

УДК 577.21

## КОРОТКИЕ РЕТРОПОЗОНЫ И ИХ ИСПОЛЬЗОВАНИЕ В ФИЛОГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

© 2009 г. Д. А. Крамеров\*, Н. С. Васецкий

Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991

Поступила в редакцию 30.12.2008 г.

Принята к печати 19.02.2009 г.

**В обзоре приведены сведения об одном из самых распространенных классов повторяющихся последовательностей в геноме эукариот – коротких ретропозонах (short interspersed elements, или SINE). Рассмотрена структура, происхождение и поведение этих ретропозонов в геноме. Нейтральность, разнообразие и большое число этих маркеров генома делает их удобным и надежным инструментом филогенетического анализа. Описаны основные методы такого анализа, на конкретных примерах обсуждаются их достоинства и ограничения.**

**Ключевые слова:** короткие ретропозоны, SINE, эволюция, филогения.

**SHORT INTERSPERSED REPETITIVE SEQUENCES (SINEs) AND THEIR USE AS A PHYLOGENETIC TOOL, by D. A. Kramerov\*, N. S. Vassetzky (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia; \*e-mail: kramerov@eimb.ru). The data on one of the most common repetitive elements of eukaryotic genomes, short interspersed elements (SINEs), are reviewed. Their structure, origin, and functioning in the genome are discussed. The variation and abundance of these neutral genomic markers makes them a convenient and reliable tool for phylogenetic analysis. The main methods of such analysis are presented, and the potential and limitations of this approach are discussed using specific examples.**

**Key words:** short interspersed elements, SINEs, evolution, phylogeny.

### КОРОТКИЕ РЕТРОПОЗОНЫ

Рассеянные по геному повторяющиеся последовательности очень многочисленны и могут составлять до половины генома эукариотических организмов. Подавляющее большинство таких повторяющихся последовательностей представлено мобильными генетическими элементами, которые делятся на два больших класса: ДНК-транспозоны и ретротранспозоны.

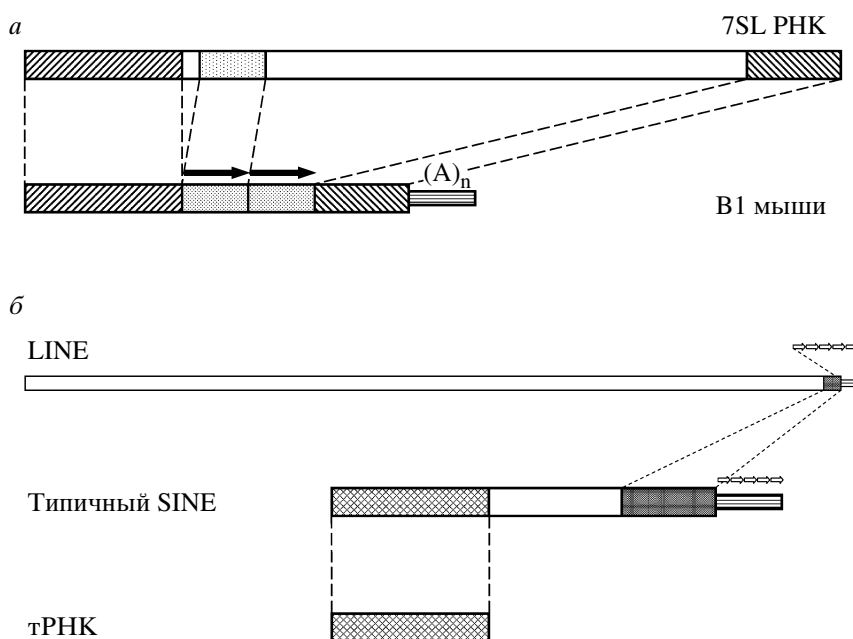
**ДНК-транспозоны** кодируют фермент транспозазу. Появление копий ДНК-транспозонов в новых местах генома происходит по механизму вырезания (*cut and paste*), т.е. транспозаза вырезает транспозон из генома, а затем катализирует его встраивание в новое место.

**Ретротранспозоны** используют совершенно иной механизм встраивания, при котором старые копии не удаляются, а копии в новых местах возникают благодаря процессу обратной транскрипции (механизм копирования, *copy and paste*). К этому классу элементов, во-первых, относятся ретротранспозоны, содержащие длинные (200–400 п.н.) концевые повторы (long terminal repeat, LTR) [1]. Такие ретротранспозоны длиной до 8 т.п.н. имеют две или три открытые рамки считывания, одна из которых кодирует обратную транскриптазу (ревертазу).

Многие из LTR-ретротранспозонов – это по сути эндогенные ретровирусы, потерявшие способность к инфицированию клеток, но способные образовывать новые провирусные копии в геноме.

Ретротранспозоны другого типа называют *длинными ретропозонами*, или LINE (long interspersed elements). Элементы этого типа имеют длину от 3 до 7 т.п.н. и не содержат длинных концевых повторов. LINE обычно содержат одну или две открытые рамки считывания, кодирующие обратную транскриптазу и РНК-связывающий белок (аналог ретровирусного Gag). Эти белки необходимы для ретропозиции данных элементов, т.е. для процесса обратной транскрипции их РНК, приводящей к образованию новых копий в геноме. К настоящему времени описано несколько десятков семейств разных LINE, которые объединяют в 15 клад [2–4]. LINE обнаружены в геномах самых разных эукариот – от беспозвоночных до человека; для одноклеточных эукариот они менее характерны [5]. Один из наиболее изученных длинных ретропозонов – элемент L1 [6]. Он присутствует в геномах всех млекопитающих, и число его копий в геноме исчисляется десятками тысяч. Нужно отметить, что подавляющее число копий L1 неполноразмерны и сильно укорочены с 5'-конца. Такие копии не способны самостоятельно размножаться, так как не кодируют полноценную обратную транскриптазу. Для их раз-

\* Эл. почта: kramerov@eimb.ru



**Рис. 1.** Структура коротких ретропозонов: B1 мыши (а) и типичный tРНК SINE (б). Стрелки обозначают повторы; другие пояснения в тексте.

множения требуется обратная транскриптаза, закодированная в немногих полноценных копиях L1. Полноценные копии называют автономными элементами, тогда как копии, потерявшие способность кодировать нормальные полипептиды, называются неавтономными. Такая ситуация наблюдается во всех семействах длинных ретропозонов.

Наконец, последний тип ретротранспозонов – *короткие ретропозоны*, или короткие рассеянные повторы (SINE, short interspersed elements). Короткие ретропозоны имеют длину от 100 до 500 п.н. Число копий разных SINE в геноме может исчисляться десятками и сотнями тысяч. Конкретные копии обычно отличаются заменами 10–30% нуклеотидов, возникшими за время существования SINE в геноме. Копии SINE, обладающие значительным сходством между собой и связанные общим происхождением, образуют семейства. В геномах млекопитающих таких семейств обычно два–четыре [7]. SINE во множестве встречаются на участках, расположенных между генами, а также в интронах, реже в экзонах (обычно в их нетранслируемых районах). SINE не кодируют полипептиды и для размножения нуждаются в обратной транскриптазе, кодируемой LINE (рис. 1) [8, 9]. Таким образом, все SINE представляют собой неавтономные мобильные элементы.

Первые известные семейства коротких ретропозонов – B1 и B2, найденные в геноме грызунов [10–13], и Alu приматов [14] – представляют два основных класса SINE, отличающихся типом нуклеотидных последовательностей, из которых они произошли. Нуклеотидные последовательности B1 и Alu

сходны с 7SL РНК, компонентом РНП-частиц, узнающих сигнальный пептид и участвующих в трансляции секреторируемых и мембранных белков [15]. При возникновении B1 произошла делеция участка длиной 180 н. 7SL РНК, дополнительные небольшие делеции и дубликации, а также многочисленные нуклеотидные замены (рис. 1а). На конце B1 имеется А-богатый участок, обычно называемый “хвостом”. Alu приматов состоит из двух сходных с B1 нуклеотидных последовательностей – левого и правого мономера. Таким образом, Alu является димерным SINE. Число семейств коротких ретропозонов, произошедших от 7SL РНК, невелико.

Элемент B2 происходит от tРНК, а именно, от аланиновой tРНК. Передний конец B2 обладает сходством с этой tРНК, далее следует район неизвестного происхождения, а на 3'-конце также располагается А-богатый хвост. Большинство известных к настоящему моменту SINE имеют аналогичный план строения (рис. 1б), хотя как tРНК-прародители, так и tРНК-неродственные районы чаще всего различаются. Этот класс коротких ретропозонов самый распространенный [7].

Наконец, существует третий класс коротких ретропозонов, происходящих от 5S рРНК. Впервые он был описан американскими исследователями у рыбы *Danio rerio* [16]. В передней части этого короткого ретропозона располагается последовательность, родственная 5S рРНК, затем следует тело SINE и его хвост. Недавно у рыб [17] и млекопитающих [18] были открыты другие представители этого класса коротких ретропозонов.

Все три типа РНК, от которых ведут свое происхождение короткие ретропозоны, синтезируются РНК-полимеразой III, а не РНК-полимеразой II, которая транскрибирует гены, кодирующие белки, в том числе гены LINE. Сами короткие ретропозоны тоже транскрибируются РНК-полимеразой III, в результате чего образуются небольшие по длине РНК [19, 20]. То, что SINE транскрибируются именно этим ферментом, неслучайно. В отличие от промотора РНК-полимеразы II, промотор РНК-полимеразы III, направляющей транскрипцию генов тРНК, 5S рРНК, 7SL РНК, а также коротких ретропозонов, располагается внутри транскрибируемой ДНК [7, 21]. Таким образом, после обратной транскрипции РНК SINE новая геномная копия имеет полноценный промотор, что делает ее способной к дальнейшим циклам транскрипции. В отличие от SINE ретропсевдогены, возникшие из белоккодирующих генов, теряют промоторы и не транскрибируются, а потому не способны к дальнейшей ретропозиции и размножению в геноме.

Несколько слов о структуре промотора РНК-полимеразы III. В SINE, произошедших из тРНК или 7SL РНК, промотор состоит из двух участков длиной по 11 п.н., называемых боксами А и В, которые разделены 30–35 п.н. В SINE, возникших из 5S рРНК, промотор состоит из трех боксов – А, IЕ и С. Боксы промотора относятся к наиболее консервативным районам коротких ретропозонов.

Сказав о промоторе, следует упомянуть и о терминаторе РНК-полимеразы III. Известно, что транскрипция, осуществляемая этой РНК-полимеразой, терминируется на участках из четырех или более Т [19, 21]. Интересно, что большинство семейств SINE не имеет собственного терминатора. Это означает, что транскрипция таких элементов продолжается за их пределы и останавливается на случайных Т-блоках в прилежащих участках ДНК.

3'-концевой район подавляющего большинства SINE млекопитающих представляет собой А-богатую последовательность. По структуре такого А-богатого хвоста все семейства коротких ретропозонов млекопитающих можно разделить на два класса [22]. У класса, в который входит В2-элемент грызунов, в состав хвоста входит терминатор транскрипции, а также гексануклеотиды ААТААА. Такие SINE относятся к классу Т<sup>+</sup>, тогда как SINE без терминатора и ААТААА – к классу Т<sup>-</sup>. Транскрипция В2 SINE заканчивается на Т-блоке, и на 3'-конце образовавшейся малой РНК синтезируется поли(А) [23]. Для такого полиаденилирования необходима последовательность ААТААА. Это уникальный пример полиаденилирования транскриптов, синтезированных РНК-полимеразой III. Поли(А), расположенная на конце транскриптов таких SINE, существенным образом увеличивает время жизни РНК в клетке и играет важную роль в процессе возникновения их новых копий в геноме.

“Хвосты” SINE у рептилий и рыб (и некоторые SINE млекопитающих) устроены несколько иначе:

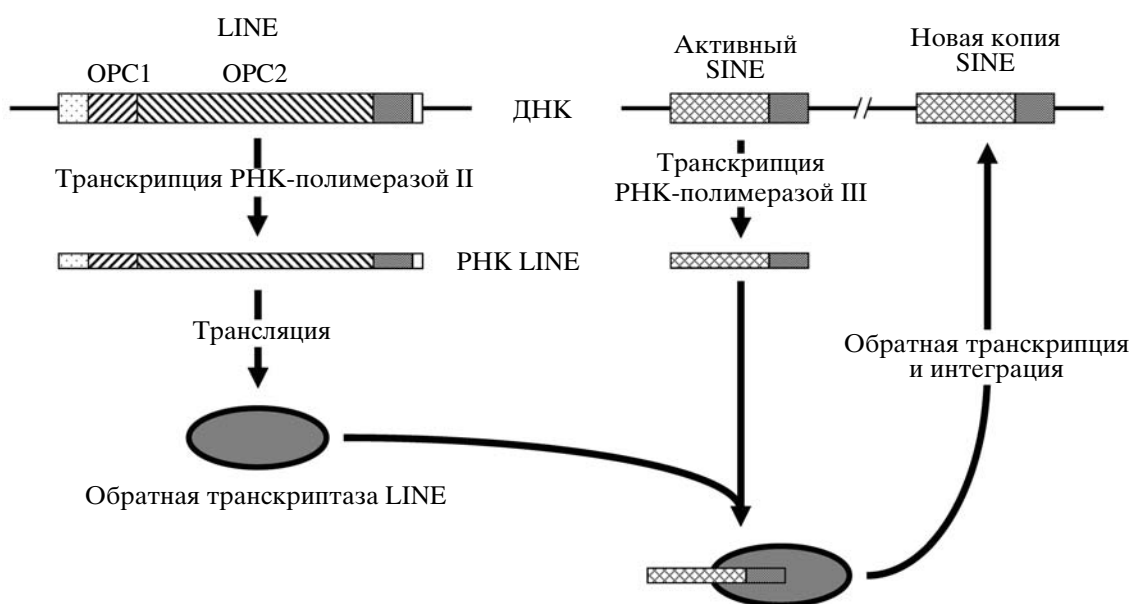
они короче и состоят из tandemных повторов длиной 3–5 н. (рис. 2). Еще одна особенность таких SINE состоит в том, что перед хвостом в них расположен участок длиной 30–100 н., гомологичный 3'-концевому району одного из LINE (рис. 2) [24], например Bov-B, CR-1, L2, RSg-1 и некоторых других. Установлено, что данный участок в РНК SINE узнается ревертазой соответствующего LINE, что приводит к обратной транскрипции этой РНК и образованию новой копии SINE. А у млекопитающих большинство семейств коротких ретропозонов, несущих А-богатый хвост, не имеет района сходства со своим LINE-партнером, а именно L1. Считается, что поли(А) или А-богатый хвост в составе РНК таких SINE выступает в качестве последовательности, узнаваемой ревертазой [25, 26].

По современным представлениям образование новых копий SINE происходит по следующему механизму [27, 28]. Комплекс РНК SINE и ревертазы LINE транспортируется в ядро, где ревертаза, обладающая также эндонуклеазной активностью, разрезает одну цепь геномной ДНК на участке, гомологичном хвосту SINE. Это приводит к тому, что хвостовая часть РНК спаривается с соответствующей цепью ДНК, 3'-конец которой служит праймером в обратной транскрипции. По завершению этой реакции и разрезания второй цепи геномной ДНК происходит достройка второй цепи и заполнение брешей клеточными ДНК-полимеразами. В результате образуется новая копия SINE, окруженная прямыми повторами длиной 8–16 п.н. (target site duplication, TSD). Вторая копия TSD образуется в результате асимметричного разрезания геномной ДНК и заполнения брешей. Почти все короткие ретропозоны окружены TSD, и они служат удобными маркерами концов копий SINE.

Несмотря на огромное число SINE в геномах, лишь единичные их копии активны и способны к размножению. Подавляющее число их производных неактивны, но они сохраняются в геноме и постепенно деградируют. С другой стороны, активность семейств коротких ретропозонов также меняется во времени. Например, Ther-1 SINE был активен у предков млекопитающих, а сейчас человеческий Ther-1 неактивен, в то время как Alu продолжает размножаться в геноме человека [29]. Более того, понятие “времени жизни” относится и к подсемействам коротких ретропозонов: так, самые старые подсемейства Alu более не размножаются; они были активны в разные периоды времени, и число копий SINE, относящихся к разным подсемействам, в геноме также значительно различается [30].

### ИСПОЛЬЗОВАНИЕ КОРОТКИХ РЕТРОПОЗОНОВ В ФИЛОГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

К началу 90-х годов прошлого века широкое распространение получили работы по построению филогенетических деревьев на основе нуклеотидных



**Рис. 2.** Схема репликации коротких ретропозонов. OPC – открытая рамка считывания (ген); другие пояснения в тексте.

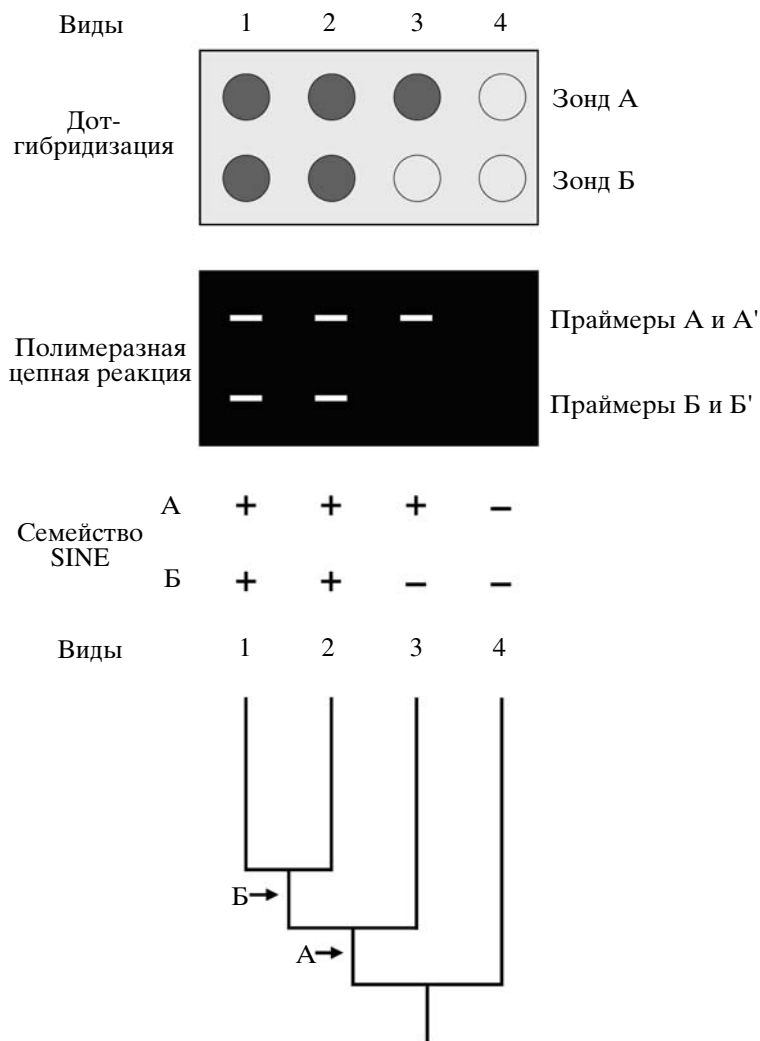
последовательностей генов или аминокислотных последовательностей кодируемых ими белков. В этих работах использовали главным образом митохондриальную ДНК. Надо сказать, что часто эти исследования приводили к парадоксальным выводам. Например, согласно одной из публикаций в журнале “Nature”, морская свинка не может быть отнесена к грызунам [31]. В последние годы с переносом акцента на ядерные гены, увеличением числа анализируемых организмов и, наконец, с развитием математических методов построения филогенетических деревьев достигнуто значительно лучшее соответствие результатов разных работ. Однако следует сказать, что множество филогенетических выводов, сделанных с помощью молекулярных методов, не согласуются с построениями, основанными на изучении морфологических признаков. Другими словами, в последние годы молекулярные систематики существенно пересмотрели традиционные взгляды на филогению тех или иных групп организмов [32–34]. И все же эволюция гена и организма могут не совпадать, и филогенетические построения, основанные на анализе генов, могут быть ошибочными. Решить эту проблему можно значительно увеличив число анализируемых генов. Однако для многих видов подобные данные недоступны, а их получение остается весьма трудоемким. Это делает востребованными другие методы, использующие независимые источники филогенетической информации.

### Метод семейств

Еще в начале 90-х мы предложили использовать SINE как источник филогенетической информации

[35–37]. Наш подход состоял в том, что нужно установить, геномы каких исследуемых видов содержат данное семейство SINE, а у каких видов его нет. Это можно сделать с помощью дот-гибридизации и полимеразной цепной реакции (ПЦР). Все виды, несущие в своих геномах SINE данного семейства, рассматриваются как более близко родственные друг другу, чем видам, в которых это семейство отсутствует. На рис. 3 виды 1, 2 и 3 содержат в своих геномах SINE семейства А, а вид 4 его не содержит. В то же время, SINE семейства Б есть у видов 1 и 2, но не у видов 3 и 4. Это позволяет построить древо, изображенное на рис. 3.

Использование этого подхода требует, чтобы в геномах организмов изучаемой таксономической группы были идентифицированы какие-либо короткие ретропозоны. Если они не были охарактеризованы ранее, то можно воспользоваться следующим сравнительно универсальным методом, предложенным нами в 1999 г. [38]. Сначала проводят ПЦР с использованием очень небольших количеств геномной ДНК и вырожденных праймеров, комплементарных боксам А и В промотора РНК-полимеразы III. В результате на многочисленных копиях SINE, предположительно представленных в данном геноме, образуется ПЦР-продукт, который можно использовать в качестве гибридационного зонда при скрининге геномной библиотеки данного или родственного вида. Секвенирование полученных клонов и анализ их нуклеотидных последовательностей позволяет идентифицировать новые семейства SINE. Наконец, зная консенсус нуклеотидной последовательности SINE, можно выбрать специфические для этой последовательности праймеры и про-



**Рис. 3.** Использование коротких ретропозонов в филогенетических исследованиях. Метод семейств. Зонды А и Б и праймеры А/А' и Б/Б' специфичны для семейств коротких ретропозонов А и Б соответственно. Стрелками на древе показаны моменты появления семейств коротких ретропозонов.

вести ПЦР с геномными ДНК целого ряда видов. Подобный анализ может быть проделан и компьютерными методами, если в банки данных депонировано достаточное число нуклеотидных последовательностей изучаемых таксонов.

Рис. 4 иллюстрирует применение этого метода для отряда грызунов. На этом “классическом” древе, взятом из монографии Romer [39], показано распределение нескольких семейств коротких ретропозонов у грызунов. Распределение одного из них, названного нами B1-dID, не полностью соответствовало представлениям о филогении грызунов. B1-dID найден у представителей трех семейств: белок (*Sciuridae*), аплодонтий (*Aplodontidae*) и сонь (*Gliridae*). На древе Romer сони располагались в ином месте – среди миоморфных (мышеподобных) грызунов (на рис. 4 отмечено звездочкой). Изучение распределения B1-dID

позволило поместить сонь в другое место на древе рядом с белками и аплодонтиями [40]. Позднее родство этих трех семейств грызунов подтвердили другие авторы при анализе ядерных генов [41].

В другом случае все началось с филогенетического анализа генов, выделившего новый надотряд млекопитающих *Afrotheria* [33]. Кроме хоботных, даманов и сирен, в него входят трубказубы, прыгунчики и тенреки со златокротами, которых ранее никогда не объединяли. Группа японских исследователей обнаружила в геномах всех этих млекопитающих семейство короткого ретропозона *AfroSINE* [42]. Это послужило надежным подтверждением правильности выделения клады *Afrotheria*, к которой систематики-морфологи относились с большим скепсисом.

Наконец, можно упомянуть еще новый надотряд *Suprprimates* (или *Euarchontoglires*) [33]. Его суще-

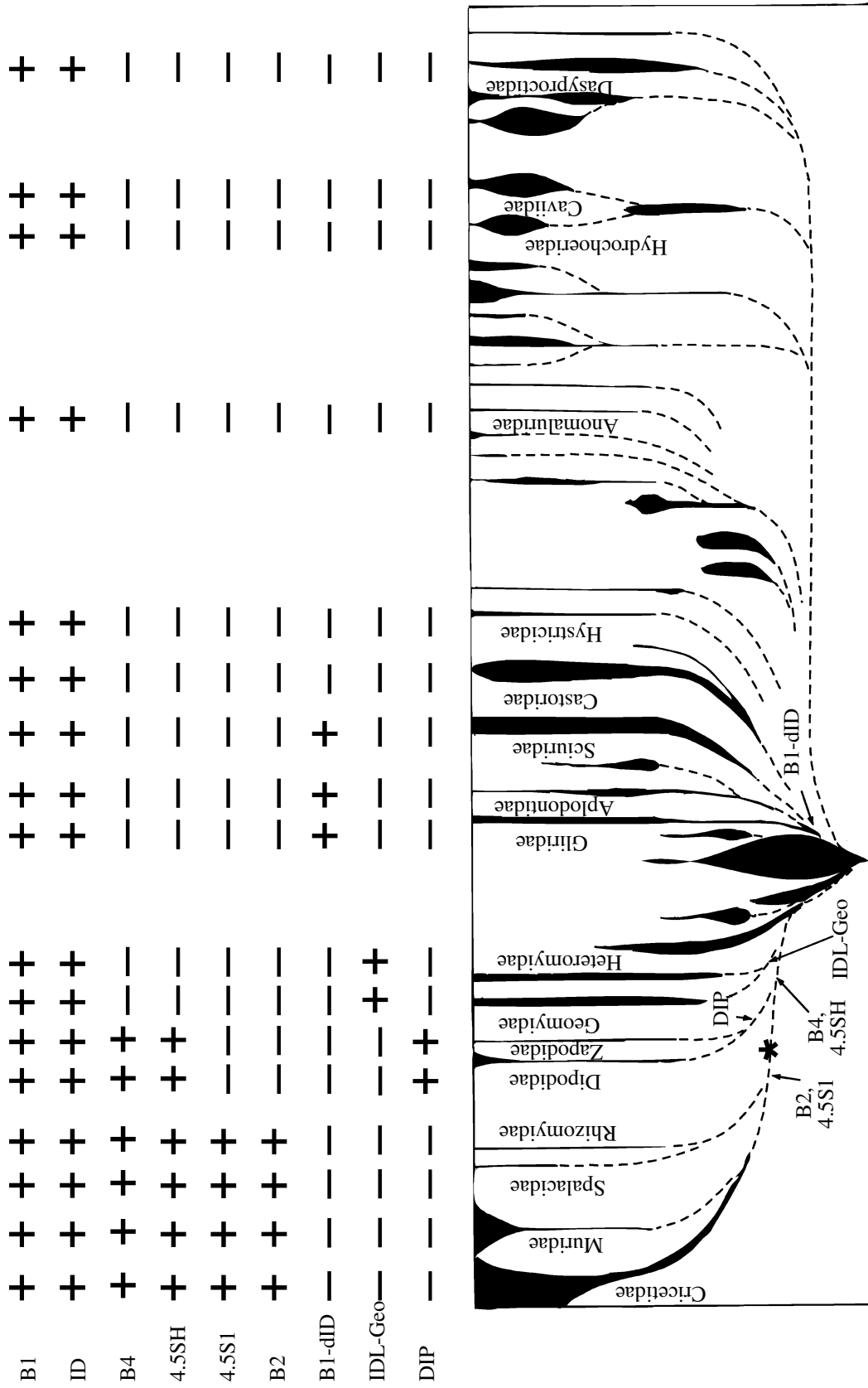


Рис. 4. Эволюционное древо грызунов. Плюсы и минусы – присутствие или отсутствие того или иного короткого ретропозона в геномах представителей соответствующих семейств грызунов [37, 40, 46, 72–74]. Стрелками на древе показаны моменты появления семейств коротких ретропозонов. Звездочкой отмечено место ветвления клады Gliridae по Romer [39].

ствование подтверждено независимо в работах японских исследователей и нашей лаборатории [43, 44]. Оказалось, что среди всех млекопитающих только представители этого надотряда, а именно грызуны, приматы и тупаи, содержат в геноме короткие ретропозоны, произошедшие от 7SL РНК. Мы, как и другие авторы [45], полагаем, что у общего предка этой группы млекопитающих, благодаря протяженной внутренней делеции, из последовательности 7SL РНК возник ретроэлемент, который эволюционировал в очень успешные SINE – B1 у грызунов и Alu у приматов.

В качестве филогенетических маркеров можно использовать не только семейства SINE, но и их подсемейства. Так, мы клонировали и секвенировали копии B1 из геномов представителей большинства семейств грызунов. Консенсусные последовательности B1 у грызунов разных семейств отличаются друг от друга заменами нуклеотидов, делециями и дупликациями. Анализ распределения этих признаков подтверждает одну из современных версий филогенетического древа грызунов [46].

### **Метод копий**

Более распространенный и разработанный подход к изучению филогении, основанный на использовании коротких ретропозонов, состоит в проверке присутствия копии SINE в конкретном локусе генома у разных видов [47]. Это можно сделать с помощью ПЦР с использованием праймеров, комплементарных нуклеотидным последовательностям, лежащим за пределами данной копии SINE (рис. 5). Амплифицированный фрагмент ДНК будет короче, если он не содержит SINE. Для построения древа необходимо исследовать по крайней мере несколько локусов, несущих SINE. Виды, содержащие копию SINE в данном локусе, филогенетически ближе друг к другу, чем к видам, у которых нет этой копии. Если доступны полные или частичные нуклеотидные последовательности геномов, то такой анализ можно провести с помощью только компьютерных методов или с использованием комбинации компьютерных и молекулярно-биологических методов.

Этот метод впервые применили для изучения филогении популяционных групп человека [48] и лососевых [49]. В дальнейшем полиморфизм (наличие/отсутствие) отдельных копий SINE Alu широко использовали для изучения филогении приматов, происхождения человека и вопросов демографии (см. обзор [30] и ссылки в нем). Впечатляющим и удивительным открытием, сделанным с помощью рассмотренного подхода, было то, что отряд китообразных (Cetacea) представляет собой внутреннюю ветвь древа отряда парнокопытных (Artiodactyla). Таким образом, гишпопотамы оказываются ближе к китам, чем к жвачным или свиньям [50, 51]. Позднее такую филогению парнокопытных и китообразных подтвердили и другими более традиционными молекулярными методами [32].

### **Интер-SINE-ПЦР**

Существует еще один метод филогенетического анализа, в котором используются короткие ретропозоны. Этот метод, называемый интер-SINE-ПЦР, заключается в ПЦР, проводимой на геномной ДНК в качестве матрицы с мечеными праймерами, комплементарными участкам SINE. Праймеры подбирают таким образом, чтобы амплифицировать участки генома, расположенные между соседними копиями SINE (рис. 6). В результате разделения продуктов ПЦР образуется набор общих и уникальных полос, и их компьютерный анализ позволяет построить эволюционные деревья. Этот метод был впервые использован для анализа эволюции парнокопытных [52]. С тех пор его с успехом применяли в филогении рукокрылых [53], насекомых [54–56], ящериц [57] и высших приматов [58].

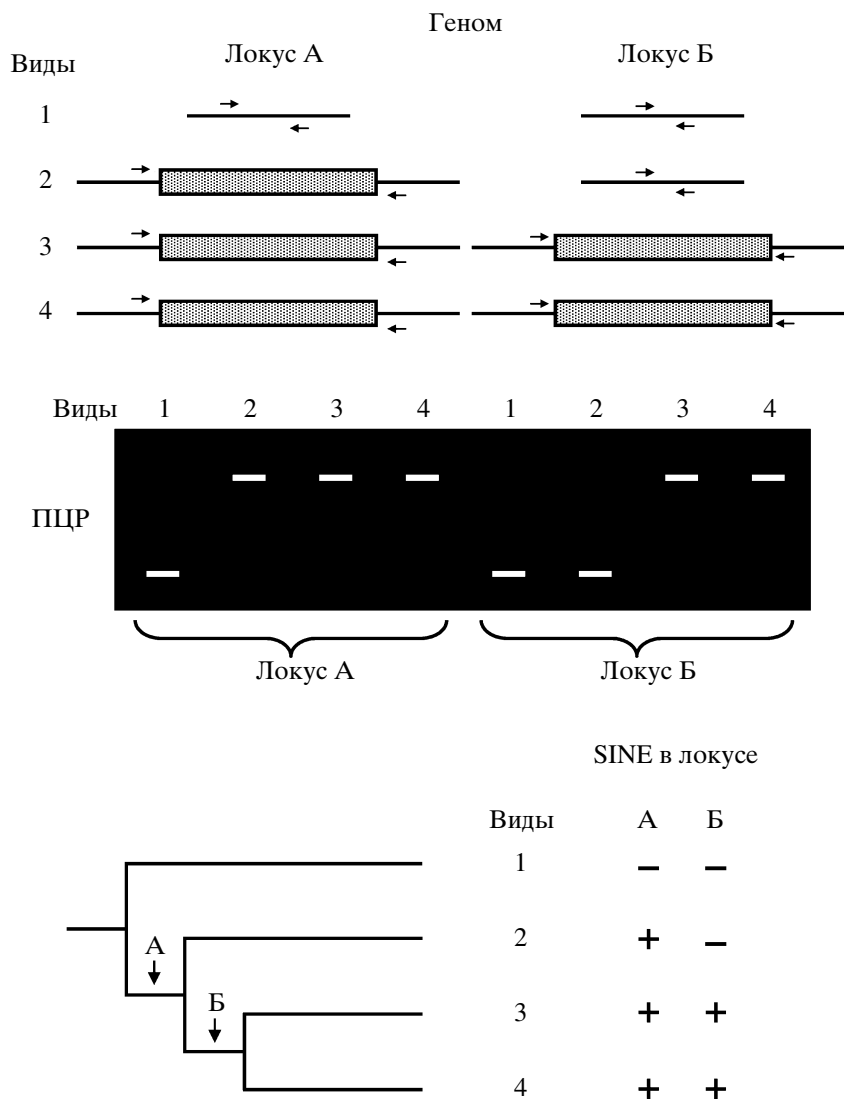
### **Достоинства и проблемы филогенетических методов, использующих короткие ретропозоны**

Хотя отдельные копии коротких ретропозонов могут быть важны для работы генов, подавляющее их большинство нефункционально, и они по сути представляют собой “геномных паразитов”. Таким образом, будучи нейтральными признаками, из всех эволюционных процессов подтвержденными только медленной деградации (за счет случайных мутаций), короткие ретропозоны могут служить очень удобными маркерами эволюции несущих их организмов. Отсутствие функции (на уровне клетки и организма) у коротких ретропозонов в целом и большинства их копий в геноме делает их признаками, не подверженными конвергентной эволюции. Это выгодно отличает их как от морфологических признаков, так и от нуклеотидных последовательностей генов.

По сути, анализ на основе коротких ретропозонов – это частный случай филогенетического анализа, использующего редкие геномные изменения (rare genomic changes), которые включают также вставки и делеции в некодирующих последовательностях, интеграции LINE и некоторых других мобильных элементов, изменение порядка генов и их дупликации, изменение генетического кода и т.п. [59]. Вместе с тем, короткие ретропозоны сравнительно небольшие и очень многокопийные, их легко детектировать. Кроме того, семейства SINE очень разнообразны, они много раз независимо возникали в эволюции (что принципиально для метода семейств). Все перечисленные свойства делают их особенно удобными и наиболее используемыми среди редких генетических изменений в филогенетических исследованиях.

С формальной точки зрения, описанные подходы к анализу эволюции организмов на основе коротких ретропозонов основаны на следующих допущениях.

1. Возникнув в геноме какого-то вида, семейство SINE передается всем видам, образовавшимся из



**Рис. 5.** Использование коротких ретропозонов в филогенетических исследованиях. Метод копий. Стрелки в верхней части рисунка обозначают праймеры, специфичные для геномных последовательностей локусов А и Б. Стрелками на древе показаны моменты интеграции коротких ретропозонов в локусы А и Б.

него. Встроившись в некий локус генома, копия SINE сохраняется в нем неопределенно долго и передается всем видам, продолжающим данную эволюционную линию.

2. Независимое возникновение одинаковых семейств коротких ретропозонов у разных видов – событие крайне маловероятное. Вероятность независимой интеграции SINE в точно тот же сайт генома разных видов очень мала.

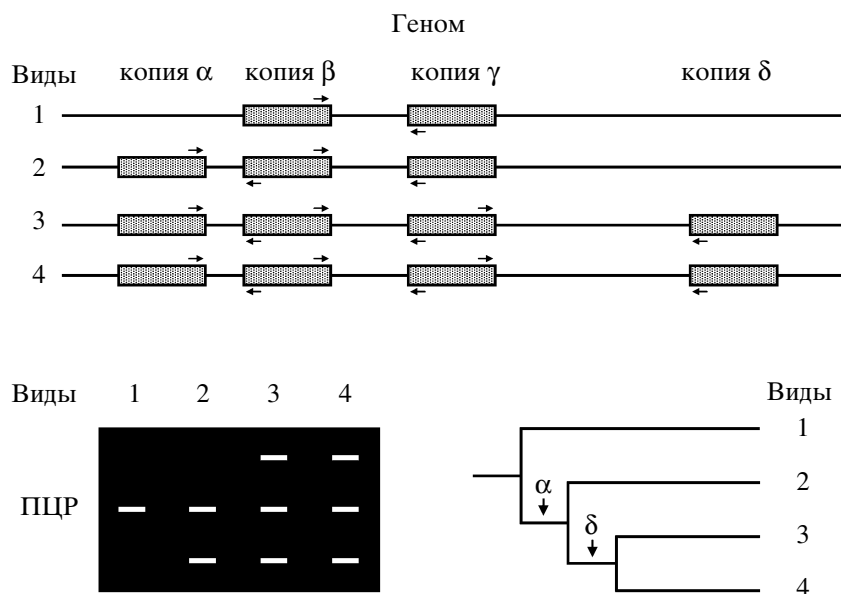
3. Семейства коротких ретропозонов не передаются горизонтально, т.е. от одних видов к другим, филогенетически им неродственным.

Данные об эволюции SINE свидетельствуют о справедливости этих допущений (с некоторыми оговорками). Так, хотя встраивание коротких ретропозонов в геном в целом носит случайный характер,

эндонуклеазы длинных ретропозонов проявляют некоторую специфичность, что увеличивает вероятность независимой интеграции SINE точно в ту же точку генома. Действительно, описано несколько случаев независимого встраивания коротких ретропозонов примерно в тот же район [60] и один случай – точно в ту же точку генома [61], однако такие события носят исключительный характер, и их влиянием на результат филогенетического анализа можно пренебречь.

Точно так же принципиально возможна элиминация конкретных копий коротких ретропозонов из генома. Такая элиминация может происходить неспецифически, т.е. захватывать районы, прилегающие к сайту встраивания (в этом случае она не влияет на результат филогенетического анализа, так как



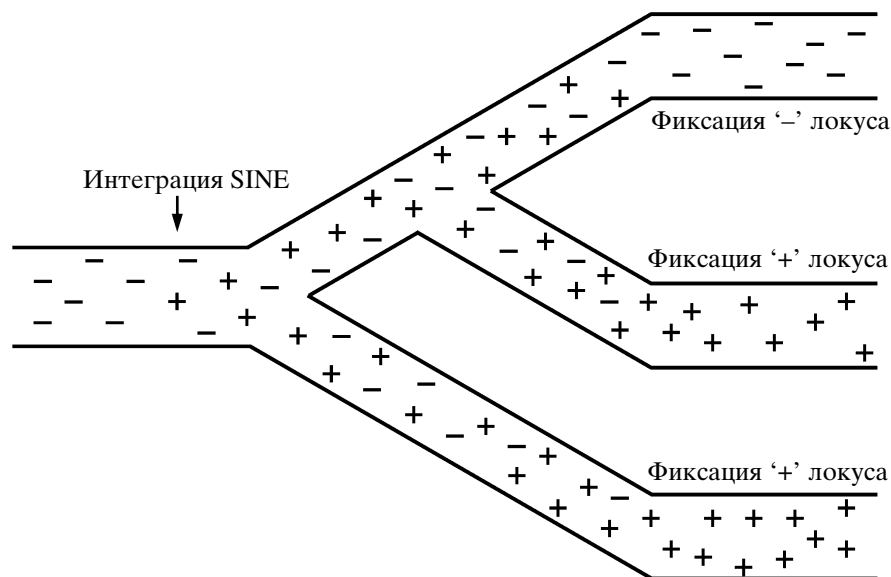


**Рис. 6.** Использование коротких ретропозонов в филогенетических исследованиях. Интер-SINE-ПЦР. Стрелки в верхней части рисунка обозначают праймеры, специфичные к SINE. Стрелками на древе показаны моменты интеграции копий  $\alpha$  и  $\delta$  данного семейства коротких ретропозонов.

ее легко обнаружить), и специфически, по всей видимости, за счет рекомбинации между TSD. По приблизительным оценкам 0.5–1% всех локусов приматов, отличающихся у разных видов наличием Alu, связано именно с такой элиминацией, а не с интеграцией копии SINE [62]. Обе эти проблемы существенны только для метода копий, впрочем, и в этом случае они легко устраняются при увеличении числа анализируемых локусов.

Короткие ретропозоны происходят из псевдогенов клеточных РНК и, чаще всего, 3'-концевой области LINE, чью ревертазу использует данный SINE. Однако их возникновение включает множество других перестроек, таких как внутренние делеции и дубликации. Особо следует отметить появление специфических нуклеотидных последовательностей неизвестного происхождения у большинства семейств коротких ретропозонов, расположенных между тРНК-родственным участком и хвостом. Все перечисленное делает практически невозможным независимое возникновение двух одинаковых или очень похожих SINE, за исключением так называемых простых коротких ретропозонов, которые состоят из последовательности, происходящей из клеточной РНК, и А-богатого хвоста [63]. Так, короткие ретропозоны ID грызунов, Vic-1 верблюдов и Das-1 броненосца, происходящие от аланиновой тРНК, имеют очень большое сходство и при этом независимое происхождение. Таким образом, простые короткие ретропозоны не следует использовать в методе семейств (к двум другим методам это не относится).

Вопрос о горизонтальном переносе ретротранспозонов регулярно обсуждается в научной литературе. Горизонтальный перенос мобильных элементов – обычное явление в случае ДНК-транспозонов и LTR-ретротранспозонов [64]. Горизонтальный перенос длинных ретропозонов также возможен, хотя, по всей видимости, присущ очень немногим семействам LINE (например, LINE клады Bov-B [18, 65]). Есть несколько публикаций, декларирующих горизонтальный перенос коротких ретропозонов (в одной из них “горизонтальный перенос SINE” даже фигурирует в заголовке [65]). На самом деле, в большинстве случаев речь идет о переносе неавтономных LINE, которые раньше причисляли к коротким ретропозонам. Горизонтальный перенос SINE также привлекали к интерпретации данных по коротким ретропозонам в случае эволюции лососевых [66]. Короткий ретропозон SmaI распространен и очень похож в геномах двух отдаленных ветвей этих рыб – сиговых, с одной стороны, кеты и горбуши, с другой, в то время как в геномах других видов его копии немногочисленны и менее сходны. Это позволило предложить [66] возможность их горизонтального переноса. Вместе с тем, на наш взгляд, существует более простое объяснение: этот короткий ретропозон был активен у предков всех лососевых и постепенно утратил активность у всех несиговых рыб, кроме кеты и горбуши, а “непохожесть” предковых копий объясняется мутационной деградацией, идущей с момента их возникновения у древних лососевых. Впрочем, та же группа исследователей предложила и другое объяснение этим данным, а именно,



**Рис. 7.** Возможный сценарий при быстром видообразовании, при котором анализ локусов коротких ретропозонов может давать неверный результат. Плюсы и минусы – присутствие или отсутствие копии SINE в определенном локусе генома в популяции.

интрогрессию [67]. Наконец, описан горизонтальный перенос коротких ретропозонов и целого ряда других ДНК лососей в геном паразита, дигенетического сосальщика *Schistosoma japonicum* [68, 69]. Описан и перенос короткого ретропозона рептилий в геном их поксвируса, также (не вполне корректно) названный “горизонтальным переносом” [70]. Активность таких перенесенных копий коротких ретропозонов не установлена и представляется крайне маловероятной, поскольку SINE – неавтономные мобильные элементы, и их размножение требует активности соответствующих автономных ретротранспозонов – LINE. Таким образом, описанные случаи переноса ДНК SINE не могут быть названы горизонтальным переносом короткого ретропозона (пока не будет доказано его размножение в геноме реципиента). В любом случае, подобный перенос ДНК SINE никак не может повлиять на результаты филогенетического анализа, проведенного с использованием коротких ретропозонов.

Однако существует ситуация, в которой результаты анализа на основе SINE могут не соответствовать настоящему дереву видов. Речь идет о явлении, называемом неполной сортировкой линий: при очень быстром видообразовании исходный полиморфизм признаков (например, локализация SINE в конкретных локусах) сохраняется некоторое время после разделения видов и их закрепление происходит позже (рис. 7). С такой ситуацией столкнулись японские исследователи, изучавшие эволюцию цихлид – рыб африканских озер [71]. Подобные результаты можно ожидать в случае межвидовой гибридизации. Стоит лишь отметить, что подобные случаи

исключительны и представляют большую проблему для всех методов филогенетического анализа.

Еще одно ограничение применения коротких ретропозонов в филогенетических исследованиях состоит в том, что с их помощью можно изучать набор только тех видов, где распространено данное семейство SINE. Можно сказать, что главное применение этих методов связано с проверкой спорных эволюционных гипотез, хотя это не исключает их использования и для построения полноценной филогении некоторых таксонов.

Несмотря на эти ограничения, описанные методы представляют собой удобный и весьма надежный инструмент филогенетических исследований. За сравнительно короткое время существования этого подхода он сумел получить признание специалистов в области эволюционной биологии, традиционно скептически относящихся к новым методам. Более того, он позволил внести несколько принципиальных исправлений в традиционную филогению млекопитающих.

Активная реализация проектов секвенирования геномов открывает большие перспективы для использования SINE в филогенетических исследованиях. Секвенирование даже одного генома в ряду рассматриваемых видов предоставляет огромный выбор локусов с короткими ретропозонами для анализа методом копий. Хотя до сих пор этот подход использовали, главным образом, для изучения эволюции млекопитающих, можно ожидать, что он найдет применение и в исследованиях эволюции других позвоночных, а также беспозвоночных и растений.

## СПИСОК ЛИТЕРАТУРЫ

1. Havecker E.R., Gao X., Voytas D.F. 2004. The diversity of LTR retrotransposons. *Genome Biol.* **5**, 225.
2. Malik H.S., Eickbush T.H. 2000. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. *Genetics.* **154**, 193–203.
3. Lovsin N., Gubensek F., Kordis D. 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol. Biol. Evol.* **18**, 2213–2224.
4. Burke W.D., Malik H.S., Rich S.M., et al. 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* **19**, 619–630.
5. Eickbush T.H., Jamburuthugoda V.K. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* **134**, 221–234.
6. Ostertag E.M., Kazazian H.H., Jr. 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538.
7. Kramerov D.A., Vassetzky N.S. 2005. Short retroposons in eukaryotic genomes. *Int. Rev. Cytol.* **247**, 165–221.
8. Kajikawa M., Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell.* **111**, 433–444.
9. Dewannieux M., Esnault C., Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48.
10. Kramerov D.A., Grigoryan A.A., Ryskov A.P., et al. 1979. Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. *Nucleic Acids Res.* **6**, 697–713.
11. Haynes S.R., Toomey T.P., Leinwand L., et al. 1981. The Chinese hamster Alu-equivalent sequence: a conserved highly repetitive, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. *Mol. Cell Biol.* **1**, 573–583.
12. Krayev A.S., Kramerov D.A., Skryabin K.G., et al. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res.* **8**, 1201–1215.
13. Krayev A.S., Markusheva T.V., Kramerov D.A., et al. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res.* **10**, 7461–7475.
14. Deininger P.L., Jolly D.J., Rubin C.M., et al. 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* **151**, 17–33.
15. Ullu E., Tschudi C. 1984. Alu sequences are processed 7SL RNA genes. *Nature.* **312**, 171–172.
16. Kapitonov V.V., Jurka J. 2003. A Novel Class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol.* **20**, 694–702.
17. Nishihara H., Smit A.F., Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16**, 864–874.
18. Gogolevsky K.P., Vassetzky N.S., Kramerov D.A. 2008. Bov-B-mobilized SINEs in vertebrate genomes. *Gene.* **407**, 75–85.
19. Haynes S.R., Jelinek W.R. 1981. Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from Alu-type dispersed repeats in Chinese hamster DNA are also found *in vivo*. *Proc. Natl. Acad. Sci. USA.* **78**, 6130–6134.
20. Kramerov D.A., Lekakh I.V., Samarina O.P., et al. 1982. The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A) + RNA. *Nucleic Acids Res.* **10**, 7477–7491.
21. Schramm L., Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620.
22. Borodulina O.R., Kramerov D.A. 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mamm. Genome.* **12**, 779–786.
23. Borodulina O.R., Kramerov D.A. 2008. Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA.* **14**, 1865–1873.
24. Ohshima K., Hamada M., Terai Y., et al. 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol. Cell Biol.* **16**, 3756–3764.
25. Dewannieux M., Heidmann T. 2005. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J. Mol. Biol.* **349**, 241–247.
26. Dewannieux M., Heidmann T. 2005. Role of poly(A) tail length in Alu retrotransposition. *Genomics.* **86**, 378–381.
27. Luan D.D., Korman M.H., Jakubczak J.L., et al. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell.* **72**, 595–605.
28. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA.* **94**, 1872–1877.
29. International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome. *Nature.* **409**, 860–921.
30. Batzer M.A., Deininger P.L. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379.
31. D'Erchia A.M., Gissi C., Pesole G., et al. 1996. The guinea-pig is not a rodent. *Nature.* **381**, 597–600.
32. Murphy W.J., Eizirik E., O'Brien S.J., et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science.* **294**, 2348–2351.
33. Murphy W.J., Eizirik E., Johnson W.E., et al. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature.* **409**, 614–618.
34. Dunn C.W., Hejnal A., Matus D.Q., et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature.* **452**, 745–749.
35. Сердобова И.М., Крамеров Д.А. 1993. Использование ретропозона B2 для изучения филогенетического родства у грызунов. *Генетика.* **29**, 1969–1981.
36. Сердобова И.М., Крамеров Д.А. 1994. Использование коротких ретропозонов в качестве филогенетических маркеров. *Докл. РАН.* **335**, 664–667.
37. Serdobova I.M., Kramerov D.A. 1998. Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. *J. Mol. Evol.* **46**, 202–214.
38. Borodulina O.R., Kramerov D.A. 1999. Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett.* **457**, 409–413.
39. Romer A.S. 1966. *Vertebrate paleontology*. Chicago: University Chicago Press.

40. Kramerov D., Vassetzky N., Serdobova I. 1999. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol. Biol. Evol.* **16**, 715–717.
41. Adkins R.M., Walton A.H., Honeycutt R.L. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol. Phylogenet. Evol.* **26**, 409–420.
42. Nikaido M., Nishihara H., Hukumoto Y., et al. 2003. Ancient SINEs from African endemic mammals. *Mol. Biol. Evol.* **20**, 522–527.
43. Nishihara H., Terai Y., Okada N. 2002. Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol. Biol. Evol.* **19**, 1964–1972.
44. Vassetzky N.S., Ten O.A., Kramerov D.A. 2003. B1 and related SINEs in mammalian genomes. *Gene*. **319**, 149–160.
45. Kriegs J.O., Churakov G., Jurka J., et al. 2007. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet.* **23**, 158–161.
46. Veniaminova N.A., Vassetzky N.S., Kramerov D.A. 2007. B1 SINEs in different rodent families. *Genomics*. **89**, 678–686.
47. Nishihara H., Okada N. 2008. Retroposons: genetic footprints on the evolutionary paths of life. *Methods Mol. Biol.* **422**, 201–225.
48. Perna N.T., Batzer M.A., Deininger P.L., et al. 1992. Alu insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.* **64**, 641–648.
49. Murata S., Takasaki N., Saitoh M., et al. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA*. **90**, 6995–6999.
50. Shimamura M., Yasue H., Ohshima K., et al. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*. **388**, 666–670.
51. Nikaido M., Rooney A.P., Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA*. **96**, 10261–10266.
52. Buntjer J.B. 1997. *Mammalian species identification by interspersed repeat PCR fingerprinting*. Utrecht Netherlands: Utrecht Univ.
53. Банникова А.А., Матвеев В.А., Крамеров Д.А. 2002. Опыт использования интер-SINE-ПЦР в изучении филогенеза млекопитающих. *Генетика*. **38**, 853–864.
54. Bannikova A.A., Bulatova N.S., Krysanov E.Y., et al. 2003. DNA Polymorphism within *Sorex araneus* and two congeneric species as inferred from inter-SINE-PCR. *Mammalia*. **67**, 263–274.
55. Bannikova A.A., Lebedev V.S., Kramerov D.A., et al. 2006. Phylogeny and systematics of the *Crociodura suaveolens* species group: corroboration and controversy between nuclear and mitochondrial DNA markers. *Mammalia*. **70**, 106–119.
56. Shafer A.B., Stewart D.T. 2007. Phylogenetic relationships among Nearctic shrews of the genus *Sorex* (Insectivora, Soricidae) inferred from combined cytochrome b and inter-SINE fingerprint data using Bayesian analysis. *Mol. Phylogenet. Evol.* **44**, 192–203.
57. Гречко В.В., Банникова А.А., Рябинина Н.Л., et al. 2006. Молекулярно-генетическое разнообразие комплекса ящериц *Darevskia raddei* (Lacertidae: Sauria): ранние этапы видообразования. *Молекуляр. биология*. **41**, 839–851.
58. Рябинина Н.Л., Банникова А.А., Шереметьева В.А., et al. 2008. Анализ ДНК высших приматов с помощью inter-SINE-ПЦР. *Генетика*. **44**, 315–322.
59. Rokas A., Holland P.W. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15**, 454–459.
60. Roy-Engel A.M., Carroll M.L., El-Sawy M., et al. 2002. Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316**, 1033–1040.
61. Rothenburg S., Eiben M., Koch-Nolte F., et al. 2002. Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*. *J. Mol. Evol.* **55**, 251–259.
62. van de Lagemaat L.N., Gagnier L., Medstrand P., et al. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15**, 1243–1249.
63. Borodulina O.R., Kramerov D.A. 2005. PCR-based approach to SINE isolation: simple and complex SINEs. *Gene*. **349**, 197–205.
64. Hua-Van A., Le Rouzic A., Maisonhaute C., et al. 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenet. Genome Res.* **110**, 426–440.
65. Kordis D., Gubensek F. 1995. Horizontal SINE transfer between vertebrate classes. *Nat. Genet.* **10**, 131–132.
66. Hamada M., Kido Y., Himberg M., et al. 1997. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the SmaI family of repeats: possible evidence for the horizontal transfer of SINEs. *Genetics*. **146**, 355–367.
67. Takasaki N., Yamaki T., Hamada M., et al. 1997. The salmon SmaI family of short interspersed repetitive elements (SINEs): interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. *Genetics*. **146**, 369–380.
68. Melamed P., Chong K.L., Johansen M.V. 2004. Evidence for lateral gene transfer from salmonids to two Schistosoma species. *Nat. Genet.* **36**, 786–787.
69. Matveev V., Nishihara H., Okada N. 2007. Novel SINE families from salmon validate *Parahucho* (Salmonidae) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3'-tails of other SINEs. *Mol. Biol. Evol.* **24**, 1656–1666.
70. Piskurek O., Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc. Natl. Acad. Sci. USA*. **104**, 12046–12051.
71. Terai Y., Takahashi K., Nishida M., et al. 2003. Using SINEs to probe ancient explosive speciation: “hidden” radiation of African cichlids? *Mol. Biol. Evol.* **20**, 924–930.
72. Gogolevskaya I.K., Kramerov D.A. 2002. Evolutionary history of 4.5SI RNA and indication that it is functional. *J. Mol. Evol.* **54**, 354–364.
73. Gogolevskaya I.K., Koval A.P., Kramerov D.A. 2005. Evolutionary history of 4.5SH RNA. *Mol. Biol. Evol.* **22**, 1546–1554.
74. Gogolevsky K.P., Kramerov D.A. 2006. Short interspersed elements (SINEs) of the Geomyoidea superfamily rodents. *Gene*. **373**, 67–74.