

Анализ генетического разнообразия с данными молекулярных маркеров: учебный модуль.

К.С. де Висенте (Bioversity International)

Ц. Лопес (Universidad Nacional Agraria
'La Molina', Перу) и

Т. Фултон (Институт геномного разнообразия,
Корнельский Университет)



Цели настоящего модуля

Цель настоящего модуля заключается в том, чтобы пользователи могли:

- Понять научные концепции генетического разнообразия посредством понимания основ популяционной генетики
- Ознакомиться с математическими выражениями, использованными для описания генетического разнообразия, и уметь выполнять необходимые расчеты на основе данных молекулярных маркеров
- Владеть основными знаниями для применения молекулярных технологий для оценки генетического разнообразия и толкования молекулярных данных соответственно.

Предисловие

Учебный модуль «Анализ генетического разнообразия при помощи данных молекулярного маркера» - это второй из двух обучающих модулей, разработанных для оказания содействия в применении молекулярных технологий для изучения биоразнообразия в обучающих программах. Он дополняет первый модуль «Использование технологии молекулярных маркеров в изучении генетического разнообразия растений: учебный модуль» и также нацелен на распространение знаний о технологии молекулярных маркеров для оценки генетического разнообразия. Таким образом, модуль обеспечивает базу для понимания анализа и толкования данных, а не является просто повторением того, как использовать технологии только потому, что это модно.

Поводом для разработки настоящего обучающего модуля послужил курс обучения по молекулярным технологиям, организованный Bioversity International в Китае при участии ученых из Китая, Индии, Индонезии, Малайзии и Филиппин, изучающих генетическое разнообразие тропических плодовых культур. В конце курса обучения было осознано, что обучение технологиям позволило подготовить очень хороших лаборантов. Однако если подготовлены ученые, компетентные в управлении и использовании генетических ресурсов, то далее партнеры Bioversity International должны быть снабжены инструментами анализа и толкования данных.

Учебный модуль предназначен в частности для ученых, которые интересуются оценкой генетического разнообразия и обладают основными знаниями в биологии и генетике, знакомы с молекулярными технологиями и нуждаются в руководстве для постановки экспериментов, а также проведения анализов и толкования их результатов.

Авторы надеются, что пользователи данных обучающих модулей не только найдут информацию в них просветительной, но также и практически полезной, и что использование технологий молекулярных маркеров может быть увлекательным и практически осуществимым подходом для любого, кто планирует исследования в области анализа генетического разнообразия.

Благодарность

Разработка учебного модуля «Анализ генетического разнообразия при помощи данных молекулярного маркера» стала возможной благодаря совместной работе Bioversity International и Института геномного разнообразия (IGD) Корнельского Университета. Авторы выражают особую благодарность за помощь следующим лицам:

Сотрудникам Регионального офиса Bioversity International для стран Тихоокеанского региона и Океании, которые в рамках проекта по тропическим плодовым культурам, финансируемого Азиатским Банком Развития (АБР), оказали частичную финансовую помощь в разработке настоящего модуля.

Феликсу Альберто Гусману, региональный офис Bioversity International для стран Латинской Америки, за помощь в упрощении предоставленной информации и улучшении окончательной версии модуля согласно полученным правкам от разных коллег и редактора.

Аманде Гаррис, IGD, и Стиву Танксли, Департамент селекции растений, оба сотрудники Корнельского Университета, за разрешение использовать информацию и иллюстрации из их слайдов.

Джоанне Лабате за «виртуальные обсуждения», Александре Гаса и Марте Хамблин, IGD, за их советы о том, какие программы должны быть обсуждены в модуле

Хумберто Гомезу Паниагуа, Международный центр тропического сельского Хозяйства (CIAT), Колумбия, за его ценные комментарии по улучшению, как содержания, так и дидактической презентации материала.

Мириам Кристине Дугу, CIAT, за ее помощь в подготовке плана данного модуля и уточнение концепций измерения генетического разнообразия.

Мерилин Варбуртон, Центр прикладной биотехнологии, CIMMYT, Мексика, за ее предложения по улучшению слайдов о методах образования кластеров.

Персоналу Отдела генетических ресурсов, CIAT, Колумбия, за предоставление изображений фасоли для использования в качестве основополагающей информации в одном из субмодулей.

Профессору Хайберу Карденас, кафедра биологии Университета Дель Валь, г. Кали, Колумбия, за апробацию обучающего модуля при помощи пяти его студентов (Иван Андрес Гонзалес, Ольга Лючио Агудело Хенао, Марта Нора Майно, Фернанд Рондон Гонзалес и Диего Маурисио Вилламарин Миранда), а также за предоставление ценных замечаний по усовершенствованию модуля.

Луиджи Гуарино, Раманаса Рао, Иссиака Зоунграна, Маргарите Баена, Зонгвен Зангу, Миккел Граму, Яну Энгельсу и Элизабет Гольдберг, сотрудникам офисов Bioversity International в различных регионах, за их замечания и предложения по усовершенствованию данного модуля и увеличению его полезности для наших партнеров.

Элизабет Л. Мак Адам, за её помощь в редактировании рукописи и ценные предложения по улучшению формата настоящего модуля, а также Лин Менендес, CIAT, за координацию работы по редактированию рукописи и полезным советам по модификации модуля.

Гледис Родригес, CIAT, за помощь в компоновке данного модуля.

Анализ генетического разнообразия с данными молекулярных маркеров: учебный модуль.

О данном модуле

Цели

I Введение

II Основные концепции популяционной генетики

III Измерения генетического разнообразия

**IV Компьютерные программы для анализа
генетического разнообразия**

V Глоссарий

VI Форма обратной связи

Что Вам необходимо знать об этом модуле

При планировании исследования очень важно, чтобы вопрос, на который Вы хотите ответить в его результате, был ясно сформулирован, даже если параллельно возникают другие вопросы, и Вы, по ходу Вашего исследования, желаете внести изменения в первоначальный план. Вы должны также знать некоторые принципы постановки экспериментов, чтобы быть уверенным в получении достоверных результатов. В контексте Вашей работы с генетическими ресурсами Вам необходимо знать основы генетического разнообразия и существующие инструменты для проверки Ваших данных, чтобы Вы могли правильно их истолковать.

Концепции и инструменты, которые мы обсуждаем, включают:

- Стратегии отбора образцов;
- Основы популяционной генетики;
- Математические измерения, используемые для описания генетического разнообразия, индексы генетической дистанции и методы, используемые для экспрессии взаимосвязанности среди образцов;
- Компьютерные программы и существующие ресурсы Интернета.

Расчеты, использованные в ключевых концепциях популяционной генетики, измерениях генетического разнообразия, индексов и методах образования кластеров иллюстрированы примерами, которые были подготовлены специально для настоящего модуля. Пользователь может сразу же увидеть, как математические экспрессии применяются. В дополнение, мы хотим показать, что даже если эти расчеты часто выполняются при помощи компьютерных программ, большинство их можно проделать вручную. Даже если используется компьютерная программа, мы верим, что исследователь должен понять, что делает компьютер, и уметь разработать свой собственный критерий, чтобы решить какой метод выбрать.

Предполагается, что настоящий учебный модуль окажет помощь тем, кто желает проводить анализы генетического разнообразия при помощи молекулярных данных. Как таковой, модуль не является исчерпывающим средством для того, чтобы научиться преподавать популяционную генетику. Однако, мы привели перечень литературных источников для поддержания важных суждений популяционной генетики, демонстрации математических расчетов и обеспечения наилучшего понимания того, как эти методы применять в реальных ситуациях исследования. Идея заключается в том, что учебный модуль может служить самостоятельным инструментом для оказания содействия обучающему процессу, в частности, помочь студентам, особенно тем, которые учатся на последних курсах Университетов, понять, как сделать выбор среди молекулярных технологий, применяющихся в изучении генетического разнообразия, и суметь использовать их в своих исследованиях. Он также может использоваться для обучения и подготовки лекций или как справочное пособие для практикующих профессионалов, которым нужно применять молекулярные и статистические методы, а также проводить анализ данных в своих исследованиях.

Пользователи настоящего модуля должны обладать основными знаниями в генетике. Кроме того, если они не знакомы с технологиями молекулярных маркеров, мы настоятельно рекомендуем начать обучение с изучения первого модуля *«Использование технологии молекулярных маркеров в изучении генетического разнообразия растений: учебный модуль»*. Как только основные принципы этих методов будут поняты, второй модуль может помочь в лучшем понимании математических алгоритмов.

Мы сконструировали учебный модуль в виде взаимодополняющих и независимых субмодулей, чтобы пользователь мог выбрать в любой момент интересующий его раздел. В некоторых случаях мы привели дополнительную информацию, добавив приложения, чтобы не усложнять основную часть, и все же приводя дополнительные математические выражения и примеры для тех, кто найдет их полезными для более глубокого понимания. Настоящий учебный модуль представлен таким образом, когда могут быть использованы либо слайды в формате презентации, либо и слайды, и сопровождающие их пояснения.

Ваши отзывы на данную работу очень важны для нас, потому что мы убеждены, что они всегда полезны. Чтобы наш ответ партнерам и другим пользователям был эффективным, мы были бы безмерно признательны, если Вы пришлете нам свои отзывы на организацию, содержание и полезность настоящего модуля по следующим адресам:

M. Carmen de Vicente
Bioversity International
Via dei Tre Denari 472/a
00057 Maccarese
Rome, Italy
E-mail: cdevicente@cqi-ar.org

Cesar Lopez
Universidad Nacional Agraria 'La Molina'
Av. La Universidad s/n Apdo. 456 Lima 12
Lima, Peru
E-mail: cflb@lamolina.edu.pe


Theresa Fulton
Institute for Genomic Diversity
130 Biotechnology Building
Cornell University
Ithaca, NY 14853
E-mail: tf12@cornell.edu

Мы искренне надеемся, что данный модуль дополняет предыдущий «*Использование технологии молекулярных маркеров в изучении генетического разнообразия растений: учебный модуль*», и что эти два модуля вместе предоставляют нашим партнерам, особенно тем, кто из развивающихся стран и имеет ограниченный доступ к современным технологиям и обширным научным публикациям или инструкциям, возможность проводить перспективные исследования в области генетического разнообразия растений, и таким образом, вносят вклад в мировые знания об этих ценных ресурсах.

M. Кармен де Висенте
Bioversity International

Cesar Lopez
Univ. 'La Molina'

Тереза Фултон
IGD, Корнельский
Университет




**Анализ генетического разнообразия
с использованием данных
молекулярных маркеров:**

Учебный модуль

Введение

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 1



Оглавление

- ▶ **Научный метод**
 - Постановка биологического вопроса
 - Разработка гипотезы и дизайн эксперимента
 - Отбор образцов внутри популяций
 - Отбор образцов внутри генома
 - Проведение эксперимента
 - Анализ и интерпретация данных
- ▶ **Уровни биологического разнообразия**
- ▶ **Измерение генетической изменчивости**
- ▶ **Взаимосвязь между фенотипом и генотипом**

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 2



Научный метод

- ▶ Постановка биологического вопроса
- ▶ Разработка гипотезы и дизайн эксперимента
- ▶ Проведение эксперимента
- ▶ Анализ и интерпретация данных

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 3

Применение научного метода начинается с постановки биологического вопроса, и таким образом, определяется причина проведения исследования. После этого начинается итерационный процесс, проходящий несколько этапов, которые ведут к анализу результата. Этот процесс может быть повторен несколько раз прежде, чем будут получены окончательные результаты. Результаты должны предоставить данные для подтверждения или отвержения гипотезы, поддерживающей дизайн эксперимента. Если гипотеза подтверждена, тогда получен удовлетворительный ответ для открытия вопроса. В случае, если гипотеза отвергнута, тогда новая гипотеза должна быть сформулирована и спроектирован другой эксперимент. Таким образом, процесс начинается снова, и повторяется до тех пор, пока не будет получена удовлетворительная интерпретация.

Иногда, когда научный интерес заключается в оценке генетического разнообразия, исходный вопрос заменяется необходимостью описания. В этом случае, гипотезы не существуют и эксперимент проектируется для сбора информации о существующей изменчивости.

На следующих слайдах будет обсужден каждый этап, упомянутый выше.



Постановка биологического вопроса

- ▶ В чем заключается проблема?
- ▶ Сбор современной информации на тему:
 - Было ли это уже изучено?
 - Существует ли объяснение?

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 4

Первый этап заключается в постановке биологического вопроса, лежащего в основе нашего научного интереса. Мы можем искать или ответ на вопрос или описание. Исследования генетического разнообразия мы часто начинаем с поиска описания, например: сколько вариаций здесь имеется? Как эта вариация организована? Результаты, которые мы получим в итоге, вероятно всего, приведут нас к вопросам, которые уже должны подразумеваться в описательном анализе.

Литературный поиск поможет нам в определении интересующего предмета, собрать информацию в более широком контексте и, соответственно, определить ключи к гипотезе, которую, возможно, нам необходимо выдвинуть.

Какого рода вопрос?

Задайте вопрос о:

- Стратегиях сохранения
- Использовании в селекции культур
- Эволюции и одомашнивании

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 5

Как ранее упомянуто, много вопросов могут привести к необходимости исследования. Примерами таких вопросов являются следующие:

- Относящиеся к сохранению:
 - Как разнообразие представлено в природе и в коллекциях?
 - Что является приоритетом для сохранения?
 - Как используя ограниченные ресурсы, мы можем застраховать себя на будущее?
 - Как мы можем количественно измерить имеющееся разнообразие для изучения генетической эрозии в будущем?
 - Существует ли характер распространения, который может быть использован для направления наших действий в проведении сбора?
 - Можем ли мы гарантировать, что наши образцы различны?
 - Принадлежат ли эти, очевидно разные, образцы к разным таксонам?
- Относящиеся к селекции культур:
 - Каким образом наш образец может быть использован в улучшении культур?
 - Являются ли эти генетические ресурсы возможными источниками аллельного разнообразия?
 - Являются ли они хорошими источниками желательных агрономических признаков?
- Относящиеся к эволюции:
 - Откуда происходит эта культура?
 - Что является предком определенной культуры?
 - Имеет ли место интрогрессия между образцами популяции различного происхождения?

Пример двух вопросов: Сколько вариаций здесь имеется? Сколько из них мы можем потерять?

Локусы	A	B	C	D	E	F	G	H	I	
Аллель	A ₁	B ₁	C ₁	D ₁	E ₁	F ₁	G ₁	H ₁	I ₁	
	A ₂	B ₂	C ₂	—	E ₂	F ₂	G ₂	—	I ₂	
	—	—	C ₃	—	E ₃	—	G ₃	—	I ₃	
	—	—	—	—	E ₄	—	—	—	—	
Генотип	A ₁ A ₁	B ₁ B ₁	C ₁ C ₁	D ₁ D ₁	E ₁ E ₁	F ₁ F ₁	G ₁ G ₁	H ₁ H ₁	I ₁ I ₁	
	A ₁ A ₂	B ₁ B ₂	C ₁ C ₂		E ₁ E ₂	F ₁ F ₂	G ₁ G ₂		I ₁ I ₂	
	A ₂ A ₂	B ₂ B ₂	C ₁ C ₃		E ₁ E ₃	F ₂ F ₂	G ₁ G ₃		I ₁ I ₃	
			⋮		⋮		⋮		⋮	
			др.		др.		др.		др.	
Общее число генотипов			$\Sigma K(K + 1)/2$							
Локусы	A	B	C	D	E	F	G	H	I	
Генотипы (кол-во)	3	3	6	1	10	3	6	1	6	58320
Одна особь	A ₁ A ₂	B ₂ B ₂	C ₂ C ₃	D ₁ D ₁	E ₂ E ₄	F ₁ F ₁	G ₃ G ₃	H ₁ H ₁	I ₂ I ₂	
Другая особь	A ₁ A ₁	B ₁ B ₂	C ₁ C ₁	D ₁ D ₁	E ₁ E ₃	F ₂ F ₂	G ₂ G ₃	H ₁ H ₁	I ₃ I ₃	

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 6

Сколько вариаций здесь имеется? и Сколько из них мы можем потерять?

Эти два вопроса являются примерами того, что мы хотели бы спросить, чтобы подсчитать количество вариаций, присутствующих в образце и, как результат, оценку числа вариаций, которые мы могли бы потерять. Слайд демонстрирует упрощенную ситуацию организма, для которого мы проанализировали 9 локусов, каждый с различным числом аллелей в образце (локусы A и B имеют по 2 аллели каждый, локус C имеет 3 аллели, локус D только 1 аллель и т.д.). Вторая часть таблицы показывает генотипы для каждого локуса: A₁ A₁, A₁ A₂, A₂ A₂, B₁ B₁ и т.д.

Подсчет предоставляет общее число генотипов на локус, где K = число аллелей в локусе. Число генотипов подсчитано для каждого локуса (например, в локусе A = 3, в локусе B = 3, в локусе C = 6, в локусе D = 1, ...). Затем, подсчитывается общее число возможных генотипов в этом организме на основании данных 9 локусов (3 x 3 x 6 x 1 x 10 x 3 x 6 x 1 x 6 = 58,320 генотипов).

В последних двух строках таблицы, показаны два генотипа из возможных 58 320. Согласно стратегии сохранения, в случае, если только два генотипа отобраны произвольно, то будет потеряна большая часть генетической изменчивости.

Разработка гипотезы и дизайн эксперимента

- ▶ Как можно наилучшим образом адресовать вопрос?
- ▶ Имеются ли альтернативные стратегии?
- ▶ Рассмотрение затрат, времени и имеющихся в наличии ресурсов
- ▶ Стратегии отбора образцов
 - Популяции в сравнении с особями
 - Геном

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 7

Хорошая гипотеза заключается в предположении изучения и должна быть поддающейся экспериментальной оценке, которая позже послужит для формулирования обширных выводов. Для постановки хорошей гипотезы в изучении генетического разнообразия требуются основные знания, например, о репродуктивных системах, пространственном и временном распространении, межвидовых взаимодействиях. Например, если мы хотим проанализировать, в основном, самоопыляющийся вид, то наша гипотеза может быть основана на факте, что этот вид будет иметь низкий уровень внутривидового полиморфизма и высокий уровень полиморфизма среди популяций. Или, если особи данного вида широко распространены в области, то наша гипотеза может предположить, что популяции в этом регионе будут иметь сходные аллельные частоты.

Как только определен основной вопрос, гипотеза поможет нам построить эксперимент. Как правило, мы могли бы использовать одну из нескольких стратегий. В конечном счете, выбранная стратегия будет зависеть от имеющихся ресурсов, как инфраструктура, опыт, время и финансовые средства. Также очень важно подумать о стратегиях отбора образцов: 1) отбор особей в нашей целевой популяции и 2) отбор генома. В исследованиях с использованием молекулярных маркеров, отбор образцов геномов предполагает принятие решения насчет количества точек вдоль нити (последовательность ДНК или хромосомы) для гарантированного получения набора данных, которые бы адекватно отвечали нашему первоначальному вопросу (например, какая точность нам нужна?)

Далее, мы обсудим отбор образцов более подробно.

Для более подробной информации по некоторым этим вопросам, обратитесь к обучающему модулю «*Использование технологий молекулярных маркеров в изучении генетического разнообразия растений*» (www.bioversityinternational.org).

Отбор образцов внутри популяций

- ▶ Сколько популяций и/или особей нам нужно для достижения цели нашего исследования?
- ▶ Реальна ли цель нашего исследования?

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 8

Каждый подход в измерении генетической изменчивости и ее структуры в популяциях может потребовать принятия решения об использовании специфичной, оптимальной стратегии отбора образцов.

Два основных критерия помогают определить оптимальную процедуру отбора, независимо от адресованного научного вопроса (Brown и Weir 1983): (1) стратегия выбора может считаться оптимальной, если варианса отбора минимизирована в каждой единице эксперимента; и (2) легкость оперирования и наличие ресурсов.

Поскольку не существует комплексной публикации, отвечающей на вопросы по всем ситуациям, разрешение некоторых положений может быть найдено в литературе. Например, Brown и Weir (1983) об отборе для оценки частоты аллелей, количества аллелей в локусе и разнообразия генов. Они, в свою очередь, упоминают основную литературу, которая может относиться к изучаемой теме. Другим примером является то, что Gregorius (1980) дает идеи о процедурах отбора и предоставляет таблицу для расчета минимальных размеров образца, необходимых для гарантирования, что все аллели с частотой выше, чем определенный порог, выявлены с учетом вероятности .

Однако, во всех случаях было бы важным оценить и сбалансировать предполагаемые результаты исследования и имеющиеся ресурсы. Для более подробной информации, см. раздел «*Заключительные выводы*» в обучающем модуле «*Использование технологии молекулярных маркеров в изучении генетического разнообразия растений*» (http://www.bioversityinternational.org/publications/pubfile.asp?ID_PUB=912).

Литература

Brown, A.H.D. и B.S. Weir. 1983. Measuring genetic variability in plant populations. Pp. 219-239 *in* Isozymes in Plant Genetics and Breeding, part A (S.D. Tanksley and T.J. Orton, eds.). Elsevier Science Publishers, Amsterdam.

Gregorius, H.R. 1980. The probability of losing an allele when diploid genotypes are sampled. *Biometrics* 36:643-652.

Отбор образцов внутри генома

- ▶ Сколько локусов должно быть проанализировано?
- ▶ Возможны две стратегии отбора:
 - Отобрать несколько высоко-информативных маркера
 - Отобрать много, слабо-информативных маркера, случайно распространенных внутри генома

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 9

С точки зрения отбора генома, идеальная ситуация могла быть одна, в случае, когда известно распространение генетического полиморфизма внутри генома. Однако, это не всегда имеет место и, как таковое, заранее не имеются знания о распространении разнообразия. Соответственно, исследования или проводятся непродуманно путем проведения экспериментов подобных уже опубликованным, или они должны быть основаны на моделировании. Во всех случаях сравнение преимуществ и недостатков разных типов маркеров является важным, потому что они имеют типично различное распространение внутри генома.

Mariette и др. (1999) представляют результаты моделирования эксперимента, в котором сила двух типов маркеров с разным количеством локусов была проанализирована: (1) AFLP эксперимент с 200 локусами, оцененный либо как доминантные, либо как кодоминантные данные, и (2) эксперимент с микросателлитами с двумя разными ситуациями – один с 5 локусами и другой с 50 локусами. Результаты были представлены в диаграмме, которая показала компромисс между числом проанализированных локусов и количеством и точностью полученной информации. Диаграмма показывает, что и количество, и содержание информации (наследование доминантности в сравнении с кодоминантностью) использованного типа маркера являются важными в обеспечении надлежащего отбора генома.

Литература

Mariette, S., V. Lecorre и A. Kremer. 1999. Sampling within the genome for measuring within-population diversity: Trade-offs between markers *in* Sampling Strategies for Marker Analysis. A compendium of the research project *Development, Optimization and Validation of Molecular Tools for the Assessment of Biodiversity in Forest Trees* as part of the European Union DGXII Biotechnology FW IV Research Programme *Molecular Tools for Biodiversity*. (E.M. Gillet, ed.).
<<http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>>

Проблемы плохого отбора образцов

1. Если количество образцов очень маленькое, мы получаем смещенную оценку частоты аллелей

Особи (число)	Генотипы			Частота аллель	
	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q
10	7	1	2	0,75	0,25
30	4	16	10	0,4	0,6
100	27	39	34	0,465	0,535
1000	299	395	306	0,4965	0,5035
1 000 000	300 000	400 000	300 000	0,5	0,5

Авторские права: Bioversity International и Корнельский Университет, 2003

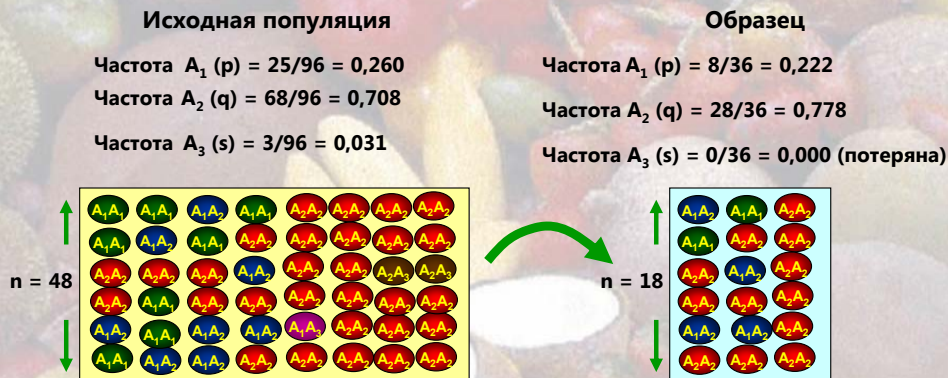
Введение 10

(продолжение на следующем слайде)

Таблица наверху демонстрирует пример, где только для одного гена (A) с двумя аллелями, частота аллелей была подсчитана для образцов, увеличивающихся в числе особей. Когда размер образца увеличился, шанс, что все генотипы могли бы быть равно представленными становится выше. Таким образом, оценка частот аллелей становится ближе к реальной ситуации, и результаты экспериментов становятся более достоверными.

Проблемы плохого отбора образцов (продолжение)


2. Некоторые аллели ускользают при отборе образцов, и они засчитываются как отсутствующие :



Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 11

Пример наверху демонстрирует ситуацию, когда суб-образец отобран в популяции таким образом, что аллель A_3 , представленный в исходной популяции, отсутствует. Суб-образец не содержит все аллели и, поэтому, не является представителем более обширной популяции. И описание генетического разнообразия более меньшего образца, как оценка более большой популяции, не будет равноценной и обоснованной.



Проведение эксперимента

- ▶ Используйте соответствующие инструменты для тестирования гипотезы
- ▶ Включите измерения контроля качества эксперимента и сбора данных

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 12

Для постановки эксперимента, исследователь должен учесть, что молекулярные инструменты могут предложить более глубокие исследования биоразнообразия и, что они обеспечивают общую основу для измерения и анализа разнообразия. Однако, данные молекулярных исследований очень часто дополняют данные других характеристик (например, морфологической, патологической), и комбинированный анализ этих данных может предложить более полную основу для интерпретации.

При выборе молекулярной технологии необходимо обратить внимание на:

- Ее потенциал для ответа на биологический вопрос
- Возможности для отбора внутри генома
- Тип и качество растительного материала
- Измерения качества данных
- Измерения для обеспечения воспроизводимости
- Квалификация, требуемая для сбора данных

Для более подробной информации по вышеупомянутым вопросам смотрите обучающий модуль «Использование технологии молекулярных маркеров в изучении генетического разнообразия растений» (www.bioversityinternational.org).



Анализ и интерпретация данных

- ▶ Поддерживают ли данные гипотезу?
- ▶ Какие аналитические инструменты должны мы использовать?
- ▶ Что нам также необходимо:
 - Научные концепции, лежащие в основе представлений о разнообразии, например:
 - Уровни биологического разнообразия
 - Изменчивость между и изменчивость внутри
 - Взаимосвязь между фенотипом и генотипом
 - Основные концепции популяционной генетики

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 13

Как только данные собраны, их анализ является следующим этапом эксперимента. Имейте в виду, что результаты основаны только на данном образце, и полученные значения будут являться только оценкой параметров популяции.

Для анализа имеются различные аналитические инструменты, а также компьютерные программы, которые будут обсуждены в последних двух разделах данного модуля. Одновременно нам необходимо знать основные научные концепции по генетическому разнообразию и популяционной генетике (которая является частью генетики, изучающей генетическую изменчивость). Последующие слайды и раздел дадут краткий обзор этих тем.

Уровни биологического разнообразия

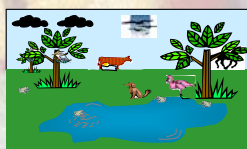
1. Внутри-популяционное разнообразие (= генетическое разнообразие)



2. Разнообразие видов



3. Разнообразие экосистем



Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 14

Разнообразие может быть разделено на категории, которые описывают различные аспекты живых систем, измеряемых учеными различными способами.

В настоящее время является общей практикой определение биоразнообразия в отношении генов (внутрипопуляционное разнообразие или генетическое разнообразие), видов и экосистем, соответствующих трем фундаментальным и иерархически связанным уровням биологической организации.

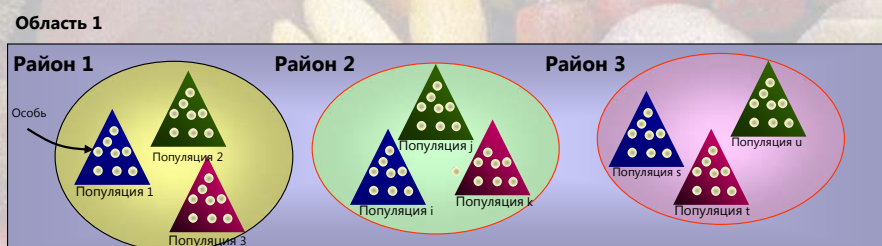
«Генетическое разнообразие» относится к вариации генов внутри видов, что является смещением генов, содержащихся в особях. Это охватывает различные популяции одного и того же вида или генетическую вариацию внутри популяции. В конечном счете, генетическое разнообразие, заключается в изменениях последовательности четырех пар оснований ДНК и являют собой генетический код. Другие типы генетического разнообразия могут быть вызваны количеством ДНК в каждой клетке, или ДНК, определяющей генетический код. Новая генетическая изменчивость создается мутациями генов или хромосом, и в организмах, размножающихся половым путем, становится распространенной по вариантам, в структуре и количестве хромосом. Отбор влияет на фонд настоящей генетической изменчивости и, соответственно, способствует эволюционному процессу и искусственному отбору. Генетическое разнообразие позволяет популяциям адаптироваться к изменяющимся условиям окружающей среды.

Уровень видов, в общем, считается наиболее естественным для рассмотрения разнообразия организма, в целом. На видах, главным образом, также сфокусирована эволюция, происхождение как и исчезновение видов является основным показателем биологического разнообразия во многих отношениях. Разнообразие на уровне видов может быть измерено несколькими способами. «Богатство видов» означает количество видов в области. «Таксономическое разнообразие» также измеряет взаимосвязь видов друг к другу.

Определение и классификация экосистем очень широко, потому что их границы обширны. Поэтому, на практике, очень трудно оценить разнообразие экосистем более чем на местном или региональном уровне, и затем, только в значительной степени, в отношении растительного покрова. Экосистемы отличаются от генов и видов не только по своему составу (они включают в себя абиотические компоненты, и частично обуславливаются почвенными и климатическими условиями), но также по своей структуре и функциям.

Измерение генетической изменчивости

- ▶ Изменчивость или полиморфизм может быть оценен на разных уровнях организации
- ▶ Распространение полиморфизма изучается для разных иерархических уровней внутри организации (области, районы, популяции, субпопуляции, особи)



Авторские права: Bioversity International и Корнельский Университет, 2003

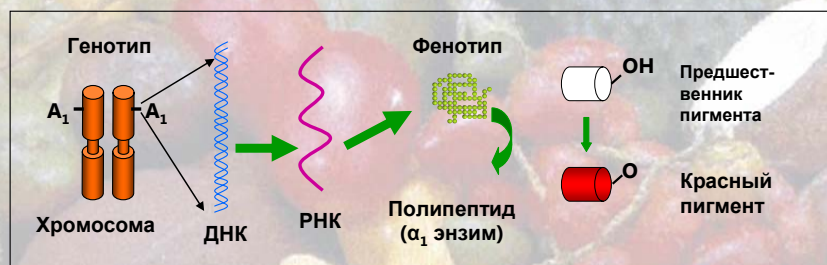
Введение 15

Генетические различия обнаружены между особями внутри популяции, и также различия выявлены в аллельных частотах между популяциями. Во всех, относительное число вариации зависит от вида, его истории и окружающей среды.

Наличие вариантов (полиморфизма) в образце может быть оценено по его генотипам, аллелям, гаплотипам или нуклеотидам. Образцы могут быть иерархически разделены на уровне вида, популяции или внутри популяции. Количество образцов должно быть достаточным, когда поставленной целью является изучение генетической изменчивости (предыдущие слайды). Стратегии отбора и размер образцов могут зависеть от организации видов:

- Несколько особей (новый интродуцированный вид)
- Много интродуцированных особей
- Особи различного географического происхождения

Взаимосвязь между фенотипом и генотипом

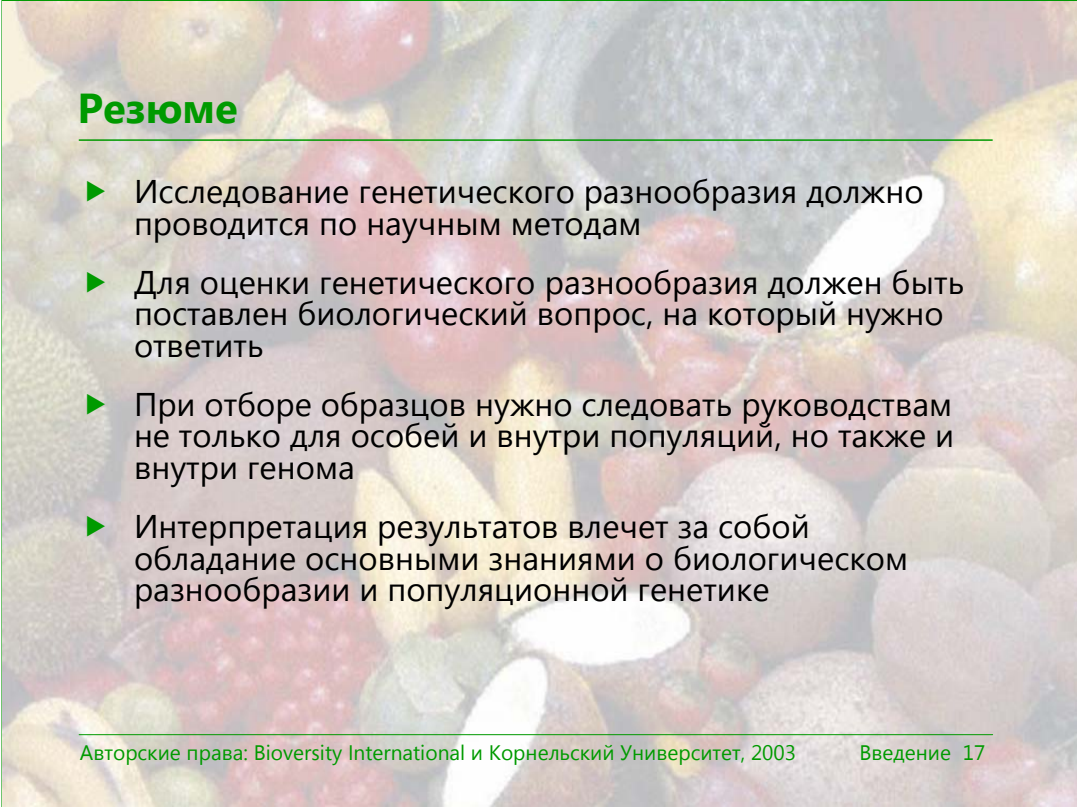


- Фенотип
 - Специфичный по признакам
- Молекулярные маркеры
 - Нейтральное разнообразие
 - ДНК или протеин
- Последовательность ДНК
 - Разнообразие аллелей
- Экспрессия
 - На уровне РНК

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 16

Проведение анализа генетической изменчивости с использованием молекулярных технологий предоставляет информацию на уровне ДНК. Это может быть нейтральное разнообразие, определенное вдоль последовательности ДНК в областях, чьи функции неизвестны, как тогда, когда мы используем анонимные типы маркеров (например, AFLP, RAPD). Или разнообразие может быть основано на известных генах в пределах кодирующих областей последовательности ДНК. Это разнообразие влияет на экспрессию этих генов и, соответственно, РНК – нуклеидная кислота отвечает за перевод информации генетического кода в протеин. Протеины, в свою очередь, являются элементами, которые формируют структуру организмов, что означает, что они ответственны за то, что мы видим - за фенотип. Таким образом, генотип и фенотип тесно взаимосвязаны. Фенотипические измерения разнообразия могут быть также использованы и, если правильно сделаны, то они могут отразить молекулярную конституцию данной особи.



Резюме

- ▶ Исследование генетического разнообразия должно проводиться по научным методам
- ▶ Для оценки генетического разнообразия должен быть поставлен биологический вопрос, на который нужно ответить
- ▶ При отборе образцов нужно следовать руководствам не только для особей и внутри популяций, но также и внутри генома
- ▶ Интерпретация результатов влечет за собой обладание основными знаниями о биологическом разнообразии и популяционной генетике



Теперь Вы должны знать ...

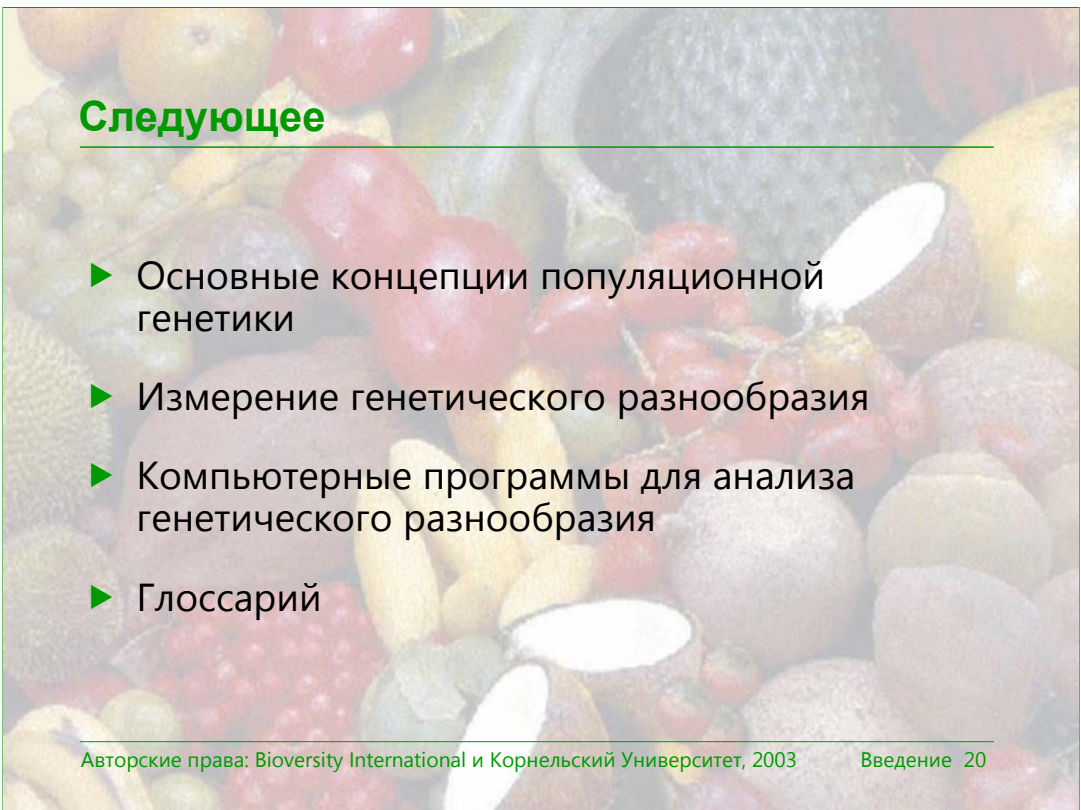
- ▶ Важные шаги в постановке эксперимента
- ▶ Важность постановки ключевых вопросов для Вашего исследования
- ▶ Основные моменты, которые необходимо учитывать при отборе образцов
- ▶ Основные концепции, лежащие в основе представлений о генетическом разнообразии

Литература

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 19

- Brown, A.H.D. и B.S. Weir. 1983. Measuring genetic variability in plant populations. Pp. 219-239 *in* Isozymes in Plant Genetics and Breeding, part A (S.D. Tanksley and T.J. Orton, eds.). Elsevier Science Publishers, Amsterdam.
- de Vicente, M.C. и T. Fulton. 2003. Using Molecular Marker Technology in Studies on Plant Genetic Diversity.
<www.bioversityinternational.org/publications/pubfile.asp?ID_PUB=912>
- Gregorius, H.R. 1980. The probability of losing an allele when diploid genotypes are sampled. *Biometrics* 36:643-652.
- Mariette, S., V. Lecorre и A. Kremer. 1999. Sampling within the genome for measuring within-population diversity: Trade-offs between markers *in* Sampling Strategies for Marker Analysis. A compendium of the research project *Development, Optimization and Validation of Molecular Tools for the Assessment of Biodiversity in Forest Trees* as part of the European Union DGXII Biotechnology FW IV Research Programme *Molecular Tools for Biodiversity*. (E.M. Gillet, ed.).
<<http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>>

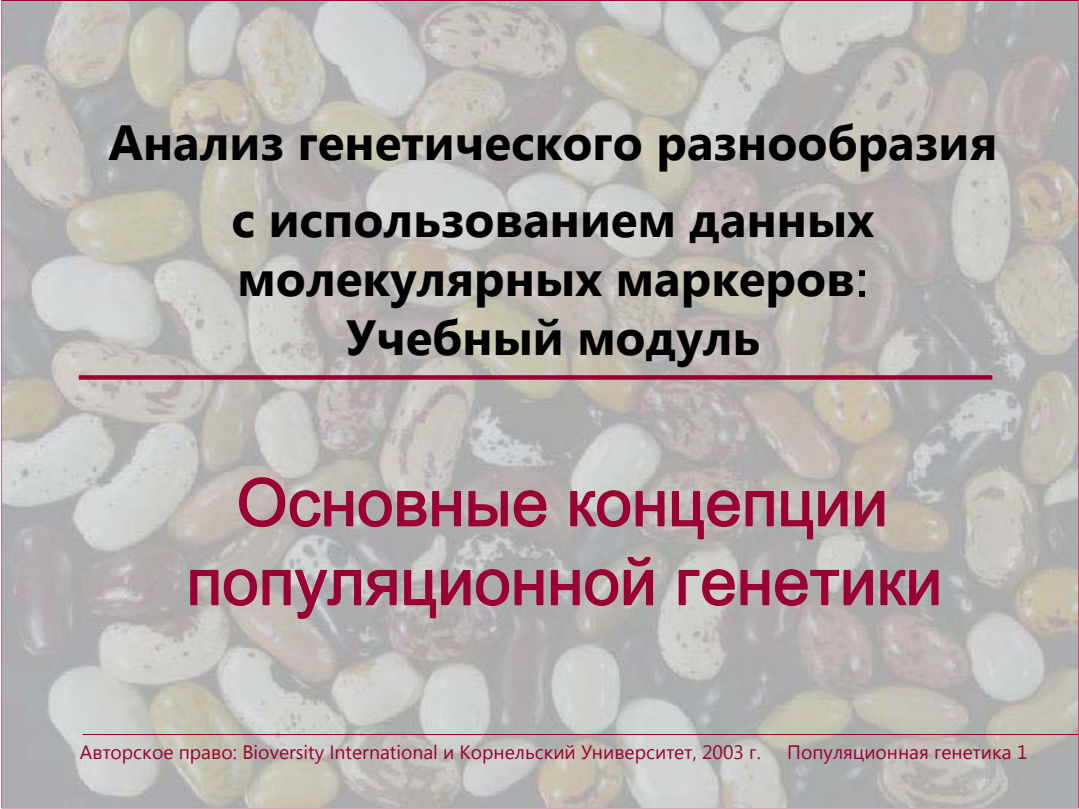


Следующее

- ▶ Основные концепции популяционной генетики
- ▶ Измерение генетического разнообразия
- ▶ Компьютерные программы для анализа генетического разнообразия
- ▶ Глоссарий

Авторские права: Bioversity International и Корнельский Университет, 2003

Введение 20



**Анализ генетического разнообразия
с использованием данных
молекулярных маркеров:
Учебный модуль**

**Основные концепции
популяционной генетики**

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 1



Содержание

- ▶ Определения
- ▶ Закон Харди-Вейнберга (The Hardy-Weinberg principle)
- ▶ Примеры подсчета частоты аллелей
- ▶ Системы репродукции и спаривания
- ▶ Силы, формирующие генетическое разнообразие
- ▶ Приложение 1: Критические значения распространения Хи-квадрата (chi-square)

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 2



Определения: Популяционная генетика

- ▶ Измерение изменчивости путем описания изменений в частоте аллелей во времени для определенного признака
- ▶ Анализ причин, вызывающих эти изменения

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 3

‘Популяционная генетика’ имеет много определений. В общем, мы можем сказать, что популяционная генетика – это изучение применения законов Менделя и других генетических принципов ко всем популяциям организмов, а не только к особям. Популяционная генетика также изучает изменения в частотах генов и, таким образом, тесно связана с эволюционной генетикой потому, что эволюция сильно зависит от изменений в частотах гена. Краткое введение в основные факторы, вероятно вызывающие изменения в генетическом разнообразии, может быть найдено на слайдах 33-42 данного раздела.

Хотя инспектирование всех генетических изменений, присутствующих в популяции, является виртуально невозможным, мы можем исследовать популяцию посредством изменений в индивидуальных фенотипах (описание определенных морфологических и фенологических признаков) и генотипах (молекулярные маркеры).



Фенотип - это ...

- ▶ Описание всех признаков особи, относящихся к ее морфологии, физиологии, экологическим взаимосвязям и поведению
- ▶ В любое определенное время, фенотип является результатом взаимодействия генов особи с окружающей средой

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 4

Фенотипические изменения могут быть или качественными (наличие или отсутствие) или количественными. Качественные признаки могут быть классифицированы. Количественные признаки могут быть измерены.

Особи принадлежат к одному и тому же фенотипу, если они выглядят подобно друг другу. Некоторые генотипы могут быть одного и того же фенотипа. Различие между генотипом и фенотипом важно для тех признаков, которые обуславливаются окружающей средой: две особи одного генотипа могут иметь разные фенотипы, обусловленные окружающей средой.

Фенотипическая изменчивость



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 5

Природные популяции различны по фенотипу. Количество фенотипического разнообразия исключительно и очевидно, даже при самом простом наблюдении.

Популяция тесно взаимосвязанных особей демонстрирует низкую изменчивость. Такая ситуация особенно опасна, если условия окружающей среды меняются и популяция не имеет вариаций для совладения с изменениями. В этом случае популяция может быстро исчезнуть.

Популяционная генетика изучает фенотипическое разнообразие, особенно если оно обусловлено различиями в генотипическом составе особей.

Ген и аллель

- ▶ Ген – это основная физическая и функциональная единица наследственности, передающая информацию от одного поколения другому
- ▶ Аллель – это любая из альтернативных форм гена, которая может существовать в единичном локусе

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 6

Не все последовательности ДНК состоят из генов. Генами являются те секции ДНК, которые выполняют известные функции. Они включают в себя секции со считанным генетическим кодом и регуляторный элемент, позволяющий считывать их код. Наличие генов устанавливается путем наблюдений за сегрегацией мутаций в потомстве от скрещиваний, полученных либо естественным, либо искусственным путем. Эти наблюдения послужили основанием для определения Грегори Менделем законов наследования в конце XIX века.

Гены могут иметь два или более аллели. В действительности, ген может иметь такое количество аллелей сколько необходимо для составления последовательности аллелей для данного гена. Аллели, принадлежащие серии, могут демонстрировать различные характеры доминирования друг над другом. Например, аллель может демонстрировать эффект доминантности, что означает, что он выражает свой фенотип, даже если он сопровождается рецессивным аллелем. Рецессивный аллель является единственным, где его генотип не выражается в гетерозиготной особи. Если аллель является кодоминантным, то его фенотипический эффект будет промежуточным в гетерозиготе и относящимся к эффекту доминанта гомозиготы и. таким образом, гомозиготного рецессива.



Генотип - это...

- ▶ Описание полного набора генов, которую особь наследует от своих родителей
- ▶ Генотип особи остается неизменным на протяжении всей ее жизни независимо от среды, окружающей и влияющей на нее

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 7

Термин «ген», в общем смысле, означает физический объект, который передается от родителей потомкам во время репродуктивного процесса и влияющий на унаследованные признаки. Генотип особи, таким образом, представляет собой сумму генов, унаследованных от родителей. Гены определяют состав протеинов, а также влияют на внешние признаки и поведение.

Генетическая изменчивость

- ▶ Генетическая изменчивость рассматривается концепцией генотипа
- ▶ Генетическая изменчивость, влияющая на признаки, существует в большинстве природных популяций, и эти признаки подвергаются воздействию со стороны аллелей многих генов в дополнение к воздействиям окружающей среды
- ▶ Проследить фенотипические различия, вызванные определенными генами, сложно

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 8

Скрытая генетическая изменчивость является даже более обширной, чем наблюдаемая в фенотипе и более настолько, что практически невозможно для двух особей в популяции иметь один и тот же генотип во всех локусах. Эта генетическая изменчивость может быть обнаружена при помощи молекулярных технологий, которые выявляют полиморфизм полезный затем в качестве генетических маркеров. Однако, даже молекулярные инструменты ограничены и, за исключением как для сравнения целых последовательностей ДНК, большинство методов ограничены до определенного количества генов или локусов. Даже в этом случае, достаточно вариаций обычно обнаруживается в образцах генов для оценки генетической изменчивости возможной для большинства популяций.

Как было упомянуто ранее, так как только маленькая часть генома обычно изучается при генетических исследованиях, то возникают вопросы о достоверности результатов для их экстраполяции на природные популяции. Это становится важным основанием для осторожного планирования экспериментов и уделения особого внимания на отбор образцов как особей, так и локусов для скрининга.



Что такое полиморфизм?

- ▶ «Наличие многих форм»
- ▶ С генетической точки зрения, полиморфизм определяется сосуществованием двух или более альтернативных фенотипов внутри популяции или среди популяций. В общем, эти разнообразные фенотипы обусловлены альтернативными аллелями одного гена
- ▶ На молекулярном уровне, полиморфизм определяется сосуществованием альтернативных связующих конфигураций или вариантов ДНК при их определении данным методом обнаружения

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 9

Говорят, что ген или фенотипический признак должен быть полиморфным, если более, чем одна форма гена или признака существует в популяции. Генетическая изменчивость, которая может вызвать эволюционные изменения, всегда присутствует.

Больше информации о полиморфизме, в общем, и молекулярном полиморфизме, в частности, приводится в обучающем модуле «Используя технологию молекулярных маркеров в изучении разнообразия генетических ресурсов растений» (www.bioversityinternational.org/publications).



Популяция - это ...

- ▶ **Экологически**
 - группа особей одного и того же вида, существующей внутри ограниченной географической области, что позволяет любым двум особям скрещиваться
- ▶ **Генетически:**
 - группа особей, разделяющих общий генофонд и имеющих возможность скрещиваться

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 10

Популяции – это исключительно сложные организмы. В популяционной генетике, фокус делается на местную единицу интербридинга более большой популяции потому, что изменения в частотах аллелей возникают внутри таких ограниченных единиц и могут привести к эволюции адаптивных признаков. Эти местные единицы интербридинга обычно называются «местные популяции», «субпопуляции» или просто «популяции». Обычно, в популяции растения вида распространены неравномерно. Подразделение популяций возникает обычно благодаря экологическими катастрофами в регионе, где они произрастают. В принципе, размер популяции не является безгранично большим и не остается постоянным.

Структура популяции

Установлены три уровня структуры популяции:

- Индивидуальные организмы
- Субпопуляции
- Вся популяция

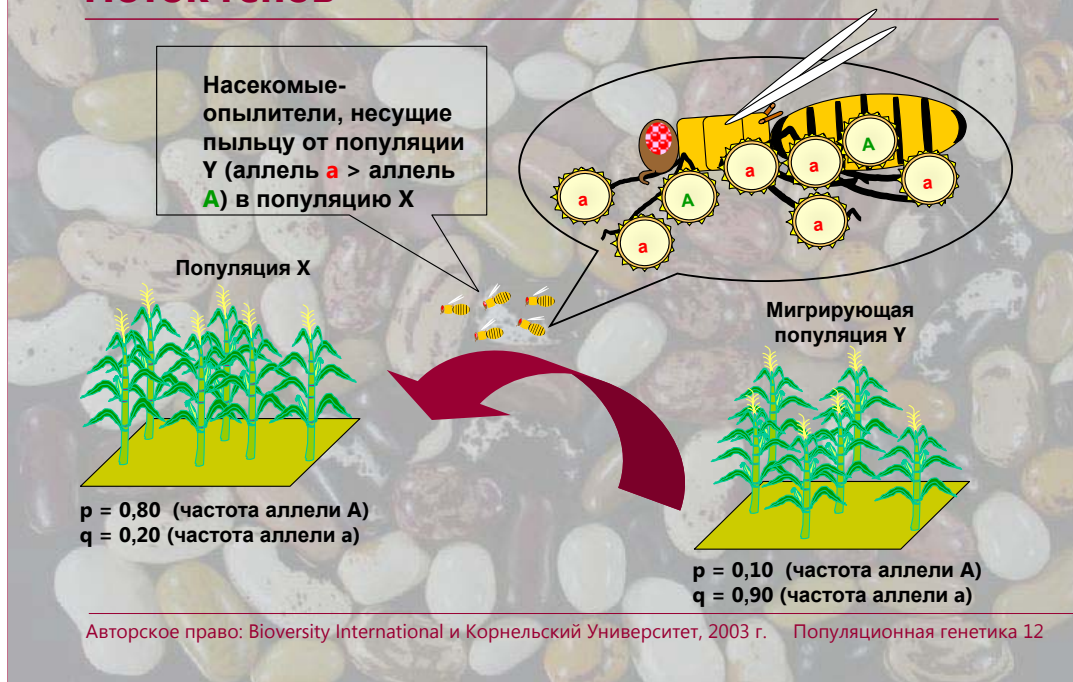
Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 11

Популяция может считаться одной единицей. Однако, для многих видов и в разных условиях, популяции подразделяются на более маленькие единицы. Такое подразделение может быть результатом экологических (прерывистый ареал) или поведенческих факторов (осознанное или неосознанное перемещение). Если популяция подразделена, то генетические связи между ее частями могут различаться в зависимости от реальной степени имеющего место потока генов.

Популяция считается структуризированной, если (1) генетический дрейф происходит в некоторых его субпопуляциях, (2) миграция не происходит равномерно по всей популяции, или (3) спаривание не происходит произвольно по всей популяции. Структура популяции влияет на степень генетической изменчивости и характер ее распространения А.

Смотрите следующие слайды и *Глоссарий* для более подробной информации о новых концепциях (например, потоке генов, миграции)

Поток генов



Поток генов - это перенос и создание генов, типичных для одной популяции, в генофонд другой путем естественной или искусственной гибридизации и обратного скрещивания (беккрасса).

На рисунке выше, популяция Y имеет более высокую частоту аллели a ($q = 0,90$). Опылители, при перелете из этой популяции в популяцию X переносят больше копий аллели a. В результате, наблюдается эффект потока генов в соответствующих поколениях популяции X как увеличение в частоте мигрирующей аллели a.

Частота аллели

- ▶ Частота аллели – это концепция, используемая для количественного определения генетической изменчивости
- ▶ Она определяется как мера общности данного аллеля в популяции и, таким образом, является удельным весом всех аллелей этого гена в популяции, являющихся специфичными для этого типа

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 13

Аллель – это альтернативная форма гена. Если ген соответствует специфичной последовательности нуклеотидов по всей молекуле ДНК, то эти аллели представляют собой разные последовательности нуклеотидов, которые возможны для этого отдельного локуса.

Часто термин «ген» используется как синоним «аллели» и, следовательно, «частота гена» иногда используется как синоним «частоте аллели». Аллельные различия в единичном локусе в популяции указывают на генетическую изменчивость. Генетическую изменчивость необходимо определить количественно для разных генов и для разных особей или популяций.

Вычисление частоты аллеля

$$P(A) = [2(AA) + (Aa)]/2n$$

- ▶ Удвоенное число гомозиготных генотипов с этим аллелем (потому что каждая гомозигота несет по две копии одного и того же аллеля),
- ▶ Плюс число гетерозиготных генотипов (потому что гетерозиготы несут только одну копию определенного аллеля),
- ▶ Деленное на удвоенное число особей в образце (потому что каждая особь несет по две аллели в каждом локусе)

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 14

Учтите, что любой результат, полученный с использованием данной формулы будет лишь оценкой общей частоты аллели в популяции, потому что обычно изучается только один образец особей. Однако, если отбор образцов особей произведен хорошо, то размер образца достаточно большой, и тогда можно предположить, что наши расчеты близки к истинной частоте аллели. Как и в правиле большого пальца, оценка частоты аллели должна быть сделана, где возможно, на образцах, взятых с 100 особей или больше.



Частота генотипа

- ▶ Это частота данного генотипа в популяции
- ▶ Частоты различных типов селекционных систем определяют математическую взаимосвязь между частотами аллеля и генотипа

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 15

Система естественного скрещивания особей может быть изучена путем исследований частот, с которыми альтернативные генотипы возникают в популяции. Когда популяция подвергается скрещиванию, которое произвольно относится к заинтересованным аллелям, определенные характеры частот генотипа могут ожидаться.

Частоты генотипа также используются для оценки количества самоопыления, возникающего в популяциях особей, которые обладают данным или смешанным типом размножения. Система спаривания или скрещивания, таким образом, имеет значительное влияние на частоту возникновения альтернативных генотипов в данной популяции.

Закон Харди-Вейнберга

- ▶ В популяции со случайным скрещиванием результат заключается в сбалансированном распространении генотипов уже через одно поколение, и таким образом, можно заключить, что генетическая изменчивость поддерживается
- ▶ Когда предположения отвечают требованиям, частота генотипа равна частотам аллели

$$\begin{array}{ccc} AA & Aa & aa \\ p^2 & 2pq & q^2 \end{array}$$

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 16

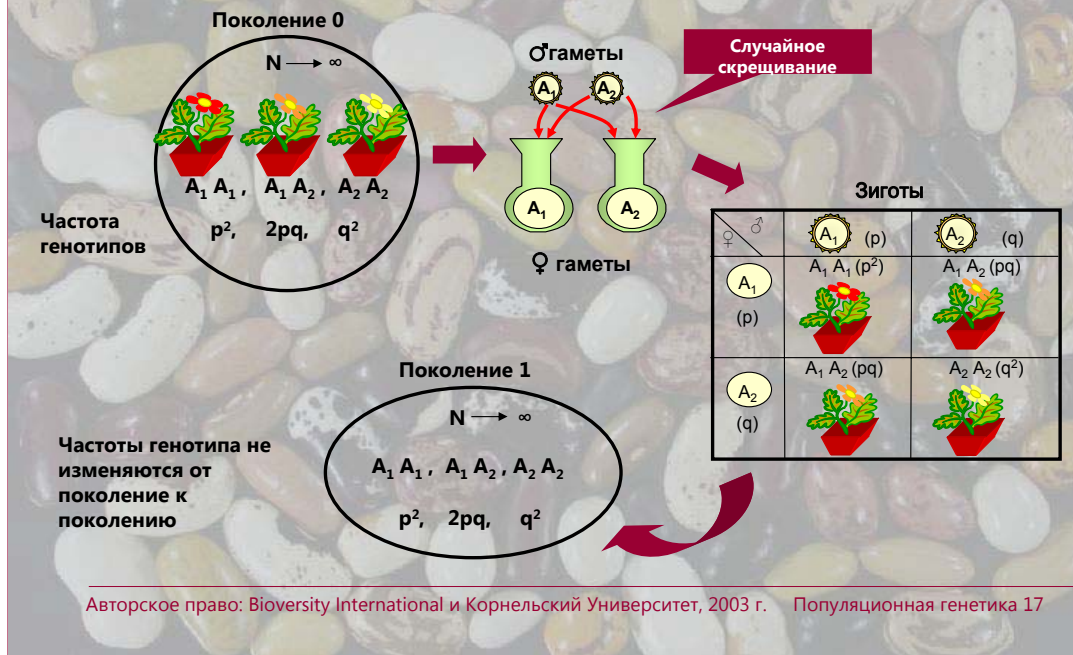
Равновесие Х-В утверждает, что половое размножение не уменьшает наследственную изменчивость от поколения к поколению; напротив, количество изменчивости остается постоянным, если отсутствуют возмущающие силы, действующие против него. Оно устанавливает взаимосвязь для вычисления частот генотипа при случайном скрещивании и, при этом, обеспечивает основу для многих исследований в популяционной генетике.

Этот принцип описывает вероятность для частот аллели в идеальной ситуации, где:

- Организм диплоиден
- Размножение половое
- Поколения не перекрываются
- Скрещивание случайное
- Размер популяции очень большой
- Миграция незначительна
- Мутации можно проигнорировать
- Естественный отбор не влияет на изучаемые аллели

Учтите, что большинство культурных растений нарушают по крайней мере одно из этих предположений.

Демонстрация принципа Х-В



Отправной точкой является поколение 0. Мы имеем ген с двумя аллелями A_1 и A_2 . Частота аллели A_1 равна p и частота аллели A_2 - q . Частоты генотипа в поколении 0 для $A_1 A_1 = p^2$, для $A_1 A_2 = 2pq$ и для $A_2 A_2 = q^2$. Если происходит случайное спаривание, то вероятность встречи для любого аллеля женского растения с любым аллелем мужского растения будет одинаковой. Таблица справа на слайде изображает четыре возможных генотипа для следующего поколения. Частота возникновения каждого генотипа приводится по продукту частоты каждого аллеля в генотипе, например, для $A_1 A_1$ он равен $p \times p = p^2$. Если суммировать результаты в таблице, что в голубой вкладке внизу слайда, то мы увидим, что частоты генотипа в поколении 1 остаются такими же, как и в предыдущем поколении.

Пример трех популяций в равновесии Харди-Вейнберга

Частоты

Популяции	Генотипы G_0			G_0		Генотипы G_1			G_1	
	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q
Популяция 1	0,6	0,2	0,2	0,7	0,3	0,49	0,42	0,09	0,7	0,3
Популяция 2	0,49	0,42	0,09	0,7	0,3	0,49	0,42	0,09	0,7	0,3
Популяция 3	0,4	0,6	0,0	0,7	0,3	0,49	0,42	0,09	0,7	0,3

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 18

Частоты генотипа популяции приведены в строках. Популяции G_0 и G_1 даны в графах. Снова мы имеем один ген с двумя аллелями A_1 и A_2 . Частота аллели A_1 равна p и частота аллели A_2 - q . Частоты генотипа разные для каждой популяции в поколении 0, например частота $A_1 A_1$ в популяции 1 равна 0,6; в популяции 2 - 0,49; и в популяции 3 - 0,4, и так далее для каждого генотипа. Мы учли, однако, что частоты аллели в трех популяциях схожи в G_0 ($p = 0,7$ и $q = 0,3$). В следующем поколении G_1 , если все необходимые условия закона Х-В соблюдены, то частоты генотипа в трех популяциях равны (теперь частота $A_1 A_1$ равна 0,49 в трех популяциях, и то же самое происходит с частотами $A_1 A_2$ и $A_2 A_2$). Частоты аллели сохранены.

Анализ Хи - квадрата

- ▶ Этот гипотетический тест полезен для определения нахождения аллельных частот в равновесии Х-В
- ▶ Процедура заключается в следующем:
 - Определить H_0 (и H_a)
 - Определить уровень значения α
 - Провести статистический тест
 - Применить критерии принятия решений

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 19

H_0 = гипотеза, говорящая, что частоты аллели для признака Q в данной популяции находятся в равновесии Х-В.

H_a = альтернативная гипотеза, говорящая, что частоты аллели для признака Q не находятся в равновесии Х-В.

Мы выбираем значительный уровень , что дает нам определенную долю уверенности в наших результатах. Статистический тест проводится по формуле

$$ST = \left[\frac{\sum [O_i - E_i]^2}{(E_i)} \right] = \chi^2_{k - mdf}$$

где:

ST = статистический тест
O = наблюдаемые частоты
E = ожидаемые частоты

k = число генотипических классов
m = число аллелей
df = степень свободы

Если наш пример предусматривает только 1 степень свободы и разница в частотах сокращена на 0,5, то фактор коррекции равен:

$$\frac{[|O_i - E_i| - 0.5]^2}{E_i}$$

Критерий принятия решений применяется следующим образом:

If $\chi^2_{cal} \leq \chi^2_{tab}$, то H_0 принимается; и, если $\chi^2_{cal} > \chi^2_{tab}$, то H_0 отвергается

где,

cal = результат вычисления ST по данным, полученным в нашем примере

tab = значение, определенное по таблице

(а χ^2 таблицу можете найти в Приложении 1, для этого щелкните мышкой [здесь](#)).

Применение теста χ^2 : пример

Генотип	AA	Aa	aa	
Число наблюдаемых	169	520	311	
Число расчетных	250	500	250	
Расчет χ^2	$(169-250-0,5)^2/250$	$(520-500-0,5)^2/500$	$(311-250-0,5)^2/250$	
Значение χ^2	+25,921	+0,760	+14,641	41,322



Критерий решения:

$$\chi^2_{\text{cal}} (41,322) > \chi^2_{\text{tab}} (3,8)$$

H_0 отклонен

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 20

В этом примере, давайте условимся, что частоты аллели были 0,429 для A_1 и 0,571 для A_2 . Каждый генотипический класс был представлен, как в таблице на слайде выше.

Гипотезы для тестирования следующие:

H_0 = это популяция в Х-В равновесии для ее частот аллели

H_a = это популяция не в равновесии Х-В для ее частот аллели

Так как число генотипических классов равно 3, и таким образом, мы имеем только 1 степень свободы, то мы применяем поправочный фактор в наших расчетах элементов χ^2

χ^2 равен 41,322. С этим значением и допустимой погрешностью в 0,05%, статистические данные отклоняют H_0 , что означает, что популяция не находится в равновесии Х-В для данного изучаемого признака.

Вычисление частот аллелей: примеры

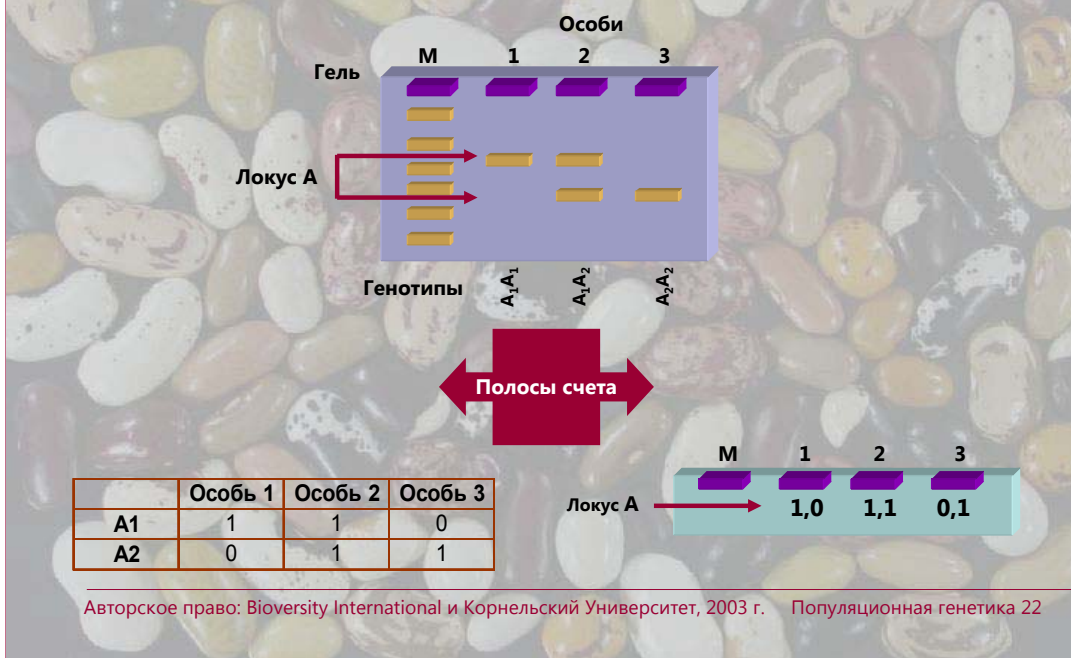
- ▶ Вычисление частот аллелей при помощи кодоминантного маркера
- ▶ Вычисление частот аллелей при помощи доминантного маркера
- ▶ Вычисление частот аллелей при помощи кодоминантного гена, имеющего множество аллелей

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 21

В следующих нескольких слайдах мы покажем примеры расчета частот аллели для результатов, полученных при помощи разных типов маркеров. Примеры приводятся с изображениями, которые имитируют натуральные гели и полосы, полученные с применением молекулярных маркеров. Для более подробной информации о технологиях молекулярных маркеров и расшифровки полос смотрите обучающий модуль «*Используя технологию молекулярных маркеров в изучении разнообразия генетических ресурсов растений*»

(http://www.bioversityinternational.org/publications/pubfile.asp?ID_PUB=912).

... при помощи кодоминантного маркера



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 22

(продолжение в следующем слайде)

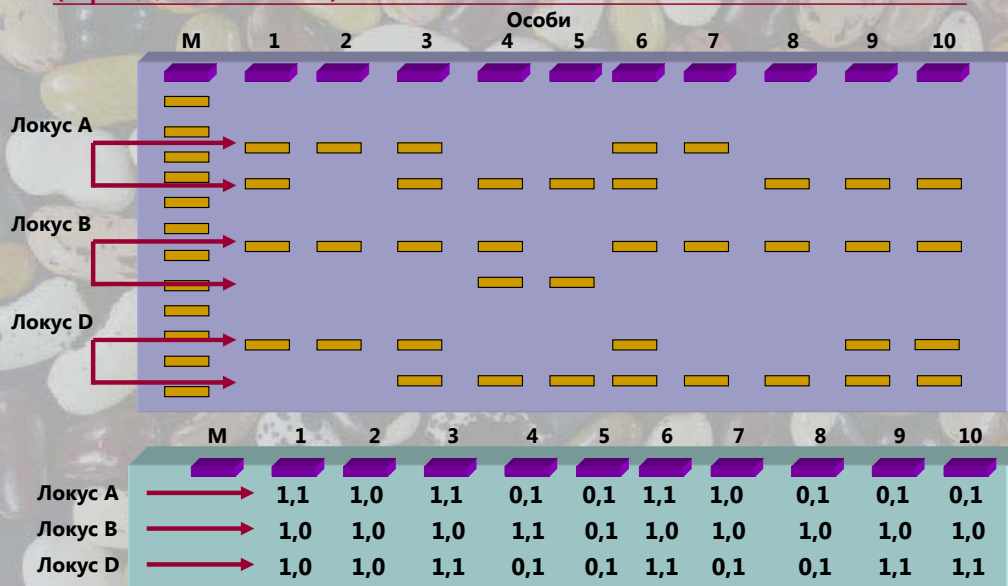
При помощи кодоминантного маркера генотипы трех генотипических классов могут быть изучены для двух гомозигот и одной гетерозиготы. На рисунке выше, в центре наверху, мы видим изображение геля с объединяющей схемой кодоминантного маркера для единичного локуса диплоидного организма. Нам необходимо подсчитать полосы в геле и преобразовать их в числа. Для этого, каждый из размеров полосы (полоса в одной и той же строке) подсчитывается и преобразуется в единицу, если она представлена или в ноль, если отсутствует. Мы можем сделать это по полосе, как в таблице внизу слева, или по генотипу, как в правом углу внизу. В таблице внизу мы можем видеть расчеты предполагаемых и наблюдаемых частот генотипа, а также частот аллели (M = маркер молекулярного веса)

Генотипы	A ₁ A ₁	A ₁ A ₂	A ₂ A ₂	Всего
Частота генотипа (ожидаемое)	p ²	2pq	q ²	1
Число особей	n ₁₁ = 40	n ₁₂ = 20	n ₂₂ = 140	n = 200
Частота генотипа (наблюдаемое)	P ₁₁ = n ₁₁ /n = 0,20	P ₁₂ = n ₁₂ /n = 0,10	P ₂₂ = n ₂₂ /n = 0,70	1

$$p = (2n_{11}/2n) + (n_{12}/2n) = P_{11} + \frac{1}{2} P_{12} = 0,20 + \frac{1}{2} (0,10) = 0,25$$

$$q = (2n_{22}/2n) + (n_{12}/2n) = P_{22} + \frac{1}{2} P_{12} = 0,70 + \frac{1}{2} (0,10) = 0,75$$

... при помощи кодоминантного маркера (продолжение)



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 23

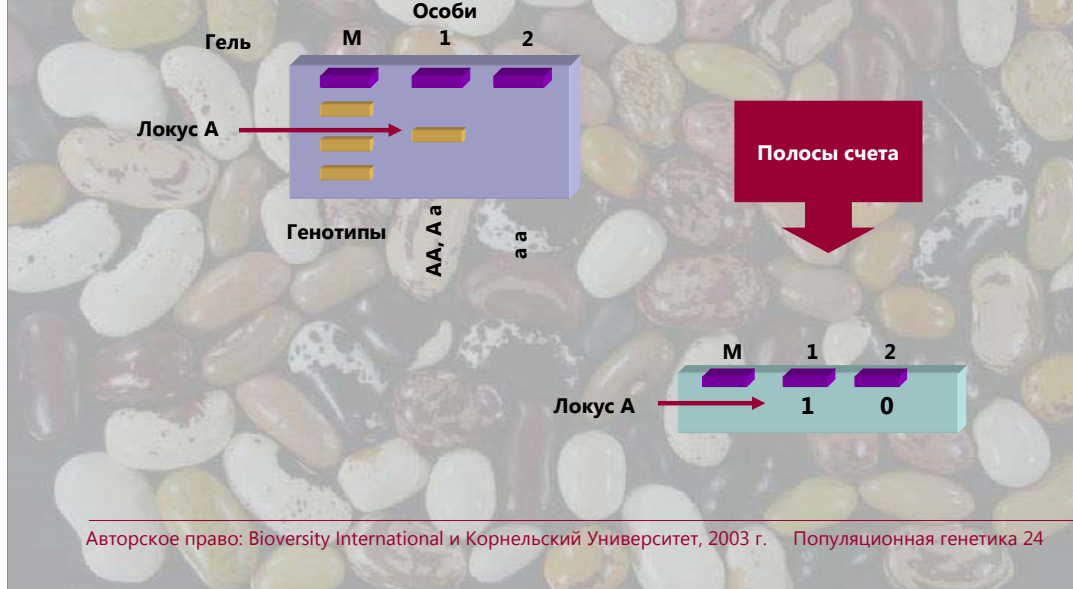
Пример в этом слайде подобен тому, что в предыдущем слайде, но для 10 особей и трех выделенных локусов (А, В и D). Для более легкой демонстрации, использован только один метод подсчета, в нижней части слайда. (М = маркер молекулярного веса)

Учтите, что гель, как в нашем примере, может быть получен только при помощи мультиплексирования, что означает помещение в одну и ту же лунку продуктов реакции по разным маркерам.

Внизу мы видим таблицу с расчетами частот генотипа и аллели.

Локус	Анализ данных					Частота аллеля	
	Генотипы	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Итого	p	q
А	7	p^2	$2pq$	q^2	1		
	Число особей	2	3	5	10		
	Частота генотипа (наблюдаемая)	$P_{11} = 0,2$	$P_{12} = 0,3$	$P_{22} = 0,5$	1	0,35	0,65
	Генотипы	$B_1 B_1$	$B_1 B_2$	$B_2 B_2$	Итого	p	q
Частота генотипа (ожидаемая)	p^2	$2pq$	q^2	1			
В	Число особей	8	1	1	10	p	q
	Частота генотипа (наблюдаемая)	$P_{11} = 0,8$	$P_{12} = 0,1$	$P_{22} = 0,1$	1		
	Генотипы	$D_1 D_1$	$D_1 D_2$	$D_2 D_2$	Итого	p	q
	Частота генотипа (ожидаемая)	p^2	$2pq$	q^2	1		
D	Число особей	2	4	4	10	p	q
	Частота генотипа (наблюдаемая)	$P_{11} = 0,2$	$P_{12} = 0,4$	$P_{22} = 0,4$	1		

... при помощи доминантного маркера



(продолжение в следующем слайде)

При помощи доминантного маркера можно исследовать только два генотипических класса: $AA + Aa$ и aa , так как один из гомозиготных классов смешался с гетерозиготным. Изображение геля со связывающей схемой доминантного маркера для единичного локуса покажет либо одну полосу, либо отсутствие полос для каждой особи. Полосы подсчитываются тем же способом как и для кодоминантного маркера, где полосы были конвертированы в 1, если полоса присутствовала или 0 в случае ее отсутствия. (M = маркер молекулярного веса)

Расчеты частот произведены как показано в таблице внизу (p , q = частоты аллели.)

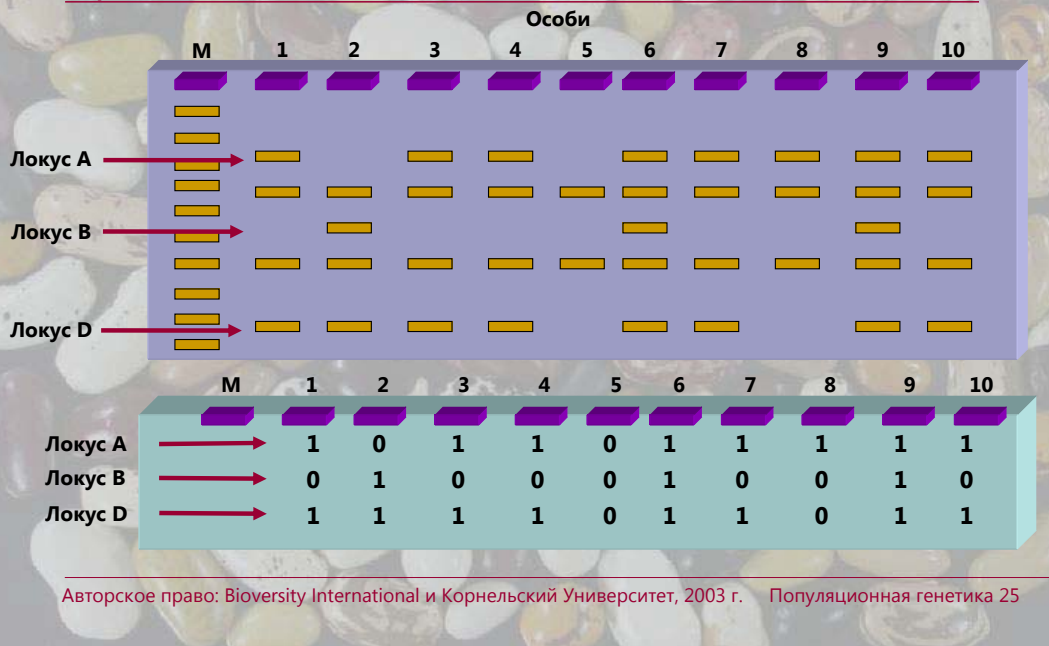
Фенотипы	A _		aa	Всего
	AA	Aa	aa	
Генотипы	AA	Aa	aa	
Частоты фенотипа (ожидаемое)	$p^2 + 2pq$		q^2	1
Число особей	$n_1 = 84$		$n_2 = 16$	$n = 100$
Частоты фенотипа (наблюдаемое)	$P_1 = n_1/n = 0,84$		$P_2 = n_2/n = 0,16$	1

$$q = \sqrt{(n_2/n)} = \sqrt{(P_2)} = \sqrt{(0,16)} = 0,4$$

$$p = (1 - q) = 0,6$$

Это смещенная оценка, так как она не учитывает рецессивные аллели в гетерозиготах

... при помощи доминантного маркера (продолжение)



Здесь мы имеем пример подобный тому, что был в предыдущем слайде, но с 10 особями и тремя выделенными локусами (А, В, и D). (M = маркер молекулярного веса)

Несегрегирующие (не разделяющиеся) полосы (мономорфические) не подсчитываются и поэтому не включаются в анализ.

Внизу мы видим таблицу с расчетами частот генотипа и аллели.

Локус	Анализ данных				Частота аллели	
	Генотипы	$A_1 _$	$A_2 A_2$	Итого	p	q
А	Частоты генотипа (ожидаемая)	$p^2 + 2pq$	q^2	1	p	q
	Число особей	8	2	10		
	Частоты генотипа (наблюдаемая)	$P_1 = 0,8$	$P_2 = 0,2$	1	0,55	0,45
	Генотипы	$B_1 _$	$B_2 B_2$	Итого	p	q
Частоты генотипа (ожидаемая)	$p^2 + 2pq$	q^2	1			
В	Число особей	3	7	10	p	q
	Частоты генотипа (наблюдаемая)	$P_1 = 0,3$	$P_2 = 0,7$	1		
	Генотипы	$D_1 _$	$D_2 D_2$	Итого	p	q
	Частоты генотипа (ожидаемая)	$p^2 + 2pq$	q^2	1		
D	Число особей	8	2	10	p	q
	Частоты генотипа (наблюдаемая)	$P_1 = 0,8$	$P_2 = 0,2$	1		

Мы не можем различить гетерозиготы, но мы можем оценить ожидаемое число гетерозигот в популяции. Например, если размер образца = 1000, тогда:
 для локуса А, число ожидаемых гетерозигот = $2pqN = 2(0,55)(0,45)(1000) = 495$
 для локуса В, число ожидаемых гетерозигот = $2pqN = 2(0,16)(0,84)(1000) = 269$
 и так далее ...

... при помощи кодоминантного маркера, имеющего множество аллелей

Генотипы	$A_1 A_1$	$A_1 A_2$	$A_1 A_3$	$A_2 A_2$	$A_2 A_3$...	$A_n A_n$	Итого
Частоты генотипа (расчетные)	p_1^2	$2p_1p_2$	$2p_1p_3$	p_2^2	$2p_2p_3$...	p_n^2	1
Число особей	n_{11}	n_{12}	n_{13}	n_{22}	n_{23}	...	n_{nn}	n
Частоты генотипа (наблюдаемые)	$P_{11} = n_{11}/n$	$P_{12} = n_{12}/n$	$P_{13} = n_{13}/n$	$P_{22} = n_{22}/n$	$P_{23} = n_{23}/n$...	$P_{nn} = n_{nn}/n$	1

$$p_1 = P_{11} + \frac{1}{2} \sum_{i \neq 1} P_{1i}$$

$$p_2 = P_{22} + \frac{1}{2} \sum_{j \neq 2} P_{2j}$$

$$p_3 = P_{33} + \frac{1}{2} \sum_{j \neq 3} P_{3j}$$

$$p_4 = P_{44} + \frac{1}{2} \sum_{j \neq 4} P_{4j}$$

$$p_n = P_{nn} + \frac{1}{2} \sum_{j \neq n} P_{nj}$$

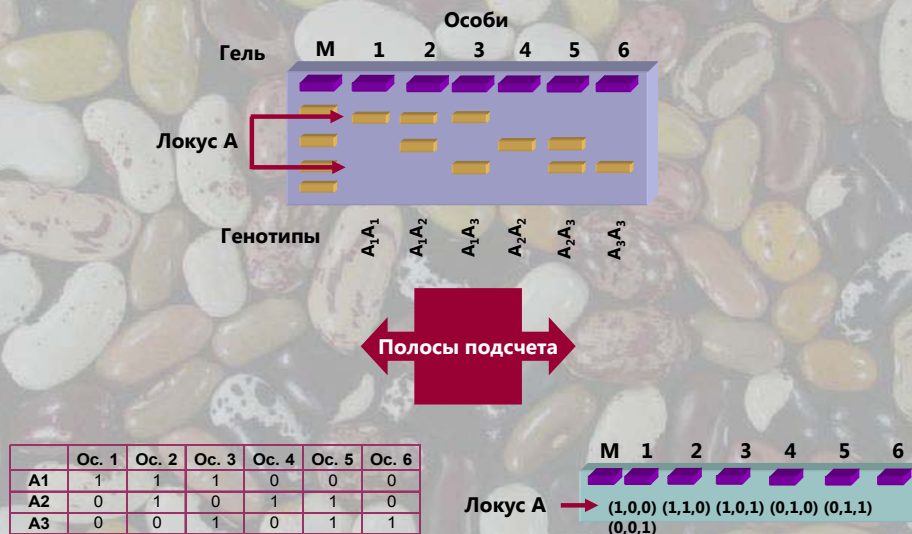
Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 26

(продолжение в следующем слайде)

Это ситуация, возникающая при использовании таких маркеров, как микросателлиты.

Мы имеем локус A с n аллелей $A_1, A_2, A_3, \dots, A_n$ и частотами аллели $p_1, p_2, p_3, \dots, p_n$, соответственно, что $A_1 = A_2 = A_3 = \dots = A_n$

**... при помощи кодоминантного маркера,
имеющего множество аллелей
(продолжение)**



Авторское право: Bioersity International и Корнельский Университет, 2003 г. Популяционная генетика 27

Снова, при помощи кодоминантного маркера генотипы трех генотипических классов могут быть изучены. На рисунке сверху, в центре верхней части, мы видим, изображение геля со связывающей схемой кодоминантного маркера с тремя аллелями (A_1 , A_2 и A_3) в диплоидном образце. Мы подсчитываем каждую полосу (каждый ряд) самостоятельно и преобразуем их в 1 в случае присутствия и 0 в случае отсутствия. Мы можем проделать это по полосам (левый угол слайда) или по генотипу (правый угол слайда). В таблице внизу, мы можем видеть расчеты ожидаемых и наблюдаемых частот генотипа, а также частот аллели (p_1 , p_2 и p_3). (M = маркер молекулярного веса)

Генотипы	$A_1 A_1$	$A_1 A_2$	$A_1 A_3$	$A_2 A_2$	$A_2 A_3$	$A_3 A_3$	Total
Частота генотипа (ожд.)	p_1^2	$2p_1p_2$	$2p_1p_3$	p_2^2	$2p_2p_3$	p_3^2	1
Число особей	$n_{11} = 4$	$n_{12} = 6$	$n_{13} = 0$	$n_{22} = 10$	$n_{23} = 2$	$n_{33} = 2$	$n = 24$
Частота генотипа (наблюд.)	$P_{11} = n_{11}/n = 0,17$	$P_{12} = n_{12}/n = 0,25$	$P_{13} = n_{13}/n = 0$	$P_{22} = n_{22}/n = 0,42$	$P_{23} = n_{23}/n = 0,08$	$P_{33} = n_{33}/n = 0,08$	1

$$p_1 = P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{13} = P_{11} + \frac{1}{2}\sum_{j \neq 1} P_{1j} = 0,17 + \frac{1}{2}(0,25 + 0,00) = 0,30$$

$$p_2 = P_{22} + \frac{1}{2}P_{21} + \frac{1}{2}P_{23} = P_{22} + \frac{1}{2}\sum_{j \neq 2} P_{2j} = 0,42 + \frac{1}{2}(0,25 + 0,08) = 0,59$$

$$p_3 = P_{33} + \frac{1}{2}P_{31} + \frac{1}{2}P_{32} = P_{33} + \frac{1}{2}\sum_{j \neq 3} P_{3j} = 0,08 + \frac{1}{2}(0,00 + 0,08) = 0,12$$

Размножение и системы скрещивания

- ▶ Аутбридинг, инбридинг или неполовое размножение
- ▶ Они влияют на:
 - Степень генетического родства среди скрещивающихся особей
 - Организацию генов скрещивающихся особей в генотипах

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 28

В принципе, аутбридинг возникает как произвольное спаривание, а инбридинг и неполовое размножение включают в себя неслучайное скрещивание.

Ауткроссинг видов, при сравнении с инбридами, может сохранять большое число вредных рецессивов, потому что положение доминирования скрывает их. Рецессивы подвергаются частой рекомбинации, в результате чего образуются новые типы гамет.

Доминирование означает ситуации, где в гетерозиготных условиях, один аллель имеет эффект фенотипа, который достаточно силен для утаивания присутствие другого (рецессивного) аллеля. В ситуации доминирования только два генотипа могут наблюдаться: доминантный фенотип, являющийся смесью гомозиготного и гетерозиготного доминанта, и рецессивный фенотип.

В перекрестно-опыляющихся видах, самоопыление ведет к инбридинговой депрессии, потому что оно увеличивает долю гомозигот, таким образом, позволяя редким рецессивным аллелям стать видимыми. Гетерозиготы в перекрестно-опыляющихся видах имеют более благоприятный эффект.

В организмах с неполовым размножением, ряд последствий имеет место в зависимости от его типа. Хотя неполовое размножение может быть постоянным способом размножения, оно может сочетаться с циклами полового размножения, которое позволяет рекомбинацию данной мутации и, как таковое, образование новых форм или комбинаций. Если происходит только неполовое размножение в популяции, то генотипические частоты не могут быть изменены.



Случайное спаривание

- ▶ Спаривание, которое происходит случайно означает, что шансы особи А скреститься с особью В не зависят от генотипов обеих особей
- ▶ Если случайное спаривание происходит, то шанс особи спариться с данным генотипом равен частоте этого генотипа в популяции

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 29

Произвольное спаривание типично для многих аутбридинговых популяций. Например, мы можем иметь популяцию, в которой генотип AA представлен 10%, Aa - 58% и aa - 32%. Если спаривание произвольное, тогда шансы особи AA спариться с другой особью AA равны $10/100$, Aa - $58/100$, и aa - $32/100$.

Неслучайное спаривание

▶ Классифицированное спаривание:

- Положительное: спаривание среди особей со схожими генотипами
- Отрицательное: спаривание среди особей с несхожими генотипами



▶ Инбридинг:

- Спаривание среди родственников



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 30

Непроизвольное спаривание возникает, когда более близко (инбридинг) или менее близко родственные особи спариваются чаще, чем можно было ожидать для популяции.

Самоопыление или инбридинг сходен с спариванием между родственниками. Оно увеличивает гомозиготность популяции и его результат распространяется на все аллели. Инбридинг сам по себе не изменяет частоты аллели, но, со временем, он приводит к гомозиготности путем медленного увеличения двух гомозиготных класса.

Коэффициент инбридинга

- ▶ Он представляет действительное отношение числа гетерозиготных генотипов к их расчетному количеству при случайном спаривании

$$F = \frac{(H_0 - H)}{H_0}$$

- ▶ F является коэффициентом инбридинга и показателем уменьшения гетерозиготности

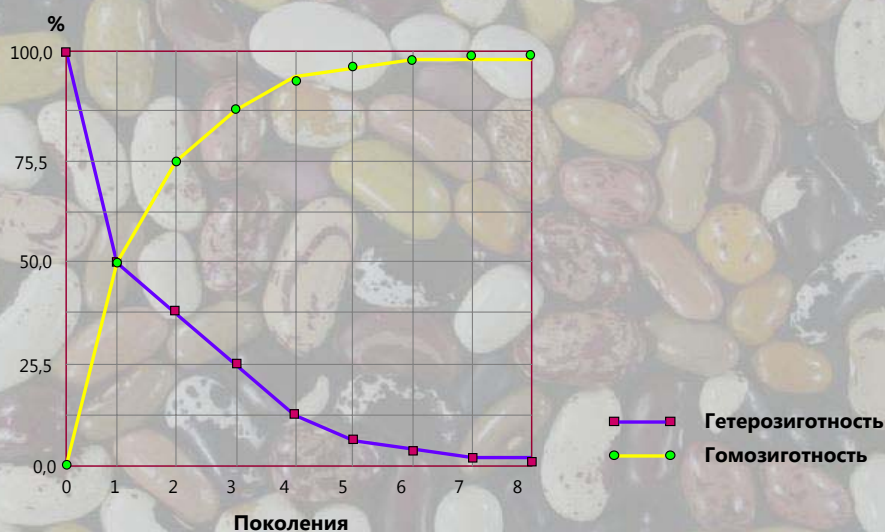
Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 31

H = фактическая частота гетерозигот в популяции

H₀ = предполагаемое число гетерозигот при произвольном скрещивании

Коэффициент инбридинга указывает на степень инбридинга в популяции

Что происходит при самоопылении?



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 32

Самоопыление является мощной инбридинговой системой, которая позволяет достичь высоких уровней гомозиготности внутри нескольких поколений. Одновременно, гетерозиготность уменьшается. Диаграмма на слайде демонстрирует данное явление, и в таблице внизу мы можем увидеть изменяющиеся значения гомозиготности и гетерозиготности в девяти поколениях (от G_0 до G_8).

Поколение	Доля самоопылившихся генотипов/поколение	Гомозиготность (%)	Гетерозиготность (%)
G_0	Aa	0	100
G_1	1AA, 2Aa, 1aa	50	50
G_2	6AA, 4Aa, 6aa	75	25
G_3	28AA, 8Aa, 28aa	87,5	12,5
G_4	120AA, 16Aa, 120aa	93,75	6,25
G_5	496AA, 32Aa, 496aa	96,875	3,125
G_6	2016AA, 64Aa, 2016aa	98,4375	1,5625
G_7	8128AA, 128Aa, 8128aa	99,21875	0,78125
G_8	32640AA, 256Aa, 32640aa	99,60938	0,390625



Факторы, формирующие генетическое разнообразие

- ▶ Мутация
- ▶ Миграция
- ▶ Рекомбинация
- ▶ Отбор
- ▶ Дрейф

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 33

Если, по любым причинам, популяция становится гомогенной, то эволюция не возникает. Таким образом, постоянное изменение, по существу, зависит от новой вариации.

Генетическая популяция – это совокупность аллельных частот всех генов в популяции. Популяции изменяются или эволюционируют, потому что частоты их генов испытывают изменение. Ряд факторов могут вызвать изменения в приспособленности – способности особи оставаться в живых, пока не произойдет размножение. Если приспособленность особи в популяции изменится, то генотипы в соответствующем поколении не будут напрямую связаны с частотами гена предыдущей популяции, таким образом, заставляя популяцию эволюционировать.

Так как изменения в популяции требуют изменений в частотах гена, то мы должны понять, как эти частоты могут изменяться. В следующих слайдах мы обсудим основные причины изменений: мутацию, миграцию, рекомбинацию, отбор и дрейф.

Мутация

- ▶ Это основной источник изменений и может быть вызван :
 - Ошибками в репликации ДНК
 - Радиационными повреждениями
- ▶ Мутация увеличивает разнообразие, но так как самопроизвольные мутации происходят редко, уровень изменений в частоте гена очень низок
- ▶ Следовательно, сама по себе мутация не двигает эволюцию популяций и видов

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 34

Самой простой мутацией является та, которая производит изменение в единичном нуклеотиде в последовательности ДНК гена. Мутация может вызвать изменение одного аллеля в другой, уже имеющийся в популяции, (от доминантного до рецессивного), или она может создать совершенно новый аллель. Мутации могут быть вредными или благоприятными. Но если они хороши для особи, тогда частоты аллели будут увеличиваться от поколения к поколению. Более того, эта мутация может мигрировать в другую популяцию и распространиться.

Примечание: Геномы могут проходить через процесс, известный как «удвоение гена». Это явление помогает особи выдержать неблагоприятную мутацию в копии гена без значительных трудностей, потому что другая копия гена все еще может функционировать соответствующим образом. Более обширные изменения могут оказать влияние на мутированный ген и обусловить разные типы адаптации особи.

Миграция

- ▶ Это движение особей или любая форма интродукции генов из одной популяции в другую
- ▶ Миграция увеличивает разнообразие и ее уровень может быть большим, вызывая значительные изменения в частоте
- ▶ Изменение в частоте гена пропорционально разнице частот популяции-получателя и среднего значения популяций-доноров

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 35

С генетической точки зрения, миграция означает не только движение особей в новые популяции, но и то, что это движение вводит новые аллели в популяцию (поток генов). Изменения в частотах гена будут возникать посредством миграции либо потому, что больше копий уже присутствующего аллеля, будут принесены, либо потому, что прибудет новый аллель.

Ряд факторов оказывают влияние на миграцию сельскохозяйственных культур:

- Селекционная система
- Родство с дикими или сорными сородичами
- Опылители
- Распространение семян

Непосредственный результат миграции заключается в увеличении генетической изменчивости популяции, и таким образом, в увеличении возможностей для популяции противостоять изменениям окружающей среды. Миграция также помогает смешивать популяции и предотвращать их дивергенцию.

Рекомбинация

- ▶ Это процесс, на основании которого клетка генерирует новые комбинации хромосом, отличных от данной клетки или клеток предков
- ▶ Она не создает новое разнообразие, но производит новые комбинации существующего разнообразия
- ▶ Генетическое изменение посредством рекомбинации при условии, что сегрегирующие аллели существуют в разных локусах, может произойти очень быстро

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 36

Генетическое разнообразие посредством рекомбинации есть результат перегруппировки генетических компонентов первоначального типа. Существуют механизмы для генерирования аллельного разнообразия (внутригенная рекомбинация) и геномного разнообразия (новые мультигенные комбинации)

Отбор

Это унаследованная способность организмов выживать и воспроизводиться. Отбор происходит таким образом, что лучшие генотипы увеличивают свою частоту в популяции со временем

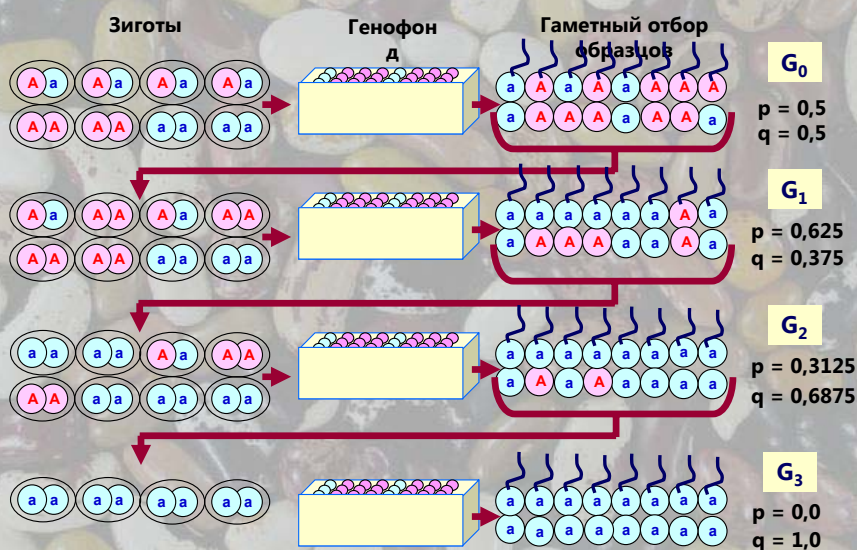
Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 37

В результате мутации развиваются новые формы. Эти формы, как объяснено, могут быть благоприятными или вредными для способности особей выжить. Если изменения являются полезными, тогда новые аллели, будучи отбираемыми в популяции, будут иметь тенденцию господствовать.

Воздействие отбора на разнообразие может быть:

- Направленное, где он уменьшает разнообразие
- Уравновешенным, где он увеличивает разнообразие. Гетерозиготы имеют самую высокую степень адаптации, и поэтому отбор благоприятствует поддержанию многочисленных аллелей.
- Зависимым от частоты, где он увеличивает разнообразие. Приспособленность является функцией частоты аллели или генотипа и изменений со временем.

Дрейф



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 38

Генетический дрейф означает колебания в частотах аллели, которые возникают случайно (особенно в маленьких популяциях) в результате случайного отбора среди гамет.

Дрейф уменьшает разнообразие внутри популяции, потому что он имеет тенденцию вызывать потерю редких аллелей, сокращая общее число аллелей.

В примере вверху, размер популяции постоянен в каждом поколении (8 особей). Каждая особь может произвести тысячи гамет, но из общего генофонда в каждом поколении требуется только $2N$ гамет на каждую особь (16 в нашем примере). Эта ситуация подобна экстрагированию маленьких образцов из двух ящиков, где в одном содержится миллион белых мячей, а в другом - миллион красных. Мы смоделируем схожую ситуацию, и в поколении G_0 10 гамет из тысячи возможных будут нести аллель A и только шесть гамет нести аллель a. В поколении G_1 из гамет, участвующих в построении зигот для следующего поколения, 5 несут аллель A и 11 - аллель a, и так далее. Эти значения варьируют произвольно. В поколении G_3 все особи сформированы аллелем a (гомозиготы) и аллель A утерян.

Эффективный размер популяции

- ▶ N_e – это число родителей, ответственных за генетический состав следующего поколения
- ▶ N_e в общем меньше, чем N из-за
 - изменения размера популяции от поколения к поколению
 - неравного соотношения полов
 - наличия перекрывающихся поколений
 - географической разбросанности популяции

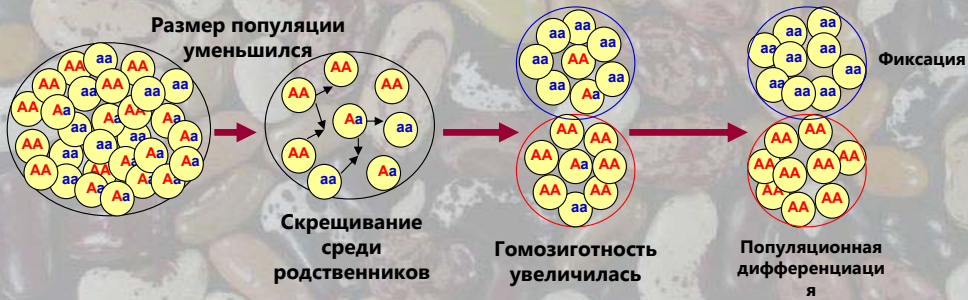
Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 39

Как велика популяция?

- Фактическое число особей в популяции называется числом учета (N). Это число часто является неточным представлением размера популяции с точки зрения генетики.
- Эффективный размер популяции (N_e) описывает размер идеальной популяции, которая показывает тот же уровень потери генетического разнообразия благодаря генетическому дрейфу, как и его увеличение для популяции.

Последствия увеличивающегося размера популяции

- ▶ Генетический дрейф со случайными изменениями частот аллелей
- ▶ Инбридинг
- ▶ Гомозиготность: фиксация и потеря аллелей
- ▶ Субпопуляционная дифференциация p



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 40

(продолжение в следующем слайде)

Факторы, приводящие к уменьшению размера популяции:

- Одомашнивание
- Субпопуляции (инбридинг, клональное размножение)
- Распространение на большие расстояния (эффект основателя)
- Возобновление коллекций генетических ресурсов

Последствия ... (продолжение)

- ▶ Дефицит возникает, когда размер популяции резко уменьшается
- ▶ Эффект колонизатора возникает, когда несколько особей колонизируют и обосновываются в новой окружающей среде



Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 41

Маленькие популяции очень подвержены угрозе исчезновения, потому что сохранившиеся образцы не являются представителем того генофонда, который был до его разрушения.

Оба эффекта зависят от числа сохранившихся особей (или колонизаторов) и уровня роста популяции.

График на слайде показывает эффект дефицита. Слева – популяция в балансе Харди-Вейнберга (Hardy-Weinberg) с частотами аллелей 0,5. В случае внезапного уменьшения и затем восстановления первоначального размера, результат может выразиться в потере одних аллелей и фиксации других. В примере справа выжил только A_2A_2 и аллель A_1 утерян.

С эффектами дефицита и колонизатора ...

- ▶ Гетерозиготность уменьшается при:

$$H_1 = (1 - 1/2N)H_0$$

- ▶ Аллели теряются при:

$$P = p^{2N} + q^{2N}$$

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 42

- Уменьшение гетерозиготности

где,

H_1 = конечная гетерозиготность

H_0 = первоначальная гетерозиготность

N = размер популяции

тогда:

если $N = 100$ и $H_0 = 0,25$, тогда $H_1 = 0,24875$

если $N = 40$ и $H_0 = 0,25$, тогда $H_1 = 0,24685$

- Потеря аллелей

где,

P = потеря аллеля

p и q = частоты аллеля

$2N$ = общее число аллелей в популяции

тогда:

если $N = 100$, $p = 0,90$ и $q = 0,1$, тогда $P = 7,05508 \times 10^{-10}$

если $N = 15$, $p = 0,90$ и $q = 0,1$, тогда $P = 0,0423911$

если $N = 10$, $p = 0,90$ и $q = 0,1$, тогда $P = 0,12157665$

Приложение 1

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 43

Для просмотра приложения, щелкните мышкой на его названии:

Приложение 1. Критические значения распределения Хи-квадрата



Вкратце

Для анализа и интерпретации данных о генетическом разнообразии мы должны знать:

- Несколько основных определений популяционной генетики
- Закон Харди-Вейнберга (The Hardy-Weinberg principle)
- Вычисление частот аллелей и генотипа
- Основные причины изменений генетического разнообразия как мутация, миграция, рекомбинация, отбор и дрейф

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 44



Теперь Вы должны знать ...

- ▶ Основные определения, используемые в популяционной генетике
- ▶ Закон Харди-Вейнберга (The Hardy-Weinberg principle)
- ▶ Как рассчитать частоту аллелей при помощи маркерных данных
- ▶ Влияние системы спаривания на разнообразие популяции
- ▶ Основные источники изменчивости и их последствия

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 45



Если Вы хотите узнать больше...

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 46

de Vicente, M.C. и T. Fulton. 2003. Using Molecular Marker Technology in Studies on Plant Genetic Diversity.

http://www.bioversityinternational.org/Publications/pubfile.asp?ID_PUB=912

Doolittle, D.P. 1987. Population Genetics: Basic Principles. Springer-Verlag, Berlin.

Falconer, D.S. и T.F.C. Mackay (eds.). 1996. Introduction to quantitative genetics (4th edn.). Longman Group, London.

Griffiths, A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin и W.M. Gelbart (eds.). 1996. An Introduction to Genetic Analysis (6th edn.). Freeman and Co., NY.

Hartl, D.L. 1988. A Primer of Population Genetics (2nd edn.). Sinauer Associates, Sunderland, MA.

Hedrick, P.W. 1985. Genetics of Populations. Jones and Barlett Publishers, Boston, MA.

Snedecor, G.W. и W.S. Cochran (eds.). 1980. Statistical Methods (7th edn.). Iowa State University Press, Ames, IO.



Следующее

- ▶ Измерения генетического разнообразия
- ▶ Компьютерные программы для анализа генетического разнообразия
- ▶ Глоссарий

Авторское право: Bioversity International и Корнельский Университет, 2003 г. Популяционная генетика 47

Приложение 1:

Основные концепции популяционной генетики

Критические значения распространения хи-квадрата

ν	α	
	0,05	0,01
1	3,84	6,64
2	5,99	9,21
3	7,81	11,34
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,92	21,67
10	18,31	23,21
11	19,68	24,72
12	21,03	26,22
13	22,36	27,69
14	23,68	29,14
15	25,00	30,58
16	26,30	32,00
17	27,59	33,41
18	28,87	34,80
19	30,14	36,19
20	31,41	37,57
21	32,67	38,93
22	33,92	40,29
23	35,17	41,64
24	36,41	42,98
25	37,65	44,31
26	38,88	45,64
27	40,11	46,96
28	41,34	48,28
29	42,56	49,59
30	43,77	50,89

**Анализ генетического
разнообразия с использованием
данных молекулярных маркеров:**

Учебный модуль

**Измерение генетического
разнообразия**

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 1

Содержание

- ▶ Основной анализ генетического разнообразия
- ▶ Типы переменных
- ▶ Количественное измерение генетического разнообразия:
 - Измерение генетического разнообразия внутри популяции
 - Измерение генетического разнообразия между популяциями
- ▶ Количественное измерение генетических взаимосвязей:
 - Разнообразие и дифференциация на нуклеотидном уровне
 - Генетическая дистанция
- ▶ Демонстрация взаимосвязей:
 - Группировка или образование кластеров
 - Классификация
- ▶ Приложения

Основной анализ генетического разнообразия

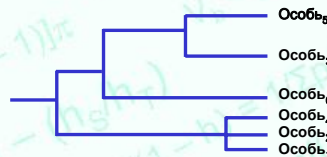
1. Описание изменчивости внутри и между популяциями, регионами, пр.

		Особи					
Д а н н ы е	1	0	1	1	0	1	
	1	0	0	0	1	1	
	0	1	1	0	1	0	
	1	0	0	0	1	1	
	0	0	1	1	0	0	
	1	1	1	0	0	0	
	1	0	1	0	1	1	

2. Оценка взаимосвязей между особями, популяциями, регионами, пр.

	01	02	03	04	05	06
01	0					
02	0,56	0				
03	0,33	0,33	0			
04	0,47	0,26	0,50	0		
05	0,32	0,43	0,37	0,28	0	
06	0,33	0,56	0,56	0,37	0,46	0

3. Экспрессия взаимосвязей между результатами, полученными от разных наборов признаков



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 3

Большинство анализов генетического разнообразия, которые мы хотели бы провести, будут включать следующие этапы:

1. *Описание разнообразия.* Это может быть сделано внутри популяции и между популяциями. Оно также может быть расширено до более больших единиц как области и регионы.
2. *Вычисление взаимосвязей между единицами, анализируемыми в первом этапе.* Оно включает в себя вычисление расстояний (геометрических или генетических) между парой исследуемых объектов.
3. *Экспрессия этих взаимосвязей при помощи любого имеющегося в распоряжении метода группировки и/или классификации.* Некоторые из этих методов позволяют сравнить результаты нашего молекулярного исследования с другими типами данных, например, географическими. На слайде, Особь₁, Особь₂, ... могут быть представлены вместо особей, популяций или регионов.

Типы переменных

- ▶ Качественные относятся к признакам или качествам и являются либо двоичными/бинарными, либо категориальными:
 - Бинарными, выбирая только два значения: имеется (1) или отсутствует (0)
 - Категориальными, выбирая значение среди многих возможных, и являются либо порядковыми, либо номинальными:
 - Порядковые: категории, имеющие последовательность
 - Номинальные: категории не взаимосвязанные друг с другом
- ▶ Количественные являются численными и либо непрерывными, либо дискретными/прерывистыми:
 - Непрерывными, выбирая значение внутри данной области
 - Дискретными, выбирая целые или десятичные числа

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 4

Примеры качественных переменных:

- Бинарные: например, опушенность листьев: имеется (1), отсутствует (0)
- Категориальные:
 - Порядковые: например, опушенность черешка: редкое (1), общее (2), обильное (3) или длина черешка: короткая (1), средняя (2), длинная (3)
 - Номинальные: например, цвет лепестка: желтый (1), красный (2), белый (3), пурпурный (4)

Примеры количественных переменных:

- Непрерывные: например, вес корней (г); длина листа (см)
- Дискретные: например, число тычинок: 2, 3, 4, ...
число плодов: 1, 2, 3, ...

Категориальные переменные могут быть переведены в бинарные, но с ограничением потому что, как мы увидим позже, некоторые коэффициенты схожести придают вес категории признака и это может оказать негативное влияние на другие оцениваемые признаки. Поэтому, чем больше категорий имеет переменная, тем больший вес она имеет при объединении с другими бинарными или категориальными переменными с несколькими категориями.

Пример перевода категориальной переменной в бинарную 1:

- Длина черешка: короткая (1), средняя (2), длинная (3)
- Короткая: имеется (1), отсутствует (0)
 - Средняя: имеется (1), отсутствует (0)
 - Длинная: имеется (1), отсутствует (0)

Количественные переменные также могут быть переведены в бинарные, например:

- от 0 до 3 плодов: имеется (1), отсутствует (0)
- от 4 до 7 плодов: имеется (1), отсутствует (0), ...

Качественное измерение генетического разнообразия: измерение генетического разнообразия внутри популяции

- ▶ На основе ряда вариантов
 - Полиморфизм или уровень полиморфизма (P_j)
 - Долю полиморфных локусов
 - Богатство аллельных вариантов (A)
 - Среднее количество аллелей на локус
- ▶ На основе частоты вариантов
 - Эффективное число аллелей (A_e)
 - Средняя предполагаемая гетерозиготность (H_e); генетическое разнообразие Нея (Nei)

Полиморфизм или уровень полиморфизма (Pj)

Ген считается полиморфным, если частота одного из его аллелей меньше или равна 0,95 или 0,99

$$P_j = q \leq 0,95 \quad \text{или} \quad P_j = q \leq 0,99$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 6

где,

Pj = уровень полиморфизма
q = частота аллели

- Это измерение обеспечивает критерий для демонстрации того, что ген проявляет изменчивость.
- Его вычисление производится путем прямого изучения на соответствие определению.
- Оно может быть использовано с кодоминантными маркерами и, очень ограниченно, доминантными маркерами потому что оценка на основе доминантных маркеров была бы смещена ниже действительных показателей.

Обычно полиморфный ген – это тот ген, для которого наиболее общий аллель имеет частоту менее 0,95. Редкие аллели определены как те, частота которых менее 0,005. Ограничение частоты аллели, которое определено на отметке 0,95 (или 0,99) является произвольным, ее цель заключается в оказании помощи при определении тех генов, в которых аллельная изменчивость является широко распространенной.

Литература

Cavalli-Sforza, L.L. и W.F. Bodmer. 1981. Genética de las Poblaciones Humanas. Ed. Omega, Barcelona.

Доля полиморфных локусов

Это число полиморфных локусов, разделенное на общее число локусов (полиморфных и мономорфных):

$$P = n_{pj}/n_{\text{общее}}$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 7

где,

P = доля полиморфных локусов

n_{pj} = число полиморфных локусов

n_{total} = общее число локусов

- Она показывает процентное содержание изменчивых локусов в популяции.
- Ее расчет основан на прямом подсчете числа полиморфных и общего числа локусов.
- Она может быть использована с кодоминантными маркерами и, очень ограничительно, с доминантными маркерами (см. предыдущий слайд для разъяснения)

Богатство аллельных вариантов (A)

- ▶ Относится к числу вариантов в образце
- ▶ Уровень разнообразия равен $(A - 1)$ вариантов потому, что внутри мономорфной популяции степень разнообразия равна нулю ($A - 1 = 0$)

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 8

- Для данного гена в образце, это измерение говорит о том, сколько аллельных вариантов могут быть найдены.
- Оно реагирует на размер образца.
- Хотя распространение аллелей не имеет значения, максимальное число аллелей имеет.
- Измерение может быть применено только с кодоминантными маркерами.

Среднее число аллелей на локус

Это сумма всех обнаруженных аллелей во всех локусах, разделенное на общее число локусов

$$n = (1/K) \sum_{i=1}^K n_i$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 9

где,

K = число локусов

n_i = число аллелей, определенных на каждый локус

- Это измерение предоставляет дополнительную информацию к информации о полиморфизме.
- Оно требует только подсчета числа аллелей в каждом локусе и затем расчета среднего значения.
- Оно наилучшим образом используется с кодоминантными маркерами, потому что доминантные маркеры не позволяют определить все аллели.

Эффективное число аллелей (A_e)

Это число аллелей, которые могут присутствовать в популяции

$$A_e = 1/(1 - h) = 1/\sum p_i^2$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 10

где,

p_i = частота i -ой аллели в локусе

$h = 1 - \sum p_i^2$ = гетерозиготность в локусе

- Измерение говорит о числе аллелей, которое можно предположить в локусе каждой популяции.
- Оно вычисляется путем инвертирования/переворачивания измерения гомозиготности в локусе.
- Оно может быть использовано с данными кодоминантных маркеров.
- На его вычисление может оказать влияние размер образца.

Это измерение разнообразия может быть информативным для разработки стратегий сбора. Например, мы определили его в данном образце. Затем мы проверили его в другом образце или на всей коллекции. Если число, полученное при проверке, меньше, чем число, полученное в первом варианте, то это означает, что нашу стратегию сбора нужно пересмотреть.

Расчет A_e : пример

Локусы (A, B, C)	Популяция 1			Популяция 2		
	Особь 1	A ₁ A ₁	B ₁ B ₁	C ₁ C ₁	A ₁ A ₁	B ₁ B ₃
Особь 2	A ₁ A ₂	B ₁ B ₂	C ₂ C ₂	A ₁ A ₁	B ₂ B ₃	C ₁ C ₁
Особь 3	A ₁ A ₁	B ₁ B ₁	C ₁ C ₃	A ₂ A ₂	B ₁ B ₄	C ₁ C ₁
Особь 4	A ₁ A ₃	B ₁ B ₃	C ₂ C ₃	A ₂ A ₂	B ₁ B ₁	C ₁ C ₁
Особь 5	A ₃ A ₃	B ₃ B ₃	C ₃ C ₃	A ₁ A ₂	B ₄ B ₄	C ₁ C ₁
Число аллелей	3	3	3	2	4	1
Частота аллели 1	0,60	0,60	0,30	0,50	0,40	1,00
Частота аллели 2	0,10	0,10	0,30	0,50	0,10	0,00
Частота аллели 3	0,30	0,30	0,40	—	0,20	0,00
Частота аллели 4	—	—	—	—	0,30	—
Гетерозиготность (h)	0,54	0,54	0,66	0,50	0,70	0,00
Эффективное число аллелей	2,17	2,17	2,94	2,00	3,33	1,00

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 11

Таблица на слайде демонстрирует пример вычисления эффективного числа аллелей. Две популяции имеют по 5 особей каждая. Для каждой особи проанализированы 3 локуса, каждый с разным числом аллелей в зависимости также от популяции (локус A имеет 3 аллели в популяции 1 и только 2 аллели в популяции 2, и так далее). Сначала, вычислены частоты аллели для каждого локуса и каждой популяции; затем гетерозиготность в каждом локусе; и наконец, A_e согласно формуле, приведенной в предыдущем слайде.

Средняя предполагаемая гетерозиготность (H_e) (генетическое разнообразие Нея (Nei) [D])

- ▶ Это вероятность того, что в единичном локусе любые две аллели, выбранные произвольно в популяции, отличаются друг от друга
- ▶ Возможны три вида расчетов :
 - Локус с двумя аллелями: $h_j = 1 - p^2 - q^2$
 - Локус j с i аллелями: $h_j = 1 - \sum p_i^2$
 - Среднее для нескольких локусов: $H = \sum_j^L h_j / L$
- ▶ Среднее H_e для всех локусов есть оценка уровня генетической изменчивости в популяции

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 12

где,

h_j = гетерозиготность на локус

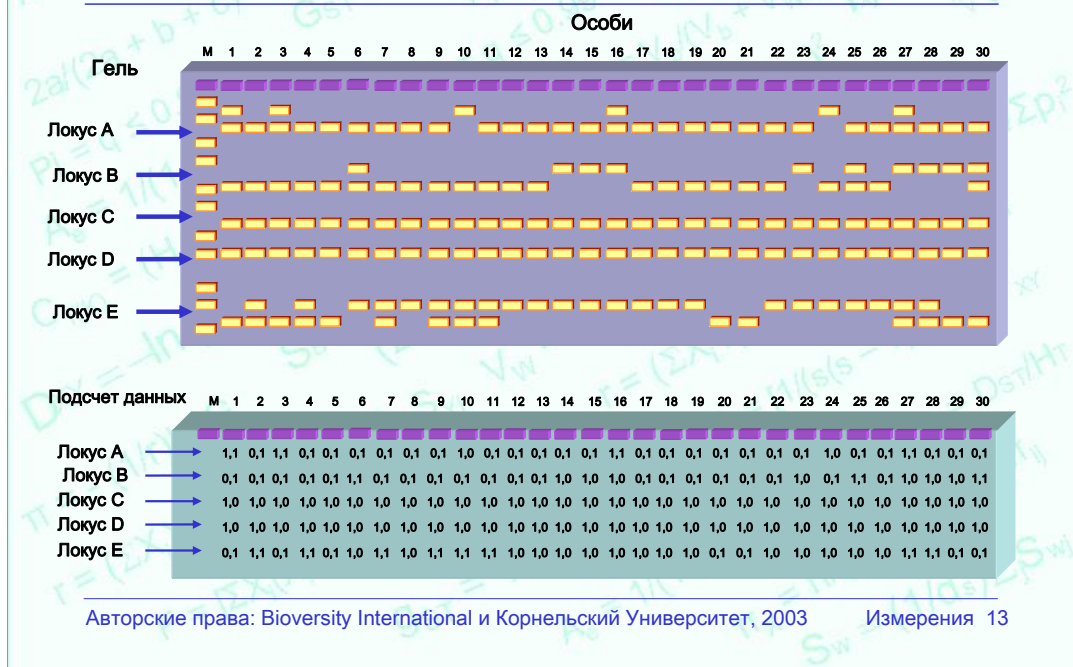
p и q = частота аллели

H = средняя гетерозиготность для нескольких локусов

L = общее число локусов

- Средняя предполагаемая гетерозиготность вычисляется путем вычитания из 1 предполагаемых частот гомозигот в локусе. Действие повторяется для всех локусов и затем вычисляется среднее значение.
- Она может быть применена ко всем маркерам как кодоминантным, так и доминантным.
- На вычисленное значение могут оказать влияние те аллели, которые показывают более высокие частоты.
- Она варьирует от 0 до 1.
- Она достигает максимального значения, когда имеется много аллелей при равных частотах.
- Как минимум 30 локусов в 20 особях каждой популяции должны быть проанализированы для уменьшения риска статистической погрешности.

Вычисление разнообразия при помощи кодоминантного молекулярного маркера



(продолжение в следующем слайде)

Верхняя половина слайда демонстрирует изображение геля с мерным маркером слева (М) и 30 особями, анализируемыми с кодоминантным маркером, которые обнаружили пять локусов (А, В, С, D и Е). Только три из этих локусов являются полиморфными (А, В и Е).

Нижняя часть слайда показывает результаты подсчета полос по каждой особи и каждому локусу. Учтите, что для облегчения изложения, не более двух аллели на локус были изображены. Хотя полосы, принадлежащие локусам С и D, были подсчитаны (1,0) для всех особей, в таком подсчете не было необходимости, так как полосы не дали информацию о разнообразии.

Смотрите следующий слайд для расчетов.

Вычисление разнообразия при помощи кодоминантного молекулярного маркера (продолжение)

Локус	Анализ данных				Частота аллели		$h_i = (1 - p^2 - q^2)$	H_i
	Генотипы	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Итого	p		
A	Генотипы	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Итого	p	q	
	Частота ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	2	4	24	30	0,13	0,87	0,23
	Частота ген. (наб.)	$P_{11} = 0,07$	$P_{12} = 0,13$	$P_{22} = 0,80$	1			
B	Генотипы	$B_1 B_1$	$B_1 B_2$	$B_2 B_2$	Итого	p	q	
	Частота ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	7	3	20	30	0,28	0,72	0,41
	Частота ген. (наб.)	$P_{11} = 0,23$	$P_{12} = 0,10$	$P_{22} = 0,67$	1			
E	Генотипы	$E_1 E_1$	$E_1 E_2$	$E_2 E_2$	Итого	p	q	
	Частота ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	15	8	7	30	0,63	0,37	0,46
	Частота ген. (наб.)	$P_{11} = 0,50$	$P_{12} = 0,27$	$P_{22} = 0,23$	1			

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 14

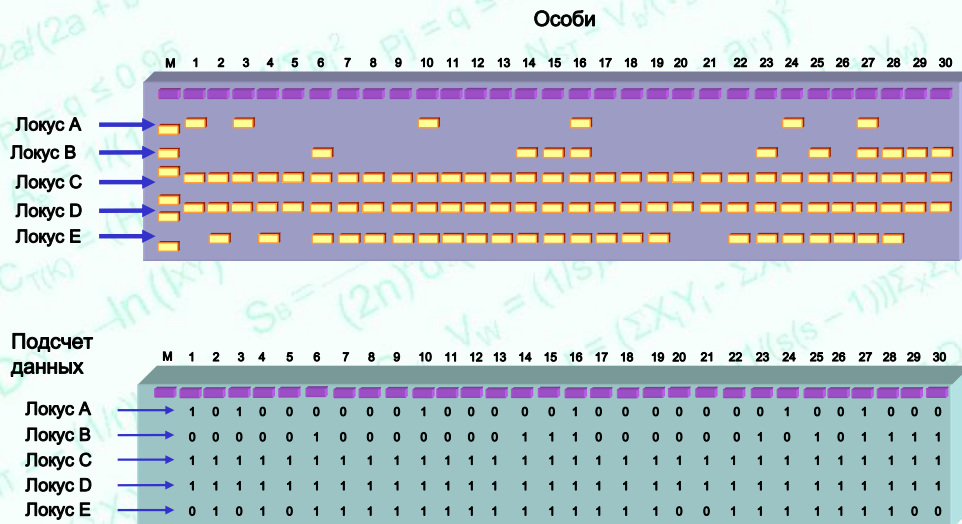
- Во-первых, мы учли, что локусы А, В и Е являются полиморфными потому, что они отвечают требованию иметь частоты аллели ниже 0,99. Локусы С и D являются мономорфными. (расч. – расчетное значение; наб.- наблюдаемое значение)
- Доля полиморфных локусов равна $P = (3/5) = 0,6$ или 60%, что получено путем деления числа полиморфных локусов на общее число анализируемых локусов.
- Для вычисления средней гетерозиготности (H_o), мы:
 - Подсчитаем сколько локусов из их общего числа являются гетерозиготными. Например, Особь₁ имеет один гомозиготный локус (А), Особь₂ также имеет один (Е), Особь₂₇ имеет 2 гомозиготных локуса (А и Е), Всего, 16 особей были мономорфными (например, они имели только одну полосу в каждом из пяти локусов), 13 особей имели 1 гетерозиготный локус и 1 особь имела 2 гетерозиготных локуса.
 - Вычислим среднее значение наблюдаемой гетерозиготности как:

$$H_o = [16(0/5) + 13(1/5) + 1(2/5)]/(30) = 0,1$$

- разнообразие гена внутри локуса (h_i) рассчитывается для каждого локуса по формуле в верхней строке таблицы, дающей на локус А = 0,23, локус В = 0,41 и локус Е = 0,46.
- Среднее расчетное разнообразие гена (H_i) рассчитывается по формуле на слайде 12:

$$H_i = (0,23 + 0,41 + 0,46)/5 = 0,22$$

Вычисление разнообразия при помощи доминантного молекулярного маркера



Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 15

(продолжение в следующем слайде)

Верхняя половина слайда демонстрирует изображение геля с мерным маркером слева (М) и 30 особями, анализируемыми с доминантным маркером. Пять локусов (А, В, С, D и E) были обнаружены. Только три из этих обнаруженных локусов являются сегрегирующими (А, В и E), в то время как два других С и D являются мономорфными.

Нижняя часть слайда показывает результаты подсчета полос по каждой особи и каждому локусу. Так как мы имеем дело с доминантным маркером, то полосы были подсчитаны как 1 в случае присутствия или 0 при отсутствии. Подсчет полос для локусов С и D можно либо опустить, либо проделать как в слайде с «1» для каждой особи.

Смотрите следующий слайд для расчетов.

Вычисление разнообразия при помощи доминантного молекулярного маркера (продолжение)

Локус	Анализ данных				Частота аллели		$h_j = (1 - p^2 - q^2)$	H_i
	Генотип	Aa	Aa	aa	Итого	p		
A	Генотип	Aa	Aa	aa	Итого	p	q	
	Част.ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	6		24	30			
	Част.ген. (набл.)	$P_1 = 0,20$		$P_2 = 0,80$	1	0,11	0,89	0,19
B	Генотип	BB	Bb	bb	Итого	p	q	
	Част.ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	10		20	30			
	Част.ген. (набл.)	$P_1 = 0,33$		$P_2 = 0,67$	1	0,18	0,82	0,30
E	Генотип	EE	Ee	ee	Итого	p	q	
	Част.ген. (расч.)	p^2	$2pq$	q^2	1			
	Число особей	23		7	30			
	Част.ген. (набл.)	$P_1 = 0,77$		$P_2 = 0,23$	1	0,52	0,48	0,50

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 16

- Во-первых, мы взглянем на полиморфизм, демонстрируемый всеми локусами. Локусы A, B и E отвечают требованию иметь частоты аллели ниже 0,99 и, таким образом, можно сказать, что они полиморфны. Локусы C и D являются мономорфными. (p. = расчетное значение; набл. = наблюдаемое значение)
- Доля полиморфных локусов (P) равна $P = (3/5) = 0,6$ или 60%. Средняя гетерозиготность (H_e) не может быть оценена потому, что доминантные маркеры не позволяют различить гетерозиготные и гомозиготные особи.
- Несмотря на вышеизложенное (2), разнообразие гена внутри локуса (h_j) может быть рассчитано для каждого локуса, используя формулу, приведенную в графе 4 верхней строки таблицы: локус A = 0,19; локус B = 0,30; и локус E = 0,50,
- Среднее разнообразие гена (H_i) рассчитано по формуле, приведенной в слайде 12 :

$$H_i = (0,19 + 0,30 + 0,50)/5 = 0,198$$

Количественное измерение генетического разнообразия: измерение генетического разнообразия между популяциями

- ▶ Межпопуляционная дифференциация для одного локуса (g_{ST})
- ▶ Межпопуляционная дифференциация для нескольких локусов (G_{ST})
- ▶ Вклад популяции в общее генетическое разнообразие
- ▶ Статистика F (Wright)
- ▶ Анализ молекулярной вариации (AMB) (Analysis of molecular variance (AMOVA))

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 17

‘Дифференциация’ означает полиморфные различия между популяциями на разных уровнях структуры (популяции и особи)

Межпопуляционная дифференциация для одного локуса (g_{ST})

$$g_{ST} = 1 - (h_S/h_T)$$

h_S = разнообразие популяции

h_T = общее разнообразие

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 18

где,

$$h_S = (\tilde{n}/(\tilde{n} - 1)) [1 - (1/s) \sum \sum x_{ij}^2 - (h_o/2\tilde{n})]$$

$$h_T = 1 - \sum [(1/s) \sum x_{ij}]^2 + (h_S/\tilde{n}s) - (h_o/2\tilde{n}s)$$

\tilde{n} = гармоническое среднее размеров популяции

s = число популяций

h_o = среднее наблюдаемой гетерозиготности

x_{ij} = оцененная частота i -ой аллели в j -ой популяции

- Формула на слайде предоставляет измерение дифференциации по числу аллелей на локус в двух и более популяциях.
- Оно колеблется от 0 до 1. Отрицательное значение может быть получено, если ошибка была допущена при проведении отбора, или была использована несоответствующая маркерная система.
- Из-за сложности компонентов, его вычисление требует специальной компьютерной программы.
- Оно может быть использовано с кодоминантными маркерами и ограничительно с доминантными, потому что это измерение гетерозиготности. Для правильного расчета действительного значения необходимы несколько поколений.

Расчет g_{ST}

Генотипы	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	p	q	$p^2 + q^2$
Поп. 1	20	30	50	0,35	0,65	0,545
Поп. 2	10	20	70	0,20	0,80	0,680
Поп. 3	60	10	30	0,65	0,35	0,545
$h_o = 1/3(0,3 + 0,2 + 0,1) = 0,20$				$s = 3$		$\sum(p^2 + q^2) = 1,77$
$1/\bar{n} = 1/n_1 + 1/n_2 + 1/n_3 = 1/100 + 1/100 + 1/100 = 0,03$					$\bar{n} = 33,33$	
$h_s = (33,33/33,33 - 1)[1 - 1/3(1,77) - (0,20/2(33,33))] = 0,4196$						
$\sum[1/3 \sum x_{ij}]^2 = (1/3(0,35))^2 + (1/3(0,65))^2 + (1/3(0,20))^2 + \dots + (1/3(0,35))^2 = 0,1967$						
$h_T = 1 - 0,1967 + [0,4196/(33,33 \times 3)] - [0,20/(2 \times 33,33 \times 3)] = 0,8065$						
$g_{ST} = 1 - (h_s/h_T) = 1 - (0,4196/0,8065) = 0,4797$						

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 19

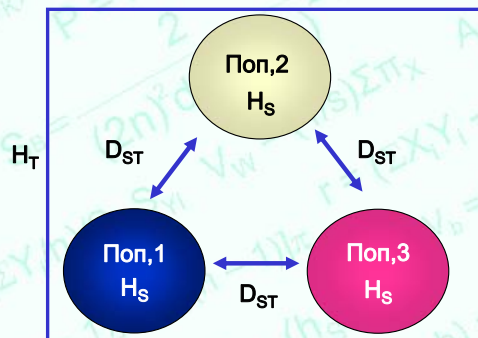
В этом примере мы имеем число особей для каждого генотипа одного локуса (A) в трех разных популяциях. Используя это число, мы хотим узнать степень дифференциации во всех трех популяциях. В таблице при расчете соблюдаются все необходимые элементы формулы, приведенной в предыдущем слайде.

Результат ($g_{ST} = 0,4797$) показывает значительную дифференциацию между популяциями в частотах аллели. Мы можем, таким образом, сказать, что высокий процент генетического разнообразия распределен между популяциями.

Межпопуляционная дифференциация для нескольких локусов (G_{ST})

G_{ST} – это коэффициент дифференциации гена

$$G_{ST} = D_{ST}/H_T$$



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 20

где,

H_T = общее генное разнообразие = $H_S + D_{ST}$

H_S = внутрипопуляционное генное разнообразие

D_{ST} = межпопуляционное разнообразие

$$(H_T/H_T) = (H_S/H_T) + (D_{ST}/H_T) = 1$$

- G_{ST} измеряет долю разнообразия гена, которое распределено между популяциями.
- Большое число локусов должно быть отобрано.
- Уравнения являются сложными и должны рассчитываться при помощи специальных компьютерных программ.

Например, предположим, что:

$$H_T = 0,263$$

$$H_S = 0,202$$

$$D_{ST} = 0,263 - 0,202 = 0,061$$

тогда, $G_{ST} = (D_{ST}/H_T) * 100 = (0,061/0,263) * 100 = 23,19\%$. Это значит, что в этом виде между популяциями существует 23% дифференциации.

Вклад популяции в общее разнообразие гена

Вклад рассчитывается путем удаления популяции таким образом, чтобы можно было оценить ее вклад в общее разнообразие гена

$$C_{T(K)} = (H_T - H_{T/K})/H_T$$

$$C_{S(K)} = (H_S - H_{S/K})/H_T$$

$$C_{ST(K)} = (D_{ST} - D_{ST/K})/H_T$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 21

где,

$C_{T(K)}$ = вклад K в общее разнообразие

$C_{S(K)}$ = вклад K в внутрипопуляционное разнообразие

$C_{ST(K)}$ = вклад K в межпопуляционное разнообразие

H_T = общее разнообразие гена

H_S = внутрипопуляционное генное разнообразие

D_{ST} = межпопуляционное разнообразие

$H_{T/K}$ = общее разнообразие гена после удаления популяции K

$H_{S/K}$ = внутрипопуляционное разнообразие гена после удаления популяции K

$D_{ST/K}$ = межпопуляционное разнообразие гена после удаления популяции K

- Измерение позволяет выразить количественно изменчивость общего разнообразия гена, когда популяция введена в или удалена из места произрастания (например, когда вводится новый сорт на поле фермера в рамках программы *in situ* сохранения).
- Оно также служит для измерения влияния потери популяции на данном участке на разнообразие гена.
- Оно также может использоваться только с кодоминантными маркерами

Статистика F (Wright)

Уравнение генетической структуры популяции равно:

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$$

$$F_{IT} = 1 - (H_I/H_T)$$

$$F_{IS} = 1 - (H_I/H_S)$$

$$F_{ST} = 1 - (H_S/H_T)$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 22

где,

H_T = общее разнообразие гена или предполагаемой гетерозиготности в общей популяции как оценено из суммированных частот аллели

H_I = внутривидовое разнообразие гена или средняя наблюдаемая гетерозиготность в группе популяций

H_S = средняя предполагаемая гетерозиготность, рассчитанная из каждой субпопуляции

Статистика F позволяет проанализировать структуры подразделенных популяций. Она также может быть использована для измерения генетической отдаленности между субпопуляциями, концепция которой основана на идее, что те субпопуляции, которые не скрещиваются между собой будут иметь разные частоты аллели в отличие от общей популяции.

Генетическая отдаленность также предоставляет способ измерения вероятности встречи между равными аллелями (эндогамии). Статистические индексы включали измерение:

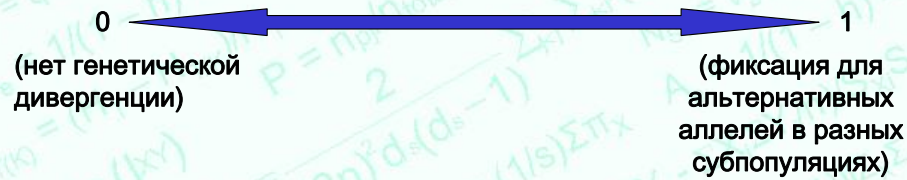
F_{IS} = дефицит или избыток средних гетерозигот в каждой популяции

F_{ST} = степень дифференциации гена между популяциями в отношении частот аллели

F_{IT} = дефицит или избыток средних гетерозигот в группе популяций

Интерпретация значений F_{ST}

Область F_{ST} равна:



когда F_{ST} :

от 0 до 0,05
от 0,05 до 0,15
от 0,15 до 0,25
>0,25

тогда генетическая дифференциация:

маленькая
средняя
большая
очень большая

Расчет статистических данных F

Поп.	Частота генотипа			p_i	q_i	$2p_iq_i$	F
	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$				
1	0,40	0,30	0,30	0,55	0,45	0,4950	0,3939
2	0,60	0,20	0,20	0,70	0,30	0,4200	0,5238
H_T	$2(0,625)(0,375) = 0,4688$			p_o	$(0,55 + 0,70)/2 = 0,625$		
H_I	$(0,3 + 0,2)/2 = 0,25$			q_o	$(0,45 + 0,30)/2 = 0,375$		
H_S	$(0,495 + 0,420)/2 = 0,4575$						

$$F_{IT} = 1 - (0,25/0,4688) = 0,4667$$

$$F_{IS} = 1 - (0,25/0,4575) = 0,4536$$

$$F_{ST} = 1 - (0,4575/0,4688) = 0,0241$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 24

(продолжение в следующем слайде)

Этот слайд демонстрирует пример из двух популяций и анализ одного локуса (A). Частоты аллели рассчитаны (p и q) как их средние значения. Переменные H_T , H_I и H_S также рассчитаны и использованы для вычисления статистики F. Анализ показывает низкую дифференциацию в частотах аллели среди двух популяций (F_{ST}). Мы можем заключить, что почти весь дефицит гетерозигот был вызван неслучайным скрещиванием внутри популяций ($F_{IS} = 0,4536$).

F = индекс фиксации (первая графа справа таблицы) есть вероятность того, что две аллели несомые одной особью будут одинаковыми. Она должна быть вычислена только с кодоминантными маркерами. Если проделать эту работу с доминантными маркерами, то результатом может быть смещенная оценка.

Расчет статистических данных F (продолжение)

Поп.	Частота генотипа			p_i	q_i	$2p_iq_i$	F
	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$				
1	0,25	0,50	0,25	0,50	0,50	0,500	0,0000
2	0,80	0,10	0,10	0,85	0,15	0,255	0,6078
H_T	$2(0,675)(0,325) = 0,4388$			p_o	$(0,50 + 0,85)/2 = 0,675$		
H_I	$(0,5 + 0,1)/2 = 0,30$			q_o	$(0,50 + 0,15)/2 = 0,325$		
H_S	$(0,500 + 0,255)/2 = 0,3775$						

$$F_{IT} = 1 - (0,30/0,4388) = 0,3163$$

$$F_{IS} = 1 - (0,30/0,3775) = 0,2053$$

$$F_{ST} = 1 - (0,3775/0,4388) = 0,1397$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 25

Это другой пример, для которого процедуры, использованные в предыдущем слайде, были соблюдены. Дифференциация в частотах аллели между двумя популяциями кажется больше ($F_{ST} = 0,1397$), с единственно средним результатом неслучайного скрещивания внутри популяций ($F_{IS} = 0,2053$).

Анализ молекулярной варiances (АМОВА)

- ▶ АМОВА (Analysis of molecular variance) – это метод изучения молекулярной варiances внутри вида
- ▶ Он основан на иерархической или гнездовой/вложенной модели
- ▶ Он отличается от анализа варiances (ANOVA) в следующем:
 - Он может содержать разные эволюционные предположения без видоизменения базовой структуры анализа
 - Движущая гипотеза использует методы пермутаций, которые не требуют предположения нормального распространения

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 26

Разные иерархические уровни разнообразия гена, изученные посредством АМОВА, могут включать:

1. Континенты, которые могут включать мелкие иерархические уровни
2. Географические регионы внутри континента
3. Области внутри региона в континенте
4. Популяции внутри области региона в континенте
5. Особи внутри популяции в области региона в континенте

Математическое описание модели для ситуаций 3 и 4 могут быть найдены в Приложении 2 и 3 соответственно. Для их просмотра щелкните мышкой [здесь](#).

Следующие два слайда иллюстрируют как провести анализ ситуации 4.

Пример АМОВА

Особь	Поп. 1		Поп. 2		Поп. 3	
	A ₁	A ₂	A ₁	A ₂	A ₁	A ₂
1	0	0	0	1	1	1
2	1	1	0	1	1	1
3	0	0	1	1	0	1
4	1	0	1	0	1	1
5	0	0	0	1	0	1
6	0	0	0	1	0	0
7	1	1	1	1	1	1
8	0	0	1	1	0	0
9	1	0	1	1	1	0
10	1	1	1	0	0	1
11	1	0	0	1	1	1
12	0	0	1	1	1	0
13	1	1	1	1	0	1
14	1	1	1	0	1	0
15	1	1	0	1	1	0

X _{...k}	15	21	18	54
X _{...k} ²	225	441	324	990
∑∑X _{i...k} ²	27	33	28	88
∑∑∑X _{ijk} ²	15	21	18	54
X _{...,2}				2916

SS _a	0,6	MS _a	0,3
SS _b	11	MS _b	0,26190476
SS _w	10	MS _w	0,22222222

A ₁ = 1	Присутствует
A ₁ = 0	Отсутствует

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 27

(продолжение на следующем слайде)

В этой таблице мы приводим данные, полученные при анализе 15 особей из каждой из трех популяций с кодоминантным маркером. Путем анализа варiances, эти данные позволят нам вычислить статистику F.

Первый шаг заключается в переводе полос, обнаруженных в гелях, в бинарные переменные со значением либо 0, либо 1. Затем суммируются присутствующие (1) и, далее мы можем продолжить с суммированием квадратов. Сначала расчет делается для одной популяции, затем для всех других, которые у нас имеются (X_{...k}). Мы имеем i = 15 особи (результат b), j = 2 аллели (результат w), k = 3 популяции (результат a).

где,

X_{...k} – это результат суммирования всех присутствующих полос (1) в особях на популяцию

X_{...k}² – это результат возведения в квадрат числа, полученного выше

∑∑X_{i...k}² – это результат складывания квадратов сумм присутствия аллели в каждой особи (например, Особь₁ в Поп.1 будет (0 + 0)² + Особь₂ в Поп.1 (1 + 1)² + Особь ...)

∑∑∑X_{ijk}² – это сумма квадрата каждого значения

SS – это сумма квадратов для результатов a, b и w

Пример расчета SS:

$$SS_a = \sum X_{...k}^2 / ij - X_{...,2} / ijk = [990 / (15 \times 2)] - [2916 / (15 \times 2 \times 3)] = 0,6$$

MS – это среднее квадратов для результатов a, b и w

Пример расчета MS: $SS_a / df_a = 0,6 / 2 = 0,3$, где df_A означает степень свободы для результата (популяции)

Пример АМОВА (продолжение)

SV	df	SS	MS	EMS
Популяции	2	0,6	0,3	$\sigma_w^2 + 2\sigma_b^2 + 2*15\sigma_a^2$
Особь/поп.	42	11	0,26190476	$\sigma_w^2 + 2\sigma_b^2$
Внутри особи	45	10	0,22222222	σ_w^2

Расчет вариансы и статистики F	
σ_a^2	0,0012698
σ_b^2	0,0198413
σ_w^2	0,22222222
σ^2	0,24333
F_{IT}	0,086758
F_{IS}	0,0819672
F_{ST}	0,0052185
$(1 - F_{IT})$	0,91324
$(1 - F_{IS})(1 - F_{ST})$	0,91324

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 28

где,

SV = источники изменчивости

df = степени свободы

SS = сумма квадратов (см. предыдущий слайд)

MS = среднее квадратов (см. предыдущий слайд)

σ^2 = сумма расчетной вариансы

EMS = предполагаемое среднее квадратов

$$\sigma_w^2 = 0,22222222$$

$$\sigma_b^2 = (MS_b - MS_w)/2 = (0,26190476 - 0,22222222)/2 = 0,0198413$$

$$\sigma_a^2 = (MS_a - MS_b)/2 * 15 = (0,3 - 0,26190476)/2 * 15 = 0,0012698$$

$$\sigma^2 = \sigma_w^2 + \sigma_b^2 + \sigma_a^2 = 0,24333 \text{ (сумма рассчитанной вариансы)}$$

Расчеты статистики F уже были разъяснены в слайде 22. Конкретно для этого примера, они могут быть следующими:

$$F_{IT} = (\sigma_a^2 + \sigma_b^2)/\sigma^2 = (0,0012698 + 0,0198413)/0,24333 = 0,086758$$

$$F_{ST} = \sigma_a^2/\sigma^2 = 0,0012698/0,24333 = 0,0052185$$

$$F_{IS} = \sigma_b^2/(\sigma_b^2 + \sigma_w^2) = 0,0198413/(0,0198413 + 0,2222222) = 0,0819672$$

Дифференциация частоты аллели среди трех популяций очень низкая ($F_{ST} = 0,0052185$) и вероятно является результатом многих случайных скрещиваний. Необходимо проанализировать больше локусов, чтобы сделать вывод.

Количественное измерение генетических взаимосвязей: разнообразие и дифференциация на нуклеотидном уровне

- ▶ Используя данные последовательности
 - Внутрипопуляционное разнообразие нуклеотида
 - Межпопуляционное разнообразие нуклеотида
- ▶ Используя данные рестрикции
 - Вариация рестриктных фрагментов (бэндов)
 - Внутрипопуляционное разнообразие нуклеотида
 - Межпопуляционное разнообразие нуклеотида

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 29

Для этих расчетов сделано предположение, что каждый нуклеотид - это локус.

Используя данные последовательности: внутрипопуляционное разнообразие нуклеотида

Оно измеряет разнообразие нуклеотида среди нескольких последовательностей в данной области генома внутри популяции (π_x)

$$\pi_x = n/(n-1) \sum X_i X_j \pi_{ij}$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 30

где,

n = число анализируемых последовательностей в особях популяции

X_i = расчетной частота i -ой последовательности в популяции

X_j = расчетной частота j -ой последовательности в популяции

π_{ij} = относительное содержание разных нуклеотидов между последовательностями i и j

- Измерение дает информацию о степени разнообразия нуклеотидов среди нескольких последовательностей в данном регионе генома. Оно эквивалентно измерению аллельного разнообразия внутри локуса.
- Оно варьирует от 0 до 1 ($0 \leq \pi_x \leq 1$).
- Факторами, ограничивающими использование аналитического инструмента, являются:

Неполные геномные последовательности должны быть в наличии
Уравнение может быть применено только к гаплоидным данным

Этот параметр дает информацию о последовательностях нуклеотида, и модель предполагает гаплотипы (гаплоидные генотипы). Даже если исследование основано на диплоидных особях, секвенирование каждой копии генома является необходимой.

Вычисление внутрипопуляционного разнообразия нуклеотида

n		Последовательность	Частота X_i
5	Seq ₁	TCC T CGAT T ATTC C CAGGGTGC C GATG A AT	5/10 = 0,5
2	Seq ₂	TCC A CGAT T ATTC G CAGGGTGC C GATG A AT	2/10 = 0,2
1	Seq ₃	TCC A CGAT C ATTC C CAGGGTGC A GATG G AT	1/10 = 0,1
2	Seq ₄	TCC G CGAT T ATTC T CAGGGTGC G GATG A AT	2/10 = 0,2
10			

$$П_{1,2} = 2/30, \quad П_{1,3} = 4/30, \quad П_{1,4} = 3/30, \quad П_{2,3} = 4/30, \quad П_{2,4} = 3/30, \quad П_{3,4} = 5/30$$

$$\begin{aligned} \Pi_X &= 10/(10 - 1) \sum X_i X_j \pi_{ij} \\ &= (10/9)[0,5 * 0,2 * (2/30) + 0,5 * 0,1 * (4/30) + \dots + 1 * 0,2 * (5/30)] \\ &= \mathbf{0,037} \end{aligned}$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 31

Этот пример имеет 10 особей в популяции X. Для каждой особи мы анализируем одну последовательность из 30 нуклеотидов и находим, что последовательности особи различаются по 5 позициям нуклеотидов (голубой). В итоге, 4 альтернативные последовательности для этих 30 нуклеотидов присутствуют в популяции. В первой графе мы замечаем число особей (n), которые имеют определенные альтернативы последовательности.

Затем мы рассчитываем число разности нуклеотидов в каждой паре последовательности внутри популяции. Например, $П_{1,2} = 2/30$ означает, что между последовательностью 1 и 2 имеются две разницы нуклеотидов (T vs. A в положении 4, и C vs. G в положении 14).

Следующее, мы рассчитываем π_X для всей популяции. Полученное число равно 0,037 или 3,7% разнообразия нуклеотида на основе проанализированной последовательности в образце из 10 особей.

Используя данные последовательности: межпопуляционное разнообразие нуклеотида

- ▶ V_{XY} измеряет дивергенцию популяции на основе степени изменчивости последовательности (1 последовательность, 2 популяции)

$$V_{XY} = d_{XY} - (\pi_X + \pi_Y)/2$$

- ▶ V_W измеряет среднее разнообразие в популяции на основе нескольких последовательностей

$$V_W = (1/s)\sum \pi_x$$

- ▶ V_b измеряет общую дифференциацию в нескольких популяциях

$$V_b = [1/(s(s-1))]\sum_X \sum_Y V_{XY}$$

- ▶ N_{ST} – это относительная дифференциация

$$N_{ST} = V_b / (V_b + V_W)$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 32

где,

V_{XY} = дивергенция между популяциями X и Y

π_X = разнообразие нуклеотида в популяции X

d_{XY} = вероятность, что два случайных нуклеотида в популяции X и Y разные

s = число популяций

- Измерение дает информацию об уровне дифференциации среди последовательностей нуклеотида в популяциях
- Оно требует данных о последовательности в образце особей для каждой популяции
- Оно требует специальной компьютерной программы, которая включает в себя функцию сравнения последовательности.

CLUSTAL W, MALIGN и PAUP* являются некоторыми из них.

Расчет межпопуляционного разнообразия нуклеотида

Дивергенция нуклеотида между X и Y

$$V_{XY} = d_{XY} - (\pi_X \pi_Y)/2 = 0,14 - (0,037 + 0,09)/2 = 0,0765$$

Общая дифференциация

$$V_b = [1/(s(s-1))] \sum_X \sum_Y V_{XY} = [1/(2(2-1))] 0,0765 = 0,03825$$

Среднее разнообразие в каждой популяции

$$V_w = (1/s) \sum \pi_X = 1/2(0,037 + 0,09) = 0,0635$$

Относительная дифференциация

$$N_{ST} = V_b / (V_b + V_w) = 0,03825 / (0,03825 + 0,0635) = 0,3759$$

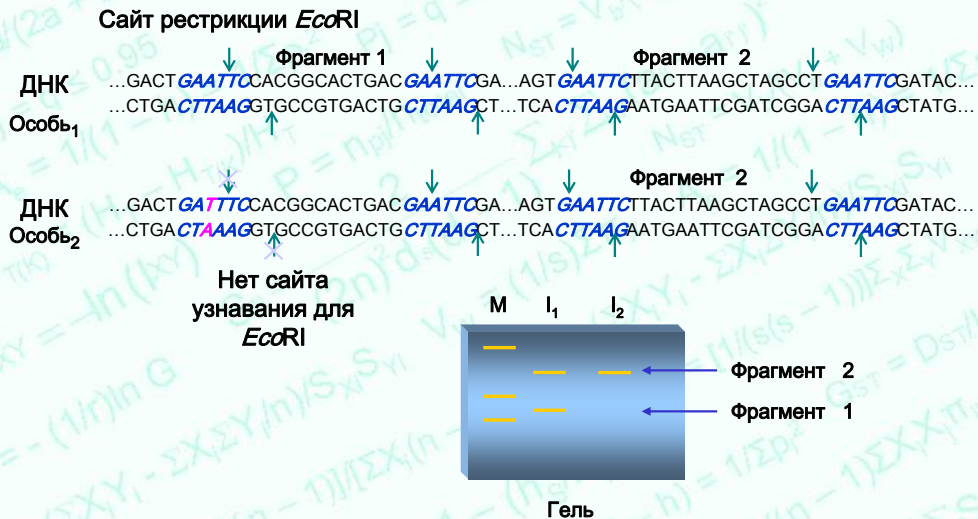
Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 33

Давайте скажем, что мы имеем другую популяцию Y, в которой разнообразие нуклеотида для той же последовательности, которая была проанализирована в слайде 31, равно $\pi_Y = 0,09$.

Мы также знаем, что вероятность того, что два нуклеотида, как произвольно взятые, являются разными в X и Y равна 0,14 (d_{XY}).

На этом слайде мы находим дивергенцию между популяциями X и Y (V_{XY}), общую дифференциацию (V_b), среднее разнообразие в каждой популяции (V_w) и относительную дифференциацию (N_{ST}).

Используя данные рестрикции: различие рестриктных фрагментов



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 34

Недостаток фрагмента 1 у Особи₂ указывает на то, что он несет разную последовательность ДНК, по крайней мере, в том сайте рестрикции. Маленькой разницы только двух нуклеотидов на рисунке вверху достаточно, чтобы заставить исчезнуть сайт узнавания для энзима.

Используя данные рестрикции: внутрипопуляционное разнообразие нуклеотида

Это измерение (π) основано на ряде фрагментов рестрикции, присутствующих в двух образцах

$$\pi = - (1/r) \ln G$$

(if $\pi < 5\%$)

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 35

где,

r = число нуклеотидов узнавания энзима рестрикции

$\ln G$ = натуральный логарифм вероятности о том, что здесь не было замены в сайте рестрикции. Расчет проводится следующим образом:

$$G = F(3 - 2G^0)^{1/4}$$

$$F = [\sum X_i(X_i n - 1)] / [\sum X_i(n - 1)]$$

F = удельный вес образуемых фрагментов

$$G^0 = F^{1/4}$$

n = число гаплоидных генотипов в популяции

X_i = расчетная частота i -го фрагмента в популяции

- Это измерение рассчитывает разнообразие в сайтах рестрикции в образце, потому что он зависит от последовательности нуклеотида сайтов рестрикции данного энзима рестрикции.
- Оно дает информацию о замещении нуклеотида в сайтах рестрикции. Оно варьирует от 0 до 1 ($0 \leq \pi_x \leq 1$).
- Уравнения выше могут быть использованы с гаплоидными образцами, митохондриальной и хлоропластной ДНК или гаплотипами.

Литература

Karp, A., P.G. Isaac и D.S. Ingram. 1998. Molecular Tools for Screening Biodiversity: Plants and Animals. Chapman & Hall, London.

Используя данные рестрикции: разнообразие нуклеотида между популяциями

- ▶ Это измерение (V_{XY}) означает дивергенцию или дифференциацию между популяциями на основе данных рестрикции

$$V_{XY} = d_{XY} - (\pi_X + \pi_Y)/2$$

- ▶ Это измерение также используется с данными маркера RAPD

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 36

где,

V_{XY} = дивергенция или дифференциация среди популяций X и Y

π_X = разнообразие рестрикции в популяции X

d_{XY} = разнообразие фрагмента среди двух популяций = $-(2/r)\ln(G_{XY})$

$G_{XY} = F_{XY}(3 - 2G^{\circ}_{XY})^{1/4}$

$G^{\circ} = F_{XY}^{1/4}$

F_{XY} = доля совместных аллелей среди популяций X и Y

$= (2\sum X_{iX}X_{iY})/(\sum(X_{iX} + X_{iY}))$

X_{iX} = расчетная частота фрагмента i в популяции X

- Оно оценивает разнообразие в сайтах рестрикции образца двух или более популяций. Оно дает информацию о замещении нуклеотида в сайтах рестрикции.
- Для этого полезны такие компьютерные программы как BIOSYS и GENEPOP. Полученные данные считаются принадлежащими к гаплоидным организмам.

Если использовать с RAPD значение 'r' замещается длиной праймера (r = 10).
Более того, сделаны некоторые предположения:

Использованы соответствующие праймеры

Полиморфизм в результате инсерции или делеции редок

Схожие фрагменты размера в разных популяциях принадлежат одному и тому же локусу

Фрагменты должны быть определены безошибочно

Обычно используются такие компьютерные программы как RAPDISTANCE и RAPDIS.

Расчет разнообразия нуклеотида между популяциями

Популяция	Пос.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Част. Xi	
A ₁		—	—				—	—					—	—									6/20 = 0,30
A ₂				—	—				—			—							—				5/20 = 0,25
A ₃						—				—	—				—	—	—		—	—	—		9/20 = 0,45
$F = \frac{[0,30(0,30 * 3 - 1) + 0,25(0,25 * 3 - 1) + 0,45(0,45 * 3 - 1)]}{0,30(3 - 1) + 0,25(3 - 1) + 0,45(3 - 1)} = 0,0325$																							
$G^{\circ} = (0,0325)^{1/4} = 0,424591$											$G = 0,0325[3 - 2(0,424591)]^{1/4} = 0,039358$												
$\pi X = -(1/6) \ln(0,039358) = 0,539176$																							

Популяция	Пос.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Част. Xi	
A ₁		—		—	—					—							—						5/20 = 0,25
A ₂		—		—			—	—		—	—	—	—	—				—	—	—	—		13/20 = 0,65
A ₃								—								—							2/20 = 0,10
$F = \frac{[0,25(0,25 * 3 - 1) + 0,65(0,65 * 3 - 1) + 0,10(0,10 * 3 - 1)]}{0,25(3 - 1) + 0,65(3 - 1) + 0,10(3 - 1)} = 0,2425$																							
$G^{\circ} = (0,2425)^{1/4} = 0,701743$											$G = 0,2425[3 - 2(0,701743)]^{1/4} = 0,272587$												
$\pi Y = -(1/6) \ln(0,272587) = 0,216633$																							

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 37

В каждой популяции мы обнаружили три фрагмента ДНК как результат рестриционного расщепления : A₁, A₂ и A₃.

Разнообразие нуклеотида в проанализированных регионах больше в популяции X ($\pi_X = 0,5392$), чем в популяции Y ($\pi_Y = 0,2166$); таким образом, X имеет больше разнообразия гена, чем популяция Y.

Между популяциями X и Y, дифференциация нуклеотида на основе сайтов рестрикции равна 0,230766.

$$F = \frac{2[0,30 * 0,25 + 0,25 * 0,65 + 0,45 * 0,10]}{(0,30 + 0,25) + (0,25 + 0,65) + (0,45 + 0,10)} = 0,14125$$

$$G^{\circ}XY = (0,14125)^{1/4} = 0,613052$$

$$GXY = 0,14125 [3 - 2(0,613052)]^{1/4} = 0,163012$$

$$dXY = -(2/6) \ln(0,163012) = 0,604643$$

$$VXY = 0,604643 - \frac{1}{2} (0,539176 + 0,216633) = 0,226739$$

$$VW = \frac{1}{2} (0,539176 + 0,216633) = 0,377905$$

$$Vb = \frac{1}{2} (0,226739) = 0,11337$$

$$NST = \frac{0,11337}{0,11337 + 0,377905} = 0,230766$$

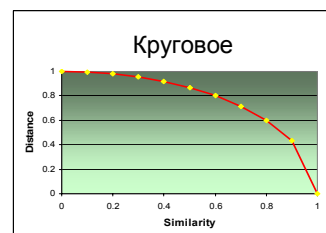
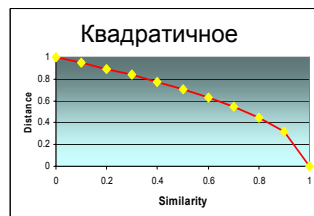
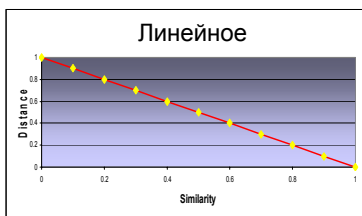
Количественное выражение генетических взаимосвязей: генетическое расстояние

- ▶ Генетическая расстояние между двумя образцами описывается как удельный вес генетических элементов (аллелей, генов, гамет, генотипов), которые не разделяются двумя образцами
- ▶ $D = 1$ когда, и только когда, два образца не имеют общих генетических элементов.

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 38

В зависимости от сходства особей возможны три типа представления расстояния (D) :

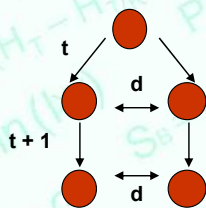
- $D = 1 - S$, известная как линейное расстояние потому, что предполагается, что взаимосвязь со сходством является линейной.
- $D = \sqrt{1 - S}$, известная как квадратичное расстояние потому, что предполагается, что взаимосвязь со сходством соответствует квадратичной функции, и поэтому, чтобы преобразовать ее в линейную необходимо вычислить квадратный корень .
- $D = \sqrt{1 - S^2}$, известная как круговое расстояние.



Модели расстояния

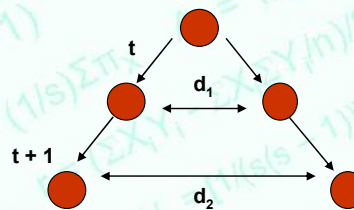
Расчет расстояния или несходства придерживается одной из двух возможных моделей:

Модель баланса



Расстояние остается постоянной во времени (баланс существует между миграцией и генетическим дрейфом)

Модель дисбаланса



Расстояние меняется со временем посредством миграции и генетического дрейфа

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 39

Для наших целей мы используем модель дисбаланса. Существует две альтернативы :

- Геометрическое расстояние
 - Не принимает в расчет процессы эволюции
 - Основана только на частотах аллели
 - Существует сложная взаимосвязь между расстоянием и временем дивергенции
- Генетическое расстояние
 - Не принимает в расчет процессы эволюции
 - Расстояние увеличивается со времени отделения от анцестральной популяции
 - Необходима генетическая модель эволюции

Когда должны мы использовать генетическое или геометрическое расстояние?

- Геометрическое расстояние используется в исследованиях разнообразия, в которых сравнения проводятся согласно морфологическим или молекулярным маркерам, собранным из оперативных таксономических единиц (ОТЕ). ОТЕ может быть особью, образцом или популяцией. Она может использоваться с доминантными маркерами (RAPD, AFLP) или кодоминантными маркерами. Так как эволюционные аспекты не учитываются, то полученные дендрограммы не могут быть интерпретированы как филогенетические древа, дающие информацию об эволюции или дивергенции среди групп.
- Наоборот, генетическое расстояние любой данной ОТЕ может быть включено в филогенетические исследования. Модель учитывает частоты аллели в ОТЕ и ее математическая основа отличается. Оно может использоваться как с кодоминантными, так и с доминантными маркерами, хотя с последними информация теряется потому, что только две аллели могут быть подсчитаны. Генетическое расстояние с доминантными маркерами, однако, требует изучения двух поколений одной и той же популяции для измерения расщепления локусов (Lynch и Milligan, 1994).

Литература

Lynch, M. и B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3:91-99.

Модели дисбаланса: геометрическое расстояние

- ▶ Она измеряет прямую взаимосвязь между индексом сходности (s) и расстоянием ($D = 1 - s$)
- ▶ Возможны разные ситуации, например:
 - Бинарные переменные
 - Количественные переменные
 - Смешанные типы переменных
 - P число переменных

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 40

(продолжение в следующем слайде)

При анализе молекулярных данных мы имеем дело с бинарными переменными (1,0). Это будет обсуждено в следующих слайдах.

В приложении 4 приводится дополнительная информация п для тех случаев когда Вы должны также работать с количественными переменными, смешанными типами переменных и разным числом переменных. В приложении 5 также приведен пример расчета геометрических расстояний с количественными переменными. Для просмотра приложений 4 и 5 щелкните мышкой [здесь](#).

Геометрическое расстояние (продолжение)

С бинарными переменными :

- Используется мультиварьируемый анализ, и матрицы сходства или дифференциации формируются между возможными парами особей или оперативными таксономическими единицами (ОТЕ)
- Две схожие особи имеют минимальное значение расстояния и максимальное значение схожести
- Расстояние и схожесть связаны обратно пропорционально
- Схожесть оценивается количеством совпадений

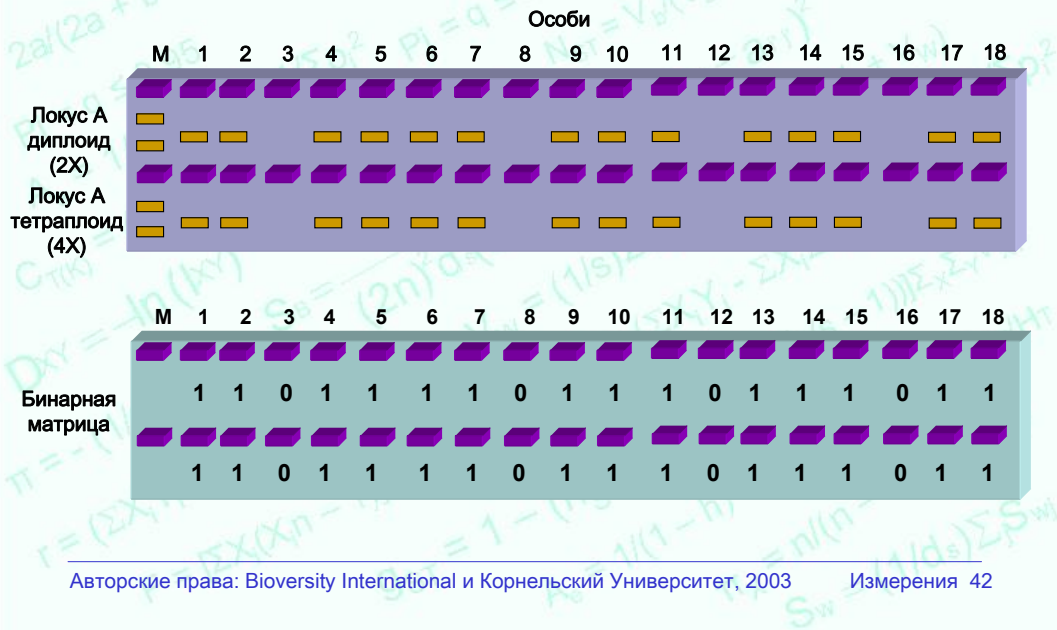
Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 41

При использовании данных молекулярного маркера и их преобразования в бинарные, необходимо учитывать следующее:

- Число пloidности вида может маскировать присутствие аллельной серии в локусе. Если это происходит, то генетическое разнообразие будет недооценено при использовании доминантных маркеров (присутствие/отсутствии).
- Если маркер кодоминантный, то нужны большие образцы для того, чтобы дать возможность обнаружить все возможные генотипы, особенно, если приходится несколько аллелей на локус.
- Нарушения расщепления – это общее явление для полиплоидных видов.
- Большинство компьютерных программ разработаны для диплоидных видов. Поэтому, если их использовать для полиплоидных видов, то могут возникнуть отклонения при оценке различных индексов генетического разнообразия.
- Репродуктивная система некоторых видов не изучена, и поэтому их тип наследования недостаточно изучен.
- Наибольший возможный охват (регионов кодирования и некодирования) генома изучаемого вида должен быть отобран и проанализирован, чтобы расчеты генетического разнообразия были достоверными.

Расчет частот аллели для диплоидов и тетраплоидов: доминантный маркер

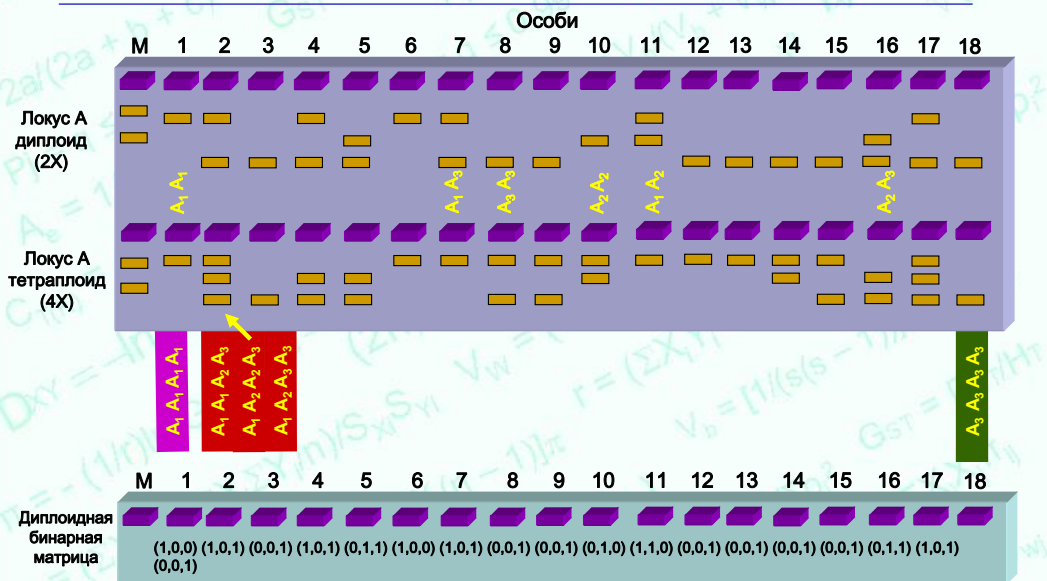


В этом примере проанализированы с доминантным маркером по 18 особей из каждого диплоидного и тетраплоидного вида. Мы предполагаем, что полученные рестриктивные фрагменты одинаковые. Полосы были преобразованы в бинарную таблицу в обоих случаях. Расчеты частот приведены в таблице ниже. Мы можем видеть, что например, в тетраплоиде, генотип 1, может быть либо АААА, АААа, ААаа и Аааа; однако, полоса будет подсчитана только как присутствует (1) также как для диплоида (АА и Аа).

Локус	Генотипы				Частоты аллели	
	Диплоид	АА, Аа	аа	Итого	p	q
А (2X)	Част.ген. (р.)	$p^2 + 2pq$	q^2	1		
	Число особей	14	4	18		
	Част.ген. (н.)	$P_1 = 0,78$	$P_2 = 0,22$	1	0,53	0,47
	Тетраплоид	АААА, АААа, ААаа, Аааа	аааа	Итого	p	q
А (4X)	Част.ген. (р.)	$p^4 + 4p^3q + 6p^2q^2 + 4pq^3$	q^4	1		
	Число особей	14	4	18		
	Част.ген. (н.)	$P_1 = 0,78$	$P_2 = 0,22$	1	0,31	0,69

Частоты аллели должны быть разными в обоих случаях; однако, потеря информации в тетраплоидной особи является значительной. Почему? Это происходит потому, что для оценки частоты рецессивной аллели а гетерозиготы АААа, ААаа, Аааа не приняты в расчет. Этот результат больше, когда число пloidий исследуемого вида неизвестно. (р. = расчетное значение; н. = наблюдаемое значение.)

Расчет частот аллели для диплоидов и тетраплоидов: кодоминантный маркер



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 43

В этом примере проанализированы с кодоминантным маркером по 18 особей из каждого диплоидного и тетраплоидного вида. Обнаружен один локус (A) с тремя аллелями в обеих ситуациях (A_1 , A_2 и A_3).

Расчет частот аллели в диплоидной особи не представляет трудности (бинарная матрица внизу слайда). Однако, для тетраплоидной особи перевод в бинарные данные затруднен тем фактом, что особи с аллелями $A_1 A_1 A_2 A_3$ не могут быть отличены от тех, что имеют другие комбинации как $A_1 A_2 A_2 A_3$ или $A_1 A_2 A_3 A_3$. Эта проблема может быть решена только как вывод, основанный на оценке числа копий фрагмента ДНК в геле.

Генотип	$A_1 A_1$	$A_1 A_2$	$A_1 A_3$	$A_2 A_2$	$A_2 A_3$	$A_3 A_3$	Итог о	p	q	r
Част.ген. (р.)	p^2	$2pq$	$2pr$	q^2	$2qr$	r^2	1	0,25	0,15	0,60
Число особей	2	1	4	1	2	8	18			
Част.ген. (н.)	$P_{11} = 0,11$	$P_{12} = 0,06$	$P_{13} = 0,22$	$P_{22} = 0,06$	$P_{23} = 0,11$	$P_{33} = 0,44$	1			

(р. = расчетное значение;
о. = наблюдаемое значение)

Сходство коэффициентов бинарных переменных: примеры

	Автор	Экспрессия	Пример значений коэффициента, если $a = 3, b = 1, c = 3, d = 2$
S1	Russel и Rao (1940)	a/n	0,333
S2	Simpson	$a/\min[(a + b), (a + c)]$	0,750
S3	Braun-Blanquet	$a/\max[(a + b), (a + c)]$	0,500
S4	Dice (1945); Nei и Li (1979)	$a/[a + (b + c)/2]$	0,600
S5	Ochiai (1957)	$a/[(a + b)(a + c)]^{1/2}$	0,612
S6	Kulczynski 2	$(a/2)[1/(a+b)] + [1/(a+c)]$	0,625
S7	Jaccard (1900, 1901, 1908)	$a/(a + b + c)$	0,429
S8	Sokal и Sneath 5 (1963)	$a/[a + 2(b + c)]$	0,273
S9	Kulczynski 1 (1928)	$a/(b + c)$	0,750
S10	Sokal и Michener (1958)	$(a + d)/n$	0,556
S11	Rogers и Tanimoto (1960)	$(a + d)/[a + d + 2(b + c)]$	0,385
S12	Sokal и Sneath 1 (1963)	$(a + d)/[a + d + (b + c)/2]$	0,714
S13	Sokal и Sneath 3 (1963)	$(a + d)/(b + c)$	1,250

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 44

		Особь _j		
		1	0	
Особь _i	1	a	b	a + b
	0	c	d	c + d
		a + c	b + d	n

где,

$$n = a + b + c + d$$

В таблице выше мы видим, что:

Индексы от S1 до S9 дают значение только в отношении информации наличия;

Индексы от S10 до S13 дают значение как о наличии, так и отсутствии

Далее мы обсудим три индекса (выделены красным в таблице выше): простого совпадения (S10), Джаккарда (Jaccard) (S7) и Нея-Ли (Nei-Li) (S4).

Индексы геометрической дистанции

Коэффициент простого совпадения, или коэффициент простого соответствия :

$$(a + d)/(a + b + c + d)$$

Коэффициент Джаккарда (Jaccard) :

$$a/(a + b + c)$$

Коэффициент Нея-Ли (Nei-Li), или Дайса (Dice):

$$2a/(2a + b + c)$$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 45

Эти три индекса отличаются по их подходу в оценке числа совпадений и различий.

Коэффициент простого совпадения полагает, что отсутствие соответствует гомозиготным локусам. Он может использоваться с данными доминантного маркера (RAPD и AFLP) так как отсутствие могло бы соответствовать гомозиготным рецессивам. Пример применения коэффициента простого совпадения для категориальных переменных приведен в Приложении 6 ([щелкните мышкой здесь](#)).

Коэффициент Джаккарда только подсчитывает полосы, присутствующие для той или другой особи (i или j). Двойное отсутствие трактуется как пропущенные данные. Если ложно-положительные или ложно-отрицательные данные возникают, то оценка индекса имеет тенденцию быть с погрешностью. Он может применяться с данными кодоминантного маркера.

Коэффициент Нея-Ли подсчитывает процентное содержание совместных полос между двумя особями и придает большую значимость тем полосам, которые присутствуют в обоих. Он полагает, что отсутствие имеет меньшее биологическое значение и, таким образом, этот коэффициент имеет полное значение в отношении схожести ДНК. Он может применяться с данными кодоминантного маркера (RFLP, SSR).

Модель дисбаланса: генетическое расстояние

- ▶ Измеряет различие между двумя генами, пропорциональное ко времени отделения от общего предка
- ▶ Возможны несколько моделей :
 - Мутация множества аллелей
например, генетическое расстояние Нея
 - Модель поэтапной мутации
Например, расстояние с использованием микросателлитов
 - Мутация в последовательности нуклеотида

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 46

- Мутации бесконечных аллелей (например, изоферменты)
 - Каждый случай мутации дает начало новой аллели.
 - Если 2 гена одинаковые, то мутация не возникает. Если 2 гена разные, то возникает неизвестное число мутаций.
 - Среднее число мутаций со времени t , когда они отошли от предка, равно $= 2\mu t$, где μ есть уровень мутации и увеличен в 2 раза так как мы имеем дело с 2 независимыми генами.
 - Вероятность, что 2 гена являются общими по наследству по прошествии времени t равно $P = e^{-2\mu t}$
- Модель поэтапной мутации (например, SSR)
 - Мутация- это прогрессивное изменение, происходящее таким образом, что фрагменты, мигрирующие на одинаковые расстояния, имеют незначительное количество мутаций.
 - В случае SSR, предполагается, что мутации изменяют число повторов, увеличивая или уменьшая их шаг за шагом. Можно продемонстрировать, что квадрат разницы в числе повторов между 2 микросателлитами пропорционален времени отхождения от общего предка.
- Мутация в последовательности нуклеотида
 - Указывает, что самое простое замещение – это мутация единичного основания.
 - Основным ограничением является потеря информации из-за незнания числа мутаций, которые могли бы иметь место в одном сайте. Для решения этой проблемы, некоторые методы предполагают вероятность транзиции (пурин → пуридин или пиридин → пиридин) и трансверсии (пурин → пиридин или пиридин → пуридин).

Расчет генетического расстояния Нея

- ▶ Стандартное генетическое расстояние Нея есть :

$$D_{XY} = -\ln(I_{XY})$$

- ▶ Он основан на концепции генетической идентичности (I_{XY}):

$$I_{XY} = \frac{J_{XY}}{\sqrt{(J_X J_Y)}}$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 47

(продолжение в следующем слайде)

где,

J_X = средняя гомозиготность в популяции X

J_Y = средняя гомозиготность в популяции Y

J_{XY} = средняя гомозиготность между популяциями

таким образом,

$I_{XY} = 1$, если две популяции имеют одинаковые частоты аллели во всех отобранных локусах

$I_{XY} = 0$, если две популяции не имеют совместных одинаковых частот аллели в отобранных локусах

- Значение D_{XY} варьирует от 0 (где популяции имеют идентичные частоты аллели) до бесконечности (∞ , где популяции не имеют ни одной совместной аллели).
- Предполагается, что уровень замещения на локус одинаковый среди всех локусов и популяций.
- Эта дистанция оценивает различия кодона на локус между двумя популяциями.

Расчет генетического расстояния Нея (продолжение)

Локус	Аллели	Частоты аллели		
		Поп,1	Поп,2	Поп,3
A	A ₁	0,80	0,74	0,65
	A ₂	0,20	0,26	0,35
Гетерозиготность локуса	h_{ijk}	0,3200	0,3848	0,4550
B	B ₁	0,86	0,81	1,00
	B ₂	0,01	0,10	0,00
	B ₃	0,13	0,09	0,00
Гетерозиготность локуса	h_{ijk}	0,2434	0,3258	0,0000
D	D ₁	0,00	1,00	0,30
	D ₂	1,00	0,00	0,70
Гетерозиготность локуса	h_{ijk}	0,0000	0,00	0,4200
Средняя гетерозиготность	H_i	0,0433	0,0547	0,0673
Средняя гомозиготность	J_i	0,9567	0,9453	0,9327
Средняя межпопуляционная гомозиготность	J_{ij}	$J_{1,2} = 0,8733$	$J_{1,3} = 0,9346$	$J_{2,3} = 0,8986$
Генетическая идентичность	I_{ij}	$I_{1,2} = 0,9183$	$I_{1,3} = 0,9894$	$I_{2,3} = 0,9570$
Генетическое расстояние	D_{ij}	$D_{1,2} = 0,0852$	$D_{1,3} = 0,0107$	$D_{2,3} = 0,0440$

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 48

Мы имеем пример, где $i = 3$ популяции, $j = 3$ полиморфных локуса и $s = 10$ мономорфными локусами. Более того, с разным числом (K) аллелей на локус (например, A и D имеют по 2 аллели каждый и B имеет 3 аллели).

Таблица показывает результаты расчета частот аллели в каждой популяции, а также гетерозиготность локуса. Затем мы рассчитываем среднюю гетерозиготность и гомозиготность ($1 -$ гетерозиготность) на популяцию.

И последнее, мы рассчитываем межпопуляционную гомозиготность и генетическую идентичность и, таким образом, мы можем оценить генетическое расстояние Нея. Расчеты производятся следующим образом:

$$j_{ijjk} = \sum_{i,j} p_{ijk} p_{ijjk}, \text{ например, } j_{1,2jk} = \text{гомозиготность среди популяций 1 и 2}$$

$$j_{1,2jk} = (0,8)(0,74) + (0,2)(0,26) + (0,86)(0,81) + (0,01)(0,10) + (0,13)(0,09) + (0,0)(1,0) + (1,0)(0,0) + 10 = 11,3533$$

$$J_{1,2} = \text{средняя межпопуляционная гомозиготность} = j_{1,2jk} / 13 = 11,3533 / 13 = 0,8733$$

$$I_{1,2} = \text{генетическая идентичность между популяциями 1 и 2} = J_{1,2} / \sqrt{(J_1 J_2)} = 0,8733 / \sqrt{(0,9567 * 0,9453)} = 0,9183$$

$$D_{1,2} = \text{генетическое расстояние между популяциями 1 и 2} = -\ln(I_{1,2}) = -\ln(0,9183) = 0,0852$$

Так как мы пока еще не разъяснили методы образования кластеров, то мы приводим дистанционную матрицу и дендрограмму этого примера в Приложении 7 ([щелкните мышкой здесь](#)).

Расчет внутрипопуляционного расстояния с использованием микросателлитов

- ▶ Расстояние внутри популяции - это среднее суммы квадратов разности в числе повторов между аллелями

$$S_{wi} = \frac{2}{2n(2n-1)} \sum_{i < i'} (a_{ii} - a_{i'i'})^2$$

- ▶ Среднее внутрипопуляционное расстояние может быть рассчитано для анализируемых локусов (d_s)

$$S_w = (1/d_s) \sum_j S_{wj}$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 49

где,

a_{ij} = размер аллели i -ой копии ($i = 1, 2, \dots, 2n$) в j -ой популяции ($j = 1, 2, \dots, d_s$)

n = число особей в образце

Два соображения:

Расчет расстояния между двумя аллелями- это трансформация числа повторов.

Трудность в использовании SSR для расчета генетических расстояний заключается в их высоком уровне мутации

Расчет межпопуляционного расстояния с использованием микросателлитов

Это межпопуляционный компонент для среднего расстояния среди сравнений всех парных аллелей

$$S_B = \frac{2}{(2n)^2 d_s (d_s - 1)} \sum_{j < j'} \sum_{i < i'} (a_{ij} - a_{i'j'})^2$$

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 50

Глобальное расстояние – это средневзвешенное среди внутри- и между популяций компонента.

$$\hat{S} = \frac{2n-1}{(2nd_s - 1)} S_w + \frac{2n(d_s - 1)}{(2nd_s - 1)} S_B$$

Эти коэффициенты представляют вероятность выбора двух разных копий одного локуса в одной и той же популяции и между двумя популяциями.

Полезными компьютерными программами являются: MICROSAT, BIOSYS, GENEPOP, GDA и POPGENE.

Демонстрация взаимосвязи: классификация или образование кластеров

Это процесс группирования (или образование кластеров) объектов в категории или классы на основании их общих признаков или взаимосвязей. Группирование может быть:

- Иерархическим :
 - Основным, который пытается раскрыть истинную природу или форму
 - Кладистическим, основанным на генеалогии и филогенетике
 - Эволюционным, основанным на филогенетике и числе эволюционных изменений
 - Фенным, основанным на наибольшем числе признаков организма и его жизненном цикле
- Неиерархическим
- Частично совпадающим

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 51

- Иерархическое: основной класс, который содержит второстепенные классы, называемые «ветви».
- Неиерархическое: каждая особь отнесена к уникальной группе путем сравнения ее с исходными классами таким образом, что ее положение является наиболее соответствующей.
- Частично совпадающее: особи могут принадлежать к более, чем одной группе.

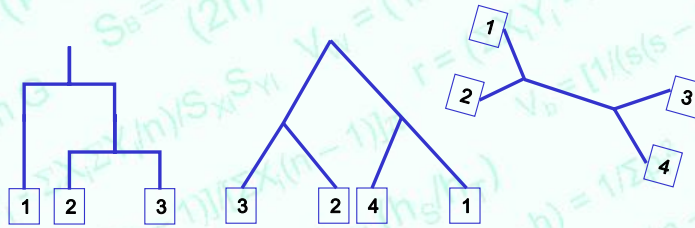
Типы классификации относятся к процедурам каталогизации объектов, организмов и т.д. и используются в нескольких областях знаний. В нашем случае мы используем иерархическую классификацию из-за природы взаимосвязей между особями, а именно, особь, популяция, образец, сорт, пр. являются единицами, которые не могут быть отнесены к двум разным группам в одно и то же время.

Литература

García, J.A., M.C. Duque, J. Tohme, S. Xu и M. Levy. 1995. SAS for Classification Analysis; Agrobiotechnology Course, October 1995. Working document. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

Классификация по фену

- ▶ Показывает взаимосвязь между образцами путем использования индекса сходства
- ▶ Метод группирования или расстояние выбраны таким образом, что древовидная схема (дендрограмма) или фенограмма (если матрица сходства содержит фенотипические данные) может быть изображена



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 52

В этом примере иерархического группирования всем признакам приданы одинаковая важность в процессе группирования.

Общее суммарное сходство между двумя группами - это сумма сходств каждого признака.

Она не учитывает генеалогию.

Фенная относится к любому признаку, использованному в процедуре классификации: или морфологическому, физиологическому, экологическому, молекулярному, или цитологическому.

Методы образования кластеров

► Этапы образования кластеров :

- Определение близости
- Оценка каждой группировки в соответствии с дистанцией
- Построение ветвей дендрограммы в каждом цикле

► Существует три метода :

- Простая связь (или «ближайший сосед»)
- Абсолютная связь (или «самый дальний сосед»)
- Средняя связь (или UPGMC)

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 53

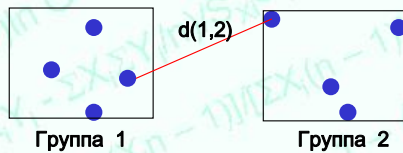
Имеются другие методы образования кластеров как:

- Метод невзвешенных парных групп с использованием центраида (Unweighted pair-group method using the centroid (UPGMC)). Он основан на дистанции между средними значениями каждой группы.
- Метод взвешенных парных групп с использованием центраида (Weighted pair-group method using the centroid (WPGMC)). Он основан на среднем значении ОТЕ в группах.
- Метод Варда. Он работает с суммой квадратов дистанций для пары ОТЕ. Он также известен как метод минимальной вариации, так как при взятии значений в квадрате он становится очень чувствительным методом (разные ОТЕ будут выглядеть более несходными, а сходные ОТЕ будут выглядеть более похожими). Он может использоваться с Эвклидовыми дистанциями и молекулярными данными, когда имеется большое число полос ДНК.

В следующих нескольких слайдах мы обсудим более подробно три метода, перечисленных в слайде выше, и приведем пример для каждого из них.

Простая связь

- ▶ Или «ближайший сосед»
- ▶ Она минимально уменьшает межгрупповое расстояние путем выбора расстояния к соседу с самой высокой схожестью
- ▶ Она работает с регулярными и компактными группами, но сильно подвергается воздействию со стороны дальних особей. Это неудобно, когда разные группы хорошо не распространены в пространстве



$d(1,2)$ = минимальное расстояние между двумя ОТЕ

Простая связь: пример

(1)

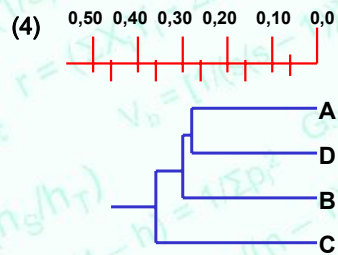
	A	B	C	D
A	0			
B	0,30	0		
C	0,43	0,35	0	
D	0,28	0,60	0,40	0

(2)

	B	C	AD
B	0		
C	0,35	0	
AD	0,30	0,40	0

(3)

	C	ADB
C	0	
ADB	0,35	0



Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 55

1. Сначала формируется матрица расстояния; затем, в первом цикле, выбирается самое короткое расстояние $d_{AD} = 0,28$.
2. Формируется новая матрица путем группировки особей A и D и расчета комбинированных расстояний:

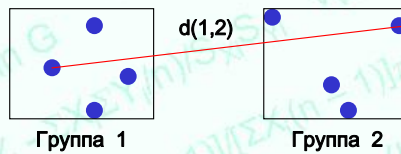
$$d_{B(AD)} = \min(d_{BA}; d_{BD}) = \min(0,30; 0,60) = 0,30$$

$$d_{C(AD)} = \min(d_{CA}; d_{CD}) = \min(0,43; 0,40) = 0,40$$
3. Формируется новая матрица путем группировки особи B с группой (AD) и расчета комбинированных расстояний

$$d_{C(ADB)} = \min(d_{AC}; d_{CD}; d_{CB}) = \min(0,43; 0,40; 0,35) = 0,35$$
4. Рисуется дендрограмма.

Абсолютная связь

- ▶ Или «самый дальний сосед»
- ▶ Она максимально увеличивает межгрупповое расстояние путем выбора расстояния к соседу с минимальной схожестью
- ▶ Она работает с регулярными и компактными группами, но опять же сильно подвергается воздействию со стороны дальних особей.



$d(1,2)$ = значительное расстояние между двумя ОТЕ

Абсолютная связь: пример

(1)

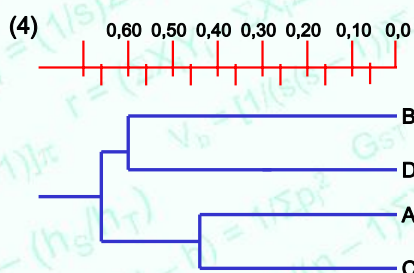
	A	B	C	D
A	0			
B	0,30	0		
C	0,43	0,35	0	
D	0,28	0,60	0,40	0

(2)

	A	C	BD
A	0		
C	0,43	0	
BD	0,30	0,40	0

(3)

	AC	DB
AC	0	
DB	0,40	0



Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 57

1. Сначала формируется матрица расстояния; затем, в первом цикле, выбирается самое длинное расстояние $d_{BD} = 0,60$,
2. Формируется новая матрица путем группировки особей B и D и расчета комбинированных расстояний:

$$d_{A(BD)} = \max(d_{BA}; d_{AD}) = \max(0,30; 0,28) = 0,30$$

$$d_{C(BD)} = \max(d_{CB}; d_{CD}) = \max(0,35; 0,40) = 0,40$$

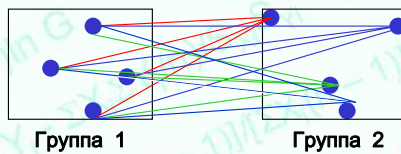
3. Формируется новая матрица с группами AC и BD и расчета комбинированных расстояний

$$d_{(AC)(DB)} = \max(d_{AD}; d_{AB}; d_{CD}; d_{CB}) = \max(0,28; 0,30; 0,40; 0,35) = 0,40$$

4. Рисуется дендрограмма.

Средняя связь

- ▶ Или «метод невзвешенного попарного группирования с использованием арифметической средней» (UPGMA)
- ▶ Он минимизирует межгрупповое расстояние путем учета среднего попарного расстояния среди всех особей образца
- ▶ Наиболее используемый метод



$d(1_i, 2_j)$ = среднее
расстояние между OTE_i и
 OTE_j групп 1 и 2

Средняя связь: пример

(1)

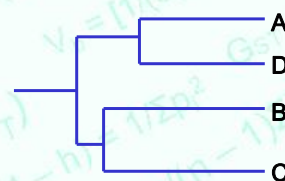
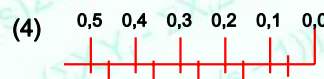
	A	B	C	D
A	0			
B	0,30	0		
C	0,43	0,35	0	
D	0,28	0,60	0,40	0

(2)

	B	C	AD
B	0		
C	0,35	0	
AD	0,45	0,415	0

(3)

	BC	AD
BC	0	
AD	0,42	0



Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 59

1. Сначала формируется матрица расстояния; затем, в первом цикле, выбирается самое короткое расстояние $d_{AD} = 0,28$.
2. Следующее, формируется новая матрица путем группировки особей A и D и расчета комбинированных расстояний:

$$d_{B(AD)} = (d_{BA} + d_{BD})/2 = (0,30 + 0,60)/2 = 0,45$$

$$d_{C(AD)} = (d_{CA} + d_{CD})/2 = (0,43 + 0,40)/2 = 0,415$$

3. Формируется новая матрица путем группировки особей с самым коротким расстоянием B с C и расчета комбинированных расстояний

$$d_{(AD)(BC)} = (d_{AB} + d_{AC} + d_{BD} + d_{BC})/4 = (0,30 + 0,43 + 0,60 + 0,35)/4 = 0,42$$

Выбирая метод образования кластеров

- ▶ Первое, соберите знания об изучаемом виде: его разнообразии, системе репродукции, плоидности и уровнях гетерозиготности
- ▶ Осторожно отберите генетические признаки для анализа
- ▶ Апробируйте разные методологии образования кластеров и оцените уровень согласия, достигнутый с каждым из них

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 60

К тому же всегда будет важно объединять как можно больше информации. Пример приведен в Приложении 8 ([щелкните мышкой здесь](#)), где имеются как морфологические, так и молекулярные данные, а также использование отдельных наборов данных сравнивается с использованием объединенных данных.

Проверка правильности анализа образования кластеров

- ▶ Внешняя проверка правильности
- ▶ Внутренняя проверка правильности
- ▶ Относительная проверка правильности
- ▶ Bootstrap анализ

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 61

Внешняя проверка правильности:

Матричное расстояние сравнивается с другой информацией, не использованной в расчетах группирования, например, генеалогия.

Внутренняя проверка правильности :

Эта технология количественно выражает искажение в результате использованного метода группирования. Она строит новую матрицу подобия или дистанции, матрицу «ко-фена», прямо из дендрограммы. Проверка правильности рассчитывается при помощи коэффициента корреляции между данными подобия или расстояния из исходной матрицы и данными новой матрицы ко-фена. Поддерживаются ли исходные расстояния оценивается после осуществления группирования (Sokal и Rohlf, 1994).

Относительная проверка правильности:

Сравнивается схожесть между методами.

Проверка правильности при помощи Bootstrap анализа :

Это метод повторного отбора образцов путем замены той же матрицей данных. Он позволяет вычислить стандартные отклонения и дисперсии и полезен в тех ситуациях, где число образцов или ресурсов, например, времени, бюджета, ограничено.

Примеры применения методов ко-фенетической корреляции и проверка правильности Bootstrap анализа приведены далее.

Литература

Sokal, R. и J. Rohlf. 1994. Biometry: The Principles and Practice of Statistics in Biological Research (3-я ред.). Freeman & Co, NY.

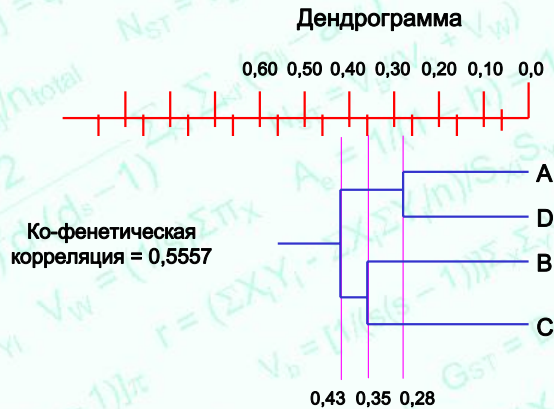
Ко-фенетическая корреляция: пример

	A	B	C	D
A	0			
B	0,30	0		
C	0,43	0,35	0	
D	0,28	0,60	0,40	0

Исходная дистанционная матрица

	A	B	C	D
A	0			
B	0,43	0		
C	0,43	0,35	0	
D	0,28	0,43	0,43	0

Ко-фенетическая матрица



Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 62

Для конструирования матрицы ко-фена мы взглянем на дендрограмму, ранее построенную с исходной матрицей (этот пример приведен в слайде 58). Мы видим, что дистанция между D и C в дендрограмме равна 0,43, и мы заполняем ячейку в матрице ко-фена. Дистанция между B и C равна 0,35 и так далее.

Расчеты для корреляции ко-фена основаны на коэффициенте корреляции:

$$r = (\sum X_i Y_i - \sum X_i \sum Y_i / n) / S_{X_i} S_{Y_i}$$

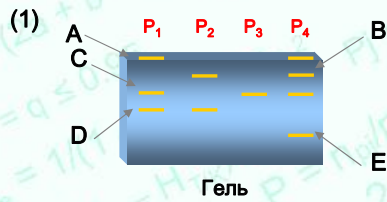
где,

X_i и Y_i – это значения сходности или дистанции исходной и ко-фенной матрицы соответственно

S_{X_i} и S_{Y_i} – это стандартные отклонения для каждой переменной

Если значение корреляции высокое, мы можем сделать вывод, что дендрограмма в действительности отражает дистанции в исходной матрице, и таким образом не имеется искажения в результате метода группирования. В примере выше мы получили значение 0,5557. Это среднее значение, которое могло бы указать, что дистанции дендрограммы не отражают данные дистанции в исходной матрице, и таким образом, существует искажение в результате использованного метода. Однако, в примере построения мы использовали очень незначительное количество данных; и даже они не были действительными результатами эксперимента, что и объясняет полученные значения.

Оценка правильности при помощи Bootstrap анализа: пример



(2)

	P ₁	P ₂	P ₃	P ₄
L ₁	1	0	0	1
L ₂	0	1	0	1
L ₃	1	0	1	1
L ₄	1	1	0	0
L ₅	0	0	0	1

Матрица данных

(3)

	P ₁	P ₂	P ₃	P ₄
P ₁	1			
P ₂	0,400	1		
P ₃	0,600	0,400	1	
P ₄	0,400	0,200	0,400	1

Матрица сходства

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 63

(продолжение в следующем слайде)

В геле выше (верхний левый угол) мы имеем 4 особи (P_i) и 5 локусов (L_j). Мы предположим, что выполняем оценку правильности в трех образцах с замещением.

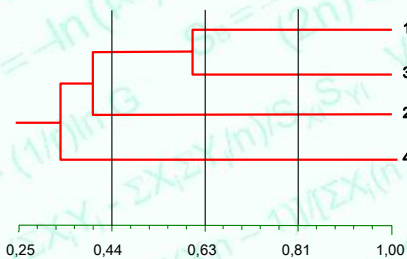
Сначала мы подсчитаем данные маркера в особях (матрица данных), и затем мы вычислим среднее сходство (простое совпадение) и его интервал.

Оценка правильности при помощи Bootstrap анализа: пример (продолжение)

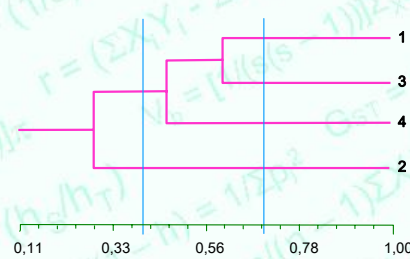
Матрица среднего сходства с стандартными отклонениями

	P ₁	P ₂	P ₃	P ₄
P ₁	1			
P ₂	0,267 ± 0,115	1		
P ₃	0,600 ± 0,000	0,400 ± 0,200	1	
P ₄	0,533 ± 0,115	0,200 ± 0,000	0,400 ± 0,200	1

Дендрограмма до замещения



Дендрограмма после замещения



Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 64

Для каждой особи, значение для каждого локуса взято по одному, с замещением, и сформирован образец с одинаковым числом локусов. Существует вероятность, что локус выбран один или более раз. Например:

M₁: L₁ L₁ L₂ L₃ L₅ (локус L₄ не нарисован)

M₂: L₁ L₂ L₃ L₄ L₃

M₃: L₃ L₁ L₅ L₂ L₄

В каждом образце вычислена матрица сходства.

Средние сходства и их стандартные отклонения рассчитаны для каждой пары особей (1 & 2, 1 & 3, 2 & 3, и так далее), и создана матрица среднего сходства.

Построена новая дендрограмма с использованием матрицы среднего сходства.

Для реальных ситуаций необходимо создать более 100 образцов замещения.

Демонстрация взаимосвязей: расположение

- ▶ Расположение – это систематизация или «упорядочение» единиц образца вдоль систем координат
- ▶ Цель расположения, а также методов классификации, заключается в интерпретации паттерна в структуре образцов

Авторские права: Bioversity International и Корнельский Университет, 2003 Измерения 65

Расположение – это многомерный метод, который дополняет образование кластеров, и обычно рассматривается как подход, который ближе к биологической реальности.

При помощи методов расположения мы хотим представить взаимосвязи образцов простым способом путем уменьшения фактической ситуации до «низкомерного пространства» (Gauch, 1982). При этом состав образцов изучается в целом, статистическая сила анализа улучшается, так как избыток так или иначе устраняется или сокращается, и относительная важность разных градиентов может быть установлена. Более всего, мы получаем графические представления, которые помогают нам интуитивно толковать взаимосвязи разных групп образцов.

В принципе, расположение – это инструмент как для исследований, так и для апробации гипотез. В любом случае, результаты, полученные при помощи методов расположения, должны всегда сравниваться с имеющимися знаниями об изучаемом образце и насколько это возможно с дополнительной информацией, относящейся к исследуемому биологическому вопросу.

Литература

Gauch, H.G., Jr. 1982. *Multivariate Analysis and Community Structure*. Cambridge University Press, UK.

Полезные методы расположения для данных молекулярного маркера

- ▶ Основной анализ координат (ОАК)
- ▶ Неметрическое многомерное масштабирование (НММ)
- ▶ Анализ аналогий (АА)

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 66

Существует много способов расположения – некоторые из них основаны на данных расстояния или на расчетах, так называемых, собственных значений (суммы всех вариантов для каждого признака в каждом компоненте). Однако, эти способы, основанные на непрерывных величинах (например, основной анализ компонента или ОАК) не подходят для использования с маркерными данными. Следовательно, мы обсудим вкратце только три способа, перечисленные в слайде выше. Подробное изложение основ этих методов потребовало бы более глубокого математического понимания использованных алгоритмов, чем тот уровень, которым как мы предполагаем обладает средний пользователь настоящего модуля. Поэтому мы поддерживаем наших читателей, которые хотят знать больше об этих методах, в поиске методов расположения в Интернете. Для обзора щелкните мышкой на сайт <<http://www.okstate.edu/artsci/botany/ordinate/overview.htm>>

Основной анализ координат (ОАК) пытается представить расстояния между образцами и может вмещать матрицы из разных несходных измерений. Он максимально увеличивает линейную корреляцию между расстояниями образца. При использовании Эвклидовых расстояний результаты идентичны ОАК.

Неметрическое многомерное масштабирование (НММ) работает путем максимального увеличения порядковой корреляции и попыткой найти наилучшую форму для вмещения данных. Этот метод раскрывает основную конфигурацию из матрицы несхожести образца. С НММ только набор точек имеет значение, а не их происхождение, и изображение может чередоваться.

Анализ соответствия (АА) повторяет средние подсчеты образца и находит места, где все образцы, охватывающие одно и то же место, настолько похожи насколько возможно и, одновременно, образцы в разных местах настолько разные, насколько это возможно.

Приложения

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 67

Для просмотра приложения щелкните мышкой на его названии:

2. [Анализ молекулярной вариансы: пример 1](#)
3. [Анализ молекулярной вариансы: пример 2](#)
4. [Геометрическое расстояние](#)
5. [Преобразование данных количественных переменных: пример](#)
6. [Применение коэффициента простого совпадения для морфологических признаков \(категориальные переменные\)](#)
7. [Расчет генетического расстояния Нея](#)
8. [Морфологические и молекулярные сходства](#)

Кратко

- ▶ Анализ генетического разнообразия и структуры популяций включает :
 - Количественное выражение разнообразия и взаимосвязей между и внутри популяциями и/или особями
 - Демонстрацию взаимосвязей
- ▶ Молекулярные данные обычно трактуются как бинарные данные
- ▶ Молекулярные данные могут быть обычно дополнены морфологическими и эволюционными данными. Для этого типы переменных могут быть преобразованы в бинарные переменные

Теперь Вы должны знать ...

- ▶ Основные этапы, вовлеченные в измерение генетического разнообразия
- ▶ Главные методы описания генетического разнообразия внутри и между популяциями
- ▶ Правильный выбор расчета расстояния для оценки взаимосвязей в изучаемом образце
- ▶ Различия между альтернативными методами образования кластеров
- ▶ Имеющиеся варианты для оценки правильности группирования
- ▶ Основные понятия, лежащие в основе концепции классификации
- ▶ Сходства и различия между образованием кластеров и расположением

Библиография

Авторские права: Bioversity International и Корнельский Университет, 2003

Измерения 70

- Cavalli-Sforza, L.L. и W.F. Bodmer. 1981. *Genética de las Poblaciones Humanas*. Ed. Omega, Barcelona.
- García, J.A., M.C. Duque, J. Tohme, S. Xu и M. Levy. 1995. SAS for Classification Analysis; Agrobiotechnology Course, October 1995. Working document. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.
- Gauch, H.G., Jr. 1982. *Multivariate Analysis and Community Structure*. Cambridge University Press, UK.
- Karp, A., P.G. Isaac и D.S. Ingram. 1998. *Molecular Tools for Screening Biodiversity: Plants and Animals*. Chapman & Hall, London.
- Lynch, M. и B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* 3:91-99.
- Sokal, R. и J. Rohlf. 1994. *Biometry: The Principles and Practice of Statistics in Biological Research* (3rd edn.). Freeman & Co, NY.

Следующее

- ▶ Компьютерные программы для анализа генетического разнообразия
- ▶ Глоссарий

Приложение 2:

Измерение генетического разнообразия

Анализ молекулярной варiances: Пример 1

Эта модель, также называемая AMOVA, измеряет разнообразие генов среди популяций с специфичным соотношением к областям региона в континенте (ситуация 3, слайд 26).

Мы имеем: i = особи, j = аллели, k = популяции

$$Y_{ki(j)} = Y + a_k + b_{k(i)} + w_{ki(j)}$$

где,

- $Y_{ki(i)}$ = значение между 0 и 1 для j -ой аллели i -ой особи k -ой популяции
- a_k = влияние k -ой популяции с дисперсией $\sigma^2 a$
- $b_{k(i)}$ = влияние i -ой особи внутри k -ой популяции, с дисперсией $\sigma^2 b$
- $w_{ki(i)}$ = влияние j -го локуса i -ой особи k -ой популяции, с дисперсией $\sigma^2 w$
- n = продукт i, j и k , есть общее число наблюдений

Источник изменений	df	SS	MS	EMS
Среди популяций	$(k - 1)$	$\sum X_{...k}^2 / ij - X_{...}^2 / jk$	MS_a	$\sigma^2 w + 2\sigma^2 b + 2n\sigma^2 a$
Среди особей/популяций	$k(i - 1)$	$\sum \sum X_{i...k}^2 / j - \sum ...k^2 / ij$	MS_b	$\sigma^2 w + 2\sigma^2 b$
Внутри особей	$ki(j - 1)$	$\sum \sum \sum X_{ijk}^2 - \sum \sum X_{i...k}^2 / j$	MS_w	$\sigma^2 w$
Итого	$kij - 1$	$\sum \sum \sum X_{ijk}^2 - X_{...}^2 / jk$		

Варианты и расчет F статистики	
$\sigma^2 a = F_{ST} \sigma^2$	$F_{IT} = (\sigma^2 a + \sigma^2 b) / \sigma^2$
$\sigma^2 b = (F_{IT} - F_{ST}) \sigma^2$	$F_{ST} = \sigma^2 a / \sigma^2$
$\sigma^2 w = (1 - F_{IT}) \sigma^2$	$F_{IS} = \sigma^2 b / (\sigma^2 b + \sigma^2 w)$
$\sigma^2 = \sigma^2 w + \sigma^2 b + \sigma^2 a$	

где,

- $\sigma^2 a$ = параметрическое значение дисперсии между популяциями, несущая идентичные аллели. Оно оценивается как $(MS_a - MS_b) / 2n$
- $\sigma^2 b$ = параметрическое значение дисперсии между особями внутри каждой популяции. Оно рассчитывается как $(MS_b - MS_w) / 2$
- $\sigma^2 w$ = параметрическое значение дисперсии внутри особей или измерение вероятности того, что аллели внутри локусов различны. Оно рассчитывается как среднее квадратичное внутри особей (CMW)

Приложение 3:

Измерение генетического разнообразия

Анализ молекулярной варiances: Пример 2

Как описано в Приложении 2, модель AMOVA измеряет разнообразие генов среди популяций на этот раз с специфичным соотношением к популяциям в пределах области региона в континенте (ситуация 4, слайд 26). Она имеет новый иерархический уровень (регион) с его соответствующими параметрическими значениями и расчетами средних квадратичных значений.

Мы имеем : i = особи, j = аллели, k = популяции, l = регионы

$$Y_{lki(j)} = Y + r_l + a_{l(k)} + b_{lk(i)} + w_{lki(j)}$$

где,

- $Y_{lki(j)}$ = значение между 0 и 1 для j -ой аллели i -ой особи k -ой популяции в l -ом регионе
- r_l = влияние l -ого региона с дисперсией $\sigma^2 r$
- $a_{l(k)}$ = влияние k -ой популяции в пределах l -ого региона с дисперсией $\sigma^2 a$
- $b_{lk(i)}$ = влияние i -ой особи в пределах k -ой популяции в l -ом регионе, с дисперсией $\sigma^2 b$
- $w_{lki(j)}$ = влияние j -го локуса в пределах i -ой особи k -ой популяции в l -ом регионе, с дисперсией $\sigma^2 c$
- n = продукт i, j и k , есть общее число наблюдений

Источник изменений	df	MS	EMS
Между регионами	$l - 1$	MS_r	$\sigma^2 w + 2\sigma^2 b + 2n\sigma^2 a + 2nl\sigma^2 r$
Между поп. внутри региона	$l(k - 1)$	MS_a	$\sigma^2 w + 2\sigma^2 b + 2n\sigma^2 a$
Между особ/поп/ рег	$lk(i - 1)$	MS_b	$\sigma^2 w + 2\sigma^2 b$
Внутри особей	$lki(j - 1)$	MS_w	$\sigma^2 w$
Итого	$lkij - 1$		
Итоговая дисперсия (%)	$\% \sigma^2 r = (\sigma^2 r / \sigma^2) * 100$		$\% \sigma^2 a = (\sigma^2 a / \sigma^2) * 100$
$\sigma^2 = \sigma^2 r + \sigma^2 w + \sigma^2 b + \sigma^2 a$	$\% \sigma^2 b = (\sigma^2 b / \sigma^2) * 100$		$\% \sigma^2 w = (\sigma^2 w / \sigma^2) * 100$

$\sigma^2 r$ – параметрическое значение дисперсии между регионами и рассчитывается как $(CMA-CMB)/2nl$.

В расчетах дисперсии добавлен знак '%', так как мы можем выразить дисперсию, отвечающую за каждый источник (регион, популяции внутри региона, особи внутри популяции) как функцию итоговой дисперсии и, таким образом, мы можем определить который из компонентов изменения является наиболее важным. Например, если значение изменения из-за регионов было высоким, а для других источников низким, то мы можем заключить, что популяции внутри регионов имеют гомогенные аллельные частоты, а популяции из разных регионов заметно различаются по своим аллельным частотам.

Приложение 4:

Измерение генетического разнообразия

Геометрическое расстояние

Количественные переменные

Геометрическое расстояние, также известное как таксономическое расстояние (Sokal, 1961) измеряется Эвклидовыми расстояниями по формуле:

$$d_{ij} = [\sum_k (X_{ik} - X_{jk})^2]^{1/2}$$

где,

X_{ik} = значение k -ой переменной i -ой особи

См. пример расчета в Приложении 5.

Смешанные переменные

Если имеются смешанные переменные, то они должны быть сначала преобразованы или стандартизированы, используя формулу ниже:

$$X_{ijstand} = \frac{X_{ij} - \bar{X}_i}{S_i}$$

где,

X_{ij} = i -ый признак j -ой особи

\bar{X}_i = среднее значение для i -ого признака

S_i = стандартизированное отклонение для i -ого признака

P число переменных

Если имеется P число переменных, то расстояние должно быть взвешено, чтобы быть независимым от числа переменных по формуле:

$$d_{ij}^2 = \frac{\sum_k \left[\frac{(X_{ik} - X_{jk})}{\sigma_k} \right]^2}{P}$$

Литература

Sokal, R. 1961. Distance as a measure of taxonomic similarity. Syst. Zool. 10(2):40-51.

Приложение 5:

Измерение генетического разнообразия

Преобразование данных из количественных переменных: пример

Мы имеем три признака для четырех особей:

- Высота растения (m)
- Вес семян (g)
- Диаметр пыльцевого зерна (μ)

Прежде чем начать расчет расстояний, мы должны стандартизировать данные по формуле:

$$m_{\text{stand}} = m - \bar{m}/\sigma$$

После стандартизации, единицы измерения теряются.

	m	m_{stand}	g	g_{stand}	μ	μ_{stand}
Особь1	1,50	0,35	0,02	0,00	80,00	-0,15
Особь2	1,20	-1,41	0,03	1,00	70,00	-1,32
Особь3	1,45	0,06	0,01	-1,00	90,00	1,02
Особь4	1,60	0,94	0,02	0,00	85,00	0,44
Среднее (X_i)	1,44	—	0,02	—	81,25	—
Отклонение (s_i)	0,17	—	0,01	—	8,54	—

Расстояния могут быть рассчитаны теперь для каждой пары особей путем применения известной уже нам формулы:

$d_{ij} = [\sum(X_{ij} - X_{kj})^2]^{1/2}$	
$d_{12} = [(0,35 - (-1,41))^2 + (0,0 - 1,0)^2 + (-0,15 - (-1,32))^2]^{1/2} = 2,34$	$d_{11} = 0$
$d_{13} = [(0,35 - 0,06)^2 + (0,0 - (-1,0))^2 + (-0,15 - 1,02)^2]^{1/2} = 1,57$	$d_{22} = 0$
$d_{14} = [(0,35 - 0,94)^2 + (0,0 - 0,0)^2 + (-0,15 - 0,44)^2]^{1/2} = 0,83$	$d_{33} = 0$
$d_{23} = [(-1,41 - 0,06)^2 + (1,0 - (-1,0))^2 + (-1,32 - 1,02)^2]^{1/2} = 3,41$	$d_{44} = 0$
$d_{24} = [(-1,41 - 0,94)^2 + (1,0 - 0,0)^2 + (-1,32 - 0,44)^2]^{1/2} = 3,10$	
$d_{34} = [(0,06 - 0,94)^2 + (-1,0 - 0,0)^2 + (1,02 - 0,44)^2]^{1/2} = 1,45$	

Как только мы получим расстояния для пар, мы можем начать поиск групп, используя метод UPGMA (для более подробной информации см. слайды 58 и 59 настоящего модуля),

Сначала, мы систематизируем наши рассчитанные значения расстояний в симметричную таблицу:

	O ₁	O ₂	O ₃	O ₄
O ₁	0			
O ₂	2,34	0		
O ₃	1,57	3,41	0	
O ₄	0,83	3,10	1,45	0

В первом цикле мы выберем самое короткое расстояние, которое в нашем случае равно $d_{1,4} = 0,83$. Новая матрица может быть сформирована таким образом путем группирования Особи₁ с Особью₄ и расчета комбинированных расстояний:

$$d_{2(1,4)} = (d_{1,2} + d_{2,4})/2 = (2,34 + 3,10)/2 = 2,72$$

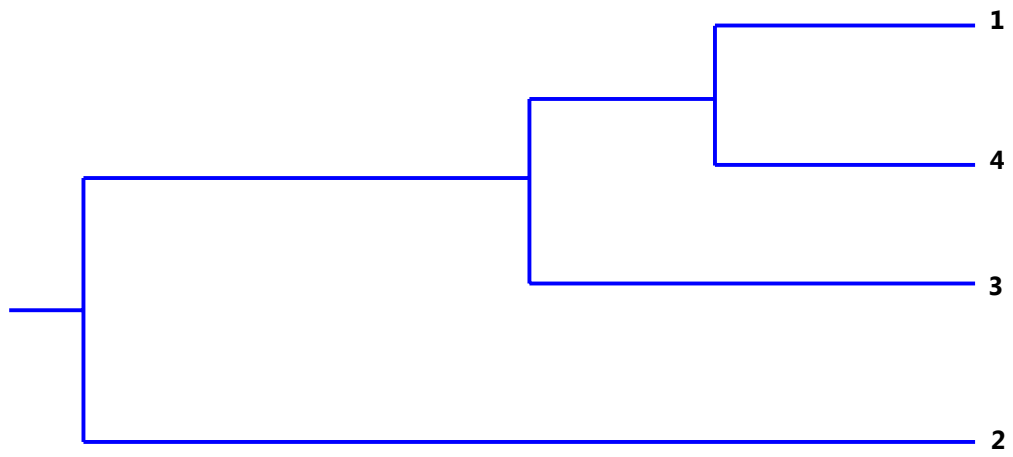
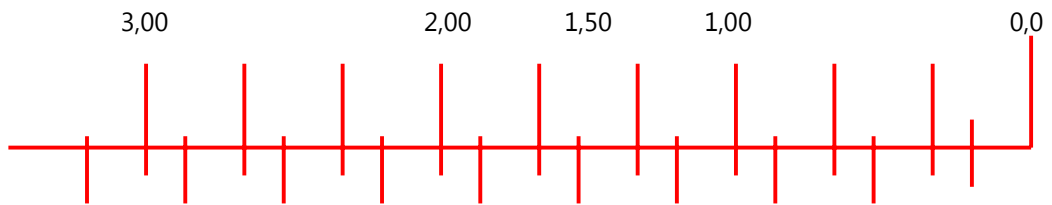
$$d_{3(1,4)} = (d_{1,3} + d_{3,4})/2 = (1,57 + 1,45)/2 = 1,51$$

	O _{1,4}	O ₂	O ₃
O _{1,4}	0		
O ₂	2,72	0	
O ₃	1,51	3,41	0

Мы видим, что наикротчайшее расстояние теперь между Особью_{1,4} и Особью₃. Далее, в новом цикле, формируется новая матрица путем группирования Особи₂ с группой O_{(1,4)3} и расчетом комбинированных расстояний $d_{((1,4)3)2} = 3,07$,

	O _{1,4(3)}	O ₂
O _{1,4(3)}	0	
O ₂	3,07	0

На основе полученных выше результатов мы можем нарисовать дендрограмму, связывающую четыре особи примера:



Приложение 6:

Измерение генетического разнообразия

Применение коэффициента простого совпадения для морфологических признаков (категориальные переменные)

Мы имеем три признака:

- Опушенность листа: редкая (1), обычная (2), обильная (3)
- Цвет лепестка: белый (1), желтый (2), красный (3)
- Длина черешка: короткий (1), средний (2), длинный (3)

Сначала мы преобразуем данные измерений в бинарные данные. Учтите, что три первоначальных признака преобразуются в 9 бинарных признаков. Это может утяжелить эти признаки в ущерб другим, используемым в анализе.

	Признак 1			Признак 2			Признак 3		
ОТЕ 1	2			1			2		
ОТЕ 2	2			3			3		
ОТЕ 3	1			2			1		
ОТЕ 4	3			3			1		
	Признак 1 (бинарный код)			Признак 2 (бинарный код)			Признак 3 (бинарный код)		
	Редк	Обычн.	Обильн	Белый	Желтый	Крас.	Корот.	Сред.	Длин.
ОТЕ 1	0	1	0	1	0	0	0	1	0
ОТЕ 2	0	1	0	0	0	1	0	0	1
ОТЕ 3	1	0	0	0	1	0	1	0	0
ОТЕ 4	0	0	1	0	0	1	1	0	0

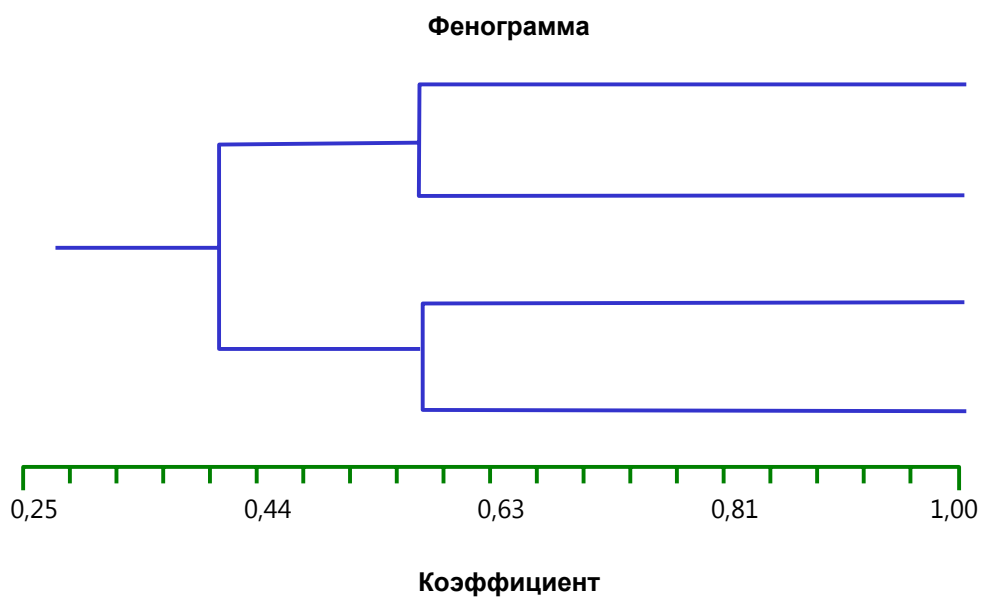
Далее мы применяем коэффициент простого совпадения для того, чтобы рассчитать попарные расстояния между особями:

Попарное сравнение для всех признаков									
ОТЕ 1 vs. 2	1	0	ОТЕ 1 vs. 3	1	0	ОТЕ 1 vs. 4	1	0	
1	a=1	b=2	1	a=0	b=3	1	a=0	b=3	
0	c=2	d=4	0	c=3	d=3	0	c=3	d=3	
ОТЕ 2 vs. 3	1	0	ОТЕ 2 vs. 4	1	0	ОТЕ 3 vs. 4	1	0	
1	a=0	b=3	1	a=1	b=2	1	a=1	b=2	
0	c=3	d=3	0	c=2	d=4	0	c=2	d=4	

vs. – в сравнении с, или

Теперь мы можем продолжить с методикой нахождения групп и рисованием соответствующей фенограммы:

	O1	O2	O3	O4
O1	0			
O2	0,56	0		
O3	0,33	0,33	0	
O4	0,33	0,56	0,56	0



Приложение 7:

Измерение генетического разнообразия

Расчет генетического расстояния Нея

Сначала, используя данные, полученные в примере (см. слайд 48) мы сформируем матрицу расстояния следующим образом:

	P ₁	P ₂	P ₃
P ₁	0		
P ₂	0,0852	0	
P ₃	0,0107	0,0440	0

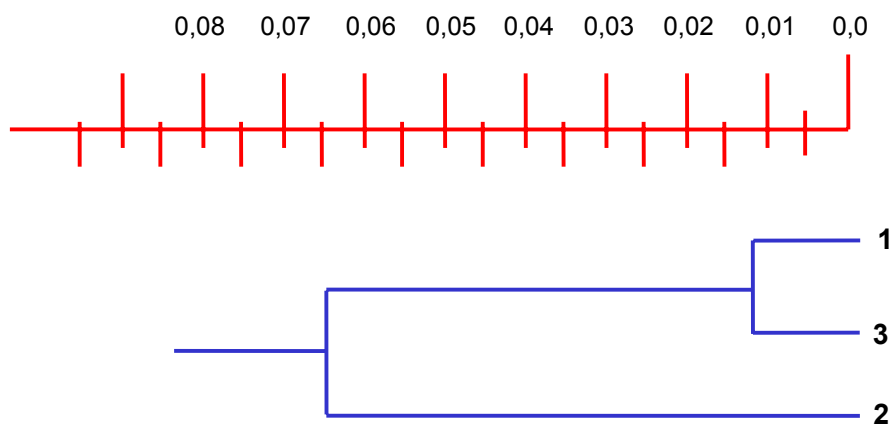
В первом цикле, мы выбрали самое короткое расстояние, $d_{1,3} = 0,0107$

Далее, во втором цикле, формируется новая матрица путем группирования Особи₁ с Особью₄ и расчетом комбинированных расстояний:

$$d_{2(1,3)} = (d_{1,2} + d_{2,3})/2 = (0,0852 + 0,044)/2 = 0,0646$$

	P _{1,3}	P ₂
P _{1,3}	0	
P ₂	0,0646	0

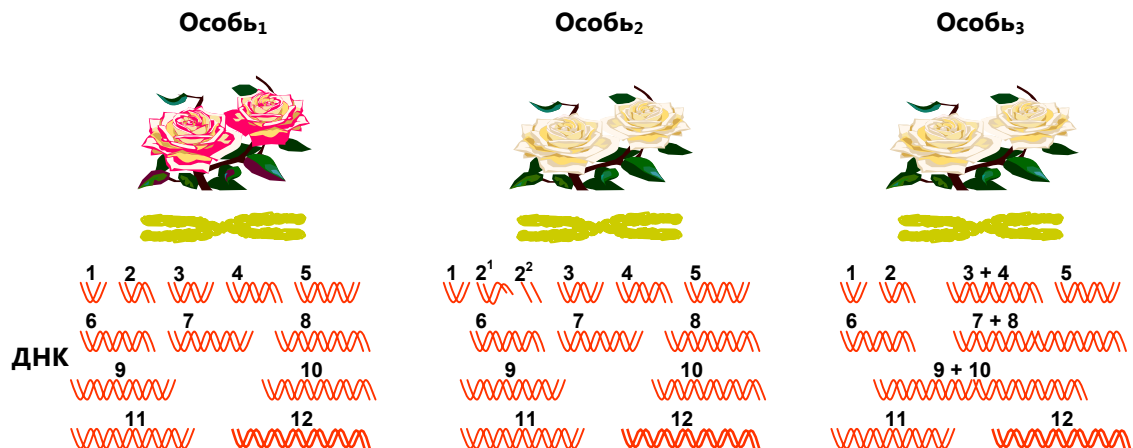
Теперь мы можем нарисовать дендрограмму:



Приложение 8:

Измерение генетического разнообразия

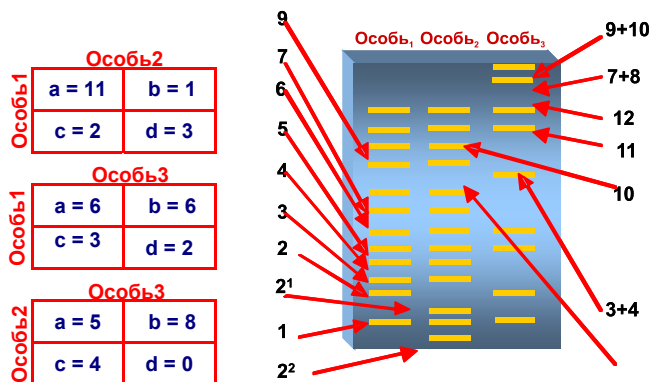
Морфологические и молекулярные сходства



Давайте представим, что мы имеем три особи роз (1, 2, 3). Номера 2 и 3 выглядят морфологически похожими, в то время как номер 1 выглядит иначе.

Если мы взглянем на фрагменты ДНК, предположительно генерированный при помощи молекулярного маркера, то мы увидим, что Особи 1 и 2 выглядят более похожими. Таким образом, что же случилось? Пример показывает важность изучения генетического разнообразия на всех возможных уровнях. Объединение информации от разных типов маркера – а именно, от тех, которые относятся к функциональным генам и тех, которые показывают полиморфизм в геномных регионах – дадут наибольшую адекватность знаниям о присутствующей генетической изменчивости. То же самое мы бы применили, если бы объединяли морфологические и молекулярные данные.

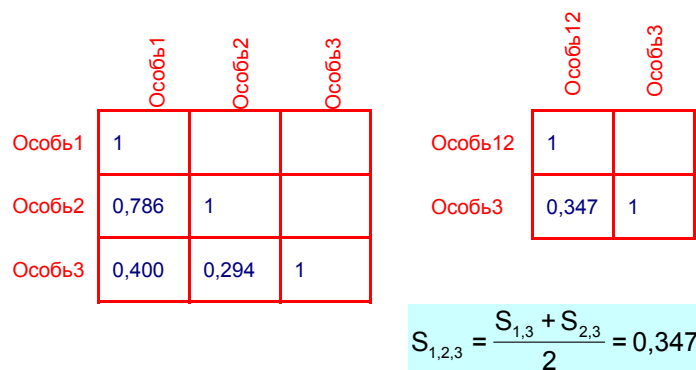
В этом Приложении мы показываем виды ошибок, которые мы можем иметь, если наши выводы будут основаны только на одном типе маркерных данных.



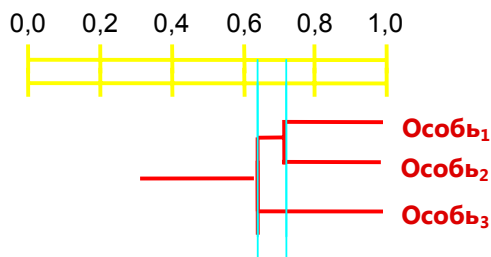
Объединенная дендрограмма показывает расстояния группирования, которые отличаются либо только от морфологической, либо только от молекулярной дендрограммы. Мы, таким образом, можем предположить, что информация, предоставленная объединенными данными, более близка к действительной ситуации. На основе профиля полосы ДНК, полученной в геле для трех особей, рассчитываются попарные расстояния с использованием коэффициента Джаккарда:

$$J = \frac{a}{a + b + c}$$

Далее мы сформируем матрицу расстояний и рисуем дендрограмму:

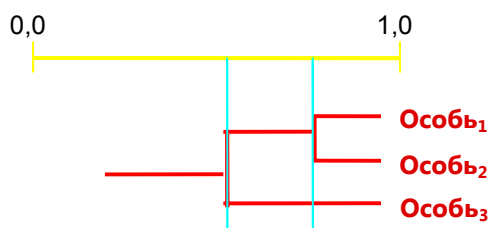


Молекулярная дондрограмма

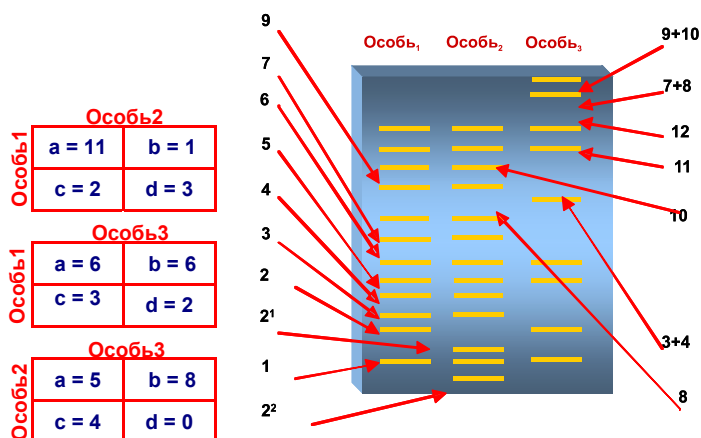


Эта дендрограмма есть результат молекулярных данных. Мы можем теперь сравнить ее с другой дендрограммой, показанной ниже, которая была разработана на основе данных морфологических наблюдений, и отмечает различия. По молекулярной дендрограмме Особи 1 и 2 ближе друг к другу, даже если морфологические данные говорят, что Особи 2 и 3 ближе друг к другу.

Морфологическая дондрограмма



Однако, мы можем использовать как молекулярные, так и морфологические данные, объединяя их и проделывая весь процесс заново с обеими типами данных одновременно.

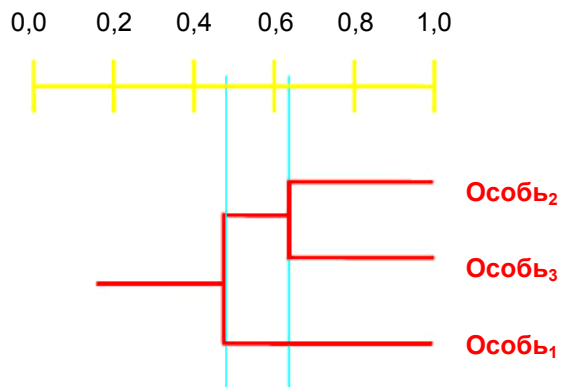


$$J_{1,2} = \frac{11}{11+2+2} = 0,733 \quad J_{1,3} = \frac{6}{6+7+3} = 0,375 \quad J_{2,3} = \frac{6}{6+8+4} = 0,333$$

	Особь1	Особь2	Особь3
Особь1	1		
Особь2	0,733	1	
Особь3	0,3375	0,333	1

	Особь12	Особь3
Особь12	1	
Особь3	0,554	1

$$S_{(1,2)3} = \frac{S_{1,3} + S_{2,3}}{2} = 0,554$$



Объединенная дендрограмма показывает расстояния группирования, которые отличаются либо только от морфологической, либо только от молекулярной дендрограммы. Мы, таким образом, можем предположить, что информация, предоставленная объединенными данными, более близка к действительной ситуации.

**Анализ генетического разнообразия
с использованием данных
молекулярных маркеров:
Учебный модуль**

**Компьютерные программы
для анализа генетического
разнообразия**

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 1

Содержание

▶ Основные характеристики

- Таблица резюме
- Некоторые компьютерные программы, их авторы, веб-сайты ...
- ... и другие свойства

▶ Пять компьютерных программ в деталях

- Arlequin
- PowerMarker
- DnaSP
- PAUP*
- MEGA

▶ Ресурсы Интернета

▶ Приложение 9: Литература по компьютерным программам

Основные характеристики

- ▶ Схожие задачи
- ▶ Основные различия
 - Интерфейс пользователя
 - Тип ввода и вывода данных
 - Платформа
 - Возможности
- ▶ Выбор на основе личных предпочтений

Авторские права: Bioversity International и Корнельский Университет, 2003. Компьютерные программы 3

Существует множество компьютерных программ для оценки генетического разнообразия. Большинство из них можно получить бесплатно через Интернет. Многие программы выполняют схожие функции и их основные различия друг от друга заключаются в пользовательском интерфейсе, типе ввода и вывода данных и платформе. Таким образом выбор программы в значительной мере зависит от личных предпочтений. В этом разделе мы описываем некоторые из существующих программ, учитывая специфические опции, которые пользователи могут найти в них предпочтительными.

Таблица резюме

Особенности	Программа					
	TFPGA ¹	Arlequin ^{1*}	GDA ¹	GENEPOP ^{1*}	GeneStrut	POPGENE ^{1*}
Разнообразие						
Гетерозиготность (наблюдаема)	x	x	x		x	x
Предполагаемая гетерозиготность	x	x	x		x	x
Число аллелей на локус		x	x		x	x
Число эффективных аллелей		x			x	x
% полиморфных локусов	x	x	x			x
S _{hannon-Weaver}						
Структура популяции						
Статистика F	x	x	x	x	x	x
Статистика G					x	x
ANOVA		x	x			
Статистика Rho		x		x		
Гомогенность	x			x		x
Миграция		x		x		x
Дистанционная изоляция				x		
Баланс						
Hardy-Weinberg	x	x	x	x	x	x
Два локуса		x	x	x	x	x
Много локусов			x			
U-тест				x		
Генетическое расстояние						
Ней (Nei)	x	x	x		x	x
Роджера (Rogers)	x				x	
Попарная F _{st}	x	x	x			
Образование кластеров						
Присоединение соседа			x			
UPGMA	x		x		x	x
Тест нейтральности		x				x

¹ Выполняет точные тесты по значимости

* Программа может содержать нулевой аллель в данных

† Пользователь может точно определить коэффициент инбридинга для оценки частоты нулевого аллеля

по Лабате (2000)

Авторские права: Bioersivity International и Корнельский Университет, 2003 Компьютерные программы 4

Джоанна Лабате (2000) написала прекрасный обзор шести компьютерных программ: TFPGA (Miller, 1997), Arlequin (Schneider и др., 1997), GDA (Lewis и Zaykin, 1999), GENEPOP (Raymond и Rousset, 1995), GeneStrut (Constantine и др., 1994), и POPGENE (Yeh и др., 1997). Ее обзор включает специфичные опции, таблицу функций каждой программ и адреса веб-сайтов, откуда они могут быть загружены. Во избежание излишеств, мы включили в наш обзор только Arlequin, который вероятно является наиболее широко используемой среди шести программ.

Для ознакомления с полным перечнем литературных источников по этим шести и другим отобраным компьютерным программам смотрите Приложение 9 ([щелкните мышкой здесь](#)).

Литература

Labate, J.A. 2000. Software for population genetic analyses of molecular marker data. *Crop Sci.* 40:1521-1528.

Некоторые компьютерные программы, их авторы, веб-сайты ...

Название	Авторы	Доступен на:
Alequin^a	Laurent Excoffier	http://lgb.unige.ch/alequin
DnaSP	Julio and Ricardo Rozas	http://www.ub.es/dnasp
PowerMarker	Kejun (Jack) Liu	http://www.powermarker.net/
MEGA2	S. Kumar и другие	http://www.megasoftware.net
PAUP*	David Swofford	http://paup.csit.fsu.edu/
TFPGA ^a	Mark Miller	http://bioweb.usu.edu/mpmbio/index.htm
GDA ^a	Paul Lewis, Dmitri Zaykin	http://lewis.eeb.uconn.edu/lewishome/software.html
GENEPOP ^a	Michel Raymond, Francois Rousset	ftp://ftp.cfece.cnrs-mon.fr/pub/PC/MSDOS/GENEPOP/Genepop.zip also at http://wbiomed.curtin.edu.au/genepop/
NTSYSpc	F.J. Rohlf	http://www.exetersoftware.com/cat/ntsyspc/ntsyspc.html
structure	Jonathan K. Pritchard	http://pritch.bsd.uchicago.edu/
GeneStrut ^t	Constantine, Hobbs & Lybery	http://www.wet.murdoch.edu.au/vetschl/imgad/GenStrut.htm
POPGENE ³	F.C. Yeh, R.-C. Yang, T. Boyle	http://www.ualberta.ca/~fyeh/index.htm
MacClade	David R. & Wayne P. Maddison	http://phylogeny.arizona.edu/macclade
PHYLIP	Joe Felsenstein	http://evolution.genetics.washington.edu/phylip.html
SITES	Jody Hey	http://lifesci.rutgers.edu/~heylab/ProgramsandData/Programs/SITES/SITES_Documentation.htm#Contents
CLUSTAL W	Thompson, Higgins и Gibson	http://www.ebi.ac.uk/clustalw
MALIGN	D. Janies и W.C. Wheeler	http://research.amnh.org/users/djanies/

^a Обсуждается в Лабате (2000)

Авторские права: Bioersity International и Корнельский Университет, 2003 Компьютерные программы 5

Мы проведем обзор других компьютерных программ, выбранных согласно их широкому использованию и отдавая приоритет тем, которые доступны бесплатно (за исключением PAUP*). Перечисленные веб-сайты и ссылки многих других существующих программ также приведены в этом и следующих слайдах. Мы включили информацию об авторах, стоимости, спецификах платформ и веб-сайты. Учтите, что наряду с тем, что некоторые программы разработаны специально только для одной платформы (обычно Windows или Macintosh), с недавним появлением «эмуляторов» (таких как SoftWindows, VirtualPC) большинство компьютерных программ могут работать на любом компьютере независимо от платформы. Хотя мы принимаем к сведению случаи, когда программа успешно использовалась с применением одного из этих эмуляторов, мы не утверждаем, что все перечисленные программы могут работать в любой платформе; просто мы знаем наверняка, что они были успешно использованы. Иногда, использование эмуляторов может быть причиной медленной работы платформы или вызвать другие проблемы. Где возможно, предпочтительно использовать платформы, для которой данная программа разработана.

Программы, перечисленные в слайде и выделенные жирным шрифтом, обсуждаются в настоящем модуле. Веб-сайты приведены по состоянию на 28 февраля 2003 г.

... и другие свойства

Название	Платформа			Стоимость (US \$)
	Windows	Macintosh	Другие	
Arlequin	X	X	X	0
DnaSP	X	с		0
PowerMarker	X			0
MEGA2	X	с	с	0
Arlequin	X	X	X	100
TEPGA	X			0
GDA	X		X	0
GENEPOP			X	0
NTSYSpc	X			230-300
sstructure	X		X	0
GeneStrut		X		0
POPGENE	X	с		0
MacClade		X		125
PHYLIP	X	X	X	0
SITES	X	X		0
CLUSTAL W	X	X	X	0
MALIGN			X	0

Авторские права: Bioversity International и Корнельский Университет, 2003. Компьютерные программы 6

с = программа работает хорошо с такими эмуляторами, как SoftWindows или VirtualPC.

В целях экономии, литературные источники по обсуждаемым программам приведены в Приложении 9 ([щелкните мышкой здесь](#)).

Пять компьютерных программ в деталях

- ▶ Arlequin
- ▶ PowerMarker
- ▶ DnaSP
- ▶ PAUP*
- ▶ MEGA

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 7

Пять компьютерных программ были выбраны для обсуждения в деталях. Этот выбор был сделан на основании результата опроса пользователей о наиболее используемых ими программах, и более полезных или представительных по их мнению программах. Опрошенные пользователи включали в себя студентов, аспирантов и докторантов, соискателей, а также профессорско-преподавательский состав. Перечень включает наиболее упомянутые программы. В случае сомнения, были выбраны программы, доступные бесплатно или те, которые казались более широко используемыми. Чтобы быть более представительным, дополнительный критерий как выполнение разных функций был использован для выбора программ.

Arlequin

- ▶ <<http://lgb.unige.ch/arlequin>>
- ▶ Имеются версии для Windows, Mac, и Linux, все - бесплатно
- ▶ Может использовать разные типы данных (но не доминантных маркеров пока)
- ▶ Пропущенные или неточные данные могут быть включены
- ▶ Данные могут быть введены как последовательности ДНК, гаплотипы RFLP, профили микросателлитов или гаплотипы мультилокуса
- ▶ Данные могут быть импортированы из файлов, созданных для других программ

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 8

(продолжение в следующем слайде)

Выпущенный в 1997 г., Arlequin (современная версия 2.001) является все еще очень популярным. Это «компьютерная программа для исследований популяционной генетики, позволяющая обрабатывать большие выборки молекулярных данных (RFLP, последовательность ДНК, микросателлиты) и в тоже время сохраняющая возможность проведения анализа обычных генетических данных (стандартные мультилокусные данные или абсолютные данные частот аллели).

Arlequin может использовать много различных типов данных таких, как молекулярные данные и частоты генотипов или гаплотипов, включая кодоминантные или рецессивные данные, но пока еще не доминантные данные. Молекулярные данные могут быть введены как последовательности ДНК, гаплотипы RFLP, профили микросателлитов или гаплотипы мультилокуса. Формат данных задается в файле ввода. Пользователь может создать файл данных на пустом месте, используя текстовый редактор и соответствующие ключевые слова, или используя «Project Outline Wizard». Данные могут импортироваться из файлов, созданных для других программ, включая MEGA, BIOSYS, GENEPOP и PHYLIP. Недостающие или неточные данные могут быть также включены. К программе прилагается очень подробное руководство для пользователя, которое включает большой объем теоретической информации, формул и литературных источников. Можно проанализировать большое число данных, и имеется опция пакетных файлов.

Авторы: Laurent Excoffier, Stefan Schneider и David Roessli, Университет Женевы, Швейцария

Литература

Schneider, S., D. Roessli и L. Excoffier. 2000. Arlequin: A Software for Population Genetics Data Analysis, Version 2.000. Genetics and Biometry Laboratory, Dept. of Anthropology, University of Geneva, Switzerland.

Arlequin (продолжение)

► Преимущества:

- Хорошая поддержка обеспечена через подробное руководство, которое включает большой объем теоретической информации, формул и библиографических источников; хорошо организованный веб-сайт с такими функциями как «Часто задаваемые вопросы»
- Графический интерфейс очень удобный для пользователя

► Недостатки:

- Многочисленные функции и опции для изучения
- Создание файла данных может быть сложным и должно быть отформатировано правильно (однако, хороший пример приведен в руководстве)

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 9

PowerMarker

- ▶ <<http://www.powermarker.net>>
- ▶ Разработан для использования данных SSR/SNP в анализах популяционной генетики
- ▶ Имеющиеся опции включают сводную статистику, древа консенсуса, тест Мантеля, построение треугольника и визуализацию результатов дисбаланса связи

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 10

(продолжение в следующем слайде)

Программа PowerMarker – это новая программа, первая официальная версия которой была выпущена в январе 2004 г. Она была разработана специально для использования данных SSR/SNP в анализах популяционной генетики. Данные могут быть импортированы из Excel и других форматов, делая установку данных очень легкой. Данные также могут быть экспортированы в форматы NEXUS и Arlequin. Она включает '2D viewer' для визуализации дисбаланса связей. Пользователь может редактировать графики в PowerMarker или экспортировать их для публикации. Программа была обширно протестирована для точности и эффективности. Прилагается полная документация. В пакет включен ряд новых моделей для изучения ассоциаций. Имеется несколько демонстрационных наборов данных для начала. Программа доступна бесплатно, но требуется обладание PHYLIP, TreeView, системой структуры Microsoft.net (все доступны бесплатно) и Excel 2000 (доступен за плату). Другим недостатком является то, что программа работает только в Windows 98 и его более поздних версиях (не в Macintosh или других системах). Поддерживается через электронную почту (только для зарегистрированных пользователей): powermarker@hotmail.com
Автор: Kejun (Jack) Liu, Университет штата Северной Калифорнии.

Литература

Liu, K. 2003. PowerMarker: New Genetic Data Analysis Software, Version 3.0. Free program distributed by the author over Internet at <<http://www.powermarker.net>>

PowerMarker (продолжение)

► Преимущества:

- Позволяет импортировать данные из Excel, что делает управление данными очень легким
- Графические интерфейсы очень удобны для пользователя
- Гарантировано, что графики имеют хорошее качество для опубликования

► Недостатки:

- Очень новая программа с возможными дефектами, требующими устранения
- Неавтономная программа, требует загрузки нескольких других компьютерных программ и покупки Excel

DnaSP

► <<http://www.ub.es/dnasp>>

► Использует данные последовательности ДНК для выполнения анализа популяционной генетики

► Выполняет очень большое количество анализов, включая измерение полиморфизма, дивергенции между популяциями, включая измерение потока генов, синонимичные и несинонимичные замещения, дисбаланс связи, рекомбинацию, и многие статистические тесты (тесты Hudson, Kreitman и Aguade, Fu и Li, Tajima, McDonald и Kreitman)

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 12

(продолжение в следующем слайде)

DnaSP, для полиморфизма последовательности ДНК, использует данные последовательности ДНК. Эта программа широко используется для анализа последовательности потому, что она делает все необходимые анализы и в то же время легка в использовании. Она была разработана исключительно для операционной системы Windows, но может работать в Macintosh с использованием эмуляторов SoftWindows или VirtualPC. DnaSP может импортировать и экспортировать несколько типов формата данных, включая FASTA и NEXUS, что очень удобно, и может обрабатывать большое число длинных последовательностей в зависимости от объема памяти компьютера. В настоящее время авторы работают над версией 4. Программа доступна бесплатно и может быть загружена с веб-сайта. Хотя руководства по пользованию программой не имеется, файл Помощи встроен в программу. В дополнение, веб-сайт включает много разъяснительного материала, а также литературных источников. Авторы опубликовали ряд материалов о программе (см. ссылки ниже).

Авторы: Julio и Ricardo Rozas

Литература

Rozas, J. и R. Rozas. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Appl. Biosci.* 11:621-625.

Rozas, J. и R. Rozas. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* 13:307-311.

Rozas, J. и R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174-175.

DnaSP (продолжение)

► Преимущества :

- Пользовательский интерфейс Windows делает программу очень легкой в использовании
- Возможность сделать очень большое число анализов

► Недостатки:

- В настоящее время доступен только для пользователей Windows (и Macintosh с использованием Windows -эмулятора)
- Не имеется пока руководства или учебного пособия
- Очевидно, что поддержка является дефицитной

PAUP*

- ▶ <<http://www.lms.si.edu/PAUP/about.html>>
- ▶ Используется для выводов и интерпретации древ эволюции
- ▶ Имеются версии для различных компьютерных платформ
- ▶ Включает: экономичность, матрицу расстояния, инварианты, методы максимального правдоподобия, много индексов и статистических анализов

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 14

(продолжение в следующем слайде)

PAUP* широко используется для того, чтобы делать заключения и интерпретировать эволюционные древа. Изначально название программы означало «Филогенетический анализ, используя экономию», но теперь она имеет много других опций. PAUP* можно получить от Sinauer Associates, Сандерленд, Массачусетс на сайте: <<http://www.sinauer.com/detail.php?id=8060>>. Хотя программа не бесплатная, она сравнительно недорогая (US\$100 на момент написания настоящего модуля). Новая версия программы, 4.0 бета, была выпущена как предварительная версия. Имеются версии для Macintosh, PowerMac, Windows и Unix/OpenVMS; версия Mac имеет некоторые дополнительные свойства. PAUP* тесно совместим с MacClade (другая программа доступная с Sinauer), так как они используют общий формат данных (NEXUS, Maddison и др. 1997).

Автор: David Swofford, Лаборатория молекулярной систематики, Национальный музей естествознания, Смитсоновский институт, Вашингтон.

Литература

Swofford, D.L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

РАУР* (продолжение)

► **Преимущества:**

Имеется хорошая поддержка как на веб-сайте (включая длинный перечень «Часто задаваемых вопросов»), так и по электронной почте

► **Недостаток:**

Не бесплатно

MEGA

- ▶ <<http://www.megasoftware.net/>>
- ▶ Использует последовательность ДНК, последовательность протеина, эволюционное расстояние или данные филогенетического дерева.
- ▶ Возможные методы анализа многочисленны, включая расчеты расстояний, формирование деревьев и много способов рассмотрения данных.

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 16

(продолжение в следующем слайде)

Программа MEGA (Molecular Evolutionary Genetics Analysis (Анализ Молекулярной Эволюционной Генетики)) широко используется с момента ее создания в 1993 г.; к настоящему моменту появилась MEGA2. Она использует последовательность ДНК, последовательность протеина, эволюционное расстояние или данные филогенетического дерева. Цель ее разработчиков заключалась в использовании преимуществ мощности компьютера и графических пользовательских интерфейсов для создания «гибкого и легкого в использовании верстака для анализа генетических данных». Хотя программа была разработана для платформы Windows, она хорошо работает и в Macintosh с эмулятором Windows, автоматизированное рабочее место Sun (с SoftWindows95) или Linux (с Windows от VMWare). Новейшая версия, 2.1, имеет много важных дополнений таких, как способность импортировать данные из NEXUS или CLUSTAL W, неограниченные размеры набора данных, и многие другие.

Книга разработчиков программы, Нея (Nei) и Кумара (Kumar) (2000) включает теоретическую информацию о статистических анализах, советы как интерпретировать результаты, полученные при помощи как их собственной, так и других программ. Подробное руководство доступно в Интернете (хотя формат не очень удобен для перелистывания страниц) вместе с доской объявлений, чтобы взаимодействовать друг с другом.

Авторы: Sudhir Kumar, Koichiro Tamura, Ingrid Jakobsen и Masatoshi Nei.

Литература

Nei, M. и S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, NY.

Sudhir, K., T. Koichiro, I.B. Jakobsen и M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 12(17):1244-1245.

MEGA (продолжение)

► Преимущества:

- Так как данная компьютерная программа существует с 1993 г., вероятно, что большинство ошибок в программе уже выявлено
- Имеется хорошая поддержка включая руководство в электронном формате в Интернете и изданную в виде книги

► Недостаток:

Это не самая легкая программа для начинающих

Ресурсы Интернета

- ▶ Сайт Европейской Лаборатории молекулярной биологии (The European Molecular Biology Laboratory)– Европейского Института Биоинформатики (European Bioinformatics Institute (EBI)) <<http://www.ebi.ac.uk/services/>>
- ▶ Страничка биологических компьютерных программ Института Пастера во Франции (Institut Pasteur) <<http://bioweb.pasteur.fr/intro-uk.html>>
- ▶ Программы филогенетики (перечисленные Джоом Фельзенштейном (Joe Felsenstein) из Университета Вашингтона <<http://evolution.genetics.washington.edu/phylip/software.html>>

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 18

(продолжение в следующем слайде)

В этом и следующем слайдах мы приводим примерный перечень ресурсов Интернета, который вы возможно найдете полезным для помещения информации, относящейся, например, к анализу генетического разнообразия, популяционной генетике и другим существующим компьютерным программам, а также линкам к полезной дополнительной информации. Для каждого представленного ресурса мы кратко описали его содержание в заметках ниже.

- Сайт Европейской Лаборатории молекулярной биологии (The European Molecular Biology Laboratory)– Европейского Института Биоинформатики (European Bioinformatics Institute (EBI)) : не только содержит линки ко многим полезным компьютерным программам и другим сайтам, но также описывает их функции, и является хорошим источником общей информации.
- Страничка биологических компьютерных программ Института Пастера в Франции (Institut Pasteur) : хотя некоторые страницы имеют информацию только на французском языке, она предоставляет очень обширный перечень компьютерных программ, доступных через Интернет, включая линки, и постоянно обновляется. Она также содержит линки к многим программам, разработанным Институтом Пастера.
- Программы филогенетики – это длинный перечень филогенетических программ, который мы когда-либо видели, насчитывающий 194 наименования. Автор добавил предупреждение, что он не пытался оценить качество или стоимость программ. Перечень не обновлялся с 2001 г., но он содержит линки к программам и сортирует их различными способами (например, по использованным методам, системам), что по-прежнему очень полезно.

Ресурсы Интернета (продолжение)

- ▶ Сайт д-ра Ed Buckler по генетике кукурузы
<<http://www.maizegenetics.net/bioinformatics/index.htm>>
- ▶ Сайт Kent Holsinger из университета Коннектикут
<<http://darwin.eeb.uconn.edu/archives.html>>
- ▶ Сайт Claire Constantine из университета Мардок (Murdoch University)
<<http://wwwvet.murdoch.edu.au/vetschl/imgad/GSLinks.htm>>
- ▶ Страничка компьютерных программ Института лесной генетики и селекции лесных деревьев (Institute of Forest Genetics and Forest Tree Breeding), Университета Геттингена (University of Göttingen), Германия
<<http://www.uni-forst.gwdg.de/forst/fg/index.htm>>

Авторские права: Bioversity International и Корнельский Университет, 2003 Компьютерные программы 19

- Сайт д-ра Эда Баклера (Ed Buckler) по генетике кукурузы содержит бесплатно доступную компьютерную программу, разработанную группой сотрудников его лаборатории. Хотя некоторая информация является характерной для кукурузы, сайт также содержит полезную информацию о геномике, ссылки к многим журналам и публикациям д-ра Баклера (Buckler) в формате PDF.
- Сайт Кента Холзингера (Kent Holsinger), Университет Коннектикут, имеет ссылки ко многим компьютерным программам, включая программы для биологии, программирования и статистики. Автор не поместил того, что он не использует регулярно и, поэтому имеется некоторая гарантия хорошего качества. На момент написания настоящего модуля сайт был недавно обновлен.
- Сайт Клер Константине (Claire Constantine), Университета Мардок (Murdoch University) хотя и не обновлялся в последнее время, но содержит ссылки к наиболее используемым программам для анализа популяционной генетики. Сайт даже очень полезен, так как он содержит сравнительную таблицу типов статистических данных, имеющихся в каждой из семи обычно используемых программ (Arlequin, GENEPOP, POPGENE, GDA, GeneStrut, DnaSP и SITES).
- Страничка компьютерных программ Института лесной генетики и селекции лесных деревьев (Institute of Forest Genetics and Forest Tree Breeding), Университета Геттингена (University of Göttingen), Германия. Этот сайт содержит 4 компьютерные программы, разработанные Институтом. Они бесплатно доступны, включают в себя хорошее описание, и страничка регулярно обновляется.

Приложение 9

Авторские права: Bioversity International и Корнельский Университет, 2003. Компьютерные программы 20

Для просмотра приложения щелкните мышкой на его названии :

Приложение 9. [Литература по компьютерным программам](#)

Вкратце

- ▶ Существует много компьютерных программ для анализа молекулярных данных для генетического разнообразия
- ▶ Большинство программ выполняют схожие задачи и их основные различия должны быть оценены в зависимости от имеющихся ресурсов и/или личных предпочтений
- ▶ В настоящее время в дополнение к имеющимся бесплатным компьютерным программам, множество ресурсов также могут быть найдены в Интернете, чтобы помочь нам получить как основную, так и специальную информацию о методах

Теперь Вы должны ...

- ▶ Быть знакомы с :
 - Отличительными особенностями имеющихся компьютерных программ для анализа генетического разнообразия
 - Преимуществами и недостатками некоторых основных компьютерных программ
- ▶ Иметь представление о некоторых ресурсах в Интернете, которые могут помочь Вам в изучении генетического разнообразия

Литература

Авторские права: Bioversity International и Корнельский Университет, 2003. Компьютерные программы 23

Литература

- Labate, J.A. 2000. Software for population genetic analyses of molecular marker data. *Crop Sci.* 40:1521-1528.
- Liu, K. 2003. PowerMarker: New Genetic Data Analysis Software, Version 1.0. Free program distributed by the author over Internet at <<http://www.powermarker.net>>
- Maddison, D. R., D.L. Swofford и W.P. Maddison. 1997. NEXUS: an extensible file format for systematic information. *Syst. Biol.* 46:590–621.
- Nei, M. и S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, NY.
- Rozas, J. и R. Rozas. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Appl. Biosci.* 11:621-625.
- Rozas, J. и R. Rozas. 1997. DnaSP, version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* 13:307-311.
- Rozas, J. и R. Rozas. 1999. DnaSP, version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174-175.
- Schneider, S., D. Roessli и L. Excoffier. 2000. Arlequin: A Software for Population Genetics Data Analysis, Version 2.000. Genetics and Biometry Laboratory, Dept. of Anthropology, University of Geneva, Switzerland.
- Sudhir, K., T. Koichiro, I.B. Jakobsen и M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 12(17):1244-1245.
- Swofford, D.L. 2002. PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

Следующее

► Глоссарий

Приложение 9:

Компьютерные программы для анализа генетического разнообразия

Литература по компьютерным программам

Arlequin

Schneider, S., D. Roessli и L. Excoffier. 2000. Arlequin: A Software for Population Genetics Data Analysis, Версия 2.000. Genetics and Biometry Laboratory, Dept. of Anthropology, University of Geneva, Switzerland.

CLUSTAL W

Thompson, J.D., D.G. Higgins и T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

DnaSP

Rozas, J. и R. Rozas. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci. 11:621-625.

GDA

Lewis, P.O. и D. Zaykin. 1999. Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, Версия 1.0 (d12). Distributed by the authors.

GENEPOP

Raymond, M. и F. Rousset. 1995. GENEPOP (версия 1.2): Population genetics software for exact tests and ecumenicism. J. Hered. 86:248-249.

GeneStrut

Constantine, C.C., R.P. Hobbs и A.J. Lymbery. 1994. FORTRAN programs for analysing population structure from multilocus genotype data. J. Hered. 85:336-337.

MacClade

Maddison, D.R. и W.P. Maddison. 2000. MacClade. Версия 4. Sinauer Associates, Sunderland, MA.

MALIGN

Janies, D. и W.C. Wheeler. 1998. MALIGN.pdf: Documentation for MALIGN, software for multiple alignments of DNA sequences. Distributed by the authors over Internet at <ftp://ftp.amnh.org/pub/molecular/malign/>.

MEGA2

Kumar, S., K. Tamura, I.B. Jakobsen и M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17(12): 1244-1245.

NTSYSpc

Rohlf, F.J. 2002. NTSYS pc: Numerical Taxonomy System, Version 2.1. Exeter Publishing, Setauket, NY.

PAUP*

Swofford, D.L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.

PHYLIP

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package), Version 3.5c. Distributed by the author.

POPGENE

Yeh, F.C., R.C. Yang, T.B.J. Boyle, Z.H. Ye и J.X. Mao. 1997. POPGENE, the User-Friendly Shareware for Population Genetic Analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

PowerMarker

Liu, J. 2003. PowerMarker: New Genetic Data Analysis Software, Version 3.0. Free program distributed by author over Internet at <http://www.powermarker.net>

SITES

Hey, J и J. Wakeley. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145:833-846.

Structure

Pritchard, J.K., M. Stephens и P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

TFPGA

Miller, M.P. 1997. Tools for Population Genetic Analysis (TFPGA), 1.3: A Windows Program for the Analysis of Allozyme and Molecular Population Genetic Data. Distributed by the author.

Анализ генетического разнообразия с использованием данных молекулярного маркера

Глоссарий

AFLP (Amplified fragment length polymorphism) - полиморфизм длины амплифицированного фрагмента. Высококчувствительный метод обнаружения полиморфизмов в ДНК. Сначала происходит расщепление ДНК ферментами рестрикции, затем выбирается набор фрагментов ДНК для ПЦР амплификации и визуализации.

АМОВА (The analysis of molecular variance) - анализ молекулярной вариации (АМВ). Это метод изучения молекулярных изменений в пределах вида.

Аллель – одна из альтернативных форм гена, которая может существовать в единичном локусе.

Аллогамия – перенос пыльцы (опыление) с пыльника цветка одного растения на рыльце цветка генетически разного растения. Называется также перекрестным опылением, ауткроссингами или ксеногамией. *См. также* Аутбридинг.

Аутбридинг – аллогамная система скрещивания, когда скрещиваются особи менее близко родственные, чем средние пары, произвольно выбранные в популяции. Называется также экзогамией и перекрестным скрещиванием (*противоположно Инбридингу*).

Ауткроссинг - см. Аллогамия.

Аутогамия:

1) перенос пыльцы (опыление) с пыльника цветка на рыльце того же цветка, или, иногда, на рыльце генетически идентичного цветка (например, того же растения или клона);

2) способность многих видов растений к естественному и успешному оплодотворению в пределах одной и той же особи. Называется также самоопылением.

Биоразнообразие – совокупность генов, видов и экосистем данного региона является ли он микроареалом или всем земным шаром. Называется также биологическим разнообразием.

Гаплоид – одинарный набор хромосом (половина полного набора генетического материала), представленный в каждой яйце- или сперматозоидной клетке животного и в каждой яйцеклетке или пыльце растения (с греч. ‘haploos’ – одинарный) (*противоположно Диплоиду*).

Гаплотип - специфичный набор аллелей в ряде локусов в определенном блоке сцепления.

Ген - основная физическая и функциональная единица наследования, передающая информацию от одного поколения другому. Ген является сегментом ДНК, включающий транскрибируемый (кодирующий) участок и регуляторный элемент, позволяющий осуществлять его транскрипцию.

Генетическая дистанция – степень связанности между подгруппами и популяциями, измеренная разными статистическими методами.

Генетическая эрозия – потеря генетического разнообразия между и внутри популяций одного и того же вида во времени, или сокращение генетической базы вида.

Генетический дрейф – непредсказуемые изменения в частоте аллели, которые возникают в маленьких популяциях.

Генетический маркер - аллель, полоса в геле или признак, экспериментально использующийся как проба для идентификации особи или одного из ее признаков.

Генетическое разнообразие – изменение в генетическом составе особей в пределах и между видами; наследуемое генетическое изменение внутри и между популяциями.

Геном - весь набор генетического материала в организме.

Гетерозигота - диплоидная особь, содержащая различные аллели в одном или более генетических локусах. (с греч. “heteros” – другой). (*противоположно Гомозиготе*).

Гомозигота – диплоидная особь, содержащая одинаковые аллели в одном или более генетических локусах (с греч. “homos” – одинаковый). (*противоположно Гетерозиготе*).

Дикий сородич – родственник культурного растения, который произрастает в природе и не используется для сельскохозяйственных целей.

Диплоид – полный набор генетического материала, состоящий из спаренных хромосом, взятых от каждого родителя. Большинство животных клеток, за исключением гамет, имеет диплоидный набор хромосом (*противоположно Гаплоиду*).

Изофермент - множественные формы фермента, синтез которого контролируется двумя или более генами.

Инбридинг – скрещивание генетически родственных особей или между родственниками. Скрещивание посредством смены родителей, принадлежащих к одному и тому же роду. Называется также эндогамией или размножением путем самоопыления (*противоположно Аутбридингу*).

Коллекция (генетических ресурсов растений):

1) сбор в одном месте одомашненных растений (примитивных, стародавних и современных сортов, а также селекционных линий) и родственных диких и сорных видов.

2) материал, собранный в результате сбора, называется термином «коллекция»

Локус - специфичный участок на хромосоме, где расположен ген или определенный отрезок ДНК.

Маркер - поддающееся определению физическое месторасположение на хромосоме, наследование которого можно проследить (например, участок гена, сайт рестрикции или RFLP маркер).

Миграция - движение особей между изолированными популяциями, размножающимися иным способом.

Микросателлитная ДНК - тип повторяющейся ДНК, состоящей из очень коротких повторов, как динуклеотиды, тринуклеотиды или тетрануклеотиды. Также называются простые повторяющиеся последовательности (SSRs).

Множественные аллели – существование нескольких известных аллелей гена.

Мутация – термин, описывающий наследуемое скачкообразное изменение фенотипа. Любое постоянное или наследуемое изменение в последовательности ДНК. Виды мутации включают в себя точковую мутацию, делецию, инсерцию и изменения в числе и структуре хромосом.

Неполовое размножение – формирование новой особи из клетки одного родителя. Оно не влечет за собой рекомбинацию или смешения родительских форм.

Одомашнивание – эволюция растений или животных либо естественно, либо путем искусственного отбора, в формы более полезные для человека, например, неосыпающиеся семена.

Пиримидин - азот-содержащее, одно-спиральное основное соединение, встречающееся в нуклеиновой кислоте. Прирамидины в ДНК – это цитозин и тимин. Пиримидины в РНК – это цитозин и урацил.

Полиморфизм - проявление различных форм, ассоциированных с различными аллелями одного гена или гомологичных одной хромосоме.

Половое размножение – производство новых особей в результате смешения в одной клетке двух различных клеток, обычно гамет и обычно от разных родителей.

Популяционная генетика – количественное изучение и измерение популяций в статистических выражениях, например, изучение генетического феномена по стандартным статистическим параметрам таких, как таблицы и распределение частот, средние значения, вариантность, стандартные отклонения.

Популяция – в генетике, это группа особей, которые разделяют общий генофонд и обладают способностью скрещиваться.

Поток генов - обмен генетическим материалом между популяциями. Это может происходить при размножении растений (например, из-за рассеивания гамет или зигот) или благодаря человеческой деятельности, как, например, введение новых сортов сельскохозяйственных культур фермерами.

Пурин – азот-содержащее, двух-спиральное основное соединение, встречающееся в нуклеиновой кислоте. Пурин в ДНК и РНК – это аденин и гуанин.

Равновесие Харди-Вейнберга (Hardy-Weinberg equilibrium) – стабильное распространение частот генотипов AA, Aa, и aa, в пропорциональном отношении p^2 , $2pq$, и q^2 , соответственно (где p и q есть частоты аллель A и a), что является следствием случайного скрещивания при отсутствии мутации, миграции, естественного отбора и случайного дрейфа.

Разнообразие видов – функция распределения и распространения видов; сходно со значением «богатство видов». В более специальной литературе включает понятие равномерного распространения видов. Об экосистеме можно сказать, что она разнообразнее в соответствии с более специальным определением, если виды, представленные в ней, имеют одинаковые размеры популяций, и экосистема считается менее разнообразной, если многие виды являются редкими, а некоторые очень распространенными.

Регенерация (коллекций генетических ресурсов):

- 1) процесс восстановления всего растения из клеток особи путем манипулирования *in vitro* культуры;
- 2) выращивание семян из образца для восстановления жизнеспособности исходного образца. Оно обычно проводится, если жизнеспособность исходного материала ниже 85%.

Рекомбинация - образование молекулы ДНК с сегментами, полученными более чем от одной родительской ДНК-молекулы. В эукариотах это достигается посредством реципрокного обмена ДНК между неродственными хроматидами гомологичной пары хромосом в течении профазы первого митотического деления. Рекомбинация позволяет хромосомам изменять свой генетический материал, таким образом, повышая потенциал генетического разнообразия. Также известна как кроссинговер.

Самооплодотворение – см. Автогамия.

Самоопыление – опыление пыльцой того же самого растения.

Селекция – размножение и генетическая манипуляция путем гибридизации или намеренного самоопыления растений в целях отбора улучшенного потомства.

Симпатрический – встречающийся в той же географической области.

Система размножения – система, посредством которой виды воспроизводятся. Для растений существует несколько естественных систем размножения, для примера *см. также* Аутбридинг и Инбридинг.

Система спаривания – характер спаривания между особями популяции, включая такие факторы, как уровень инбридинга, моногамии, ряда синхронных спариваний. Система спаривания имеет существенное значение в определении как генетической структуры, так и эволюционного потенциала природных популяций.

Сорное растение:

- 1) в сельском хозяйстве, отдельное растение или вид, растущий там, где он нежелателен;
- 2) в экологии, растение, которое адаптировано для произрастания в нарушенных или открытых ареалах, например, на местах пожарищ или на территориях, нарушенных в результате деятельности человека.

Сохранение – это управление антропогенным использованием биосферы таким образом, чтобы она могла обеспечивать максимальную устойчивую полезность нынешним поколениям наряду с сохранением своего потенциала удовлетворять потребности и желания будущих поколений человечества. Следовательно, сохранение охватывает охрану, поддержание, устойчивое использование, восстановление и усиление функций естественной окружающей среды.

Сохранение *ex-situ*:

- 1) метод сохранения, связанный с удалением ресурсов гермоплазмы (семян, пыльцы, сперматозоидов, отдельных организмов) из их первоначального местообитания или природной среды.
- 2) хранение компонентов биоразнообразия в живом виде за пределами их первоначального местообитания или природной среды (противоположно *in situ* сохранению).

Сохранение *in-situ* - метод сохранения, предназначенный для обеспечения генетической целостности генетических ресурсов путем их сохранения в эволюционно динамичных экосистемах их первоначального местообитания или природной среды (*противоположно сохранению ex-situ*).

Фермент рестрикции - эндонуклеаза, распознающая конкретную последовательность и разрезающая цепь ДНК в этой точке.

Филогенетика – история эволюции видов. Диаграмма, иллюстрирующая историю эволюции популяций и родственных видов.

Характеристика – оценка признаков растения, которые высоко наследуемы, легко распознаваемы невооруженным глазом, одинаково выражены в любой окружающей среде и применимы для различения фенотипов.

RAPD (Random amplified polymorphic DNA) - случайно амплифицированная полиморфная ДНК. Метод амплификации анонимных отрезков ДНК при помощи ПЦР с произвольными праймерами.

RFLP (Restriction Fragment Length Polymorphism) - полиморфизм длины рестрикционного фрагмента (ПДФ). Различия между индивидуумами определяются как разница в размерах фрагментов ДНК после рестрикции.

SSR (Simple-sequence repeats) - см. Микросателлитная ДНК.

Форма обратной связи

Характер тематики настоящего модуля таков, что постоянно меняется, и поэтому желательно периодически обновлять его содержание. Вот почему мы предлагаем этот модуль в электронном формате. Мы были бы благодарны за замечания и просьбы пользователей, чтобы помочь нам в выполнении этой задачи и дальнейшем совершенствовании модуля. Ваши замечания относительно полезности данного инструмента также очень помогут нам.

Ниже мы предлагаем Вам ответить на несколько вопросов. Вы можете отправить нам свои ответы, посредством нажатия окна “Отправить” в конце этой страницы, или по факсу (№ 57-2 445 0096), или по почте Вашему контактному лицу в Bioversity International, или по электронной почте – на адреса:

cdevicente@cgiar.org, cflb@lamolina.edu.pe или tf@cornell.edu

Вопросы:

1. Расскажите, как Вы используете данный модуль.....

2. Считаете ли Вы, что данный модуль:

a. Представляет обзор основных научных концепций популяционной генетики для оценки генетического разнообразия?

Да
Нет
Почему?

b. Показывает, как технологии молекулярных маркеров могут быть практически использованы любым, кто планирует исследование по анализу генетического разнообразия

Да
Нет
Почему?

c. Дает информацию о том, как проводить анализ и толковать данные молекулярных маркеров

d. Дает список современных библиографических источников по каждой из обсуждаемых тем?

Да
Нет
Почему?

3. Расскажите, что Вам понравилось, и что Вы хотели бы изменить в модуле:

- a. Содержание
- b. Структуру разделов
- c. Рисунки, диаграммы
- d. Приложения
- e. Другое

4. Что Вы добавили бы в модуль, чтобы сделать его более полезным в Вашей конкретной ситуации?

- a. Список основных библиографических источников
- b. Примеры применения
- c. Дополнительную информацию о:
 - I. Концепциях популяционной генетики
 - II. Измерениях генетического разнообразия
 - III. Описанных компьютерных программах
- d. Другое

5. Какие еще темы Вы хотели бы увидеть в данном модуле?

6. Было ли трудно загружать модуль или его части? Почему?

Отправить

Bioversity International является независимой международной научной организацией, которая стремится улучшить благополучие нынешнего и будущих поколений человечества посредством оказания содействия сохранению и широкому использованию сельскохозяйственного биоразнообразия на фермерских полях и в лесах. Bioversity является одним из 15 центров, поддерживаемых Консультативной группой международных сельскохозяйственных исследований (CGIAR), ассоциации общественных и частных членов, которые поддерживают усилия по мобилизации достижений современной науки для устранения голода и бедности, улучшения питания и состояния здоровья населения, а также защиты окружающей среды. Штаб-квартира Bioversity находится в селе Маккарисе, недалеко от Рима, в Италии, а также его офисы имеются еще в 20 странах мира. Институт работает по четырем программам: Разнообразие для жизнеобеспечения; Понимание и управление биоразнообразием; Глобальное партнерство; и Товары для жизнеобеспечения.

Bioversity присвоен международный статус в соответствии с Учредительным Договором, который к январю 2006 года подписан и ратифицирован правительствами Австралии, Алжира, Бельгии, Бенина, Боливии, Бразилии, Буркина Фасо, Венгрии, Гвинеи, Греции, Дании, Египта, Израиля, Индии, Индонезии, Ирана, Италии, Иордании, Камеруна, Кении, Кипра, Китая, Конго, Коста Рики, Кот-Дивуара, Мавритании, Марокко, Малайзии, Мали, Норвегии, Пакистана, Панамы, Перу, Польши, Португалии, России, Румынии, Сенегала, Сирии, Словакии, Судана, Туниса, Турции, Уганды, Украины, Чешской Республики, Чили, Швейцарии и Эквадора.

Финансовая поддержка программе исследований Bioversity предоставляется более, чем 150 донорами, включая правительства стран, частные фонды и международные организации. Для подробной информации о донорах и исследовательской деятельности института, пожалуйста, смотрите годовые отчеты Bioversity, которые доступны по запросу на адрес bioversity-publications@cgiar.org или на веб-сайте Bioversity (www.bioversityinternational.org).

Географические названия, использованные в данной публикации, не отражают мнения Bioversity или CGIAR относительно юридического статуса какой-либо страны, территории, города, района и их органов власти, или относительно делимитации их границ. Подобным же образом, взгляды авторов не обязательно совпадают с мнением этих организаций по данному вопросу.

Использование запатентованного названия не означает полного согласия с данной публикацией и приводится только для информации.

Авторские права ©2004 Bioversity International и Института геномного разнообразия (IGD) Корнельского Университета.

Авторскими и всеми другими правами относительно настоящей работы владеют Bioversity International и Корнельский университет, за исключением случаев их иного конкретного использования. Частные лица могут свободно снимать копию и распечатывать материалы для образовательных и некоммерческих целей без получения предварительного разрешения со стороны владельцев авторских прав. Ссылка на источник материалов обязательна.

Ссылка

Анализ генетического разнообразия с данными молекулярных маркеров: обучающий модуль. Под ред. К.С. де Висенте и Т. Фултон. Bioversity International. 2004. Рим, Италия.

М. Кармен де Висенте, PhD, является специалистом в области молекулярной генетики растений и работает в региональном офисе Bioversity International для стран Латинской Америки, г. Кали, Колумбия. Цезарь Лопес, MSc – преподаватель генетики в Universidad Nacional Agraria 'La Molina' в г. Лима, Перу, и Тереза Фултон, PhD, директор отдела внедрения разработок Института геномного разнообразия (IGD) Корнельского Университета в г. Итака, штат Нью-Йорк, США.

Русская версия модуля подготовлена под редакцией А.А. Абдуллаева, кандидата биологических наук, старшего научного сотрудника Института Генетики и Экспериментальной Биологии Растений Академии Наук Республики Узбекистан, и М.К. Турдиевой, сотрудница офиса Bioversity International для стран Центральной Азии в г. Ташкент, Узбекистан.