

УДК 575.174:599.9

Alu-ПОВТОРЫ В ГЕНОМЕ ЧЕЛОВЕКА© 2003 г. И. Ю. Хитринская, В. А. Степанов*, В. П. Пузырев¹Научно-исследовательский институт медицинской генетики Томского научного центра
Сибирского отделения Российской академии медицинских наук, Томск, 634050¹Сибирский государственный медицинский университет, Томск, 634050

Поступила в редакцию 24.12.2001 г.

Поступила после доработки 18.11.2002 г.

Высокоповторяющиеся участки ДНК занимают более 50% генома человека. Два наиболее распространенных повтора у млекопитающих – длинные (LINE) и короткие (SINE) диспергированные элементы – представлены семействами L1 и *Alu* соответственно. *Alu*-элементы представляют собой димеры сходных, но не одинаковых фрагментов с общей длиной примерно 300 п.н., происходящих от гена, кодирующего 7SL РНК. Они содержат двучленный промотор для РНК-полимеразы III, поли(А)-участок между мономерами, 3'-поли(А)-концевой участок, многочисленные CpG-островки и фланкированы короткими прямыми повторами. *Alu*-повторы составляют более 10% всего генома человека и могут перемещаться посредством ретропозиции. По-видимому, они сыграли важную роль в эволюции генома. Интеграция *Alu*-повтора в функционально значимые районы генов, а также другие вызываемые ими нарушения нормальных функций генов являются причиной некоторых наследственных заболеваний, в том числе, вероятно, они связаны с процессом канцерогенеза. Выделяют 14 подсемейств *Alu*, определяемых по диагностическим мутациям. Некоторые из них, представленные в геноме человека, полиморфны и встроились в новые локусы сравнительно недавно. Те из копий *Alu*-повторов, перемещение которых пришлось на время этнической дивергенции современного человека, являются полезными маркерами для эволюционно-генетических исследований.

Ключевые слова: *Alu*-повтор, эволюция, ретропозиция, 7SL РНК.

Повторяющаяся ДНК, как существенная часть генома эукариот, обнаружена в лаборатории Бриттена в середине 60-х годов XX века [1], где предположили, что количество ДНК-повторов пропорционально генетической сложности организма. Повторы составляют ~17% ДНК *Caenorhabditis elegans*, 67% ДНК *Nicotiana tabacum* [2], а у человека – более 50% генома [3]. По результатам полного секвенирования генома человека выделяют пять основных классов повторяющихся элементов: повторы, возникшие из транспозонов (диспергированные элементы); неактивные (частично) копии генов (процессированные псевдогены); простые прямые короткие повторы, такие как (A)_n, (CA)_n, (CGG)_n; сегментные дубликации, содержащие блоки размером в 10–300 т.п.н.; блоки тандемно повторяющихся последовательностей, такие как в центромерах, теломерах, коротких плечах акроцентрических хромосом и кластерах рибосомальных генов [3].

Большая часть из перечисленных повторов генома человека относится к первому классу (45%). Значительная часть уникальной ДНК также происходит от древних копий транспозонных элементов, но они сильно дивергировали и уже не рас-

познаются как транспозоны [4]. У млекопитающих почти все транспозоны относятся к одному из четырех типов, три из которых перемещаются через этап образования РНК-посредника (процесс ретропозиции), а один – непосредственно как ДНК. Это короткие (SINE) и длинные (LINE) диспергированные повторы, элементы с длинными концевыми повторами (LTR-транспозоны) и ДНК-транспозоны [3].

Повторы класса SINE имеются в геномах всех млекопитающих, а также других позвоночных. Большинство семейств этого класса, включая древний и ретропозиционно неактивный MIR-повтор (Mammalian-wide Interspersed Repeat) и B2-повтор грызунов, произошли от тРНК [5, 6]. Другие SINE, типа B1-повтора грызунов, такие как *Alu*-повторы, встречаются только у приматов и происходят от 7SL РНК [7]. В данном обзоре рассмотрены свойства и эволюция *Alu*-семейства, одного из наиболее обширных членов класса SINE.

**ПРОИСХОЖДЕНИЕ И СТРУКТУРА
Alu-ЭЛЕМЕНТОВ**

Alu-повторы получили свое название благодаря тому, что большинство из них содержат тетра-нуклеотид AGCT (170 п.н. от начала повтора), ко-

* Эл. почта: vadimst@img.tsu.ru

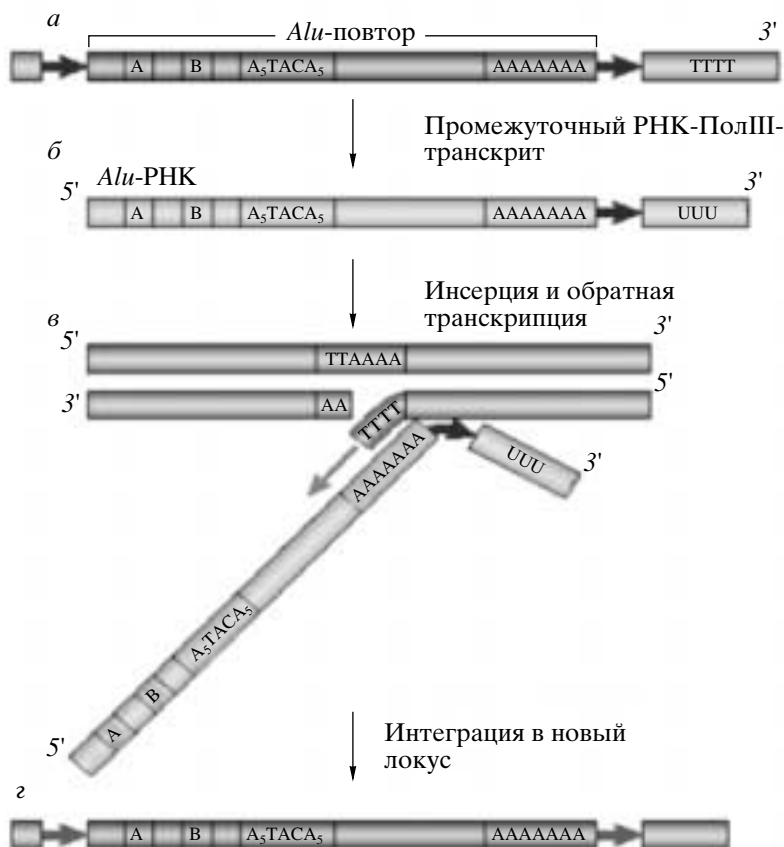


Рис. 1. Структура *Alu*-повтора в геноме человека и механизм его ретропозиции. *a* – Строение типичного *Alu*-повтора (пояснение в тексте); *б* – промежуточный РНК-Пол-III-транскрипт; *в* – пример обратной транскрипции. Направление транскрипции показано серой стрелкой; *z* – Копия *Alu*-повтора в новом участке генома [4].

торый может расщепляться рестриктазой *AluI*. *Alu*-повторы состоят из двух тандемно-расположенных форм древнего мономера (FAM), а именно, из свободного левого (FLAM) и правого (FRAM) мономеров [8]. У приматов объединение FLAM-FRAM произошло примерно 60 млн. лет назад [9]. Древний *Alu*-мономер произошел от гена 7SL РНК путем делеции 141 п.н. и появления поли(А)-участка на 3'-конце [7].

Структура *Alu*-повтора (рис. 1) [4] характеризуется несколькими особенностями. Левая половина (FLAM) длиной 140 п.н. соединена через поли(А) с более длинной правой половиной (FRAM), которая содержит дополнительно 31 п.н. [8]. В большинстве *Alu*-повторов присутствуют концевые поли(А)-участки, способствующие, вероятно, обратной транскрипции *Alu*-РНК [10]. Несоответствие в длине между правым и левым мономерами определяется делецией, произошедшей во время их эволюции от FAM [8, 11].

Левый мономер содержит два промоторных элемента для РНК-полимеразы III, блок А и блок В, каждый из которых имеет длину около 10 п.н.

[11]. Они располагаются в положениях 10–25 и 70–90 н. соответственно [12, 13]. Промотор инициирует транскрипцию из положения А, а точность инициации транскрипции определяет блок В. Предполагают [4], что для транскрипции необходим только блок В, так как среди членов Ya5 (HS) подсемейства гораздо большая гомология наблюдается именно в этом блоке.

Alu-повторы перемежаются многочисленными CpG-повторами, которых особенно много у членов молодых подсемейств *Alu*-повторов [14]. Структура *Alu*-подсемейств, а также структура членов отдельных подсемейств консервативна. *Alu*-повторы гомологичны друг другу, хотя и не полностью (различия до 24%) [4].

ЛОКАЛИЗАЦИЯ И РАСПРОСТРАНЕНИЕ В ГЕНОМЕ ЧЕЛОВЕКА

В геноме человека насчитывается 1090000 копий *Alu*-повторов, что составляет 10.6% ядерной ДНК [3]. Подобно другим SINE они часто встречаются внутри некодирующих регионов (межген-

ные регионы, интроны и др.) [15]. Значительная их доля сосредоточена в R-сегментах хромосом, где в основном располагаются тканеспецифичные гены [16]. Кроме того, высказывается предположение, что *Alu*-элементы “предпочитают” интегрироваться в регионы, богатые АТ-повторами [17].

Высокая концентрация *Alu*-повторов в насыщенных генами областях хромосомы создает широкие возможности для дубликации участков генома между отдельными повторами или для их элиминации, а также для хромосомных перестроек. Возможно, быстрая эволюция приматов связана с резким возрастанием генетического разнообразия за счет рекомбинации *Alu*-повторов [18].

В нескольких локусах обнаружены кластеры *Alu*-повторов, составленные из различных подсемейств. Возможно, эти кластеры, образовались за счет ретропозиционных событий, которые происходили в различные периоды эволюции приматов [19, 20]. *Alu*-повторы могут располагаться поодиночке, а также парами в прямой и инвертированной ориентации [21].

РЕТРОПОЗИЦИЯ *Alu*-ПОВТОРОВ

Ретропозиция *Alu*-повторов включает несколько последовательных событий. Согласно одной из предложенных моделей, транскрипция *Alu*-повторов инициируется РНК-полимеразой III и продолжается вдоль всего элемента до фланкирующей последовательности (поли(T)), где происходит терминация транскрипции (рис. 1) [4]. Для осуществления обратной транскрипции необходим праймер. Во время терминации транскрипции на 3'-конце РНК-транскрипта образуется урациловый “хвост”, который спаривается с обогащенным аденином промежуточным регионом *Alu*-повтора, и образовавшаяся шпилька служит праймером для обратной транскрипции. Затем фермент, обратная транскриптаза, связывается с праймером и синтезирует цепь ДНК, которая комплементарна РНК-транскрипту. Для интеграции кДНК в новый локус требуются разрывы в геномной ДНК (по одному в каждой цепочке ДНК) и фермент интегразы. Репарация брешей после инсерции осуществляется лигазой и заканчивается формированием прямых фланкирующих повторов [22].

В другой модели предполагается альтернативный механизм ретропозиции [23, 24], который начинается с того, что эндонуклеаза делает надрезы ДНК в последовательности 5'-ТТАААА-3'. При этом возникает серия тиминов на комплементарной нити в сайте расщепления, который способен связываться с поли(А)-участком РНК-транскрипта и начинать обратную транскрипцию [23].

Позднее действительно был выявлен общий сайт расщепления 5'(Pu)_n(Pu)_n3' длиной 6 п.н. [4].

ФУНКЦИИ *Alu*-ЭЛЕМЕНТОВ

Alu-повторы оказывают влияние на состав, организацию и экспрессию генома. Они могут усиливать транскрипцию соседнего локуса за счет своей промоторной или энхансерной активности [25, 26]. *Alu*-повторы могут также уменьшать транскрипционную активность соседних регионов, содействуя сборке нуклеосом в этой области [27, 28]. Кроме того, *Alu*-повторы способствуют метилированию близлежащих локусов, таким образом обеспечивая еще один механизм контроля экспрессии генов [29]. Хотя метилирование обычно приводит к угнетению транскрипции, есть случаи, когда метилирование *Alu*-повторов увеличивает транскрипционную активность. Например, метилированные CpG-участки, расположенные в пределах двух *Alu*-повторов в локусе GPHa (α -субъединица гликопротеидного гормона), стимулируют экспрессию этого гена [30].

В *Alu*-повторах могут, в принципе, возникать различные вторичные и третичные структуры за счет триплексных, крестообразных и других неканонических структур ДНК [16]. Можно предположить, что эти потенциально возможные трехмерные структуры могут выполнять структурирующую роль, связываясь с белками хромосом и создавая локальную пространственную структуру, удобную для функционирования генов и/или для регуляции их активности. Возможно также, что *Alu*-повторы, удаленные в линейной структуре ДНК друг от друга на достаточное расстояние, образуют между собой пространственные комплексы, что может приводить к компактизации ДНК в хромосоме [18].

Уровень транскрипции также регулируется вторичной структурой ДНК *Alu*-повторов [31]. Например, локус человека CD8- α содержит два *Alu*-повтора, которые могут связываться друг с другом, образуя крестообразную структуру и, таким образом, ингибировать транскрипцию. *Alu*-элементы, содержащиеся в пределах интронов или в 5'- и 3'-нетранслируемых регионах (UTR), могут затрагивать процессинг пре-мРНК и, тем самым, видоизменять продукт гена [32, 33]. *Alu*-повторы могут инактивировать или изменять функции генных продуктов, создавая альтернативные сайты сплайсинга или вмешиваясь в его механизм.

Эти изменения, которые часто заканчиваются серьезными генетическими последствиями, могут быть вызваны как точковыми мутациями в ранее существовавших *Alu*-элементах, так и инсерциями, образовавшимися *de novo* [17, 33, 34]. Присутствие *Alu*-повторов, так же, как и других ретропо-

зиционных элементов (B1, B2, Mir, LINE), в пре-мРНК может затрагивать полиаденилирование транскриптов, а также влиять на эффективность трансляции [6, 35, 36].

Присутствие *Alu*-повторов в экзонах может негативно влиять на экспрессию гена [37, 38]. Инсерция в кодирующую или регуляторную область гена может привести к развитию заболевания. Описано три патогенетических механизма, связанных с *Alu*-повторами [39]: ретропозиция *Alu*-повторов в гены *de novo*: инсерция *Alu*-повторов в мРНК при сплайсинге; гомологичная рекомбинация между *Alu*-повторами, которая приводит к хромосомной перестройке. Наиболее распространенным из этих механизмов является, по-видимому, первый.

Так, обнаружена инсерция *Alu*-повтора *de novo* в пятом экзоне гена фактора свертываемости крови XI у пациента с гемофилией В [40]. *Alu*-повтор длиной 322 п.н. расположен в кодирующей области и прерывает рамку считывания глутаминовой кислоты белка фактора XI, что приводит к образованию преждевременного стоп-кодона. Последовательность нуклеотидов встроенного *Alu*-повтора отличается от *Alu*-семейства Ya (HS) только одним дополнительным адениновым остатком, фланкирована прямыми повторами (15 п.н.) и содержит на 3'-конце поли(А)-участок из 78 п.н. Прямые повторы почти не содержат аденина и тимина, однако последовательность, фланкирующая повторы, состоит преимущественно из А- и Т-остатков.

В ДНК пациента с ахолиноэстераземией обнаружили инсерцию *Alu*-повтора во 2-ой экзон гена ChE. *Alu*-повтор длиной 342 п.н. включает поли(А)-участок из 38 п.н. и фланкирован дубликациями сайта мишени (15 п.н.) с обеих сторон. В нормальной последовательности гена ChE мутация отсутствует. Последовательность *Alu*-повтора имеет 93% гомологии с эволюционно самым молодым *Alu*-подсемейством человека; это позволяет предположить, что инсерция появилась у этого пациента в результате механизма ретроинверсии [41].

Аналогичный механизм генной инактивации обнаружен и при некоторых других заболеваниях, включая нейрофиброматоз типа 1, гиперпаратиреозидизм новорожденных, болезнь Гентингтона, семейную гиперкальцемию и другие. Инсерция *Alu*-повтора *de novo* в район потенциальных онкогенов и генов-супрессоров опухоли может спровоцировать процесс канцерогенеза. В частности, показано, что одной из патогенетических мутаций при лимфоме В-клеток является интеграция *Alu*-повтора в локус *Mlvi-2*.

Делеции в генах вследствие гомологичной рекомбинации между двумя *Alu*-повторами обнаружены в локусах рецептора липопротеинов низкой плотности при семейной гиперхолестеринемии, в гене α -цепи β -гексозаминазы и в гене адено-

зиндезаминазы, при тяжелом комбинированном иммунодефиците, а также в гене ингибитора С1 при наследственной ангионевротической эдеме, при некоторых других наследственных заболеваниях и хромосомных перестройках [39].

Механизм, связанный с инсерцией *Alu*-повтора в мРНК в процессе сплайсинга, достаточно редок. Описано лишь несколько патологических состояний, вызванных сохранением интегрированной части *Alu*-повтора в процессе сплайсинга мРНК. Так, обнаружена инсерция 142 нуклеотидов в месте соединения экзона 3 и 4 в мРНК гена орнитин-d-аминотрансферазы (ОАТ). Выстроившаяся последовательность представляет собой 3'-концевую часть *Alu*-повтора, присутствующего в норме в интроне 3 гена ОАТ. Трансверсия С–G в *Alu*-повторе привела к возникновению нового потенциального донорного сайта сплайсинга внутри *Alu*-элемента, активации криптического акцепторного сайта и переходом части *Alu*-элемента в зрелую мРНК ОАТ [32].

ЭВОЛЮЦИЯ ПОДСЕМЕЙСТВ

В результате анализа последовательности многочисленных *Alu*-повторов показано, что они несут большое число точечных мутаций, возникших в течение эволюции. Спектр точечных мутаций, характерный для определенного подсемейства *Alu*-повторов называют “диагностическими” мутациями. Насчитывается, по крайней мере, 14 основных подсемейств [4, 42, 43] (рис. 2), которые можно объединить в три группы: молодые, промежуточные и древние, исходя из времени начала ретропозиции. В членах древних подсемейств число диагностических мутаций невелико, наблюдается большая гомология с предковым димером (по сравнению с более молодыми группами) [42, 44]. С другой стороны, *Alu*-повторы древних подсемейств характеризуются большой гетерогенностью, снижением доли CpG-повторов, а также деградацией обеих фланкирующих последовательностей и поли(А)-участков [45]. Эти особенности связаны с возникновением многочисленных мутаций (не в диагностических позициях), которые накопились с момента ретропозиции. Исследования показали, что два древнейших подсемейства, Jo и Jb, произошли около 81 млн. лет назад [42], что согласуется со временем дивергенции между грызунами и приматами.

Подсемейства S(Sx, Sp, Sq, Sc) являются промежуточными по возрасту (48–35 млн. лет назад) и CpG-составу. Предполагают, что различия в последовательности *Alu*-повторов подсемейств Sp и Sc обусловлено их происхождением от двух различных предковых последовательностей [45, 46]. На основании анализа множества *Alu*-повторов приматов можно реконструировать древние ретропозиционные события, которые произошли

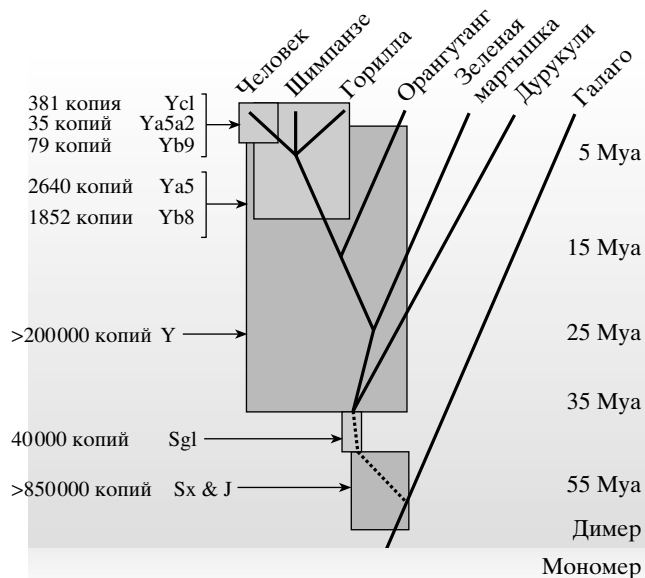


Рис. 2. Происхождение и распространение *Alu*-повторов у приматов. События экспансий *Alu*-подсемейств (Yc1, Ya5a2, Yb8, Y, Sg1, Sx и J) наложены на дерево эволюции приматов. Слева от древа показано приблизительное число копий каждого *Alu*-подсемейства. Mya (million years ago) – миллион лет назад [4].

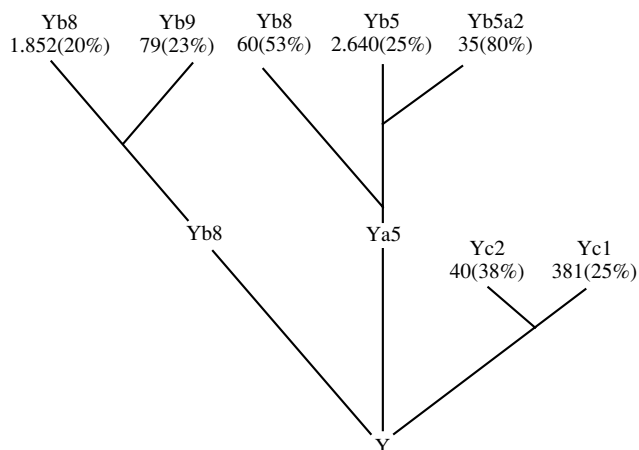


Рис. 3. Экспансия “молодых” *Alu*-подсемейств человека. Несколько подсемейств *Alu*-повторов распространились одновременно в геноме человека. Прежде всего, это три Y-подсемейства, которые названы “Ya”, “Yb” и “Yc” в соответствии со стандартной *Alu*-номенклатурой. Цифрами показано приблизительное число копий каждого подсемейства, оцененное с помощью компьютерного анализа нуклеотидной последовательности генома человека. В скобках дан процент полиморфных инсерций в каждом семействе. *Alu*-подсемейства с меньшим числом копий и более высоким уровнем полиморфизма, вероятно, являются наиболее молодыми в геноме человека [4].

свыше 30 млн. лет назад и в которые было вовлечено каждое из подсемейств (Sx, Sp, Sq) [47]. Скорость амплификации *Alu*-повторов значительно снизилась и сейчас это редкое событие, которое происходит только в группе молодых подсемейств [48].

Почти все недавно интегрировавшие в геном человека *Alu*-повторы относятся к одному из нескольких небольших и тесно взаимосвязанных “молодых” *Alu*-подсемейств (Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb9) (рис. 3) [4, 46]. Например, средний возраст молодых подсемейств Ya (ранее называли PV (Predicted Variant) или HS (Human Specific)) и *Alu*Yb8 (которое ранее называли Sb2) составляет примерно 5 млн. лет [42, 49]. Некоторые члены *Alu*-подсемейств встроились в геном человека недавно, чем обусловлен полиморфизм по их наличию или отсутствию. Такие повторы могут существовать в одной популяции, в семье, или, в случае инсерции *de novo*, у единственного человека [50]. Вероятность того, что *Alu*-повтор будет утерян первоначально, достаточно высока. Это зависит от размера популяции (чем больше популяция, тем больше вероятность, что *Alu*-повтор будет утерян), но в течение короткого промежутка времени частота может по какой-либо причине резко возрасти [4].

Для объяснения происхождения *Alu*-семейств из модифицированного РНК-транскрипта гена 7SL РНК предложено 2 модели (рис. 4): транспозонная и модель мастер-гена. Транспозонная модель говорит о том, что многие SINE-элементы способны генерировать новые, которые обладают транспозонной активностью. По мере амплификации новые последовательности все больше отличаются от первоначальной. Последовательности нуклеотидов у членов *Alu*-семейства отличаются вследствие накопления различных мутаций. Показано, что скорость амплификации не является экспоненциальной, как следовало бы ожидать, если транспозонная модель верна [13]. В действительности скорость амплификации *Alu*-элементов высоко вариабельна. За единицу скорости амплификации принято число фиксированных инсерций в 5000 лет, что примерно в 100 раз медленнее, чем 40–50 млн. лет назад.

Согласно модели мастер-гена [44], большинство SINE происходят от одного или нескольких активных участков генома. Эта модель предполагает линейную скорость амплификации, которая контролируется мастер-геном. Мутации в мастер-генах генерируют новые подсемейства и являются основной причиной различий в скорости амплификации. Например, исходный мастер-ген генерировал семейство элементов с идентичной последовательностью. Через некоторое время мастер-ген мутировал и копии этой “вторичной” последовательности образовали новое подсемей-

ство. Каждая новая мутация в мастер-гене не только генерирует новое подсемейство, но и в большинстве случаев воздействует также на скорость амплификации семейства.

Alu-ПОВТОРЫ КАК ГЕНЕТИЧЕСКИЙ МАРКЕР

Alu-повторы как генетические маркеры широко применяются для картирования генома, в клинической диагностике и для характеристики геномных перестроек. Кроме того, явление полиморфизма Alu-повторов используют для изучения разнообразия генома человека на различных уровнях, для анализа родословных и в судебной экспертизе. Они применяются также при анализе геномных перестроек, что способствует пониманию некоторых процессов эволюции человека [19, 51] и приматов [52, 53].

Некоторые свойства полиморфных Alu-повторов делают их очень удобными генетическими маркерами. Это и высокая стабильность Alu-повторов, и низкий уровень инсерций *de novo*, и отсутствие механизма их удаления из специфического локуса. С высокой степенью надежности можно рассматривать каждый случай инсерции Alu-повтора в любой из локусов как независимое событие, произошедшее лишь однажды. Кроме того, природа перемещения Alu-повторов позволяет достаточно уверенно выявлять начальное (отсутствие Alu-повтора в данном локусе) и конечное (инсерция повтора) аллельное состояние локуса. Иными словами, для Alu-повторов в отличие от других полиморфных диаллельных систем всегда известно предковое состояние и направляющие мутации. Наконец, генотипирование полиморфных Alu-повторов отличается методической простотой [13, 54].

В настоящее время полиморфные Alu-повторы, наряду с другими системами генетических маркеров (микросателлиты, митохондриальная ДНК, Y-хромосома, однонуклеотидные полиморфизмы (SNP) широко используются для анализа филогении, эволюции и характеристики структуры генофондов современных популяций человека. Исследования по генетическому разнообразию с помощью Alu-повторов ведутся в нескольких исследовательских центрах США и Европы, а также в России [49, 54–59]. Результаты позволяют охарактеризовать спектр генетического разнообразия по Alu-повторам у населения мира и подтвердить основные миграционные пути заселения территории земного шара человеком.

Так, данные по распределению Alu-повторов в популяциях соответствуют гипотезе африканского происхождения современного человека [49, 54]. Генетическое разнообразие африканских популяций выше, чем в популяциях других континентов.

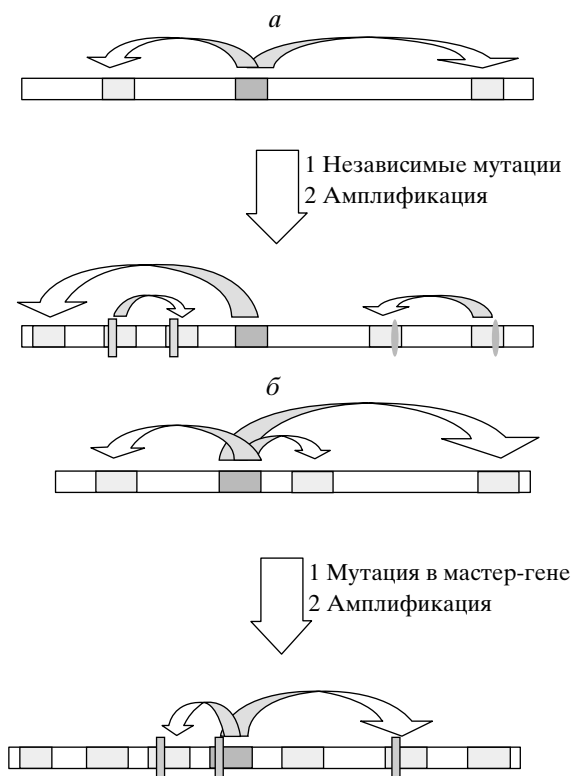


Рис. 4. Модели происхождения Alu-семейств. *а* – Транспозонная модель, *б* – модель мастергена. Круглыми стрелками показана ретропозиция Alu-повтора в новые участки генома. Черным цветом обозначен исходный Alu-повтор, серым – копии Alu-повтора, вертикальными прямоугольниками и овалами – мутации, возникающие в Alu-повторе.

Время разделения африканских и не африканских популяций, согласно данным по Alu-повторам, соответствует 137000 ± 15000 лет. Показано, что популяции Австралии и Новой Гвинеи почти так же близки к гипотетическому предку, как и африканские популяции, что, вероятно, свидетельствует о ранней экспансии современного человека в Австралазию.

Из немногочисленных примеров региональных исследований генофонда с помощью Alu-повторов можно упомянуть работы по уточнению генетических взаимоотношений между различными группами американских индейцев [60] и между некоторыми кавказскими этносами [56]; работы по реконструкции историко-генетических связей между коренными этническими группами Индостана [57], а также наши собственные исследования, посвященные реконструкции эволюционной истории населения севера Евразии [58, 59, 61].

В представленном обзоре мы попытались обобщить современные данные о происхождении и эволюции Alu-повторов, о механизмах их ретропозиции и об использовании в качестве генетиче-

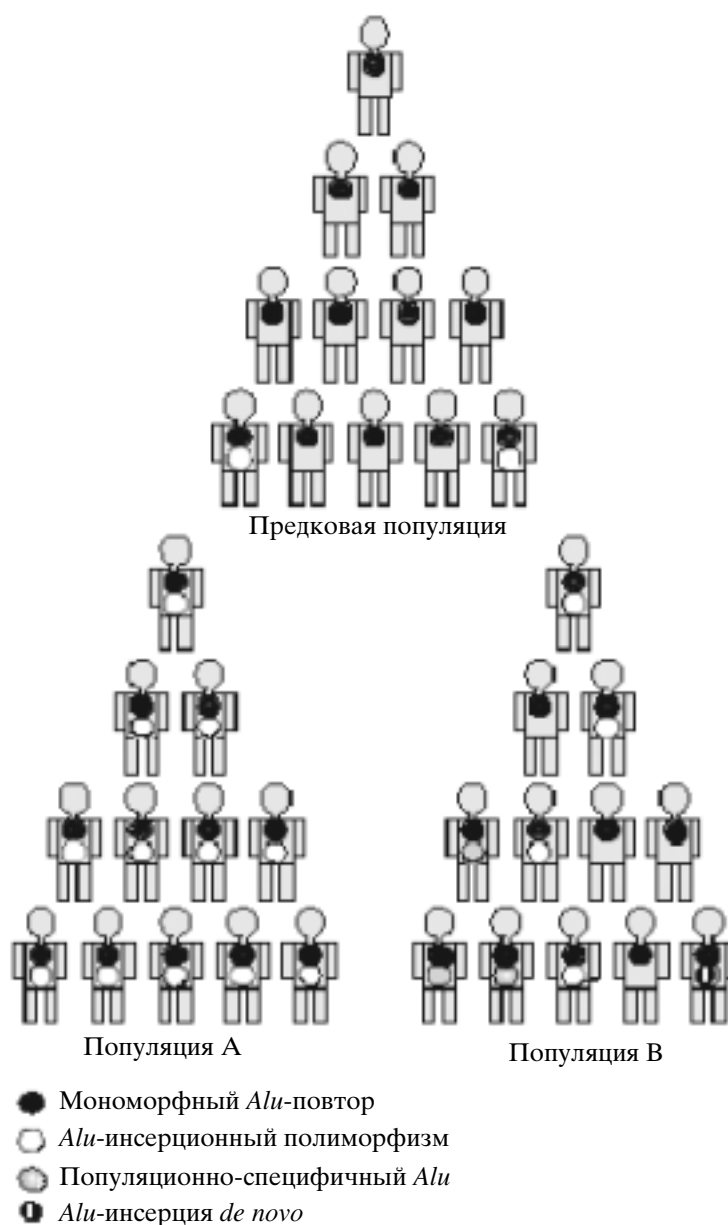


Рис. 5. Распространение *Alu*-инсерций. Сверху показана предковая популяция человека, а ниже – две дочерние субпопуляции. У всех членов популяции имеется мономорфная *Alu*-вставка – черный кружок. *Alu*-инсерции в предковой популяции и в дочерней субпопуляции – белый кружок, специфичный для конкретной популяции *Alu*-элемент – серый кружок и *Alu*-инсерции *de novo* – кружок с полоской [4].

ских маркеров в медицинской и популяционной генетике. С *Alu*-повторами связано постепенное повышение геномного разнообразия, а их амплификация привела к тому, что сформировалось самое большое семейство мобильных элементов в геноме человека. Несколько тысяч копий *Alu*-повторов встроилось в геном человека после расхождения приматов и человека, и они несут “вредные” мутации. Свой вклад в генетическое разнообразие человека вносит рекомбинация между *Alu*-повторами, что часто связано с различными

генетическими болезнями человека. Многие *Alu*-повторы влияют на экспрессию генов путем их метилирования. Хотя некоторые из *Alu*-повторов несут “вредные” мутации, подавляющее большинство недавно интегрировавшихся повторов не оказывает видимого негативного влияния на геном и, видимо, отражают нейтральные мутационные события. Именно эти полиморфные *Alu*-повторы являются удобными генетическими маркерами для изучения взаимодействия между популяциями и их происхождения.

Перспективы изучения *Alu*-повторов в геноме человека связаны, на наш взгляд, с несколькими основными направлениями. Одним из них является исследование биологических и генетических функций самих повторяющихся фракций ДНК, в том числе, и *Alu*-повторов, в качестве структурных и регуляторных элементов генома. Выявление закономерностей распределения *Alu*-повторов в геноме человека, вероятно, позволит уточнить наши представления о структурной организации генома и эволюции хромосом. Это перспективное направление исследований близко работам группы Блинова [16, 18]. *Alu*-повторы могут оказаться весьма полезным инструментом изучения проблем сравнительной генетики. Наконец, выявление полиморфных локусов *Alu*-повторов в геноме человека может пролить свет на вопросы реконструкции эволюционной истории популяций человека и их расселения по территории земного шара.

Работа получила финансовую поддержку инициативных и молодежных проектов Российского фонда фундаментальных исследований (02-04-49166, 00-04-48506, 01-04-63076, 00-15-9786, 02-04-0604); фонда Wener-Gren Foundation (6801); фонда “Базы данных о генофондах человека, животных, растений и микроорганизмов”; Федерального центра научно-технических программ “Исследования и разработки по приоритетным направлениям науки и техники 2002–2004 гг.”.

СПИСОК ЛИТЕРАТУРЫ

1. Waring M., Britten R.J. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science*. **154**, 791–794.
2. Smit A.F. 1996. The origin of interspersed repeats in the human genome. *Current Opin. Genet. Develop.* **6**, 743–748.
3. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921.
4. Batzer M.A., Deininger P.L. 2002. *Alu* repeats and human genomic diversity. *Nature Rev. Genetics*. **3**, 370–379.
5. Krayev A.S., Markusheva T.V., Kramerov D.A., Ryskov A.P., Skryabin K.G., Bayev A.A., Georgiev G.P. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res.* **10**, 7461–7475.
6. Smit A.F., Riggs A.D. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23**, 98–102.
7. Ullu E., Tschudi C. 1984. *Alu* sequences are processed 7SL RNA genes. *Nature*. **312**, 171–172.
8. Quentin Y. 1992. Origin of the *Alu* family: a family of *Alu*-like monomers gave birth to the left and the right arms of the *Alu* elements. *Nucleic Acids Res.* **20**(13), 3397–3401.
9. Zietkiewicz E., Richer C., Sinnett D., Labuda D. 1998. Monophyletic origin of *Alu* elements in primates. *J. Mol. Evol.* **42**, 172–182.
10. Economou E.P., Bergen A.W., Warren A.C., Antonarakis S.E. 1990. The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. USA*. **87**, 2951–2954.
11. Jurka J., Zuckerkandl E. 1991. Free left arms as precursor molecules in the evolution of *Alu* sequences. *J. Mol. Evol.* **33**, 49–56.
12. Knight A., Batzer M.A., Stoneking M., Tiwari H.K., Scheer W.D., Herrera R.J., Deininger P.L. 1996. DNA sequences of *Alu* elements indicate a recent replacement of the human autosomal genetic complement. *Proc. Natl. Acad. Sci. USA*. **93**, 4360–4364.
13. Nomick G.E., Batzer M.A., Deininger P.L., Herrera R.J. 1996. The mobile genetic element *Alu* in the human genome. *Bioscience*. **46**, 32–41.
14. Chen C., Gentles A.J., Jurka J., Karlin S. 2002. Genes, pseudogenes, and *Alu* sequence organization across human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA*. **99**, 2930–2935.
15. Batzer M.A., Kilroy G.E., Richard P.E., Shaikh T.H., Desselte T.D., Hoppens C.L., Deininger P.L. 1990. Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res.* **18**, 6793–6798.
16. Блинов В.М., Ресенчук С.М., Уваров Д.Л., Чирикова Г.Б., Денисов С.И., Киселев Л.Л. 1998. *Alu*-элементы генома человека. Инвариантная вторичная структура левого и правого мономеров. *Молекуляр. биология*. **32**, 84–92.
17. Bailey A.D., Shen C.K. 1993. Sequential insertion of *Alu* family repeats into specific genomic sites of higher primates. *Proc. Natl. Acad. Sci. USA*. **90**, 7205–7209.
18. Блинов В.М., Денисов С.И., Сараев Д.В., Шевцов Д.В., Уваров Д.Л., Опарина Н.Ю., Сандахчиев Л.С., Киселев Л.Л. 2001. Структурная организация генома человека, распределение нуклеотидов, *Alu*-повторов и экзонов в хромосомах 21 и 22. *Молекуляр. биология*. **35**, 1032–1038.
19. Toda Y., Tomita M., 1997. *Alu* elements as an aid in deciphering genome rearrangements. *Gene*. **205**, 173–176.
20. Kulski J.K., Gaudieri S., Bellgard M., Balmer L., Giles K., Inoko H., Dawkins R.L. 1997. The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J. Mol. Evol.* **45**, 599–609.
21. Batzer M.A., Gudi V.A., Mena J.C., Foltz D.W., Herrera R.J., Deininger P.L. 1991. Amplification dynamics of human-specific (HS) *Alu* family members. *Nucleic Acids Res.* **19**, 3619–3623.
22. Jagadeeswaran P., Forget B.G., Weissman S.M. 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? *Cell*. **26**, 141–142.
23. Jurka J., Klonowski P. 1996. Integration of retroposable elements in mammals: selection of target sites [letter]. *J. Mol. Evol.* **43**, 685–689.
24. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in the integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA*. **94**, 1872–1877.
25. Vansant G., Reynolds W.F. 1995. The consensus sequence of a major *Alu* subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. USA*. **92**, 8229–8233.

26. Britten R.J. 1997. Mobile elements inserted in the distant past have taken of important functions. *Gene*. **205**, 177–182.
27. Deininger P.L., Batzer M.A. 1999. *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193.
28. Englander E.W., Howard B.H. 1995. Nucleosome positioning by human *Alu* elements in chromatin. *J. Biol. Chem.* **270**, 10091–10096.
29. Heller H., Kammer C., Wilgenbus P., Doerfler W. 1995. Chromosomal insertion of foreign (adenovirus type 12, plasmid, or bacteriophage lambda) DNA is associated with enhanced methylation of cellular DNA segments. *Proc. Natl. Acad. Sci. USA*. **92**, 5515–5519.
30. Cox G.S., Gutkin D.W., Haas M.J., Cosgrove D.E. 1998. Isolation of an *Alu* repetitive DNA binding protein and effect of CpG methylation of binding to its recognition sequence. *Biochim. Biophys. Acta*. **1396**, 67–87.
31. Hanke J.H., Hambor J.E., Kavathas P. 1995. Repetitive *Alu* elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *J. Mol. Biol.* **246**, 63–73.
32. Mitchell G.A., Labuda D., Fontaine G., Saudubray J.M., Bonnefont J.P., Lyonnet S., Brody L.C., Stell G., Obie C., Valle D. 1991. Splice-mediated insertion of an *Alu* sequence inactivates ornithine delta-aminotransferase: a role for *Alu* elements in human mutation. *Proc. Natl. Acad. Sci. USA*. **88**, 815–819.
33. Knebelmann B., Forestier L., Drouot L., Quinones S., Chuet C., Benassy F., Saus J., Antignac C. 1995. Splice-mediated insertion of an *Alu* sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **4**, 675–679.
34. Wallace M.R., Anderson L.B., Saulino A.M., Gregory P.E., Glover T.W., Gollins F.S. 1991. A *de novo* *Alu* insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866.
35. Harendza C.J., Johnson L.F. 1990. Polyadenylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive element downstream of the open reading frame. *Proc. Natl. Acad. Sci. USA*. **87**, 2531–2535.
36. Schmid C.W. 1996. *Alu* structure, origin, evolution, significance, and function of one-tenth of human DNA. *Progr. Nucleic Acids Res. Mol. Biol.* **53**, 283–319.
37. Berquin I.M., Ahram M., Sloane B.F. 1997. Exon 2 of human cathepsin B derives from an *Alu* element. *FEBS Lett.* **419**, 121–123.
38. Szmulewicz M.N., Novick G.E., Herrera R.J. 1998. Effects of *Alu* insertions on gene function. *Electrophoresis*. **19**, 1260–1264.
39. Miki Y. 1998. Retrotranspositional integration of mobile genetic elements in human diseases. *J. Hum. Genet.* **43**, 77–84.
40. Vidaud D., Vidaud M., Bahnak B.R., Siguret V., Sanchez S., Laurian Y., Meyer D., Goosens M., Lavergne J.M. 1993. Haemophilia B due to a *de novo* insertion of a human-specific *Alu*-subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30–36.
41. Muratani K. 1991. Inactivation of the cholinesterase gene by *Alu* insertion: Possible mechanism for human genetransposition. *Proc. Natl. Acad. Sci. USA*. **88**, 11315–11319.
42. Kapitonov V., Jurka J. 1996. The age of *Alu* subfamilies. *J. Mol. Evol.* **42**, 59–65.
43. Roy-Engel A.M. 2002. Non-traditional *Alu* evolution and primate genomic diversity. *J. Mol. Biol.* **316**, 1033–1040.
44. Deininger P.L., Batzer M.A., Hutchinson C., Edgell M.H. 1992. Master genes in mammalian repetitive DNA amplification. *Trends. Genet.* **8**, 307–312.
45. Jurka J., Milosavljevic A. 1991. Reconstruction and analysis of human *Alu* genes. *J. Mol. Evol.* **32**, 105–121.
46. Roy-Engel A.M., Carrol M.L., Vogel E., Garber R.K., Nguyen S.V., Salem A.-H., Batzer M.A., Deininger P.L. 2001. *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics*. **159**, 279–290.
47. Britten R.J. 1994. Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proc. Natl. Acad. Sci. USA*. **91**, 6148–6150.
48. Shaikh T.H., Deininger P.L. 1996. The role and amplification of the HS *Alu* subfamily founder gene. *J. Mol. Evol.* **42**, 15–21.
49. Batzer M.A., Deininger P.L., Hellmann-Blumberg U., Jurka J., Labuda D., Rubin C., Schmid C.W., Zietkiewicz E., Zuckerkandl E. 1996. Standardized nomenclature for *Alu* repeats. *J. Mol. Evol.* **42**, 3–6.
50. Roy A.M. 2000. Potential gene conversion and source genes for recently integrated *Alu* elements. *Genome Res.* **10**, 1485–1595.
51. Mnukova-Fajdelova M., Satta Y., O'huigin C., Mayer W.E., Figueroa F., Klein J., 1994. *Alu* elements of the primate major histocompatibility complex. *Mamm. Genome*. **5**, 405–415.
52. Szmulewicz M.N., Andino L.M., Reategui E.P., Woolley-Barker T., Disotell T.R., Herrera R.J., 1999. An *Alu* insertion polymorphism in a baboon hybrid zone. *Am. J. Phy. Anthropol.* **109**, 1–8.
53. Hamdi H., Nishio H., Zielinski R., Dugiczy A. 1999. Origin and phylogenetic distribution of *Alu* DNA repeats: Irreversible events in the evolution of primates. *J. Mol. Biol.* **289**, 861–871.
54. Stoneking M., Fontius J.J., Clifford S.L., Soodyall H., Arcot S.S., Saha N., Jenkins T., Tahir M.A., Deininger P.L., Batzer M.A. 1997. *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* **7**, 1961–1071.
55. York D.S., Blum V.M., Low J.A., Rowold D.J., Puzyrev V.P., Saliukov V., Odinkova O., Herrera R.J., 1999. Phylogenetic signals from point mutations and polymorphic *Alu* insertions. *Genetica*. **107**, 163–170.
56. Nasidse I., Risch G.M., Robichaux M., Sherry S.T., Batzer M.A., Stoneking M. 2001. *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur. J. Hum. Gen.* **9**, 267–272.
57. Majumder P.P. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their evolutionary implications. *Eur. J. Hum. Genet.* **7**, 435–446.
58. Степанов В.А., Хитринская И.Ю., Пузырев В.П. 2001. Генетическая дифференциация населения Тувы по полиморфным *Alu*-инсерциям. *Генетика*. **37**, 563–569.

59. Хитринская И.Ю., Степанов В.А., Пузырев В.П. 2001. Анализ полиморфизма *Alu*-инсерций в бурятских популяциях. *Генетика*. **37**, 1553–1558.
60. Novick G.E., Novick C.C., Yunis J., Yunis E., Antunez de Mayolo P., Scheer W.D., Deininger P.L., Stoneking M., York D.S., Batzer M.A., Herrera R.J. 1998. Polymorphic *Alu* insertions and the Asian origin of native American populations. *Hum. Biol.* **70**, 23–39.
61. Степанов В.А., Пузырев В.П., Спиридонова М.Г., Хитринская И.Ю. 1999. Анализ полиморфизма *Alu*-инсерций в городской и сельской русской популяции Сибири. *Генетика*. **35**, 1138–1143.

Alu Repeats in the Human Genome

I. Yu. Khitrinskaya¹, V. A. Stepanov¹, and V. P. Puzyrev²

¹*Institute of Medical Genetics, Tomsk Research Center, Siberian Division, Russian Academy of Sciences, Tomsk, 634050 Russia; E-mail: vadimst@img.tsu.ru*

²*Siberian State Medical University, Tomsk, 634050 Russia*

Highly repetitive DNA sequences account for more than 50% of the human genome. The L1 and *Alu* families harbor the most common mammalian long (LINEs) and short (SINEs) interspersed elements. *Alu* elements are each a dimer of similar, but not identical, fragments of total size about 300 bp, and originate from the 7SL RNA gene. Each element contains a bipartite promoter for RNA polymerase III, a poly(A) tract located between the monomers, a 3'-terminal poly(A) tract, and numerous CpG islands, and is flanked by short direct repeats. *Alu* repeats comprise more than 10% of the human genome and are capable of retroposition. Possibly, these elements played an important part in genome evolution. Insertion of an *Alu* element into a functionally important genome region or other *Alu*-dependent alterations of gene functions cause various hereditary disorders and are probably associated with carcinogenesis. In total, 14 *Alu* families differing in diagnostic mutations are known. Some of these, which are present in the human genome, are polymorphic and relatively recently inserted into new loci. *Alu* copies transposed during ethnic divergence of the human population are useful markers for evolutionary genetic studies.