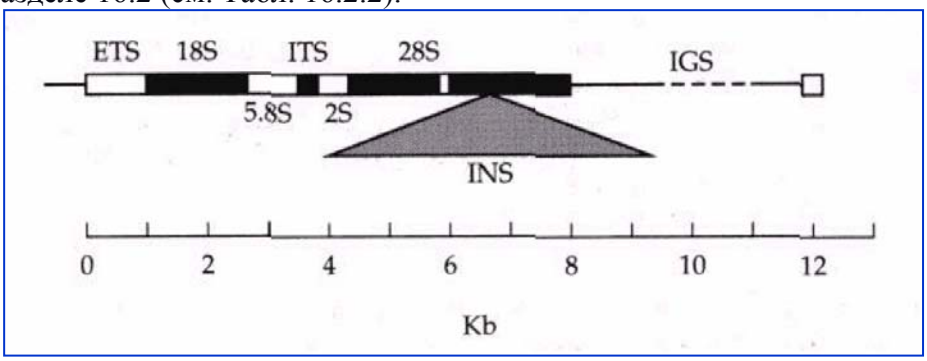


Рис. 16.4.4. Организация рРНК гена у Drosophila melanogaster - тандемно повторенного мультигенного семейства.

В добавок к 18S, 5.8S, 2S и 28S рРНК генам, повторяющаяся единица содержат внешний транскрибируемый спейсер (external transcribed spacer, ETS), внутренний транскрибируемый спейсер (internal transcribed spacer, ITS), а также внутригенную последовательность. Некоторые копии также содержат вставки (Insertions, INS) (По Hartl, Clark, 1989).

Fig. 15.4.4. Organization of the rRNA genes in Drosophila melanogaster, a tandemly repeated multigene family. In addition to the 18S, 5.8S, 2S, and 28S rRNA genes, the repeated unit contains an

external transcribed spacer (ETS), an internal transcribed spacer (ITS), and an intergenic sequence. Some copies also contain insertion sequences (INS) (From Hartl, Clark, 1989).

Эволюция мультигенных семейств.

Между особями одного вида различные копии генов мультигенного семейства обычно обладают большим сходством друг к другу, чем к членам гомологичных генных семейств, представленных у других родственных видов (см. рис. 16.2.2, Табл. 16.3.1). Число этих генов увеличивается посредством дупликаций, об основных механизмах которых было сказано ранее. Дополнительно возможно их увеличение за счет механизма РНК-обусловленной транспозиции, амплификации генов и их обмена (Maeda, Smithies, 1986). Используя методики молекулярной филогенетики можно реконструировать родословную индивидуальных членов мультигенного семейства, как это сделано для глобинового семейства (рис. 16.4.5).

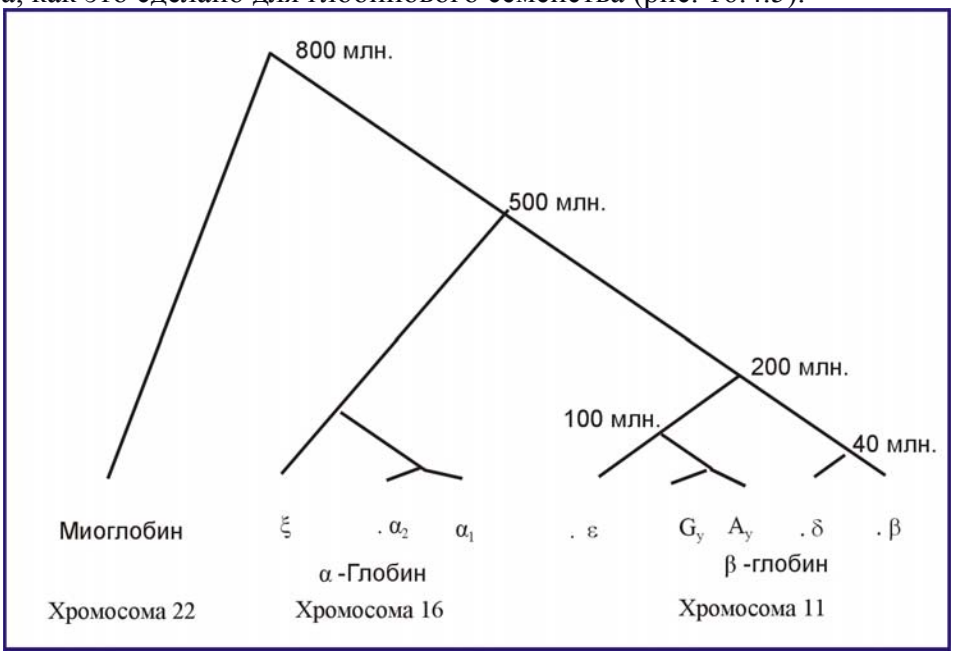


Рис. 16.4.5. Эволюционная история мультигенного семейства глобина.

Около 800 миллионов лет назад (млн.), дупликационное событие в предковом гене дало начало двум филогенетическим линиям. Одна из них привела к гену миоглобина, который локализован у человека на хромосоме 22. Другая линия, претерпев второе дупликационное событие около 500 млн., дала начало предкам альфа- и бетта-подсемейств этого семейства. Дупликации около 200 млн. произвели гамма- и бетта-глобиновые подсемейства. У человека α-глобиновый ген локализован на хромосоме 16, а β-глобиновые гены – располагаются на хромосоме 11 (По Klug, Cummings, 2002).

Fig. 15.4.5. Evolutionary history of the globin gene superfamily. About 800 million years ago (MYA), a duplication event in an ancestral gene gave rise to two lineages. One led to the myoglobin gene, which in humans is located on chromosome 22. The other lineage underwent a second duplication event about 500 MYA, giving rise to the ancestors of the alpha and beta subfamilies. Duplications about 200 MYA produced the gamma and beta globin subfamilies. In humans, the α-globin genes are located on chromosome 16 and the β-globin genes are on chromosome 11 (From Klug, Cummings, 2002).

Механизм согласованной эволюции

Члены мультигенных семейств имеют тенденцию сохранять больше сходства, чем можно ожидать при случайной, дрейфовой дифференциации. Они стремятся эволюционировать совместно (согласованно) поскольку существует механизм, который ведет к единообразию последовательности внутри семейства. Эта тенденция к единообразию известна, как **согласованная эволюция** (concerted evolution). Предложено два механизма согласованной

эволюции (рис. 16.4.6). Первый – это **конверсия гена**, процесс, при котором спаривание нуклеотидов между двумя достаточно гомологичными генами сопровождается утерей всей или части нуклеотидной последовательности в одном гене и ее заменой репликой нуклеотидной последовательности другого гена (рис. 16.4.6a). Формально это ведет к тому, что один ген обращает (конвертирует) другую последовательность в себе-подобную последовательность. Второй механизм – это согласованная эволюция за счет неравного кроссинговера (рис. 16.4.6b).

В тандемных мультигенных семействах может возникать неточное спаривание генов в ходе мейоза, и кроссинговер приводит к образованию гамет, которые могут иметь либо увеличенное, либо уменьшенное число копий. Неравный кроссинговер сам по себе не дает постоянного увеличения числа копий, скорее он приводит к увеличению изменчивости в числе копий. Таким образом, последующий кроссинговер может либо скорректировать число копий, либо далее увеличить их число. Формально конечный результат в этом случае будет эквивалентен эффекту при конверсии гена, поскольку нуклеотидные последовательности меняются таким же способом.

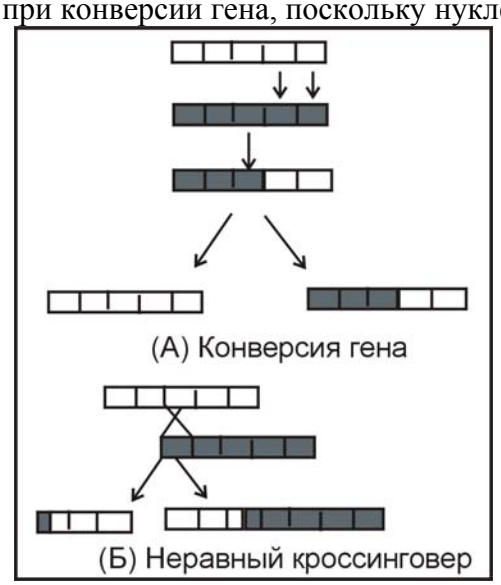


Рис. 16.4.6. Два механизма согласованной эволюции, которые могут единообразить последовательности членов мультигенных семейств.

(А) Конверсия гена, при которой генетическая информация в одном гене удваивается, и затем замещает генетическую информацию его гомолога. (Б) Неравный кроссинговер. Увеличение числа членов генного семейства осуществляется, поскольку неравный кроссинговер может исправляться последующим неравным кроссинговером, дающим идентичные последовательности ДНК среди нескольких соседних членов мультигенного семейства (По Hartl, Clark, 1989).

Fig. 15.4.6. Two mechanisms of concerted evolution that can homogenize the sequences of members of multigene families, (a) Gene conversion, in which genetic information in one gene is replicated and replaces genetic information in its homolog. (b) Unequal crossing over. The increase in the number of members of the gene family because of unequal crossing over can be corrected by a subsequent unequal crossing over, resulting in identical DNA sequences among several adjacent members of the multigene family (From Hartl, Clark, 1989).

Роль неравного кроссинговера в увеличении числа генов была постулирована давно (Bridges, 1936). Однако только после того как молекулярные данные стали многочисленными, стало ясно, что он и в действительности играет важную роль в увеличении или уменьшении содержания ДНК и, соответственно, числа генов. В частности для таких мультигенных семейств, как семейства генов иммуноглобулина и рРНК, неравный кроссинговер может быть главным фактором увеличения их численности (Hood et al., 1975). Положения о том, что конверсия играет существенную роль для увеличения размера генома, также основательно проработаны (Radding, 1982). Отмечено (Slightom et al., 1980), что часть второго интрона человеческого A_γ глобинового гена является очень сходной с последовательностью негомологичного гена G_γ той же самой особи. Цитированные авторы предположили, что часть гена A_γ была «конвертирована» межгенным обменом и стала более похожей на ген G_γ. Конверсия гена была обнаружена также и у обыкновенной мидии (Rawson et al., 1996).

Работа механизмов конверсии гена и неравного кроссинговера поддерживает сходство между копиями генов в мультигенных семействах. Теоретическое осмысливание и анализ этих механизмов представлены в нескольких работах (Ohta, 1982; Nagylaki, 1984 и др.).

Псевдогены

Эволюционные изменения псевдогенов представляют особенный интерес, поскольку они могут быть использованы для тестирования теории нейтральных мутаций. Поскольку псевдогены не имеют функциональной нагрузки, то темп нуклеотидных замен для них должен быть выше, чем для структурных генов. Как упоминалось ранее, псевдогены в действительности эволюционируют быстрее (Лекция 6, Табл. 6.5.1). Как только ген становится не функционирующим, его очень трудно оживить. Дело в том, что имеется несколько путей для нарушения работы гена и обратные мутации, которые могли бы его восстановить, должны происходить точно по тем же самым точкам, где происходили прямые мутации; вероятность таких событий весьма мала. Более того, после того как ген перестал функционировать, в нем продолжают возникать мутации, многие из которых являются вредными. Поэтому вероятность восстановления псевдогена будет уменьшаться во времени.

Эндогенные ретровирусы у млекопитающих

Различные организмы, как подчеркнуто выше, включая млекопитающих, содержат в норме множественные копии ДНК нескольких различных ретровирусов. Последовательность ДНК, которая гомологична ретровирусам, и которая становится фиксированной в геноме вида, известна, как эндогенная ретровирусная последовательность, или вирогены (virogenes). В геноме позвоночных животных обнаружено несколько типов эндогенных последовательностей. Вирогены обычно представлены 10-100 копиями и репродуцируются посредством обратной транскрипции РНК и её реинтеграции в ДНК хозяина. При подходящих условиях в клетках эндогенные вирусы могут давать зрелые вирусные частицы. Не часто, но такие вирусы способны давать инфекции не только в организме своего хозяина, но также и у других особей, в том числе и размножаясь в клетках организмов близких видов. Высвободившиеся вирусы ведут себя как типичные ретровирусы и их небольшая часть, через механизм обратной транскрипции вновь встраивается в ДНК хозяина. Эти и другие наблюдения предполагают, что эндогенные ретровирусные последовательности могут быть полезными для хозяина, создавая некоторую степень иммунности против других вирусных инфекций (Benveniste, 1985).

Эндогенные вирусы, как выяснилось, сохраняются в геноме длительное время. Например, тип С эндогенного ретровируса бабуина обнаружен у всех видов обезьян Старого Света, но не был обнаружен у обезьян Нового Света. Данный тип ретровируса, соответственно, внедрился в линию предков обезьян уже после их дивергенции в связи с расхождением континентов и сохранился у обезьян Старого Света минуя многочисленные видообразования в этой группе. Тип С эндогенного вируса бабуина также иллюстрирует быструю скорость эволюции, обычно обнаруживаемую для вирогенов (рис. 16.4.7). Приведенные данные суммируют гибридизационные эксперименты на ДНК с двумя типами проб ДНК бабуина *Papio cynocephalus,* при их гибридизации с ДНК других видов приматов (рис. 16.4.7). Первая проба была клеточной ДНК бабуина и по ней построена линия, с равными значениями для обеих осей (рис. 16.4.7, нижняя сплошная линия). Заметим, что значения для ДНК всех видов попадают на эту линию, что указывает на относительное постоянство нуклеотидных замен, служащее молекулярными часами и отражаемое в величине ΔTm, при усреднении по всему не повторяющемуся геному.

Надо окончить эту, бесконечную как сама жизнь, тему об эволюции генома. Мы, безусловно, понимаем, что жизнь и геном, как её важнейшая составляющая, эволюционировали от простого к более сложному в своей организации. Простые кольцевые ДНК прокариот, с 250-350 кодирующими генами, со временем превратились в такие тонко структурированные образования, как хромосомы эукариот, включающих до 26588 кодирующих генов, как, например, у человека. Функции генов у прокариот и эукариот также имеют много различий. Эти функции тоже

развивались в ходе эволюции, в основном, по линии усложнения, что в итоге привело к иерархическим, каскадным, гомеостазисным и сете-подобным принципам регуляции на клеточном, индивидуальном и популяционном уровнях у эукариот. Эволюция многих генов и их частей порождается случайными или стохастическими процессами. Однако большинство генных функций, и геном в целом, высоко консервативны и постоянно находятся под действием естественного отбора, чаще всего очищающего отбора. Как мы имели возможность убедиться в ходе данного курса лекций, большую роль играют и другие виды естественного отбора, в том числе и направленного, дающего начало новым генам с их новыми функциями. Более обычны в природе различные взаимодействия случайных и систематических факторов.

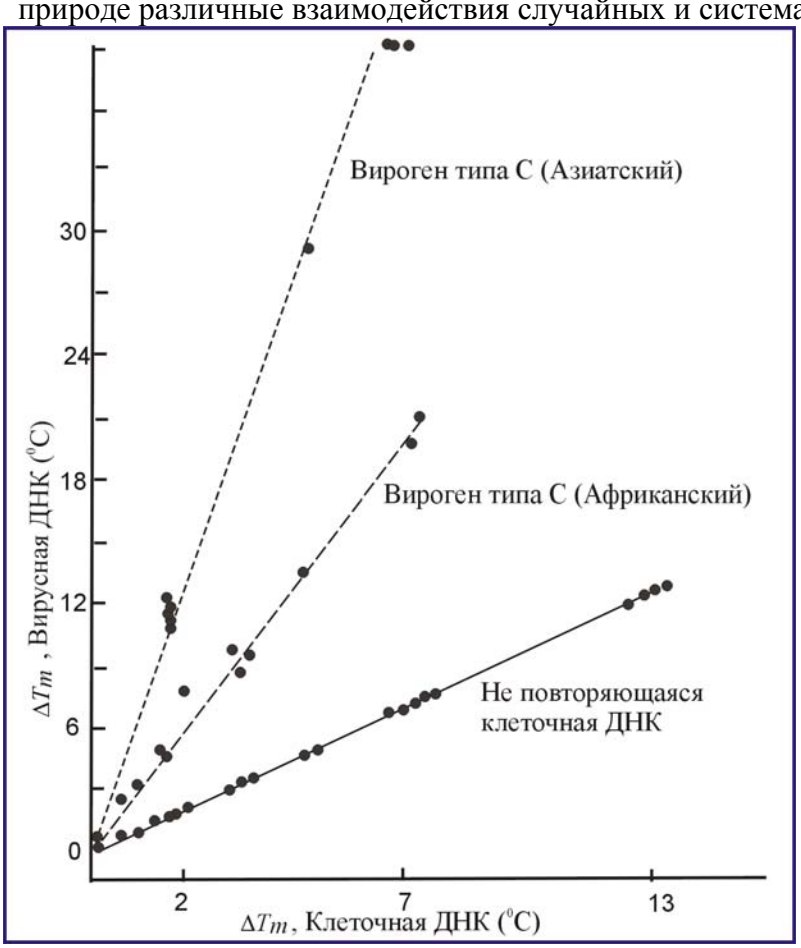


Рис. 16.4.7. Скорость эволюции эндогенного ретровируса типа С бабуина (пунктирные линии) при сравнении с не повторяющимися клеточными ДНК приматов (сплошная линия).

У африканских приматов эти вирогены эволюционировали почти в три раза быстрее, чем не повторяющиеся клеточные ДНК. В сравнении с азиатскими приматами вирогены эволюционировали в шесть раз быстрее (верхняя линия; по Benveniste 1985; заимствовано из Hartl, Clark, 1989 с упрощениями).
Fig. 15.4.7. Evolutionary rate of baboon Type C endogenous retrovirus (dashed lines) as compared with nonrepetitive cellular DNA (solid). In African primates, these virogenes have evolved almost threefold faster than nonrepetitive cellular DNA, and in Asian primates they have evolved about sixfold faster (After Benveniste 1985; adopted from Hartl, Clark, 1989).

16.5. ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 16

1. Прочтите Приложения II и IV. Выберите темы своих рефератов, консультируясь с преподавателем и своим научным руководителем.
2. Консультация по всем темам Практических занятий.

Foreword *3*

Introduction *4*

Chapter 1. EVOLUTIONARY HISTORY OF LIFE. *4*

1.1. Main Stages in Earth Geological Evolution.

1.2. Paleontology Dating of Evolution.

1.3. Molecular Biological Dating.

1.4. Self-Education Tasks & Training.

Chapter 2. THE MATERIAL BASIS OF HEREDITY. *13*

2.1. Chromosome Structure.

2.2. Meiosis and Gamete Formation.

2.3. Mendel's or Transmission Genetics. Mono-, Di-, and Trihybrid Crosses. Mendel's Lows.

2.4. Fine Gene Structure. Mutations and Modifications.

2.5. Training Course, #2.

Chapter 3. TRANSLATION OF GENETIC INFO. AN INTRODUCTION TO PROTEOMICS.

3.1. Components that Critical for Protein Synthesis. *23*

3.2. The process of Translation: From RNA to Polypeptide.

3.3. Heredity, Proteins and Function.

3.4. Training Course, #3.

Chapter 4. IMMUNOGENETIC ANALYSIS AND EVOLUTIONARY DISTANCES. *39*

4.1. Introduction to Immunogenetics.

4.2. Inheritance of Immunogenetic Traits.

4.3. Immunogenetic Dating of Evolution.

4.4. Dating of Evolution by Amino Acid Sequence Analysis.

4.5. Training Course, #4.

Chapter 5. BIOCHEMICAL GENETICS. *54*

5.1. Introduction to the Methods of Detection of Protein Variability.

5.2. Interpreting of Protein Variability.

5.3. Genetics of Isozymes.

5.4. Evolution of a Genome Regulation.

5.5. Training Course, #5.

Chapter 6. GENES AND MUTATIONS. *67*

6.1. Introduction to Mutation Study.

6.2. Structure and Functions of a Gene.

6.3. Type of Mutations on DNA Level.

6.4. Mutations and Amino Acid Substitutions.

6.5. Mutation Rate.

6.6. Training Course, #6.

Chapter 7. GENETIC ASPECTS OF SPECIATION. *82*

7.1. General Genetic Approach: Advances and Limitations.

7.2. What Data are Necessary? What the Data Base is?

7.3. Species Concept. The Review of Literature Data on Heterozygosities and Distances.

7.4. Speciation Modes (SM): Population Genetic View.

7.5. Training Course, #7.

Chapter 8. POPULATION GENETIC STRUCTURE OF A SPECIES. *103*

8.1. Population Structure in Bisexual Species.

8.2. Modes of Intraspecies Structure.

8.3. Methods and Principles in Studies of Population Genetic Structure.

8.4. A Field Population Genetic Research.

8.5. Training Course, #8.

Chapter 9. ANALYSIS OF GENETIC DIVERSITY WITHIN A SPECIES: THEORY AND EXPERIENCE. *118*

9.1. Main Population Parameters.

9.2. Quantitative Measures of Differentiation.

9.3. Models of Population Structure.

9.4. Empirical Estimates of Parameters.

9.5. Training Course, #9.

Chapter 10. NATURAL HYBRIDIZATION AND INTROGRESSION IN WILD SPECIES *134*

10.1. What Hybrid and Hybridization are?

10.2. Methods of Detecting Hybridization.

10.3. Genetic Interpretation of Hybridized Populations.

10.4. Empirical Investigations of Hybrid Zones.

10.5. Training Course, #10.

Chapter 11. QUANTITATIVE TRAITS: INHERITANCE AND EVOLUTION. *157*

11.1. Quantitative Variation and its Lows.

11.2. The Multiple Factor Hypothesis.

11.3. Main Tasks of Genetic Research of Quantitative Traits: σ^2_{PH}, σ^2_G, Cv, h^2, and other Variables.
11.4. Heritability.

11.5. Mapping of Quantitative Trait Loci (QTL). QTL Evolution.

11.6. Training Course, #11.

Chapter 12. GENETIC HOMEOSTASIS, HETEROSIS AND HETEROZYGOSITY. *176*

12.1. Concept of Genetic Homeostasis.

12.2. Heterosis and Heterozygosity.

12.3. Empirical Results and Interpreting.

12.4. Training Course, #12.

Chapter 13. DNA PLYMORPHISM WITHIN AND AMONG POPULATIONS. *191*

13.1. DNA Sequence Polymorphism. Nucleotide Diversity and *p*-distance.

13.2. DNA Polymorphism Estimated from Restriction Site Data.

13.3. DNA Length Polymorphism. Theory and Observations

13.4. DNA Divergence Among Populations.

13.5. Training Course, #13.

Chapter 14. PHYLOGENETIC TREES. *209*

14.1. Types of Phylogenetic Trees.

14.2. Method of Distance Matrix.

14.3. Parsimony and Maximal Likelihood Methods.

14.4. Role of Population genetic Theory in a Tree Construction.

14.5. Training Course, #14.

Chapter 15. PRACTICAL TRAINING WITH NUCLEOTIDE SEQUENCES: EDITION, SUBMISSION, ALIGNMENT, TREE BUILDING AND ANALYSIS OF PHYLOGENY. *227*
15.1. Edition of sequences and their submission to a gene bank.

15.2. Data formats and gene banks available.

15.3. Alignment of sequences.

15.4. Chousing an optimal substitution model.

15.5. Tree building using MEGA-3 (MEGA-4) program package. An annotation on programs PAUP, MrBayes and some others.
15.6. Training Course, #15.

Chapter 16. GENOMIC EVOLUTION. *246*

16.1. Origin and Change of Genome.

16.2. Evolutionary Change of Genome Size.
16.3. Formation of New Genes.
16.4. Repetitive DNA and Multigene Families.
16.5. Training Course, #16.
Content in English *265*
Subject Index *269*
Appendix I. Methodical Instruction For Self-Education. *270*
Appendix II. Exam Questions for the Course *274*
Appendix III. Main Literature & Web-Support. *278*
Appendix IV. Explanation of Basic Terms. *279*
Appendix V. Complete List of References. *283*

Предметный указатель

Аллель – 14,15,21,42,44,93,98,112,119,142,150,167.
Аллельные продукты – 58.
Аллельные варианты – 74.
Белки – 23,28,30,32-37,47,53,54,57-60,65,67,72-74,77,79,191,196,198.
Ген – 13,17,20,21,23,28,29,30,42,55,75,76,119,147,148,149,189,191,192,193,194,195,199,200.
Законы Менделя – 13,17.
Время дивергенции – 12,65,78,79,81,149.
Датировки времени дивергенции – 4,6, .
Датировки эволюции – 4,7.
Интрон – 13,20,21,24,58,140,189,200.
Колинеарность – 23,31,55.
Комплементация – 13,20.
Локус – 14,21,42,43,47,52,53,55,56,57,58,59,60,61,63,64,65,73,74,75,77,79,82,83,84,85,86, 89,90,93,95,96,100,102,106,107,111,115-117,123,125-129,130,130-140,142,150,152,161,167, 173,192
Молекулярно-биологические датировки – 4,9.
Молекулярные часы – 4,6,10,168.
мРНК – 24,26-28,31,35,56,58-60,68,70,194.
Мутации – 13,20,21,23,28,29,48,56,65-69,71,72-75,78-80,91,96,110,112,114,139,140,142,150, 165–167,169,190,192,196,203.
Модификации – 13,20,21,31,33,36,57,79,89.
Овогенез – 16,17.
Период – 9.
Плейотропия – 13,20.
Полигения – 13,20.
Промотор – 13,20,21.
Протеомика – 23,35,36.
Сперматогенез – 16.
Сцепление – 13,19.
Терминатор – 13,20.
Трансляция – 23,26,28.
тРНК – 10,23-28,68,189,194,195,198,201,206.
Частота аллеля – 95,98,99,110-112,116,118,137.
Цистрон – 13,20,21.
Экзон – 13,20,21,23,34,35,57-59,67,195.
Эпистаз – 13,20,121,124,125.
Эпоха – 4,9.
Эра – 4,5,7.

ПРИЛОЖЕНИЕ I

МЕТОДИЧЕСКАЯ ИНСТРУКЦИЯ

ДЛЯ САМООБРАЗОВАНИЯ СТУДЕНТОВ ПО КУРСУ
«МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА»

Ю.Ф. Картавцев

Задача методической инструкции – углубить знания по основополагающим, наиболее важным аспектам спецкурса. Для этого предлагается проработать самостоятельно одну из предлагаемых ниже тем рефератов, с обязательным освещением основных положений темы и использованием прилагаемой специальной литературы.

Темы рефератов:

1. Структура хроматина, подвижность генома и эволюция организации генома.
2. Хромосомы: их морфология и эволюция. Роль хромосом в видообразовании.
3. Популяционно-генетическая структура вида у гидробионтов в сравнении с наземными организмами.
4. Методы и принципы биохимической генетики в применении к популяционной, эволюционной и экологической генетике.
5. Применение молекулярно-генетических маркеров (МГМ) в морской биологии и в биологии в целом. Достижения и перспективы использования МГМ.
6. Генетические аспекты видообразования у животных.
7. Генетика в аквакультуре.
8. Филогенетика.

Главные вопросы для проработки и литература:

ТЕМА 1. СТРУКТУРА ХРОМАТИНА, ПОДВИЖНОСТЬ ГЕНОМА И ЭВОЛЮЦИЯ ОРГАНИЗАЦИИ ГЕНОМА.

Вопросы, которые необходимо проработать в реферате:
1. Компапктизация и декомпактизация хроматина.. 2. Регуляция репликации, транскрипции и трансляции. 3. Структура гена у прокариот и эукариот.. 4. Изменения в организации генома от низших к высшим организмам.

Литература:

1. *W.S. Klug, M.R. Cummings. Essential genetics. 4th Edition, Prentice Hall, Inc., Upper Saddle River, NJ 07458, 2002. (Where to purchase: Kyobo Book Store, Tel.: 051-806-3501, 30000 Won).Daniel L. Hartl, Elizabeth W. Jones. Essential genetics. 3rd Edition, Jones And Bartlett Publishers, Sudburry, Massachusetts, Boston-Toronto-London-Singapore, 2002.*
3. *Б. Льюин Гены. Ред. Г.П. Георгиев. М.: Мир. 1987. 544 с.*
4. *А.В. Жимулев. Общая и молекулярная генетика. Новосибирск: Изд-во Новосибирского. Унив., 2002. 459 с.*

ТЕМА 2. ХРОМОСОМЫ: ИХ МОРФОЛОГИЯ И ЭВОЛЮЦИЯ. РОЛЬ ХРОМОСОМ В ВИДООБРАЗОВАНИИ.

Вопросы, которые необходимо проработать в реферате:

1. Методы окрашивания хромосом. Морфология хромосом, кариограммы. 2. Основные положения о хромосомных числах у рыб и беспозвоночных и эволюции их кариотипа. 3. Кариотипические различия и их роль в установлении репродуктивных изолирующих барьеров. Полиплоидия и генное видообразование. Проблема парапатрического (стасипатрического) видообразования у рыб.

Литература:

1. *С. Оно. Генетические механизмы прогрессивной эволюции. М.: Мир, 1973. 227 с.*
2. *King, Max. Species Evolution: the role of chromosome change. Cambridge University Press, 1993. Published by the Press Syndicate of the University of Cambridge, 1993. The Pitt Building, Trumpington Street, Cambridge CB2 1RP 40 West 20th Street, New York, NY 10011-4211, USA 10 Stamford Road, Oakleigh, Melbourne 3166, Australia.*
3. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1987.*
4. *Ю.П. Алтухов. Генетические процессы в популяциях. М.: Наука, 1989. 2-е изд., 328 с.*

ТЕМА 3. ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКАЯ СТРУКТУРА ВИДА У ГИДРОБИОНТОВ В СРАВНЕНИИ С НАЗЕМНЫМИ ОРГАНИЗМАМИ.

Вопросы, которые необходимо проработать в реферате:

1. Основные внутривидовые группировки у животных. 2. Специфика популяционно-генетической структуры у рыб и беспозвоночных. 3. Модели популяционно-генетической организации вида. Факторы популяционно-генетической динамики и стационарности. Генетические основы рационального использования природных ресурсов.

Литература:

1. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1990.*
2. *Ю.П. Алтухов. Генетические процессы в популяциях. М.: Наука, 1989. 2-е изд., 328 с.*
3. *Алтухов Ю.П., Салменкова Е.А., Омельченко В.Т. Популяционная генетика лососевых рыб. М.: Наука, 1997. 288 с.*
4. *Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.*
5. *Ли К.-Ч. Введение в популяционную генетику. М.: Мир. 1978. 555 с.*
6. *D.L. Hartl, A.G. Clarke. Principles of Population Genetics. Sinauer Assoc. Inc. Publishers, Sunderland, Massachusetts 01375, 1989 (Second edition).*
7. *Жизнеспособность популяций. Природоохранные аспекты. Ред. М.Е. Сулей. М.: Мир, 1989.*

ТЕМА 4. МЕТОДЫ И ПРИНЦИПЫ ИССЛЕДОВАНИЯ ИЗОФЕРМЕНТОВ В ПРИМЕНЕНИИ К ПОПУЛЯЦИОННОЙ, ЭВОЛЮЦИОННОЙ И ЭКОЛОГИЧЕСКОЙ ГЕНЕТИКЕ.

Вопросы, которые необходимо проработать в реферате:

1. Электрофоретические методы исследования белков как маркеров генов. 2. Генетика изоферментов. Принципы генетической интерпретации изменчивости изоферментов и аллоферментов. 3. Измерение гетерозиготности и генетических расстояний и применение этих мер в биологии и экологии.

Литература:

1. *Р. Левонтин. Генетические основы эволюции. М.: Мир, 1978. 351 с.*

2. *Evolution of Genes and Proteins, Eds. M. Nei, R.K. Koehn, Sunderland (Mass.): Sinauer Assoc., 1983.*
3. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1990.*
4. *Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.*
5. *Ю.П. Алтухов. Генетические процессы в популяциях. М.: Наука, 1989. 2-е изд., 328 с.*

ТЕМА 5. ПРИМЕНЕНИЕ МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКИХ МАРКЕРОВ (МГМ) В МОРСКОЙ БИОЛОГИИ И В БИОЛОГИИ В ЦЕЛОМ. ДОСТИЖЕНИЯ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ МГМ.

Вопросы, которые необходимо проработать в реферате:
1. Методы исследования изменчивости фрагментов ДНК как маркеров генов. 2. Полиморфизм ДНК внутри- и между популяциями; гаплотипическое и нуклеотидное разнообразие. 3. Достижения и перспективы использования МГМ.

Литература:
1. *Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.*
2. *Park L.K., Moran P. Developments in molecular genetic techniques in fisheries // Rev. Fish. Biol., 1994. V.4. P.272-279.*
3. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1990.*

ТЕМА 6. ГЕНЕТИЧЕСКИЕ АСПЕКТЫ ВИДООБРАЗОВАНИЯ У ЖИВОТНЫХ.

Вопросы, которые необходимо проработать в реферате:
1. Основные способы видообразования. 2. Гетерозиготность и генетические расстояния. Филогенетические деревья. 3. Генетическая дивергенция таксонов различного ранга. 4. генетические аспекты возникновения новых видов. Разобрать примеры для наиболее изученных групп.

Литература:
1. *King, Max. Species Evolution: the role of chromosome change. Cambridge University Press, 1993. Published by the Press Syndicate of the University of Cambridge, 1993. The Pitt Building, Trumpington Street, Cambridge CB2 1RP 40 West 20th Street, New York, NY 10011-4211, USA 10 Stamford Road, Oakleigh, Melbourne 3166, Australia.*
2. *Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.*
3. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1990.*
4. *Левонтин Р. Генетические основы эволюции. М.: Мир, 1978. 351 с.*
5. *Картавцев Ю.Ф., Свиридов В.В., Ханзава Н., Сазаки Т. Генетическая дивергенция видов дальневосточных красноперок рода Tribolodon (Pisces, Cyprinidae) и близких таксонов // Генетика, 2002. Т.38. №11. С.1518-1531.*

ТЕМА 7. ГЕНЕТИКА В АКВАКУЛЬТУРЕ.

Вопросы, которые необходимо проработать в реферате:

1. Полузамкнутые системы воспроизводства (ПЗСВ) и полноцикличные системы воспроизводства (ПЦСВ). 2. Основные генетические задачи и их специфика при ПЗСВ и ПЦСВ; восстановление и интродукция с генетической точки зрения. 4. Генетические основы разведения рыб при ПЦСВ; определение наследуемости признаков и методы селекции.

Литература:

1. *I.R. Calaprice. Mariculture - Ecological and genetic aspects of production. J. Fish. Research Board of Canada, 1976, v.1, № 4.*
2. *Krueger et al. Genetic aspects of fisheries rehabilitation. Canadian J. Fish. a. Aquat. Sci., 1981, v.38, № 12.*
3. *Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1990.*
4. *Алтухов Ю.П., Салменкова Е.А., Омельченко В.Т. Популяционная генетика лососевых рыб. М.: Наука, 1997. 288 с.*

ТЕМА 8. МОЛЕКУЛЯРНАЯ ФИЛОГЕНЕТИКА.

Вопросы, которые необходимо проработать в реферате:

1. Предмет филогенетики как научной дисциплины. 2. Типы филогенетических деревьев. 3. Метод матриц расстояний. Методы парсимонии и максимального правдоподобия. 4. Молекулярные часы и оценки времени дивергенции genetic Программные средства конструирования деревьев.

Литература:

1. Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.
2. D.L. Hartl, A.G. Clarke. Principles of Population Genetics. Sinauer Assoc. Inc. Publishers, Sunderland, Massachusetts 01375, 1989 (Second edition).
3. Nei M., Kumar S. Molecular evolution and phylogenetics. Oxford Univ. Press, N.Y., 2000. 333p.
4. Swofford D.L., Olsen G.J., Waddel P.J., Hillis D.M. Phylogenetic inference // Hillis D.M., Moritz C., Mable B., Eds. Molecular Systematics. Sunderland, Massachusetts (01375-0407, USA): Sinauer. Assoc. Inc., 1996. pp.407-514.

ПРИЛОЖЕНИЕ II

ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ
«МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА»
Ю.Ф. Картавцев

Раздел 1: Определите концепцию.

1. Основные этапы геологической эволюции Земли.
2. Палеонтологические датировки эволюции.
3. Молекулярно-биологические датировки времени возникновения видов.
4. Строение хромосомы.
5. Мейоз и формирование гамет.
6. Менделевская или классическая генетика. Моно-, ди- и тригибридные скрещивания. Законы Менделя.
7. Тонкая структура гена. Мутации, модификации, хромосомные перестройки.
8. Типы изучаемых признаков, их отличие.
9. Компоненты, критичные для синтеза белков.
10. Процесс трансляции: от ДНК – к полипептиду. Наследственность, белки и функции.
11. Понятие об иммуногенетике.
12. Наследование иммуногенетических признаков.
13. Иммуногенетические датировки эволюции. Датирование эволюции через анализ первичной последовательности аминокислот.
14. Методы выявления изменчивости белков.
15. Генетическая интерпретация белковой изменчивости.
16. Генетика изоферментов..
17. Основные модели видообразования. Генетические аспекты видообразования. 18. Популяционная структура перекрестно размножающихся видов.
19. Феноменологический и научный подходы для описания внутривидовой структуры.
20. Методы и принципы исследования внутривидовой генетической структуры.
21. Приведите пример популяционно-генетического исследования природного вида.
22. Главные популяционно-генетические параметры.
23. Количественные меры генетической дифференциации.
24. Модели популяционно-генетической структуры.
25. Эмпирические оценки популяционно-генетических параметров.
26. Что такое гибрид и гибридизация?
27. Методы обнаружения гибридизации.
28. Генетическая интерпретация гибридизирующих популяций.
29. Эмпирические исследования гибридных зон.
30. Количественная изменчивость и ее законы.
31. Гипотеза множественных факторов.
32. Главные генетические задачи исследования количественных признаков (КП): σ^2_{PH}, σ^2_G, Cv, h^2 и другие переменные.
33. Наследуемость количественных признаков: концепция и практика.
34. Картирование локусов количественных признаков.
35. Концепция генетического гомеостаза.
36. Гетерозис и гетерозиготность.
37. Эмпирические результаты о связи гетерозиготности и количественных признаков.

38. Исследования мутаций. Структура и функция гена.
39. Мутации и замены аминокислот.
40. Типы мутаций на уровне ДНК. Темп мутаций.
41. Полиморфизм первичной последовательности ДНК. Нуклеотидное разнообразие.
42. Оценка полиморфизма ДНК по сайтам рестрикции (ПДРФ).
43. Полиморфизм длины ДНК. Теория и оценки.
44. Дивергенция ДНК между популяциями.
45. Типы филогенетических деревьев.
46. Использование метода матриц расстояний при построении филогенетических деревьев.
47. Методы парсимонии и максимального правдоподобия при построении филогенетических деревьев.
48. Роль популяционно-генетической теории при построении филогенетических деревьев.
49. Возникновение и эволюция генома.
50. Эволюционные изменения размера генома.
51. Формирование новых генов.
52. Повторяющиеся ДНК и мультигенные семейства.

Раздел 2: Проявите оперативное понимание изученного предмета. Выберете правильный ответ.

1. Генетический словарь: дайте определение следующим обозначениям.

(i) AA, Aa, BbCc: все эти три генотипа являются гетерозиготами.
(ii) AA, Aa, BbCc: AA - это гомозигота, но Aa, BbCc являются гетерозиготами.

2. Генетический словарь: дайте определение следующим обозначениям.

(i) P, F1, Fa – являются обозначениями для родителей, первого и возвратного скрещивания.
(ii) P, F1, Fa – являются обозначениями для родителей, первого и анализирующего скрещивания.

3. Объясните термины: ***локус*** и ***ген***.

(i) ***Локус –*** *место на хромосоме, где располагается определенный ген.*
(ii) ***Локус –*** *часть гена, место которого на хромосоме известно.*
(i) ***Ген -*** *не кодирующая последовательность ДНК, являющаяся дискретной единицей наследственности.*
(ii) ***Ген -*** *кодирующая последовательность ДНК, которая определяет элементарную функцию организма, например, кодирование полипептида, и является дискретной единицей наследственности.*

4. Объясните термины: ***аллель*** и ***цистрон.***

(i) ***Аллель*** *– вариант гена с определенными особенностями функции (например, различный заряд белка).*
(ii) ***Аллель*** *– продукт различных генов с определенным сходством функции (например, теплоустойчивость белка).*
(i) ***Цистрон –*** это единица генетической регуляции онтогенеза, которая включает кодирующую последовательность ДНК и регуляторные элементы для переключения молекулярного синтеза.
(ii) ***Цистрон –*** это единица генетической функции, которая включает кодирующую последовательность ДНК и регуляторные элементы для молекулярного синтеза.

5. Что такое ***сцепление***? Отметьте правильный ответ.

(1) Совместное наследование пары или большего числа генов, кодирующих различные признаки. Основополагающие работы, доказавшие факт сцепления были выполнены Морганом, Стертевантом и Бриджесом. [ ]

(2) Совместное наследование пары или большего числа генов, кодирующих различные признаки. Основополагающие работы, доказавшие факт сцепления были выполнены Де-Фризом, Вейсманом и Бэтсоном. [ ]

6. Почему возникает сцепление? [ ] (1) Потому, что гены локализованы в хромосомах – одна группа сцепления представляет одну хромосому; [ ] (2) Потому, что гены локализованы в половых хромосомах – одна группа сцепления представляет один тип половых хромосом.

Определите правильный выбор.

7. Что такое плейотропия и эпистаз, оба являющиеся вариантом взаимодействия неаллельных генов?

Плейотропия – это случай действия гена, на более, чем один признак.

Эпистаз – это межлокусное взаимодействие, когда один ген подавляет действие другого гена.

Обнаруживаются ли эти взаимодействия неаллельных генов при наследовании чешуи карпа? Оба или один из них?

8. (1) ***Вид –*** *это биологическая единица, которая репродуктивно изолирована от других таких единиц и состоящая из одной или большего числа самовоспроизводящихся популяций организмов, размножающихся половым путем и занимающих определенную территорию в природе*. (2) ***Вид –*** *это репродуктивное сообщество популяций (репродуктивно изолированных от других), которые занимают определенную нишу в природе.*

Принадлежат ли эти два определения к **биологической концепции вида**?

9. (1) ***Локальное стадо (локальная популяция)*** *– это система субпопуляций, которая эквивалентна теоретической модели подразделенной популяции и является наименьшей стабильной репродуктивной единицей вида.* (2) Популяция – это репродуктивная единица организмов, которая обладает общим генофондом. Кто: является автором первого и второго определений: Dobzhansky (1951, 1970), Алтухов (1974)?

10. Классификация типов внутривидовой структуры на основе средств популяционной генетики. Правильно ли определить две модели: (1) ***панмиктической популяции*** и (2) ***подразделенной популяции***?

11. ***Частота аллеля, p.*** Эмпирически при кодоминировании или неполном доминировании ***p*** может быть оценено прямым подсчетом генотипов в выборке. Какая формула является правильной для определения ***p*** (1) или (2)?

(1) $p_{A1} = (2D\ A_1A_1 + H\ A_1A_2) / 2n = (D\ A_1A_1 + H/2\ A_1A_2) / n$, ***(2)*** $(p + q)^2 = p^2 + 2pq + q^2$ или в общем виде для множественных аллелей $(\Sigma p_k A_{ij})^2 = \Sigma p^2_k A_iA_i + \Sigma 2p_k(1-p_k)\ A_iA_j$.

12. Закон Харди-Вайнберга может быть выражен следующим образом, как показано ниже. Какое выражение правильно: (1), (2), (3), (1) и (2), все?

(1) $(p + q)^2 = p^2 + 2pq + q^2$.

(2) $(\Sigma p_k A_{ij})^2 = \Sigma p^2_k A_iA_i + \Sigma 2p_k(1-p_k)\ A_iA_j$.

(3) $(p + q + r)^2 = p^2+q^2+r^2+2pq+2pr+2qr$.

<u>*Раздел 3. Проверка умения и навыков работы с темами по курсу. Решите пример (из тем практических занятий).*</u>

1. Хи-квадрат тест независимости частот по качественным признакам. $X^2 = \Sigma[(f - f') / f']$, f' и f – ожидаемые и наблюдаемые численности, d.f. = n - 1, где n – число фенотипических классов. Найдите значение, определите значимость отклонений частот и укажите генотип для **P**.

P		F_2		n	d.f.
C_1C_2	C_2C_2	C_1C_2	C_2C_2		
-	-	105	95	200	1
-	-	64	77	141	1
-	-	63	86	149	1
-	-	81	69	150	1

-	-	313	327	640	1

Примечание. $X^2 = \Sigma\,[(f-f')^2 / f']$; $X^2_{.95}$ (1) = 3.84; Данные по локусу ***GPD**** в скрещиваниях горбуши (Картавцев и др., 1985).

2. Частота аллеля и ее стандартная ошибка. Гетерозиготность.

Используя формулу (1) из Раздела 2, Пример 11, рассчитайте частоту аллеля, стандартную ошибку и ожидаемую гетерозиготность для выборки: AA = 100, Aa = 70, aa = 30.

4. Хи-квадрат тест оценки соответствия наблюдаемых и ожидаемых по формуле Харди-Вайнберга (Х-В) частот генотипов. Ожидаемые и наблюдаемые численности генотипов определены для природных популяций улитки, *Littorina brevicula* по цвету и пятнистости раковины (Картавцев, 1976).

Проверьте ожидания Х-В в одном численном примере ниже.

Выборка	**Фенотип/генотип**			**N**	**p**	**q**	**X^2**
	A	AB	B				
1	38 (33.7)	43 (51.6)	24 (19.7)	105	0.567	0.433	2.92
2	35 (31.6)	43 (49.8)	23 (19.6)	101	0.559	0.441	1.88
3	70 (78.2)	11 (94.3)	20 (28.4)	201	0.624	0.376	6.40
4	41 (43.0)	51 (47.1)	11 (12.9)	103	0.646	0.354	0.70
5	44 (38.9)	60 (70.3)	37 (31.8)	141	0.525	0.475	1.75
6	29 (32.0)	66 (61.8)	27 (29.9)	124	0.508	0.492	1.18
7	39 (46.1)	98 (79.7)	29 (35.2)	160	0.531	0.469	3.82
8	47(59.1)	93 (68.8)	8 (20.0)	148	0.632	0.368	18.19
9	117(137.6)	204(163.0)	28 (48.3)	349	0.628	0.373	21.93

Примечание. $\mathbf{X^2}$ – наблюдаемые значения в хи-квадрат тесте; d.f. = n-k, где n – число генотипических классов, k - . число аллелей. $X^2_{.95}$ (d.f.=1) = 3.84.

5. Проверка навыков работы с программным пакетом SPECSTAT.

На примере исходной таблицы данных для трех локусов тихоокеанской мидии (файл Edspec.dt1) SPECSTAT (Картавцев, Соловьев, 1992) рассчитайте выходные таблицы (Tables). Расскажите, как интерпретировать результаты расчетов для первой таблицы.

6. Хи-квадрат тест и G-тест гетерогенности частот аллелей в программном пакете SPECSTAT.

На примере исходной таблицы данных для трех локусов тихоокеанской мидии (файл Edspec.dt1) SPECSTAT (Картавцев, Соловьев, 1992) рассчитайте выходные таблицы (Tables). Расскажите, как интерпретировать результаты расчетов для второй таблицы.

7. Оценки генетических расстояний и построение дендрограмм невзвешенным парногрупповым методом (НПГМ).

Используя пакет программ BIOSYS и файл-пример в этом пакете получите матрицу расстояний и НПГМ-дендрограмму. Поясните этапы объединения таксонов (групп).

8. Проверка навыков работы с пакетом программ POPULUS.

Продемонстрируйте действие дрейфа генов, используя соответствующие программные модули.

9. Проверка навыков работы с пакетом программ STATISTICA.

Используя программные модули Basic Statistics оцените по t-критерию Стьюдента, значимость отличия средних арифметических одного из признаков раковины моллюска *Nucella heyseana* между парой выборок. Используйте файл: Nu0001.sta.

ПРИЛОЖЕНИЕ III

ОСНОВНАЯ ЛИТЕРАТУРА И ИНТЕРНЕТ-ПОДДЕРЖКА

«МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА»

Ю.Ф. Картавцев

1. Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.
2. Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аммер. М.: Наука, 1990.
3. D.L. Hartl, A.G. Clarke. Principles of Population Genetics. Sinauer Assoc. Inc. Publishers, Sunderland, Massachusetts 01375, 1989 (Second edition).
4. Ли К.-Ч. Введение в популяционную генетику. М.: Мир. 1978. 555 с.
5. Жизнеспособность популяций. Природоохранные аспекты. Ред. М. Сулей, М.: Мир, 1989.
6. King M. Species evolution. The role of chromosome change. Cambridge University Press, N.Y. 1993.
7. Левонтин Р. Генетические основы эволюции. М.: Мир, 1978. 351 с.8. Алтухов Ю.П. Генетические процессы в популяциях. М.: Наука, 1989. 2-е изд., 328 с.
9. POPULUS 1.4. Don Alstad Department of Ecology, Evolution & Behavior University of Minnesota 318 Church St. SE Minneapolis, MN, USA 55455-0302/ Tel: 612-625-0488. E-mail: DNA@UMNACVX.bitnet .
10 . BYOSIS. Swofford D.L., Selander R.B. Biosys-1: A Fortran program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J. Hered., 1981. V.72. No 4. P.281-283.
11. SPECSTAT. Картавцев Ю.Ф., Соловьев А.А. Програмный мини-комплекс SPECSTAT для статистической обработки данных по популяционной генетике // Генетика, 1992. Т.28. № 3. С.194-197.
12. Raymond M., Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity, 1995. V.86: P.248-249.
13. Statistica for Windows. Users Guide. StatSoft Inc., Tulsa OK, East 14th Street. 1994. 1064 pp.
http://www.jbpub.com/genetics http://csep10.phys.utk.edu/genetics_demo
http://www.gene.com.achttp://gened.emc.maricopa.edu/bio/BIO181/BIOBK/BioBookTOC.html
http://www.columbia.edu/cu/biology/courses/c3032/

ПРИЛОЖЕНИЕ IV

ОБЪЯСНЕНИЕ ОСНОВНЫХ ТЕРМИНОВ

«МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ И ПОПУЛЯЦИОННАЯ ГЕНЕТИКА»
Ю.Ф. Картавцев

- **Аддитивный аллель.** Каждый локус, детерминирующий КП, может быть занят, либо аддитивным аллелем, который вносит свой вклад в детерминацию, либо не аддитивным аллелем, не вносящим такого вклада.
- **Аллель** – это вариант гена, с определенными особенностями функции (например, изменен заряд молекулы белка).
- **Аллозимы** – это изоферменты, образуемые в результате аллельной изменчивости одного и того же локуса.
- **Балансовая гипотеза или гипотеза сверхдоминантности** предсказывает, что сами локусы или тесно сцепленные с ними маркеры влияют на взаимосвязь H_O и количественного признака.
- **Белок** – это линейный неветвящийся полимер аминокислот. Первичная структура белка определяется последовательным порядком аминокислот в его молекуле. Различают также еще три других уровня организации белка, связанных с конформацией молекулы, это – вторичная, третичная и четвертичная структуры белка.
- **Вид** – это биологическая группировка, состоящая из одной или нескольких популяций перекрестно размножающихся особей, которые репродуктивно изолированы от других таких группировок и стабильны в природе, занимая определенный ареал.
- **Ген** – это кодирующая последовательность ДНК, которая определяет элементарную функцию организма, такую, например, как синтез фермента.
- **Гаметическое неравновесие.** См. неравновесие по сцеплению.
- **Генетический гомеостаз** может быть определен, как свойство популяции балансировать свой генетический состав и противостоять неожиданным изменениям.
- **Гетерозис** гибридная сила или увеличение «мощности» потомства F_1 по сравнению с обеими родительскими формами.
- **Гипотеза инбредной депрессии** предполагает, что гены сами по себе нейтральны и являются только маркерами геномной гетерозиготности.
- **Гибрид** – это генетическая помесь или потомство от скрещивания между генетически различными организмами. Гибридом можно также считать особь со смешанной родословной.
- **Гибридизация** – это процесс, посредством которого появляются гибриды.
- **Гибридная зона** – это географическое пространство, где встречаются гибриды естественного происхождения между предполагаемыми родительскими формами.
- **Дисперсия.**

Генетическая дисперсия – это компонента изменчивости КП, которая может быть отнесена к генетическому контролю. Она имеет сложную природу.

Средовая дисперсия – это компонента изменчивости КП, которая не может быть отнесена к генетическому контролю.

- **Доминирование** – это один из вариантов взаимодействие аллелей одного генетического локуса.
- **Доминантный** аллель, - это аллель, который всегда проявляется в контролируемом им признаке вне зависимости от генотипа. Наоборот, **рецессивный** аллель проявляется в признаке только в отсутствие первого, то есть, - в гомозиготном состоянии.

- **Закон Харди-Венберга** определяет, что в большой, свободно скрещивающейся популяции перекрестно размножающихся организмов соотношение между частотами гамет и частотами генотипов постоянно и соответствует биномиальным пропорциям. Это соответствие нарушается лишь при воздействии систематических эволюционных факторов – естественного отбора, миграции, дрейфа или других «возмущающих» факторов, как ассортативное скрещивание или инбридинг.
- **Изолокусы** – локусы, белковые продукты которых не отличаются при электрофорезе.
- **Информативный нуклеотидный сайт.** Нуклеотидный сайт является **информативным** для построения дерева только тогда, когда имеются, по меньшей мере, два типа полиморфных нуклеотидов, каждый из которых представлен, по меньшей мере, два раза. Нуклеотиды, которые существуют в ОТЕ уникально, называются **единичными нуклеотидами** (singular nucleotides) и они не являются информативными.
- **Кладограмма –** это укорененное дерево, в котором наоборот филогенетические взаимосвязи основываются на генетических генеалогиях и рассмотрении различных путей эволюции, следуя определенным правилам и с конструированием наилучшего из возможных деревьев.
- **Клина** – это постепенное (градиентное) или резкое изменение частот аллелей в прямом или обратном направлении: вид 1 (популяция 1) → гибриды → вид 2 (популяция 2), поддерживаемое балансом между расселением отбором против гибридов.
- **Клон** – совокупность генетически идентичных организмов.
- **Комплементация** – это вариант межгенного взаимодействия, при котором продукты генов взаимодополняют друг друга и при этом может возникать новый фенотип.
- **Локус –** это место на хромосоме где располагается данный ген. В тоже время термин локус употребляется как синоним термина ген.
- **Множественные аллели** – это более чем два аллеля, встречающихся в популяции по одному и тому же локусу.
- **Множественные молекулярные формы ферментов** (ММФФ) представлены изоферментами (аллозимами), генетически детерминированными молекулярными формами, а также - посттрансляционными модификациями.
- **Мультигенные семейства.** В некоторых случаях средне повторяющаяся ДНК включает функциональные гены, представленные тандемно множественными копиями – это **мультигенные семейства**. Члены мультигенных семейств имеют тенденцию сохранять больше сходства, чем можно ожидать при случайной, дрейфовой дифференциации во времени. Они стремятся эволюционировать совместно (согласованно) поскольку существует механизм, который ведет к единообразию последовательности внутри семейства. Эта тенденция к единообразию известна, как **согласованная эволюция** (concerted evolution) и осуществляется за счет конверсии генов и неравного кроссинговера.
- **Мутация -** скачкообразное изменение гена или его части.

Спонтанные мутации – это мутации, которые случайно возникают в природе, и для них не обнаруживается какого-либо специфического действующего агента. Наоборот, мутации, которые вызываются действием какого-либо специфического агента (излучение, химическое вещество), называются индуцированными.

Биохимические или **мутации пищевых цепей** – это мутации, которые изменяют соответствующие биохимические признаки.

Мутации, которые происходят в синонимических кодонах, называются **синонимическими** или **молчащими мутациями**, тогда как другие - **несинонимическими** или **аминокислотно-замещающими мутациями**.

Мутации жизнеспособности могут быть классифицированы, как **летальные мутации, нейтральные, полезные** и **вредные.**

- **Молчащий полиморфизм** – это полиморфизм ДНК, который не проявляется на уровне

последовательности аминокислот в белке.

- **Наследуемость** – это доля генотипической дисперсии в суммарной дисперсии количественного признака: $h^2_B = \sigma^2_G / \sigma^2_{PH}$.
- **Неравновесие по сцеплению (гаметическое неравновесие)** – это неслучайная ассоциация аллелей различных локусов в гаметах.
- **Нормальные/Иммунные сыворотки** – это сыворотки выработанные организмом без антигена и, наоборот, в присутствие антигена.
- **Нормальное распределение** – это распределение плотности вероятности с двумя параметрами, **математическое ожидание** (μ) и **дисперсия** (σ^2). Оценками этих параметров при практическом исследовании признака в выборочной процедуре являются **средняя** и **стандартное отклонение**.
- **Нуклеотидное разнообразие** – это среднее число нуклеотидных различий на сайт между двумя последовательностями.
- **Нуклеон** – это весь фрагмент ДНК, подвергаемый исследованию. **Гаплотип** или нуклеоморфа – это определенная последовательность нуклеотидов в ДНК, полученная в ходе секвенирования или рестриктазного анализа.
- **Нулевой аллель** – это аллель локуса, который не дает активного (полноценного) продукта и, соответственно, для иммуногенетических признаков, не производит какого либо антигена.
- **Островная модель популяционной структуры –** это одна из моделей подразделенной популяции, которая наиболее популярна у исследователей**.**
- **Полигенный признак** – это признак, который контролируется несколькими генами.
- **Популяция** – это репродуктивная единица организмов, которые имеют общий генофонд. Локальные популяции или локальные стада обычно представлены подразделенными популяциями, которые состоят из демов или субпопуляций.
- **ПДРФ** (Restriction Fragment Length Polymorphism – RFLP), полиморфизм длины рестрикционных фрагментов. ПДРФ – это методика, основанная на использовании ферментов эндонуклеаз или рестриктаз, способных разрезать ДНК в определенных местах (сайтах) узнавания.
- **РАПД** (Randomly Amplified Polymorphic DNAs – RAPD) рэндемично амплифицируемые полиморфные ДНК. РАПД – это молекулярный полиморфизм, идентифицируемый случайно, и используемый для маркирования изменчивости ДНК и картирования генов.
- **Репродуктивная изоляция** между биологическими видами значит отсутствие какого бы то ни было потока генов (в смысле, что F_1, и особенно F_2 или F_b, неплодовиты или не жизнеспособны; низко плодовиты, слабо жизнеспособны).
- **Стандартизированная дисперсия частот аллелей.** Обозначается как *Fst (Gst)***.** Эта мера используется для оценки относительного уровня межпопуляционной или внутрипопуляционной дифференциации. **Gst** и **Fst** используются как средство для организации иерархического анализа популяционной структуры. **Gst** называется также коэффициентом генной дифференциации.
- **Тандемные дупликации** – это дупликации, повторяющиеся одна за другой. Они могут возникать в результате действия трех механизмов: (1) **неравный кроссинговер** – рекомбинация между гомологичными парами хромосом, при которой на одном из гомологов участок ДНК дуплицируется, а на другом утрачивается; (2) **неравный обмен сестринских хроматид** – рекомбинационное событие сходное с неравным кроссинговером, которое возникает между двумя хроматидами одной и той же хромосомы; (3) **ошибки репликации** – при репликации матричная нить молекулы ДНК может сжиматься, вызывая вставку короткого фрагмента во вторую синтезируемую нить, что может поменять ОРС, то есть – дать новый ген.

- **Трансляция** – это биологическая полимеризация аминокислот в полипептидную цепь на матрице мРНК.
- **Фенограмма** – это укорененное дерево, в котором при определении связей организмов принимается в расчет только фенотипическое сходство (различие).
- **Цистрон –** это единица генетической функции, которая включает кодирующий участок молекулы ДНК и регуляторные элементы для синтеза макромолекул живых организмов.
- **Электрофорез** – это процесс разделения макромолекул на основе их электрического заряда; обычно он проводится в геле.
- **Элонгация гена.** Иной способ изменения гена – это его удлинение или **элонгация**. Элонгация гена может происходить за счет частичной или полной его дупликации.
- **Эпистаз** – это вариант межгенного взаимодействия, при котором один ген подавляет действие другого гена.

ПРИЛОЖЕНИЕ V

ПОЛНЫЙ СПИСОК ЛИТЕРАТУРЫ

Абелев Г.И. Основы иммунитета // Соросовский образовательный журнал, 1996. №5. С.4-10.

Айала Ф.Дж. Введение в популяционную и эволюционную генетику. М.: Мир, 1984. 232с.

Алиханян С.И., Акифьев А.П., Чернин Л.С. Общая генетика. М.: Высш. Шк., 1985. 448 с.

Алтухов Ю.П. Генетические процессы в популяциях. М.: Наука, 1983. 1-е изд., 280 с.

Алтухов Ю.П. Генетические процессы в популяциях. М.: Наука, 1989. 2-е изд., 328 с.

Алтухов Ю.П. Вид и видообразование. Соросовский образовательный журнал, 1997. №4. С.2-10.

Алтухов Ю.П., Рычков Ю.Г. Популяционные системы и их структурные компоненты. Генетическая стабильность и изменчивость // Журн. общ. биол., 1970. Т.31. №5. С.507-526.

Алтухов Ю.П., Рычков Ю.Г. Генетический мономорфизм видов и его возможное биологическое значение // Журн. общ. биол., 1972. Т.33. С.281-301.

Алтухов Ю.П., Пудовкин А.И. Салменкова Е.А., Коновалов С.М. Стационарность распределений частот генов лактатдегидрогеназы и фосфоглюкомутазы в системе субпопуляций локального стада рыб (на примере *Oncorhynchus nerka* (Walb.). Сообщение 2. Случайный дрейф генов, миграция и отбор как факторы стационарности // Генетика, 1975. Т.11. №4. С.54-62.

Алтухов Ю.П., Лившиц Г.М. Факторы дифференциации и интеграции генофонда изолированной популяции моллюска *Chondrus bidens* // Докл. АН СССР, 1978. Т.238. No 24. С.955-958.

Алтухов Ю.П., Варнавская Н.В. Адаптивная генетическая структура и ее связь с внутрипопуляционной дифференциацией по полу, возрасту и скорости роста у тихоокеанского лосося - нерки Oncorhynchus nerka (Walb.) // Генетика, 1983. Т.19. №5. С.796-807.

Алтухов Ю.П., Духарев В.А., Животовский Л.А. Отбор против редких электрофоретических вариантов белков и темпы спонтанного мутационного процесса в популяциях // Генетика, 1983. Т.19. №2. С.264-276.

Алтухов Ю.П., Салменкова Е.А. Трансплантации, системная организация и рациональное хозяйственное использование популяций рыб // Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер, М.: Агропромиздат, 1991. С.387-398.

Алтухов Ю.П., Салменкова Е.А., Картавцев Ю.Ф. Связь аллозимной гетерозиготности с жизнеспособностью и скоростью роста горбуши // Цитология и генетика, 1991. Т.25. №1. С.47-51.

Алтухов Ю.П., Салменкова Е.А., Омельченко В.Т. Популяционная генетика лососевых рыб. М.: Наука, 1997. 288 с.

Алтухов Ю.П., Шереметьева В.А. Гетерозиготность генома и долголетие человека // Докл АН России, 2000. Т.371. № 5. С.710-713.

Аттер Ф.М., Эберсолд П., Уинанс Г. Интерпретация генетической изменчивости, выявляемой методом электрофореза // Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер, М.: Агропромиздат, 1991. С.37-67.

Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. М.: Мир, 1982. 488 с.

Безруков В.Ф. Алгебраические закономерности соотношения гетерозиготности со средним значением и дисперсией количественного признака // Генетика, 1989. Т.25. №7. С.1310-1319.

Бейли Н. Математика в биологии и медицине. М.: Мир, 1970. 326 с.

Боркин Л.Я., Даревский И.С. Сетчатое (гибридогенное) видообразование у животных // Журн. общей биол., 1980. Т.41. No 4. С.485-505.

Васильев В.П. Эволюционная кариология рыб. М.: Наука, 1985. 209 с.

Воронцов Н.Н. Синтетическая теория эволюции, ее источники, основные постулаты и нерешенные проблемы // Журнал всес. хим. общ-ва им. Д.И. Менделеева 1980. Т. 25. №3. С.295-314.

Гершензон С.М. Основы современной генетики. 2-е изд., Киев: Наукова думка, 1983. 560с.

Гриценко О.Ф. О популяционной структуре горбуши Oncorhynchus gorbuscha (Walbaum)//Вопр. ихтиол.,1981. Т.21. Вып. 5. С.787-799.

Дуброва Ю.Е., Сусков И.И. Сопряженная изменчивость росто-весовых характеристик и биохимических маркеров генов у новорожденных//Цитология и генетика, 1990. Т.24. №2. С.46-52. 1990.

Дуброва Ю.Е., Карафет Т.М., Сукерник Р.И., Гольцова Т.В. Изучение связи гетерозиготности с параметрами плодовитости у лесных ненцев и планасан // Генетика, 1990. Т.26. №1. С.122-129.

Животовский Л.А., Афанасьев К.И., Рубцова Г.А. Селективные процессы по ферментным локусам у горбуши Oncorhynchus gorbuscha (Walbaum) // Генетика, 1987. Т.23. №10. С.1876-1883.

Животовский Л.А. Интеграция полигенных систем в популяциях. М.: Наука, 1984. 183 с.

Животовский Л.А. Популяционная биометрия. М.: Наука, 1991. 271 с.

Жизнеспособность популяций. Природоохранные аспекты. Ред. М. Сулей, М.: Мир, 1989.

Жимулев А.В.. Общая и молекулярная генетика. Новосибирск: Изд-во Новосибирского. Унив., 2002. 459 с.

Жученко А.А. Экологическая генетика культурных растений (адаптация, рекомбиногенез, агробиоценоз). Кишинев: Штиинца, 1980. 588 с.

Иванков В.Н. Изменчивость и микроэволюция рыб. Владивосток: Изд-во Дальневост. ун-та, 1998. 124 с.

Ильин В.Е., Коновалов С.М., Шевляков А.Г. Коэффициент миграции и пространственная структура тихоокеанских лососей // Биологические основы развития лососевого хозяйства в водоемах СССР. М.: Наука, 1983. С.9-31.

Картавцев Ю.Ф. Полиморфизм окраски раковины брюхоногого моллюска *Littorina brevicula* (Philippi) // В кн.: Биологические исследования залива Восток. Владивосток: ДВНЦ АН СССР, 1976. С.93-98.

Картавцев Ю.Ф. Возможное определение сбалансированности полиморфизма локусов, кодирующих изоферменты // Биохимическая и популяционная генетика рыб. Ред. Кирпичников В.С., Ленинград, Ин-т цитологии АН СССР, 1979. С.36-40.

Картавцев Ю.Ф. Аллозимная гетерозиготность и морфологический гомеостаз у горбуши Oncorhynchus gorbuscha (Pisces: Cyprinidae). Генетика, 1990. Т.26. №8. С.1399-1407.

Картавцев Ю.Ф. Генетическая изменчивость и дифференциация в популяциях морских животных // Диссертация докт. биол. наук., защищена в С.-Петербургском гос. Университете, 1995. 530 с. (Рукопись).

Картавцев Ю.Ф. Анализ сопряженности аллозимной изменчивости с приспособительно значимыми признаками у молоди горбуши Oncorhynchus gorbuscha (Pisces, Salmonidae) при массовых скрещиваниях // Биология моря, 1998. Т.24. № 1. С.34-37.

Картавцев Ю.Ф. Связь между гетерозиготностью и количественным признаком: внутрилокусные взаимодействия и мультилокусное усреднение // Генетика, 2004. Т.40. №10. С.00-00.

Картавцев Ю.Ф., Заславская Н.И. Генетическая структура, факторы интеграции и дифференциации популяций обыкновенной мидии (*Mytilus edulis*) // Генетика, 1982. Т.18. №10. С.1653-1666.

Картавцев Ю.Ф., Мамонтов А.М. Электрофоретическая оценка белковой изменчивости и сходства омуля двух форм сигов (Coregonidae) и хариуса (Thymallidae) озера Байкал//Генетика, 1983. Т.19. №11. С.1895-1902.

Картавцев Ю.Ф., Глубоковский М.К., Черешнев И.А. Генетическая изменчивость и дифференциация симпатричных видов гольцов (Salvelinus, Salmonidae) Чукотки//Генетика, 1983. Т.19. №4. С.584-593.

Картавцев Ю.Ф., Картавцева И.В., Воронцов Н.Н. Сообщ. 6. Степень генетического сходства пяти видов различных родов палеарктических хомяков (Mammalia, Cricetinae) // Генетика, 1984. Т.20. No 6. С.961-967.

Картавцев Ю.Ф., Полякова Н.Е., Карпенко А.И. Наследование, количественная и качественная изменчивость аллозимов малатдегидрогеназы, малик энзима и лактатдегидрогеназы кеты *Oncorhynchus keta* // Генетика, 1985. Т.21. No 5. С.845-853.

Картавцев Ю.Ф., Салменкова Е.А., Рубцова Г.А., Афанасьев К.И. Семейный анализ аллозимной изменчивости и её взаимосвязь с размером тела и выживаемостью потомства у горбуши Oncorhynchus gorbuscha (Walb.) // Генетика, 1990. Т.26. №9. С.1610-1619.

Картавцев Ю.Ф. Изменчивость частот аллелей в пространстве и во времени в популяциях горбуши Oncorhynchus gorbuscha (Walbaum) // Вопросы ихтиол., 1991. Т.31. No 3. С.487-495.

Картавцев Ю.Ф., Соловьев А.А. Программный мини-комплекс SPECSTAT для статистической обработки данных по популяционной генетике // Генетика, 1992. Т.28. № 3. С.194-197.

Картавцев Ю.Ф., Никифоров С.М.. Оценка эффективно- репродуктивного размера популяции тихоокеанской мидии из залива Петра Великого Японского моря // Генетика, 1993. Т.29. №3. С.476-489.

Картавцев Ю.Ф., Ситников А.В., Никифоров С.М., Чичвархин А.В. Аллозимная и морфометрическая изменчивость хищного брюхоногого моллюска Nucella heyseana (MOLLUSCA, GASTROPODA) в условиях загрязнения среды и в норме // Генетика, 1998. Т.34. №10. С.1425-1433.

Картавцев Ю.Ф., Рыбникова И.Г. Генетическое и морфобиологическое исследование популяций тихоокеанской сельди *Clupea pallasi* из Японского и Охотского морей // Генетика, 1999. Т.35. №8. С.1093-1103.

Картавцев Ю.Ф., Рыбникова И.Г., Ситников А.В., Амачаева Е.Ю., Свиньина О.В.. Генетическая и морфометрическая изменчивость брюхоногого моллюска Nucella heyseana (Mollusca, Gastropoda) в средовом оптимуме и пессимуме // Генетика, 2000. Т.36 . №1. С.1340-1347.

Картавцев Ю.Ф., Свиридов В.В., Ханзава Н., Сазаки Т. Генетическая дивергенция видов дальневосточных красноперок рода Tribolodon (Pisces, Cyprinidae) и близких таксонов // Генетика, 2002. Т.38. №11. С.1518-1531.

Кирпичников В.С. Генетические основы селекции рыб. Л.: Наука, 1979. 392 с.

Кирпичников В.С. Генетика и селекция рыб. Л.: Наука, 1987. 520 с.

Колмогоров А.Н. Уклонения от формул Харди при частичной изоляции // Докл. АН СССР, 1935. Т.3. No 3, С.129-132.

Коновалов С.М. Популяционная биология тихоокеанских лососей. Л.: Наука, 1980. 237 с.

Корочкин Л.И., Серов О.Л., Манченко Г.П. Понятие об изоферментах // Генетика изоферментов. Ред. Л.И. Корочкин. М.: Наука, 1977. С.5-17.

Корочкин Л.И. Введение в генетику развития. М.: Наука, 1999. 252 с.

Красилов В.А. Эволюция и биостратиграфия. М.: Наука, 1977. 256 с.

Кусакин О.Г., Дроздов А.Л. Филема органического мира. Ч.1. Пролегомена к построению филемы. Санкт-Петербург: Наука, 1994. 272 с.

Кусакин О.Г., Дроздов А.Л. Филема органического мира. Ч.2. Prokaryota, Eukaryota, Microsporobiontes, Archeomonadobiontes, Euglenobiontes, Mixobiontes, Rhodobiontes, Alveolates, Heterokontes. Санкт-Петербург: Наука, 1998. 359 с.

Ланде Р., Бэрроуклаф Дж.Ф. Эффективная численность популяции, генетическая изменчивость и их использование для управления популяциями // В кн.: Жизнеспособность популяций: природоохранные аспекты. Ред. М. Сулей, М.: Мир, 1989. С.116-157.

Лебедев Н.В. Элементарные популяции рыб. М.: Пищевая пром-сть, 1967. 211 с.

Левонтин Р. Генетические основы эволюции. М.: Мир, 1978. 351 с.

Ли К.-Ч. Введение в популяционную генетику. М.: Мир. 1978. 555 с.

Линдберг Г.У. О влиянии смены фаз трасгрессий и регрессий на эволюцию рыб и рыбообразных. Докл АН СССР, 1948. Т.63. №1. С.93-95.

Льюин Б. Гены. Ред. Г.П. Георгиев. М.: Мир. 1987. 544 с.

Майр Э. Систематика и происхождение видов с точки зрения зоолога. М.: Иностранная литература, 1947. 502 с.

Майр Э. Зоологический вид и эволюция. М.: Мир, 1968. 398 с.

Меттлер Л., Грегг Т. Генетика популяций и эволюция. М.: Мир, 1972. 323 с.

Мина М.В. Микроэволюция рыб: Эволюционные аспекты фенетического разнообразия. М.: Наука, 1986. 207 с.

Нейфах А.А., Лозовская Е.Р. Гены и развитие организма. М.: Наука, 1984. 188 с.

Никольский Г.В. Об относительной стабильности вида и некоторых вопросах таксономии // Зоол. Журн., 1968. Т.47. №6. С.860-874.

Оно С. Генетические механизмы прогрессивной эволюции. М.: Мир, 1973. 227 с.

Оммани Ф. Рыбы. М.: Мир. 1975. 192 с.

Попова Н.А. Иммунология. Новосибирск: Изд-во НГУ, 1999. 96 с.

Популяционная генетика и управление рыбным хозяйством. Ред. Н. Риман, Ф. Аттер. М.: Наука, 1987.

Рао С.Р. Кластер-анализ в применении к изучению перемешивания рас в популяциях людей//Классификация и кластер, Ред. Дж. Ван-Райзин. М.: Мир, 1980. С.148-167.

Ратнер В.А. Хроника великого открытия: идеи и лица // Природа, 1998. №4. С.68-79. №11. С.18-28.

Решетников Ю.С. Экология и систематика сиговых рыб. М.: Наука, 1980. 310 с.

Рухлов Ф.Н., Любаева О.С. Результаты мечения молоди горбуши *Oncorhynchus gorbuscha* (Walb.) на рыбоводных заводах Сахалинской области в 1976 г. // Вопр. ихтиол., 1980. Т.20. №1. С.134-144.

Рычков Ю.Г. Система древних изолятов человека в Северной Азии в свете проблем стабильности и эволюции популяций // Вопросы антропологии, 1973. Вып.44. С.3-22.

Рычков Ю.Г., Шереметьева В.А. Генетика циркумполярных популяций Евразии в связи с проблемой адаптации человека // Ресурсы биосферы. Л.: Наука, 1976. Вып. 3. С.10-41.

Салменкова Е.А. Генетика изоферментов рыб // Успехи соврем. биол., 1973. Т.75. №2. С.217-235.

Салменкова Е.А. Основные результаты и задачи популяционно-генетических исследований лососевых рыб // Генетика в аквакультуре. Ред. В.С. Кирпичников, Л.: Наука, 1989. С.7-29.

Ситников А.В., Картавцев Ю.Ф., Никифоров С.М. Оценка эффективного размера популяции травяной креветки *Pandalus kessleri* темпоральным методом и прямым учетом производителей в сезон размножения // Генетика, 1998. Т.34. №10. С.1425-1433.

Скурихина Л.А., Картавцев Ю.Ф., Панькова М.В., Чичвархин А.Ю. Исследование двустворчатых моллюсков *Mytilus trossulus* и *Mytilus galloprovincialis (Bivalvia, Mytilidae)* с

помощью ПЦР-маркеров в зоне их гибридизации в заливе Петра Великого Японского моря // Генетика, 2001. Т.37. №11. С.1556-1559.

Струнников В.А. Генетические основы гетерозиса и комбинационной способности у тутового шелкопряда // Генетика, 1986. Т.22. №5. С.666-677.

Тимофеев-Ресовский М.В., Яблоков А.В., Глотов Н.В. Очерк учения о популяции. М.: Наука, 1973. 277 с.

Тимофеев-Ресовский М.В., Воронцов Н.Н., Яблоков А.В. Краткий очерк теории эволюции. М.: Наука, 1977. 301 с.

Фокс С., Дозе У. Молекулярная эволюция и возникновение жизни. М.: Мир, 1975. 374 С.

Фролов С.В. Изменчивость и эволюция кариотипа лососевых рыб. Владивосток: Дальнаука, 2000. 229 с.

Хочачка П., Сомеро Дж. Биохимическая адаптация. М.: Мир, 1988. 568 с.

Чичвархин А.Ю., Картавцев Ю.Ф., Кафанов А.И. Генетические взаимосвязи некоторых видов Mytilidae (Mollusca: Bivalvia) северной части тихого океана // Генетика. 2000. Т.36. №9. С.1206-1220.

Яблоков А.В. Популяционная биология. М.: Высш. шк., 1987. 303с.

Янулов К.П. О группировках окуня-клювача (*Sebastes mentella* Travin) в Лабрадорско-Ньюфаундлендском районе // Советские рыбохозяйственные исследования с северо-западной части Атлантического океана. М., 1962. С.285-296.

Air J.M. Sequence relationships among of the hemaglutinin genes of 12 subtypes of influenza A virus // Proc. Nat. Acad. Sci. U.S.A., 1981. V.78. P.7639-7643.

Allendorf F.W. Electromorphs or alleles? // Genetics, 1977. V.87. P.821-822.

Allendorf F.W., Thorgaard G.H. Tetraploidy and the evolution of salmonid fishes // In: Evolutionary Genetics of Fishes. Ed. B. Turner. N.Y.: Plenum Press, 1984. P.1-53.

Allendorf F., Utter F., Ryman N. Genetics and fishery management: Past, present, and future // In: Population genetics and fishery management. N. Ryman, F. Utter, Eds. Washington Sea Grant Programs, Univ. Seattle-London: Washington Press, 1987. P.1-20.

Anderson P.K. Ecological structure and gene flow in small mammals // Symp. Zool. Soc. London, 1970. V.26. P. 299-325.

Arnold M.L., Emms S.K. Paradigm lost: Natural hybridization and evolutionary innovation // In: Endless forms: Species and speciation. D.J. Howard, S.H. Berlocher, Eds., Oxford-N.Y.: Oxford Univ. Press, 1998. P.379-389.

Aquadro C.F., Avise J.C. Genetic divergence between rodent species assessed by using two-dimensional electrophoresis // Proc. Nat. Acad. Sci. U.S.A., 1981. V.78. P.3784-37-88.

Atchley W.R., Rutledge J.J., Cowley D.E. A multivariate statistical analysis of direct and correlated response to selection in the rat // Evolution, 1982. V.36. P.677-698.

Avise J.C. Genetic differentiation during speciation // Molec. Evol., Ed. F.J. Ayala, Massachusetts: Sunderland, 1976. P.106-122.

Avise J.C. Molecular markers, natural history and evolution. New-York: Chapman & Hall, 1994.

Avise J.C., Smith M.H. Biochemical genetics of sunfish. II. Genetic similarity between hybridizing species // Amer. Nat., 1974. V.108. P.458-472.

Avise J.C., Patton J.C., Aquadro C.F. Evolutionary genetics of birds. III. Comparative molecular evolution in New World warblers and rodents // Heredity, 1980. V.71. P.303-310.

Avise J.C., Aquadro C.F. A comparative summary of genetic distances in the vertebrates: Pattern and correlations // Evol. Biol., 1982. V.15. P.151-185.

Avise J.C., Van Den Avyle M.J. Genetic analysis of reproduction of hybrid white bass × stripped bass in the Savannah River // Transactions Amer. Fish. Soc., 1984. V.113. P.563-570.

Avise J.C., Saunders N.C. Hybridization and introgression among species of sunfish (*Lepomis*): Analysis by mitochondrial DNA and allozyme markers // Genetics, 1984. V.108. P.237-250.

Avise J.C., Shapira J.F., Daniel S.W., Aquadro C.F., Lansman R.AQ. Mitochondrial DNA differentiation during the speciation process in *Peromyscus* // Mol. Biol. Evol., 1983. V.1. P.38-56.

Avise J.C., Bermingham E., Kessler L.G., Saunders N.C. Characterization of mitochondrial DNA variability in a hybrid swam between subspecies bluegill sunfish *(Lepomis macrochirus)* // Evolution, 1984. V.38. P.931-941.

Avise J.C., Wollenberg K. Phylogenetics and origin of species // Proc. Nat. Acad. Sci. USA, 1997. V.94. P.7748-7755.

Awramik S.M., Schopf J.W., Walter M.R. Filamentous fossil bacteria from the Arhean Western Australia // Precambr. Res., 1983. V.20. P.357-374.

Ayala F.J., Tracey M.L., Hedgecock D., Richmond R.C. Genetic differentiation during the speciation process in Drosophila // Evolution, 1974. V.28. P.576-592.

Ayala F.J. Scientific hypotheses, natural selection and neutrality theory of protein evolution // The Role of Natural Selection in Human Evolution, Ed. F.M. Salzano. North-Holland Publ. Co., 1975. P.19-42.

Bailey N.T.J. The mathematical approach to biology and medicine. London-N.Y.-Sydney: John Wiley and sons, 1967.

Baker R.L., Chapman A.B., Wardell R.T. Direct response to selection for post weaning weight gain in the rat // Genetics, 1975. V.80. P.171-189.

Bell J.I., Selby M.J., Butter W.J. The highly polymorphic region near the human insulin gene is composed of simple randomly repeated sequences // Nature, 1982. V.295. P.31-35.

Bams R.A. Survival and propensity for homing as affected by presence or absence of locally adapted paternal genes in two transplanted populations of pink salmon (*Oncorhynchus gorbuscha*) //J. Fish. Res. Bd. Can., 1976. V.33. P.2716-2725.

Barghoorn E.S., Schopf J.W. Microorganisms three billion years old from the Precambrian of South Africa // Science, 1966. V.152. P.758-763.

Barton N.H. The dynamics of hybrid zones // Heredity, Lond., 1979. V.43. P.341-359.

Barton N.H. Multilocus clines // Evolution, 1983. V.37. P.454-471.

Barton N.H., Hewitt G.M. Analysis of hybrid zones // Ann. Rev. Ecol. Syst., 1985. V.16. P.113-148.

Beacham T.D., Withler R.E. Heterozygosity and morphological variability of the pink salmon (Oncorhynchus gorbuscha) from the Southern British Columbia and Puget Sound // Can. J. Genet. Cytol., 1985. V.27. P.571-579.

Beaumont A.R., Gosling E.M., Beveridge C.M., Budd M.D., Burnell G.M. Studies of heterozygosity and size in the scallop *Pecten maximus* (L.) // Proc. 19-th European Mar. Biol. Symp., Ed. Gibbs R.E., Cambridge U.K.: Cambr. Univ. Press, 1985. P.443-455.

Begon M. The effective size of natural *Drosophila subobscura* populations // Heredity, 1977. V.38. P.13-18.

Beland K.F., Roberts F.L., Sanders R.L. Evidence of *Salmo salar* x *Salmo trutta* hybridization in a North American river // Can. J. Fish. Aquat. Sci., 1981. V.38. P.552-554.

Benjamin D.C., Berzovsky J.A., East I.J. and 12 other authors. The antigenic structure of proteins: A rappraisal // Ann. Rev. Immun., 1984 V.2. P.67-101.

Bennet J.H. On the theory of random mating // Annals of Eugenics, 1954. V.11. P.311-317.

Benveniste R.E. The contribution of retroviruses to the study of mammalian evolution // In: Molecular Evolutionary Genetics. Ed. R.J. McIntire. N.Y.: Plenum, 1985. P.359-417.

Beverley S.M., Wilson A.C. Molecular evolution in Drosophila and higher Diptera. II. A time scale for the evolution // J. Mol. Evol., 1984. V.21. P.1-13.

Beverley S.M., Wilson A.C. Ancient origin of Hawaian Drosophilinae inferred from protein comparison // Proc. Nat. Acad. Sci. U.S.A., 1985. V.82. P.475304757.

Bhattia K.K. Blake N.M., Kirk R.L. The frequency of private electrophoretic variants in Australian aborigines and indirect estimates of mutation rate // Amer. J. Hum. Genet., 1979. V.31. P. 731-740.

Blake C.C.F. Exons and the evolution of proteins // Intl. Rev. Cytol., 1985. V.93. P.149-185.

Brannon E.L. Orientation mechanisms of homing salmonids // Proc. Salmon a. Trout Migratory Behav. Symp, Ed. E.L. Brannon, E.O. Salo. Seattle: Univ. Wash., 1982. P.219-227.

Brenner S., Jacob F., Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis // Nature, 1961. V.190. P.575-580.

Brererton J. Evolved regulatory mechanisms of population control // Evolution of Living Organisms. Melbourne, 1962. P.81-93.

Bridges C.B. Genes and chromosomes // Teaching Biol., 1936. V.11. P.17-23.

Brown W.M. Evolution of animal mitochondrial DNA // In: Evolution of Genes and Proteins. Eds. M. Nei, R.K. Koehn. Sunderland, Mass.: Sinnauer Assoc., 1983. P.62-88.

Brown W.M., George M., Wilson A.C. Rapid evolution of animal mitochondrial DNA // Proc. Nat. Acad. Sci. U.S.A., 1979. V.76. P.1967-1971.

Brown W.M., Pragner E.M., Wang A., Wilson A.C. Mitochondrial DNA sequences of primates: Tempo and mode of evolution // J. Mol. Evol., 1982. V.18. P.225-239.

Buri P. Gene frequency in small populations of mutant Drosophila // Evolution, 1956. V.10. P.367-402.

Burroker N.E. Gene flow in mainland and insular populations of *Crassostrea* (Mollusca) // Biol. Bull., 1984. V.166. P.550-557.

Busack C.A., Torgaard J.H., Bannon M.P., Gall G.A.E. An elelctrophoretic karyotypic and meristic characterization of the Eagle Lake trout, *Salmo gairdnery aquilarum* // Copeia, 1980. No 3. P. 418-424.

Busack C.A., Gall G.A.E. Introgressive hybridization in populations of Paiute cutthroat trout (*Salmo clarki seleniris*) // Can. J. Fish. Aquat. Sci., 1981. V.38. P.939-951.

Bush G.L. Modes of animal speciation // Ann. Rev. Ecol. Syst., 1975. V.6. P.339-364.

Bush R.M., Smouse P.E., Ledig F.T. The fitness consequences of multiple-locus heterozygosity: the relationships between heterozygosity and growth rate in pitch pine (*Pinus rigida* Mill.) // Evolution, 1987. V.41. P.787-798.

Butlin R. What do hybrid zones in general and the *Chortippus paralellus* zone in particular, tell us about speciation // In: Endless Forms: Species and speciation. Eds. D.J. Howard, S.H. Berlocher. Oxford – N.Y.: Oxford Univ. Press, 1998. P.367-378.

Calaprice I.R. Mariculture - Ecological and genetic aspects of production. J. Fish. Research Board of Canada, 1976, v.1, № 4. P.1068-1087.

Calder N. Timescale: An atlas of the fourth dimension. New York: Viking Press, 1983.

Campton D.E., Utter F.M. Natural hybridization between steelhead trout (*Salmo gairdneri*) and coastal cutthroat trout (*Salmo clarki clarki*) in two Puget Sound streams // Can. J. Fish. Aquat. Sci., 1985. V.42. No1. P.110-119.

Campton D.E., Johnston J.M. Electrophoretic evidence for a genetic admixture of native and non-native trout in Yakima River, Washington // Transaction of Amer. Fish. Soc., 1985. V.114. 782-793.

Campton D.E. Natural hybridization and introgression in fishes. Method of detection and genetic interpretation // In Population Genetics & Fishery Management, Eds. N. Ryman and F. Utter, 1987. P.161-192.

Cann R.L. The evolution of human mitochondrial DNA // Ph.D. Thesis. Berkeley, Univ. California, 1982.

Cann R.L., Brown W.M., Wilson A.C. Evolution of human mitochondrial DNA: A preliminary report // In: Human Genetics. Part A.: The Unfolding Genome. Ed. B. Bonne-Tamir. N.Y.: Alan R. Liss, 1982. P.157-165.

Canon W.B. The wisdom of the body. Norton, N.Y. 1932. 312 p.

Carson H.L. Genetics of speciation at the diploid level // Amer. Nat., 1975. V.109. P.83-92.

Cavalli-Sforza L.L., Barrai J., Edwards A.W.F. Analysis of human evolution under random genetic drift // Cold Spring Harbor Symp. Quant. Biol., 1964. V.29. P.9-20.

Cavalli-Sforza L.L., Edwards A.W.F. Phylogenetic analysis. Models and estimation procedures // Amer J. Hum. Genet., 1967. V.19. P.233-257.

Cavalli-Sforza L.L., Bodmer W.F. The genetics of human populations. San Francisco: W.H. Freeman and Co., 1971. 965 p.

Cavender J.A. Test of phylogenetic hypotheses under generalized model //Math. Biosci., 1981. V.54. P.217-229.

Chakraborty R. Estimation of time of divergence from phylogenetic studies // Can. J. Genet. Cytol., 1977. V.19. P.217-223.

Chakraborty R. The distribution of the number of heterozygous loci in an individual in natural populations // Genetics, 1981. V.98. P.461-466.

Chakraborty R., Fuerst P.A., Nei M. Statistical studies on protein polymorphism in natural populations. II. Gene differentiation between populations // Genetics, 1978. V.88. P.367-390.

Chakraborty R., Haag M., Ryman N., Stahl G. Hierarchical gene diversity analysis and its application to brown trout population data // Hereditas, 1982. V.97.P.17-21.

Chakraborty R., Ryman N. Relationships of mean and variance of genotypic values with heterozygosity per individual in a natural population // Genetics, 1983. V.103. P.149-152.

Chakraborty R., Leimar O. Genetic variation within a subdivided population // In: Population Genetics & Fishery Management, Eds. N. Ryman and F. Utter, 1987. P.89-120.

Chakravarti A., Buetow K.H., Antonarakis S.E., Waber P.G., Boehm C.D., Kazazian H.H. Nonuniform recombination within a human β–globin gene cluster // Amer. J. Hum. Genet., 1984. V.36. P.1239-1258.

Champion A.B., Prager E.M., Wachter D., Wilson A.C. Microcomplement fixation // In: Biochemical and Immunological Taxonomy of Animals. Ed. C.A. Wright. London: Acad. Press, 1974. P.397-416.

Chapman B.S., Vincent K.A., Wilson A.C. Persistence or rapid generation of DNA length polymorphism at the zeta locus of humans // Genetics, 1986. V.112. P.79-92.

Chow S., Fujio Y. Comparison of intraspecific genetic diversity levels among local populations in Decapod Crustacean species; with some references of phenotypic diversity // Nippon Suisan Gakkaishi, 1987. V.53. No 5. P.691-693.

Chromas-pro, 2007. http://www.flu.org.cn/en; http://www.technelysium.com.au/chromas.html.

Cockerham C.C., Weir B.S. Digenic descent measures for finite populations // Genetic Res., 1977. V.30. P.121-127.

Coen E.S., Thoday J.M., Dover G. Rate of turnover of structural variants in the rDNA gene family of Drosophila melanogaster // Nature, 1982. V.295. P.564-568.

Collier G.E., O'Brien S.J. A molecular phylogeny of the Felidae: Immunological distance // Evolution, 1985. V.39. P.473-487.

Comesana A.S., Toro J.E., Innes D.J., Thompson R.J. A molecular approach in the ecology of a mussel (*Mytilus edulis – M. trossulus*) hybrid zone on the east coast of Newfoundland, Canada // Mar. Biol., 1999. V.133. P.213-221.

Coyne J.A. Gel electrophoresis and cryptic protein variation // Isozymes, 1982. V.5. P.1-32.

Coyne J.A. Genetics and speciation // Nature, 1992. V.355. P.511-515.

Crick F.H.S. On protein synthesis // In: The Biological Replication of Macromolecules: 12th Symp. Soc. Expr. Biol. Cambidge: Cambr. Univ. Press, 1958. P.138-163.

Crick F.H.S. Codon-anticodon paring: The wobble hypothesis // J. Mol. Biol., 1966. V.19. P.548-555.

Crow J.F. Breeding structure of populations. 2. Effective population number // Statistics and Mathematics in Biology. Ames: Iowa State College Press, 1954. P.543-556.

Crow J.F., Kimura M. An introduction to population genetics theory. N.Y.: Harper & Row, 1970.

Cracraft J. Species concepts and speciation analysis // Current Ornithology, 1983. V.1. P.159-187.

Danzman R.G., Ferguson M.M., Allendorf F.W. Genetic variability and components of fitness in hatchery strains of rainbow trout // J. Fish Biol., 1989. V.35. Suppl. A. P.313-319.

Darnell D.W., Klotz K.M. Subunit constitution of proteins: A table // Arch. Biochem and Biophys., 1975. V.166. P.651-682.

Darnell J.E. Implication of RNA-RNA splicing in evolution of eukaryotic cells // Science, 1978. V.202. P.1257-1260.

Dayhoff M.O. Ed. Atlas of protein sequence and structure. Silver Springs, Md.: Natl. Biomed. Res. Found, 1969. V.4.

Dayhoff M.O. Atlas of protein sequence and structure. Silver Springs, Md.: Natl. Biomed. Res. Found, 1972. V.5.

Dayhoff M.O. Survey of new data and computer methods of analysis // In: Atlas of Protein Sequence and Structure. Ed. M.O. Dayhoff. Natl. Biomed. Res. Found, Silver Springs, Md., 1978. V.5. Suppl. 2. P.2-8.

Dayhoff M.O., Park C.M. Cytochrome c: Building a phylogenetic tree // In: Atlas of Protein Sequence and Structure. Ed. M.O. Dayhoff. Natl. Biomed. Res. Found, Silver Springs, Md., 1969. V.4. P.7-16.

Dayhoff M.O., Barker W.S. Mechanisms in molecular evolution: Examples // In: Atlas of Protein Sequence and Structure. Ed. M.O. Dayhoff. Natl. Biomed. Res. Found, Silver Springs, Md., 1972. V.5. P.41-45.

Delaney M.E., Bloom S.E. Replication banding patterns in the chromosomes of rainbow trout // Journal of Heredity, 1984. V.75. P.431-434.

De Qeiros K. The general lineage concept of species, species criteria and process of speciation: A conceptual unification and terminological recommendations // In: Endless Forms: Species and speciation. Eds. D.J. Howard, S.H. Berlocher. Oxford – N.Y.: Oxford Univ. Press, 1998. P.57-75.

Dickerson R.E. The structure of cytochrome c and the rates of chromosome evolution // J. Mol. Evol., 1971. V.1. P.26-45.

Diehl W.J., Koehn R.K. Multiple locus heterozygosity, mortality, and growth in a cohort of Mytilus edulis // Marine Biol., 1985. V.88. P.265-271.

Diehl W.J., Gaffney P.M., Koehn R.K. Physiological and genetical aspects of growth in the mussel Mytilus edulis. 1. Oxygen consumption, growth, and weight loss // Physiol. Zool., 1986. V.59. No 2. P.201-211.

Dobzhansky T. Genetics and the origin of species. N.Y.: Columbia Univ. Press, 1937.

Dobzhansky T. Genetics of natural populations. 9. Temporal changes in the composition of populations of Drosophila pseudoobscura // Genetics (US), 1943. V.28. P.162-186.

Dobzhansky T. Genetics and the origin of species. 3rd edn. New York: Columbia Univ. Press., 1951.

Dobzhansky Th. Evolution, Genetics and Man. John Wiley & Sons, Inc. – N.Y., Chapman & Hall, Limited – London, 1955. 398 p.

Dobzhansky Th. Genetics of the evolutionary process. N.Y., L.: Columbia Univ. Press, 1970. 505 p.

Dobzhansky Th. Species of Drosophila // Science, 1972. V.177. P.664-669.

Donaldson L.R., Allen G.H. Return of silver salmon, Oncorhynchus kisutch (Walbaum) to point of release // Trans. Amer. Fish. Soc., 1958. V.87. P.13-22.

Doolittle R.F. The genealogy of some recently evolved vertebrate proteins // Trends in Biochem. Sci., 1985. V.10. P.233-237.

Doolittle W.F. Genes in pieces: Where they ever together? // Nature, 1978. V.272. P.581-582.

Dowling T.E., Moore W.S. Level of reproductive isolation between two cyprinid fishes, *Notropis cornutus* and *N. Crysocephalus* // Copeia, 1984. P.617-628.

Dowling T.E., Moore W.S. Evidence for selection against hybrids in the family Cyprinidae (genus *Notropis*) // Evolution, 1985. V.39. P.152-158.

Drake J.W. Spontaneous mutations accumulating in bacteriophage T4 in the complete absence of DNA replication // Proc. Nat. Acad. Sci. U.S.A., 1966. V.55. P.738-743.

Drozdov A.L., Kussakin O.G. State of the art and problems of megasystematics // In: Problems of Evolution. Eds. A.P. Kryukov, L.V. Yakimenko. Vladivostok: Dalnauka, 2002. P.18-29.

Dubrova Y.E., Salmenkova E.A., Altukhov Y.P., Kartavtsev Y.P., Kalkova E.V., Omeltchenko V.T. Family heterozygosity and progeny body length in pink salmon *Oncorhynchus gorbuscha* (Walbaum) // Heredity, 1995. V.75. P.281-289.

Durand P, Blanc F. Divergence genetique chezun bivalve marine tropical: Punctada margaritifera // Coll. Nat. CNRS "Biologie des Populations", Lyon, 1986. P.323-330.

East E.M. Heterosis // Genetics, 1936. V.21. P.376-397.

Eck R.V., Dayhoff M.O. Atlas of protein sequence and structure Silver Springs, Md.: Natl. Biomed. Res. Found, 1966.

Edgel M.H., Hardies S.C., Brown B., and 7 other authors. Evolution of the mouse β–globin complex locus // In: Evolution of Genes and Proteins. Eds. M. Nei, R.K. Koehn. Sunderland, Mass.: Sinnauer Assoc., 1983. P.1-13.

Ehrlich P.R. The population biology of the butterfly, Euphydryas editha. 2. Structure of the Jasper Ridge colony // Evolution, 1965. V.19. P.327-336.

Elandt-Johnson R.C. Probability models and statistical methods in genetics. N.Y.: Wiley, 1970.

Eldredge N., Cracraft J. Phylogenetic patterns and the evolutionary process. N.Y.: Columbia Univ. Press, 1980.

Elston R.C. The estimation of admixture in racial hybrids // Annals of Human Genetics, 1971. V.35. P.9-17.

Engels W.R. Estimating genetic divergence and genetic variability with restriction endonucleases // Proc. Nat. Acad. Sci. U.S.A., 1981. V.78. P.6329-6333.

Evolution of Genes and Proteins. Eds. M. Nei, R.K. Koehn. Sunderland, Mass.: Sinauer Assoc., 1983.

Faith J.P. Distance methods and the approximation of most parsimonious trees // Syst. Zool., 1985. V.34. P.312-325.

Farris J.S. Estimating phylogenetic trees from distance matrices // American Naturalist. 1972. V.106. P.645-668.

Farris J.S. On the phonetic approach to vertebrate classification. In: Major Patterns of Vertebrate Evolution. Eds. M.K. Hecht, P.C. Goody, B.M. Hecht. New York: Plenum Press, 1977. P.823-850.

Felsenstein J. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters // Syst. Zool., 1973. V.22. P.240-249.

Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach // J. Mol. Evol., 1981. V.17. P.368-376.

Felsenstein J. Numerical methods for inferring evolutionary trees // Quart. Rev. Biol., 1982. V.57. P.379-404.

Felsenstein J. Confidence limits on phylogenies with a molecular clock // Syst. Zool., 1985. V.34. P.152-161.

Felsenstein J. PHYLIP (Phylogeny Inference Package) Version 3.57c. University Washington, 1995.

Felsenstein J. Inferring phylogenies. Sanderland, Massachusetts: Sinauer Associates Inc., 2004. 664 p.

Ferris S.D., Whitt G.S. Evolution of the differential regulation of duplicate genes after polyploidization // J. Mol. Evol., 1979. V.12. No 3. P.267-317.

Ferris S.D., Sage R.D., Huang C.-M., Nielsen J.T., Ritte U., Wilson A.C. Flow of mitochondrial DNA across a species boundary // Proc. National Acad. Sci. U.S.A., 1983. V.80. P.2290-2294.

Fitch W.M. Evidence suggesting a non-random character to nucleotide replacement in naturally occurring mutations // J. Mol. Biol., 1967 V.26. P.499-507.

Fitch W.M., Margoliash E. Construction of phylogenetic trees // Science, 1967a. V.155. P.279-284.

Fitch W.M., Margoliash E. A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case // Biochem. Genet., 1967b. V.1. P.65-71.

Fitch W.M. Toward defining the course of evolution: Minimum change for a species tree topology // Syst. Zool., 1971a. V.20. P.406-416.

Fitch W.M. Evolution of clupeine Z, a probable crossover product // Nature New Biol., 1971b. V.229. P.245-247.

Fitch W.M. On the problem of discovering of most parsimonious tree // Amer. Natur., 1977. V.3. P.223-257.

Fitch W.M., Leiter J.M.E., Li X., Palese P. Positive Darwinian evolution in human influenza A viruses. Proc. National Acad. Sci. USA, 1991. V.88. P.4270-4274.

Fox G.E., Magrum L.J., Balch W.E., Wolf R.S., Woese C.R. Classification of methanogenic bacteria by16S ribosomal RNA characterization // Proc. Nat. Acad. Sci. U.S.A., 1977. V.74. P.4537-4541.

Gaffney P.M. Enzyme heterozygosity, growth rate, and viability in Mytilus edulis: another look // Evolution, 1990. V.44. No 1. P.204-210.

Gaffney P.M., Scott T.M., Koehn R.K., Diehl W.J. Interrelationships of heterozygosity, growth rate and heterozygote deficiencies in the coot clam, Mulinia lateralis // Genetics, 1990. V.124. No 3. P.687-699.

Gardner J.P.H., Skibinski D.O.F. Historical and size-dependent genetic variation in hybrid mussel populations // Heredity, 1988. V.61. P.93-105.

Garton D.W. Relationship between multiple locus heterozygosity and physiological energetics of growth in estuarine gastropod Thais hemastoma // Physiol. Zool., 1984. V.57. No 5. P.530-543.

Garton D.W., Koehn R.K., Scott T.M. Multiple locus heterozygosity and physiological energetics of growth of the coot clam, Mulinia lateralis, from a natural population // Genetics, 1984. V.108. No 2. P.445-455.

Geist V. Mountain sheep, a study in behavior and evolution. Chicago: Univ. Chicago Press, 1971. 383 p.

GenBank, 2006. (http://www.ncbi.nlm.nih.gov)

Gentini M.R., Beaumont A.R. Environmental stress, heterozygosity and growth rate in Mytilus edulis // J. Exper. Mar. Biol. Ecol., 1988. V.120. P.145-153.

Gilbert W. Why genes in pieces? // Nature, 1978. V.271. P.501.

Glass B., Li C.C. The dynamics of racial intermixture – an analysis based on American Negro // Amer. J. Hum. Genet., 1953. V.5. P.1-20.

Glass B. on the unlikelihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States // Amer. J. Hum. Genet., 1955. V.7. P.368-385.

Go M. Correlation of DNA exonic regions with protein structural units in hemoglobin // Nature, 1981. V.291. P.90-92.

Gojobory T., Li W.-H., Graur D. Patterns of nucleotide substitution in pseudogenes and functional genes // J. Mol. Evol., 1982. V.18. P.360-369.

Gojobory T., Yokoyama S. Rates of evolution of retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues // Proc. Nat. Acad. Sci. U.S.A., 1985. V.82. P.4198-4201.

Gold J.R. Cytogenetics // In: Fish Physiology. London – N.Y., 1979. V.8. P.353-405.

Goldman D., Giri P.R., O'Brien S.J. A molecular phylogeny of the hominid primates as indicated by two-dimensional protein electrophoresis // Proc. Nat. Acad. Sci. U.S.A., 1987. V.84. P.3307-3311.

Goodman M., Moore G.W., Matsuda G. Darwinian evolution in the genealogy of hemoglobin // Nature, 1975. V.253. P.603-608.

Goodman M., Romerro-Herrara A.E., Dene H., Czelusniak J., Tashian R.E. Amino acid sequence evidence on the phylogeny of primates and other eutherians // In: Macromolecular Sequences in Systematic and Evolutionary Biology. Ed. M. Goodman. N.Y.: Plenum Press, 1982. P.115-191.

Gosling E.M. Genetic heterozygosity and growth rate in a cohort of Mytilus edulis from the Irish coast // Marine Biol., 1989. V.100. P.211-215.

Grant V. The evolutionary process: A critical review of evolutionary theory. N.Y.: Columbia Univ. Press, 1985.

Green R.H., Sing S.M., Hicks B., McCuaig J.M. An arctic intertidal population of Macoma baltica (Mollusca pelecypoda): genotypic and phenotypic components of population Structure // Can. J. Fish. Aquat. Sci., 1983. V.40. No 9. P.1360-1371.

Greenfield D.W., Greenfield T. Introgressive hybridization between *Gila orcutti* and *Hesperoleucus symmetricus* // (Pisces: Cyprinidae) in the Guyama River Basin, California . I. Meristics, morphometrics, and breeding // Copeia, 1972. P.849-859.

Greenfield D.W., Abdel-Hameed F., Deckert G.D., Finn R.R. Hybridization between *Chrosomus erythrogaster* and *Notropis cornutus* (Pisces: Cyprinidae) // Copeia, 1973. P.54-60.

Greenwood J.J.D. Effective population number in the snail Cepea nemoralis // Evolution, 1975. V.28. P.513-526.

Greenwood J.J.D. Effective population number in Cepea: a modification // Ibid, 1976. V.30. P.186.

Hall B. Phylogenetic trees made easy: A how-to manual for molecular biologists. Synauer Assoc. Inc., Sunderland, Massachusetts, 2001. 179 p.

Haldane J.B.S.1932. The causes of evolution. London: Longmans and Green.

Harrison R.G. Linking evolutionary pattern and process: The relevance of species concepts for the study of speciation // In: Endless Forms: Species and speciation. Eds. D.J. Howard, S.H. Berlocher. Oxford – N.Y.: Oxford Univ. Press, 1998. P.19-31.

Harris H., Hopkinson D.A., Robson E.B. The incidence of rare alleles determining electrophoretic variants: Data on 43 enzyme loci in man // Ann. Hum. Genet., 1974. V.37. P.237-253.

Harris H., Hopkinson D.A. Handbook of enzyme electrophoresis in human genetics. Amsterdam: North Holland Publ. Co., 1976.

Hartl D.L. Genetic dissection of segregation distortion. III. Unequal recovery of reciprocal recombinants // Genetics, 1980. V.96. P.986-996.

Hartl D.L., Clarke A.G. Principles of Population Genetics. Sunderland, Massachusetts 01375: Sinauer Assoc. Inc. Publishers, 1989 (Second edition).

Hartl D.L., Clarke A.G. Principles of Population Genetics. Sunderland, Massachusetts 01375: Sinauer Assoc. Inc. Publishers, 1997 (Third edition).

Hartl D.L., Jones E.W. Essential genetics. 3rd Edition, Boston-Toronto-London-Singapore: Jones and Bartlett Publishers, Sudburry, Massachusetts, 2002.

Hartmann W.L., Raleigh R.V. Tributary homing of sockeye salmon at Brooks and Karluk Lakes Alaska // J. Fish. Res. Bd. Can., 1964. V.21. P.485-504.

Hays J.D., Imbrie J., Shackleton N.J. Variation in earth's orbit: pacemaker of the ice ages // Science, 1976. V.194. P.1121-1132.

Heath D.A., Rawson P.D., Hilbish T.J. PCR-based nuclear markers identify alien blue mussel (Mytilus spp.) genotypes on the west coast of Canada // Can. J. Fish. Aquat. Sci., 1995. V.52. P.2621-2627.

Hedrick P.W., Jain S., Holden L. Multilocus systems in evolution // In: Evolutionary Biol., V.11. Eds. M.K. Hecht, W.C. Steere, B. Wallace. N.Y.: Plenum Press, 1978. P.101-184.

Hedrick P.W. Genetics of populations. Science Books International, Boston, Mass., 1983. 629p.

Hill W.G. Estimation of linkage disequilibrium in randomly mating populations // Heredity, 1974. V.33. P.229-239.

Hillis D.M., Mable B.K., Moritz C. Application of molecular systematics: The state of the field and a look to the future // Hillis D.M., Moritz C., Mable B. (Eds.). Molecular Systematics. Sunderland, Massachusetts (01375-0407, USA): Sinauer. Assoc. Inc., 1996. Ch. 12, P.515-543.

Hines N.O. Fish of rare breeding. Salmon and trout of Donaldson strains. City of Washington: Smitsonian Inst. Press, 1976. 167 p.

Hoffmann H.J., Schopf J.W. Early proterozoic microfossils // In: Earth's Ealiest Biosphere: Its Origin and Evolution. Ed. J.W. Schopf. Princeston, N.J.: Princeston Univ. Press, 1983. P.329-360.

Holland J., Spindler K., Horodyski F., Grabau E., Nichol S., VandePol S. Rapid evolution of RNA genomes // Science, 1982. V.215. P.1577-1585.

Hood L., Campbell J.H., Elgin S.C.R. The organization, Expression, and evolution of antibody genes and other multigene families // An. Rev. Genet., 1975. V.9. P.305-353.

Hori H., Osawa S. Evolutionary change in 5S RNA secondary structure and a phylogenetic tree 54 5S RNA species // Proc. Nat. Acad. Sci. U.S.A., 1979. V.76. P.381-385.

Howard D.J., Warning G.L. Topographic diversity, zone width, and the strength of reproductive isolation in the zone of overlap and hybridization // Evolution, 1991. 45(5). P.1120-1135.

Howard D.J. Unanswered questions and future directions in the study of speciation // In: Endless Forms: Species and Speciation. Eds. Howard D.J., Berlocher S.H. Oxford – N.Y.: Oxford Univ. Press, 1998. P.439-448.

Hubbs C.L., Kuronuma K. Hybridization in nature between two genera of flounders in Japan //Papers of Michigan Acad. Sci., Arts and Letters, 1942.V.27. P.267-306.

Hubbs C.L., Hubbs L.C., Johnson R.E. Hybridization in nature between species of Catostomid fishes // Contribution of the Laboratory of Vertebrate Biology, Univ. of Michgan, 1943. V.22. P.1-77.

Hubbs C., Kuehne R.A., Ball J.C. The fishes of upper Guadelupe River, Texas // Texas Journal of Science, 1953. V.5. P.216-244.

Hubbs C.L. Hybridization between fish species in nature // Syst. Zoology, 1955. V.4. P.1-20.

Hubbs C.L., Lagler K. Fishes of the Great Lake region. Ann Arbor, Michigan: University of Michigan Press, 1970. 213 p.

Hubby J.L., Lewontin R.C. A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura // Genetics, 1966. V.54. P.577-594.

Hulsenbeck JP, Rondquist F. 2001. Mr. BAYES: Bayesian inference of phylogeny. Bioinformatics **17**: 754-755.

Hunt J.A., Hall T.J., Britten R.J. Evolutionary distances in Hawaian *Drosophila* measured by DNA association // J. Mol. Evol., 1981. V.17. P.361-367.

Hunter R.L., Markert C.L. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels // Science, 1957. V.125. P.1294-1295.

Huxley J.S. The evolutionary process // Evolution as a process, Ed. J. Huxley, A.C. Hurdy, E.B. Ford. London: Georg Allen and Unwin, 1954.

Ingram V.M. The hemoglobins in genetics and evolution. N.Y.: Columbia Univ. Press., 1963.

Johnson M.J., Wallace D.C., Farris C.D., Rattazi M.C., Cavalli-Sfortza L.L. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns // J. Mol. Evol., 1983. V.19. P.255-271.

Kahler A.L., Allard R.W., Miller R.D. Mutation rates for enzyme and morphological loci in barley (*Hordeum vulgare* L.) // Genetics, 1984. V.106. P.729-734.

Kaplan N., Langley C.H. A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mapping // J. Mol. Evol., 1979. V.13. P.295-304.

Kartavtsev Y.P., Berenboim B.I., Zgurovsky K.A. Population genetic differentiation in the pink shrimp *Pandalus borealis* Kroyer from the Barents and Bering Seas // J. Shellfish Res., 1991. V.10. No 2. P.333-339.

Kartavtsev Y.P. Allozyme heterozygosity and morphological homeostasis in pink salmon fry *Oncorhynchus gorbuscha* (Pisces: Salmonidae): evidences from the family analysis // J. Fish Biol., 1992. V.40. No 1. P.17-24.

Kartavtsev Y.P., Zgurovsky K.A., Fedina Z.M. 1993. Spatial structure of the pink shrimp *Pandalus borealis* from the Far Eastern seas as it proved by methods of population genetics and morphometrics // J. Shellfish Res., 1993. V.12(1). P.81-87.

Kartavtsev Yu. Ph. Wide-scale genetic differentiation among the pink shrimp Pandalus borealis (Crustacea: Pandalidae) populations // In: Genetics a. Evol. Aquat. Organisms, Ed. A.R. Beaumont, L.: Chapman & Hall, 1994. pp. 41-52.

Kartavtsev Yu. Ph. Allozyme heterozygosity and morphological variability in pink shrimp Pandalus borealis // Isozyme Bull., 1996. V.29. P.29.

Kartavtsev Y.P., Svinyna O.V. Allozyme markers and morphometric variability in gastropod mollusk *Nucella heyseana* (Mollusca, Gastropoda) and their association with environmental change // Korean J. Genetics, 2003. V.25 No 4. P.1-12.

Kartavtsev Y. Association between heterozygosity level and quantitative trait score: Intra-locus interaction and multiple loci averaging // In: International Conference "Marine Environment: Nature, Communication, and Business", Vladivostok, 2003. Busan: Publ. Korea Maritime Univ., 2003. P.17 (Abstract).

Kartavtsev Y.Ph., Zaslavskaya N.I., Svinyna O.V., Kijima A. Allozyme and morphometric variability in the whelk *Nucella heyseana* (Mollusca, Gastropoda) from Russian and Japanese waters: evidence for a single species under different names // J. Molluscan Studies, 2004 (Submitted).

Kato I., Kohr W.J., Laskowski M.J. Evolution of aves ovomucoids // In: Proc. 11th Feder. Eur. Biol. Sci.. Eds. S. Magnuson, M. Ottesen, B. Taltman, K. Dano, H. Neurath. N.Y.: Pergamon Press, 1978. P.197-206.

Kerster H.W. Neighborhood size in the rusty lizard, *Sceloporus olivacens* // Evolution, 1964. V.18. P.445-457.

Kim Y.J., Gorman G.C., Papenfuss T., Roychoudhury A.K. Genetic relationships and genic variation in the amphisbaenian genus Bipes // Copeia, 1976. V.1. P.120-124.

Kimura M. "Stepping-stone" model of population // Ann. Rep. Nat. Inst. Genet. Mishima, 1953. V.3. P.63-65.

Kimura M. A stochastic model concerning the maintenance of genetic variability in quantitative characters // Proc. Nat. Acad. Sci. U.S.A., 1965. V.54. P.731-736.

Kimura M. Genetic variability maintained in a finite population due to mutational production of neutral or nearly neutral isoalleles // Genet. Res., 1968. V.11. P.247-269.

Kimura M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations // Genetics, 1969. V.61. P.893-903.

Kimura M. Theoretical foundation of population genetics at the molecular level // Theor. Popul. Biol., 1971. V.2. P.174-208.

Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge Univ. Press. 1983a.

Kimura M. The neutral theory of molecular evolution In: // Evolution of Genes and Proteins, Eds. M. Nei, R.K. Koehn, Sunderland, Mass.: Sinauer Assoc., 1983b. P.208-233.

Kimura M. Rare variant alleles in the light of neutral theory // Mol. Biol. Evol., 1983c. V.1. P.84-93.

Kimura M., Crow J.F. The measurement of effective population number // Evolution, 1963. V.17. P.279-288.

Kimura M., Weiss G.H. The stepping-stone model of population structure and the decrease of genetic correlation with distance // Genetics, 1964. V.49. P.561-576.

Kimura M., Ohta T. Theoretical aspects of population genetics. Princeton, New Jersey: Princeton Univ. Press., 1971. 219p.

Kimura M., Ohta T. Eukaryotes-prokaryotes divergence estimated by 5S ribosomal RNA sequences // Nature New Biol., 1973. V.243. P.199-200.

King J.L., Jukes T.H. Non-Darwian evolution // Science, 1969. V.164. P.788-798.

King J., Ohta T. Polyallelic mutation equilibria // Genetics, 1975. V.79. P.681-691.

King M. Species Evolution: the role of chromosome change. Cambridge: Cambridge University Press, 1993. 336 p.

Klug W.S., Cummings M.R. Essential genetics. 4th Edition, Upper Saddle River, NJ 07458: Prentice Hall, Inc., 2002.

Koehn R.K., Eanes W.F. Subunit size and genetic variation of enzyme in natural populations of Drosophila // Theor. Popul. Biol., 1977. V.11. P.330-341.

Koehn R.K., Gaffney P.M. Genetic heterozygosity and growth rate in Mytilus edulis // Marine Biol., 1984. V.82. P.1-7.

Koehn R.K., Diehl W.J., Scott T.M. The differential contribution by individual enzymes of glycolysis and protein catabolism to the relationship between heterozygosity and growth rate in the coot clam, Mulinia lateralis // Genetics, 1988. V.118. P.121-130.

Kondo S. Evolutionary considerations of DNA repair and mutagenesis. In: Molecular Evolution and Polymorphism. Ed M. Kimura Mishima, Japan: National Institute of Genetics, 1977. P.313-331.

Kondrashev F.A. Rogozin I.B, Wolf Y.I., Koonin E.V. 2002. Selection in the evolution of gene duplications // Genome Biol. 3, Res. 0008.1-0008.9.

Kreitman M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster* // Nature, 1983.V.304. P.412-417.

Krueger C.C., Gharrett A.J., Dehring J.R., Allendorf F.W. Genetic aspects of fisheries rehabilitation. Canadian J. Fish. a. Aquat. Sci., 1981, v.38, № 12.

Kubitschek H.E. Introduction to research with continuous cultures. Engellwood Cliffs, N.J.: Prentice-Hall, 1970.

Kuby J. Immunology. 2nd. ed. N.Y.: W.H. Freeman, 1994. 660 p.

Kumar S, Tamura K, Nei M. 1993. MEGA: Molecular evolutionary genetics analysis (with a 130-page printed manual). University Park: Pennsilvania State Univ. MEGA3. Web-base β-version, 2005 update.

Lamotte M. Recherches sur la structure genetique des populations naturelles de Cepea nemoralis (L.) // Bull. Biol. France et Belg., 1951. Supl. 35. P.1-239.

Lande R. The maintenance of genetic variability by mutation in a polygenic character with linked loci // Genet. Res., 1975. V.26. P.221-234.

Lande R. The influence of the mating system on the maintenance of genetic variability in polygenic characters // Genetics, 1977. V.86. P.485-498.

Lande R. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry // Evolution, 1979. V.33. P.402-416.

Lande R. The genetic covariance between characters maintained by pleiotropic mutations // Genetics, 1980. V.94. P.203-215.

Lane S., McGregor A.J., Taylor S.G., Gharrett A.J. Genetic marking of an Alaskan pink salmon population, with an evaluation of the mark and the marking process // Amer. Fish. Soc. Symp., 1990. V.7. P.395-406.

Langley, C.H., Fitch W.M.. An examination of the constancy of the rate of molecular evolution. J. Mol. Evol., 1974. V.3. P.161-177.

Langley, C.H., Montgomery E.A., Qattlebaum W.F. Restriction map variation in the Adh region of *Drosophila* // Proc. Nat. Acad. Sci. U.S.A., 1982. V.79. P.5631-5635.

Laurie-Ahlberg C.C., Maroni G., Buley G.C., Lucchesi J.C., Weir B.S. Qauntative genetic variation of enzyme activities in natural populations of Drosophila melanogaster // Proc. Nat. Acad. Sci. U.S.A., 1982. V.77. P.1073-1077.

Leary R.F., Allendorf F.W., Knudsen K.L. Developmental stability and enzyme heterozygosity in rainbow trout // Nature, 1983. V.301. P.71-72.

Leary R.F., Allendorf F.W., Knudsen K.L., Thorgaard G.H. Heterozygosity and developmental stability in gynogenetic diploid and triploid rainbow trout // Heredity, 1985. V.54. P.219-225.

Leary R.F., Allendorf F.W., Knudsen K.L. Differences in inbreeding coefficients do not explain the association between heterozygosity at allozyme loci and developmental stability in rainbow trout // Evolution, 1987. V.41. No 6. P.1413-1415.

Lerner I.M. The genetic basis of selection. N.Y.: John Wiley and Son, 1958.

Lerner M. Genetic homeostasis. N.Y.: Dover Publ. Inc., 1970. 134 p.

Leslie J.F. Linkage analysis of seventeen loci in poecilid fish (genus Poecilopsis) // J. Hered., 1982. V.73(1). P.19-23.

Levinton J.S., Koehn R.K. Population genetics of mussels // In: Marine Mussels: Their Ecology and Physiology, Ed. B.L. Bayne. London and New York: Cambridge Univ. Press, 1976. P.357-384.

Lewontin R.C. The interaction of selection and linkage. I. General considerations. Heterotic models // Genetics, 1964. V.49. P.49-67.

Lewontin R.C., Kojima K. The evolutionary dynamics of complex polymorphism // Evolution, 1960. V.14. P.458-472.

Lewontin R.C. The Genetic Basis of Evolutionary Change. New York and London: Columbia University Press, 1974.

Li C.C. Population genetics. Chicago: Univ. Chicago Press, 1955. 346 p.

Li W.-H. Evolution of duplicate genes and pseudogenes // In: Evolution of Genes and Proteins. Eds. M. Nei, R. Koehn. Sunderland, Mass: Sinauer Assoc., 1983. P.14-37.

Li W.-H. Evolutionary change of restriction cleavage sites and phylogenetic inference // Genetics, 1986. V.113. P.187-213.

Li W.-H., Gojobory T., Nei M. Pseudogenes as a paradigm of neutral evolution // Nature, 1981. V.292. P.237-239.

Li W.-H., Wu C.-I., Luo C.-C. Nonrandomness of pint mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications // J. Mol. Evol., 1984. V.21. P.58-71.

Livshits G., Kobiliansky E. Lerner's concept of developmental homeostasis and the problem of heterozygosity level in natural populations // Heredity, 1985. V.55. P.341-353.

Lynch J.D. The gauge of speciation: on the frequencies of modes of speciation // In: D. Otte, J. Endler, Eds. Speciation and its consequences. Sunderland, Mass.: Sinauer, 1989. 527-553.

Lynch J.D., Conery J.S. The evolutionary fate and consequences of duplicate genes // Science, 2000. V.290. P.1151-1155.

Maeda N., Smithies O. The evolution of multigene families: Human haptoglobin genes // Ann. Rev. Genet., 1986. V.20. P.81-108.

Magni G.E. Spontaneous mutations. In: Proc. 12th Int. Congr. Genet., Tokyo, 1969. V.3. P.247-259.

Manchenko G.P. Handbook of Detection of Enzymes on Electrophoretic Gels., Publisher: CRC Pr I Llc. 2002.

Margoliash E., Smith E.L. Structural and functional aspects of cytochrome c in relation to evolution // In: Evolving Genes and Proteins. Eds. V. Bryson, H.J. Vogel. N.Y.: Academic Press, 1965. P.221-242.

Marshall D.R., Brown A.H.D. The charge-state model of protein polymorphism in natural populations // J. Mol. Evol., 1975. V.6. P.149-163.

Marr J.C. The problem of defining and recognizing subpopulations of fishes // U.S. Fish and Wildlife Service, Spec. Sci. Rep., Fish, 1957. V. 208. P.1-6.

Mather K., Harrison B.J. The manifold effect of selection // Heredity, 1949. V.3. P.1-52, 131-162.

Maxson L.R., Wilson A.C. Convergent morphological evolution detected by studying proteins of tree frog in the *Hyla eximia* species group // Science, 1974. V.185. P.66-68.

Maxson L.R., Wilson A.C. Albumin evolution and organismal evolution in tree frogs (Hylidae) // Syst. Zool., 1975. V.24. P.1-15.

Maxson L.R., Sarich V.M., Wilson A.C. Continental drift and the use of albumin as an evolutionary clock // Nature, 1975. V.255. P.397-400.

Maxson L.R. Molecular probes of phylogeny and biogeography in toads of the widespread genus Bufo // Mol. Biol. Evol., 1984. V.1. P.345-356.

Mayr E. Process of speciation in animals. In: Mechanisms of speciation. Ed. C. Barigozzi. N.Y.: Alan R. Liss, 1982. P.1-20.

McCune A.R., Lovejoy N.R. The relative rate of sympatric and allopatric speciation in fishes: Tests using DNA sequence divergence between sister species and among clades // In: Endless Forms: Species and speciation. Eds. D.J. Howard, S.H. Berlocher. Oxford – N.Y.: Oxford Univ. Press, 1998. P.172-185.

McLaughlin P.J., Dayhoff M.O. Eukaryotes versus prokaryotes: An estimate of evolutionary distance // Science, 1970. V.168. P.1469-1470.

McLaughlin P.J., Dayhoff M.O. Evolution of species and proteins: A time scale // In: Atlas of protein Sequence and Structure. Ed. M.O. Dayhoff. Silver Springs, Md.: Natl. Biomed. Res. Found, 1972. V.5. P.47-66.

McLellan T., Ames F.-L. Nikaido K. Genetic variation in proteins: Comparison of one-dimensional and two-dimensional electrophoresis // Genetics, 1983. V.104. P.381-390.

McLellan T. Molecular charge and electrophoretic mobility in cetacean hemoglobins of known sequence // Biochem. Genet., 1984. V.22. P.181-200.

Mitton J.B., Piers B.A. The distribution of individual heterozygosity in natural populations // Genetics, 1980. V.95. No 4. P.1043-1054.

Mitton J.B., Lewis W.M. Relationships between genetic variability and life-history features of bony fishes // Evolution, 1989. V.43. No 8. P.1712-1723.

Moore W.S. An evolution of narrow hybrid zone in vertebrates // Quarterly Rev. Biol., 1977. V.52. P.263-277.

Mukai T., Cockerham C.C. Spontaneous mutation rates at enzyme loci in Drosophila melanogaster // Proc. Nat. Acad. Sci. U.S.A., 1977. V.74. P.2514-2517.

Muller H.J. Advances in radiation mutagenesis through studies on Drosophila // In: Progress in Nuclear Energy. N.Y.: Pergamon Press, 1959. Ser.6. V.2. P.146-160.

Nagylaki T. The evolution of multigene families under interchromosomal gene conversion // Genetics, 1984. V.106. P.529-548.

Neave F. The origin and speciation of Oncorhynchus // Trans. Roy Soc. Canada., 1958. V.52. P.25-49.

Neel J.V. “Private” genetic variants and the frequency of mutation among South American Indians // Proc. Nat. Acad. Sci. U.S.A., 1973. V.70. P.3311-3315.

Neel J.V., Rothman E.D. Indirect estimates of mutation rate in tribal American Indians // Ibid, 1978. V.75. P.5585-5588.

Neel J.V., Mohrenweiser H.W., Meisler M.H. Rate of spontaneous mutation at human loci encoding protein structure // Proc. Nat. Acad. Sci. U.S.A., 1980. V.77. P.6037-6041.

Neel J.V., Satoh C., Goriki K., Fujita M., Takahashi N., Asakawa J.-T., Hazama R. The rate with which spontaneous mutation alerts the electrophoretic mobility of polypeptides // Proc. Nat. Acad. Sci. U.S.A., 1986. V.83. P.389-393.

Neff N.A., Smith G.R. Multivariate analysis of hybrid fishes // Syst. Zool., 1979. V.28. P.176-196.

Nei M. Genetic distances between populations // Amer. Nat., 1972. V.106. No 949. P.283-292.

Nei M. Analysis of gene diversity in subdivided populations // Proc. Nat. Acad. Sci. USA, 1973. V.70. P.3321-3323.

Nei M. Molecular population genetics and evolution. Amsterdam-Oxford-N.Y.: North Holland Publ. Co., Amer Elsevier Publ. Co., Inc., 1975. 288 p.

Nei M. Estimation of mutation rate from rare protein variants // Amer. J. Hum. Genet., 1977a. V.29. P.225-232.

Nei M. F-statistics and analysis of gene diversity in subdivided populations // Ann. Hum. Genet. L., 1977b. V.41. P.225-233.

Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals // Genetics, 1978. V.89. P.583-590.

Nei M. Genetic polymorphism and the role of mutations in evolution // In: Evolution of Genes and Proteins. Eds. M. Nei, R. Koehn. Sunderland, Massachusetts: Sinauer Assoc., 1983. P.165-190.

Nei M. Human evolution at molecular level // In: Population Genetics and Molecular Evolution. Eds. T. Ohta, K. Aoki. Tokyo: Japan Sci. Soc. Press, 1985. P.41-64.

Nei M., Chakraborty R. Genetic distance and electrophoretic identity of proteins between taxa // J. Mol. Evol., 1973. V.2. P.323-328.

Nei M., Tajima F. DNA polymorphism detectable by restriction endonucleases // Genetics, 1981. V.97. P.145-163.

Nei M., Tajima F. Evolutionary change of restriction cleavage sites and phylogenetic inference for man and apes // Mol. Biol. Evol., 1985. V.2. P.189-205.

Nei M., Roychoudhury A.K. Genetic relationship and evolution of human races // Evol. Biol., 1982. V.14. P.1-59.

Nei M., Tajima F., Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data // J. Mol. Evol., 1983. V.19. P.153-170.

Nei M., Stephens J.C., Saitou N. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes // Mol. Biol. Evol., 1985. V.2. P.66-85.

Nei M. Molecular evolutionary genetics. N.Y.: Columbia Univ. Press, 1987. 512 p.

Nei M., Kumar S. Molecular evolution and phylogenetics. N.Y.: Oxford Univ. Press, 2000. 333p.

Nelson J.S. Hybridization between two cyprinid fishes, *Hybopes plumbea* and *Rhinichthys cataractae*, in Alberta // Can. J. Zool., 1966. V.44. P.963-968.

Nelson J.S. Occurrence of hybrids between longnose sucker (*Catastomus catastomus*) and white sucker (*C. commersoni*) in upper Canananaskis Resorvoir, Alberta // J. Fish. Res. Board. Can., 1973. V.30. P.557-560.

Nevo E., Cleave H. Genetic differentiation during speciation // Nature, 1978. V.275. P.125.

Nevo E., Beiles A., Ben-Shlomo R. The evolutionary significance of genetic diversity: ecological, demographic and life history correlates // Evolutionary Dynamics of Genetic Diversity. Ed. G.S. Mani. Lecture notes in biomathematics, 1984. V.53. P.4-213.

Novick A., Szilard L. Experiments with hemostat on spontaneous mutation of bacteria // Proc. Nat. Acad. Sci. U.S.A, 1950. V.36. P.708-719.

Ohnishi S., Kawanishi M., Watanabe T.K. Biochemical phylogenies of Drosophila: Protein differences detected by two-dimensional electrophoresis // Genetica, 1983. V.61. P.55-63.

Ohno S. So much "junk" DNA in our genome // Brookbaven Symp. Biol., 1972. V.23. P.366-370.

Ohno S. Original domain for the serum albumin family arose from repeated sequences //Proc. Nat. Acad. Sci. USA, 1981. V.78. P.7657-7661.

Ohno S. Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes // J. Mol. Evol., 1984. V.20. P.313-321.

Ohta T. Linkage disequilibrium due to random genetic drift in finite subdivided population //Proc. Nat. Acad. Sci. USA, 1982. V.79. P.1940-1944.

On-Line Biology Book. 2002. Text: M.J. Farabee. http://gened.emc.maricopa.edu/bio/BIO181/BIOBK/BioBookTOC.html

Park L.K., Moran P. Developments in molecular genetic techniques in fisheries // Rev. Fish. Biol., 1994. V.4. P.272-279.

Paterson H.E.H More evidence against speciation by reinforcement // South African J. Sci., 1978. V.74. P.369-371.

Paterson H.E.H The recognition concept of species // In: Species and Speciation. Ed. E.S. Vrba. Pretoria: Transvaal Museum Monograf, 1985. P.21-29.

Penney R.W., Hart M.J. Distribution, genetic structure, and morphometry of Mytilus edulis and M. Trossulus within a mixed species zone. J. Shellfish Res., 1999. V.18(2). P.367-374.

Pereira V. Gene history repeat itself // Heredity, 2004. V.93. P.3-4.

Phillipp D.P., Childers W.F., Whitt G.S. A biochemical genetic evaluation of the northern and Florida subspecies of largemouth bass // Transactions of Amer. Fisheries Soc., 1983. V.112. P.1-20.

Phillips R.B., K.D. Zajicek. Q band chromosomal banding polymorphisms in lake trout (*Salvelinus namaycush*) // Genetics, 1982. V.101. P.227-234.

Phillips R.B., K.D. Zajicek, Utter F.M. Q band chromosome polymorphism in chinook salmon (Oncorhynchus tshawytscha) // Copeia, 1985. P.273-278.

Pilbeam D. The descent of hominids and hominids // Sci. Amer., 1984. V.250(3). P.84-96.

POPULUS 1.4. Don Alstad Department of Ecology, Evolution & Behavior University of Minnesota 318 Church St. SE Minneapolis, MN, USA 55455-0302/ Tel: 612-625-0488. E-mail: DNA@UMNACVX.bitnet_.

Post T.J., Uzzel T. The relationship of Rana sylvatica and the monophyly of Rana boylei group // Syst. Zool., 1981. V.30. P.170-180.

Powell J.R. Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: Evidence from *Drosophila* // Proc. Nat. Acad. Sci USA, 1983. V.80. P.492-495.

Prager E.M., Wilson A.C. The dependence of immunological cross-reactivity upon sequence resemblance among lysozymes // J. Biol. Chem., 1971. V.246. P.5978-5989.

Prager E.M., Brush A.H., Nolan R.A., Wilson A.C. Slow evolution of transferrin and albumin in birds according to micro-complement fixation analysis // J. Mol. Evol., 1974. V.3. P.243-262.

Radding C.M. Strand transfer in homologous genetic recombination // Ann. Rev. Genet., 1982. V.16. P.405-437.

Ramshaw J.A., Coyne J.A., Lewontin R.C. The sensitivity of gel electrophoresis as a detector of genetic variation // Genetics, 1979. V.93. P.1019-1037.

Rasmussen D.I. Blood group polymorphism and inbreeding in natural populations of the deer mouse, *Peromiscus maniculatus* // Evolution, 1964. V.18. P.219-229.

Raymond M., Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism // .J Heredity, 1995. V.86: P.248-249.

Remington C.L. Suture-zones of hybrid interaction between recently joined biotas // In: Evolutionary Biology. Eds. T. Dobhansky, M.K. Hecht, W.C. Steere. N.Y.: Appleton-Century Crofts, 1968. V.2. P.321-428.

Roberts D.F. The dynamics of racial intermixture in the Americasn Negro: Some anthropological considerations // Amer. J. Human Gen., 1955. V.10. P.117-144.

Ronquist F, Huelsenbeck JP. Mr.BAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 2003. V.19. P.1572-1574.

Rodriguez-Trelles F., Tarrio R., Ayala F. Convergent neofuntionalization by positive Darwinian selection after ancient recurrent duplication for xantine dehydrogenase gene // Proc. Natl. Acad. Sci. U.S.A., 2003. V.100.13413-13417.

Rogers J.S. Measures of genetic similarity and genetic distance // Univ. Texas Publ., 1972. V.7213. P.145-153.

Ross M.R., Cavender T.M. Morphological analysis of four experimental intergeneric cyprinid hybrid crosses // Copeia, 1981. P.377-387.

Rawson P.D., Secor C.L., Hilbish T.J. The effect of natural hybridization on the regulation of doubly uniparental MtDNA inheritance in blue mussels (Mytilus spp.) // Genetics, 1996. V.144. P.241-248.

Rawson P.D., Hilbish T.J. Asymmetric introgression of mitochondrial DNA among European populations of Blue mussels (Mytilus spp.) // Evolution, 1998. V.52. No 1. P.100-108.

Russell P.J. Genetics. 5th ed. Menlo Park, California: Addison Wesley Longman Inc., 1998.

Ryman N.F., Allendorf F.W., Stahl G. Reproductive isolation with little genetic divergence in sympatric populations of brown trout (Salmo trutta) // Genetics, 1979. V.92. P.247-262.

Saitou, N., M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees // Mol. Biol. Evol. 1987. V. 4. P. 1406-1425.

Saldana P.H. Gene flow from white into Negro population in Brazil // Amer. J. Hum. Genet., 1957. V.9. P.299-309.

Sanders B., Wright J. Immunogenetic studies in two trout species of the genus Salmo // Annals of the New York Acad. Sci., 1962. V. 97. P.116-130.

Sarich V.M., Wilson A.C. Qauntative immunochemistry and the evolution of primate albumins: Micro-complement fixation // Science, 1966. V.154. P.1563-1566.

Sarich V.M., Wilson A.C. Immunological time scale for hominid evolution // Science, 1967. V.158. P.120-1203.

Schopf J.M., Walter M.R. Archean microfossils: New evidence of ancient microbes In: Earth's Earliest Biosphere: Its Origin and Evolution. Ed. M.H. Smith, T. Toule. Athens: Univ. of Georgia Press, 1983. P.60-99.

Schopf J.M., Hayes J.M., Walter M.R. Evolution of Earth's earliest ecosystems: Recent progress and unsolved problems // In: Earth's Earliest Biosphere: Its Origin and Evolution. Ed. J.W. Schopf. Princeston, N.J.: Princeston Univ. Press, 1983. P.361-384.

Schwartz F.J. World literature on to fish hybrids, with an analysis by family, species and hybrid // Publications of the Gulf Coast Res. Lab. Amuseum, 1972. No 3. 328 p.

Schwartz F.J. World literature on to fish hybrids, with an analysis by family, species and hybrid; Supplement 1 // NOAA Techn. Report NMFS SSRF-750, U.S. Dep. Commerce, 1981. 507 p.

Setzer P.Y. An analysis of natural hybrid swarm by means of chromosome morphology // Transactions of Amer. Fisheries Soc., 1970. V.99. P.139-146.

Shah D.M., Langley C.H. Inter- and intraspecific variation in restriction maps of Drosophila mitochondrial DNAs // Nature, 1979. V.271. P.696-699.

Shaw C.R. Electrophoretic variation in enzymes // Science, 1965. V.149. P.936-943.

Shaw C.R., Prasad R. Starch gel electrophoresis of enzymes. A compilation of recipes // Biochem. Genet., 1970. V.4. P.292-520.

Shull G.H. Duplicate genes for capsule in *Bursa bursapastoris* // Z. Ind. Abst. Vererb., 1914. V.12. P.97-149.

Sibley C.G., Ahlquist J.E. The phylogeny of hominid primates as indicated by DNA-DNA hybridization // J. Mol. Evol., 1984. V.20. P.2-15.

Simon R.C., Noble R.E. Hybridization in *Oncorhynchus* (Salmonidae). I. Viability and inheritance in artificial crosses of chum and pink salmon // Transactions of Amer. Fisheries Soc., 1968. V.97. P.109-118.

Simpson G.G. Principles of animal taxonomy. The species and lower categories. N.Y.: Columbia Univ. Press, 1961.

Singh S.M., Zouros E. Genetic variation associated with growth rate in the American oyster (Crassostrea virginica) // Evolution, 1978. V.32. No 2. P.342-353.

Singh S.M. Enzyme heterozygosity associated with growth at different development stages in oysters // Can. J. Genet. Cytol., 1982. V.24. No 4. P.451-458.

Singh R.S., Lewontin R.C., Felton A.A. Genetic heterogeneity within electrophoretic "alleles" of xantine dehydrogenases in Drosophila pseudoobscura // Genetics, 1976. V.84. P.609-629.

Slightom J.L., Blechl A.E., Smithies O. Human fetal $^{G}\gamma$- and $^{A}\gamma$–globin genes: Complete nucleotide sequence suggest that DNA can be exchanged between these duplicated genes // Cell, 1980. V.21. P.627-638.

Skibinski D.O.F., Beardmore J.A., Cross T.F. Aspects of the population genetics of *Mytilus* (Mytilidae: Mollusca) in the British Isles // Biol. J. Linn. Soc., 1983. V.19. P.173-183.

Smith K. Racial investigations. 6. Statistical investigations on inheritance in Zoarces viviparus L. // In: Compte Rendue Traveau Lab. Carlsberg, Copenhagen, 1921. V.14(11). P.1-60.

Smouse P.E. The fitness consequences of multiple-locus heterozygosity under the multiplicative over dominance and inbreeding depression models // Evolution, 1986. V.40. P.946-957.

Smouse P.E., Chakraborty C. The use of restriction fragment length polymorphism in paternity analysis // Amer. J. Hum. Genet., 1986. V.38. P.918-939.

Sneath P.H., Sokal R.R. Numerical taxonomy. San Francisco: Freeman, 1973.

Sokal R.R., Michener C.D. A statistical method for evaluating systematic relationships. University of Kansas Sci. Bull., 1958. V.28. P.1409-1438.

Sola L., Cataudella S., Capanna E. New developments in vertebrate cytotaxonomy. III. Karyology of bony fishes: A review // Genetics, 1981. V.54. P.283-328.

Soloman D.J., Child A.R. Identification of juvenile natural hybrids between Atlantic salmon (*Salmo salar* L.) and trout (*S. trutta* L.) // J. Fish Biol., 1978. V.12. P.499-501.

Soule M. Allozyme variation, its determinants in space and time. In Molecular Evolution, Ed. F.J. Ayala. Sunderland, Massachusetts: Sinauer Assoc., Inc., 1976. P.60-77.

Sourdis J., Krimbas C. Accuracy of phylogenetic trees estimated from DNA sequence data // Mol. Biol. Evol., 1987. V.4. P.159-166.

Spolsky C., Uzzell T. Natural interspecies transfer of mitochondrial DNA in amphibians // Proc Nat. Acad. Sci. U.S.A., 1984. V.81. P.5802-5805.

Statistica for Windows. Users Guide. StatSoft Inc., Tulsa OK, East 14th Street. 1994. 1064 pp.

Stevenson M.M., Buchanon T.M. An analysis of hybridization between the cyprinodont fishes, *Cyprinodon variegates* and *C. elegans* // Copeia, 1973. P.682-692.

Stein J.P., Catterall J.F., Cristo P., Means A.R., O'Malley B.W. Ovomucoid intervining sequences specify functional domains and generate protein polymorphism // Cell, 1980. V.21. P.681-687.

Stephens J.C., Nei M. Phylogenetic analysis of polymorphic DNA sequences at the Adh locus in Drosophila melanogaster and its sibling species // J. Mol. Evol., 1985. V.22. P.289-300.

Swofford D.L., Selander R.B. Biosys-1: A Fortran program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J. Hered., 1981. V.72. No 4. P.281-283.

Swofford D.L., Olsen G.J., Waddel P.J., Hillis D.M. Phylogenetic inference // Hillis D.M., Moritz C., Mable B., Eds. Molecular Systematics. Sunderland, Massachusetts (01375-0407, USA): Sinauer. Assoc. Inc., 1996. pp.407-514.

Tabachnik W.J., Powell J.R. Genetic structure of East African domestic population of *Aedes aegypti* // Nature, 1978. V.272. P.535-537.

Tajima F. Evolutionary relationships of DNA sequences in finite populations // Genetics, 1983. V.105. P.437-460.

Takahata N., Slatkin M. Mitochondrial gene flow. Proc. Nat. Acad. Sci., U.S.A., 1984. V.81. P.1764-1767.

Tateno Y., Nei M., Tajima F. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species // J. Mol. Evol., 1982. V.18. P.387-404.

Templeton A.R. Mechanisms of speciation - population genetic approach // Ann. Rev. Ecol. Syst., 1981. V.12. P.23-48.

Templeton A.R. Phylogenetic inference from restriction endonuclease cleaverage site maps with particular reference to the evolution of humans and apes // Evolution, 1983. V.37. P.221-244.

Templeton A.R. Species and speciation: Geography, population structure, ecology, and gene trees // In: Endless Forms: Species and Speciation. Eds. D.J. Howard, S.H. Berlocher N.Y.-Oxford: Oxford Univ. Press, 1998. P.32-43.

Thompson E.A. The Icelandic admixture problems // Annals of Human Genetics, 1973. V.37. P.69-80.

Thorpe J.P. The molecular clock hypothesis: Biochemical evolution, genetic differentiation, and systematics // Ann. Rev. Ecol. Syst., 1982. V.13. P.139-168.

Tinkle D.W. Population structure and effective size of a lizard populations // Evolution, 1965. V.19. P.569-573.

Tontana F., Tagliarini J., Congiu L., Lanfredi M., Chicca M., Laurente C., Rossi R. Karyotypic characterization of Great sturgeon, *Huso huso*, by multiple staining techniques and fluorescent in situ hybridization // Marine Biology, 1998. V. 132. P.495-501.

Topal M.D., Fresco J.R. Complementary base paring and the origin of substitution mutations // Nature, 1976. V.263. P.285-289.

Turelli M., Ginzburg L. Should individual fitness increase with heterozygosity? // Genetics, 1983. V.104. P.191-209.

Turelli M. Heretable genetic variation via mutation-selection balance: A population genetic perspective // Theor. Pop. Biol., 1984. V.25. P.138-193.

Utter F. Aebersold P., Winans G. Interpreting genetic variation detected by electrophoresis // In: Population Genetics and Fishery Management. Eds. N. Ryman, F. Utter. Seattle – London: Washington Univ. Press, 1987. P.21-46.

Uzzel T., Corbin K.W. Fitting discrete probability distribution to evolutionary events // Science, 1971. V.172. P.1089-1096.

Van Valen L. Ecological species, multispecies, and oaks // Taxon, 1976. V.25. P.233-239.

Venter J.C., Adams M.D., Mayers E.W., et al. The sequence of the human genome // Science, 2001. 291:1304-1351.

Vawter A.T., Rosenblatt R., Gorman G.C. Genetic divergence among fishes of the Eastern Pacific and the Carribean: Support for molecular clock // Evolution, 1980. V.34. P.705-711.

Vernon E.H. Morphometric comparison of three races of kokanee (*Oncorhynchus nerka*) within a large British Columbia Lake // J. Fish. Res. Bd. Can., 1957. V.14. P.573-598.

Voelker R.A., Schaffer H.E., Makai T. Spontaneous allozyme mutations in *Drosophila melanogaster*: Rate of occurrence and nature of mutants // Genetics, 1980. V.94. P.961-968.

Vogel F. Non-randomness of base replacement in point mutation // J. Mol. Evol., 1972. V.1. P.334-367.

Vogel F., Motulsky A.G. Human genetics. 2nd ed. Berlin: Springer, 1986. 807 p.

Verspoor E., Fraser N.H.C., Youngson A.F. Protein polymorphism in Atlantic salmon within a Scottish river: evidence for selection and estimates of gene flow between tributaries // Aquaculture, 1991. V.98. P.217-230.

Vorontsov N.N. The problem of species and speciation // International Studies in the Phylosophy of Science, 1989. V.3(2). P.173-189.

Wahlund S. Zusamensetzung von populationen und correlation-sercheinungen vom standpunkt der vererbungslehre aus betrachtet // Hereditas, 1928. V.11. P.65-106.

Walsh M.M., Lowe D.R. Filamentous microfossils from the 3,500 MY old overwacht group, Barberton Mountain Land, South Africa // Nature, 1985. V.314. P.530-532.

Ward R.D. Biochemical genetic variation in genus Littorina (Prosobranchia: Mollusca). Hydrobiologia, 1990. V.193. P.53-69.

Ward R.D., Skibinski D.O.F., Woodwark M. Protein heterozygosity, protein structure, and taxonomic differentiation // Evolutionary Biology. Eds. M.K. Hecht et al., N.Y.: Plenum Press. 1992. V.26. P.73-159.

Watterson G.A. On the number of segregating sites in genetical models without recombination // Theor. Popul. Biol., 1975. V.7. P.256-276.

Weir B.S. Inferences about linkage disequilibrium //Biometrics, 1979. V.35. P.235-254.

Weiss G.H., Kimura M. A mathematical analysis of the stepping stone model of genetic correlation // J. Appl. Probab., 1965. V.2. P.129-149.

Weir B.S. Inferences about linkage diequilibrium // Biometrics, 1979. V.35. P.235-254.

Whitmore D.H. Introgressive hybridization of smallmouth bass (*Micropterus dolomieui*) and Gaudelupe bass // Copeia, 1983. P.672-679.

Whitney J.B., Cobb R.R., Popp R.A., O'Rourke T.W. Detection of neutral amino acid substitutions in proteins // Proc. Nat. Acad. Sci. U.S.A., 1985. V.82. P.7646-7650.

Whittaker R.H. New concepts of kingdoms or organisms // Science, 1969. V.163. P.150-160.

Wiley E.O. Phylogenetics. The theory and practice of phylogenetic systematics. N.Y.: John Wiley and Sons, 1981.

Wilhelm R. Genotype-specific selection within a hybrid population of the mussel genus *Mytilus* Master thesis. Columbia: Univ. of South Carolina, 1993.

Wilson A.C. Gene regulation in evolution // Molecular Evolution, Ed. F.J. Ayala. Sunderland, Massachusetts: Sinauer Assoc., Inc., 1976. P.225-234.

Wilson A.C., Carlson S.S., White T.J. Biochemical evolution // Ann. Rev. Biochem., 1977. V.46. P.573-639.

Wilton, A.N., C.C. Laurie-Ahlberg, T.H. Emigh and J.W. Curtsinger. 1982. Naturally occurring enzyme activity variation in *Drosophila melanogaster*. II. Relationship among enzymes // Genetics 102: 207-221.

Wolfe K., Shields D. Molecular evidence for an ancient duplication of the entire yeast genome. Nature, 1997. V.387. P.708-713. Woodruff D.S. Natural hybridization and hybrid zones // Syst. Zool., 1973. V.22. P. 213-218.

Workman P.L., Niswander J.D. Population studies on the Southwestern indian tribes. II. Local genetic differentiation in the Papago // Amer. J. Hum. Genet., 1970. V.1. P.24-29.

Workman P.L., Harpending H., Lalouel J.M., Lynch C., Niswander J.D., Singleton R. Population studies on the Southwestern indian tribes. VI. Papago population structure: A comparison of genetic and migration analysis // Genetic Structure of Populations, Ed. N.E. Morton. Honolulu: Haw. Univ. Press, 1973. P.166-194.

Wright J.E., Johnson K., Hollister A., May B. Meiotic models to explain classical linkage, pseudolinkage, and chromosome pairing in tetraploid derivative salmonid genomes. Isozymes: Curr. Top. Biol. Med. Res., 1983. V.10. P.239-260.

Wright S. Systems of mating // Genetics, 1921. V.6. P.111-178.

Wright S. Evolution in Mendelian populations. Genetics, 1931. V.16. P.97-159.

Wright S. The analysis of variance and correlations between relatives with respect to deviations from an optimum // J. Genet., 1935. V.30. P.243-256.

Wright S. Size of populations and breeding structure in relation to evolution // Science, 1938. V.87. P.430-431.

Wright S. Isolation by distance // Genetics, 1943. V.28. P.114-138.

Wright S. The genetical structure of populations // Ann. Eugenics, 1951. V.15. P.323-354.

Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating // Evolution, 1965. V.19. P.395-420.

Wright S. Evolution and the genetics of population. Vol. 1. Genetic and biometric foundations. Chicago: Univ. Chicago Press, 1968.

Wright S. Evolution and the genetics of population. Vol. 2. The theory of gene frequencies. Chicago: Univ. Chicago Press, 1969. 511 p.

Wright S. Variability within and among populations. Evolution and the genetics of populations. Vol. 4. Chicago – London: University of Chicago Press, 1978.

Wu C.-I., Hollocher H. Subtle is nature: The genetics of species differentiation and speciation // In: Endless Forms: Species and Speciation. Eds. D.J. Howard, S.H. Berlocher. N.Y.-Oxford: Oxford Univ. Press, 1998. P.339-351.

Wyles J.S., Kunkel J.G., Wilson A.C. Birds, behavior, and anatomical evolution // Proc. Nat. Acad. Sci. U.S.A., 1983. V.80. P.4394-4397.

Yager L.N., Kaumeyer J.F., Weinberg E.S. Evolving sea urchin histone genes – nucleotide polymorphisms in the H4 gene and spacers of *Strongylocentritus purpuratus* // J. Mol. Evol., 1984. V.20. P.215-226.

Yang S.Y., Patton J.L. Genetic variability and differentiation in the Galapagos finches // The Auk, 1981. V.98. P.230-242.

Yonekawa H., Moriwaki K., Gotoh O., Hayashi J.-I., Watanabe J., Miyashita N., Petras M.L., Tagashira Y. Evolutionary relationships among five subspecies of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA // Genetics, 1981. V.98. P.801-816.

Yonekawa H., Tsuda K., Tsuchia K., Yakimenko L.V., Korobitsyna K.V., Chelomina G.N., Spiridonova L.N., Frisman L.V., Kryukov A.P., Moriwaki K. Genetic diversity, geographic distribution and evolutionary relationships of *Mus musculus* subspecies based on polymorphism of mitochondrial DNA // In: Problems of Evolution. Ed. A.P. Kryukov, L.V. Yakimenko, Vladivostok: Dalnauka, 2000. P.90-108.

Yunis J.J., Prakash O. The origin of man: A chromosomal pictorial legacy // Science, 1982. V.215. P.1525-1530.

Zouros E. Mutation rates, population sizes, and amount of electrophoretic variation of enzyme loci in natural populations // Genetics, 1979. V.92. P.623-646.

Zouros E., Singh S.M., Miles H.E. Growth rate in oysters: an overdominant phenotype and its possible explanations // Evolution, 1980. V.34. No 5. P.856-867.

Zouros E., Folts D.W. Minimal selection requirements for the correlation between heterozygosity and growth, and for the deficiency of heterozygotes, in oyster populations // Dev. Genet., 1984. V.4. P.393-405.

Zouros E. On the relation between heterozygosity and heterosis: An evaluation of the evidence from marine mollusks // Isozymes: Current Topics in Biol. Med. Res., 1987. V.15. P.255-270.

Zouros E., Folts D.W. The use of allelic isozyme variation for the study of heterosis // Isozymes: Current Topics in Biol. Med. Res., Eds. M.C. Ruttazi, J.S. Scandalios, G.S. Whitt. N.Y.: Alan Liss., 1987, V.13, P.1-59.

Zouros E., Mallet A.L. Genetic explanations of the growth/ heterozygosity correlation in marine mollusks. In Reproduction Genetics and Distribution of Marine Organisms. Eds. J.S. Ryland, P.A. Tyler, Fredensborg: Olsen & Olsen, 1989. pp. 317-323.

Zuckerkandl E., Pauling L. Molecular disease, evolution, and genetic heterogeneity. In: Horizons in Biochemistry. Eds. M. Kasha, B. Pullman. N.Y.: Academic Press, 1962. P.189-225.

Zuckerkandl E., Pauling L. Evolutionary divergence and convergence in proteins // In: Evolving Genes and Proteins. Eds. V. Bryson, H.J. Vogel. N.Y.: Academic Press, 1965. P.97-166.