



УДК 577.113.5:578.828

## НОВОЕ СЕМЕЙСТВО ГЕНОВ KIAA1245, РАЗЛИЧАЮЩИХСЯ ПО ПРИСУТСТВИЮ В ИНТРОНАХ LTR HERV-K

© 2002 г. Т. В. Виноградова<sup>#</sup>, П. А. Жулидов, А. Э. Илларионова, Е. Д. Свердлов*Институт биоорганической химии им. М.М. Шенякина и Ю.А. Овчинникова РАН,  
117997 ГСП, Москва, В-437, ул. Миклухо-Маклая, 16/10*

Поступила в редакцию 23.07.2001 г. Принята к печати 12.10.2001 г.

При исследовании транскрибирующихся длинных концевых повторов (LTR) эндогенных ретровирусов человека (HERV) семейства К в нормальных и опухолевых тканях нами обнаружен транскрипт, содержащий LTR и последовательность, гомологичную фрагменту мРНК KIAA1245. Среди геномных последовательностей, депонированных в GenBank, нами было идентифицировано 10 структур, имеющих высокий уровень идентичности с мРНК KIAA1245. Для всех обнаруженных последовательностей была определена экзон-интронная структура и проведено сравнение степени идентичности экзонных последовательностей. Установлено, что последовательности различаются как по степени идентичности экзонных структур, так и по наличию/отсутствию LTR HERV-K в третьем интроне. Идентифицированные последовательности образуют новое семейство генов, которое состоит, как минимум, из четырех подсемейств. Два из этих подсемейств содержат LTR, тогда как два других – нет. Методом ПЦР показано, что интеграция LTR в интрон произошла после отделения эволюционной ветви орангутана от остальных гоминоидов, но до отделения ветви гориллы, то есть между 8 и 13 млн. лет назад.

*Ключевые слова:* LTR; HERV-K; семейства генов.

### ВВЕДЕНИЕ

Анализ практически полной последовательности генома человека показал, что значительную его часть составляют повторяющиеся элементы, такие, как короткие диспергированные повторы (SINE, Short Interspersed Nuclear Elements) (~13% генома), длинные диспергированные повторы (LINE, Long Interspersed Nuclear Elements) (~20%) и ретроэлементы, содержащие ретровирусные длинные концевые повторы (LTR, Long Terminal Repeats) – около 8% [1–3], функциональная роль которых в настоящее время является предметом интенсивных исследований. LTR ретровирусов образуются при превращении вируса в провирус и предназначены для контроля транскрипции ретровирусной ДНК, встроенной в геном клетки-хозяина. В процессе эволюции происходили, по-видимому, многократные интеграции провирусных ДНК в геном приматов, в результате чего в настоящее время в геноме человека присутствуют многочисленные семейства эндогенных ретровирусов человека (HERV). Со временем процесс мутирования таких провирусов привел к утере ими способности к синтезу полноценных вирусных частиц, и они превратились в стабильные элементы генома, наследу-

емые как обычные гены. Помимо мутировавших в разной степени провирусов, геном насыщен десятками тысяч индивидуальных, не связанных с вирусными генами LTR, образовавшихся в результате рекомбинаций между двумя LTR, окаймляющими провирусную последовательность.

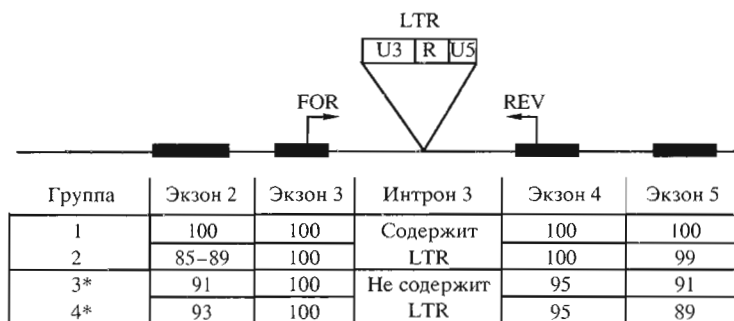
Типичный LTR включает в себя весь комплекс регуляторных элементов, необходимых для эффективной транскрипции: промотор и энхансер, необходимые для инициации транскрипции, и сигналы терминации и полиаденилирования РНК [4, 5]. Легко представить себе, что внедрение LTR рядом с каким-либо клеточным геном может добавить к собственным контрольным элементам этого гена регуляторные элементы LTR и, таким образом, повлиять на регуляцию экспрессии этого гена. В геноме человека LTR часто встречаются в непосредственной близости от генов [6]. Существуют примеры, подтверждающие участие LTR в регуляции транскрипции [7].

Таким образом, LTR эндогенных ретровирусов, оказываясь в процессе эволюции вблизи клеточных генов, могут в силу своего регуляторного потенциала влиять на экспрессию этих генов, и, тем самым, влиять на процессы видообразования.

При изучении роли LTR в регуляции транскрипции нами было обнаружено множество РНК, содержащих в своем составе LTR. Один из таких транскриптов, экспрессирующийся в герминоген-

Сокращения: LTR – длинный концевой повтор; HERV-K – эндогенные ретровирусы человека семейства К.

<sup>#</sup> Автор для переписки (тел.: (095) 330-69-92; факс: (095) 330-65-38; эл. почта: tv@humgen.siocb.ras.ru).



**Рис. 1.** Схема фрагмента экзон-интронной структуры, общей для четырех групп нуклеотидных последовательностей, гомологичных мРНК KIAA1245 ([8], GenBank AB033071). Черными прямоугольниками обозначены экзоны, в таблице под каждым из экзонов для всех групп генов указана степень их идентичности (%) с мРНК KIAA1245. LTR – длинный концевой повтор, включающий в себя области U3, U5 и R. Стрелками обозначены праймеры FOR и REV. Группы генов, практически не различающиеся по степени идентичности экзонных областей, но различающиеся по степени идентичности интронов, отмечены\*.

ной опухоли семиноме в 10 раз слабее, чем в нормальной паренхиме яичка, был подвергнут более детальному анализу, результаты которого являются предметом данного сообщения.

Обсуждаемая РНК оказалась высокогомологичной мРНК KIAA1245 [8], аннотированной в GenBank под номером AB033071. При анализе геномных последовательностей из GenBank, имеющих гомологию с мРНК KIAA1245, мы обнаружили, что идентифицированные последовательности образуют новое семейство генов, которое состоит, как минимум, из четырех подсемейств. Два из этих подсемейств содержат LTR, тогда как два других – не содержат. Методом филогенетического анализа мы показали, что интеграция LTR произошла после отделения эволюционной ветви орангутана от остальных гоминоидов, но до отделения ветви гориллы, то есть между 8 и 13 млн. лет назад.

### РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

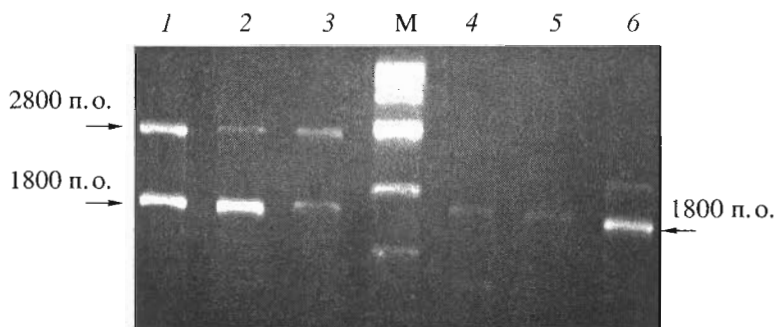
Для идентификации РНК, содержащих в своем составе LTR, мы использовали метод селективной амплификации кДНК, основанный на эффекте ПЦР-супрессии. Метод представляет собой модификацию опубликованной нами ранее селективной амплификации геномных фрагментов, содержащих LTR [9]. Детали метода будут опубликованы отдельно. LTR-содержащие фрагменты кДНК, полученные в результате супрессивной амплификации, разделяли в денатурирующем ПААГ, элюировали и определяли их первичную структуру, которую далее сопоставляли с последовательностями, депонированными в GenBank. Среди множества полученных таким образом LTR-содержащих последовательностей одна оказалась гомологичной мРНК KIAA1245, экспрессия которой является тканеспецифичной и повышена в некоторых отделах головного мозга [8]. LTR, входящие в состав тканеспецифично экс-

прессирующихся генов, являются наиболее вероятными кандидатами на роль регуляторных LTR, и поэтому данная последовательность была подвергнута более детальному анализу.

При анализе нуклеотидных последовательностей, содержащихся в GenBank, мы обнаружили, что мРНК KIAA1245 имеет гомологию с 10 геномными последовательностями человека. Оказалось, что наряду с ожидаемыми последовательностями, содержащими LTR, в геномной базе данных аннотированы четыре последовательности, высокогомологичные мРНК KIAA1245 и не содержащие LTR (рис. 1). Сопоставление геномных последовательностей с последовательностью мРНК KIAA1245 позволило установить частичную интрон-экзонную структуру, характерную для всех идентифицированных гомологичных последовательностей (рис. 1). В генах, гомологичных мРНК KIAA1245, нами идентифицировано шесть экзонов. Реально гены содержат большее число интронов и экзонов, но их анализ на основе доступной в настоящее время информации невозможен.

Для того чтобы установить, имеем ли мы дело с дуплицированными (паралогичными) геномными локусами, часть из которых подверглась интеграции LTR уже после дупликации, с аллельными вариантами, или, наконец, с ошибками в сборке нуклеотидных геномных последовательностей при секвенировании генома человека, было необходимо провести более детальный анализ всех обнаруженных гомологичных последовательностей.

Существование LTR-содержащих и LTR-несодержащих вариантов генов было экспериментально подтверждено с помощью ПЦР-амплификации геномной ДНК человека с праймерами FOR и REV (рис. 1), которые комплементарны последовательностям 3-го и 4-го экзонов, фланкирующим LTR, и позволяют различить эти группы генов. Мы получили продукты амплификации длиной около 2800 и 1800 п.о. (рис. 2, дорожка 1),



**Рис. 2.** Электрофоретическое разделение в 0.8% агарозном геле продуктов ПЦР-амплификации с праймерами FOR и REV на матрице ДНК человека (дорожка 1) и приматов (дорожка 2 – шимпанзе, 3 – горилла, 4 – орангутан, 5 – гиббон, 6 – мандрил Сфинкс). М – маркер длин ДНК “1kb DNA Ladder” (Promega, США). Стрелками указаны продукты ПЦР-амплификации с их размерами. Фрагмент длиной 2800 п. о. содержит LTR, а фрагмент длиной 1800 п. о. не содержит LTR.

соответствующие ожидаемым фрагментам, содержащему и не содержащему LTR (длина LTR около 1000 п. о.). Присутствие этих двух продуктов свидетельствует о том, что данный геномный локус, имеющий гомологию с мРНК KIAA1245, представлен в геноме человека, как минимум, в двух вариантах.

Для выяснения степени идентичности последовательности мРНК KIAA1245 и экзонных частей упомянутых геномных фрагментов было прове-

дено попарное выравнивание их нуклеотидных последовательностей (рис. 1).

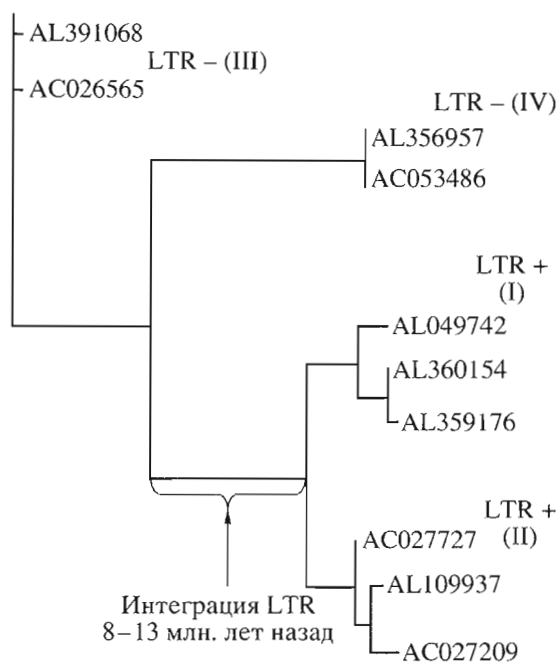
В результате анализируемые последовательности были разделены на три группы, различающиеся по степени идентичности с мРНК KIAA1245 в области четырех экзонов, структура которых наиболее надежно установлена. К первой группе были отнесены последовательности с практически 100%-ной идентичностью по всем четырем экзонам. Последовательности второй группы показывают сравнительно низкую (не более чем 89%) степень идентичности по второму экзону при высокой гомологии по остальным. В последовательностях третьей группы идентичны (100%) только третьи экзоны, тогда как для остальных трех экзонов степень идентичности варьирует от 89 до 95%.

Первая и вторая группы содержат LTR в интроне между третьим и четвертым экзонами, а у третьей группы LTR отсутствует.

Полученные данные не оставляют сомнения в том, что мРНК KIAA1245 кодируется одним из генов довольно многочисленного и сравнительно недавно образовавшегося семейства. Все обнаруженные члены семейства расположены на хромосоме 1, хотя и в разных ее локусах.

Для дальнейшего анализа структурных гомологий между членами этого семейства было проведено множественное выравнивание по интронным участкам, расположенным между третьим и четвертым и между четвертым и пятым экзонами, где нуклеотидные последовательности всех 10 членов семейства были установлены с высокой степенью надежности. Такое выравнивание позволяет достовернее установить эволюционные взаимосвязи между изучаемыми последовательностями, поскольку интроны являются селективно нейтральными участками генома и, вследствие этого, скорость накопления мутаций в них меньше зависит от окружения гена.

Анализ полученной дендрограммы ближайших соседей (рис. 3) показал, что существует четыре варианта локусов, два из которых соответ-



**Рис. 3.** Дендрограмма ближайших соседей для геномных последовательностей, гомологичных мРНК KIAA1245. Дендрограмма построена по интронным участкам, расположенным между третьим и четвертым и между четвертым и пятым экзонами генов. Указаны номера депонирования последовательностей в GenBank. Римскими цифрами обозначены группы генов: содержащие LTR (I, II) и не содержащие LTR (III, IV). Стрелка указывает время интеграции LTR.

ствуют первой и второй группам, идентифицированным ранее, тогда как два других являются следствием разделения идентифицированной выше третьей группы на две отдельные четко различающиеся группы.

Приблизительное время интеграции LTR мы определили методом ПЦР-амплификации на матрице геномной ДНК различных приматов с праймерами, соответствующими прилегающим к LTR геномным последовательностям. Результаты такого филогенетического анализа приведены на рис. 2. Амплификация двух фрагментов разной длины (2800 и 1800 п. о.) свидетельствует, что в геномах человека, шимпанзе и гориллы присутствуют как содержащие LTR, так и не содержащие LTR варианты генов исследуемого семейства. В геномах орангутана, гиббона и мандрила был обнаружен только короткий фрагмент, соответствующий членам семейства, не содержащим LTR.

Зная время расхождения эволюционных ветвей приматов, мы можем предположить, что интеграция данного LTR произошла после отделения предковой эволюционной ветви орангутана от остальных гоминоидов, но до отделения ветви гориллы, то есть между 8 и 13 млн. лет назад.

Полученные данные создают основу для дальнейшего анализа роли LTR в регуляции экспрессии генов данного семейства путем определения различий в тканеспецифичности экспрессии LTR-содержащих и LTR-несодержащих генов.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Селективная амплификация LTR-содержащих фрагментов кДНК** проводили по разработанному нами ранее методу, позволяющему селективно амплифицировать геномные фрагменты, содержащие LTR [9], используя в качестве матрицы для ПЦР кДНК. Синтез кДНК проводили на матрице суммарной РНК, используя набор реактивов "SMART PCR cDNA Synthesis Kit" (Clontech, США), согласно прилагаемому протоколу, в качестве затравки для синтеза первой цепи кДНК обратной транскриптазой использовали праймер GGGCTGGGGACGGTCAGGT, комплементарный последовательности LTR HERV-K. Разделение селективно амплифицированных LTR-содержащих фрагментов кДНК проводили электрофорезом в 6% денатурирующем ПААГ по стандартной методике [10]. Для элюции фрагментов кДНК из геля вырезали полоску геля и инкубировали 2 ч при 55°C в буфере, содержащем 50 мкл 10 мМ Трис-НСl, рН 8.0. Первичную структуру фрагментов определяли с помощью автоматического секвенатора "ALFexpress II automated DNA sequencer" (Amersham-Pharmacia Biotech).

**Геномная ПЦР.** В качестве матрицы использовали геномную ДНК человека или приматов

(шимпанзе, гориллы, орангутана, гиббона, мандрила). Предварительно смешивали 1.75 ед. акт. Taq-ДНК-полимеразы (ИБХ, Россия) и 0.7 мкг TaqStart-антител (Clontech, США) в буфере, содержащем 50 мМ КCl, 10 мМ Трис-НСl, рН 7.0, инкубировали при комнатной температуре 5–10 мин и добавляли в инкубационную смесь для ПЦР, содержащую буфер (50 мМ КCl, 10 мМ Трис-НСl, рН 9.0, 0.1% Тритон X-100), 2.5 мМ MgCl<sub>2</sub>, 0.125 мМ dNTPs, 0.4 мМ праймеры FOR (TATATTTTGGACAGCAGTTTTTCC) и REV (CCTTGCCACCTGAAGAAGAC) и 10 нг геномной ДНК. Суммарный объем реакционной смеси – 50 мкл, проводили 28 циклов ПЦР в режиме: 94°C – 30 с, 58°C – 30 с, 72°C – 60 с. Из реакционной смеси отбирали аликваты по 10 мкл для анализа продуктов ПЦР электрофорезом в 0.8% агарозном геле.

**Картирование, поиск гомологий, анализ геномного окружения.** Поиск гомологичных последовательностей проводили в базе данных GenBank NCBI при помощи компьютерной программы MEGABLAST (<http://ncbi.nlm.nih.gov/BLAST>). Повторяющиеся последовательности анализировали с использованием программы RepeatMasker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). Поиск клеточных генов и фрагментов кДНК проводили с использованием программ Human Genome BLAST (<http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html>) и Draft Human Genome Browser (<http://genome.ucsc.edu/goldenPath/hgTracks.html>).

**Множественное выравнивание и построение дерева ближайших гомологий.** Парное сравнение последовательностей осуществляли с помощью программы BLAST2 (<http://ncbi.nlm.nih.gov/BLAST>). Множественное выравнивание осуществляли с использованием программы ClustalW [11].

Для подсчета матриц расстояний на основе множественного выравнивания была использована программа DNADIST. Подсчет проводили с использованием матрицы KIMURA. Дерево ближайших гомологий было построено с помощью программы FITCH. Все эти программы входят в пакет программ Phylip (<http://bioweb.pasteur.fr/seqanal/phylogeny/phylic-uk.html>). Визуализацию результатов построения дерева ближайших гомологий осуществляли с помощью программы TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>).

**Олигонуклеотидные праймеры.** Дизайн праймеров проводили с использованием программы Primer3 WWW сервера Whitehead Institute ([http://www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)).

## БЛАГОДАРНОСТИ

Авторы выражают признательность В.К. Потапову и Н.В. Скапцовой за синтез олигонуклео-

тидов, Л.П. Леппик и Л.Г. Николаеву за помощь в работе над статьей.

Исследования проведены при поддержке грантов ГНТП "Геном человека", РФФИ (грант № 01-04-48900) и INTAS-991-1143.

#### СПИСОК ЛИТЕРАТУРЫ

1. Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W. et al. // *Nature*. 2001. V. 409. P. 860–921.
2. Smit A.F. // *Curr. Opin. Genet. Dev.* 1999. V. 9. P. 657–663.
3. Brosius J. // *Gene*. 1999. V. 238. P. 115–134.
4. Leib-Mosch C., Seifarth W. // *Virus Genes*. 1995. V. 11. P. 133–145.
5. Sverdlov E.D. // *FEBS Lett.* 1998. V. 428. P. 1–6.
6. Vinogradova T., Volik S., Lebedev Y., Shevchenko Y., Lavrentyeva I., Khil P., Grzeschik K.H., Ashworth L.K., Sverdlov E. // *Gene*. 1997. V. 199. P. 255–264.
7. Britten R.J. // *Gene*. 1997. V. 205. P. 177–182.
8. Nagase T., Ishikawa K., Kikuno R., Hirosawa M., Nomura N., Ohara O. // *DNA Res.* 1999. V. 6. P. 337–345.
9. Lavrentieva I., Broude N.E., Lebedev Y., Gottesman I.I., Lukyanov S.A., Smith C.L., Sverdlov E.D. // *FEBS Lett.* 1999. V. 443. P. 341–347.
10. Маниатис Т., Фрич Э., Сэмбрук Дж. Методы генетической инженерии. Молекулярное клонирование. М.: Мир, 1984.
11. Thompson J.D., Higgins D.G., Gibson T.J. // *Nucl. Acids Res.* 1994. V. 22. P. 4673–4680.

### A New Family of KIAA1245 Genes with and without the HERV-K LTRs in Their Introns

T. V. Vinogradova<sup>#</sup>, P. A. Zhulidov, A. E. Illarionova, and E. D. Sverdlov

<sup>#</sup>Phone: +7 (095) 330-6992; fax: +7 (095) 330-6538; e-mail: tv@humgen.siobc.ras.ru  
Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences,  
ul. Miklukho-Maklaya 16/10, GSP Moscow, 117997 Russia

A transcript containing the long terminal repeat (LTR) and the sequence homologous to the KIAA1245 mRNA fragment were revealed among the transcribed LTRs of human endogenous viruses of the K family in normal and tumor tissues. Ten other sequences with a high level of homology to the KIAA1245 mRNA were found in the GenBank. The intron–exon structures were determined for all the sequences, and their exon sequences were compared. The comparison showed that they differ both in the extent of the exon homology and in the presence or absence of the HERV-K LTR in the third intron. The revealed sequences form a new gene family that comprises at least four subfamilies. Two of these subfamilies have the LTR, and the other two do not. We showed by PCR that the LTR was integrated into the introns after the divergence of the orangutan evolutionary branch from other hominoids but before the divergence of the gorilla branch, i.e., 8–13 million years ago. The English version of the paper: *Russian Journal of Bioorganic Chemistry*, 2002, vol. 28, no. 4; see also <http://www.maik.ru>.

*Key words:* HERV-K, LTR, gene families