МИКРОСАТЕЛЛИТНАЯ ИЗМЕНЧИВОСТЬ В ПОПУЛЯЦИЯХ ЧЕЛОВЕКА И МЕТОДЫ ЕЕ ИЗУЧЕНИЯ

Л.А. Животовский

Институт общей генетики им. Н.И. Вавилова РАН, Москва, Россия,
e-mail: levazh@gmail.com

Статья представляет собой обзор по микросателлитной изменчивости в популяциях челоека. Рассмотрены принципы выявления микросателлитных аллелей и их номенклатуры, особенности мутационного процесса по микросателлитным локусам. Даны общие сведения об эволюции микросателлитной изменчивости. Приведены некоторые статистические методы, основанные на данных о микросателлитном разнообразии, в том числе, при оценке возраста гаплогрупп Y-хромосомы и времени дивергенции популяций.

Микросателлиты – особый класс ДНК-маркеров. Они представляют собой фрагменты ДНК с большим количеством – до сотни и даже выше – тандемно повторяющихся идентичных «мотивов», обычно называемых «повторами»: короткими последовательностями из нескольких (как обычно принято считать – от одной до шести) пар нуклеотидов (Tautz, 1993). Микросателлиты высокополиморфны, с десятками аллелей в каждом локусе и высокими темпами мутирования (Levinson, Gutman, 1987; Jeffreys *et al*., 1988; Kelley *et al*., 1991; Henderson, Petes, 1992). Аллели микросателлитного локуса отличаются друг от друга длиной, в основном числом повторов. Небольшие размеры микросателлитных локусов позволяют применить метод полимеразной цепной реакции в целях генотипирования и обеспечить высокую воспроизводимость результатов в различных лабораториях (Weber, May, 1989; Weber, Wong, 1993). Для типирования микросателлитов требуется небольшое количество ДНК, которую можно экстрагировать даже из сильно деградированного биологического материала. Микросателлиты встречаются в большом количестве у всех эукариотических организмов (Kashi *et al*., 1990) и сейчас используются для изучения как природных популяций, так и популяций сельскохозяйственных животных и растений. У человека они распределены по всему геному, что позволяет использовать их, как и мононуклеотидные полиморфизмы (SNP), для анализа ассоциаций, сцепления и картирования генов (Holmans, 2001; Ott, 1999; Kong *et al*., 2002), в качестве маркеров наследственных заболеваний (Ashley, Warren, 1995). Значительный полиморфизм микросателлитов обеспечивает возможность с высокой вероятностью идентифицировать личность и выявить биологическое родство (Evett, Weir, 1998). Высокие темпы мутирования в этих локусах приводят к накоплению популяционно-специфических мутаций, что позволяет проводить детальный анализ популяционной структуры (Bowcock *et al*., 1994; Jorde *et al*., 2000; Rosenberg *et al*., 2002; Zhivotovsky *et al*., 2003). Данная статья – это обзор основных известных на сегодня особенностей микросателлитной изменчивости в популяциях человека и некоторых методов ее анализа.

Микросателлитные аллели и их определение

Микросателлиты буквально рассыпаны по геному человека: число их достигает нескольких десятков тысяч. В качестве примера на рис. 1 представлено распределение охарактеризованных микросателлитных локусов по хромосоме 15 человека; из него видно, что микросателлиты встречаются во всех сегментах хромосомы. К настоящему времени у человека исследовано около десятка тысяч мик-

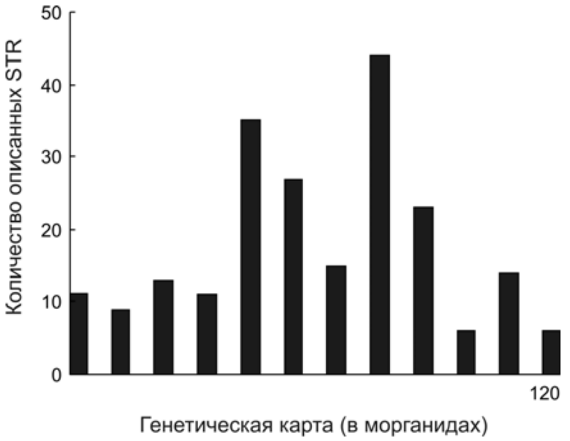


Рис. 1. Распределение описанных микросателлитных локусов в хромосоме 15 (по данным Marshfield La).
Генетическая карта построена по средней частоте рекомбинации у обоих полов.

росателлитных локусов (Dib *et al.*, 1996; Weber, Broman, 2001; The Mammalian Genotyping Service: http://research.marshfieldclinic.org/genetics/sets/combo.html).

В зависимости от длины повтора микросателлиты классифицируют на локусы с моно-, ди-, три-, тетра-, пента-, и гексануклеотидными повторами. Например, с-протоонкоген вируса саркомы кошки fes/fps, локализованный в хромосоме 15 человека, содержит тетрануклеотидные повторы АТТТ в одном из интронов этого гена. Они и определяют микросателлитный локус, обозначаемый как *FES/FPS* (рис. 2). Микросателлитные локусы высокополиморфны – т. е. для каждого из них имеется много аллелей. Например, в популяциях обнаружено до десятка аллелей локуса *FES/FPS* (с числом повторов от 7 до 15) и нередки локусы с гораздо бóльшим числом аллелей (рис. 3).

Микросателлитные фрагменты выявляют методом полимеразной цепной реакции (ПЦР), обеспечивающим многократное увеличение числа копий (амплификацию) данного фрагмента ДНК. Амплификация достигается в ходе циклического (30–40 циклов) комплементарного достраивания ДНК-матрицы с помощью термостабильной *Taq*-полимеразы. Цикличность обеспечивается программируемыми термостатами (амплификаторами) с автоматической сменой температур режимов плавления ДНК и ее репликации. Синтез данного фрагмента ДНК инициируется ДНК-затравками в виде пары *праймеров* – синтетических олигонуклеотидов, комплементарных нуклеотидным последовательностям на границах исследуемого фрагмента. Поскольку микросателлитные аллели короткие и вместе с праймерами обычно не превышают 200–300 п.н., то даже сильно деградированный биологический материал может содержать полноразмерные интактные копии исследуемого фрагмента ДНК, обеспечивая высокую вероятность их успешной амплификации. Именно по этой причине ПЦР небольших участков ДНК, в том числе микросателлитов, оказался особенно важным для судебно-медицинских исследований и для анализа древних костных остатков (правда, более успешным в этих случаях пока является анализ митохондриальной ДНК (мтДНК) из-за высокой копийности молекул

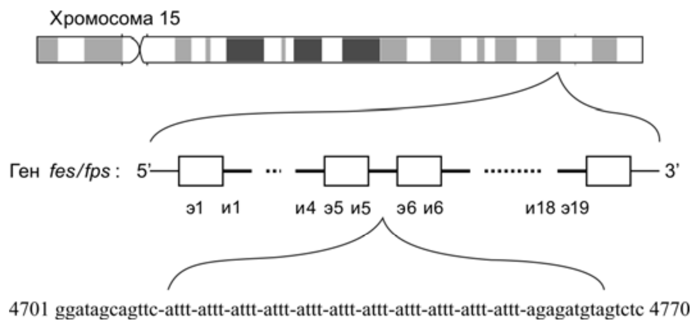


Рис. 2. Микросателлитный локус *FES/FPS*.
Локус *FES/FPS*, состоящий из серии тетрануклеотидных повторов АТТТ, содержится в интроне с-протоонкогена *fes/fps*. На рисунке дана нуклеотидная последовательность между позициями 4701 и 4770, где располагаются повторы АТТТ (полная длина гена *fes/fps* превышает 12 тыс. пар нуклеотидов). Данный *FES/FPS*-аллель имеет 11 повторов.

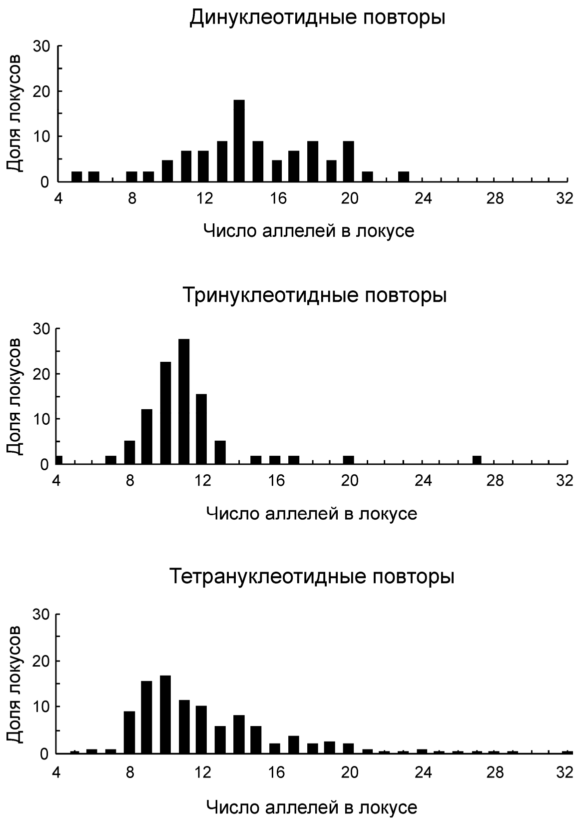


Рис. 3. Число аллелей в аутосомных микросателлитных локусах с ди-, три- и тетрануклеотидными повторами, обнаруженных в выборке более чем 1 тыс. индивидов из 52 этнических групп основных регионов мира (Zhivotovsky *et al.*, 2003).

Видно, что подавляющее число локусов имеют 8 и более аллелей. Локусы с динуклеотидными повторами в среднем более вариабельны, чем локусы с более длинными мотивами.

мтДНК в митохондриях). Для исследуемого микросателлитного локуса подбирают такую пару праймеров, чтобы комплементарные им концевые (фланкирующие) участки ДНК были высокоспецифичны – отсутствовали в других участках генома человека или других организмов, способных «засорить» исследуемую ДНК, – и обеспечивали надежность амплификации. Для этого они должны быть достаточной длины (20–30 п.н.), их 3′-концы не должны быть комплементарными друг другу и не должны содержать C/G-трактов. Кроме того, они должны быть сбалансированными по содержанию нуклеотидов для равенства температур плавления ДНК. В качестве примера на рис. 4 показана одна из используемых пар праймеров для локуса *FES/FPS*. Были разработаны эффективные методы анализа микросателлитов с использованием праймеров, меченных флуоресцентными красителями, с последующей детекцией продуктов реакции с помощью автоматических секвенаторов ДНК (Ziegle *et al*., 1992).

Уточним определение понятий микросателлитного локуса и микросателлитных аллелей. Участок ДНК с определенной геномной локализацией, содержащий короткие тандемные повторы, называют *микросателлитным локусом* (или *микросателлитом*), часто *STR-локусом* (или просто *STR* – от английского Short Tandem Repeats), или же *SSR* – Simple Sequence Repeats. *Аллель* STR-локуса – это окаймленный данной парой праймеров фрагмент ДНК, содержащий определенное число исследуемых повторов. Поскольку выбор праймеров во многом произволен, то термины «локус» и «аллель» здесь имеют чисто символический характер, указывая лишь на то, что этот локус содержит участок ДНК с исследуемыми повторами. Можно выбрать иную пару праймеров, окаймляющую этот же фрагмент ДНК с повторами, – и это будет тот же STR-локус, если целью исследования является выявление числа или нуклеотидного содержания тех же самых повторов в данном участке ДНК. При этом соответствующие аллели, выявляемые разными парами праймеров, отличаются друг от друга на одно и то же число нуклеотидов, определяемое только различием в положении концевых участков амплифицируемых фрагментов ДНК. Независимо от того, определяют ли разные пары праймеров бóльшие или

[4621] 5′… TTGGCCTGGAGCAGCTGGAAGATGGAGTGGCTGTTA ATTCATGTAGGGAAGGCTGTGGGA
AGAAGAGGTTTAGGAGACAAGGATAGCAGTTC**ATTTATTTATTTATTTATTTATTTATTT**
ATTTATTTATTTATTTAGAGATGTAGTCTC ATTCTTTCGCCAGGCTGGAG TGCAGTGGCG …3′ [4800]

Рис. 4. Фрагмент ДНК, содержащий ATTT-повторы локуса *FES/FPS* и фланкирующие участки.

На рисунке показан фрагмент между позициями 4621 и 4800 с-протоонкогена *fes/fps* (см. рис. 1). 11 тетрануклеотидных повторов ATTT выделены жирным шрифтом. Участки для одной из используемых пар праймеров подчеркнуты.

меньшие амплифицированные фрагменты ДНК, важно, чтобы эти фрагменты содержали весь участок с исследуемыми повторами, которые и являются основной характеристикой микросателлитного локуса. Поэтому формула локуса обычно указывает на структуру повторов и их число. Так, формула STR-локуса *FES/FPS* записывается как $[ATTT]_n$, где n – число повторов; аллель в этом локусе (приведенный на рис. 2), обозначается как $[ATTT]_{11}$ – в нем содержится 11 повторов ATTT.

В особо важных случаях – например, в судебно-медицинской экспертизе – генотипирование (т. е. выявление аллелей STR-локуса у исследуемых индивидов) проводят по двум или даже трем парам праймеров. Делают это потому, что не исключена мутация в районе праймера (либо новая, возникшая у данного индивида, либо возникшая раньше и достигшая определенной частоты в популяции), ассоциированная с определенным аллелем, при этом амплификация этого аллеля невозможна – он ведет себя как необнаруживаемый «нуль-аллель». Указанием на присутствие такой мутации у индивида является его гомозиготность по данному аутосомному микросателлитному локусу, которая оказывается в таком случае ложной, поскольку «нуль-аллель» не амплифицируется. Как правило, микросателлиты высокополиморфны и потому подавляющее большинство людей гетерозиготны по ним; лишь небольшой процент индивидов – реальные гомозиготы, имеющие идентичные по нуклеотидному составу аллели в обоих гомологичных микросателлитных локусах. Использование другой пары праймеров, не включающей эту мутацию, позволяет провести ПЦР и определить реальный генотип индивида по данному микросателлитному локусу. Конечно, микросателлитные локусы гаплоидных тканей и органов или гемизиготной части генома (например, Y-хромосомы) не показывают гетерозиготного генотипа, хотя нуль-аллели могут встречаться и по ним; в этом случае они легко обнаруживаются по отсутствию продукта амплификации. Разработаны так называемые мультиплексные реакции, позволяющие проводить одновременную амплификацию и фрагментный анализ по нескольким локусам; они различимы на фореграмме окраской и положением, обеспечиваемыми флуоресцентными праймерами и неперекрывающейся вариацией размеров аллелей разных локусов.

Номенклатура микросателлитных локусов не имеет единого стандарта. Название STR-локуса может быть связано с тем геном, в котором он находится (например, *FES/FPS*), а может отражать хромосомную локализацию и определенный номер (например, *D11S1986* в хромосоме 11 и *DYS392* в Y-хромосоме). Даже обозначение повтора может быть разным: иметь различные варианты перестановок или указывать на разные комплементарные цепи ДНК. Например, на рис. 3 видно, что, поскольку участок с тетрануклеотидными повторами окаймлён нуклеотидами A, повторы в локусе *FES/FPS* могут быть обозначены не только как ATTT, но и как TTTA, а исходя из нуклеотидной последовательности комплементарной цепи – еще и как AAAT или TAAA.

Структура микросателлитных аллелей может быть разной сложности и иметь вариации как по участку с повторами, так и во фланкирующих последовательностях (см. классификацию: Urquhart *et al*., 1994; Gill *et al*., 1995). Например, все аллели локуса *FES/FPS* имеют только так называемые совершенные, или простые повторы: их регулярность ничем не прерывается. В то же время в этом конкретном локусе встречаются вариации, которые отличаются от стандартных аллелей заменой нуклеотида A на нуклеотид C в одной из позиций 5′-фланкирующего района и выявляются прямым секвенированием или рестрикционным анализом. Локус *HUMTH01* (первый интрон гена тирозингидроксилазы) имеет большинство аллелей с простыми повторами $[AATG]_n$, но некоторые его аллели имеют более сложную структуру, как, например, $[AATG]_5ATG[AATG]_3$ или $[AATG]_6ATG[AATG]_4$. Эти аллели содержат между основными тетрануклеотидными повторами тринуклеотид ATG, видимо, путем делеции возникший из основного повтора. Вариации в структуре повторов в этом локусе могут образоваться и за счёт замены, встраивания или выпадения отдельного нуклеотида, например, в аллеле $[AATG][AACT][AATG]_8ATG[AATG]_3$ участок ATG образовался из повтора AATG вследствие потери нуклеотида A. Наконец, встречаются сложные аллели, составленные из двух и более

5'..CCAACTCTCATCTGTATTATCTATG TGTG*TCTGTCTGTCTGTCTGTCTGTCTAT*
*CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA*TCATTATACCTACTTCTGTA
T CCAACTCTCATCTGTATTATCTATG TATCTGTCTGTCTGTCTAT *CTATCTATCTAT*
*CTATCTATCTATCTATCTATCTATCTATCCCTCCCTCTATCAATCTATCTATTTATC*TA
GCAGTCCATCATCTATCTATGACATTCTTCTGCTACTCAGGGATAATTGTGTTCCTCA
AGTAACACTTGGCAATGTCTGGAAACAATTTTGGCAGTTGCACTGGGATGGGTGTTC
TCGCATCTGGTGGGTGGAGATAAGA..3'

Рис. 5. Структура локуса *DYS389I/II*.
Для анализа двух частей этого локуса, в каждом из которых наблюдается вариабельное число повторов, используют как внешние, так и внутренний праймеры.

типов повторов. Например, один из аллелей локуса *vWF*, используемого в судебно-генетических исследованиях, имеет подобную сложную структуру: $[TCTA]_2[TCTG]_4$ $[TCTA]_3TCCA[TCTA]_3$. Такие сложные локусы нередко разбивают на части, соответствующие своим группам повторов, которые затем анализируют раздельно с использованием как окаймляющих, так и внутренних праймеров (рис. 5).

Обычно длина аллеля, выраженная в числе нуклеотидных пар (без определения числа повторов), достаточна для сравнительного анализа. Однако для эволюционных оценок и датировок нередко необходима информация о размере аллеля, выраженная именно в числе повторов. В таком случае необходимо знать полную нуклеотидную последовательность аллеля определенного эталонного образца или использовать в каждой серии генотипирования образцов так называемые лэддеры – фрагменты ДНК аллелей с известным числом повторов. Это позволяет рассчитать размер аллеля в терминах числа повторов.

Мутационный процесс по микросателлитным локусам

Подавляющая доля мутаций в микросателлитных локусах возникает за счет специфической ошибки репликации ДНК в районе микросателлита – проскальзывания (англ. slippage) ДНК-полимеразы вдоль гомополимерной последовательности на число нуклеотидов, кратное длине повтора. Темпы возникновения микросателлитных мутаций (μ) гораздо выше, чем частота точковых мутаций у эукариот: если скорость возникновения последних – порядка величин 10^{-9}–10^{-8} на нуклеотид и порядка 10^{-6} на ген, то для изменения числа повторов она гораздо выше: от 10^{-6} до 10^{-2} (Schlötterer, 2000). Именно по этой причине вариабельность STR-локусов характеризуется числом повторов, а основным методом генотипирования является фрагментный анализ, позволяющий различать размеры аллеля. На рис. 6 приведена родословная, в которой выявлена вновь возникшая мутация по одному из микросателлитных локусов Y-хромосомы.

Если микросателлит находится в кодирующем участке генома, то его мутации либо летальны, если они прерывают синтез важного белка, либо кодируют дефектные полипептидные цепи. Последнее обычно связано с тринуклеотидными повторами, поскольку они не сдвигают рамку считывания, а лишь уменьшают, а чаще увеличивают длину кодируемой этим геном полипептидной цепи. Например, хорея Гентингтона, вызывающая поражение центральной нервной системы, обусловлена дисфункцией белка гентингтина вследствие увеличения в нем полиглютаминовой зоны. Это связано с ростом копийности триплетных повторов CAG до 40 ÷ 100 в гене, который в норме имеет менее 30 повторов. Другое заболевание – фрагильный X-синдром – обусловлено возрастанием числа CGG-повторов: от нормы, менее 50, до многих сотен копий у больных (Ashley, Warren, 1995).

Анализ более 5 тысяч аутосомных микросателлитов с диаллельными повторами показал, что средняя частота их возникновения $\mu = 6{,}2 \times 10^{-4}$ на локус на поколение (Dib *et al.*, 1996). Темпы мутирования в локусах с три- и тетрануклеотидными повторами ниже (Chakraborty *et al.*, 1997). Основная часть микросателлитных мутаций изменяет количество повторов как в сторону увеличения, так и в сторону уменьшения ровно на единицу. Однако часть мутантных аллелей может терять или приобретать по два и даже более

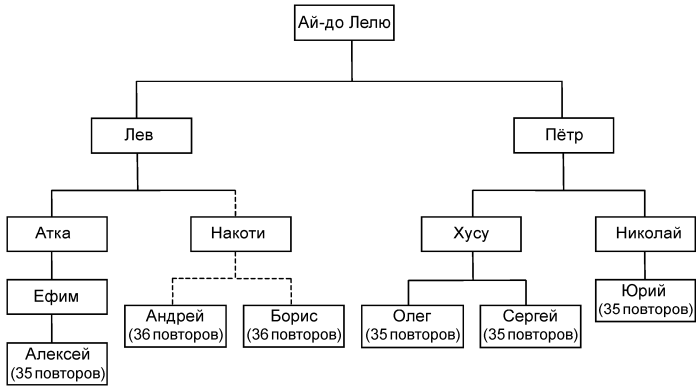


Рис. 6. Пример вновь возникшей мутации в STR-локусе Y-хромосомы. (По данным, предоставленным автору Л.П. Осиповой, Т.М. Карафет и М. Хаммером).

Родословная мужчин в семье лесных ненцев, имеющих гаплогруппу N-P63. У типированных индивидов указано число повторов по локусу CDY-1. По остальным исследованным локусам все типированные индивиды в этой семье имели идентичные аллели. Поскольку возникновение двух мутаций маловероятно, то данные можно интерпретировать так, что родоначальник семьи Ай-до Лелю и его сыновья Лев и Петр имели одинаковые Y-гаплотипы с аллелем из 35 повторов по локусу CDY-1. Этот аллель мутировал к 36 повторам у сына Льва – Накоти, который передал его своим сыновьям Андрею и Борису; пунктиром показан путь этого аллеля.

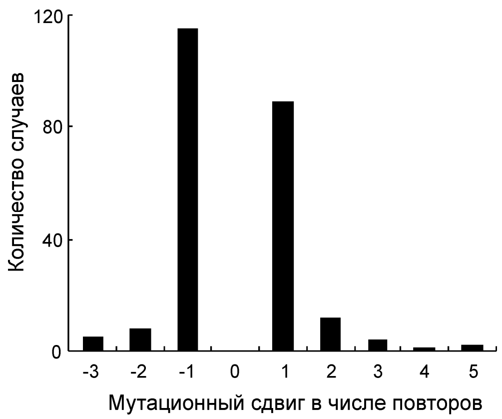


Рис. 7. Пример спектра мутаций в микросателлитных локусах (по данным Xu *et al*., 2000).

Изучено 287786 локус-трансмиссий «родитель/потомок» в полных семьях по 273 микросателлитам с тетрануклеотидными повторами. Среди них было обнаружено 508 менделевских несоответствий, из которых повторным генотипированием были надежно установлены 236 мутаций в половых клетках. Как видно на представленном рисунке, в большинстве этих мутаций количество повторов изменяется ровно на единицу, однако нередки мутационные изменения на два и более повторов (таких мутаций было 16,4 %).

Мутационная дисперсия $\sigma_m^2 = \sum_{i=-3}^{5} n_i i^2 \Big/ n$ равна 1,83,

где n – число исследованных мутаций (в нашем случае – 236), i – мутационный сдвиг в числе повторов (в нашем случае – от –3 до +5), а n_i – число мутаций с изменением числа повторов на величину i.

повторов (рис. 7). Как следует из рисунка 7, дополнительным параметром мутационного процесса по микросателлитным локусам, помимо темпов мутирования – μ, является разброс мутационных изменений в числе повторов (мутационная дисперсия) – σ_m^2, который определяется как $\sigma_m^2 = \frac{1}{n}\sum_i n_i i^2$, где n – число исследованных мутационных событий, i – мутационный сдвиг в числе повторов, а n_i – число мутаций с изменением числа повторов на величину i. При симметричных мутациях, когда средние значения мутационных сдвигов в минус- и плюс-стороны одинаковы (т. е. $\frac{1}{n}\sum_{i=-3}^{5} n_i i = 0$, σ_m^2) – это обычная дисперсия. Если мутации изменяют размер аллеля ровно на единицу, тогда $\sigma_m^2 = 1$; при бóльших мутационных изменениях (как, например, на рис. 7) $\sigma_m^2 > 1$.

В эволюционных исследованиях базовым является произведение этих двух параметров (μ и σ_m^2): $w = \mu\sigma_m^2$ (Slatkin, 1995; Zhivotovsky, Feldman, 1995; Di Rienzo *et al*., 1998), которое было названо *эффективным темпом мутирования* (effective mutation rate: Zhivotovsky, 1999; Zhivotovsky *et al*., 2000). Связано это с тем, что чем больше возникает мутантных аллелей, отстоящих от родительского аллеля на несколько повторов (т. е. чем больше величина σ_m^2), тем быстрее нарастает общая изменчивость популяций и тем выше темпы дифференциации популяций по числу повторов. Анализ данных о более чем пяти тысячах аутосомных микросателлитных локусов с динуклеотидными повторами (Dib *et al*., 1996) позволил оценить мутационную дисперсию σ_m^2: для этих данных она оказалась равной порядка 2,5 (Feldman *et al*., 1999), что отвечает эффективному темпу мутирования в среднем на динуклеотидный локус $w = 1{,}52 \times 10^{-3}$ (Zhivotovsky *et al*., 2000). Сопоставление популяционной изменчивости по аутосомным динуклеотидным микросателлитным локусам с изменчивостью по аутосомным локусам с три- и тетрануклеотидными повторами у тех же индивидов позволило оценить средние эффективные темпы мутирования для три- и тетрануклеотидных маркеров. Для двух различных наборов локусов и различных популяционных выборок оценки эффективных темпов мутирования в аутосомных микросателлитных локусах с три- и тетрануклеотидными повторами оказались несколько различающимися: $0{,}85 \times 10^{-3}$ и $0{,}93 \times 10^{-3}$ (Zhivotovsky *et al*., 2000) и $0{,}71 \times 10^{-3}$ и $0{,}70 \times 10^{-3}$ (Zhivotovsky *et al*., 2003).

Темпы мутирования могут значительно варьировать от локуса к локусу (Di Rienzo *et al*., 1998; Cooper *et al*., 1999; Zhivotovsky *et al*., 2001). Более того, аллели разной длины в одном и том же локусе также могут отличаться по темпам мутирования. Например, более длинные аллели могут чаще мутировать, подчас предпочтительно к аллелям с меньшим числом повторов (Zhang *et al*., 1994; Xu *et al*., 2000; Huang *et al*., 2002; Dupuy *et al*., 2004), что может приводить к росту дисперсии по числу повторов с увеличением среднего размера локуса (Kayser *et al*., 2004). Возможность того, что аллели большей длины могут предпочтительно мутировать в короткие аллели, и наоборот, короткие аллели – в длинные, может являться эволюционным механизмом, удерживающим изменчивость по микросателлитным локусам в определенных пределах (Garza *et al*., 1995; Nauta, Weissing, 1996; Feldman *et al*., 1997; Zhivotovsky *et al*., 1997; Xu *et al*., 2000). Отбор против индивидов-носителей аллелей с большим числом повторов, например, ассоциированных с тяжелыми заболеваниями, также может ограничивать микросателлитную изменчивость.

Особо следует сказать о мутационном процессе по микросателлитным локусам Y-хромосомы. По локусам с три- и тетрануклеотидными повторами, наиболее часто используемым для филогеографического и популяционного анализа, средние скорости возникновения новых мутаций – $(2{,}0 \div 3{,}0) \times 10^{-3}$ на локус на поколение (Heyer *et al*., 1997; Kayser *et al*., 2000; Weale *et al*., 2001). Эти «трансмиссионные», или «семейные» оценки получены по парам «отец–сын» и генеалогическим данным. Они значительно превышают «эволюционные» оценки темпов мутирования по данным о микросателлитных вариациях в медианной сети – $0{,}26 \times 10^{-3}$ на 20 лет или $0{,}33 \times 10^{-3}$ на 25 лет (Forster *et al*., 2000) и по сводным данным (о микросателлитной изменчивости в популяциях с документированной историей и сравнительной изменчивости микросателлитных локусов Y-хромосомы и аутосом) – $0{,}69 \times 10^{-3}$ на локус на 25 лет (Zhivotovsky *et al*., 2004). Следует подчеркнуть, что данные оценки – это усредненные по локусам скорости возникновения мутаций; между микросателлитными локусами Y-хромосомы различия в темпах мутирования значительны: стандартное отклонение составляет $0{,}57 \times 10^{-3}$ (Zhivotovsky *et al*., 2004). Столь большое различие между «трансмиссионной» и «эволюционной» оценками вызваны, вероятно, тем, что большое число вновь возникших мутаций не дают никакого эволюционного вклада и быстро исчезают из популяции. Причиной этого могут быть как случайная элиминация, так и ассоциативный отбор STR-мутаций, обусловленный селективными процессами по нерекомбинантной части гаплоидного генома Y-хромосомы. В связи с этим стоит указать, что еще большее по величине различие между «семейными» и «эволюционными» оценками обнаружено и по другой нерекомбинантной гаплоидной системе – митохондриальной ДНК (Heyer *et al*., 2001). Причины подобных различий ещё не выяснены до конца ни для мтДНК, ни для Y-хромосомы (см. Di Giacomo *et al*., 2004; Zhivotovsky, Underhill, 2005).

Знание «трансмиссионных» оценок темпов мутационного процесса по микросателлитам Y-хромосомы важно для судебно-медицинских генетических экспертиз и исследования наследственных болезней, в то время как «эволюционные» оценки скоростей мутирования важны для датирования древних популяционных событий.

Эволюция микросателлитной изменчивости

Задолго до открытия микросателлитной изменчивости описание ее эволюции было предвосхищено теоретическими исследованиями (Ohta, Kimura, 1973; Moran, 1975), в которых изучались модели генетического дрейфа по локусам с пошаговыми мутациями. Широкое исследование микросателлитной изменчивости в популяциях человека началось с публикации Bowcock *et al*. (1994), поддержавшей гипотезу африканского происхождения человечества (Cann *et al*., 1987). К этому времени были даны оценки темпов мутирования по микросателлитным локусам (Weber, Wong, 1993), после чего появились первые методические исследования: по особенностям микросателлитной изменчивости у человека (Valdes *et al*., 1993), генетическому расстоянию $(\delta\mu)^2$, позволившему оценить время в модели африканского происхождения человечества (Goldstein *et al*., 1995a), R_{ST}-статистике как мере дифференциации популяций (Slatkin, 1995), дан общий теоретический анализ микросателлитной изменчивости (Zhivotovsky, Feldman, 1995). Это позволило развить методы оценки параметров роста численности древних популяций человека по данным о микросателлитах в современных популяциях (Kimmel *et al*., 1998; Reich, Goldstein, 1998; Pritchard *et al*., 1999; Gonser *et al*., 2000; Jin *et al*., 2000; King *et al*., 2000; Zhivotovsky *et al*., 2000), ввести новые генетические расстояния и статистики для оценки времени дивергенции популяций (Bowcock *et al*., 1994; Deka *et al*., 1995; Goldstein *et al*., 1995a, b; Zhivotovsky, 1999, 2001), разработать методы выявления подразделенности популяций и интенсивности генных потоков между ними (Slatkin, 1995; Michalakis, Excoffier, 1996; Rousset, 1996; Feldman *et al*., 1999). Было показано, что для надежной (с малой стандартной ошибкой) оценки времени свершения древних эволюционных событий по данным о микросателлитной изменчивости в современных популяциях человека могут потребоваться многие сотни локусов (Zhivotovsky, Feldman, 1995; Goldstein *et al*., 1996; Jorde *et al*., 1997).

Для изучения эволюционной истории популяций важно, чтобы используемые маркеры были селективно нейтральны, не влияли бы на приспособленность особей. В противном случае филогенетические древа популяций и оценки времени свершения древних популяционных событий, основанные на данных о частотах аллелей и гаплотипов, окажутся статистически смещенными. Результаты изучения белковых маркеров свидетельствуют, что их можно рассматривать как нейтральные лишь условно, поскольку они кодируют функционально важные белки, вовлеченные в адаптивные процессы. Микросателлиты в целом более нейтральны, чем белковые маркеры, поскольку в популяционные исследования вовлекаются STR-локусы, локализованные в интронах и других некодирующих участках генома. Необходимо отметить, что вследствие значительной скорости возникновения мутаций, равновероятно увеличивающих или уменьшающих размер аллеля на один или более мотивов, микросателлитная изменчивость характеризуется высоким уровнем *гомоплазии* – наличием аллелей, идентичных по числу повторов, но не идентичных по происхождению (т. е. мутировавших от аллелей с иным числом повторов).

Допущение селективной нейтральности позволяет рассматривать модели эволюции микросателлитной изменчивости, учитывающие лишь два основных в данном случае фактора: случайный генетический дрейф и мутационный процесс. Еще одним существенным фактором популяционной динамики, способным повлиять на распределение аллельных частот, является изменение численности, поскольку популяции человека не стационарны: человечество в целом растёт, а численность отдельных популяций может и уменьшаться.

Разработана широкая панель аутосомных микросателлитных локусов у человека (см. http://research.marshfieldclinic.org/genetics/sets/combo.html), позволяющая как картировать гены, так и изучать эволюцию популяций человека в глобальном масштабе. По одной из таких панелей нами (Rosenberg *et al*., 2002, Zhivotovsky *et al*., 2003) было проведено сравнительное исследование представителей разных народов мира, позволившее ответить на вопрос, можно ли по образцу ДНК конкретного индивида определить его регионально-этническую принадлежность и дать глобальную картину генетической дифференциации и эволюции человечества (рис. 8–10).

Статистический анализ микросателлитной изменчивости

Основной тип полиморфизма микросателлитных локусов связан с вариабельностью аллелей по числу повторов. На рис. 11 видно, что распределение аллелей, упорядоченных по возрастанию числа повторов, похоже на гистограмму распределений для количественных признаков. Гистограммы распределений аллелей микросателлитных локусов нередко унимодальны и внешне напоминают нормальное распределение. Однако на самом деле они часто далеки от нормального распределения, и статистические тесты показывают, что отклонения от нормального распределения по числу повторов часто вызваны «тяжелыми хвостами» распределения – аллелями с крайними величинами числа повторов. Количественно это выражается в величине коэффициента эксцесса (отношения четвертого центрального момента к квадрату дисперсии): она равна 3 для нормального распределения, в то время как ожидаемое его значение в равновесной популяции превышает 5 (Zhivotovsky, Feldman, 1995). Более того, «ненормальность» по числу микросателлитных повторов нередко связана с многовершинностью распределений и даже с разрывом гистограммы распределения на части. Нередко наличие таких разрывов объясняют тем, что в данной популяции могла возникнуть «многошаговая» мутация, т. е. мутация, изменяющая размер аллеля на два или большее число повторов. Это одно из возможных объяснений. Однако нельзя исключать и другое объяснение: случайный генетический дрейф также может привести к многовершинным и разорванным распределениям даже при одношаговых мутациях, т. е. мутациях, изменяющих число повторов ровно на единицу (рис. 12).

Возможность использовать подходы и методы количественной генетики позволяет исследовать микросателлитную изменчивость и различия между популяциями двумя разными взаимодополняющими подходами. С одной стороны, можно применить методы, развитые для классических полиморфных систем (белков, групп крови и др.) – с учетом их частоты в популяциях безотносительно к тому, сколько повторов стоит за этими аллелями (Животовский, 1991; Вейр, 1995; Weir, 1996). Например, в судебно-медицинских генетических исследованиях размер аллеля сам по себе несуществен; здесь важно лишь то, что микросателлитные локусы намного более удобны и надежны, имеют бóльшую идентификационную мощность и лучше удовлетворяют критериям доказательности, чем применявшиеся до того полиморфные белковые системы, группы крови и мультилокусные ДНК-фингерпринты (NRC, 1996). С другой стороны, популяционную изменчивость по микросателлитным локусам можно характеризовать в терминах количественных признаков с использованием хорошо развитого математического аппарата количественной генетики. Представление микросателлитной изменчивости как количественной удобно и естественно для оценки популяционно-генетических параметров, описания эволюционной динамики и датирования древних популяционных событий по данным о генотипическом составе ныне живущих популяций. Это возможно по той причине, что микросателлитный повтор является основной мутационной единицей, а особенности и

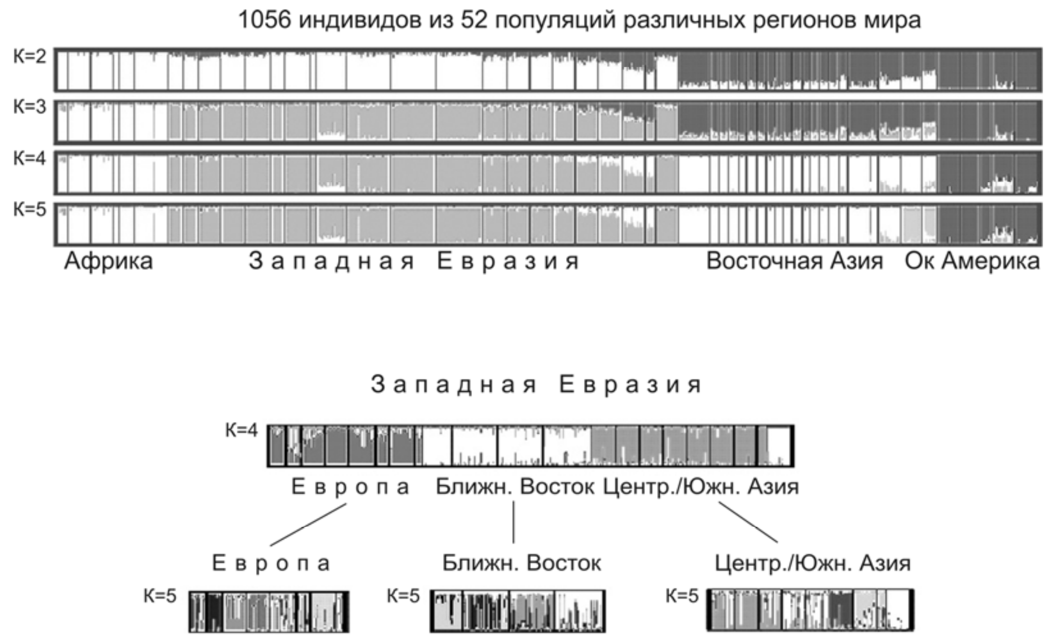


Рис. 8. Разбиение 1056 индивидов из 52 популяций на группы согласно их сходству друг с другом по 377 микросателлитным локусам (Rosenberg *et al*., 2002).

Были исследованы 1056 человек из 52 этнических групп из различных регионов мира – экваториальной, южной и северной Африки, западной, центральной и восточной Азии, Европы, Океании, Центральной и Южной Америки, биологические образцы которых хранятся в виде клеточных культур (The HGDP-CEPH Human Genome Diversity Cell Line Panel, Cann *et al*., 2002). Индивиды относились к определенной этнической группе согласно их самоопределению и были генотипированы по 404 STR-локусам в The Mammalian Genotyping Service (см. http://research.marshfieldclinic.org/genetics/sets/combo.html, Marshfield Panel #10), охватывающим все хромосомы, включая половые; из них были выбраны 377 аутосомных локусов (данные доступны по адресу http://research.marshfieldclinic.org/genetics/Freq/FreqInfo.htm). После того как все индивиды были генотипированы, полученные данные анализировали следующим образом. Задавали определенное число K (от 2 до 5 ÷ 6) и затем использовали алгоритм «structure» (Pritchard *et al*., 2000), чтобы подразделить всех 1056 индивидов на K «ДНК-групп» в соответствии со степенью их сходства по исследованным 377 локусам – без использования информации об их этнической принадлежности. После того как были сформированы ДНК-группы, определяли, в какую из них попали индивиды той или иной этнической группы. На диаграммах оригинала (Rosenberg *et al*., 2002; см. также Животовский, 2004б) группы были представлены разным цветом. Здесь диаграммы могли быть воспроизведены лишь в черно-белом варианте. На диаграммах все популяции отделены друг от друга вертикальной чертой. В пределах популяции каждый индивид представлялся вертикальной полосой с одной или несколькими цветами пропорционально тому, какая доля локусов отвечает соответствующей группе. Например, при $K = 5$ обе популяции из Океании (Ок), хотя они отчетливо отличаются от других групп, содержат индивидов, имеющих локусы, характерные для народов Восточной Азии и Западной Евразии, а в одной из популяций экваториальной Африки (пигмеи племени биака) имеется индивид, скорее относящийся к группе популяций из Западной Евразии. Оказалось, что по сходству ДНК практически все индивиды распределились по большим группам соответственно их географической принадлежности. А именно: при разбиении выборки на две группы ($K = 2$) в одну из них попали все представители негроидной (африканской, или черной) и европеоидной (евразийской, или белой) рас, а в другую – монголоидной (Восточная Азия и американские индейцы) и океанийской рас. При $K = 3$ первая группа разделилась на африканскую и евразийскую расы, при $K = 4$ из второй группы отделились американские индейцы, а при $K = 5$ выделилась океанийская раса. Многие индивиды несли большинство признаков своей расы. Но некоторые при этом имели значительную долю признаков другой расы. Дальнейший анализ (не показан на рисунке из-за невозможности воспроизведения цветной гаммы) показал, что в пределах каждой географической расы выделение групп популяций не было уже столь явным. Например, евразийская раса хорошо подразделяется на народы Европы, Ближнего Востока и Центральной/Южной Азии, но в их пределах наблюдается значительное перекрытие (см. Rosenberg *et al*., 2002; Животовский, 2004а, б, 2005а). Хорошо идентифицируются все аборигенные популяции – вероятно, вследствие генетического дрейфа, вызванного их небольшой численностью. Например, среди африканских выборок отчетливо выделяются представители старейших (из тех, что изучены здесь) племен – охотников-собирателей: пигмеев (биака и мбути) и сан, однако гораздо более многочисленные банту-говорящие народы не разделяются данным методом; американские индейцы четко разбиваются на все пять из имевшихся в данном исследовании племен: пима и майя из Центральной Америки, колумбийцы из севера Южной Америки, суруи и каритиана из бассейна Амазонки; два исследованных народа Океании также хорошо отделяются один от другого.

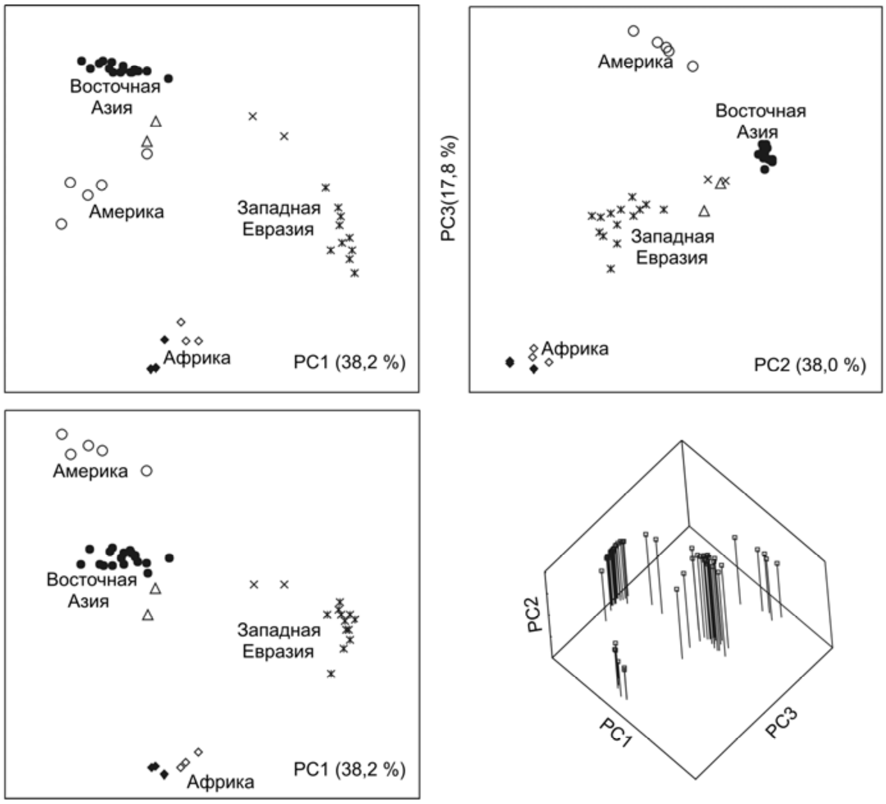


Рис. 9. Расположение выборок из основных регионов мира в пространстве трех главных компонент (по данным Zhivotovsky *et al*., 2003).

Ромбы – Центральная/Южная Африка (черные – охотники-собиратели, белые – банту-говорящие); звездочки – Западная Евразия (крестиком выделены уйгуры и хазарейцы); кружки черные – Восточная Азия; кружки светлые – Америка; треугольники – Океания.

Исследованы выборки из 52 этнических групп по 374 микросателлитным локусам (см. рис. 8). Главные компоненты были определены по 52×52-матрице парных значений коэффициента корреляции между средними значениями числа повторов в выборках (Zhivotovsky, 1999). А именно, если $\overline{r}_{A1}, \overline{r}_{A2}, ..., \overline{r}_{AL}$ и $\overline{r}_{B1}, \overline{r}_{B2}, ..., \overline{r}_{BL}$ – средние значения числа повторов в популяциях *A* и *B* по *L* исследованным локусам, то вычисляется коэффициент корреляции r_{AB} между ними.

Полученная картина дифференциации согласуется с данными, представленными на рис. 8. Детали распределения популяций (см. Zhivotovsky *et al*., 2003). Исследованные популяционные выборки образуют большие географические кластеры: Южная Африка, Западная Евразия, Восточная Азия, Океания и Америка. Положение популяций в пределах каждого кластера соответствует их популяционному статусу. Например, банту-говорящие народы Африки оказываются генетически ближе друг к другу, чем к племенам охотников-собирателей сан и мбути; на диаграмме сан и мбути оказались на границе африканского кластера, относительно обособясь от другого племени пигмеев – биака, оказавшегося ближе к подкластеру народов банту, что имеет свое объяснение (см. Zhivotovsky *et al*., 2003). Западная Евразия, которая включает Ближний Восток, Европу, Центральную и Южную Азию и Северную Африку, четко отделяется от других больших групп и имеет выраженную внутреннюю структуру. Уйгуры (выборка взята из Западного Китая) и хазарейцы занимают положение, промежуточное между Евразией и Восточной Азией, что отражает их родство с монголами. Этнические группы Восточной Азии формируют отдельный кластер, в котором также прослеживается определенная внутренняя структура. Океания отделяется от Восточной Азии. Популяции Америки отстоят от других континентальных групп. Одна из амазонских популяций – суруи – более обособлена от других американских популяций – вероятно, из-за изоляции и значительного генетического дрейфа, вызванного чрезвычайно малой численностью этого племени. Центральноамериканская популяция майя, напротив, показывает бóльшую, чем другие американские популяции, близость к неамериканским кластерам, что может отражать влияние миграций на американский континент в постколумбову эпоху.

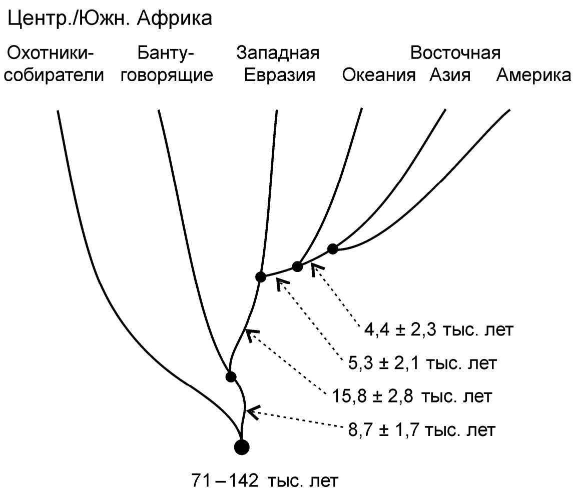


Рис. 10. Эволюционное древо популяций человека по данным об аутосомных микросателлитных локусах (Zhivotovsky *et al*., 2003).

Первичное разделение популяций человека современного анатомического типа и отделение друг от друга популяций началось в Африке – в среднем, около 100 тыс. лет назад. Численность той популяции, которая определила генетический пул современного населения земного шара, была порядка двух тысяч человек (см. Zhivotovsky *et al*., 2003; Животовский 2004б, 2005а). Затем одна из ветвей растущего в численности человечества вышла из Африки и стала делиться на континентальные группы в ходе пространственной экспансии и усиливающейся географической изоляции между этими группами. Стрелками указано минимальное время, прошедшее между отделением эволюционных ветвей. Следует иметь в виду, что отделение ветви не означает еще физического присутствия популяций в этом регионе. Например, ветвь, ведущая свое начало от азиатских популяций к американским индейцам, показывает, когда эта ветвь отделилась генетически; но нужно было еще время, за которое отделившиеся группы достигнут Берингии, дождутся конца оледенения и проникнут вглубь Америки.

темпы мутационного процесса по микросателлитным локусам, необходимые в целях датирования, известны сейчас даже лучше, чем для минисателлитов и классических полиморфных систем.

Ниже мы обсуждаем некоторые показатели (статистики) микросателлитной изменчивости, основанные на методах количественной генетики, поскольку на сегодня они описаны только в специальной литературе. Введем вначале необходимые обозначения. Пусть в популяционной выборке объема n аллелей (для диплоидных организмов $n = 2N$,

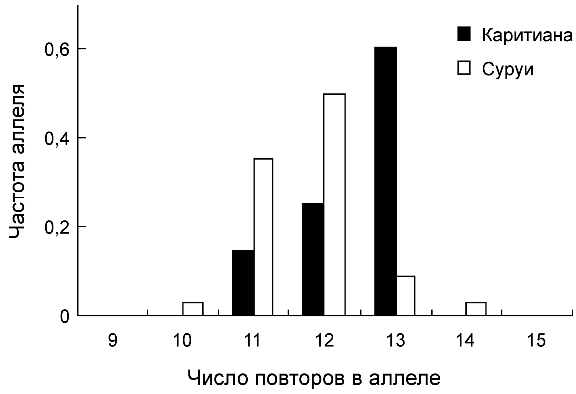


Рис. 11. Частоты аллелей тетрануклеотидного локуса *D2S2944* в выборках из южноамериканских племен бассейна Амазонки – каритиана и суруи (по Zhivotovsky *et al*., 2003).

Аллели по оси абсцисс расположены в порядке возрастания числа повторов. Поэтому гистограмма распределений в каждой выборке напоминает распределение количественного признака. Обе исследованные выборки отличаются друг от друга как по среднему числу повторов, так и по их дисперсии.

где N – количество особей в выборке) по данному микросателлитному локусу обнаружено k различных аллелей, каждый в количестве n_i ($I = 1, 2, \ldots, k$); $n = n_1 + n_2 + \ldots + n_k$, причем i-й аллель имеет r_i повторов. Имея эти данные, можно получить следующие, хорошо известные оценки параметров изменчивости (см. Крамер, 1975):

$$p_i = \frac{n_i}{n}$$ – частота аллеля с i повторами;

$$\bar{r} = \frac{\sum_{i=1}^{k} n_i r_i}{n} \quad \text{или} \quad \bar{r} = \sum_{i=1}^{k} p_i r_i$$ – среднее

число повторов; $V = \dfrac{\sum_{i=1}^{k} n_i r_i^2 - n\bar{r}^2}{n}$ или

$$V = \left(\sum_{i=1}^{k} p_i r_i^2 - \bar{r}^2 \right)$$ – дисперсия числа повторов; (эта оценка смещена; несмещенная оценка: $\dfrac{n}{n-1}V$);

$$K=\frac{n^2-2n+3}{(n-1)(n-2)(n-3)}\sum_{i=1}^{k}n_i(r_i-\overline{r})^4-\frac{3n(2n-3)}{(n-1)(n-2)(n-3)}V^2 \text{ или}$$

$$K=\frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{k}p_i(r_i-\overline{r})^4-\frac{3n(2n-3)}{(n-1)(n-2)(n-3)}V^2.$$

V, r и *K* – это начальные буквы английских терминов *variance, repeat* и *kurtosis*.

На этих параметрах основан ряд популяционно-генетических статистик по микросателлитным локусам. Перечислим некоторые из них. Определена мера подразделенности популяции – R_{ST} (Slatkin, 1995), являющаяся аналогом показателя F_{ST} С. Райта. Дан индекс дисбаланса $(-\ln\hat{\beta})$, вычисляемый по данным о гетерозиготности и дисперсии числа повторов как тест на рост численности популяций и прохождения ею «горлышка бутылки» (Kimmel *et al*., 1998; King *et al*., 2000). Другой подобный тест основан на индексе экспансии S_k, определяемом по данным о величинах эксцесса и дисперсии (Zhivotovsky *et al*., 2000); при этом можно оценить эффективную численность прапопуляции и время, когда она стала возрастать (Zhivotovsky *et al*., 2003, см. также уравнения (7), (8) из Zhivotovsky *et al*., 2000).

Для оценки генетических расстояний между популяциями была предложена мера $\delta\mu^2$, определяемая как эвклидово расстояние между средними значениями числа повторов в сравниваемых популяциях (Goldstein *et al*., 1995a):

$$\delta\mu^2=\frac{\sum_{l=1}^{L}(\overline{r}_{1l}-\overline{r}_{2l})^2}{L},$$

где $\overline{r}_{1l}$ и $\overline{r}_{2l}$ – это среднее число повторов по локусу l, а L – число исследованных микросателлитных локусов в популяциях, обозначенных как 1 и 2. Это расстояние связывает время дивергенции популяций от общего предка (T) со скоростью мутаций (Goldstein *et al*., 1995a; Zhivotovsky, Feldman, 1995): $T = \delta\mu^2/2w$. Однако эта формула справедлива лишь при условии, что: 1) обе популяции находятся в генетическом равновесии; 2) численность их одинакова и не меняется во времени; 3) нет генных потоков между популяциями. При нарушении этих теоретических допущений приведенная статистика дает смещенные оценки времени дивергенции. Для популяций человека ни одно из этих допущений не выполняется. В частности, если популяция растет в численности, то приведенная формула дает резко заниженное время дивергенции (Zhivotovsky, 2001). Поэтому для оценки времени, прошедшего от разделения популяций, была введена модифицированная оценка T_D:

$$T_D=\frac{\sum_{l=1}^{L}\left[(\overline{r}_{1l}-\overline{r}_{2l})^2+V_{1l}+V_{2l}-2V_{0l}\right]\Big/L}{2w},$$

где V_{1l} и V_{2l} – дисперсии числа повторов по локусу l в сравниваемых популяциях, а V_{0l} – это дисперсия числа повторов в родительской популяции, от которой дивергировали популяции 1 и 2 (Zhivotovsky, 2001). Статистическая межлокусная стандартная ошибка оценки T_D (которую обозначим se_T) вычисляется следующим образом. Вначале оценивается каждый компонент числителя T_D: $t_l=(\overline{r}_{1l}-\overline{r}_{2l})^2+V_{1l}+V_{2l}-2V_{0l}$. Затем вычисляется $se_t=\sqrt{\left(\sum_{l=1}^{L}t_l^2-L\overline{t}^2\right)\Big/L(L-1)}$, где $\overline{t}$ – среднее значение всех t_l (оно совпадает с числителем T_D). Окончательно $se_T = se_t/2w$.

Оценка T_D показывает время в числе поколений, если w – эффективная скорость возникновения мутаций на локус на поколение. В этом случае его можно пересчитать в единицах времени, приняв определенную длительность поколения; часто ее принимают равной 25 годам. Если w – темпы мутирования за определенную единицу времени (как, например, 0,69 × 10^{-3} – скорость возникновения мутаций в микросателлитных

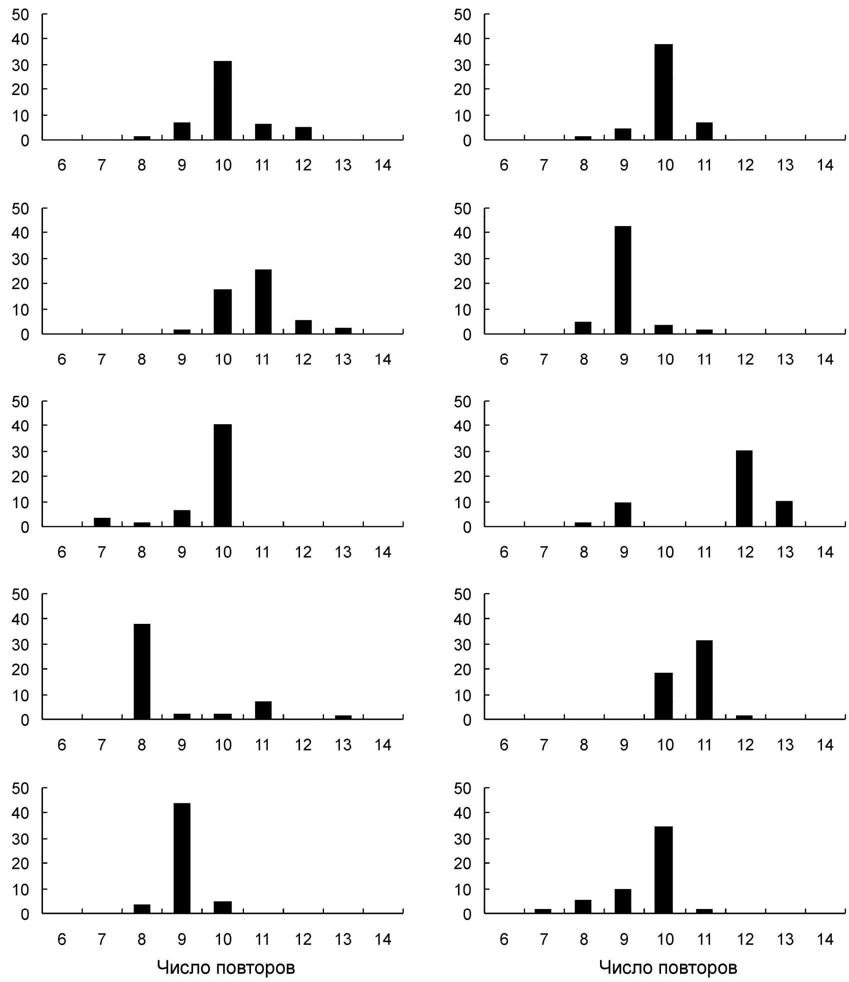


Рис. 12. Распределения частот микросателлитных аллелей, достигнутые в десяти моделированных популяциях, после их дивергенции от общей предковой популяции.

Под действием генетического дрейфа распределение частот аллелей с течением времени может остаться унимодальным и без разрывов, а может стать мультимодальным и даже с разрывами.

Исходная родительская популяция была мономорфна по аллелю с числом повторов 10. Численность каждой предковой популяции – 25 диплоидных особей (50 гаплоидных геномов). Факторы динамики: мутационный процесс и случайный генетический дрейф. В целях экономии расчетного времени допустили высокий темп мутирования – 0,01 на локус на поколение, мутации равновероятно увеличивают или уменьшают число повторов на единицу. Число поколений независимой дивергенции предковых популяций – 100.

локусах Y-хромосомы за 25 лет), то величина T_D, умноженная на 25, дает оценку в годах. Следует отметить, что поскольку в числителе этой формулы стоит усредненная по локусам величина, то и для оценки времени дивергенции требуется знание темпов мутирования не по отдельным локусам, а лишь в среднем по всем используемым локусам. Это значительно облегчает анализ данных: хотя темпы мутирования сильно варьируют от локуса к локусу, оценка средней частоты мутирования относительно стабильна для большого набора локусов и мало зависит от этого набора, если выбор локусов в определенном смысле случаен – не определяется темпами мутирования.

Оценка T_D не требует условий генетического равновесия и стационарной численности и устойчива по отношению к слабым миграциям. Однако у нее есть свой очевидный недостаток: величины V_{0l} неизвестны. Было предложено несколько подходов (Zhivotovsky, 2001), позволяющих обойти этот недостаток. Один из них – оценивать верхнюю и нижнюю границы для T_D. Например, положив V_{0l} равными нулю, мы получим верхнюю границу T_D.

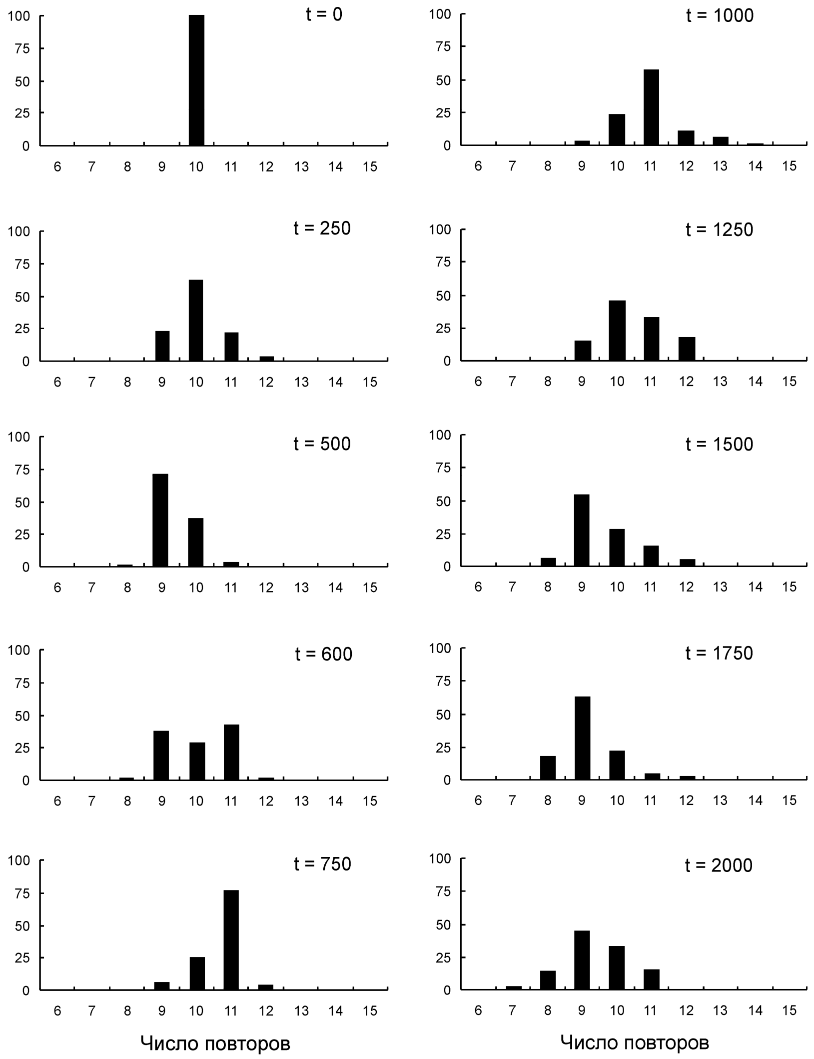
Однако эта верхняя граница может быть намного больше реального времени дифференциации, если изменчивость велика, а популяции недавно дивергировали друг от друга. Далее можно получить нижнюю границу для T_D, приняв в качестве V_{0l} средние значения дисперсий в сравниваемых или иных популяциях. Это будет действительно нижней границей только в том случае, когда отделившиеся от общего предка популяции в среднем росли в численности, поскольку их дисперсии при этом увеличивались во времени за счет мутаций. (В случае, если популяции уменьшались в численности за оцениваемое время, этот подход даст верхнюю границу T_D). Еще один подход к оценке T_D – принять одну (или несколько) из ныне существующих популяций в качестве модельной, предположительно воспроизводящей вариабельность предковой популяции, и использовать дисперсию числа повторов в такой модельной популяции в качестве V_{0l} (см., например, Zhivotovsky *et al*., 2003, P. 1178). Можно применить показатель T_D к микросателлитным аллелям, сцепленным с уникальными мутациями, например, к микросателлитной изменчивости в пределах одной Y-хро-мосомной гаплогруппы (Zhivotovsky *et al*., 2004) или к изменчивости в пределах аутосомных SNPSTR-систем (Ramakrishnan, Mountain, 2004). Для микросателлитной изменчивости в гаплогруппах Y-хромосомы интерпретация T_D и других статистик рассмотрены ниже. Имеются другие индексы, оценивающие сходство популяций по микросателлитным локусам (Bowcock *et al*., 1994; Shriver *et al*., 1995; Michalakis, Excoffier, 1996).

Анализ микросателлитной изменчивости в гаплогруппах Y-хромосомы

В реконструкции генетической истории человечества, наряду с использованием полиморфизма митохондриальной ДНК, важной стала разработка методик и номенклатуры гаплогрупп Y-хромосомы (Underhill *et al*., 2000; Hammer et al., 2001; Y Chromosome Consortium, 2002). *Гаплогруппа* Y-хромосомы определяется композицией уникальных мутаций (инсерций/делеций и нуклеотидных замен), возникших в различных участках Y-хромосомы. В момент возникновения гаплогруппы ее численность равна единице, поскольку определяющая ее мутация представлена только у одного индивида, а затем она начинает стохастически меняться. Вновь возникшая гаплогруппа с большой вероятностью может исчезнуть из популяции, если мужской род, ведущий свое начало от «родоначальника» этой мутации, прервется в каком-либо поколении. Например, в большой стационарной по численности популяции число сыновей в семье (k) распределено по закону Пуассона: $e^{-1}/k!$. Значит, вероятность элиминации этой гаплогруппы уже к следующему поколению равна $1/e$, т. е. почти 0,37, а вероятность для нее исчезнуть в ближайших поколениях ещё выше. Таким образом, большинство новых гаплогрупп элиминируются стохастическим процессом генетического дрейфа. Однако некоторые из гаплогрупп могут остаться в популяции по чисто случайным причинам и увеличить свою частоту – вплоть до столь значительной, чтобы попасть даже в небольшую по объему популяционную выборку. На фоне стохастического роста численности гаплогруппы в ней начинает накапливаться изменчивость по микросателлитным локусам. Исходно в Y-хромосоме основателя данная гаплогруппа была представлена одним определенным *гаплотипом* – набором аллелей в различных STR-локусах. Затем у потомков в каждом локусе, помимо аллеля-основателя, появляются другие аллели за счет мутаций – возникают новые гаплотипы. По данным о накопленной гаплотипической изменчивости можно оценить возраст этой гаплогруппы. Следует сказать, что подразделение изменчивости Y-хромосомы на уникальные гаплогруппы с последующим анализом «молекулярных часов» в виде быстромутирующих микросателлитных локусов является очень эффективным методом изучения тонкой структуры и эволюционной истории популяций. Аналогичный метод применяют при изучении митохондриальной ДНК. В обеих ситуациях успешная применимость метода обусловлена тем, что обе системы (Y-хромосома и мтДНК) представляют собой интактные, нерекомбинирующие генетические единицы. Недавно для изучения эволюционной истории по аутосомной части генома было предложено использовать тесно сцепленные пары локусов, каждая из которых состоит из STR-локуса и уникальной мутации, такой, как SNP или делеция/инсерция (Mountain *et al*., 2002).

Для оценки возраста гаплогруппы (точнее – возраста STR-изменчивости, наблюдаемой в пределах рассматриваемой гаплогруппы в данной популяции) важной является статистика ASD_0 (*A*verage *S*quared *D*ifference *zero*). Этот показатель близок к дисперсии (введенной как средний квадрат различий между аллелями (Goldstein, 1995b), но, в отличие от нее, определяется как среднеквадратичное отклонение от числа повторов в *аллеле-основателе* (Zhivotovsky *et al*., 2004), а именно: ASD_0 определяется как $\frac{\sum_{i=1}^{k} n_i \left(r_i - f\right)^2}{n}$, где f – это число повторов в аллеле-основателе. Практически ее величину можно вычислить как $ASD_0 = V + \left(\bar{r} - f\right)^2$, где V – это указанная выше смещенная оценка дисперсии. Данный подход и статистику ввели Thomas *et al*., (1998), рассматривая возможность только одношаговых мутаций (т. е. изменяющих число повторов на единицу и не больше) и используя в качестве f число повторов у модального, т. е. наиболее частого гаплотипа (см. также Stumpf, Goldstein, 2001; Zhivotovsky *et al*., 2004; Животовский, 2005б).

Метод оценки возраста STR-изменчивости по статистике ASD_0 основан на следующем математическом факте. Пусть исследуются L микросателлитных локусов и f_l – это число повторов в гаплотипе-основателе по l-му локусу. Обозначим через N_{li} и $p_{li} = \frac{N_{li}}{N_l}$ численность и частоту i-го аллеля, имеющего r_{li} повторов в l-ом локусе в пределах данной гаплогруппы; здесь N_l – суммарное количество типированных аллелей в данной выборке по l-му локусу в этой гаплогруппе. Тогда, если от момента происхождения гаплогруппы (нуле-вой STR-изменчивости) прошло T_0 поколений, то ожидаемое значение величины $\sum_{l=1}^{L} \sum_i p_{li} \cdot \left(r_{li} - f_l\right)^2$ равно wLT_0, где w – средний по локусам эффективный темп мутирования. Отсюда следует, что, если имеется популяционная выборка, индивиды которой отнесены к гаплогруппам и типированны по L микросателлитным локусам, то оценка возраста STR-изменчивости в данной гаплогруппе такова: $T_0 = \frac{\frac{1}{L}\sum_{l=1}^{L} \sum_i p_{li} \cdot \left(r_{li} - f_l\right)^2}{w}$ (Zhivotovsky *et al*., 2004). Оценку эту можно представить в виде $T_0 = \frac{1}{wL} \cdot \sum_l t_l$, где $t_l = \sum_i p_{li} \cdot \left(r_{li} - f_i\right)^2$, и на основе этого определить приближенную стандартную ошибку для T_0: $se_T = \frac{1}{w} \cdot \sqrt{\frac{\sum_l \left(t_l - T_0\right)^2}{L(L-1)}}$ (Zhivotovsky *et al*., 2004). Существенным в применении указанной оценки является определение гаплотипа-основателя. Как видно на рис. 13, на протяжении первых поколений нарастания изменчивости аллель-основатель еще остается наиболее представительным в выборке: частоты других аллелей еще не вырастают до значительных величин. Именно поэтому в качестве основателя принимают модальный гаплотип, т. е. гаплотип, частота которого максимальна в исследуемой выборке (Thomas *et al*., 1999; Stumpf, Goldstein, 2001). Однако нередко распределение частот аллелей полимодальное, а если даже оно унимодальное, то, как известно из статистики (например, Крамер, 1975), оценка максимума распределения – моды – статистически неустойчива. Более того, как видно на рис. 13, несколько сот поколений мутаций и дрейфа могут привести к тому, что аллель-основатель перестает быть модальным; еще быстрее перестает быть модальным гаплотип-основатель из-за мутаций в составляющих его локусах. Поэтому более подходящей является медианная оценка числа повторов в гаплотипе-основателе (Животовский, неопубликованные данные): для каждого локуса она вычисляется как медиана числа повторов по всем аллелям, которые обнаружены в данной выборке. Например, пусть имеется следующий набор из одиннадцати аллелей (упорядоченных по числу повторов): 5, 6, 6, 6, 6, 7, 7, 8, 8, 8, 8. Если а


t = 0
t = 250
t = 500
t = 600
t = 750
t = 1000
t = 1250
t = 1500
t = 1750
t = 2000
Число повторов
Число повторов

б


Поколения

по этим данным построить гистограмму частот, то будет видно, что распределение бимодальное, с пиками на числе повторов 6 и 8; поэтому модальный класс в данном примере просто отсутствует. Медианой же здесь будет $f=7$ (при четном объеме выборки аллелей вычисляют среднее двух срединных аллелей). Медианная оценка статистически гораздо более устойчива, чем модальная, но и она не совершенна, поскольку на эволюционно больших временах распределение числа повторов в гаплотипах может отклоняться от аллельной композиции основателя, и тогда медианная оценка будет приводить к заниженной оценке возраста гаплогруппы. В подобных случаях следует сопоставлять друг с другом медианные оценки, полученные по данным о выборках из географически различных популяций – в случае их совпадения можно быть более уверенным, что это и есть реальный гаплотип-основатель, хотя и здесь есть свои тонкости, а именно: в каждой популяции микросателлитная изменчивость в данной гаплогруппе могла происходить от своего основателя – одного мужчины или группы мужчин с тем же гаплотипом. Поэтому гаплотипы-основатели, определенные медианным методом, так же, как и основанные на них оценки возраста гаплогруппы, сделанные по данным одной выборки или небольшого числа выборок из географически ограниченной области, могут относиться не к истинному моменту возникновения этой гаплогруппы, а к моменту появления ее на данной территории. В случае же если группа мигрантов на момент вхождения в репродуктивный пул данной популяции имела по данной гаплогруппе ненулевую микросателлитную изменчивость, оценка возраста гаплогруппы будет относиться не к этой популяции, а к той, из которой эти мигранты пришли. Именно по этим причинам оценка T_0 интерпретируется как возраст микросателлитной изменчивости, наблюдаемой по данной гаплогруппе в данной популяции, но не как время происхождения гаплогруппы.

Анализ микросателлитной изменчивости в пределах гаплогруппы Y-хромосомы можно проводить, используя и другие методы. Например, применить оценку T_D для определения времени, прошедшего с момента отделения ныне существующих популяций от общей предковой популяции (в пределах данной гаплогруппы). Поскольку в начале процесса накопления STR-изменчивости ее дисперсия равна нулю, а численность гаплогруппы в среднем за время, прошедшее с момента ее возникновения, росла, то, положив $V_{0l} = 0$ в приведенной выше формуле для T_D, мы получим верхнюю границу для времени дивергенции популяций (точнее их частей, соответствующих изучаемой гаплогруппе). Эта граница тем ближе к реальной, чем меньше момент разделения популяций отстоит от момента возникновения STR-изменчивости в данной

Рис. 13. Моделирование стохастической динамики частоты аллеля-основателя.

а – изменение гистограммы частот аллелей в ряду поколений. По оси ординат – частоты аллелей (в %).

б – случайные изменения частоты аллеля-основателя под действием мутаций и генетического дрейфа. По оси ординат – частота аллеля-основателя (в %).

На рисунке представлена одна из реализаций динамики частот микросателлитных аллелей в модели с пошаговыми мутациями. В исходном поколении $t = 0$ популяция была мономорфна по аллелю-основателю, число которых было принято равным 10. Темп мутирования – величина порядка 10^{-3} на поколение, мутации – пошаговые: с равной вероятностью увеличивают или уменьшают число повторов ровно на единицу; численность популяции – 500 диплоидных особей. Вначале наиболее частым был аллель-основатель (см. рис. А, t = 250, и рис. б). Однако затем, после трехсот-четырехсот поколений, наиболее частым стал аллель с 9 повторами (см. $t = 500$), затем он сменился на аллель с числом повторов 11 (см. $t = 750$). В данной реализации стохастической динамики переход от аллеля 9 к аллелю 11 сопровождался бимодальностью частотного распределения аллелей: см. $t = 600$ (следовательно, бимодальность может вызываться не только мутациями, изменяющими число повторов на несколько единиц, но и чисто стохастическими процессами). Затем наиболее частым вновь стал аллель с 10 повторами и затем вновь аллель 9. Таким образом, в ряду поколений частоты аллелей стохастически меняются и аллель-основатель может быть не самым частым. В ряде других реализаций наиболее частыми на какое-то время становились также аллели, отстоящие от аллеля 10 на два и более повтора; в некоторых реализациях процесс изменения распределения аллелей был не колеблющимся, как на рис. а, а как бы направленным – или в сторону уменьшения числа повторов, или в сторону их увеличения; при этом частота аллеля-основателя снижалась до незначительной величины. Кроме того, следует иметь в виду, что возврат к распределению с максимальной частотой аллеля 10 (как, например, на рис. а, $t = 1250$) может быть вызван не только стохастическим увеличением частоты аллеля-основателя, но и гомоплазией – обратными мутациями к нему.

гаплогруппе. Однако для недавно разошедшихся популяций эта верхняя граница может значительно отстоять от реального времени дивергенции. Поэтому вместо нее можно использовать показатель T_D, в котором значение V_0 оценивается исходя из возраста гаплогруппы и внутрипопуляционной дисперсии по числу повторов – детали применения можно найти в следующих работах (Zhivotovsky *et al*., 2004; Zegura *et al*., 2004; Semino *et al*., 2004; Rootsi *et al*., 2004; Gonçalves *et al*., 2005; Sengupta *et al*., 2006).

Географическое происхождение гаплогрупп Y-хромосомы можно установить по следующему эмпирическому критерию: в том месте или популяции, где возникла данная гаплогруппа, ее частота и STR-дисперсия (или возраст STR-изменчивости) максимальны по сравнению с другими популяциями (Sengupta *et al*., 2006). При их несовпадении (когда максимум частоты гаплогруппы приходится на один географический регион, а максимум дисперсии – на другой), место возникновения гаплогруппы становится неопределенным, но при необходимости сделать предварительное заключение предпочтение следует отдавать дисперсии как статистически и эволюционно более устойчивому показателю.

Заключение

Микросателлитная изменчивость огромна. Наличие в популяциях десятка, а нередко и десятков различных аллелей по каждому микросателлитному локусу обязано высоким темпам мутирования и селективной нейтральности, поскольку для эволюционного исследования выбираются те микросателлиты, которые лежат вне кодирующих областей генома. Однако именно их высокая изменчивость требует осторожности в интерпретации результатов анализа популяционных данных. Например, существуют значительные межлокусные различия: в популяциях по одним локусам может быть, скажем, пять аллелей, а по другим – двадцать (рис. 3). Соответственно, по одним локусам популяции могут мало отличаться друг от друга, а по другим – значительно. Как это интерпретировать? Два основных фактора могут привести к межлокусным различиям. Первый – это вариация в темпах мутирования: аллельное разнообразие по числу повторов прямо связано с частотой возникновения микросателлитных мутаций. Поэтому локусы с низкой скоростью мутирования получают больше «эволюционных шансов» на то, что в них будет иметься меньше аллелей, чем в быстромутирующих локусах. Второй фактор – «генетическая выборочность»: по каждому отдельному локусу эволюционная траектория представляет собой сильно колеблющийся во времени случайный процесс и потому для данного вида или популяции конкретная реализация этого процесса в данном локусе может привести к малому числу аллелей и малой дисперсии по числу повторов, а может, напротив, привести к большой аллельной вариации. Теория показывает, что в популяции величина внутрилокусной дисперсии по числу повторов может варьировать от локуса к локусу с коэффициентом вариации, превышающим 100 % (!), даже если темпы мутирования в этих локусах идентичны. Эффекты этих двух факторов – вариации в темпах мутирования и генетической выборочности – перекрываются и потому трудноотделимы друг от друга. В частности, разделение микросателлитных локусов на группы мало- и сильновариабельных может привести к тому, что локусы окажутся сгруппированными не только по темпам мутирования, но и во многом по их «случайной» эволюционной судьбе у данного вида. Поэтому выводы о внутри- и межпопуляционных различиях, сделанные по каждому из исследованных микросателлитных локусов в отдельности, следует соотнести с доступной информацией о скорости мутирования в каждом из этих локусов. При отсутствии такой информации следует ориентироваться на популяционно-генетический анализ, основанный на сводных данных о совокупности микросателлитных локусов. Конечно, если исследованные микросателлитные локусы можно подразделить на четко различимые группы, возможно, отличающиеся друг от друга мутационными свойствами (например, с ди- и тетрануклеотидными повторами), то такие группы локусов можно анализировать раздельно. Если же такое разделение затруднительно (например, из-за малого числа локусов), а различия между локусами по степени аллельной вариабельности значительны, то для сравнительного анализа индивидуальных или межпопуляционных различий можно применить нормировку на дисперсию числа повторов или иные преобразования, сглаживающие межлокусную вариацию. В любом случае популяционно-генетический анализ данных и интерпретация его результатов должны учитывать особенности микросателлитной изменчивости.

Работа поддержана программой Президиума РАН «Фундаментальные науки – медицине» (направление № 5 «Полиморфизм человека») и грантом РФФИ 04-04-48639.

Литература

Вейр. Б. Анализ генетических данных. М.: Мир, 1995. 399 с.

Животовский Л.А. Популяционная биометрия. М.: Наука, 1991. 267 с.

Животовский Л.А. Исследование генетической структуры и эволюции популяций человека по данным об аутосомных микросателлитных локусах // Матер. VII популяционного семинара. Сыктывкар, 2004. Сыктывкар 2004а. (в печати).

Животовский Л.А. Гены и расы: Все мы одного роду-племени // Наука в России. 2004б. № 4. С. 33–38.

Животовский Л.А. ДНК-различия между этническими группами и их эволюционное становление // Эволюция, поведение, общество: этология человека / Ред. М.Л. Бутовская. М.: Ин-т антропологии и этнографии РАН, 2005а. (в печати).

Животовский Л.А. ДНК-датирование древних популяционных событий // Эволюция, поведение, общество: этология человека / Ред. М.Л. Бутовская. М.: Ин-т антропологии и этнографии РАН, 2005б. (в печати).

Крамер Г. Математические методы статистики. М.: Мир, 1975.

Ashley C.T., Warren S.T. Trinucleotide repeat expansion and human disease // Annu. Rev. Genet. 1995. V. 29. P. 703–728.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J. *et al*. High resolution of human evolutionary trees with polymorphic microsatellites // Nature. 1994. V. 368. P. 455–457.

Cann H.M., de Toma C., Cazes L. *et al*. A human genome diversity cell line panel // Science. 2002. V. 296. P. 261–262.

Cann R.L., Stoneking M., Wilson A.C. Mitochondrial DNA and human evolution // Nature. 1987. V. 325. P. 31–36.

Chakraborty R., Kimmel M., Stivers D.N. *et al*. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 1041–1046.

Cooper G., Amos W., Bellamy R. *et al*. An empirical exploration of the $(\delta\mu)^2$ genetic distance for 213 human microsatellite markers // Am. J. Hum. Genet. 1999. V. 65. P. 1125–1133.

Deka R., Jin L., Shriver M.D. *et al*. Population genetics of dinucleotide $(dC–dA)_n{\cdot}(dG–dT)_n$ polymorphisms in world populations // Am. J. Hum. Genet. 1995. V. 56. P. 461–474.

Dib C., Faure S., Fizames C. *et al*. A comprehensive genetic map of the human genome based on 5,264 microsatellites // Nature. 1996. V. 380. P. 152–154.

Di Giacomo F., Luca F., Popa O. *et al*. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe // Hum. Genet. 2004. V. 115. P. 357–371.

Di Rienzo A., Donnelly P., Toomajian C. *et al*. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories // Genetics. 1998. V. 148. P. 1269–1281.

Dupuy B.M., Stenersen M., Egeland T., Olaisen B. Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci // Human Mutat. 2004. V. 23. P. 117–124.

Evett I.W., Weir B.S. Interpreting DNA Evidence. Sinauer Associates, Inc. Sund, Mass, 1998.

Feldman M. W., Bergman A., Pollock D.D., Goldstein D.B. Microsatellite genetic distances with range constraints: analytic description and problems of estimation // Genetics. 1997. V. 145. P. 207–216.

Feldman M.W., Kumm J., Pritchard J.K. Mutation and migration in models of microsatellite evolution // Microsatellites: Evolution and Applications / Eds D.G. Goldstein, C. Schlotterer. Oxford Univ. Press, 1999. P. 98–115.

Forster P., Röhl A., Lünnemann P. *et al*. A short tandem repeat-based phylogeny for the human Y chromosome // Am. J. Hum. Genet. 2000. V. 67. P. 182–196.

Garza J.C., Slatkin M., Freimer N.B. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size // Mol. Biol. Evol. 1995. V. 12. P. 594–603.

Gill P., Kimpton C.P., Urquhart A. *et al*. Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future // Electrophoresis. 1995. V. 16. P. 1543–1552.

Goldstein D.B., Linares A.R., Cavalli-Sforza L.L., Feldman M.W. Genetic absolute dating based on microsatellites and the origin of modern humans // Proc. Natl Acad. Sci. USA. 1995a. V. 92. P. 6723–6727.

Goldstein D.B., Linares A.R., Cavalli-Sforza L.L., Feldman M.W. An evaluation of genetic distances for use with microsatellite loci // Genetics. 1995b. V. 139. P. 463–471.

Goldstein D.B., Zhivotovsky L.A., Nayar K. *et al.* Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes // Mol. Biol. Evol. 1996. V. 13. P. 1213–1218.

Gonçalves R., Freitas A., Branco M., Rosa A., Fernandes A.T., Zhivotovsky L.A. *et al.* Y-chromosome lineages from Portugal, Madeira and Acores record elements of Shepardim and Berber ancestry // Ann. Hum. Genet. 2005. V. 69. P. 443–454.

Gonser R., Donnelly P., Nicholson G., Di Rienzo A. Microsatellite mutations and inferences about human demography // Genetics. 2000. V. 154. P. 1793–1807.

Hammer M.F., Karafet T.M., Redd A.J. *et al.* Hierarchical patterns of global human Y-chromosome diversity // Mol. Biol. Evol. 2001. V. 18. P. 1189–1203.

Henderson S.T., Petes T.D. Instability of simple sequence DNA in *Saccharomyces cerevisiae* // Mol. Cell. Biol. 1992. V. 12. P. 2749–2757.

Heyer E., Puymirat J., Dietjes P. *et al.* Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees // Hum. Mol. Genet. 1997. V. 6. P. 799–803.

Heyer E., Zietkiewicz E., Rochowski A. *et al.* Phylogenetic and familial estimates of mitochondrial substitution rates: Study of control region mutations in deep-rooting pedigrees // Am. J. Hum. Genet. 2001. V. 69. P. 1113–1126.

Holmans P. Nonparametric linkage // Handbook of Statistical Genetics / Ed. D.J. Balding *et al.* John Wiley, Sons, Ltd, 2001. P. 487–505.

Huang, Q.Y., Xu F.H., Shen H. *et al.* Mutation patterns at dinucleotide microsatellite loci in humans // Am. J. Hum. Genet. 2002. V. 70. P. 625–634.

Jeffreys A.J., Royle N.J., Wilson V., Wong Z. Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA // Nature. 1988. V. 332. P. 278–281.

Jin L., Baskett M.L., Cavalli-Sforza L.L., Zhivotovsky L.A. *et al.* Microsatellite evolution in modern humans: a comparison of two data sets from the same populations // Ann. Hum. Genet. 2000. V. 64. P. 117–134.

Jorde L.B., Rogers A.R., Bamshad M. *et al.* Microsatellite diversity and the demographic history of modern humans // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 3100–3103.

Jorde L.B., Watkins W.S., Bamshad M.J. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data // Am. J. Hum. Genet. 2000. V. 66. P. 979–988.

Kashi Y., Tikochinsky Y., Genislav E. *et al.* Large restriction fragments containing poly-TG are polymorphic in a variety of vertebrates // Nucl. Acids Res. 1990. V. 18. P. 1129–1132.

Kayser M., Kittler R., Erler A. *et al.* A comprehensive survey of human Y-chromosomal microsatellites // Am. J. Hum. Genet. 2004. V. 74. P. 1183–1197.

Kayser M., Roewer L., Hedman M. *et al.* Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs // Am. J. Hum. Genet. 2000. V. 66. P. 1580–1588.

Kelley R., Gibbs M., Collick A., Jeffreys A.J. Spontaneous mutation at the hypervariable mouse minisatellite locus Ms6-hm: flanking DNA sequence and analysis of germline and early somatic mutation events // Proc. R. Soc. Lond. B. 1991. V. 245. P. 235–245.

Kimmel M., Chakraborty R., King J.P. *et al.* Signatures of population expansion in microsatellite repeat data // Genetics. 1998. V. 148. P. 1921–1930.

King J.P., Kimmel M., Chakraborty R. A power analysis of microsatellite-based statistics for inferring past population growth // Mol. Biol. Evol. 2000. V. 17. P. 1859–1868.

Kong A., Gudbjartsson D.F., Sainz J. *et al.* A high-resolution recombination map of the human genome // Nat. Genet. 2002. V. 31. P. 241–247.

Levinson G., Gutman G.A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution // Mol. Biol. Evol. 1987. V. 4. P. 203–221.

Michalakis Y., Excoffier L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci // Genetics. 1996. V. 142. P. 1061–1064.

Moran A.P. Wandering distributions and the electrophoretic profile // Theor. Popul. Biol. 1975. V. 8. P. 318–330.

Mountain J.L., Knight A, Jobin M. *et al.* SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes // Genome Res. 2002. V. 12. P. 1766–1722.

Nauta M.J., Weissing E.J. Constraints on allele size at microsatellite loci: implications for genetic differentiation // Genetics. 1996. V. 143. P. 1021–1032.

NRC (National Research Council). The Evaluation of Forensic DNA Evidence. National Academy Press. Washington, D.C., 1996. 254 p.

Ohta T., Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population // Genet. Res. 1973. V. 22. P. 201–204.

Ott J. Analysis of Human Genetic Linkage. Baltimore; London: The Johns Hopkins University Press, 1999. 382 p.

Pritchard J.K., Seielstad M.T., Pe´rez-Lezaun A., Feldman M.W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites // Mol. Biol. Evol. 1999. V. 16. P. 1791–1798.

Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data // Genetics. 2000. V. 155. P. 945–959.

Ramakrishnan U., Mountain J.L. Precision and accuracy of divergence time estimates from STR and SNPSTR variation // Mol. Biol. Evol. 2004. V. 21. P. 1960–1971.

Reich D.E., Goldstein D.B. Genetic evidence for a Paleolithic human population expansion in Africa // Proc. Natl Acad. Sci. USA. 1998. V. 95. P. 8119–8123.

Rootsi S., Magri Ch., Kivisild T., Benuzzi G., Help H., Bermisheva M., Kutuev I., Barać L., Peričić M., Balanovsky O., Pshenichnov A., Dion D., Grobei M., Zhivotovsky L.A. *et al*. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe // Am. J. Hum. Genet. 2004. V. 75. P. 128–137.

Rosenberg N.A., Pritchard J.K., Weber J.L. *et al*. Genetic structure of human populations // Science. 2002. V. 298. P. 2381–2385.

Rousset F. Equilibrium values of measures of population subdivision for stepwise mutation processes // Genetics. 1996. V. 142. P. 1357–1362.

Schlötterer C. Evolutionary dynamics of microsatellite DNA // Chromosoma. 2000. V. 109. P. 365–371. [Erratum in: Chromosoma. 2001. V. 109. P. 571].

Semino O., Magri Ch., Benuzzi G., Lin A.A., Al-Zahery N., Battaglia V., Maccioni L., Triantaphyllidis C., Shen P., Oefner P.J., Zhivotovsky L.A. *et al*. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the Neolithization of Europe and later migratory events in the Mediterranean area // Am. J. Hum. Genet. 2004. V. 74. P. 1023–1034.

Sengupta S., Zhivotovsky L.A., King R. *et al*. Polarity and temporality of high resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists // Am. J. Hum. Genet. 2006. (In press).

Shriver M.D., Jin R., Boerwinkle E. *et al*. A novel measure of genetic distance for highly polymorphic tandem repeat loci // Mol. Biol. Evol. 1995. V. 12. P. 914–920.

Slatkin M. A measure of population subdivision based on microsatellite allele frequencies // Genetics. 1995. V. 139. P. 457–462.

Stumpf M.P.H., Goldstein D.B. Genealogical and evolutionary inference with human Y chromosome // Science. 2001. V. 291. P. 1738–1742.

Tautz D. Notes on the definition and nomenclature of tandemly repetitive DNA sequences // DNA Fingerprinting: State of the Science / Eds S.D.J. Pena, J.T. Chakraborty, J.T. Epplen, A.J. Jeffreys. Birkhäuser Verlag, Basel, Switzerland, 1993. P. 21–28.

Thomas M.G., Skorecki K., Ben-Ami H. *et al*. Origins of Old Testament priests // Nature. 1998. V. 394. P. 138–140.

Underhill P.A., Shen P., Lin A.A. *et al*. Y chromosome sequence variation and the history of human populations // Nat. Genet. 2000. V. 26. P. 358–361.

Urquhart A., Kimpton C.P., Downes T.J., Gill P. Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers // Int. J. Leg. Med. 1994. V. 107. P. 13–20.

Valdes A.M., Slatkin M., Freimer N.B. *et al*. Allele frequencies at microsatellite loci: the stepwise mutation model revisited // Genetics. 1993. V. 133. P. 737–749.

Weale M.E., Yepiskoposian L., Jager R.F., Armenian Y chromosome haplotypes reveal strong regional structure within a single ethnonational group // Hum. Genet. 2001. V. 109. P. 659–674.

Weber J.L., Broman K.W. Genotyping for human whole-genome scans: past, present, and future // Advances in Genet. 2001. V. 42. P. 77–96.

Weber J., May P. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction // Am. J. Hum. Genet. 1989. V. 44. P. 388–396.

Weber J.L., Wong C. Mutation of human short tandem repeats // Hum. Mol. Genet. 1993. V. 2. P. 1123–1128.

Weir B.S. Genetic Data Analysis II. Methods for Discrete Population Genetic Data. Sinauer, Sunderland, Mass. 1996.

Xu X., Peng M., Fang Z., Xu X. The direction of microsatellite mutations is dependent upon allele length // Nat. Genet. 2000. V. 24. P. 396–399.

Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups // Genome Research. 2002. V. 12. P. 339–348.

Zegura, S.L., Karafet T.M., Zhivotovsky L.A., Hammer M.F. High resolution SNPs and microsatellite haplotypes point to a single, recent entry of native American chromosomes into the Americas // Mol. Biol. Evol. 2004. V. 21. P. 164–175.

Zhang L., Leeflang E.P., Yu J., Arnheim N. Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene // Nat. Genet. 1994. V. 7. P. 531–535.

Zhivotovsky L.A. A new genetic distance with application to constrained variation at microsatellite loci // Mol. Biol. Evol. 1999. V. 16. P. 467–471.

Zhivotovsky L.A. Estimating divergence time with the use of microsatellite genetic distances: impacts of pop-ulation growth and gene flow // Mol. Biol. Evol. 2001. V. 18. P. 700–709.
Zhivotovsky L.A., Bennett L., Bowcock A.M., Feldman M.W. Human population expansion and microsatellite variation // Mol. Biol. Evol. 2000. V. 17. P. 757–767.
Zhivotovsky L.A., Feldman M.W. Microsatellite variability and genetic distances // Proc. Natl Acad. Sci. USA. 1995. V. 92. P. 11549–11552.
Zhivotovsky L.A., Feldman M.W., Grishechkin S.A. Biased mutations and microsatellite variation // Mol. Biol. Evol. 1997. V. 14. P. 926–933.
Zhivotovsky L.A., Goldstein D.B., Feldman M.W. Genetic sampling error of distance $(\delta\mu)^2$ and variation in mutation rate among microsatellite loci // Mol. Biol. Evol. 2001. V. 18. P. 2141–2145.
Zhivotovsky L.A., Rosenberg N.A., Feldman M.W. Features of evolution and expansion of modern humans inferred from genome-wide microsatellite markers // Am. J. Hum. Genet. 2003. V. 72. P. 1171–1186.
Zhivotovsky L.A., Underhill P.A. On the evolutionary mutation rate at Y-chromosome STRs: Comments on paper by Di Giacomo *et al*. (2004) // Hum. Genet. 2005. V. 116. P. 529–532.
Zhivotovsky L.A., Underhill P.A., Cinnioğlu C. *et al*. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time // Am. J. Hum. Genet. 2004. V. 74. P. 50–61.
Ziegle J.S., Su Y., Corcoran K.P. *et al*. Application of automated DNA sizing technology for genotyping microsatellite loci // Genomics. 1992. V. 14. P. 1026–1031.

MICROSATELLITE VARIATION IN HUMAN POPULATIONS AND THE METHODS OF THEIR ANALYSIS

L.A. Zhivotovsky

N.I. Vavilov Institute of General Genetics, Moscow, Russia, e-mail: levazh@gmail.com

Summary

The paper reviews the features of microsatellite variation in human populations. It gives the nomenclature of microsatellite alleles and the approaches to their determination, describes the specificity of mutation process at microsatellite loci, and considers the dynamics of microsatellite parameters under mutation and genetic drift. Some microsatellite statistics are considered, including those related to the history of Y-chromosome haplogroups.