

АНАЛИЗ ДУПЛИКАЦИИ ГЕНОВ миРНК В ГЕНОМЕ ЧЕЛОВЕКА И РОЛЬ ЭВОЛЮЦИИ ТРАНСПОЗОНОВ В ЭТОМ ПРОЦЕССЕ

И.И. Титов^{1,2}, П.С. Ворожейкин¹

¹ Государственное образовательное учреждение высшего профессионального образования Новосибирский государственный университет, Новосибирск, Россия, e-mail: titov@bionet.nsc.ru;

² Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия

Одним из механизмов эволюции миРНК в геномах являются дубликации их генов. В работе показывается, что миРНК человека в основном организованы в кластеры размером до 5 тыс. нуклеотидов, а их копии случайно рассеяны по геному в среднем через $4,3 \times 10^6$ п.о. Мы провели сравнение окрестностей копий миРНК с последовательностями мобильных элементов и установили, что подавляющее большинство (96 %) гомологов миРНК распространилось ДНК-транспозонами и ретроэлементами. Среди них наибольший вклад вносят приматоспецифичные элементы *madel*, копии которого содержат миРНК семейства *hsa-mir-548*. Исследование эволюции *madel* показывает, что эти мобильные элементы возникли и взрывообразно размножились в геноме общего предка приматов и в дальнейшем разошлись со средними для геномов скоростями. Тогда же независимо друг от друга возникли современные миРНК семейства *hsa-mir-548* как результат адаптивных мутаций в части копий транспозона, которые и были рекрутированы в систему молчания генов. Задokumentированные в работе изменения в транспозонах при их эволюции в гены миРНК так же, как и многочисленные найденные последовательности копий миРНК, могут быть полезными для предсказания новых миРНК.

Ключевые слова: миРНК, мобильные элементы, ДНК-транспозон, *madel*, компьютерный анализ, вторичная структура, эволюция.

Введение

МикроРНК (миРНК) – семейство малых, около 22 нуклеотидов (нт), одноцепочечных некодирующих РНК последовательностей, участвующих в посттранскрипционной регуляции экспрессии генов (Bartel, 2004). В геноме человека миРНК встречаются в интронах, экзонах и межгенных областях. Около 42 % миРНК человека относят к кластерам размером до 3 тыс. нт, а рядом с известными миРНК наблюдаются консервативные пре-миРНК-подобные элементы (Altuvia *et al.*, 2005).

Одним из путей эволюции миРНК является дубликация. Некоторые гены миРНК млекопитающих, включая человека, происходят от мобильных генетических элементов (Smalheiser *et al.*, 2005; Piriyaongsa *et al.*, 2007). В частности,

7 генов миРНК из *hsa-mir-548* семейства предположительно эволюционировали от элемента *madel* (Piriyaongsa *et al.*, 2007). Известно, что мобильные элементы составляют до 45 % генома человека (Consortium, 2001). Большая распространенность и повторяемость мобильных элементов могут объяснить обширное распространение генов миРНК и гомологов миРНК в геноме человека и помочь обнаружению новых миРНК.

В работе проведен компьютерный анализ внутригеномного распределения миРНК человека: оценен размер кластеров миРНК и исследована роль мобильных элементов в распространении генов миРНК по геному. Особое внимание уделяется эволюции приматоспецифичного транспозона *madel*, ответственного за распространение миРНК семейства *hsa-mir-548*.

Материалы и методы

В работе используются 695 известных пре-миРНК человека из базы данных miRBase (Griffiths-Jones *et al.*, 2008), релиз 12.0. Пре-миРНК содержат 343 зрелые миРНК из 5'-ветви предшественника и 349 – из 3'-ветви. Используются геномы *Homo sapiens* (релиз hg19 соответствует релизу 37.1 NCBI), *Pan troglodytes* (релиз panTro2), *Pongo abelii* (релиз ponAbe2), *Macaca mulatta* (релиз rheMac2) и *Callithrix jacchus* (релиз Cal Jac3).

Сравнение последовательностей рассматриваемых миРНК, их генов и повторяющихся геномных последовательностей (в том числе мобильных элементов) проводится с помощью веб-серверов RepeatMasker (Smith, 1999) и UCSC Table Browser (Karolchik *et al.*, 2004). Внутригеномный поиск нуклеотидных последовательностей проводится с помощью программы BLASTN (Altschul *et al.*, 1997). Для построения вторичных структур РНК последовательностей и вычисления их свободной энергии использовались программы RNAfold (Hofacker *et al.*, 1994) и GArna (Titov *et al.*, 2002). Множественное выравнивание нуклеотидных последовательностей осуществляется с помощью программы CrustalW (Jeanmougin *et al.*,

1998). Построение филогенетических деревьев для нуклеотидных последовательностей проводится с помощью программы MEGA4 (Tamura *et al.*, 2007) методом максимальной парсимонии (Sober, 1983).

Результаты

Взаимное расположение миРНК и их копий

Корреляционная функция парных расстояний является удобным способом характеризовать взаимную зависимость расположения объектов наблюдения. Используя позиции миРНК в геноме человека, мы рассчитали вероятность наблюдения ближайших пар миРНК на определенном расстоянии друг от друга в одной цепи ДНК. Оказалось, что на расстояниях до 5 тыс. нуклеотидов эта функция примерно обратно пропорциональна расстоянию $y \propto x^{-1}$ (рис. 1, а), что свидетельствует о корреляциях в расположении миРНК, например об организации в кластеры. Медленный степенной закон спада автокорреляций нуклеотидов ранее наблюдался в ДНК (Лобзин, Четкин, 2000), а одним из объяснений является блочная организация генома (Karlin *et al.*, 1993; Капитонов,

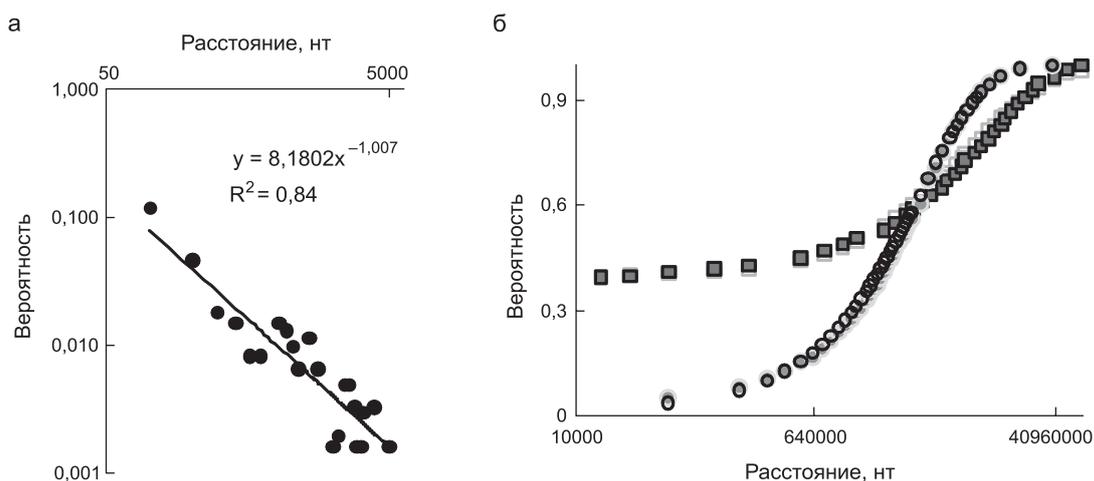


Рис. 1. Функция распределения вероятностей (по оси ординат) парных расстояний (по оси абсцисс) между соседними миРНК.

Расстояния рассчитаны в отдельности для каждой из цепей генома человека. Слева в двойных логарифмических координатах построен график для расстояний до 5 тыс. нт. (а). Аппроксимация линейной функцией отвечает степенному закону спада от расстояния (на врезке). Справа показаны функции распределения вероятностей парных расстояний при расстояниях свыше 10 тыс. нт (б). Функции для миРНК и их копий отмечены закрашенными квадратами и кругами соответственно. Незакрашенные символы показывают аппроксимации функций «затянутой» экспонентой.

Титов, 1994). Отметим, что присутствия двух миРНК в одной пре-миРНК недостаточно для объяснения корреляционного эффекта, поскольку область расстояний, в которой наблюдается степенная зависимость, значительно больше 80 нт – типичной длины пре-миРНК.

На расстояниях более 10^4 нт миРНК располагаются разреженно, поэтому для больших расстояний вместо функции вероятности мы построили ее кумулятивную функцию. Здесь функция распределения по ближайшим расстояниям превращается в «затянутую» экспоненту: кумулятивная функция имеет вид:

$$F = 1 - e^{-\left(\frac{x-G}{D}\right)^E}$$
 (рис. 1, б) (параметры распределения приведены в табл. 1). Подобное распределение при $E = 0,5-0,7$, т. е. с экспоненциальным «толстым хвостом», присуще объектам с флуктуационно-неоднородным распределением и в данном случае характеризует взаимное расположение кластеров миРНК.

Для сравнения с помощью поиска по гомологии мы нашли точные копии миРНК и их позиции и с их помощью построили аналогичную функцию ближайших расстояний между копиями миРНК (рис. 1, б) (подробнее анализ копий описывается в следующих разделах). Показатель степени E для распределения расстояний между копиями оказался близким к единице (табл. 1), что приближенно соответствует случайно-равномерному расположению со средним расстоянием между копиями, равным $4,3 \times 10^6$ п.о. Интересно, что пары миРНК с ее копией находятся в одной и той же хромосоме в малом (164) числе случаев.

Таблица 1

Параметры аппроксимации кумулятивных функций распределения ближайших расстояний между миРНК и между точными копиями миРНК на расстояниях больше 10^4 нт

Тип распределения	C	D	E
Расстояния между миРНК	-2×10^6	$6,2 \times 10^6$	0,6
Расстояния между точными копиями миРНК	$-5,3 \times 10^4$	$4,3 \times 10^6$	0,9

Таким образом, наблюдается сильная скоррелированность расположения миРНК внутри хромосом на расстояниях до 5 тыс. нт в согласии с наблюдениями об организации миРНК в кластеры (Altuvia *et al.*, 2005). В отличие от миРНК их копии располагаются независимо как друг от друга, так и от оригиналов.

Содержание миРНК и их копий в хромосомах: роль мобильных элементов в копировании миРНК

Для исследования механизма распространения миРНК мы рассмотрели распределение известных зрелых миРНК человека и их копий по отдельным хромосомам. Для этого сначала мы нашли все последовательности, которые совпадают с известными миРНК. Общее число найденных последовательностей – 2122 шт., приблизительно поровну в каждой из цепей ДНК.

Далее мы разделили набор из этих 2122 последовательностей миРНК вместе с их окрестностями на две части в зависимости от степени совпадения их окрестностей с известными пре-миРНК. Рассматриваемые окрестности каждой миРНК по размеру соответствовали гену этой миРНК. В группу оригинальных генов входят 785 последовательностей, которые являются либо известными пре-миРНК (695 шт.), либо их точными копиями (90 шт.). Вторую группу копий генов составляют остальные 1337 последовательностей, окрестности которых отличаются от известных пре-миРНК.

Затем мы определили содержание последовательностей из этих двух групп в отдельных хромосомах (рис. 2). Из рис. 2, а видно, что межхромосомное распределение генов миРНК сильно неоднородно. 14-я, Y- и особенно 19-я хромосомы характеризуются наибольшим числом (рис. 2, а) и удельным (в расчете на длину хромосомы рис. 2, в) количеством пре-миРНК. В отличие от самих пре-миРНК распределение их копий более однородно (рис. 2, б), а плотность почти одинакова за исключением 19-й хромосомы (рис. 2, г). Иначе говоря, размер хромосомы является основным параметром, который определяет содержание в ней гомологов пре-миРНК, а интенсивность копирования между хромосомами примерно равна.

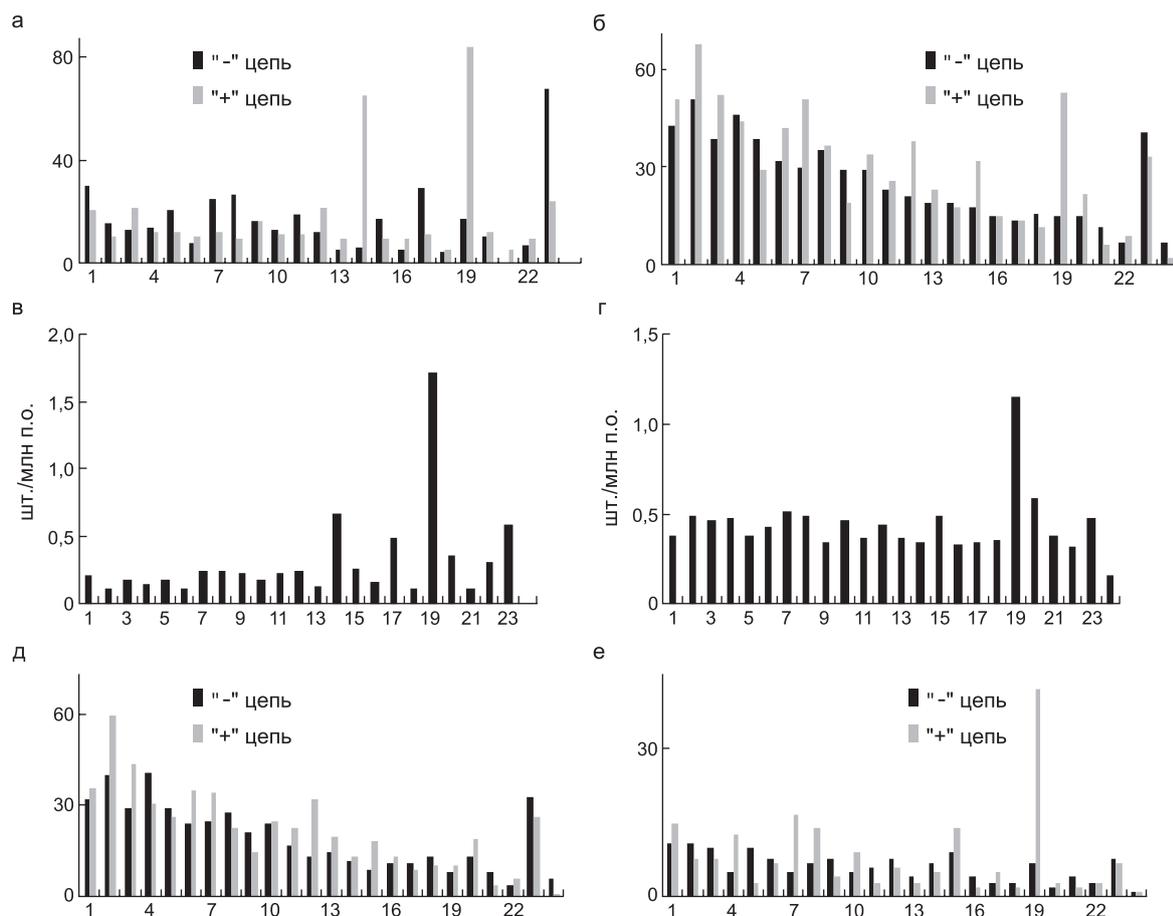


Рис. 2. Межхромосомное распределение генов миРНК и сходных с ними последовательностей в геноме человека.

а – количество генов миРНК; б – количество копий генов миРНК; в – средняя плотность генов миРНК в хромосоме в штуках на 1 млн нт; г – средняя плотность копий генов миРНК; д – количество копий генов миРНК из семейства *hsa-mir-548*; е – количество копий генов миРНК без учета миРНК семейства *hsa-mir-548*. По оси *x* отображается номер хромосомы (позиция 23 соответствует X-хромосоме, 24 – Y-хромосоме). По оси *y* отображается число наблюдений.

Большую часть 19-й хромосомы составляют повторяющиеся последовательности: 55 % против 44,8 % в среднем по геному (Grimwood *et al.*, 2004). Из этих повторов 19-й хромосомы большую часть составляют SINE элементы (Grimwood *et al.*, 2004). Для выявления связи между копированием миРНК и мобильными генетическими элементами мы сравнили последовательности генов миРНК и мобильных элементов посредством веб-сервера RepeatMasker (Smith, 1999), параметры поиска по умолчанию.

Для 116 из 695 известных пре-миРНК мы нашли высокую гомологию с известными ДНК-транспозонами и ретроэлементами. Эти пре-миРНК имеют подавляющее число копий (1285 из 1337, или 96 % от общего числа обна-

руженных копий генов). Наиболее часто гены миРНК копируются мобильными элементами *mde1*, *FLAM_A*, *AluYh9* (рис. 3).

Из этих 1285 копий пре-миРНК большинство (1030, или 83%) являются копиями пре-миРНК из семейства *hsa-mir-548* и принадлежат приматоспецифичному ДНК-транспозону *mde1*, а остальные соответствуют *SINE*, *LINE* и другим ретроэлементам. Межхромосомные распределения копий пре-миРНК из семейства *hsa-mir-548* и остальных копий показаны на рис. 2, д и 2, е соответственно.

Таким образом, обширная распространенность копий миРНК по геному человека объясняется присутствием в мобильных элементах последовательностей некоторых пре-миРНК. В хромосоме 19 гомологи пре-миРНК, в основном

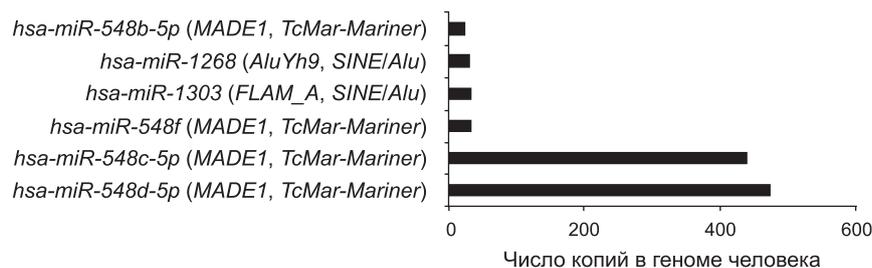


Рис. 3. Пять миРНК с наибольшим числом копий и соответствующие им мобильные элементы.

Ось *y* – название миРНК (название мобильного элемента, семейство), ось *x* – число обнаруженных гомологов миРНК в геноме человека.

копии генов недавно обнаруженных миРНК, *hsa-mir-1268* и *hsa-mir-1303*, контролируются внутрихромосомными дупликациями SINE элементов. В целом самую значительную роль в геномной экспансии гомологов пре-миРНК играет транспозон *madel*, который равномерно копирует в другие хромосомы гены семейства *hsa-mir-548*. Особенности эволюции *madel* рассмотрены ниже.

Эволюция мобильного элемента *madel*

Копии транспозона *madel* обнаружены в геномах приматов *H. sapiens*, *P. troglodytes*, *P. abelii*, *M. mulatta* и *C. jacchus*. *Madel* элементы относятся к семейству неавтономных ДНК транспозонов *Tc1/Mariner* и имеют длину 80 п.н. с двумя инвертированными концевыми повторами по 37 п.н. и вставкой 6 п.н. между ними (Morris, 2008). Согласно базе UCSC Table Browser, в геномах этих организмов присутствует более 6 тыс. копий *madel*. Для каждого генома мы рассчитали число копий в зависимости от величины сходства с полной последовательностью транспозона, а также число совпадающих копий между геномами. Оказалось, что это распределение фрагментов приблизительно одинаково для всех пяти геномов (рис. 4), что свидетельствует о сходстве крупномасштабной эволюции *madel* у всех видов.

Само число повторов между геномами изменяется незначительно, что делает маловероятным такой сценарий размножения повторов, при котором ближайший общий предок приматов наследовал небольшое количество *madel*. При этом парные числа совпадающих копий сильно различны: доля консервативных копий меняется от 0,3 до 38 % (табл. 2). Данные из табл. 2 соот-

ветствуют эволюции со средними для геномов скоростями по филогенетическому дереву приматов около 10^{-3} замен на позицию за 1 млн лет (Steiper, Young, 2006). Так, простые вычисления показывают, что ожидаемое при нейтральном расхождении число совпадений между человеком и шимпанзе равно 2935 фрагментам (ср. с наблюдаемым 2949) (табл. 2) для различия их геномов в 0,0012 нуклеотидов на позицию (Mikkelsen *et al.*, 2005) и $7715 = (7823 + 7607)/2$ фрагментам с наблюдаемым средним размером в 60 нт.

В настоящее время в базе miRBase (релиз 12.0) присутствует 29 генов миРНК человека, 18 генов шимпанзе и 6 макаки из семейства *hsa-mir-548*, сходство последовательностей которых с *madel* свидетельствует об их происхождении от этого мобильного элемента. Близки друг к другу и сами миРНК этого семейства, отлича-

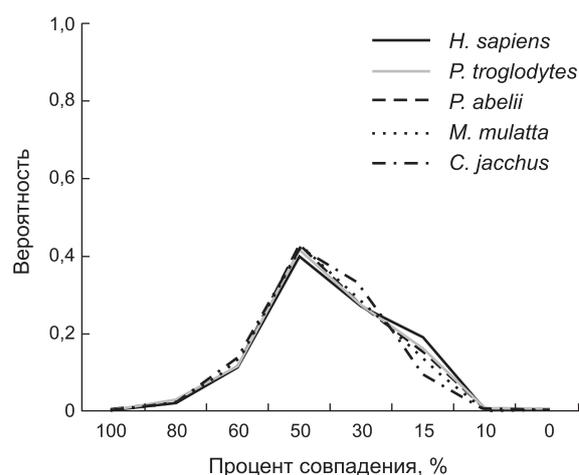


Рис. 4. Распределение вероятностей сходства *madel* и его фрагментов из пяти геномов приматов. Поиск проведен с помощью веб-сервера RepeatMasker (Smith, 1999) с параметрами по умолчанию.

Таблица 2

Количество полностью совпадающих фрагментов *made1* между геномами 5 приматов

Приматы	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>P. abelii</i>	<i>M. mulatta</i>	<i>C. jacchus</i>
<i>H. sapiens</i>	7823	2949	880	183	32
<i>P. troglodytes</i>	2949	7607	839	191	29
<i>P. abelii</i>	880	839	7705	187	28
<i>M. mulatta</i>	183	191	187	6706	21
<i>C. jacchus</i>	32	29	28	21	6446

ясь не более чем на 4 позиции. Мы сравнили последовательности *made1* и человеческих генов семейства *hsa-mir-548*, нашли одиночные замены и сравнили районы мобильного элемента по числу замен. Оказалось, что мобильный элемент в целом однородно вариабелен: район миРНК и область вне ее имеют одинаковое число замен на позицию. Эволюция пре-миРНК более чем в два раза быстрее среднегеномной и происходит со скоростью $2,3 \times 10^{-3}$ замен на позицию за миллион лет, близкой к скорости быстро эволюционирующей фракции *Alu*-повторов (Ruiz-Narvaez, Campos, 2008).

На основании множественного выравнивания последовательностей пре-миРНК этого семейства и полных последовательностей *made1* нами были построены укорененные филогенетические деревья, отдельно для «+» и «-» цепей ДНК-транспозона (рис. 5).

Из рис. 5 видно, что пре-миРНК семейства *hsa-mir-548* высокомологичны транспозону, а

ветвления дерева определяются малым числом замен, что делает точную топологию дерева вблизи корня малодостоверной, но всегда приближенно звездообразной для разных миРНК. С учетом одинаковой скорости замен в пре-миРНК и консервативности района миРНК вблизи концов дерева можно заключить, что мутации именно в этом районе отвечают за расхождение вблизи корня дерева.

Часто одни и те же миРНК можно обнаружить у разных видов (Griffiths-Jones, 2004). Для семейства *hsa-mir-548* межвидовой консерватизм проявляется в том, что к настоящему моменту эти миРНК обнаружены также у макаки, орангутанга и шимпанзе. В частности, 11 известных миРНК шимпанзе из этого семейства, в том числе 4 пре-миРНК, полностью совпадают с человеческими. Перечисленные в этом разделе обстоятельства приводят к выводу, что включение мобильного элемента *made1* в систему молчания генов произошло у общего предка

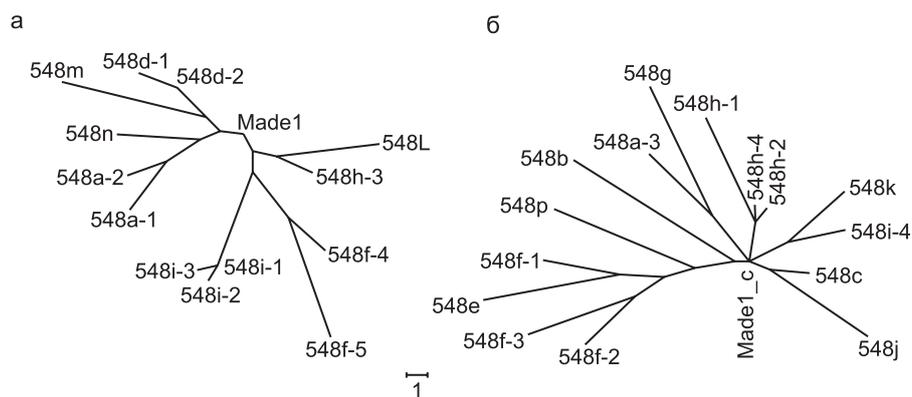


Рис. 5. Укорененные филогенетические деревья для семейства генов *hsa-mir-548* и плюс (а) и минус (б) цепей *made1*, построенные программой MEGA4 методом максимальной парсимонии.

Деревья построены программой MEGA4 методом максимальной парсимонии. Длина ребер пропорциональна величине отличия последовательностей. Не включен ген *hsa-mir-548o*, из-за вставки отличающийся более чем на 60 % от *made1*. Гены с числовым суффиксом, например, *hsa-mir-548d-1* и *hsa-mir-548d-2*, содержат одинаковые миРНК.

приматов образованием современных миРНК в канонической форме повтора. Разнообразие миРНК было достигнуто не только нуклеотидными заменами, но и сдвигом позиций миРНК. Незатронутыми мутациями остались короткие повторы U_n и A_n вблизи краев миРНК (рис. 6).

Благодаря палиндромной структуре элемента *madel*, он способен образовывать шпильки, которые сходны по своим характеристикам с вторичными структурами пре-миРНК (рис. 6). Мы провели анализ сохранения канонической шпильки после копирования *madel* в геноме человека. Для этого из почти 8 тысяч копий *madel* в геноме человека мы выделили около 2900 последовательностей, накладывая ограничения, которые характерны для генов миРНК человека (Nam *et al.*, 2005). Затем мы провели их выравнивание с последовательностью полного *madel*-транспозона и среди обнаруженных более 20 тысяч замен рассмотрели те из них, которые изменяют оба нуклеотида в комплементарных парах канонической структуры. Всего в транспозоне было обнаружено 2170 замен пары, из них 1624 сохраняют комплементарность, т. е. являются коадаптивными (рис. 6). Для сравнения мы ввели то же число случайных одиночных замен на том же количестве идентичных последовательностей *madel*. Всего в тесте Монте Карло было получено около 850 замен пары, из которых около 300 сохраняют комплементарность: обе величины примерно на 1300 меньше наблюдаемых значений.

Таким образом, эволюционирующие копии *madel* склонны сохранять каноническую палиндромную структуру, которая способствует формированию транскриптом устойчивой вторичной структуры в виде шпильки (рис. 6).

Обсуждение и выводы

Полицистронная организация миРНК известна для многих организмов, но лучше изучена в человеке (Bartel, 2004; Cullen, 2004). Однако число миРНК в кластерах и типичный размер кластеров остаются предметом для дискуссии. С одной стороны, при компьютерном поиске кластеров миРНК ранее использовался порог в 3 тыс. нт (Altuvia *et al.*, 2005), с другой, была обнаружена коэкспрессия миРНК, которые находятся на расстояниях вплоть до 50 тыс. нт

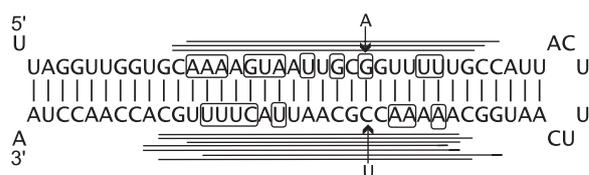


Рис. 6. Вторичная структура транскрипта «+» цепи *madel*.

Горизонтальными линиями обозначены расположения миРНК семейства *hsa-mir-548*. Обведены консервативные нуклеотиды в области миРНК. Стрелками показана коадаптивная замена $GC \rightarrow AU$, наблюдаемая во фрагменте *madel*.

(Baskerville, Bartel, 2005). Здесь при помощи статистических аргументов мы получили промежуточную между упомянутыми выше оценку в 5 тыс. нт. Поэтому число миРНК, которых отнесли к кластерам при компьютерном анализе (Altuvia *et al.*, 2005), является недооцененным.

Дупликации являются основным механизмом распространения миРНК по геному. Главную роль здесь играют короткие приматоспецифичные ДНК-транспозоны и ретроэлементы. Кластеры миРНК в 19-й хромосоме распространяются через *Alu*-повторы, а в целом по геному основная часть копий миРНК дублирована мобильными элементами *madel* в случайные позиции в среднем через $4,3 \times 10^6$ п.о. По нашим наблюдениям, распространение копий *madel* в приматах произошло взрывообразно до расхождения на *H. sapiens*, *P. troglodytes*, *P. abelii*, *M. mulatta* и *C. jacchus*. Для окончательной проверки такого сценария необходимо сравнение позиций мобильного элемента между геномами.

Тогда же независимо друг от друга возникли современные миРНК семейства *hsa-mir-548* как результат адаптивных мутаций в части копий транспозона, которые и были рекрутированы в механизм молчания генов. Мутации не затронули короткие повторы U_n и A_n вблизи краев миРНК и привели к смещениям позиций референтной миРНК, образовав целое семейство *hsa-mir-548*. Ранее для 7 генов миРНК из *hsa-mir-548* семейства было известно их происхождение от *madel* (Piriyarongsa *et al.*, 2007). В этой работе мы обнаружили 29 таких генов.

Некоторые семейства миРНК имеют высокий уровень экспрессии, древнее происхождение, выполняют важные функции и эволюцион-

но-инвариантны. Вариабельность функций миРНК контролируется как переключениями между мишенями консервативных миРНК, так и узкоконсервативной видоспецифичностью миРНК. Мобильные элементы, соответствующие наиболее копируемому миРНК (*Made1*, *FLAM_A*, *AluYh9*), встречаются только у приматов, что позволяет предположить участие происходящих от них миРНК в фундаментальных регуляторных процессах в данных видах. Интерес представляет проявление этих миРНК в фенотипе, специфичном для приматов. Так, миРНК, которые распространились через *Alu*-повторы по 19-й хромосоме, предпочтительно экспрессируются в плаценте и мозге (Noguer-Dance *et al.*, 2010). МиРНК семейства *hsa-mir-548* предположительно причастны к регулированию экспрессии генов при раке (Piriyaopongsa *et al.*, 2007).

Устойчивая шпилечная структура и широкая распространенность мобильных элементов, которые содержат миРНК, свидетельствуют о существовании большого числа молчащих генов миРНК. В связи с этим исследование механизма превращения транспозона в пре-миРНК оказывается важным как для диагностики и терапии, так и для развития компьютерных методов предсказания миРНК.

Благодарности

Работа поддержана Междисциплинарным интеграционным проектом СО РАН 119 «Постгеномная биоинформатика: компьютерный анализ и моделирование молекулярно-генетических систем» и Госконтрактом № 10104-37/П-24/110-323/020610-004 по теме «Эволюция гео-биологических систем» по подпрограмме 2 программы 24 фундаментальных исследований Президиума РАН «Происхождение биосферы и эволюция гео-биологических систем».

Литература

Капитонов В.В., Титов И.И. Порядок интронов и дальние корреляции в нуклеотидных последовательностях // Докл. АН. 1994. Т. 337. С. 810–812.
Лобзин В.В., Четкин В.Р. Порядок и корреляции в геномных последовательностях ДНК. Спектральный анализ // Усп. физ. наук. 2000. Т. 17. Вып. 1. С. 57–81.

Altschul S.F., Madden T.L., Schäffer A.A. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucl. Acids Res. 1997. V. 25. P. 3389–3402.
Altuvia Y., Landgraf P., Lithwick G. *et al.* Clustering and conservation patterns of human microRNAs // Nucl. Acids Res. 2005. V. 33. № 8. P. 2697–2706.
Bartel D.P. MicroRNAs: genomics, biogenesis, mechanism, and function // Cell. 2004. V. 116. № 2. P. 281–297.
Baskerville S., Bartel D. Microarray profiling of human microRNAs reveals frequent coexpression with neighboring miRNAs and host genes // RNA. 2005. V. 11. P. 241–247.
Consortium I.H.G.S. Initial sequencing and analysis of the human genome // Nature. 2001. V. 409. № 6822. P. 860–921.
Cullen B.R. Transcription and processing of human microRNA precursors // Mol. Cell. 2004. V. 16. № 6. P. 861–865.
Griffiths-Jones S. The microRNA Registry // Nucl. Acids Res. 2004. V. 32. P. D109–D111.
Griffiths-Jones S., Saini H.K., van Dongen S., Enright A.J. miRBase: tools for microRNA genomics // Nucl. Acids Res. 2008. V. 36(Database Issue). P. D154–D158.
Grimwood J., Gordon L.A., Olsen A. *et al.* The DNA sequence and biology of human chromosome 19 // Nature. 2004. V. 428. P. 529–535.
Hofacker I.L., Fontana W., Stadler P.F. *et al.* Fast folding and comparison of RNA secondary structures // Monatshefte f. Chemie. 1994. V. 125. P. 167–188.
Jeanmougin F., Thompson J.D., Gouy M. *et al.* Multiple sequence alignment with Clustal X // Trends Biochem. Sci. 1998. V. 23. № 10. P. 403–405.
Karlin S., Brendel V. Patchiness and correlations in DNA sequences // Science. 1993. V. 259. P. 677–680.
Karolchik D., Hinrichs A.S., Furey T.S. *et al.* The UCSC Table Browser data retrieval tool // Nucl. Acids Res. 2004. V. 32. P. D493–D496.
Lee R.C., Feinbaum R.L., Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* // Cell. 1993. V. 75. P. 843–854.
Mikkelsen T.S., Hillier L.W., Eichler E.E. *et al.* Initial sequence of the chimpanzee genome and comparison with the human genome // Nature. 2005. V. 437. P. 69–87.
Morris K.V. RNA and the regulation of gene expression: a hidden layer of complexity. Norfolk: Horizon Scientific Press, 2008. 228 p.
Nam J.W., Shin K.R., Han J. *et al.* Human microRNA prediction through a probabilistic co-learning model of sequence and structure // Nucl. Acids Res. 2005. V. 33. № 11. P. 3570–3581.
Noguer-Dance M., Abu-Amero S., Al-Khtib M. *et al.* The primate-specific microRNA gene cluster

- (C19MC) is imprinted in the placenta // Hum. Mol. Genet. 2010. V. 19. № 18. P. 3566–3582.
- Piriyapongsa J., Jordan I.K. A family of human microRNA genes from miniature inverted-repeat transposable elements // PLoS ONE. 2007. V. 2. N e203. P. 1–11.
- Ruiz-Narváez E.A., Campos H. Evolutionary rate heterogeneity of *Alu* repeats upstream of the *APOA5* gene: do they regulate *APOA5* expression? // J. Hum. Genet. 2008. V. 53. № 3. P. 247–253.
- Smalheiser N.R., Torvik V.I. Mammalian microRNAs derived from genomic repeats // Trends Genet. 2005. V. 21. P. 322–326.
- Smith A., Green P. RepeatMasker. 1999. available at <http://repeatmasker.org>.
- Sober E. Parsimony in systematics: philosophical issues // Annu. Rev. Ecol. Syst. 1983. V. 14. P. 335–357.
- Steiper M.E., Young N.M. Primate molecular divergence dates // Mol. Phylogenet. Evol. 2006. V. 41. P. 384–394.
- Tamura K., Dudley J., Nei M., Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 // Mol. Biol. Evol. 2007. V. 24. P. 1596–1599.
- Titov I.I., Vorobiev D.G., Ivanisenko V.A., Kolchanov N.A. A fast genetic algorithm for RNA secondary structure analysis // Russ. Chem. Bull. 2002. V. 51. № 7. P. 1135–1144.

THE ANALYSIS OF miRNA DUPLICATION IN HUMAN GENOME AND THE ROLE OF TRANSPOSON EVOLUTION IN THIS PROCESS

I.I. Titov^{1,2}, P.S. Vorozheykin¹

¹ Novosibirsk State University, Novosibirsk, Russia, e-mail: titov@bionet.nsc.ru;

² Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

Summary

MiRNA duplication is one of the modes of their evolution. In this paper we show that human miRNAs lie mainly in clusters of two or more of up to 5 000 nt size while their copies are randomly scattered with the average pair distance of $4,3 \times 10^6$ bp. We have compared neighborhood of miRNA copies with the sequences of the mobile elements and have found that 96 % of miRNA homologs propagate by the DNA transposons and retroelements. Amongst them the primate-specific element *made1*, copies of which contain a family of miRNAs *hsa-mir-548*, brings the greatest contribution. We speculate that *made1* evolution is the following. These mobile elements have emerged and multiplied explosively over the genome of primate common ancestor, and diverged later with the average genomic rates. Simultaneously and independently of each other a modern family of miRNAs *hsa-mir-548* emerged as a result of adaptive mutagenesis in transposon copies, which were recruited into the silencing system. Nucleotide changes found in the transposon during its transformation into the miRNA genes, as well as the numerous miRNA copies could be useful for predicting the new miRNAs.

Key words: miRNA, transposable element, DNA transposon, *made1*, computer analysis, secondary structure, evolution.