

УДК 577.21

## ПОВТОРЯЮЩИЕСЯ ПОСЛЕДОВАТЕЛЬНОСТИ ГЕНОМА ТУПАЙИ (MAMMALIA, SCANDENTIA)

© 2006 г. О. А. Тен, О. Р. Бородулина, Н. С. Васецкий, Н. Ю. Опарина, Д. А. Крамеров\*

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

Поступила в редакцию 21.07.2005 г.

Клонированы, секвенированы и изучены копии двух высокоповторяющихся последовательностей генома обыкновенной тупайи (*Tupaia glis*). Одна из них, Tu III, длиной около 260 п.н. относится к коротким ретропозонам, или коротким рассеянным по геному элементам (SINE). 5'-концевой район этого короткого ретропозона, видимо, ведет свое происхождение от глициновой тРНК. Tu III содержит протяженные пурино- и пиримидино-богатые участки, которые вносят существенный вклад в образование вторичной структуры РНК Tu III. Этот короткий ретропозон присутствует также в геноме гладкохвостой тупайи, относящейся к роду *Dendrogale*. Tu III, по-видимому, свойствен только отряду Scandentia (тупайи), так как он не обнаружен в ДНК других млекопитающих. Вторая повторяющаяся последовательность, Tu-SAT1, представляет собой тандемный повтор с мономером длиной около 365 п.н. Ряд черт Tu-SAT1 указывает на то, что это центромерная сателлитная ДНК.

**Ключевые слова:** тупайя, Scandentia, ретротранспозон, короткий ретропозон, SINE, тандемные повторы, сателлитная ДНК, вторичная структура РНК.

REPETITIVE SEQUENCES OF THE TREE SHREW GENOME (MAMMALIA, SCANDENTIA), by O. A. Ten, O. R. Borodulina, N. S. Vassetzky, N. Yu. Oparina, D. A. Kramerov\* (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia; \*e-mail: kramerov@eimb.ru). Copies of two repetitive elements of the genome of common tree shrew (*Tupaia glis*) were cloned and sequenced. The first element, Tu III, is a ~260 bp long short interspersed element (SINE) with the 5'-end derived from glycine RNA. Tu III carries long polypurine- and polypyrimidine-rich tracts, which may contribute to the specific secondary structure of Tu III RNA. This SINE was also found in the genome of smooth-tailed tree shrew of another genus (*Dendrogale*). Tu III seems to be confined to the order Scandentia (tree shrews) since it was not found in DNA of other tested mammals. The second element Tu-SAT1 is a tandem repeat with a monomer length of 365 bp. Some properties of its nucleotide sequence suggest that Tu-SAT1 is a centromeric satellite.

**Key words:** tree shrew, Scandentia, retrotransposon, short retroposon, SINE, tandem repeats, satellite DNA, RNA secondary structure.

Повторяющиеся последовательности составляют значительную долю геномов эукариот и подразделяются на два основных класса – тандемных и рассеянных по геному повторов. Тандемные повторы (сателлитная ДНК) занимают протяженные области в центромерных и прицентромерных частях хромосом. Длина повторяющейся единицы (мономер) сателлитной ДНК может варьировать у разных видов от 2 до 200 п.н. и более [1, 2]. Число мономеров сателлитных последовательностей в геноме обычно исчисляется миллионами копий. В пределах одного биологического вида копии мономеров значительно меньше отличаются друг от друга, чем от сателлитных мономеров, выделяемых из родственных видов.

Это происходит благодаря процессу так называемой согласованной (concerted) эволюции, основанной на неравном кроссинговере [3, 4]. В эволюции происходит довольно быстрая смена одних сателлитных ДНК другими, поэтому виды, принадлежащие разным родам (например, *Rattus* и *Mus*), могут не иметь общих сателлитных последовательностей [5, 6]. Хотя значение сателлитной ДНК остается не вполне ясным, она, вероятно, вовлечена в формирование хромосомных центромер и, тем самым, должна играть определенную роль в митозе и мейозе [7, 8].

Рассеянные по геному повторяющиеся последовательности представлены, главным образом, различными видами мобильных генетических элементов, среди которых следует отметить ДНК-транспозоны, ретротранспозоны (или LTR-содержащие транспозоны), а также длинные (ДР) и короткие ретропозоны (КР), более известные

Принятые сокращения: КР – короткие ретропозоны; ДР – длинные ретропозоны; SINE – (short interspersed element) короткий рассеянный по геному элемент.

\*Эл. почта: kramerov@eimb.ru

под названиями LINE (long interspersed elements) и SINE (short interspersed elements) соответственно [9, 10]. Мобильные элементы, за исключением ДНК-транспозонов, используют для размножения своих копий процесс обратной транскрипции. Причем, если ретротранспозоны и ДР кодируют свои собственные обратные транскриптазы, то КР используют обратную транскриптазу ДР. В геноме млекопитающих имеется лишь одно семейство активных ДР – L1 (не считая Bov-B крупного рогатого скота), тогда как КР разнообразны, и к настоящему времени известно не менее 20 семейств активных КР млекопитающих. КР (100–400 п.н.) значительно меньше, чем полно-размерные копии других мобильных генетических элементов (4–7 т.п.н.). Обычно геном содержит от одного до трех семейств КР [11], и каждое семейство, как правило, представлено  $10^5$ – $10^6$  копиями, что заметно больше этой величины для семейств других мобильных генетических элементов. За исключением Alu-элементов приматов и B1-элементов грызунов, которые произошли от 7SL РНК (малой цитоплазматической РНК, участвующей в синтезе секретируемых белков), и SINE3 полосатого данио, возникшего из 5S рРНК, КР ведут свое происхождение от тех или иных видов тРНК. На 5'-конце таких КР располагается участок, обладающий сходством с тРНК. Затем следует участок неизвестного происхождения, специфичный для каждого семейства КР. На 3'-конце КР находится А-богатая последовательность (млекопитающие и растения) или участок, состоящий из нескольких очень коротких прямых повторов (рептилии и рыбы). Многие свойства КР, как и других мобильных генетических элементов, свидетельствуют об их эгоистической или паразитической природе [12]. Однако существование каждого семейства КР в геноме в течение многих миллионов лет делает КР важными факторами эволюции. Например, встраиваясь около генов или в их интроны, они могут изменять транскрипцию и сплайсинг генов [13]. КР могут играть важную роль в защите клеток от стресса [14].

КР оказались очень хорошими маркерами для изучения филогенеза. Маркерами могут служить как целые семейства КР [15], так и их индивидуальные копии, интегрированные в те или иные локусы генома [16, 17]. Этот подход основан на том, что возникновение одинаковых семейств КР или интеграция КР в одно и то же место генома – события крайне маловероятные. КР использовали для изучения филогенеза человека [18], грызунов [19], парнокопытных и китов [16], млекопитающих групп Afrotheria [20] и Euarchontoglires [21].

В настоящей работе клонированы и изучены повторяющиеся последовательности, наиболее представленные в геноме тупайи (*Tupaia glis*). Одно из семейств – сателлитная ДНК, другое – КР.

Изучение генома тупай и их повторяющихся последовательностей представляет несомненный интерес. Первоначально тупай относили к отряду насекомоядных. Затем их стали рассматривать как семейство в составе отряда приматов, что в отечественной литературе часто продолжают делать до сих пор. Позднее тупай выделили в отдельный отряд Scandentia [22, 23], что, видимо, вполне оправдано. Дальнейшее изучение, в частности, ядерных генов и повторяющихся последовательностей необходимо для окончательного выяснения родственных отношений между приматами, тупайями, шерстокрылами, грызунами и зайцеобразными, образующими кладу Euarchontoglires. Изучение генома тупай также представляет интерес, поскольку их довольно часто используют как лабораторных животных в физиологических и медицинских исследованиях. Ценность тупай обусловлена тем, что они, видимо, ближе к приматам, чем любые другие лабораторные животные, а их содержание проще и дешевле, чем содержание приматов.

#### ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Фиксированные в спирте ткани *Tupaia glis* (Вьетнам, окрестности Далата) и *Dendrogale murina* любезно предоставлены О. Лихновой (ИПЭЭ РАН) и В. Матвеевым (МГУ). ДНК выделяли, используя обработку тканей протеиназой К с последующей экстракцией лизата смесью фенола с хлороформом.

**Клонирование и отбор повторяющихся последовательностей.** Геномную ДНК гидролизвали рестриктазами HindIII и EcoRI и после инактивации ферментов инкубировали с T4 ДНК-лигазой и плазмидой pGEM7Zf, предварительно расщепленной теми же рестриктазами. Клетки *Escherichia coli* XL1-Blue трансформировали полученным материалом и проводили скринирование клонотеки посредством гибридизации колоний с меченой суммарной геномной ДНК *T. glis*. ДНК метили, используя [ $\alpha$ - $^{32}$ P]dATP и набор для мечения случайными затравками (Prime-ItII, “Stratagene”). Условия гибридизации аналогичны описанным ранее [24]. Отобрали 20 положительных клонов и проанализировали их ДНК-вставки при помощи перекрестной гибридизации, что позволило разделить все клоны на две группы. (Впоследствии установили, что одна группа содержала КР, а другая – сателлитную ДНК.) ДНК-вставки секвенировали сразу или после клонирования их рестрикционных фрагментов, присутствие повторяющейся последовательности в которых определяли при помощи блот-гибридизации.

Для клонирования мономеров сателлитной ДНК геномную ДНК *T. glis*, обработанную BglII, лигировали с плазмидой pSL1190 и отбирали кло-

ны, гибридизирующиеся с меченой ДНК-вставкой одного из клонов второй группы.

**Секвенирование.** Нуклеотидные последовательности определяли при помощи модифицированного дидезоксинуклеотидного метода, используя ДНК-полимеразу T7 (Sequenase 2.0, “Amersham”) и [ $\alpha$ - $^{32}$ P]dATP [25]. Использовали стандартные M13-затравки или в случае длинных клонированных фрагментов – затравки, специфичные для КР тупайи (Tgl-SINE.dir: 5'-TGGYTATGTYGCTGGACTC-CCAT-3' и Tgl-SINE.rev: 5'-TGGCCATRCYRCAAA-TCCCTT-3'). Секвенированные копии КР Tu III депонированы в GenBank под номерами AY661722-AY661731, а копии сателлитной ДНК Tu-SAT1 – под номерами AY661714-AY661721.

**Полимеразную цепную реакцию (ПЦР)** использовали для изучения распространения КР Tu III у млекопитающих. В реакционную смесь добавляли 0.1 нг геномной ДНК тупайи (*T. glis* или *D. murina*) или 10 нг ДНК других млекопитающих. В качестве затравок использовали Tgl-SINE.dir и Tgl-SINE.rev. Проводили 25 циклов ПЦР, включающих инкубацию в течение 1 мин при 95, 64 и 72°C. Продукты ПЦР разделяли при помощи электрофореза в 3%-ном агарозном геле (2% NuSieve и 1% обычной агарозы). ДНК окрашивали бромистым этидием. Амплифицированный фрагмент ДНК *D. murina* элюировали из геля и клонировали в векторе pGEM-T (“Promega”).

**Блот-гибридизация сателлитной ДНК.** Геномную ДНК *T. glis* обрабатывали BglIII таким образом, чтобы получить препараты как частично, так и полностью переваренной ДНК. После разделения в 1%-ном агарозном геле ДНК окрашивали и переносили с помощью стандартного капиллярного блотинга на нейлоновый фильтр. Затем проводили гибридизацию, используя в качестве зонда меченный  $^{32}$ P клонированный мономер сателлитной ДНК *T. glis*.

**Компьютерный анализ последовательностей.** При создании консенсуса КР TuIII гипервариабельные сайты CpG/TrG/CpA, возникающие при метилировании, были превращены в CpG. Вторичную структуру РНК предсказывали при помощи программы mfold 3.2 [26] со стандартными параметрами. Конформации ДНК моделировали с использованием сервера Plot.it (<http://hydra.icgeb.tri->

[este.it/~kristian/dna/plot\\_it.html](http://kristian/dna/plot_it.html)) и программы TwistFlex (<http://margalit.huji.ac.il/TwistFlex/index.html>). Все расчеты проводили в окне 30 п.н. Проанализировали изменения угла поворота (twist) в В-ДНК [27] и общую степень гибкости ДНК по консенсусному набору параметров для тринуклеотидов [28].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Ранее с целью получить свидетельства родства грызунов, приматов и тупай мы провели поиск КР, гомологичных 7SL РНК, в геноме обыкновенной тупайи *T. glis* [21]. Такие КР действительно были обнаружены, хотя число их копий в геноме *T. glis* (500–1000) значительно ниже, чем число V1-элементов в геноме мыши ( $10^5$ ) или Alu-элементов в геноме человека ( $10^6$ ). Затем мы осуществили скрининг геномной клонотеки *T. glis* не только 7SL-специфичным зондом, но и меченой геномной ДНК этого вида. В результате обнаружили целый ряд клонов, которые гибридизовались со вторым зондом, указывая на присутствие в составе клонированных фрагментов ДНК высокоповторяющихся нуклеотидных последовательностей. В настоящей работе, используя перекрестную гибридизацию клонированных ДНК-вставок, мы разделили эти клоны на две группы. Последующее секвенирование выявило рассеянную по геному повторяющуюся последовательность в клонах одной группы и tandemный повтор – в клонах другой группы.

### Основной короткий ретропозон генома тупай

На рис. 1 представлены сходные участки в клонах первой группы (семь верхних последовательностей). Сравнительно небольшая длина этих участков (260 п.н.), А- и В-боксы промотора РНК-полимеразы III в 5'-концевой области, А-богатая последовательность в 3'-концевой области и короткие фланкирующие повторы свидетельствуют от том, что обнаруженная повторяющаяся последовательность представляет собой типичный КР. Поиск в базах нуклеотидных последовательностей выявил в интроне гена *M6P/IGF2R* две неаннотированные копии этого КР *T. glis* (AF339164) и еще шесть копий (AB090261–66) из генома другого вида тупайи (*T. belangeri*), описанных ранее

**Рис. 1.** Выравнивание нуклеотидных последовательностей короткого ретропозона Tu III из генома тупай *Tupaia glis*, *T. belangeri* и *Dendrogale murina*. Копии Tgl и Dmu секвенированы в настоящей работе, остальные последовательности взяты из GenBank (копии T1301, T1302, T1307, T1401, T1505 и T1811 [29]; копии CFnm секвенированы в составе кДНК). Консенсус Tu III составлен на основе данных копий, а также 36 EST и 12 полученных нами продуктов ПЦР-амплификации ДНК *D. murina*, содержащих неполные последовательности Tu III. Нуклеотидные последовательности Tu III приведены прописными буквами, прямые повторы во фланкирующих последовательностях (строчные буквы) подчеркнуты. Стрелками подчеркнуты прямые повторы внутри Tu III, образовавшиеся благодаря дубликациям или утерям в результате делеций. Сверху отмечены полипуриновый и полипиримидиновый участки. Глициновая тРНК<sup>(CCC)</sup> из базы геномных тРНК GtRNAdb (human chr16 #34) [32] представлена как возможный эволюционный предшественник Tu III.

Поли (Pu)

Бокс В

Бокс А

Консекус Ту III  
rPHK-Gly

Tg1-5  
Tg1-6  
Tg1-9  
Tg1-10  
Tg1-32

AF339164a  
AF339164b

T11811  
T1505  
T1301  
T1307  
T1401  
T1302  
CF525779  
CF526281  
CF525703  
CF526193  
CF527388  
CF526210

Dmu-1  
Dmu-2  
Dmu-3c

*T. gilv*  
*T. belangeri*  
*D. murina*

Секвенсы генома, содержащие повторяющиеся последовательности, сгруппированы по видам и типам элементов. Каждый элемент представлен в виде блока с номером и названием. Блоки содержат нуклеотидные последовательности, выделенные жирным шрифтом. Блоки сгруппированы в группы: Консекус Ту III, rPHK-Gly, Tg1-5, Tg1-6, Tg1-9, Tg1-10, Tg1-32, AF339164a, AF339164b, T11811, T1505, T1301, T1307, T1401, T1302, CF525779, CF526281, CF525703, CF526193, CF527388, CF526210, Dmu-1, Dmu-2, Dmu-3c. Группы Tg1-5, Tg1-6, Tg1-9, Tg1-10, Tg1-32, AF339164a, AF339164b, T11811, T1505, T1301, T1307, T1401, T1302, CF525779, CF526281, CF525703, CF526193, CF527388, CF526210, Dmu-1, Dmu-2, Dmu-3c относятся к *T. gilv*. Группы T1307, T1401, T1302, CF525779, CF526281, CF525703, CF526193, CF527388, CF526210 относятся к *T. belangeri*. Группы Dmu-1, Dmu-2, Dmu-3c относятся к *D. murina*.

Поли (Pu)

Tu-III  
Tg1-5  
Tg1-3  
Tg1-6  
Tg1-9  
Tg1-10  
Tg1-32

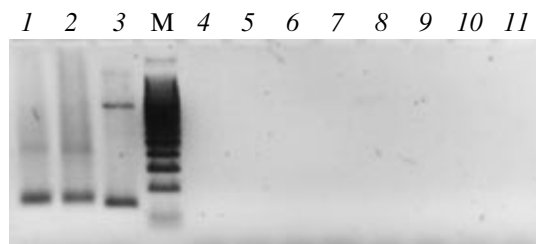
AF339164a  
AF339164b

T11811  
T1505  
T1301  
T1307  
T1401  
T1302  
CF525779  
CF526281  
CF526193  
CF527388  
CF526210

Dmu-1  
Dmu-2  
Dmu-3c

*T. gilv*  
*T. belangeri*  
*D. murina*

Секвенсы генома, содержащие повторяющиеся последовательности, сгруппированы по видам и типам элементов. Каждый элемент представлен в виде блока с номером и названием. Блоки содержат нуклеотидные последовательности, выделенные жирным шрифтом. Блоки сгруппированы в группы: Tu-III, Tg1-5, Tg1-3, Tg1-6, Tg1-9, Tg1-10, Tg1-32, AF339164a, AF339164b, T11811, T1505, T1301, T1307, T1401, T1302, CF525779, CF526281, CF526193, CF527388, CF526210, Dmu-1, Dmu-2, Dmu-3c. Группы Tu-III, Tg1-5, Tg1-3, Tg1-6, Tg1-9, Tg1-10, Tg1-32, AF339164a, AF339164b, T11811, T1505, T1301, T1307, T1401, T1302, CF525779, CF526281, CF526193, CF527388, CF526210 относятся к *T. gilv*. Группы T1307, T1401, T1302, CF525779, CF526281, CF526193, CF527388, CF526210 относятся к *T. belangeri*. Группы Dmu-1, Dmu-2, Dmu-3c относятся к *D. murina*.



**Рис. 2.** Распространение Tu III в геноме млекопитающих. ПЦР проводили, используя праймеры, специфичные для Tu III. В реакционную смесь добавляли 0.1 нг геномной ДНК тупайи или 10 нг ДНК других млекопитающих. 1 – Обыкновенная тупайя (*Tupaia glis*); 2 – северная гладкохвостая тупайя (*Dendrogale murina*); 3 – Tu III-содержащий клон (Tgl-10); 4 – еж (*Mesechinus dauricus*); 5 – мышь (*Mus musculus*); 6 – кролик (*Oryctolagus cuniculus*); 7 – собака (*Canis familiaris*); 8 – кенгуру (*Dendrolagus bennettianus*); 9 – бык (*Bos taurus*); 10 – человек (*Homo sapiens*); М – маркер ДНК (“лестница с шагом 100 п.н.”)

[29] под названием “Tu type III SINE”. Однако в этой работе авторы ограничились очень кратким анализом данного КР, сосредоточив свое внимание на двух низкокопийных (~200 копий на геном) КР, Tu типа I и II, каждый из которых включает участки (мономеры), обладающие сходством с тРНК и 7SL РНК. Кроме того, в базе данных EST мы обнаружили шесть полноразмерных и 36 неполных (оборванных) нуклеотидных последовательностей Tu III, аннотированных как фрагменты мРНК головного мозга *T. belangeri*. Эти полноразмерные последовательности вместе со всеми упомянутыми выше геномными копиями Tu III выравнены и представлены на рис. 1. Различия между копиями Tu III двух видов рода *Tupaia* не превышали различий между копиями внутри генома одного вида.

Круг распространения Tu III выясняли при помощи ПЦР с геномными ДНК ряда млекопитающих. Продукт ПЦР ожидаемого размера (140 п.н.) наблюдался в случае тупайи обоих родов (*Tupaia* и *Dendrogale*), но не у представителей других проанализированных отрядов млекопитающих (рис. 2). Это указывает на распространение КР Tu III среди видов отряда Scandentia и его отсутствие в геномах других млекопитающих. ПЦР-продукт геномной ДНК *D. murina* клонировали, 12 клонов секвенировали и подтвердили, что амплифицированные нуклеотидные последовательности соответствуют центральному району копий Tu III

(данные не представлены). Чтобы однозначно доказать присутствие в геноме *D. murina* копий типичного Tu III, геномную клонотеку этого вида тупайи скринировали Tu III-специфичным зондом и три гибридизующихся клон секвенировали. Это позволило выявить копии Tu III *D. murina* (рис. 1) и сделать вывод о том, что КР Tu III из геномов тупайи двух разных родов очень сходны. Таким образом, этот КР свойствен, видимо, всем тупайям, но его распространение ограничено отрядом Scandentia.

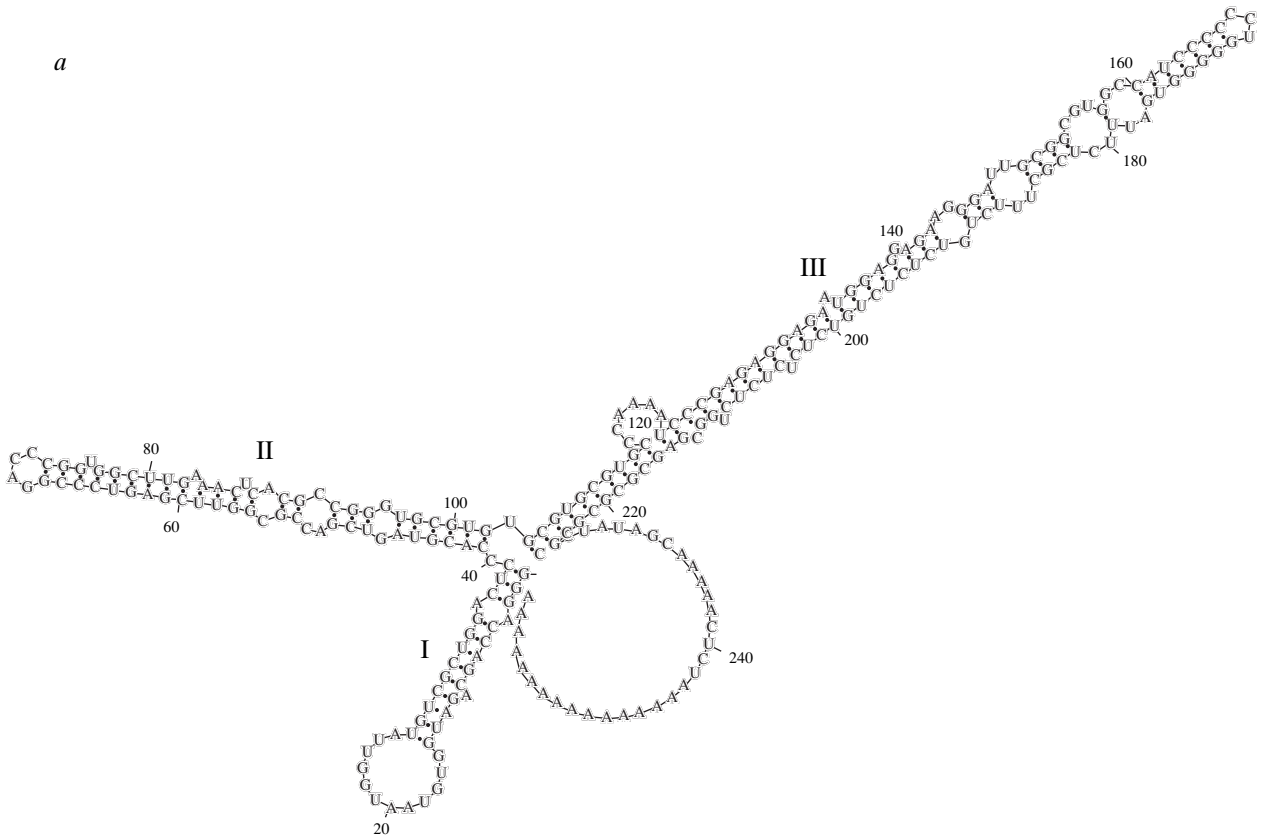
Вид тРНК, послуживший эволюционным предшественником Tu III, не смогли идентифицировать в работе [29]. Используя иной метод сравнения последовательностей, мы обнаружили сходство (70%) между 5'-концевой частью Tu III и одним из вариантов глициновой тРНК<sup>(CCC)</sup> (рис. 1). Таким образом, эта тРНК может рассматриваться в качестве вероятного эволюционного предшественника Tu III.

Неродственный тРНК район Tu III начинается с участка длиной 42 п.н. без каких-либо особенностей, кроме последовательности (YGTG)<sub>3</sub>. Затем следует полипуриновый блок длиной 24 п.н., а еще через 31 п.н. располагается полипиримидиновый участок длиной 38 п.н., содержащий только три пуриновых остатка. Как и другие КР млекопитающих, Tu III заканчивается А-богатым участком. Большинство копий фланкировано короткими прямыми повторами (рис. 1), которые образовались путем дупликации геномной ДНК в точке интеграции КР.

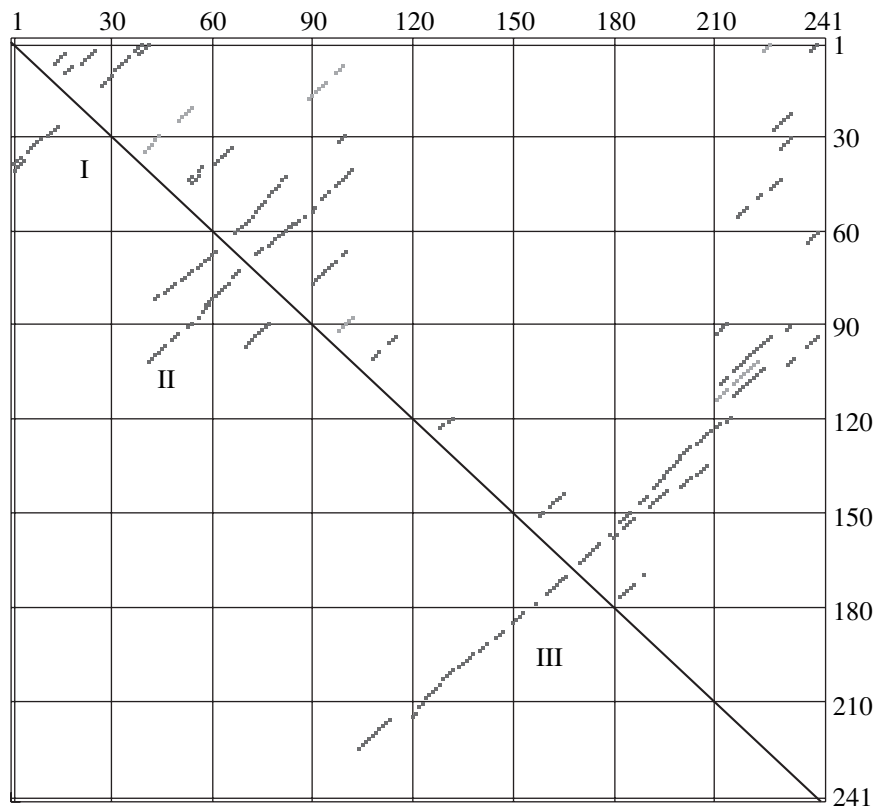
Если полипиримидиновые районы свойственны многим семействам КР млекопитающих [24], внутренние (“нехвостовые”) полипуриновые участки ранее не отмечались. В составе обоих этих участков Tu III встречаются короткие тандемные повторы (две–три копии), причем в некоторых случаях наблюдается вариабельность числа этих повторов (отмечены горизонтальными стрелками на рис. 1). Такая вариабельность может быть обусловлена проскальзыванием (slippage) ДНК-полимеразы в ходе репликации. Этот путь возникновения делеций и вставок хорошо известен для микро- и минисателлитных ДНК [30]. Недавно, анализируя структуру и эволюцию КР B1, мы обнаружили в его составе внутренние дупликации и делеции, фланкированные очень короткими (3–6 п.н.) прямыми повторами [21]. Предположили, что эти повторы провоцируют образование дупликаций и делеций путем пере-

**Рис. 3.** Вторичная структура РНК Tu III. а – Вариант сворачивания консенсусной последовательности РНК Tu III с наименьшей свободной энергией  $\Delta G$  (-105). Три шпильчатые структуры обозначены римскими цифрами. б – Анализ консервативных элементов предсказанных вторичных структур. Линии, состоящие из точек и расположенные выше диагонали, соответствуют шпильчатым структурам в конкретных предсказанных вариантах сворачивания РНК Tu III. Аналогичные линии, расположенные ниже диагонали, – консервативные шпильчатые структуры, присутствующие во всех или в большинстве вариантов. Все три шпильки в представленной вторичной структуре относятся к консервативным, хотя район шпильки II может быть свернут и иным образом.

*a*



*б*



скока обратной транскриптазы с одного такого повтора на другой в процессе синтеза ДНК на матрице РНК КР [11]. Анализ копий Tu III также выявил несколько делеций, в образовании которых могли участвовать короткие повторы, расположенные на некотором расстоянии друг от друга (рис. 1, копии Tgl\_32, Dmi\_1, повторы подчеркнуты стрелками). Отметим, что одна из последовательностей повтора всегда входит в состав делеции. Возможно, что и tandemные повторы Tu III, о которых говорилось выше, могли возникать и исчезать в процессе обратной транскрипции.

На рис. 3а показан вариант предсказанной вторичной структуры РНК Tu III с наименьшей свободной энергией ( $\Delta G = -105$ ). Если тРНК-родственная часть образует две отдельные шпильки, то остальная часть (кроме А-богатого участка) формирует длинную шпильку, лишь в нескольких местах прерываемую небольшими неспаренными участками. Сравнительный анализ различных вариантов предсказанных вторичных структур РНК Tu III со свободной энергией, близкой к минимальной (рис. 3б) показал, что все варианты сохраняют длинную шпильку, образованную, главным образом, за счет спаривания между полипуриновым и полипиримидиновым блоками. Весьма вероятно, такая вторичная структура РНК Tu III важна для ее функционирования.

#### Основная сателлитная ДНК генома *Tupaia glis*

Из клонотеки, полученной лигированием плазмидного вектора с геномной ДНК *T. glis*, гидролизованной HindIII и EcoRI, выделено значительное число клонов, содержащих повторяющуюся последовательность, отличную от КР Tu III. Хотя размеры вставок в таких клонатах существенно различались, их частичное секвенирование показало, что они содержат сходные, видимо, tandemно организованные повторяющиеся последовательности ДНК (данные не представлены). Поскольку клонированные фрагменты были слишком велики для полного их секвенирования с использованием лишь стандартных праймеров, мы создали клонотеку на основе геномной ДНК *T. glis*, обработанной BglII. Рестриктазу BglII выбрали потому, что каждый мономер обнаруженного tandemного повтора содержал ее сайт узнавания. После гибридационного скрининга клонотеки секвенировали вставки ДНК семи клонов (рис. 4). Шесть из них имели сходные размеры (361–367 п.н.) и нуклеотидные последовательности. Наибольшим сходством обладали клоны Tu\_SAT-07, -13, -14 и -17 (группа 1), а также 06 и 11 (группа 2). Фрагмент клона Tu\_SAT-25 отличался от других несколько сильнее, в том числе, в нем нарушен стандартный сайт BglII, благодаря чему он имел большую длину. “Дополнительный” участок этого фрагмента (Tu\_SAT-25') также выравнивался с другими кло-

нированными последовательностями (рис. 4), свидетельствуя о tandemной организации данного повтора.

Tandemность повтора дополнительно подтвердили полным и частичным гидролизом геномной ДНК *T. glis* рестриктазой BglII с последующим электрофорезом в агарозном геле (рис. 5а) и блот-гибридизацией (рис. 5б). При гибридазации наблюдалась “лесенка” полос, характерная для tandemных повторов. Таким образом, судя по всему, обнаруженная нами в геноме обыкновенной тупайи нуклеотидная последовательность представляет собой сателлитную ДНК.

Анализ этой сателлитной ДНК, названной нами Tu-SAT1, показал следующее. В Tu-SAT1 отсутствуют участки сходства с изученными сателлитными ДНК других видов, что типично для центромерных сателлитов. Несмотря на относительно высокое содержание А + Т (61.7%), характерное для сателлитных ДНК, в нуклеотидных последовательностях Tu-SAT1 присутствуют GC-богатые сегменты (GGCCC) (рис. 4). По-видимому, по крайней мере, часть этих участков может взаимодействовать с белками. Так, 17-нуклеотидный участок связывания центромерного белка CENP-B – наиболее GC-богатый фрагмент  $\alpha$ -сателлита приматов [31].

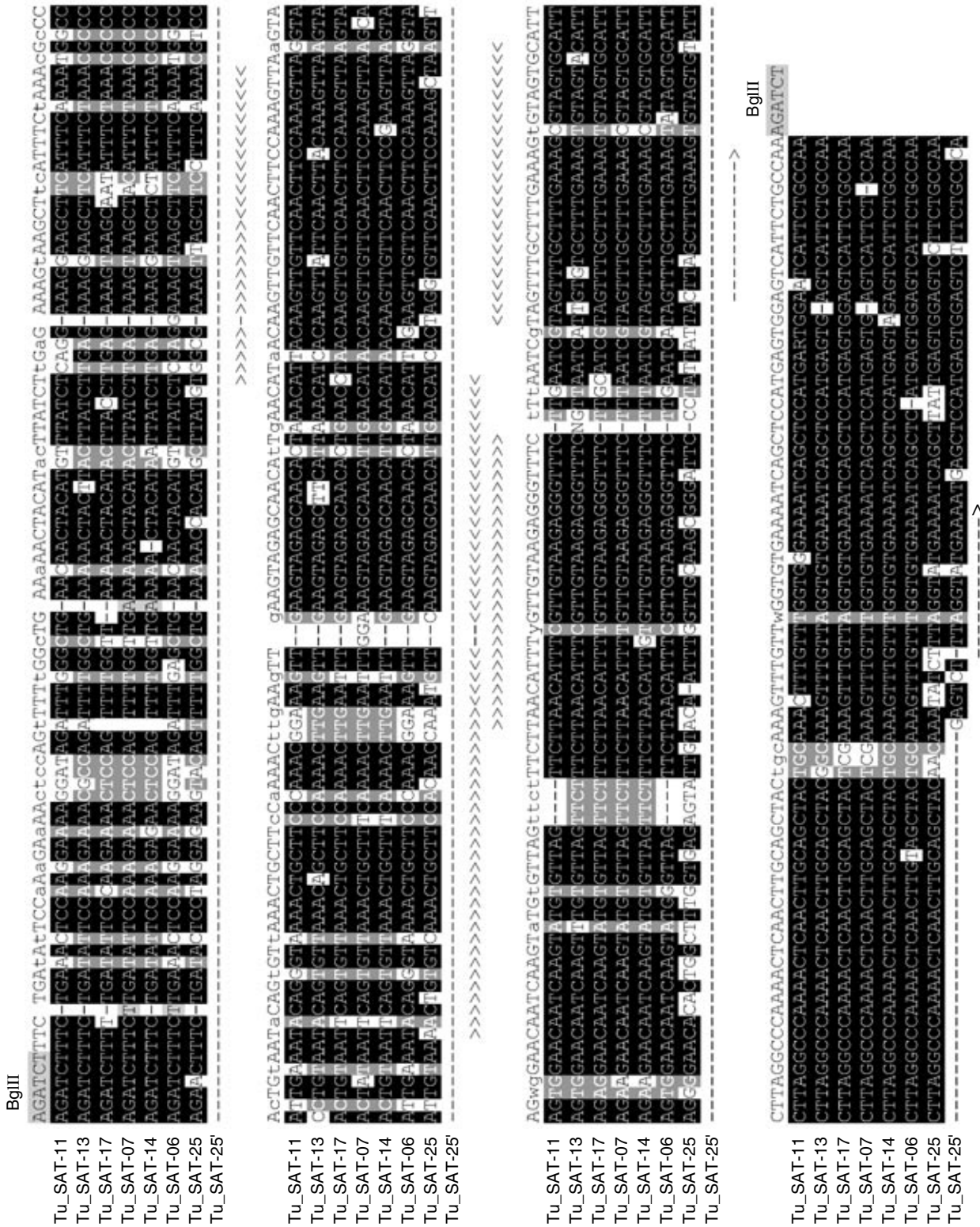
Нуклеотидная последовательность мономера Tu-SAT1 насыщена короткими простыми последовательностями, в частности, моно- и динуклеотидными периодами, что также типично для сателлитной ДНК (рис. 4). Предполагается, что, по крайней мере, некоторые сателлиты образовались в результате дубликаций короткого повтора. В таком случае его вырожденные копии можно найти по всей длине мономера сателлита (например, вырожденный повтор 5 п.н. в  $\alpha$ -сателлитной ДНК приматов). Несмотря на присутствие отдельных прямых и инвертированных повторов (рис. 4), нам не удалось подтвердить подобное происхождение Tu-SAT1.

Анализ потенциальной гибкости ДНК в мономере Tu-SAT1 показал, что Tu-SAT1 характеризуется высоким уровнем гибкости, при этом в мономере чередуются сегменты, различающиеся по уровню гибкости. Подобная неравномерная структура достаточно распространена среди сложных сателлитных ДНК, в том числе, и среди центромерных сателлитов.

Важно также отметить, что длина мономера Tu-SAT1 (360–367 п.н.) приблизительно соответствует длине ДНК двух нуклеосом. Это свойство (т.е. кратность размера мономера длине ДНК нуклеосомы) также типично для сателлитных ДНК (например, 171 п.н. у человека, 180 п.н. у риса, 340 п.н. у свеклы и т.п.) [5, 6].

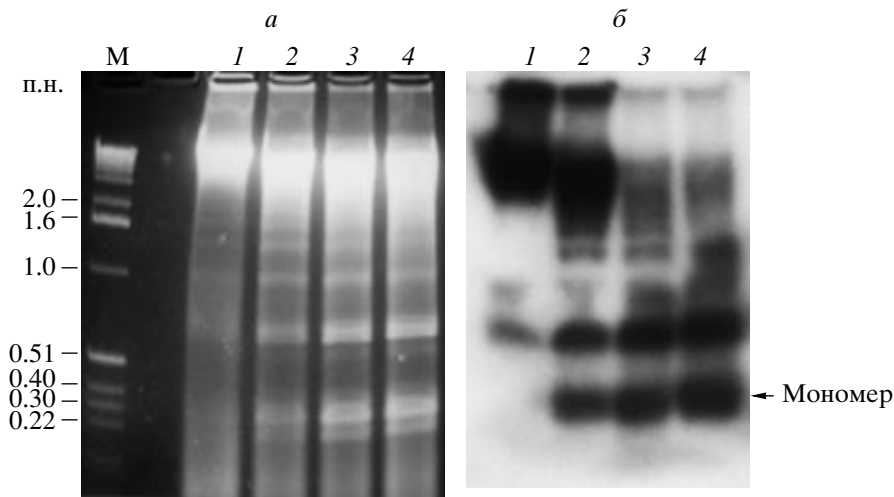
В данной работе описаны два главных повтора генома тупайи. Один из них, Tu III, это типичный КР, однако ему присущи некоторые особенности,





**Рис. 4.** Выравнивание BglII-номеров повтора Tu-SAT1 генома обыкновенной тупай *Tupaia glis*. Представлены нуклеотидные последовательности семи клонированных мономеров и их консенсуса. Сайт BglII на правом конце клонированных последовательностей не показан. В клоне Tu\_SAT-25 левый сайт BglII мутирован, благодаря чему клонированная последовательность более длинная. Дополнительный участок (Tu\_SAT-25) выравнивается с другими мономерами, что указывает на tandemную организацию повтора Tu-SAT1. Инвертированные повторы отмечены знаками >>> или <<< под выравниванием, прямые повторы – ----->





**Рис. 5.** Тандемная организация повтора Tu-SAT1. Геномную ДНК *Tupaia glis* обрабатывали BglII в течение 5, 15, 30 и 60 мин (1, 2, 3 и 4 соответственно). *a* – Электрофорез ДНК в 1.5%-ном агарозном геле. *б* – Гибридизация ДНК, перенесенной на нейлоновый фильтр, с меченым Tu-SAT1-специфичным зондом. Наблюдаемая “лесенка” полос свидетельствует о тандемной организации мономеров Tu-SAT1. М – маркер размера ДНК (“Gibco BRL”).

в частности, он содержит полипиридиновый участок, а транскрипт этого КР обладает выраженной вторичной структурой. Его распространение ограничено отрядом тупай (Scandentia), что подтверждает монофилию этого отряда. Второй повтор, Tu-SAT1, проявляет свойства типичного центромерного сателлита. Открытие повторяющихся последовательностей может послужить отправной точкой в изучении генома этого перспективного лабораторного животного.

Работа поддержана Российским фондом фундаментальных исследований (05-04-49553 и 02-04-48644).

#### СПИСОК ЛИТЕРАТУРЫ

1. Беридзе Т.Г. 1982. *Сателлитная ДНК*. М.: Наука.
2. Willard H.F. 1990. Centromeres of mammalian chromosomes. *Trends Genet.* **6**, 410–416.
3. Elder J.F., Jr., Turner B.J. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quant. Rev. Biol.* **70**, 297–320.
4. Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Human Genet.* **64**, 24–30.
5. Henikoff S., Ahmad K., Malik H.S. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science.* **293**, 1098–1102.
6. Ugarkovic D., Plohl M. 2002. Variation in satellite DNA profiles—causes and effects. *EMBO J.* **21**, 5955–5959.
7. Lamb J.C., Birchler J.A. 2003. The role of DNA sequence in centromere formation. *Genome Biol.* **4**, 214.
8. Енукашвили Н.И., Кузнецова И.С., Подгорная О.И. 2003. Организация центромер у млекопитающих. *Цитология.* **45**, 255–270.
9. Smit A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opin. Genet. Dev.* **9**, 657–663.
10. Deininger P.L., Batzer M.A. 2002. Mammalian retroelements. *Genome Res.* **12**, 1455–1465.
11. Kramerov D.A., Vassetzky N.S. 2005. Short Retroposons (SINEs) in Eukaryotic Genomes. *Int. Rev. Cytol.* **247**, 165–221.
12. Orgel L.E., Crick F.H. 1980. Selfish DNA: the ultimate parasite. *Nature.* **284**, 604–607.
13. Makalowski W. 2000. Genomic scrap yard: how genomes utilize all that junk. *Gene.* **259**, 61–67.
14. Allen T.A., von Kaenel S., Goodrich J.A., Kugel J.F. 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nature Struct. Mol. Biol.* **11**, 816–821.
15. Сердобова И.М., Крамеров Д.А. 1994. Использование коротких ретрозонов как филогенетических маркеров. *Докл. РАН.* **335**, 664–667.
16. Shimamura M., Yasue H., Ohshima K., Abe H., Kato H., Kishiro T., Goto M., Munechika I., Okada N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature.* **388**, 666–670.
17. Shedlock A.M., Okada N. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays.* **22**, 148–160.
18. 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**, 1061–1071.
19. 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.

20. Nikaido M., Nishihara H., Hukumoto Y., Okada N. 2003. Ancient SINEs from African endemic mammals. *Mol. Biol. Evol.* **20**, 522–527.
21. Vassetzky N.S., Ten O.A., Kramerov D.A. 2003. B1 and related SINEs in mammalian genomes. *Gene*. **319**, 149–160.
22. Carrol R.L. 1988. *Vertebrate Paleontology and Evolution*. NY.: Freeman and Company.
23. Павлинов И.Я. 2003. *Систематика современных млекопитающих*. М.: МГУ.
24. Borodulina O.R., Kramerov D.A. 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mammal. Genome*. **12**, 779–786.
25. Redston M.S., Kern S.E. 1994. Klenow co-sequencing: a method for eliminating “stops”. *Biotechniques*. **17**, 286, 288.
26. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.
27. Sarai A., Mazur J., Nussinov R., Jernigan R.L. 1989. Sequence dependence of DNA conformational flexibility. *Biochemistry*. **28**, 7842–7849.
28. Gromiha M. 2000. Structure based sequence dependent stiffness scale for trinucleotides: a direct method. *J. Biol. Phys.* **26**, 43–50.
29. 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.
30. Debrauwere H., Gendrel C.G., Lechat S., Dutreix M. 1997. Differences and similarities between various tandem repeat sequences: minisatellites and microsatellites. *Biochimie*. **79**, 577–586.
31. Masumoto H., Nakano M., Ohzeki J. 2004. The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. *Chromosome Res.* **12**, 543–556.
32. Schattner P., Brooks A.N., Lowe T.M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–689.