

ОБЗОРНЫЕ
И ТЕОРЕТИЧЕСКИЕ СТАТЬИ

УДК 575.174:577.113.5

ПРЕДУБЕЖДЕНИЯ О МИКРОСАТЕЛЛИТНЫХ
ИССЛЕДОВАНИЯХ И КАК ИМ ПРОТИВОСТОЯТЬ

© 2019 г. Т. В. Галинская^{1, 2, *}, Д. М. Щепетов^{1, 3, 4}, С. Н. Лысенков⁵

¹Московский государственный университет им. М.В. Ломоносова, кафедра энтомологии, Москва, 119234 Россия

²Всероссийский центр карантина растений, Московская область, Раменский р-н, пос. Быково, 140150 Россия

³Институт биологии развития им. Н.К. Кольцова Российской академии наук, Москва, 119334 Россия

⁴Центр фундаментальных исследований Национальный исследовательский университет,
Высшая школа экономики, Москва, 101000 Россия

⁵Московский государственный университет им. М.В. Ломоносова,
кафедра биологической эволюции, Москва, 119234 Россия

*e-mail: nuha1313@gmail.com

Поступила в редакцию 11.09.2018 г.

После доработки 26.12.2018 г.

Принята к публикации 11.01.2019 г.

В данной статье обобщены основные понятия и методики, применяющиеся в микросателлитных исследованиях. Тезисно изложены основные стадии работы, даны ссылки на наиболее часто используемые программы для обработки микросателлитных данных. В статье также кратко разбирается биологический смысл индексов, наиболее часто встречающихся в микросателлитных исследованиях.

Ключевые слова: популяции, эволюция, микросателлиты, карантинные виды, популяционная генетика, программное обеспечение.

DOI: 10.1134/S0016675819060043

Высокополиморфные микросателлитные локусы являются удобным инструментом для исследования микроэволюционных процессов [1], генетической структуры популяций и взаимодействий между популяциями [2, 3] из-за большого разнообразия аллелей и высокого уровня гетерозиготности.

С помощью исследования микросателлитных локусов можно решать следующие задачи: выявить наличие или отсутствие популяционной структуры (пространственной, временной и др.) внутри интересующего вас вида животных или растений; выявить пути расселения видов-вредителей или карантинных видов; выяснить, насколько эти популяции связаны между собой потоком генов; оценить эффективную численность популяции и внутривидовую генетическую изменчивость; узнать, проходила ли популяция недавно через “бутылочное горлышко”; выяснить, коррелируют ли популяционные структуры вида-симбионта и вида-хозяина.

Микросателлит — участок ДНК, в котором последовательно повторяется фрагмент длиной от 2 до 9 нуклеотидов. Общая длина такого повтора при этом составляет обычно менее 300–400 нуклеотидов. Их также называют короткими тандем-

ными повторами (Short Tandem Repeat, STR) или простыми повторяющимися последовательностями (Simple Sequence Repeat, SSR). Как правило, число повторов в одном микросателлите составляет от 5 до 40, но их может быть и больше. Разница в числе повторов порождает аллели разной длины. Микросателлиты могут находиться как в кодирующих, так и в некодирующих регионах, в том числе в регуляторных последовательностях [4].

Частота мутаций в микросателлитных локусах в среднем в 10–100000 раз больше, чем в других областях генома. Мутации в микросателлитах, как правило, затрагивают не единичные нуклеотиды, а уменьшают или увеличивают количество повторяемых фрагментов [5]. Эти особенности способствуют высокой степени полиморфизма микросателлитных участков [1].

Популяционные исследования на основе анализа микросателлитов часто вызывают опасение у наших коллег и редко проводятся: на сайте elibagu.ru поиск по статьям по слову “микросателлит” в разделе “биология” дает только 31 статью. Для успешного проведения таких исследований вам может пригодиться наше краткое руководство.

Здесь мы приведем краткий план исследования, а в самой статье рассмотрим каждый этап подробно.

1. Формулировка задачи.
2. Подбор праймеров (если они не были подобраны ранее вами или вашими коллегами).
3. Подбор условий ПЦР (если они не были подобраны ранее вами или вашими коллегами).
4. Обработка результатов капиллярного фореза.
5. Выгрузка данных по длинам микросателлитных повторов в виде таблицы Excel.
6. Проверка, не собрали ли вы образцы тканей с одного объекта несколько раз (этот этап нужен не для всех исследований).
7. Выявление границ популяций и их число.
8. Выявление ошибок в результатах ПЦР микросателлитных участков.
9. Выявление нарушения равновесия Харди–Вайнберга, не зависящего от ошибок генотипирования, а отражающего реальное положение дел в природе (этот этап нужен не для всех исследований).
10. Изучение изменчивости локусов, включенных в ваш анализ, а также анализ распределения генотипов в популяциях.
11. Оценка эффективной численности популяции (этот этап нужен не для всех исследований).
12. Выявление недавнего прохождения популяции через “бутылочное горлышко” (этот этап нужен не для всех исследований).
13. Выявление миграции между популяциями (этот этап нужен не для всех исследований).

Отдельную проблему составляет отсутствие общепринятой русскоязычной терминологии либо ее незнание. В обиходе исследователи активно используют стихийно возникший жаргон, состоящий из не всегда удачных заимствований и калек с английских терминов. В настоящей статье мы иногда используем эти жаргонные названия, а также приводим англоязычные термины, чтобы читателям было проще сопоставить эту статью со своим опытом.

1. Формулировка задачи и методология

Для начала четко сформулируйте задачу – предполагаете ли вы наличие популяционной структуры, зависящей от географического или временного распределения, или же, например, наличие популяционной структуры паразитов в популяциях млекопитающих.

Для верного выбора методики проведения исследования желательно иметь представление о биологии объекта исследования.

Вид, недавно прошедший через “бутылочное горлышко” (или другие резко изменяющие частоту

генов неадаптивные популяционные процессы), даже расселившись на большой территории, может иметь низкое генетическое разнообразие и не показать никакой популяционной структуры даже на значительной выборке [6, 7].

Для вида, характеризующегося большим числом потомков и мультивольтинностью (наличием нескольких поколений в сезон), может быть затруднено определение истинного значения эффективной численности популяции и будет необходимо взять большую выборку [8].

Для паразитических видов (например, насекомых), живущих небольшими популяциями, напротив, может быть достаточно небольшой выборки для статистически достоверного исследования [9].

Старайтесь собирать материал в разных точках в одно и то же время, иначе из-за разных поколений степень внутривидовой генетической изменчивости может увеличиться, так же как и рассчитанная эффективная численность популяции, а степень дифференциации популяций, наоборот, уменьшится [10].

В общем случае при подготовке материала советуют изучать около 48–96 особей из каждой предполагаемой вами популяции [11–13], это число часто берется для технического удобства при плащечном выделении ДНК и постановке ПЦР. Выборка более 96 особей из популяции считается избыточной. В случае, если у вас не хватает материала, стоит стремиться к максимально возможной выборке из популяции. Работать можно и с маленькими выборками (желательно не менее 20 особей из популяции), но стоит быть готовым к тому, что при обработке результатов у вас может не хватить данных для выявления статистически значимых различий между популяциями [11–13].

Образцы тканей собираются согласно инструкциям, подходящим именно вашему объекту. Например, с целью минимизации разрушения ДНК образцы тканей насекомых помещают в 96%-ный этиловый спирт без предварительного замаривания (умерщвления) и хранят при температуре –20°C. При таких условиях хранения для микросателлитных исследований подходит давность сбора образцов до 10 лет.

2. Подбор праймеров

В случае, если похожие исследования были проведены до вас, вы можете использовать праймеры из статей коллег.

Если никто ранее не проводил исследований для вашего объекта, вы можете подобрать праймеры для микросателлитов самостоятельно, используя программу Websat [14, 15] на основе полногеномного секвенирования вашего объекта или близкородственного вида, которые можно найти в

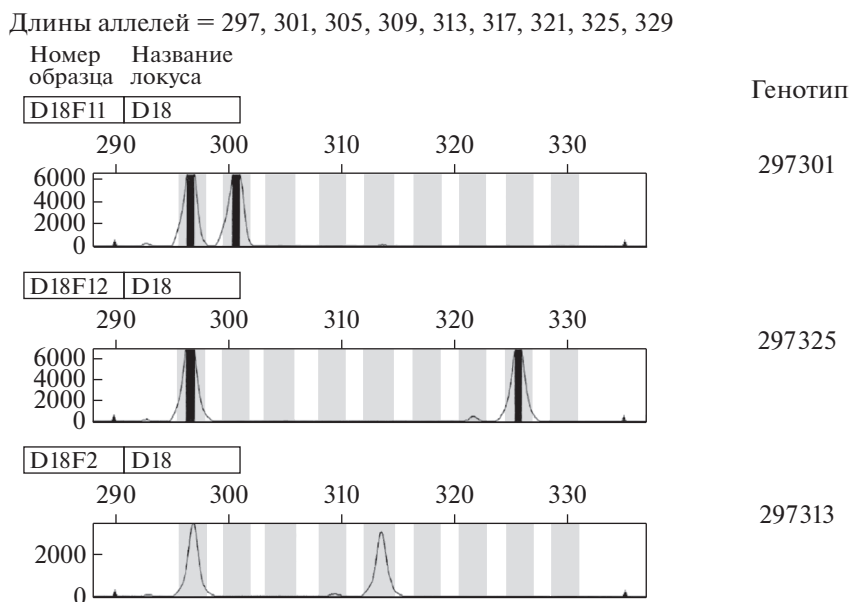


Рис. 1. Серии аллелей и некоторые генотипы проиллюстрированы здесь для тетра nukлеотидного кодоминантного микросателлита или STR локуса D18 с повторным мотивом [AGAA]_n, который широко используется в судебной экспертизе человека (по [11], стр. 7, с изменениями).

GenBank [16]. При выборе микросателлитных участков стоит выбирать четырехбуквенные и более длинные повторы сходной, но не одинаковой длины (например общей длиной от 200 до 300 нуклеотидов и различающиеся между собой на несколько нуклеотидов: 230, 234, 242 и так далее) (рис. 1).

Двухбуквенные повторы чаще подвержены “закисанию” (“проскальзыванию” во время амплификации, см. раздел 8 данной статьи) во время ПЦР, а трехбуквенные повторы чаще могут встраиваться внутрь гена (потому, что не сдвигают рамку считывания) и, соответственно, находиться под отбором. Стоит также учитывать, что скорости накопления мутаций зависят от длины повтора. Локусы с многобуквенными повторами могут не показать различий, в то время как по более коротким повторам различия будут наблюдаться. Старайтесь, чтобы в используемой панели маркеров локусы с разной длиной повтора были представлены равномерно.

Перед обработкой все праймеры проверяйте на вероятность образования шпилек и палиндромов при помощи программы OligoCalc [17, 18]. Праймеры из статей коллег также проверяйте на вероятность образования шпилек и димеров – возможно, они не учли этого в своей работе.

Как показывают исследования [9], среди заново подобранных праймеров 80–90% оказываются нерабочими. Нужно проверить их работоспособность на небольшой выборке вашего объекта, около 20 экземпляров [11]. Это можно сделать, заказав обычные, не меченые праймеры в любой

коммерческой фирме (в России наиболее популярны Евроген, Синтол и т.д.).

Выделение ДНК проводится методом, наиболее подходящим для вашего объекта с учетом способа сбора материала. В случае крупной выборки мы советуем пользоваться плащечным выделением [19, 20].

3. Подбор условий ПЦР

Исходя из данных программы Websat [15], выбирают несколько исходных температур отжига. Например, программа Websat рекомендует температуру 56°C, для дальнейшей отработки мы берем температуры 52, 54, 56, 58 и 60°C. Одна и та же реакционная смесь обрабатывается при пяти разных температурах отжига.

В процессе оптимизации условий амплификации для уменьшения количества неспецифического ПЦР-продукта массу ДНК, добавляемую в смесь, иногда изменяют со 100 до 50 мкг. Концентрацию праймеров также можно уменьшать в 2 раза для увеличения четкости сигнала в микросателлитном анализе.

Качество полученного ПЦР-продукта затем оценивается методом электрофореза в агарозном и/или полиакриламидном гелях, на основании чего выбирается температура отжига праймеров. Оптимальной считается такая температура, после отжига на которой ПЦР-продукт показывает наиболее яркую и четкую полосу после электрофореза.

ПЦР-продукт реакции с каждой парой праймеров должен показать на фореze либо одну четкую полосу (гомозигота по данному аллелю), либо две полосы (гетерозигота по данным аллелям). В случае, если на фореze нет полос, более двух полос, отсутствует изменчивость между особями (все полосы у всех особей одинаковые), имеется значительная полоса неспецифического продукта — эта пара праймеров вам не подойдет.

Локусы, демонстрирующие высокую степень полиморфизма, пригодны для их использования в популяционных исследованиях. Неизменяемые локусы (например, локусы, одинаковые у всех изученных образцов) убираются из анализа.

Наиболее полиморфные локусы являются и наиболее информативными.

В общем случае для популяционного исследования необходимо хотя бы 9–10 микросателлитных участков [21]. Исключения составляют некоторые специфические работы, к примеру изучение локуса, сцепленного с интересующим вас моногенным признаком [22].

В дальнейшей работе используются праймеры, меченные флуоресцентными красителями трех цветов (FAM, R6G, TAMRA), их также можно заказать в любой коммерческой фирме (Евроген, Синтол и т.д.). ПЦР проводят, смешивая в одной пробирке набор из трех праймеров каждого цвета и получая мультиплекс. Дополнительно подбирая условия и диапазоны длин ампликонов, можно создать мультиплексы с использованием локусов, меченных одним цветом, но разного размера. В случаях, когда совместное использование праймеров из разных пар приводит к образованию неспецифического продукта, целесообразно проводить ПЦР независимо и смешивать уже готовые ампликоны.

Анализировать результаты ПЦР мультиплексов удобно на капиллярном фореze. К образцам предварительно добавляется раствор формамида (HiDi) с набором флуоресцентно меченых фрагментов заранее известной длины с шагом в 20 нуклеотидов (Ladder, “лэддер”, “лесенка”), который можно заказать в коммерческой фирме. Калибровка по лэддеру позволяет при обработке определить длину изучаемых ПЦР-продуктов ДНК с точностью до нуклеотида.

В большинстве работ, посвященных вариативности длин коротких tandemных повторов, вариативность повтора не учитывается — вместо этого считается его суммарная длина, которая затем переводится в число повторов [11–13].

В случае вероятности ошибок из-за плохой сохранности ДНК можно провести ПЦР несколько раз [6].

4. Обработка результатов капиллярного фореza

Обработка производится в программе GeneMarker [23]. После калибровки полученных данных по Ladder составляются панели со всеми представленными в выборке аллелями (длинами повторов) каждого микросателлитного локуса. Это возможно при наложении всех пиков одного цвета на общей картинке. После того как на панели будут обозначены все варианты длин ПЦР-продукта, полученная панель загружается в перечень панелей программы, после чего GeneMarker автоматически анализирует фореz каждого образца и представляет соответствующую длину повтора. Во избежание ошибок каждый образец проверяется вручную, при необходимости результаты автоматического анализа корректируются или дополняются (рис. 1, 2).

5. Выгрузка данных по длинам микросателлитных повторов в виде таблицы Excel

Данные по длинам микросателлитных повторов выгружаются для дальнейшей статистической обработки в виде таблицы Excel, из которой данные удобно переформатировать в другие форматы файлов. Для удобства просмотра текстовых файлов мы рекомендуем программу Notepad++ [24].

Стоит помнить, что все микросателлитные данные — кодоминантные, т.е. в результатах анализа мы всегда можем отличить гетерозиготу от любой из гомозигот.

Для обcчетов микросателлитных данных в настоящее время существует множество программ. Ниже мы порекомендуем некоторые из них, но советуем быть внимательными. В разных программах один и тот же индекс может иметь разные условные обозначения, например, индекс Шеннона обозначается как I в программе GenAlEx [25] и как H в пакете “poppr” пространства R [26]. Наоборот, одно и то же обозначение может иметь разный смысл в разных программах, например, в программе GenAlEx эффективное число аллелей (No. of Effective Alleles) обозначается как “Ne”, т.е. так же, как и эффективная численность популяции, вычисляемая в программе NeEstimator [27–29].

Подробно о форматах, которые может обработать программа, и о специфике ее работы смотрите в инструкции. Этот совет применим ко всем описанным ниже программам.

6. Проверка, не собрали ли вы образцы тканей с одного объекта несколько раз

Если вы собирали материал в одной точке несколько лет подряд, не убивая объект (вы брали пробы крови, экскрементов, шерсти или слюны), а ваши объекты живут больше года, нужно прове-

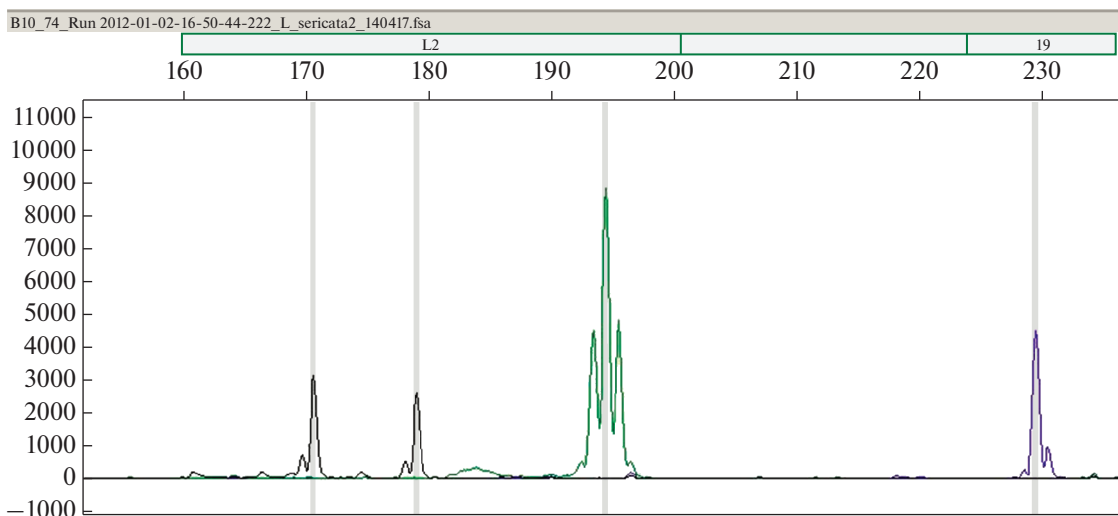


Рис. 2. Визуализация данных капиллярного электрофореза в программе GeneMarker. Сигнал продукта, амплифицированного с трех различных локусов, обозначен, соответственно, зеленым, синим и черным цветом. Аллель отмечен серой вертикальной линией. Примерную длину продукта можно оценить по горизонтальной шкале сверху. Наличие двух пиков говорит о гетерозиготности образца по данному локусу (черный). Несколько пиков (зеленый) иллюстрируют пример “заикания” во время ПЦР (“статеры” – жаргонизм).

речь, не собрали ли вы образцы тканей с одного объекта несколько раз с помощью сравнения генотипов в программе Cervus [30, 31] и/или программе Pedigree v. 2.0 [32, 33]. Чтобы избежать ложного исключения экземпляров из анализа, рекомендуется вручную перепроверять все случаи совпадения генотипов. Также совпадение генотипов можно проверить в программе Mstools [34]. Наличие сибсов, т.е. потомков одних родителей, можно проверить с помощью программы GIMLET [35, 36].

7. Выявление границ популяций и их число

Если вы продолжаете давнее исследование и добываете экземпляры из популяций, уже ранее исследованных с помощью микросателлитного анализа, то в этом случае вы уже примерно представляете границы популяций. Тогда в начале исследования имеет смысл проверить данные на соответствие равновесию Харди–Вайнберга внутри популяций (см. ниже, раздел 8). Если вы только начинаете свое исследование, сначала надо постараться определить примерные границы популяций.

Чтобы выявить границы популяций и их число, возможно использовать различные программы. Сейчас используются программы, основанные на определении популяционной структуры и наиболее вероятном разделении данных на кластеры с применением алгоритма Монте-Карло по схеме марковских цепей (MCMC) для байесовской статистики. Это программы Geneland [37, 38] и STRUCTURE [39, 40].

В обеих программах (Geneland и STRUCTURE) для микросателлитных данных основное допущение состоит в том, что предполагаемые популяции находятся в приблизительном равновесии Харди–Вайнберга с равновесием по сцеплению (HWLE: Hardy–Weinberg Linkage Equilibrium).

Программа Geneland может обрабатывать генотипические данные (микросателлиты – microsatellites; однонуклеотидные замены (“снимпы” – жаргонизм) – SNPs), фенотипические данные, а также опционально может использовать пространственные координаты экземпляров (или популяций).

В программу внедрены алгоритмы, основанные на разных моделях. Самый популярный алгоритм основан на пространственной модели (spatial model) и может использовать не только генотипы, но и пространственные координаты экземпляров (или популяций).

Пример настроек программы Geneland: number of populations 1...10 (число предполагаемых популяций ставим от единицы до максимального числа, которое мы можем предположить, с некоторым запасом; например, мы предполагаем, что максимальное число популяций может быть 8; в таком случае ставим от одного до десяти); number of iterations: 1000000 (в большинстве случаев этого числа итераций MCMC достаточно для стабилизации всех параметров модели и накопления необходимого количества данных); thinning: 100 (мы записываем каждый 100-й шаг для экономии места на жестком диске); allele frequency model: uncorrelated (если мы не уверены, что в разных популяциях частоты одних и тех же аллелей сход-

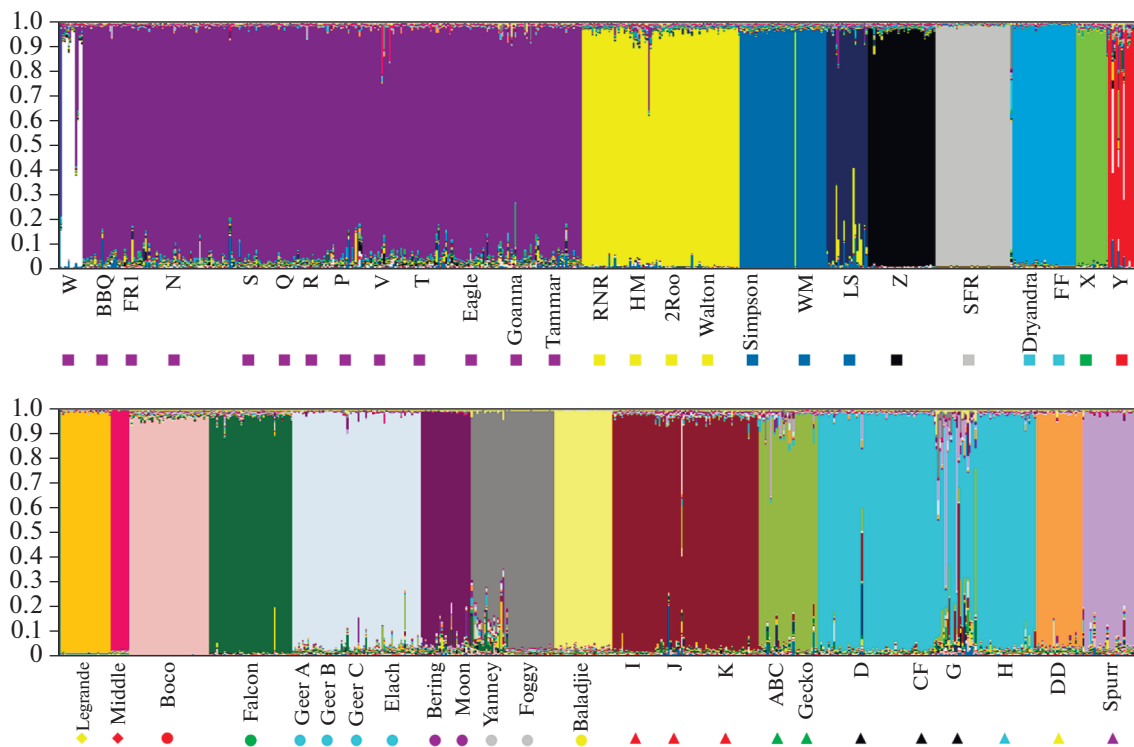


Рис. 3. Результаты обсчета данных в STRUCTURE для ящериц *Stenophorus ornatus*, $K = 23$. Каждый столбец представляет собой отдельный экземпляр, каждая вертикальная линия представляет долю членства особи в обозначенном цветом кластере (по [119]).

ны между собой); spatial model: false (если в нашем анализе нет географических координат); null allele model: true (если мы не уверены в отсутствии нулевых аллелей на данном этапе обработки данных). Рекомендуется провести несколько независимых запусков процесса (multiple independent runs) для каждого предполагаемого числа популяций, например 5 или 7 запусков, чтобы проверить конвергентность результатов. После завершения расчетов рекомендуется убрать из анализа первые 10% результатов (до стабилизации параметров модели), чтобы убрать шум из результатов: recalculate with burnin 1000.

В Geneland может быть учтено предполагаемое присутствие нулевых аллелей для диплоидных данных (null allele model: true). Обратите внимание, что если эта опция используется, все двойные отсутствующие генотипы будут интерпретироваться как двойные нулевые аллели. Это может вызвать чрезмерную оценку нулевых частот аллелей, если некоторые из недостающих данных не являются нулевыми аллелями (т.е. не из-за проблем с амплификацией в ПЦР), а, например, если некоторые локусы отсутствуют для всех индивидуумов в определенных выборках.

Число кластеров для каждого из независимых запусков программа Geneland определяет автоматически. В случаях, когда результаты не совпада-

ют или преобладание плотности одного из вариантов в алгоритме MCMC незначительно, следует повторить анализ с большим числом итераций.

Программа STRUCTURE (рис. 3–5) сходна по принципу работы с программой Geneland и их часто используют одновременно в одной статье, чтобы убедиться в результатах выделения границ популяций. Параметры работы программы STRUCTURE подбирают на основании соответствующих статей [41, 42]. Для определения истинного количества кластеров используется метод, предложенный Эванно в 2005 г. [41] и реализованный сервисом STRUCTURE HARVESTER [43]. На рис. 5 показан результат применения этой программы к тем же данным, что и на рис. 4. Пик на приведенном графике отражает наиболее вероятное число исходных кластеров, которое оказалось равно одному. Таким образом, по рисунку можно заключить, что достоверная разница между исследуемыми популяциями не была обнаружена и разделения генетической структуры на кластеры не наблюдается.

Пример настроек программы STRUCTURE: анализ проводят с использованием модели генетического смешения (admixture) и скоррелированности частот аллелей в разных популяциях (т.е. предполагается, что частоты аллелей в разных популяциях сходны — например, из-за ми-

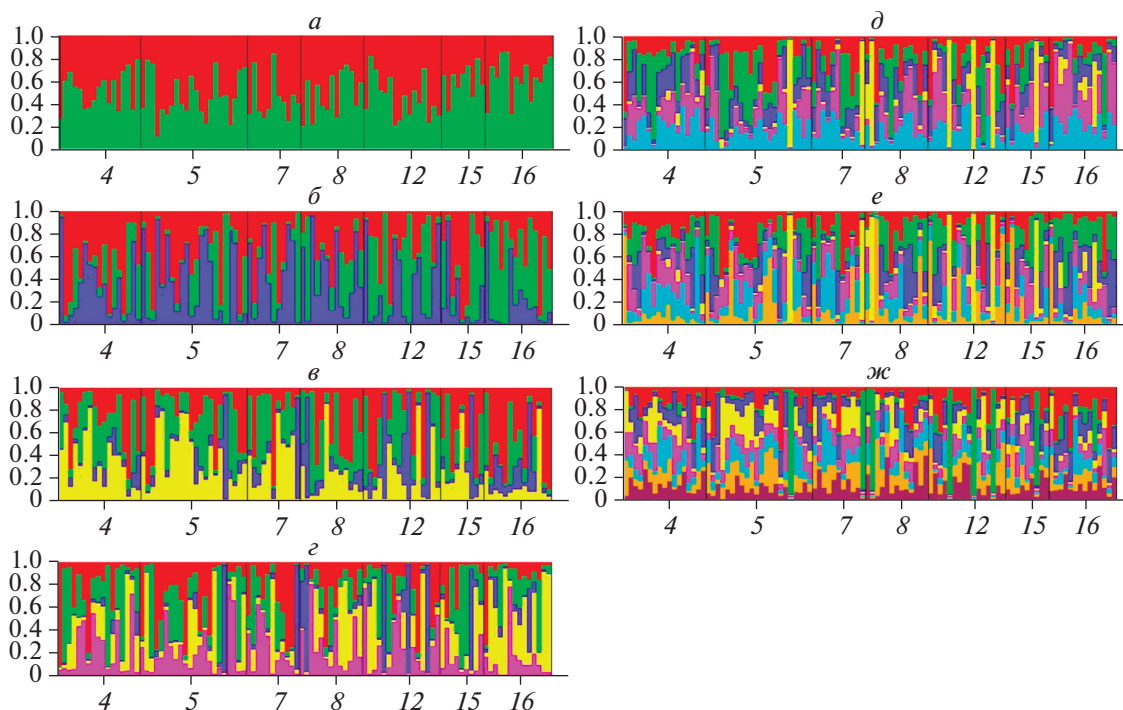


Рис. 4. Результаты обчета данных в STRUCTURE для мух *Lucilia sericata* для $K = 2$ (а), 3 (б), 4 (в), 5 (г), 6 (д), 7 (е), и 8 (ж) кластеров соответственно. Каждая вертикальная линия представляет долю членства особи в обозначенном цветом кластере. Цифры отражают географическую принадлежность особей (4 – Евпатория, 5 – Коктебель, 7 – Зуя, 8 – Севастополь, 12 – Алексин, 15 – Волгоград, 16 – Липецк). Как можно видеть из приведенной иллюстрации, соотнести географическую приуроченность образцов с принадлежностью к кластерам не представляется возможным ни при одном смоделированном количестве кластеров (от 2 до 8) (по [120]).

граций или недавнего разделения). Предварительный выбор стартовой точки марковской цепи (burn-in) проводят в течение 500000 итераций, с последующим построением марковской цепи в течение 1000000 итераций для предполагаемого числа групп в выборке K , равного от 1 до 11 в шести повторностях для каждой величины K [44, 45].

8. Выявление ошибок в результате ПЦР микросателлитных участков

Ошибки в результатах ПЦР микросателлитных участков могут быть вызваны низкими концентрациями матричной ДНК [46], что может привести к тому, что аллель не будет амплифицироваться из-за случайной ошибки в ходе ПЦР (“выпадение аллеля” (“аллельный выпад” – жаргонизм), “allelic dropout” [47]). Ошибки генотипирования также могут возникать из-за преимущественного усиления коротких аллелей, т.е. “отсева длинных аллелей”, или “доминирования коротких аллелей” [48], когда более длинный аллель специфически не амплифицируется. Кроме того, может возникать “проскальзывание” во время ПЦР-амплификации (“заикание”), которое приводит к образованию дополнительных ПЦР-продуктов, отличающихся от нужного продукта кратно

длине повторяющейся единицы микросателлита [49]. Такие “заикания” часто встречаются в локусах динуклеотидов (рис. 2), что затрудняет различение гомозигот и гетерозигот. Наконец, когда

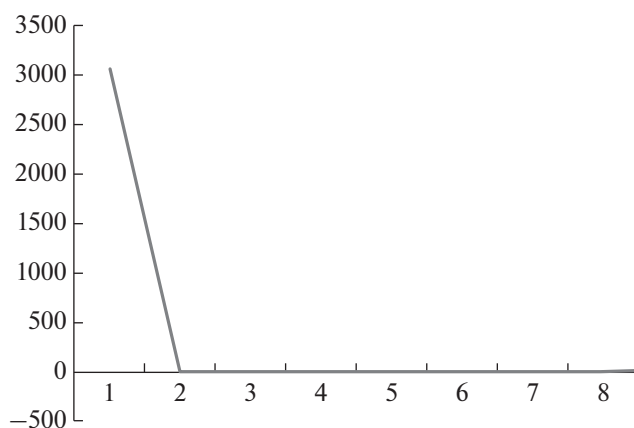


Рис. 5. Значения функции ΔK (ось ординат) от возможного числа кластеров (ось абсцисс). Значение, отложенное по оси X , указывает на истинное количество кластеров (в данном случае, 1). Высота пика отражает выраженность популяционной структуры (по [120]).

мутации происходят на сайтах праймеров, некоторые аллели не могут быть амплифицированы (нуль-аллели), что приводит к ложным гомозиготам [50].

Такие ошибки генотипирования могут вызывать отклонения в распределении аллелей внутри популяций, выявленных в программе GeneLand или STRUCTURE, от равновесия Харди–Вайнберга (Hardy–Weinberg Equilibrium, HWE) и равновесия по сцеплению (Linkage Equilibrium, LE), в частности, это может быть дефицит гетерозигот [50], изменяющий генетический анализ популяций. Эти отклонения часто очень похожи на те, которые вызваны инбридингом, ассортативным скрещиванием или эффектом Валунда (Wahlund effect) [51].

Закон Харди–Вайнберга сформулирован для изолированной бесконечно большой панмиктической популяции. На практике отклонения от этого закона могут возникать из-за ошибок генотипирования, а могут отражать реальное распределение генотипов в популяции.

Ошибки генотипирования: “доминирование коротких аллелей”, “заикание” и нуль-аллели дают свою собственную специфическую аллельную “подпись” (т.е. недостатки и избытки конкретных генотипов), в отличие от отсева аллелей в случае непанмиксии, который, как предполагается, в значительной степени не зависит от размера аллеля [47, 52]. Таким образом, можно различать отклонения, вызванные непанмиксией, и те, которые вызваны различными ошибками генотипирования.

Локусы возможно протестировать на наличие ошибок генотипирования в программе Micro-Checker [53, 54]. Мы советуем использовать установочную версию программы (Installed Version), а не переносную (Stand-alone Version).

Micro-Checker может помочь выявить ошибки генотипирования из-за нуль-аллелей, “доминирования коротких аллелей” и “заикания”. Там, где доступны мультилокусные генотипы, программа может различать инбридинг и эффект Валунда и отклонения от равновесия Харди–Вайнберга, вызванные непанмиксией. Кроме того, она может идентифицировать возможные ошибки при ручном вводе данных. Примеры ошибок генотипирования и рекомендации по их исправлению приведены в руководстве пользователя. В случае наличия нуль-аллелей Micro-Checker может оценивать частоту нулевого аллеля и соответственно корректировать наблюдаемые частоты аллелей и генотипов. Эти отрегулированные частоты аллелей могут использоваться впоследствии для дальнейшего генетического анализа популяции, например, в программах Genepop [55, 56], Arlequin [57, 58] или Fstat [59, 60].

Нулевые аллели с большей вероятностью встречаются у популяций с большой эффективной численностью (N_E) и с высокой скоростью мутаций во фланкирующих областях генома (областях, где садятся праймеры); а также в тех популяциях, которые генетически разошлись с популяцией, на основе экземпляров из которой были разработаны праймеры. При значительном расхождении популяций коэффициент инбридинга F_{ST} (см. ниже, раздел 10) и генетические расстояния некорректно завышаются при наличии нулевых аллелей [61].

В настоящее время существует множество других программ, которые помогают выявлять ошибки генотипирования. Gimlet [35] может идентифицировать ложные гомозиготы (нуль-аллели) и ложные аллели (“заикание”), но эта программа требует повторного генотипирования каждого образца. Аналогичным образом, повтор генотипирования также требуется для использования метода максимального правдоподобия [52]. Программное обеспечение, такое как Pedmanager [62, 63], выполняет проверку ошибки менделевского наследования, которая требует записей о предках исследуемых экземпляров. Программа Cervus [30] устанавливает наличие нулевых аллелей, анализируя отклонения от равновесия Харди–Вайнберга с применением критерия хи-квадрат, и использует алгоритм, основанный на различии наблюдаемой и ожидаемой частоты гомозигот для оценки частоты нулевых аллелей [64]. Однако Cervus не может отличить нуль-аллели от других ошибок генотипирования и не предоставляет альтернативных методов для оценки нулевых аллелей.

Отклонения от соответствия распределения аллелей внутри предполагаемых популяций равновесию Харди–Вайнберга могут быть выявлены с помощью пакета “pegas” пространства R [65]. При использовании пространства R [66–68] удобнее применять графический интерфейс [69]. Подробная инструкция для популяционно-генетических расчетов в пространстве R есть по адресу [70].

Отклонения от равновесия Харди–Вайнберга могут быть также выявлены с помощью программы GENETIX [71, 72] или Genepop [55, 56].

Оценка частот нулевых аллелей и поправки на них показателей разнообразия и индексов фиксации могут быть проведены в программе FreeNa [61, 73, 74] и в программе ML-NullFreq [75, 76].

Оценки частот нулевых аллелей могут быть проведены в программе Geneland [38].

9. Выполнение нарушения равновесия Харди–Вайнберга, не зависящего от ошибок генотипирования, а отражающего реальное положение дел в природе

На практике часто наблюдается нарушение равновесия Харди–Вайнберга, не зависящее от ошибок генотипирования, а отражающее реальное положение дел в природе. Эти отклонения объясняются нарушением тех или иных предположений, характеризующих идеальную популяцию: а) конечные размеры, б) действие отбора на данный локус; в) неслучайное скрещивание особей, в том числе наличие в популяции какой-либо подразделенности, г) миграция. Впрочем, хотя каждый из этих факторов может вызывать такие отклонения, их действие отнюдь не гарантирует их наличия.

Следует отметить, однако, что, основываясь только на данных по распределению генотипов по данному локусу в данной популяции, нельзя сделать выбор в пользу одного из этих объяснений. Поэтому не стоит сразу интерпретировать отклонение от равновесия Харди–Вайнберга как указание на действие отбора.

Для начального знакомства с методами детекции отбора в геноме советуем обратиться в блог *evolgen* [77] (это последний пост по этой теме, в нем есть ссылки на предыдущие посты).

10. Изучение изменчивости локусов, включенных в ваш анализ, а также анализ распределения генотипов в популяциях

Понять, насколько изменчивы локусы, включенные в ваш анализ, а также провести анализ разнообразия и равномерности распределения генотипов в популяциях можно с помощью программы GenAlEx [25]. Вы можете рассчитать такие показатели, как N – объем выборки, N_a – число различных аллелей, N_E – эффективное число аллелей, наблюдаемую (H_O) и ожидаемую (H_E) гетерозиготность, индекс фиксации Райта (F_{IS}), а также генетическое расстояние Неи ($Nei D$) между отдельными популяциями, и др.

В некоторых статьях вы можете увидеть, как эти параметры подсчитывают для “популяций”, выделяя их не на основании результата математических обчислений в Geneland и STRUCTURE, а по географическому принципу. Такой подход распространен, но не совсем корректен, так как выделение групп особей по чисто географическому принципу не всегда верно отражает популяционную структуру вида.

Сразу можем порекомендовать руководства по программе GenAlEx [11], которые также включают подробное разъяснение сути и смысла индексов, которые мы будем обсуждать ниже в этом разделе.

Что означают эти вычисленные параметры?

Одна из мер генетического разнообразия – число аллелей в популяции. Число аллелей легко измерить, но его недостаток в том, что он зависит от численности выборки, по которой оценивается разнообразие. Это затрудняет сравнение изменений, полученных для выборок разной численности [78].

Генетический дрейф может изменять частоты аллелей и может необратимо удалять аллели из популяции. Однако этому снижающему разнообразие эффекту дрейфа противодействует мутационный процесс – и в результате уровень разнообразия устанавливается на некотором уровне, определяемым мутационно-дрейфовым равновесием [78]. Подчеркнем, что это равновесие – динамическое: новые аллели возникают в ходе мутаций, а старые вымываются дрейфом.

Эффективное число аллелей – это нижняя оценка числа аллелей, одновременно присутствующих в популяции, при условии, что генетический дрейф и мутагенез уравновешены.

Еще одна мера генетического разнообразия – *доля гетерозигот* в популяции (H), и она не так прямо зависит от численности выборки [78].

H_O – наблюдаемая гетерозиготность внутри популяции, т.е. доля образцов, которые гетерозиготны в данном локусе.

H_E – ожидаемая гетерозиготность внутри популяции, т.е. доля гетерозиготности, ожидаемая при случайном скрещивании и вычисляемая при помощи равновесия Харди–Вайнберга.

Гетерозиготность связана со скоростью мутагенеза и численностью популяции и может быть в дальнейшем использована для выяснения эффективной численности популяции (см. ниже, раздел 11).

Можно вычислить ожидаемую и наблюдаемую гетерозиготности, усредненные по субпопуляциям, и общую ожидаемую гетерозиготность для всей выборки.

В некоторых статьях используют разницу ($H_O - H_E$): если она меньше нуля, то считается, что наблюдается дефицит гетерозигот [79]. Однако этот параметр не является необходимым, так как дублирует F -статистику Райта [80–82].

Генетический дрейф действует независимо в разных (суб)популяциях и потому может приводить к фиксации в них разных аллелей из исходно присутствующих в генофонде. На основе этого разработаны индексы, характеризующие степень генетической дифференциации.

При соблюдении равновесия Харди–Вайнберга вероятность *фиксации* аллеля равна его текущей частоте. Скорость фиксации – это произведение числа вновь порожденных аллелей за поколение на вероятность фиксации. Насколько быстро и как

часто новые аллели достигают фиксации, определяется временем до фиксации и скоростью фиксации [78].

Пожалуй, наиболее распространенными статистическими показателями в популяционной генетике является F -статистики Райта [80–82].

F_{IS} – коэффициент инбридинга у особей в популяции. Он измеряет уменьшение гетерозиготности индивида из-за неслучайного скрещивания в пределах его субпопуляции.

Индекс фиксации F_{IS} внутри популяции имеет значения от -1 до $+1$. Значения, близкие к нулю, ожидаются при случайном скрещивании (панмиксии). Существенные положительные значения указывают на инбридинг или необнаруженные нуль-аллели. Отрицательные значения указывают на превышение гетерозиготности из-за отрицательного ассортативного скрещивания или отбора в пользу гетерозигот.

F_{IT} – коэффициент инбридинга у особей в общей выборке. Эта статистика учитывает эффекты как неслучайного скрещивания внутри субпопуляций, так и генетической дифференциации среди субпопуляций.

F_{ST} – коэффициент инбридинга в популяциях по отношению к общей выборке. Этот коэффициент отражает меру генетической дифференциации между субпопуляциями, т.е. долю общего генетического разнообразия (гетерозиготности), которая распределяется среди субпопуляций. F_{ST} имеет значения от 0 до $+1$, но при некоторых подходах к расчетам возможно получение отрицательных значений. В последнем случае следует считать, что F_{ST} равен нулю. Если все популяции находятся в равновесии Харди–Вайнберга с одинаковыми частотами аллелей, то $F_{ST} = 0$.

На практике F_{ST} редко превышает 0.5 и часто намного меньше. Райт [83] предложил такую интерпретацию: значения $F_{ST} \geq 0.25$ считаются очень большими различиями между субпопуляциями; диапазон от 0.15 до 0.25 указывает на умеренное дифференцирование; в то время как дифференциация является незначительной, если F_{ST} составляет 0.05 или менее. Однако интерпретация величины F_{ST} более сложна, чем простая ссылка на это количественное руководство. Хедрик [84] показал, что с современными гипервариабельными маркерами, характеризующимися многими аллелями, значения F_{ST} могут быть значительно ниже, чем для генетических маркеров с очень небольшим количеством аллелей. Поэтому в современных популяционно-генетических методах более важный вопрос заключается в том, можем ли мы обнаружить статистически значимую генетическую дифференциацию ($F_{ST} > 0$) или нет. С помощью AMOVA возможно провести такие статистические тесты [11].

Для рассчитанного по вашим данным индекса F_{ST} может быть также вычислена выраженная через p -значение статистическая значимость отличия от нулевой гипотезы об отсутствии различий между популяциями. Интерпретация значимости в этом случае стандартна: если p -значение меньше 0.05, то популяции значимо различаются.

В некоторых статьях вы можете увидеть заключение о том, что $F_{ST} = 0.11$ означает высокую генетическую дифференциацию [10]. Несмотря на кажущееся противоречие с вышеописанными критериями Райта, такие фразы часто означают, что в других работах, проведенных на том же объекте (например, на том же виде), значения этих индексов были еще ниже, чем в упомянутой статье.

Индекс разнообразия Шеннона [85] широко используется в экологии [86], но менее распространен в популяционной генетике. Значения индекса Шеннона для микросателлитных маркеров считаются высокими, если они не менее 1.5. В таком случае они указывают на достаточно высокое генетическое разнообразие в исследованных популяциях [79].

Генетическая идентификация Неи (Nei's I) может принимать значения от 0 до 1. Генетическое расстояние Неи (Nei's D) варьируется от 0 до бесконечности [87, 88]. Оценка статистической значимости Nei's I и Nei's D доступна в программе GenAlEx. Хедрик [84] предполагает, что эта оценка может дать недостоверные результаты, когда гомозиготность низкая и численность выборки мала. Генетическое расстояние Nei может необоснованно принимать слегка отрицательные значения, которые должны интерпретироваться как ноль.

Графическое представление дистанций Неи между популяциями в виде дерева может быть выполнено с помощью пакета "poppr" пространства R [89].

Изменчивость локусов, анализ относительного обилия, разнообразия и равномерности распределения генотипов в популяциях также можно рассчитать с помощью пакета "poppr" пространства R [26, 90].

Аллельное разнообразие (AR) может также быть вычислено в программе FSTAT [59, 60]; ожидаемая (H_E) и наблюдаемая (H_O) гетерозиготность по локусу и по популяции может быть вычислена в программе Arlequin [57, 58], а также в программах PopGene [91], FSTAT [59, 60] и GenStat [92].

Степень генетической дифференциации популяций также может быть вычислена на основании параметра G_{ST} , который по своей сути сходен с индексами F_{ST} , R_{ST} и D . Существуют давние дебаты о том, какой из этих четырех параметров целесообразнее использовать [93, 94]. С нашей точки

зрения, и G_{ST} , и F_{ST} , и R_{ST} , и D использовать правомерно.

Параметр G_{ST} может быть подсчитан с помощью пакета “poppr” пространства R [95].

Индекс F_{ST} может быть также вычислен в программе GENETIX [72].

Точный тест Фишера (Fisher exact test) может быть проведен в программе Genepop [56] и также может быть использован для выявления степени генетической дифференциации популяций [96], границы которых были определены с помощью байесовских методов кластеризации в программах Geneland и STRUCTURE.

11. Оценка эффективной численности популяции

С помощью программы NeEstimator [27–29] можно оценить эффективную численность популяции (N_E).

Программа NeEstimator позволяет использовать несколько методов оценки эффективной численности популяции: “Heterozygote-excess” [97], “Linkage Disequilibrium” [98] и “Molecular Coancestry” [27].

Эффективная численность популяции имеет большой биологический смысл и тесно связана с действием генетического дрейфа. Генетический дрейф по данному локусу останавливается только в случае утраты разнообразия, т.е. фиксации одного из аллелей. Таким образом, утрата или фиксация аллеля вследствие дрейфа неизбежна, но для больших популяций любой из этих исходов может потребовать очень длительного времени [78]. Важно, что этот процесс происходит и в отсутствие каких-либо селективных различий между аллелями, что часто верно в случае молекулярной эволюции [99, 100]. Время, за которое гетерозиготность под действием дрейфа уменьшается в 2 раза, обратно пропорционально эффективной численности популяции [78]. Эффективная численность популяции в первом приближении — это число особей, участвующих в размножении. Более строго, это параметр, характеризующий скорость утраты гетерозиготности в популяции, и потому она зависит, кроме числа размножающихся особей, от других факторов, влияющих на эту скорость [51, 101]. Эффективная численность популяции обычно гораздо меньше, чем фактическое число особей [78].

В силу описанной выше связи эффективной численности с генетическим дрейфом, потеря разнообразия вследствие дрейфа особенно существенна для малых популяций [78]. Выявление достоверной эффективной численности популяций охраняемых видов, находящихся под угрозой исчезновения [21], имеет значение для восстановления малых популяций этих видов, поскольку эта оценка обеспечивает основу для прогнозирования

стойкости этих популяций и их способности адаптироваться к изменениям окружающей среды.

В популяциях конечного размера положительный отбор и генетический дрейф действуют в противоположных направлениях. Показано, что для вероятности фиксации вновь возникшего полезного аллеля в популяции имеет решающее значение не селективное преимущество само по себе, а его отношение к эффективной численности популяции [78] — таким образом, в малых популяциях даже полезный аллель может с достаточно высокой вероятностью исчезнуть.

При свободном скрещивании в небольшой группе за одно поколение теряется $1/(2N)$ доля гетерозиготности (N — эффективная численность популяции), и окончательным уделом такой группы является полная гомозиготность, если нет потока иммигрантов (и не возникает новых мутаций — но в малой популяции утрата генетического разнообразия идет настолько быстро, что этим можно пренебречь). С другой стороны, иммигранты (см. ниже, раздел 13) будут препятствовать достижению полной гомозиготности в популяции [51]. Поэтому в случае редких или вымирающих видов [21] при обнаружении изолированных популяций с небольшой эффективной численностью решением проблемы может стать организация “коридоров” или переселение особей между популяциями для обеспечения миграции и противодействию генетическому дрейфу.

Иногда в программе NeEstimator вы можете получить результат “infinite”, то есть эффективный размер популяции был оценен как “бесконечно большой”. Авторы программы предлагают интерпретировать данный результат как наличие единой панмиктической популяции и, вероятно, активного обмена генами внутри нее. Способность особей некоторых видов преодолевать большие расстояния и их высокая плодовитость способствуют тому, что миграции даже небольшого количества особей хватает для нивелирования эффектов генетического дрейфа и географической удаленности точек.

Однако необходимо учитывать, что определению эффективной численности популяций как “бесконечно большого” могли послужить также ошибка выборки и возможное наличие перекрывающихся популяций в сборах [8, 102].

Оценка эффективной численности популяции может быть проведена, помимо вышеупомянутых, в программе LDNe [98, 103, 104] и в программе OneSamp [105, 106].

12. Выявление недавнего прохождения популяции через “бутылочное горлышко”

Анализ микросателлитов позволяет выявить недавнее прохождение популяции через “буты-

лочное горлышко” [6], т.е. резкое сокращение численности, вымывающее изменчивость, с последующим восстановлением. Предположение о “бутылочном горлышке” можно проверить с помощью программы Bottleneck [107, 108].

13. Выявление миграции между популяциями

Достаточно часто бывает, что популяции не полностью изолированы друг от друга. Частоты аллелей могут изменяться вследствие миграции особей [51]. Миграцию между популяциями можно вычислить с помощью программ BayesAss [109, 110], STRUCTURE или Geneclass [111, 112]. Математическая модель возможных миграций между локальными популяциями на основе числа мигрантов на поколение (Nm) реализована в R-пакете divMigrate [113, 114].

Наиболее очевидное следствие продолжительных межпопуляционных миграций состоит в том, что частоты аллелей в разных группах все более и более приближаются друг к другу и в отсутствие каких-либо противодействующих сил делают вид в целом более гомогенным [78].

ЗАКЛЮЧЕНИЕ

В случае, если вы столкнулись с проблемой, не освещенной в данной статье, мы советуем внимательно посмотреть еще раз инструкцию к используемой программе. Также мы можем посоветовать обращаться к коллегам, которые уже проводили сходные по методике исследования.

Не забывайте о том, что у всех используемых методов есть некоторые исходные допущения, которые могут быть неадекватны вашим данным. Поэтому рекомендуем подробнее разбираться в сути применяемых расчетов, чтобы оценить, насколько правомочны выводы, сделанные по вашим данным.

Для самостоятельного изучения на начальном этапе работы можем порекомендовать книги с описанием популяционных процессов: книга Ли [51], глава 10 в книге Хаубольда и Вие [78] и глава 2 в книге Эвиса [115].

Про Байесову статистику вы можете почитать в книгах Ли и Вагенмэйкера [116], Гелмана и др. [117] и Крашке [118].

Работа Т.В. Галинской выполнена за счет гранта Российского научного фонда (проект № 18-74-00035). Работа С.Н. Лысенкова выполнена в рамках госзадания Ч. 2 № ЦИТИС АААА-А16-116021660031-5.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Bowcock A., Ruiz-Linares A., Tomfohrde J. et al.* High resolution of human evolutionary trees with polymorphic microsatellites // *Nature*. 1994. V. 368. P. 455–457. doi 10.1038/368455a0
2. *Jarne P., Lagoda P.J.L.* Microsatellites, from molecules to populations and back // *Trends Ecol. Evol.* 1996. V. 11. № 10. P. 424–429. doi 10.1016/0169-5347(96)10049-5
3. *Li Y.-C., Korol A.B., Fahima T., Nevo E.* Microsatellites within genes: Structure, function, and evolution // *Mol. Biol. Evol.* 2004. V. 21. № 6. P. 991–1007. doi 10.1093/molbev/msh073
4. *López-Flores I., Garrido-Ramos M.A.* The repetitive DNA content of eukaryotic genomes // *Genome Dynamics*. 2012. V. 7. P. 1–28. doi 10.1159/000337118
5. *Pumpernik D., Oblak B., Borštnik B.* Replication slippage versus point mutation rates in short tandem repeats of the human genome // *Mol. Genet. Genom.* 2008. V. 279. № 1. P. 53–61. doi 10.1007/s00438-007-0294-1
6. *Alasaad S., Soriguer R.C., Chelomina G. et al.* Siberian tiger’s recent population bottleneck in the Russian Far East revealed by microsatellite markers // *Mammal Biol.* 2011. V. 76. № 6. P. 722–726. doi 10.1016/j.mambio.2011.01.011
7. *Henry P., Miquelle D., Sugimoto T. et al.* In situ population structure and ex situ representation of the endangered Amur tiger // *Mol. Ecol.* 2009. V. 18. № 15. P. 3173–3184. doi 10.1111/j.1365-294X.2009.04266.x
8. *Diakova A.V., Schepetov D.M., Oyun N.Y. et al.* Assessing genetic and morphological variation in populations of Eastern European *Lucilia sericata* (Diptera: Calliphoridae) // *Europ. J. Entomol.* 2018. V. 115. P. 192–197. doi 10.14411/eje.2018.017
9. *Ascunce M.S., Touns M.A., Kassu G. et al.* Nuclear genetic diversity in human lice (*Pediculus humanus*) reveals continental differences and high inbreeding among worldwide populations // *PLoS One*. 2013. V. 8. № 2. P. 1–12. doi 10.1371/journal.pone.0057619
10. *Pérez-Alvarez M.J., Olavarría C., Moraga R. et al.* Microsatellite markers reveal strong genetic structure in the endemic Chilean dolphin // *PLoS One*. 2015. V. 10. № 4. P. 1–15. doi 10.1371/journal.pone.0123956
11. GenAlEx Tutorials [cited 2018 Sep 07]. Available from: <http://biology-assets.anu.edu.au/GenAlEx/Tutorials.html>
12. *Peakall R., Smouse P.E.* GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update // *Bioinformatics*. 2012. V. 28. № 19. P. 2537–2539. doi 10.1093/bioinformatics/bts460
13. *Peakall R., Smouse P.E.* GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research // *Mol. Ecol. Notes*. 2006. V. 6. № 1. P. 288–295. doi 10.1111/j.1471-8286.2005.01155.x
14. *Martins W.S., Lucas D.C.S., Neves K.F.S., Bertoli D.J.* WebSat – A Web Software for MicroSatellite Marker Development // *Bioinformation*. 2009. V. 3. № 6. P. 282–283.
15. Websat software (cited 2018 Sep 07). Available from: <http://wsmartins.net/websat/>.

16. GenBank: genome (cited 2018 Sep 07). Available from: <https://www.ncbi.nlm.nih.gov/genome/>.
17. *Kibbe W.A.* OligoCalc: An online oligonucleotide properties calculator // *Nucl. Acids Res.* 2007. V. 35. № Web Server Issue. P. W43–W46. doi 10.1093/nar/gkm234
18. OligoCalc software (cited 2018 Sep 07). Available from: <http://www.bio.bsu.by/molbiol/oligocalc.html>.
19. *Ivanova N.V., de Waard J., Hebert P.D.N.* An inexpensive, automation-friendly protocol for recovering high-quality DNA // *Mol. Ecol. Notes.* 2006. V. 6. № 4. 998–1002. doi 10.1111/j.1471-8286.2006.01428.x
20. CCDB DNA Extraction (cited 2018 Sep 07). Available from: http://ccdb.ca/site/wp-content/uploads/2016/09/CCDB_DNA_Extraction.pdf.
21. *Dou H., Yang H., Feng L. et al.* Estimating the population size and genetic diversity of Amur tigers in North-east China // *PLoS One.* 2016. V. 11. № 4. P. 1–14. doi 10.1371/journal.pone.0154254
22. *Reisser C.M.O., Fasel D., Hurlimann E. et al.* Transition from environmental to partial genetic sex determination in *Daphnia* through the evolution of a female-determining incipient W chromosome // *Mol. Biol. Evol.* 2017. V. 34. № 3. P. 575–588. doi 10.1093/molbev/msw251
23. GeneMarker software (cited 2018 Sep 07). Available from: <https://softgenetics.com/GeneMarker.php>.
24. Notepad++ software (cited 2018 Sep 07). Available from: <https://notepad-plus-plus.org>.
25. GenAlEx 6.502 software (cited 2018 Sep 07). Available from: <http://biology.anu.edu.au/GenAlEx/>.
26. *Everhart S.E., Kamvar Z.N., Grünwald N.J.* Genotypic richness, diversity, and evenness (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/Genotypic_EvenRichDiv.html.
27. *Nomura T.* Estimation of effective number of breeders from molecular coancestry of single cohort sample // *Evol. Appl.* 2008. V. 1. № 3. P. 462–474. doi 10.1111/j.1752-4571.2008.00015.x
28. *Do C., Waples R.S., Peel D. et al.* NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data // *Mol. Ecol. Resources.* 2014. V. 14. № 1. 209–214. doi 10.1111/1755-0998.12157
29. NeEstimator software (cited 2018 Sep 07). Available from: <http://www.molecularfisherieslaboratory.com.au/neestimator-software/>.
30. *Marshall T.C., Slate J., Kruuk L.E.B., Pemberton J.M.* Statistical confidence for likelihood-based paternity inference in natural populations // *Mol. Ecol.* 1998. V. 7. № 5. P. 639–655. doi 10.1046/j.1365-294x.1998.00374.x
31. Cervus software (cited 2018 Sep 07). Available from: http://www.fieldgenetics.com/pages/aboutCervus_Using.jsp.
32. *Butler K., Field C., Herbinger C.M., Smith B.R.* Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data // *Mol. Ecol.* 2004. V. 13. № 6. P. 1589–1600. doi 10.1111/j.1365-294X.2004.02152.x
33. Pedigree v. 2.0 software (cited 2018 Sep 07). Available from: <https://www.progenygenetics.com>.
34. Mstools software (cited 2018 Sep 07). Available from: <http://peterslab.org/MSTools/>.
35. *Valière N.* GIMLET: A computer program for analysing genetic individual identification data // *Mol. Ecol. Notes.* 2002. V. 2. № 3. P. 377–379. doi 10.1046/j.1471-8286.2002.00228.x-i2
36. GIMLET software (cited 2018 Sep 07). Available from: <http://pbil.univ-lyon1.fr/software/Gimlet/gimlet%20frame1.html>.
37. *Guillot G., Santos F.* A computer program to simulate multilocus genotype data with spatially auto-correlated allele frequencies // *Mol. Ecol. Resources.* 2009. V. 9. № 4. P. 1112–1120. doi 10.1111/j.1755-0998.2008.02496.x
38. Geneland software (cited 2018 Sep 07). Available from: <http://www2.imm.dtu.dk/~gigu/Geneland/>.
39. *Pritchard J.K., Stephens M., Donnelly P.* Inference of population structure using multilocus genotype data // *Genetics.* 2000. V. 155. № 2. P. 945–959.
40. STRUCTURE software (cited 2018 Sep 07). Available from: <https://web.stanford.edu/group/pritchard-lab/structure.html>.
41. *Evanno G., Regnaut S., Goudet J.* Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study // *Mol. Ecol.* 2005. V. 14. № 8. P. 2611–2620. doi 10.1111/j.1365-294X.2005.02553.x
42. *Wang J.* The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse // *Mol. Ecol. Resources.* 2016. V. 17. № 5. P. 981–990. doi 10.1111/1755-0998.12650
43. *Earl D.A., van Holdt B.M.* STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method // *Conserv. Genet. Resources.* 2012. V. 4. № 2. P. 359–361. doi 10.1007/s12686-011-9548-7
44. *Копылов-Гуськов Ю.О.* Анализ разнообразия перистых ковылей группы родства *Stipa dasycphylla* на территории европейской части России: Дис. ... канд. биол. наук. М.: МГУ им. М.В. Ломоносова, 2016. 209 с.
45. *Крамина Т.Е., Шанцер И.А.* Анализ спонтанной гибридизации двух диплоидных видов *Lotus* (Fabaceae) в Волгоградской области // *Бот. журн.* 2010. Т. 95. № 6. С. 93–106.
46. *Wandeler P., Smith S., Morin P.A. et al.* Patterns of nuclear DNA degeneration over time – a case study in historic teeth samples // *Mol. Ecol.* 2003. V. 12. № 4. P. 1087–1093. doi 10.1046/j.1365-294X.2003.01807.x
47. *Miller C.R., Waits L.P.* The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): Implications for conservation // *Proc. Natl Acad. Sci. USA.* 2003. V. 100. № 7. P. 4334–4339. doi 10.1073/pnas.0735531100
48. *Wattier R., Engel C.R., Saumitou-Laprade P., Valero M.* Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta) // *Mol. Ecol.* 1998. V. 7. № 11. P. 1569–1573. doi 10.1046/j.1365-294x.1998.00477.x

49. *Shinde D., Lai Y.L., Sun F.Z., Arnheim N.* Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites // *Nucl Acid Res.* 2003. V. 31. № 3. P. 974–980. doi 10.1093/nar/gkg178
50. *Shaw P.W., Pierce G.J., Boyle P.R.* Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers // *Mol. Ecol.* 1999. V. 8. № 3. P. 407–417. doi 10.1046/j.1365-294X.1999.00588.x
51. *Лу Ч.* Введение в популяционную генетику. М.: Мир, 1978. 555 с.
52. *Miller C.R., Joyce P., Waits L.P.* Assessing allelic dropout and genotyping reliability using maximum likelihood // *Genetics.* 2002. V. 160. № 1. P. 357–366.
53. *Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P.* MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data // *Mol. Ecol. Notes.* 2004. V. 4. № 3. P. 535–538. doi 10.1111/j.1471-8286.2004.00684.x
54. Micro-Checker software (cited 2018 Sep 07). Available from: <http://www.nrp.ac.uk/nrp-strategic-alliances/elsa/software/microchecker/>.
55. *Raymond M., Rousset F.* GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism // *J. Heredity.* 1995. V. 86. № 3. P. 248–249. doi 10.1093/oxfordjournals.jhered.a111573
56. Genepop software (cited 2018 Sep 07). Available from: <http://genepop.curtin.edu.au>.
57. *Excoffier L., Lischer H.E.L.* Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows // *Mol. Ecol. Resources.* 2010. V. 10. № 3. P. 564–567. doi 10.1111/j.1755-0998.2010.02847.x
58. Arlequin software (cited 2018 Sep 07). Available from: <http://cmpg.unibe.ch/software/arlequin35/>.
59. *Goudet J.* FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). 2001. Available from: <http://www2.unil.ch/popgen/softwares/fstat.htm>.
60. FSTAT software (cited 2018 Sep 07). Available from: <https://www2.unil.ch/popgen/softwares/fstat.htm>.
61. *Chapuis M.-P., Estoup A.* Microsatellite null alleles and estimation of population differentiation // *Mol. Biol. Evol.* 2007. V. 24. № 3. P. 621–631. doi 10.1093/molbev/msl191
62. *Ewen K.R., Bahlo M., Treloar S.A. et al.* Identification and analysis of error types in high-throughput genotyping // *Amer. J. Human Genet.* 2000. V. 67. № 3. P. 727–736. doi 10.1086/303048
63. Pedmanager software (cited 2018 Sep 07). Available from: <https://github.com/gaow/genetic-analysis-software/blob/master/pages/PEDMANAGER.md>.
64. *Summers K., Amos W.* Behavioral, ecological and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus* // *Behavioral Ecol.* 1997. V. 8. № 3. P. 260–267. doi 10.1093/beheco/8.3.260
65. *Kamvar Z.N., Everhart S.E., Grünwald N.J.* Locus stats, heterozygosity, HWE. Подзаголовок “Hardy–Weinberg equilibrium” (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/Locus_Stats.html.
66. The R Project for Statistical Computing (cited 2018 Sep 07). Available from: <https://www.r-project.org/>.
67. The Comprehensive R Archive Network (cited 2018 Sep 07). Available from: <https://cran.r-project.org/>.
68. R/(programming/ language. [https://en.wikipedia.org/wiki/R/\(programming/_language](https://en.wikipedia.org/wiki/R/(programming/_language).
69. RStudio is an integrated development environment (IDE) for R (cited 2018 Sep 07). Available from: <https://github.com/rstudio/rstudio>.
70. Population genetics and genomics in R (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/index.html.
71. *Belkhir K., Borsa P., Chikhi L. et al.* GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. 2004. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
72. GENETIX software (cited 2018 Sep 07). Available from: <http://kimura.univ-montp2.fr/genetix/>.
73. *Chapuis M.-P., Lecoq M., Michalakis Y. et al.* Do out-breaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles // *Mol. Ecol.* 2008. V. 17. № 16. P. 3640–3653. doi 10.1111/j.1365-294X.2008.03869.x
74. FreeNa software (cited 2018 Sep 07). Available from: <http://www1.montpellier.inra.fr/CBGP/software/FreeNA/>.
75. *Kalinowski S.T., Taper M.L., Marshall T.C.* Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment // *Mol. Ecol.* 2007. V. 16. № 5. P. 1099–1106. doi 10.1111/j.1365-294X.2007.03089.x
76. ML-NullFreq software (cited 2018 Sep 07). Available from: <http://www.montana.edu/kalinowski/software/null-freq.html>.
77. Detecting Natural Selection (Part 7) (cited 2018 Sep 07). Available from: <http://scienceblogs.com/evolution/2006/01/14/detecting-natural-selection-pa-1/>.
78. *Хаубольд Б., Вие Т.* Введение в вычислительную биологию: эволюционный подход. М.: Ижевск: НИЦ “Регулярная и хаотическая динамика”, Ижевский ин-т компьютер. исслед., 2011. 456 с.
79. *Харзинова В.Р., Гладырь Е.А., Степанов А.И. и др.* Разработка мультиплексной панели микросателлитов для оценки достоверности происхождения и степени дифференциации северного оленя *Rangifer tarandus* // *С.-х. биология.* 2015. Т. 50. № 6. С. 756–765. doi 10.15389/agrobiology.2015.6.756rus
80. *Wright S.* Isolation by distance under diverse systems of mating // *Genetics.* 1946. V. 31. № 1. P. 39–59.
81. *Wright S.* The genetical structure of populations // *Ann. Eugen.* 1949. V. 15. № 1. P. 323–354. doi 10.1111/j.1469-1809.1949.tb02451.x
82. *Wright S.* The interpretation of population structure by F-Statistics with special regard to systems of mating // *Evolution.* 1965. V. 19. № 3. P. 395–420. doi 10.2307/2406450

83. *Wright S.* Evolution and the Genetics of Populations. Variability Within and Among Natural Populations. V. 4. Chicago: Univ. Chicago Press, 1978. 590 p.
84. *Hedrick P.W.* Genetics of Populations. Boston: Jones and Bartlett, 2000. 675 p.
85. *Shannon C.E.* A mathematical theory of communication // Bell System Technical J. 1948. V. 27. № 3. P. 379–423, 623–656. doi 10.1002/j.1538-7305.1948.tb01338.x
86. *Шутиков В.К., Розенберг Г.С., Зинченко Т.Д.* Количественная гидроэкология: методы системной идентификации. Тольятти: ИЭВБРАН, 2003. 463 с.
87. *Nei M.* Genetic distance between populations // Amer. Naturalist. 1972. V. 106. № 949. P. 283–392. doi 10.1086/282771
88. *Nei M.* Estimation of average heterozygosity and genetic distance from a small number of individuals // Genetics. 1978. V. 89. № 3. P. 583–590.
89. Population structure: G_{ST} , genetic distance, and clustering. Подзаголовок “Genetic Distance” (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/Pop_Structure.html.
90. *Kamvar Z.N., Everhart S.E., Grünwald N.J.* Locus stats, heterozygosity, HWE. Подзаголовок “Locus summary statistics” (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/Locus_Stats.html.
91. PopGene software (cited 2018 Sep 07). Available from: <https://sites.ualberta.ca/~fyeh/popgene.html>.
92. GenStat software (cited 2018 Sep 07). Available from: <https://www.vsni.co.uk/software/genstat/>.
93. Should I use F_{ST} , G_{ST} or D ? Posted on 2 March, 2011 by Nolan Kane (cited 2018 Sep 07). Available from: <http://www.molecularecologist.com/2011/03/should-i-use-fst-gst-or-d-2/>.
94. *Meirmans P.G., Hedrick P.W.* Assessing population structure: F_{ST} and related measures // Mol. Ecol. Resources. 2011. V. 11. № 1. P. 5–18. doi 10.1111/j.1755-0998.2010.02927.x
95. Population structure: G_{ST} , genetic distance, and clustering. Подзаголовок “ G_{ST} an example with *Felis catus* data” (cited 2018 Sep 07). Available from: https://grunwaldlab.github.io/Population_Genetics_in_R/Pop_Structure.html.
96. *Rousset F.* Genepop ‘007: a complete re-implementation of the genepop software for Windows and Linux // Mol. Ecol. Resources. 2008. V. 8. № 1. P. 103–106. doi 10.1111/j.1471-8286.2007.01931.x
97. *Zhdanova O., Pudovkin A.I.* Nb_HetEx: A Program to Estimate the Effective Number of Breeders // J. Heredity. 2008. V. 99. № 6. P. 694–695. doi 10.1093/jhered/esn061
98. *Waples R.S., Do C.* Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution // Evol. Appl. 2010. V. 3. № 3. P. 244–262. doi 10.1111/j.1752-4571.2009.00104.x
99. *Kimura M.* Evolutionary rate at the molecular level // Nature. 1968. V. 217. P. 624–626. doi 10.1038/217624a0
100. *King J.L., Jukes T.H.* Non-Darwinian evolution // Science. 1969. V. 164. № 3881. P. 788–798. doi 10.1126/science.164.3881.788
101. *Futuyma D.J.* Evolution. Sunderland (Massachusetts): Sinauer Associates, 2005. 603 p.
102. *Waples R.S., Antao T., Luikart G.* Effects of overlapping generations on linkage disequilibrium estimates of effective population size // Genetics. 2014. V. 197. № 2. P. 769–780.
103. *Waples R.S., Do C.* LDNE: a program for estimating effective population size from data on linkage disequilibrium // Mol. Ecol. Resources. 2008. V. 8. № 4. P. 753–756. doi 10.1111/j.1755-0998.2007.02061.x
104. LDNe software (cited 2018 Sep 07). Available from: <https://conserver.iugo-cafe.org/user/Robin%20Waples/LDNe>.
105. *Heim M., Tallmon D., Boucher C. et al.* ONeSAMP 2.0 computes the effective population size of gene data sets. Copyright. 2016.
106. OneSamp software (cited 2018 Sep 07). Available from: <https://github.com/kingufl/ONeSAMP>.
107. *Piry S., Luikart G., Cornuet J.M.* Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data // J. Heredity. 1999. V. 90. № 4. P. 502–503. doi 10.1093/jhered/90.4.502
108. Bottleneck software (cited 2018 Sep 07). Available from: <http://www1.montpellier.inra.fr/CBGP/software/Bottleneck/bottleneck.html>.
109. *Wilson G.A., Rannala B.* Bayesian inference of recent migration rates using multilocus genotypes // Genetics. 2003. V. 163. № 3. P. 1177–1191.
110. BayesAss software (cited 2018 Sep 07). Available from: <http://www.rannala.org/inference-of-recent-migration/>.
111. *Piry S., Alapetite A., Cornuet J.-M. et al.* GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection // J. Heredity. 2004. V. 95. № 6. P. 536–539. doi 10.1093/jhered/esh074
112. GeneClass software (cited 2018 Sep 07). Available from: <http://www1.montpellier.inra.fr/CBGP/software/GeneClass/>.
113. *Alcala N., Goudet J., Vuilleumier S.* On the transition of genetic differentiation from isolation to panmixia: What we can learn from G_{ST} and D // Theor. Populat. Biol. 2014. V. 93. P. 75–84. doi 10.1016/j.tpb.2014.02.003
114. *Sundqvist L., Zackrisson M., Kleinhans D.* A new approach to estimate directional genetic differentiation and asymmetric migration patterns // Ecol. Evol. 2013. V. 6. № 11. P. 3461–3475. doi 10.1002/ece3.2096
115. *Avise J.C.* Phylogeography: The History and Formation of Species. Cambridge, MA: Harvard Univ. Press, 2000. 447 p.
116. *Lee M.D., Wagenmakers E.-J.* Bayesian Cognitive Modeling: A Practical Course. Cambridge University Press, 2013. 280 p.
117. *Gelman A., Carlin J.B., Stern H.S. et al.* Bayesian Data Analysis. Chapman and Hall/CRC, 2013. 675 p.

118. *Kruschke J.K.* Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Burlington: Acad. Press, 2010. 672 p.
119. *Levy E., Kennington W.J., Tomkins J.L., LeBas N.R.* Phylogeography and Population Genetic Structure of the Ornate Dragon Lizard, *Ctenophorus ornatus* // PLoS One. 2012. V. 7. № 10. e46351. doi 10.1371/journal.pone.0046351
120. *Дьякова А.В.* Изучение популяций *Lucilia sericata* (Meigen, 1826) (Diptera, Calliphoridae) на территории России: Автореф. дис. ... магистра биол. наук. М.: МГУ им. М.В. Ломоносова, 2017. 67 с.

Prejudices Against Microsatellite Studies and How to Resist Them

T. V. Galinskaya^{a, b, *}, D. M. Schepetov^{a, c, d}, and S. N. Lysenkov^e

^aDepartment of Entomology, Lomonosov Moscow State University, Moscow, 119234 Russia

^bAll-Russian Plant Quarantine Center, Moscow oblast, Vykovo, 140150 Russia

^cKoltzov Institute of Developmental Biology Russian Academy of Sciences, Moscow, 119334 Russia

^dCentre of Fundamental Research, National Research University Higher School of Economics, Moscow, 101000 Russia

^eDepartment of Biological Evolution, Lomonosov Moscow State University, Moscow, 119234 Russia

*e-mail: nuha1313@gmail.com

In this article the basic concepts and techniques used in microsatellite studies are summarized. The main stages of the work are briefly described, references to the most frequently used programs for proceeding microsatellite data are given. The article also briefly analyzes biological meaning of the indices most often encountered in microsatellite studies.

Keywords: populations, evolution, microsatellites, quarantine species, population genetics, software.