

НОВОЕ ОГРАНИЧЕНИЕ МИКРОСАТЕЛЛИТНЫХ МАРКЕРОВ ДЛЯ ИХ ПРИМЕНЕНИЯ В ПОПУЛЯЦИОННЫХ ИССЛЕДОВАНИЯХ (НА ПРИМЕРЕ ПАНМИКТИЧЕСКИХ ПОПУЛЯЦИЙ)

В.В. Горбачев

Институт биологических проблем севера ДВО РАН, Магадан, Россия,
e-mail: genetic2@yandex.ru

С применением коалесцентных моделей была показана возможность ошибки I рода при использовании микросателлитов в популяционно-генетических исследованиях. Установлено, что даже для панмиктических популяций величина ошибки может составлять CI 95% ~ (11–18%), что делает малоубедительным мнение о возможности универсального применения микросателлитных маркеров для оценки популяционной дифференциации. Высказывается мнение, что возможной причиной полученных результатов является высокое аллельное разнообразие, вызванное нестабильностью микросателлитных маркеров.

Ключевые слова: микросателлиты, популяция, коалесцентное моделирование, статистическая ошибка I рода.

Введение

В настоящее время многие авторы высказывают мнение о том, что микросателлитные последовательности (далее STR) являются мощными молекулярными маркерами (Sarrì *et al.*, 2006) и наиболее удобным инструментом для использования их в популяционно-генетических исследованиях (Malcolm *et al.*, 1997; Whittakera *et al.*, 2003; Gomes *et al.*, 2011). Подобное мнение основано на возможности STR-маркеров быстро накапливать замены вследствие самых высоких мутационных скоростей среди всех известных к настоящему времени молекулярно-генетических маркеров (Schlötterer *et al.*, 1998; Животовский, 2006). Высокая изменчивость очень часто позволяет получить достоверные различия между выборками, что делает их весьма «удобными молекулярными локусами». Отметим, что практически у всех маркеров рано или поздно выявлялись ограничения, которые не позволяли использовать их в качестве универсальных генетических инструментов (Avisе, 2004). Однако в настоящее время в литературных источниках встречается лишь «робкая критика» данного вида маркеров, а выявляемые недостатки в основном

относятся к области методических разработок (нуль-алели, трудности разработки новых маркеров и т. д.) (Avisе, 2004; Хрусталева и др., 2010). Широкое применение STR-локусов породило уверенность в их абсолютной универсальности, что делает необходимыми исследования, направленные на изучение области применимости микросателлитов и их потенциальных ограничений. Подобную теоретическую проверку достаточно просто провести с применением коалесцентных моделей. Коалесцентное моделирование предполагает использование ретроспективного подхода для оценки популяционных генетических процессов, позволяя проследивать моделируемые генеалогии аналогично тому, как это делается для дендрограмм. При этом коалесцентные модели могут достаточно полноценно использоваться при статистических проверках различных гипотез.

Одними из основных понятий, связанных со статистическими проверками, являются ошибки I и II рода. Статистические ошибки представляют собой либо неверное отклонение правильной нулевой гипотезы (далее H_0) в первом случае или низкую распознающую способность применяемого метода по отношению к подобной

гипотезе (во втором случае) (Sokal, Rohlf, 1994). Таким образом, если создать модель, для которой заведомо будет выполняться H_0 и только она, то любое значимое от нее отклонение (при $p < 0,05$) будет, соответственно, расценено как статистическая ошибка I рода. При выявлении достаточно высокой доли статистических ошибок (более 5%) возможно предполагать ограничение применимости метода для тех или иных целей.

Целью нашего исследования является оценка возможности применения микросателлитных маркеров для дифференциации популяций.

Теоретическое обоснование модели

Предположим полностью панмиктическую популяцию достаточно большой постоянной численности, для того чтобы нивелировать любое влияние дрейфа генов. Если по окончании моделирования произвести случайные выборки особей из разных частей этой популяции, то возможно попарно оценить, насколько они генетически дифференцированы друг от друга. В соответствии с теоретическими построениями (панмиктическая популяция) следует ожидать либо полное отсутствие дифференциации, либо такое ее малое значение, которым можно пренебречь (менее 5%). В случае если число ошибок I рода превысит пороговый предел, то возможно говорить о тенденции STR локусов к завышению значений и, как следствие, искажению получаемых результатов.

Описание модели

Моделирование проводилось с применением программы для симулирования коалесцентных процессов SimCoal 2.0 (Laval, Excoffier, 2004). Модель предполагала гаплоидную популяцию из 11 групп, каждая из которых была константной численности по 10 тыс. особей. Для каждой локальности производились самостоятельные коалесцентные процедуры. Общая продолжительность жизни популяции – 250 поколений, в течение 50 первых поколений от начала существования популяции поток мигрантов между группами составлял не более 1% на поколение (для накопления гетерогенности внутри группы). По истечении указанного числа поколений и на протяжении 200 остав-

шихся (вплоть до настоящего времени) поток мигрантов резко увеличивался и составлял не менее 99% между каждой группой. Таким образом, формировалась полностью панмиксная популяция, и из каждой группы делались случайные выборки в 11, 21, 51, 101 и 201 особь. Все выборки повторялись для 2, 5, 10, 15 и 25 микросателлитных локусов. Частота мутаций по умолчанию составляла $1,2 \times 10^{-3}$ на локус на поколение (данные темпы находятся в нижнем диапазоне мутационных скоростей микросателлитных локусов) (Malcolm *et al.*, 1997; Животовский, 2006). В среднем для 30 аллелей на локус модель предполагала пошаговые мутации (step-wise mutation model). В целом было проанализировано 25 моделей (при попарном комбинировании размера выборки и числа локусов), каждая из которых повторялась по 110 раз. Суммарно было проверено 2750 моделей. Полученные результаты моделирования использовались для расчета R_{st} по Слаткину – аналога генетических дистанций F_{st} для микросателлитных маркеров (Wright, 1951; Slatkin, 1995), с применением программы Arlequin v.3.0 (Excoffier *et al.*, 2005). Всякое значимое отклонение ($p < 0,05$) от нулевой гипотезы (H_0 – выборки в панмиктической популяции не отличаются на достоверном уровне) при попарном сравнении выборок между собой использовалось для оценки доли ошибок I рода.

Результаты моделирования

Суммарное число статистических ошибок использовалось для оценки вероятности их появления. Исходя из полученных значений рассчитывался 95%-й доверительный интервал, который находился в диапазоне от 10,9 до 17,5% ошибок I рода, средняя вероятность появления ошибки составляла 14,2% на модель. Результаты моделирования представлены в виде графа (рис.).

Из представленного выше графа легко заметить общую тенденцию к завышению неверных значений. Только в одном случае процент ошибок I рода был незначительно ниже пороговой величины (~ 4,9%), во всех остальных случаях полученные результаты превышали обозначенный 5%-й предел. Столь высокая вероятность появления статистических ошибок не позво-

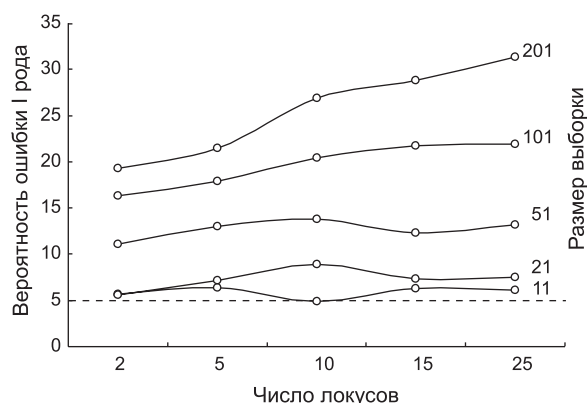


Рис. Вероятность появления ошибки I рода в %.

Пунктирной линией, параллельной оси абсцисс, отсечен диапазон значений, не превышающих 5 %.

ляет пренебрегать данной тенденцией и если не подвергает сомнению саму возможность применения **STR маркеров в популяционно-генетических исследованиях**, то делает это использование малоубедительным (для объективной оценки дифференциации популяций).

Укажем также на одну особенность полученных результатов – увеличение выборки и количества используемых локусов находится в прямой корреляции с величиной получаемой ошибки. Таким образом, вопреки бытующему мнению увеличение выборки в данном случае само по себе не способствует повышению объективности получаемых результатов.

Мы полагаем, что основной причиной является высокое аллельное разнообразие, в свою очередь, обусловленное нестабильностью аллелей микросателлитных локусов (изменчивостью). Указанная особенность, по нашему мнению, является ключевой для объяснения полученных результатов. Ведь в настоящее время известно, что **STR-маркеры являются одними из самых мутабельных локусов**, что приводит к достаточно быстрому накоплению изменений. Следует отметить, что весь выявляемый при популяционных исследованиях сигнал разделяется на популяционный (признаки, характерные для популяции) и индивидуальный (Павлинов, 2005).

Если предположить такие признаки, которые способны достаточно быстро изменяться, то по мере существования популяции индивидуальный сигнал станет превышать популяционный, что в конечном итоге приведет к увеличению филогенетического шума. Любая выборка или

несколько выборок в конечном счете будут отражать не популяционные особенности, а индивидуальные (все индивидуумы будут различаться между собой). С данным предположением согласуется и тот факт, что увеличение размера выборки, количества анализируемых локусов увеличивает вероятность ошибки I рода, что, по всей видимости, связано с возрастанием общей тенденции к появлению в выборке уникальных аллелей. Таким образом, использование нестабильных (мутабельных) аллелей в популяционных исследованиях, скорее, будет оценивать не популяционную дифференциацию, а изменчивость маркеров. Все указанные особенности во многом справедливы и для микросателлитных последовательностей.

Еще одной возможной причиной, объясняющей полученные результаты, является математический аппарат, оценивающий генетические дистанции (**Rst**). Как было показано в недавнем исследовании, **Gst и аналогичные генетические меры** оценивают вовсе не генетические расстояния между популяциями, а их генетическую изменчивость (Jost, 2008).

В целом полученные результаты позволяют говорить о новом ограничении **STR-маркеров** – завышении значений при оценке генетической дифференциации популяций, по всей видимости, связанном с аллельным разнообразием данных последовательностей. С учетом представленных результатов мы склонны считать, что указанный тип локусов подходит более для оценки биологического разнообразия, нежели популяционно-генетической дифференциации. По крайней мере, получаемые результаты всегда оставляют место для сомнения, является ли выявленная дифференциация между выборками следствием реально протекающих генетических процессов или появилась по причине использования нестабильных последовательностей.

Мы не умаляем возможности применения микросателлитов там, где это бесспорно: в криминалистике (при идентификации личностей), в сельском хозяйстве (при типировании сортов), в медицине (при установлении отцовства и материнства). Однако мы полагаем, что применение **STR-маркеров в популяционных исследованиях** требует осторожности и проверки получаемых результатов с помощью других, более стабиль-

ных, последовательностей (яДНК, мтДНК). В случае противоречивости результатов мы рекомендуем исследователям отдавать приоритет «старой доброй» мтДНК или иным, менее изменчивым, маркерам ядерного генома.

Автор выражает благодарность Л.А. Животовскому, А.Н. Татаренкову, а также рецензентам «Вавиловского журнала генетики и селекции» за существенные комментарии на ранние варианты данной рукописи.

Литература

- Животовский Л.А. Микросателлитная изменчивость в популяциях человека и методы ее изучения // Информ. вестник ВОГиС. 2006. Т. 10. № 1. С. 74–96.
- Павлинов И.Я. Введение в современную филогенетику (кладогенетический аспект). М.: Изд-во КМК, 2005. 391 с.
- Хрусталева А.М., Волков А.А., Стоклицкая Д.С. и др. Сравнительный анализ изменчивости STR- и SNP-локусов в популяциях нерки (*Oncorhynchus nerka*) Восточной и Западной Камчатки // Генетика. 2010. Т. 46. № 11. С. 1544–1555.
- Avise J.C. Molecular Markers, Natural History and Evolution. N.Y.; London: Chapman and Hall, 2004. 541 p.
- Excoffier L., Laval G., Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis // Evol. Bioinformatics Online. 2005. V. 1. P. 47–50.
- Just L. Gst and its relatives do not measure differentiation // Mol. Ecol. 2008. V. 17. P. 4015–4026.
- Gomes E.V., Breseguello L., Augusto M. *et al.* Microsatellite markers reveal genetic variation within *Sclerotinia sclerotiorum* populations in irrigated dry bean crops // Brazil J. Phytopathol. 2011. V. 159. I. 2. P. 94–99.
- Sarri V., Baldoni L., Porceddu A. *et al.* Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations // Genome. 2006. V. 49. P. 1606–1615.
- Malcolm D., Mackay T., Aquadro C.F. Low mutation rates of microsatellite loci in *Drosophila melanogaster* // Nat. Genet. 1997. V. 15. P. 99–102.
- Sokal R.R., Rohlf F.J. Biometry. 3-rd ed. Freeman and Co., San Francisco, 1994. 880 p.
- Slatkin M. A measure of population subdivision based on microsatellite allele frequencies // Genetics. 1995. V. 139. P. 457–462.
- Whittaker J., Harborda R., Boxall N. *et al.* Likelihood-based estimation of microsatellite mutation rates // Genetics. 2003. V. 164. P. 781–787.
- Schlötterer C., Ritter R., Harr B., Brem G. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates // Mol. Biol. Evol. 1998. V. 15. N 10. P. 1269–1274.
- Wright S. The genetical structure of population // Annu. Eugen. 1951. V. 15. P. 323–354.
- Laval G., Excoffier L. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history // Bioinformatics. 2004. V. 20. N 15. P. 2485–2487.

A NEW LIMITATION OF MICROSATELLITE MARKERS FOR THEIR USE IN POPULATION RESEARCH BY THE EXAMPLE OF PANMICTIC POPULATIONS

V.V. Gorbachev

Institute of Biological Problems of the North, Far East Branch of the Russian Academy of Sciences, Magadan, Russia, e-mail: genetic2@yandex.ru

Summary

Coalescent models demonstrate that type I statistical errors may occur in population genetics studies using microsatellites. Even in panmictic populations, the frequency of type I errors with 95 % confidence interval was 11–18 %, which calls in question the opinion about the universal applicability of microsatellite markers to the estimation of population differentiation. One of the possible causes of these results is the high allelic diversity due to the instability of microsatellite markers.

Key words: microsatellites, population, coalescent modeling, type I statistic error.