

Выравнивание последовательностей РНК с вторичной структурой: предшественники микроРНК

Хазиев Г.А.^{*1}, Зверков О.А.^{†1}, Селиверстов А.В.^{‡1}, Любецкий В.А.^{§1}

¹Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук, Москва, Россия

Аннотация. Цель работы состоит в анализе изменений в ходе эволюции тех пре-микроРНК, которые связаны у млекопитающих с долговременной памятью и обучением. Большие изменения наблюдаются в пре-микроРНК miR-195 и miR-206 у мышевидных грызунов (надсемейство Muridae) и у опоссума. Но изменения малы у всех приматов. Более того, для miR-206 показана связь изменений с логарифмом максимальной видовой продолжительности жизни. Это позволяет предполагать ослабление действия стабилизирующего отбора на miR-206 у видов с малой максимальной продолжительностью жизни. Напротив, обе miR-134 и miR-182 более консервативные, а небольшие изменения никак не связаны с продолжительностью жизни. Также рассмотрены изменения вторичной структуры пре-микроРНК. Для этого исследования авторами был разработан алгоритм для усечения исходной последовательности до несовершенного палиндрома, который близок к совершенному и получается удалением префикса или суффикса, когда это возможно. Такой несовершенный палиндром приблизительно соответствует шпильке на РНК. Существенно, что границы палиндрома устанавливаются в ходе оптимизации. Алгоритм реализован пакетом программ на языке Python. Программы доступны по адресам <http://lab6.iitp.ru/-/pali> и <http://lab6.iitp.ru/-/trimmers>. Поиск совершенного палиндрома, близкого в метрике Левенштейна к исходной последовательности, часто даёт несколько оптимальных палиндромов. На этом пути для miR-195 показана связь числа оптимальных палиндромов с логарифмом от максимальной видовой продолжительностью жизни. Для miR-134, miR-182 и miR-206 такой зависимости нет. Отметим, что miR-195 и miR-206 различаются между собой по типу изменений.

Ключевые слова: микроРНК, долговременная память, продолжительность жизни, выравнивание, палиндром, шпилька, редакционное расстояние, млекопитающие, биоинформатика.

ВВЕДЕНИЕ

МикроРНК (miR) — некодирующие РНК длиной от 18 до 25 нуклеотидов. МикроРНК участвуют в регуляции экспрессии генов, комплементарно связывая мРНК. Зрелые микроРНК образуются из пре-микроРНК, которые имеют большую длину и обычно имеют структуру несовершенного палиндрома, формирующего шпильку РНК [1].

Существует несколько методов для предсказания вторичной структуры РНК по одной последовательности [2, 3]. При этом одна последовательность может формировать

*khaziev@iitp.ru

†zverkov@iitp.ru

‡slvstv@iitp.ru

§lyubetsk@iitp.ru

разные структуры с близкими значениями свободной энергии [4, 5]. Роль альтернативных структур уточняется при выравнивании различных последовательностей, учитывающем их вторичную структуру [6, 7]. Рассмотрена эволюция пре-микроРНК, участвующих в обучении и формировании долговременной памяти [8, 9]. В частности, сверхэкспрессия miR-195 в гиппокампе крыс защищает от развития деменции, а её ингибирование приводит к нарушению пространственной памяти [10]. Усиление экспрессии кластера микроРНК miR-183/96/182 в гиппокампе способствует формированию долговременной памяти. Снижение содержания miR-183/96/182 приводит к ухудшению памяти в старости, но память может быть улучшена при сверхэкспрессии этого кластера [11]. С другой стороны, в амигдале при формировании амигдала-зависимой долговременной памяти отмечается снижение содержания miR-182, а её сверхэкспрессия приводит к нарушению долговременной памяти [12]. МикроРНК miR-134 и miR-206 негативно влияют на формирование долговременной памяти. Сверхэкспрессия miR-134 в CA1 области гиппокампа приводит к значительному ухудшению формирования долговременной памяти [13]. Нарушение miR-134 и miR-124-зависимой регуляции лежит в основе когнитивных дисфункций при болезни Альцгеймера [14, 15]. Повышенный уровень miR-206 наблюдается в височной коре головного мозга человека при болезни Альцгеймера [16]. Снижение уровня miR-206 приводит к улучшению памяти [17].

Редакционное расстояние (или расстояние Левенштейна [18]) между последовательностями и наибольшая общая подпоследовательность используются в биоинформатике; они легко вычислимы за квадратичное время [19]. Известно несколько алгоритмов для точного вычисления редакционного расстояния за немного меньшее время [20, 21]. Математическое ожидание длины наибольшей общей подпоследовательности двух случайных двоичных последовательностей рассмотрено в работе [22]. Аппроксимировать редакционное расстояние можно за существенно меньшее время [23]. Для фиксированной верхней границы на редакционное расстояние известны алгоритмы, время работы которых линейно зависит от длины последовательностей [24]. Также созданы параллельные алгоритмы для поиска наибольшей общей подпоследовательности [25, 26]. Перечисление всех наибольших общих подпоследовательностей рассмотрено в работе [27]. Для перечисления хордовых диаграмм, используемых для описания несовершенных палиндромов, также применяется гафниан симметричной матрицы [28]. Поиск наибольших подстрок, то есть наибольших общих подпоследовательностей, элементы которых идут подряд, в частности, с данным числом несовпадений рассмотрен в работе [29]. Наряду с редакционным расстоянием используют и другие оценки сходства последовательностей [30].

Анализ структуры РНК связан с поиском несовершенных палиндромов [31, 32]. В статье [33] обсуждаются различные определения палиндромов. Для приложений в биоинформатике удобно считать, что алфавит состоит из четырёх букв $\{A, C, T, G\}$ или $\{A, C, U, G\}$, которые связаны отношением комплементарности: A комплементарна к T или к U, C комплементарна к G. Для конкатенации двух последовательностей выполнено $c(xy) = c(y)c(x)$. Иными словами, комплементарная (reverse complement) последовательность получается одновременной перестановкой букв в обратном порядке и заменой букв на комплементарные буквы. Например, $c(AACT) = AGTT$. Совершенным (perfect) палиндромом называется последовательность x , которая комплементарна самой себе: $c(x) = x$. Поскольку никакая буква не комплементарна себе, совершенный палиндром имеет чётную длину и имеет вид $zc(z)$. Далее всюду предполагается это ограничение. Последовательность, близкую в метрике Левенштейна к палиндрому, называют несовершенным (imperfect) палиндромом.

МАТЕРИАЛЫ И МЕТОДЫ

Усечение последовательностей, содержащих вырожденный палиндром

Для оценки сходства последовательности с палиндромом удобно использовать функцию $\text{imp}(x)$, значение которой равно отношению редакционного расстояния от оптимального палиндрома до исходной последовательности к длине исходной последовательности.

$$\text{imp}(x) = \frac{\min\{\text{dist}(x, wc(w)) \mid x = wz\}}{|x|}.$$

Корректность такого определения следует из равенства $\text{imp}(x) = \text{imp}(c(x))$.

Значение функции $\text{imp}(x)$ говорит, насколько последовательность x отличается от совершенного палиндрома, и равно нулю для совершенных палиндромов. Алгоритм вычисления функции $\text{imp}(x)$ служит модификацией алгоритма поиска скрытого палиндрома из статьи [32]. Дополнительное уменьшение времени работы достигается, поскольку заранее известна матрица редакционных расстояний между префиксами одной последовательности. Между префиксами, длины которых равны j и k , редакционное расстояние вычислимо по формуле $|j - k|$.

Также разработан алгоритм для выделения из последовательности x более короткого несовершенного палиндрома y со значением $\text{imp}(y) < \text{imp}(x)$, получаемого удалением из исходной последовательности x префикса или суффикса, когда это возможно. Эта операция называется усечением. Основной идеей наших алгоритмов является проверка, верно ли, что одна из оптимальных длин скрытого палиндрома в последовательности x отличается от длины x . Если так, то программа усекает x на половину абсолютной величины разности длин исходной последовательности x и скрытого палиндрома. Например, в случае, когда исследуемая последовательность $x = wc(w)z$ равна конкатенации палиндрома и короткого хвоста, палиндром $wc(w)$ будет оптимальным для нашего алгоритма. При этом оптимальный палиндром короче исходной последовательности. С другой стороны, для последовательности $x = ywc(w)$ с достаточно коротким префиксом y оптимальный палиндром будет длиннее исходной последовательности. Однако величина $\text{imp}(x)$ не зависит от того, расположен ли палиндром в начале или в конце исходной последовательности. Хотя от этого зависит длина оптимального палиндрома.

Листинги программ доступны по адресам <http://lab6.iitp.ru/-/pali> и <http://lab6.iitp.ru/-/trimmers>. Использование библиотеки Numba позволяет значительно уменьшить время работы программы.

Вторичная структура РНК предсказана программой RNAstructure [4, p. 4]. Деревья построены методом максимального правдоподобия (maximum likelihood), реализованным в программе MEGA [34]. Для визуализации применялся веб-сервис iTOL [35].

Пре-микроРНК

К названию микроРНК, которые отличаются на один или два нуклеотида, приписывается дополнительная строчная буква. Пре-микроРНК, дающие начало идентичным микроРНК, но локализованные в разных местах генома, имеют в названии дополнительную цифру, отделённую дефисом. Когда две зрелые микроРНК образуются из двух различных концов исходной пре-микроРНК, добавляется суффикс $-3p$ или $-5p$. Приставка *mir*- используется для обозначения пре-микроРНК, а *miR*- обозначает зрелую микроРНК. Рассмотрено около 38.5 тысяч пре-микроРНК, доступных в базе данных miRBase [36]. Из рассмотрения были исключены вырожденные последовательности (120

последовательностей), в которых некоторые позиции определены неоднозначно.

Таблица 1. Максимальная видовая продолжительность жизни (МПЖ) в годах и числа рассмотренных mir-206, mir-195, mir-182 и mir-134

Вид	Русское название	МПЖ	206	195	182	134
<i>Homo sapiens</i>	Человек	120	1	1	1	1
<i>Pan troglodytes</i>	Обыкновенный шимпанзе	68	1	1	1	1
<i>Pan paniscus</i>	Карликовый шимпанзе	55		1		1
<i>Gorilla gorilla</i>	Западная горилла	60.1	1	1	1	1
<i>Pongo pygmaeus</i>	Калимантанский орангутан	59	1	1	1	1
<i>Nomascus leucogenys</i>	Белощёкий хохлатый гиббон	44.1		1	1	
<i>Macaca mulatta</i>	Макак-резус	40	1	1	1	1
<i>Macaca nemestrina</i>	Свинохвостый макак	37.6	1			1
<i>Callithrix jacchus</i>	Обыкновенная игрунка	22.8	1		1	1
<i>Microcebus murinus</i>	Мышиный лемур	18.2		1	1	
<i>Otolemur garnettii</i>	Галаго Гарнетта	20		1	1	
<i>Tupaia chinensis</i>	Тупайя	(12)		1		
<i>Mus musculus</i>	Домовая мышь	4	1	2	1	1
<i>Cricetulus griseus</i>	Китайский хомячок	(5.3)	1	1	1	1
<i>Cavia porcellus</i>	Морская свинка	12	1	1	1	1
<i>Rattus norvegicus</i>	Серая крыса	3.8	1	1	1	1
<i>Oryctolagus cuniculus</i>	Дикий кролик	13	1	1	1	1
<i>Canis familiaris</i>	Собака	27	1	1	1	1
<i>Bos taurus</i>	Бык	20	1	1	1	1
<i>Capra hircus</i>	Домашняя коза	20.8	1	1	1	1
<i>Ovis aries</i>	Домашняя овца	22.8				1
<i>Sus scrofa</i>	Кабан	27	1	1	1	
<i>Equus caballus</i>	Домашняя лошадь	57	1	1	1	1
<i>Eptesicus fuscus</i>	Аргентинский кожан	19	1		1	
<i>Pteropus alecto</i>	Бавеанская летучая лисица	20.3	1	1	1	
<i>Dasypus novemcinctus</i>	Девятипоясный броненосец	22.3	1	1	1	1
<i>Monodelphis domestica</i>	Домовый опоссум	5.1	1	1	1	
<i>Ornithorhynchus anatinus</i>	Утконос	22.6	1		1	

Максимальная продолжительность жизни

Сведения о максимальной видовой продолжительности жизни (МПЖ) получены из базы данных AnAge [37], значение Longevity. Для человека принято значение МПЖ 120 лет. В табл. 1 собраны данные о МПЖ у рассматриваемых видов.

Если МПЖ не указана в базе данных AnAge, то принято значение для близкого вида, оно указано в скобках. У китайского хомячка указана МПЖ для барабинского хомячка *Cricetulus barabensis*. Для разных видов тупайи МПЖ составляет: *Tupaia belangeri* 11.1, *T. glis* 12.4, *T. minor* 12.2, *T. tana* 11.6 лет.

РЕЗУЛЬТАТЫ

Поиск палиндромов

Среднее значение величины $\text{imp}(x)$ для всех рассмотренных невырожденных последовательностей равно 0.19175. Медиана равна 0.19186, среднеквадратичное отклонение равно 0.04148. Минимальное значение равно нулю для последовательности hsv1-mir-H11 из *Herpes Simplex Virus 1*. Это подтверждает высокое сходство большинства этих РНК с палиндромами, что согласуется с образованием этими РНК шпилек, которые важны для формирования микроРНК. При этом длина оптимального палиндрома может заметно отличаться от длины исходной последовательности.

Далее подробно рассмотрены четыре пре-микроРНК, которые связаны с формированием долговременной памяти, у млекопитающих. В частности, последовательности пре-микроРНК mir-206 различаются по длине, но вблизи петли на шпильке РНК почти всегда выравниваются без вставок. При сравнении ортологичных пре-микроРНК не обнаружено значимой связи между величиной $\text{imp}(x)$ и МПЖ у млекопитающих. Изменения $\text{imp}(x)$ маленькие, что говорит о консервативности вторичной структуры РНК. Среднее значение $\text{imp}(x)$ для пре-микроРНК mir-134, mir-182, mir-195 и mir-206 равно 0.193, что близко к среднему по всем пре-микроРНК.

При поиске совершенного палиндрома, близкого в метрике Левенштейна к исходной последовательности, часто бывает несколько оптимальных палиндромов. Но в рассматриваемых случаях изменение числа оптимальных палиндромов мало влияет на стабильность вторичной структуры РНК, поскольку не приводит к формированию альтернативных структур с близкой свободной энергией. Среди рассмотренных у млекопитающих пре-микроРНК mir-134, mir-182, mir-195 и mir-206 максимальное число оптимальных палиндромов равно пяти, медиана для разных пре-микроРНК от двух до трёх. Для mir-195 подтверждена связь числа оптимальных палиндромов с логарифмом от МПЖ. Коэффициент корреляции Пирсона равен 0.61, а p -значение равно 0.0015. Для mir-134, mir-182 и mir-206 такой зависимости нет ($p > 0.5$).

Анализ последовательностей

На рис. 1 показано дерево пре-микроРНК mir-206 у 22 млекопитающих, построенное методом максимального правдоподобия. Видно, что наибольшие отличия от других видов наблюдаются у мышевидных грызунов, а также у опоссума. Наблюдается статистически значимая связь между логарифмом МПЖ и длиной пути до листа в дереве. Результаты собраны в табл. 2.

Большие отличия во вторичной структуре пре-микроРНК mir-206 вблизи петли шпильки наблюдаются у мыши и хомячка. Напротив, у мышиноного лемура замена нуклеотида сохраняет вторичную структуру пре-микроРНК mir-206. Здесь пара нуклеотидов U · G, связанных в шпильке РНК, заменяется на пару комплементарных нуклеотидов C · G.

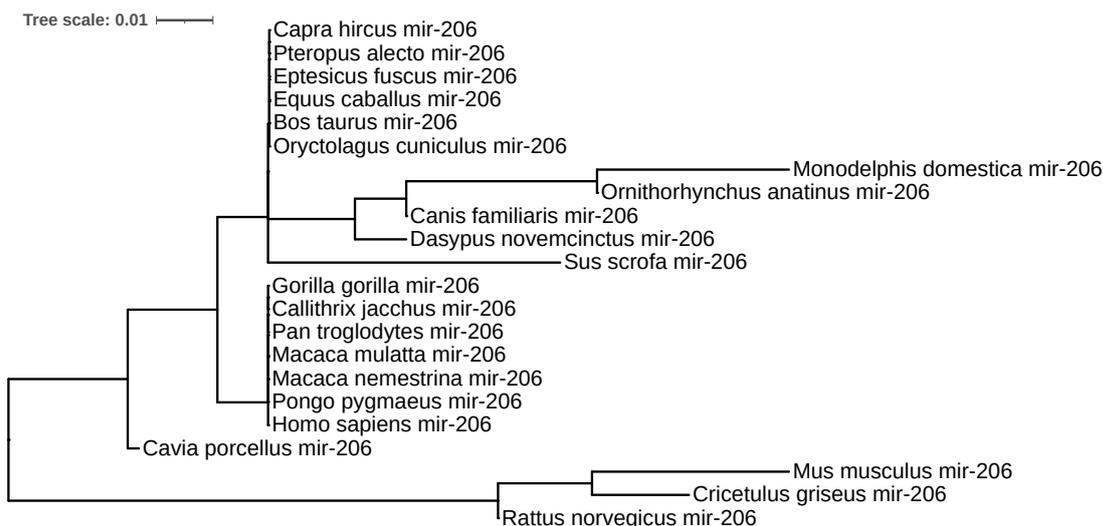


Рис. 1. Дерево пре-микроРНК mir-206, построенное методом максимального правдоподобия.

Таблица 2. Число последовательностей n , коэффициент корреляции Пирсона r между логарифмом МПЖ и длиной пути до листа в дереве для пре-микроРНК mir-206, mir-195, mir-182 и mir-134, а также p -значение

Пре-микроРНК	n	r	p
mir-206	22	-0.67	0.0006
mir-195	24	-0.35	0.09
mir-182	24	0.12	0.6
mir-134	19	0.06	0.8

Последовательности пре-микроРНК mir-195 различаются по длине. У мыши два варианта микроРНК, из которых mir-195b содержит вставку внутри консервативной части. На рис. 2 показано дерево пре-микроРНК mir-195 у 23 млекопитающих, построенное методом максимального правдоподобия. Если исключить mir-195b у мыши, то наибольшие отличия наблюдаются у опоссума. Также выделяются грызуны, особенно мышь и крыса, мышиный лемур, галаго, тупайя, парнокопытные (коза, бык и кабан). Но связь изменений в mir-195 с логарифмом МПЖ выражена слабо, см. табл. 2.

При этом большие отличия во вторичной структуре пре-микроРНК mir-195 вблизи петли шпильки наблюдаются у мыши. Напротив, у мышинного лемура, морской свинки и крысы замены нуклеотидов мало влияют на топологию вторичной структуры РНК. У морской свинки для mir-195 замена нуклеотида сохраняет вторичную структуру пре-микроРНК. Здесь пара комплементарных нуклеотидов G · C, связанных в шпильке РНК, заменяется на пару G · U, также связанных в шпильке.

Замены нуклеотидов в пре-микроРНК mir-195 обычно происходят на большем расстоянии от петли шпильки, чем в mir-206, где замены часто происходят вблизи петли.

На рис. 3 показано дерево пре-микроРНК mir-182 у 24 млекопитающих. На рис. 4 показано дерево пре-микроРНК mir-134 у 19 млекопитающих. В том и другом случае различия пре-микроРНК между видами маленькие и не связаны с МПЖ.

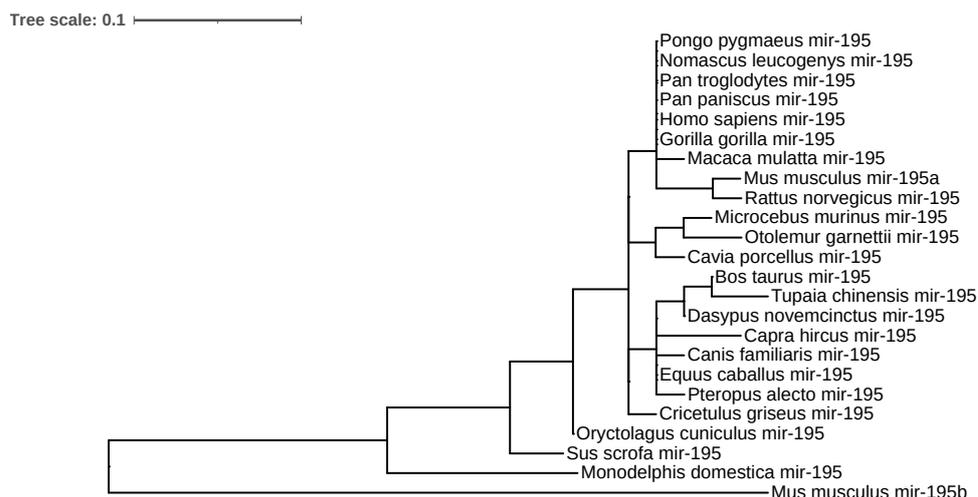


Рис. 2. Дерево пре-микроРНК mir-195, построенное методом максимального правдоподобия.

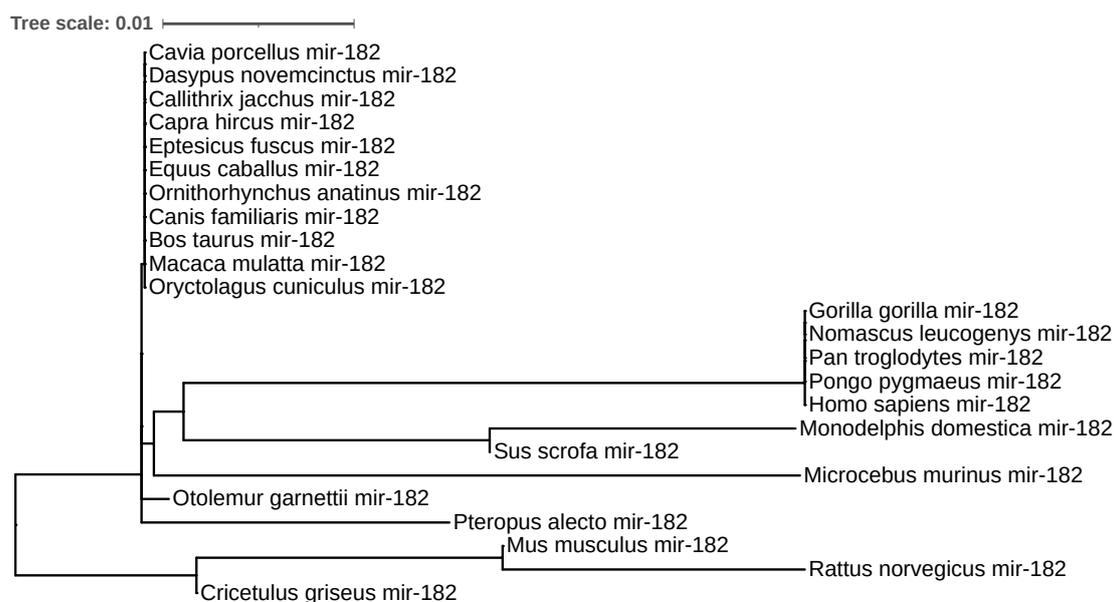


Рис. 3. Дерево пре-микроРНК mir-182, построенное методом максимального правдоподобия.

ОБСУЖДЕНИЕ

Поскольку структура шпильки участвует в процессинге пре-микроРНК [1, 8], изменение этой структуры может оказать влияние на уровень зрелой микроРНК. Наше исследование показало, что в пре-микроРНК mir-195 и mir-206 некомпенсированных отличий, влияющих на вторичную структуру РНК, у мышевидных грызунов (надсемейство Muroidea) и у опоссума заметно больше, чем у других видов. Напротив, у морской свинки для mir-195 и у мышиноного лемура для mir-206 наблюдаются замены нуклеотидов, сохраняющие вторичную структуру пре-микроРНК.

У мышевидных грызунов очень мала продолжительность жизни, максимальная около

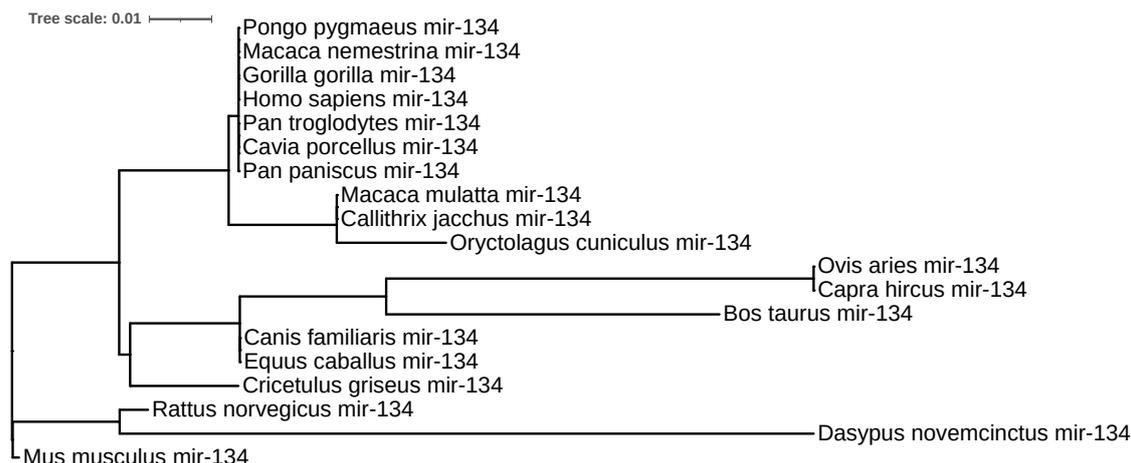


Рис. 4. Дерево пре-микроРНК mir-134, построенное методом максимального правдоподобия.

4 лет. У опоссума МПЖ 5.1 лет. Морская свинка относится к грызунам, но сравнительно далека от Muroidea, её МПЖ 12 лет. Мышиный лемур может прожить ещё дольше. Это позволяет предположить, что у мышевидных грызунов и у опоссума, имеющих малую МПЖ, происходит быстрое накопление замен, заметно влияющих на вторичную структуру пре-микроРНК mir-195 и mir-206.

Рост скорости эволюции с уменьшением продолжительности жизни легко объяснить увеличением числа поколений за время, прошедшее после расхождения видов. Например, у мыши размножение может начаться на 42-й день, а беременность длится 19 дней [37]. У опоссума размножение может начаться на 122-й день, а беременность длится 15 дней [37]. Однако сравнение с другими пре-микроРНК показывает, что изменения в mir-195 и mir-206 гораздо больше, чем в других пре-микроРНК. Это позволяет предполагать ослабление действия стабилизирующего отбора на mir-206 у видов с малой МПЖ. С меньшей уверенностью это можно предполагать и для mir-195.

Росту МПЖ способствует развитие способов защиты от врагов и освоение защищенных экониш [38], что связано с развитием когнитивных способностей. Возможно, при малой продолжительности жизни долговременная память имеет меньшее значение для выживания. С другой стороны, большая продолжительность жизни предполагает длительное сохранение способности к обучению, что предполагает не только запоминание нового, но также отказ от стереотипов, изменение поведения в меняющихся условиях.

ЗАКЛЮЧЕНИЕ

Показано быстрое накопление изменений в структуре пре-микроРНК mir-195 и mir-206 у некоторых видов. При этом mir-195 и mir-206 различаются между собой по типу изменений. Но различается и роль зрелых микроРНК miR-195 и miR-206 и в формировании долговременной памяти. Наибольшие отличия от консенсуса наблюдаются у видов с малой максимальной продолжительностью жизни, включая мышевидных грызунов и опоссума. Напротив, пре-микроРНК mir-182 и mir-134 более консервативны, а небольшие изменения не связаны с продолжительностью жизни.

Работа выполнена в рамках государственного задания ИППИ РАН, утвержденного Минобрнауки России.

СПИСОК ЛИТЕРАТУРЫ

1. Назипова Н.Н. Разнообразие некодирующих РНК в геномах эукариот. *Матем. биология и биоинформ.* 2021. Т. 16. № 2. С. 256–298. doi: [10.17537/2021.16.256](https://doi.org/10.17537/2021.16.256) (Nazipova N. Variety of non-coding RNAs in Eukaryotic genomes. *Math. Biol. Bioinf.* 2021. V. 16. No. 2. P. 256–298.)
2. Fu L., Cao Y., Wu J., Peng Q., Nie Q., Xie X. Ufold: fast and accurate RNA secondary structure prediction with deep learning. *Nucleic Acids Research.* 2022. V. 50. No. 3. P. e14. doi: [10.1093/nar/gkab1074](https://doi.org/10.1093/nar/gkab1074)
3. Chen C.C., Chan Y.M. REDfold: accurate RNA secondary structure prediction using residual encoder-decoder network. *BMC Bioinformatics.* 2023. V. 24. Article no. 122. P. 1–13. doi: [10.1186/s12859-023-05238-8](https://doi.org/10.1186/s12859-023-05238-8)
4. Lorenz R. (ed.) *RNA Folding: Methods and Protocols*. Springer, Humana Press, 2024. doi: [10.1007/978-1-0716-3519-3](https://doi.org/10.1007/978-1-0716-3519-3)
5. Mittal A., Turner D.H., Mathews D.H. NNDB: An expanded database of nearest neighbor parameters for predicting stability of nucleic acid secondary structures. *Journal of Molecular Biology.* 2024. V. 436. No. 168549. doi: [10.1016/j.jmb.2024.168549](https://doi.org/10.1016/j.jmb.2024.168549)
6. Ройтберг М.А., Симеоненков М.Н., Таболина О.Ю. Парето-оптимальные выравнивания символьных последовательностей. *Биофизика.* 1999. Т. 44. № 4. С. 581–594.
7. Malik A., Zhang L., Gautam M., Dai N., Li S., Zhang H., Mathews D.H., Huang L. LinearAlifold: linear-time consensus structure prediction for RNA alignments. *Journal of Molecular Biology.* 2024. V. 436. No. 17. Article no. 168694. doi: [10.1016/j.jmb.2024.168694](https://doi.org/10.1016/j.jmb.2024.168694)
8. Гринкевич Л.Н. Роль микроРНК в обучении и долговременной памяти. *Вавиловский журнал генетики и селекции.* 2020. Т. 24. № 8. С. 885–896. doi: [10.18699/VJ20.687](https://doi.org/10.18699/VJ20.687) (Grinkevich L.N. The role of microRNAs in learning and long-term memory. *Vavilov Journal of Genetics and Breeding.* 2020. V. 24. No. 8. P. 885–896.)
9. Vasiliev G.V., Ovchinnikov V.Y., Lisachev P.D., Bondar N.P., Grinkevich L.N. The expression of miRNAs involved in long-term memory formation in the CNS of the mollusk *Helix lucorum*. *International Journal of Molecular Sciences.* 2023. V. 24. No. 1. Article no. 301. P. 1–21. doi: [10.3390/ijms24010301](https://doi.org/10.3390/ijms24010301)
10. Ai J., Sun L.-H., Che H., Zhang R., Zhang T.-Z., Wu W.-C., Su X.-L., Chen X., Yang G., Li K., Wang N., Ban T., Bao Y.-N., Guo F., Niu H.-F., Zhu Y.-L., Zhu X.-Y., Zhao S.-G., Yang B.-F. MicroRNA-195 protects against dementia induced by chronic brain hypoperfusion via its anti-amyloidogenic effect in rats. *Journal of Neuroscience.* 2013. V. 33. No. 9. P. 3989–4001. doi: [10.1523/JNEUROSCI.1997-12.2013](https://doi.org/10.1523/JNEUROSCI.1997-12.2013)
11. Jawaid A., Woldemichael B.T., Kremer E.A., Laferriere F., Gaur N., Afroz T., Polymenidou M., Mansuy I.M. Memory decline and its reversal in aging and neurodegeneration involve miR-183/96/182 biogenesis. *Mol. Neurobiol.* 2019. V. 56. No. 5. P. 3451–3462. doi: [10.1007/s12035-018-1314-3](https://doi.org/10.1007/s12035-018-1314-3)
12. Griggs E.M., Young E.J., Rumbaugh G., Miller C.A. MicroRNA-182 regulates amygdala-dependent memory formation. Version 2. *J. Neurosci.* 2013. V. 33. No. 4. P. 1734–1740. doi: [10.1523/JNEUROSCI.2873-12.2013](https://doi.org/10.1523/JNEUROSCI.2873-12.2013)
13. Gao J., Wang W.Y., Mao Y.W., Gräff J., Guan J.S., Pan L., Mak G., Kim D., Su S.C., Tsai L.H. A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature.* 2010. V. 466. No. 7310. P. 1105–1109. doi: [10.1038/nature09271](https://doi.org/10.1038/nature09271)
14. Wang X., Liu D., Huang H.Z., Wang Z.H., Hou T.Y., Yang X., Pang P., Wei N., Zhou Y.F., Dupras M.J., Calon F., Wang Y.T., Man H.Y., Chen J.G., Wang J.Z., Hébert S.S., Lu Y., Zhu L.Q. A novel microRNA-124/PTPN1 signal pathway mediates synaptic and

- memory deficits in Alzheimer's disease. *Biol. Psychiatry*. 2018. V. 83. No. 5. P. 395–405. doi: [10.1016/j.biopsych.2017.07.023](https://doi.org/10.1016/j.biopsych.2017.07.023)
15. Baby N., Alagappan N., Dheen S.T., Sajikumar S. MicroRNA-134-5p inhibition rescues long-term plasticity and synaptic tagging/capture in an A β (1-42)-induced model of Alzheimer's disease. *Aging Cell*. 2020. V. 19. No. 1. Article no. e13046. P. 1–14. doi: [10.1111/acer.13046](https://doi.org/10.1111/acer.13046)
 16. Lee S.T., Chu K., Jung K.H., Kim J.H., Huh J.Y., Yoon H., Park D.K., Lim J.Y., Kim J.M., Jeon D., Ryu H., Lee S.K., Kim M., Roh J.K. miR-206 regulates brain-derived neurotrophic factor in Alzheimer disease model. *Ann. Neurol*. 2012. V. 72. P. 269–277. doi: [10.1002/ana.23588](https://doi.org/10.1002/ana.23588)
 17. Wang C.N., Wang Y.J., Wang H., Song L., Chen Y., Wang J.L., Ye Y., Jiang B. The anti-dementia effects of Donepezil involve miR-206-3p in the hippocampus and cortex. *Biol. Pharm. Bull*. 2017. V. 40. No. 4. P. 465–472. doi: [10.1248/bpb.b16-00898](https://doi.org/10.1248/bpb.b16-00898)
 18. Левенштейн В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов. *Доклады АН СССР*. 1965. Т. 163. № 4. С. 845–848. (Levenshtein V.I. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. 1966. V. 10. No. 8. P. 707–710.)
 19. Needleman S.B., Wunsch Ch.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*. 1970. V. 48. No. 3. P. 443–453. doi: [10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4)
 20. Tiskin A. Semi-local longest common subsequences in subquadratic time. *Journal of Discrete Algorithms*. 2008. V. 6. No. 4. P. 570–581. doi: [10.1016/j.jda.2008.07.001](https://doi.org/10.1016/j.jda.2008.07.001)
 21. Grabowski S. New tabulation and sparse dynamic programming based techniques for sequence similarity problems. *Discrete Applied Mathematics*. 2016. V. 212. P. 96–103. doi: [10.1016/j.dam.2015.10.040](https://doi.org/10.1016/j.dam.2015.10.040)
 22. Tiskin A. The Chvátal–Sankoff problem as a problem of symbolic dynamics. *Записки научных семинаров ПОМИ*. 2023. Т. 528. С. 214–237. (*Zapiski Nauchnykh Seminarov POMI*. 2023. V. 528. P. 214–237.)
 23. Chakraborty D., Das D., Goldenberg E., Koucký M., Saks M. Approximating edit distance within constant factor in truly sub-quadratic time. *Journal of the ACM*. 2020. V. 67. No. 6. Article no. 36. P. 1–22. doi: [10.1145/3422823](https://doi.org/10.1145/3422823)
 24. Kociumaka T., Radoszewski J., Starikovskaya T. Longest common substring with approximately k mismatches. *Algorithmica*. 2019. V. 81. No. 6. P. 2633–2652. doi: [10.1007/s00453-019-00548-x](https://doi.org/10.1007/s00453-019-00548-x)
 25. Тетуев Р.К., Пятков М.И., Панкратов А.Н. Параллельный алгоритм глобального выравнивания протяжённых аминокислотных и нуклеотидных последовательностей. *Матем. биология и биоинформ.* 2017. Т. 12. № 1. С. 137–150. doi: [10.17537/2017.12.137](https://doi.org/10.17537/2017.12.137) (Tetuev R.K., Pyatkov M.I., Pankratov A.N. Parallel algorithm for global alignment of long aminoacid and nucleotide sequences. *Math. Biol. Bioinf.* 2017. V. 12. No. 1. P. 137–150.)
 26. Mishin N., Berezun D., Tiskin A. Efficient parallel algorithms for string comparison. *ICPP '21: Proceedings of the 50th International Conference on Parallel Processing*. 2021. No. 50. P. 1–10. doi: [10.1145/3472456.3472489](https://doi.org/10.1145/3472456.3472489)
 27. Conte A., Grossi R., Punzi G., Uno T. Enumeration of maximal common subsequences between two strings. *Algorithmica*. 2022. V. 84. P. 757–783. doi: [10.1007/s00453-021-00898-5](https://doi.org/10.1007/s00453-021-00898-5)
 28. Efimov D.B. Enumeration of even and odd chord diagrams. *Journal of Applied and Industrial Mathematics*. 2024. V. 18. No. 2. P. 216–226. doi: [10.1134/S1990478924020042](https://doi.org/10.1134/S1990478924020042)
 29. Amir A., Charalampopoulos P., Pissis S.P., Radoszewski J. Dynamic and internal longest common substring. *Algorithmica*. 2020. V. 82. P. 3707–3743. doi:

[10.1007/s00453-020-00744-0](https://doi.org/10.1007/s00453-020-00744-0)

30. Бурдонов И., Максимов А. Двадцать функций подобия двух конечных последовательностей. *Программирование*. 2023. № 5. С. 3–18. doi: [10.31857/S0132347423050035](https://doi.org/10.31857/S0132347423050035) (Burdonov I., Maksimov A. Twenty similarity functions for two finite sequences. *Programming and Computer Software*. 2023. V. 49. P. 373–387. doi: [10.1134/S0361768823050031](https://doi.org/10.1134/S0361768823050031))
31. Alzamel M., Hampson C., Pliopoulos C.S., Lim Z., Pissis S., Vlachakis D., Watts S. Maximal degenerate palindromes with gaps and mismatches. *Theoretical Computer Science*. 2023. V. 978. Article no. 114182. P. 1–16. doi: [10.1016/j.tcs.2023.114182](https://doi.org/10.1016/j.tcs.2023.114182)
32. Зверков О.А., Селиверстов А.В., Шиловский Г.А. Выравнивание скрытого палиндрома. *Матем. биология и биоинформ.* 2024. Т. 19. № 2. С. 427–438. doi: [10.17537/2024.19.427](https://doi.org/10.17537/2024.19.427) (Zverkov O.A., Seliverstov A.V., Shilovsky G.A. Alignment of a hidden palindrome. *Math. Biol. Bioinf.* 2024. V. 19. No. 2. P. 427–438.)
33. Mieno T., Funakoshi M., Nakashima Y., Inenaga S., Bannai H., Takeda M. Computing maximal palindromes in non-standard matching models. *Information and Computation*. 2025. V. 304. Article no. 105283. doi: [10.1016/j.ic.2025.105283](https://doi.org/10.1016/j.ic.2025.105283)
34. Kumar S., Stecher G., Suleski M., Sanderford M., Sharma S., Tamura K. MEGA12: Molecular Evolutionary Genetics Analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution*. 2024. V. 41. P. 1–9. doi: [10.1093/molbev/msae263](https://doi.org/10.1093/molbev/msae263)
35. Letunic I., Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*. 2024. V. 52. No. W1. P. W78–W82. doi: [10.1093/nar/gkae268](https://doi.org/10.1093/nar/gkae268)
36. Kozomara A., Birgaoanu M., Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019. V. 47. P. D155–D162. doi: [10.1093/nar/gky1141](https://doi.org/10.1093/nar/gky1141)
37. de Magalhaes J.P., Costa J. A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*. 2009. V. 22. No. 8. P. 1770–1774. doi: [10.1111/j.1420-9101.2009.01783.x](https://doi.org/10.1111/j.1420-9101.2009.01783.x)
38. Shilovsky G.A., Putyatina T.S., Markov A.V. Evolution of longevity in tetrapods: Safety is more important than metabolism level. *Biochemistry (Moscow)*. 2024. V. 89. No. 2. P. 322–340. doi: [10.1134/S0006297924020111](https://doi.org/10.1134/S0006297924020111)

Рукопись поступила в редакцию 11.06.2025.

Переработанный вариант поступил 01.10.2025.

Дата опубликования 01.10.2025.

Alignment of RNA sequences with secondary structure: microRNA precursors

Georgii Khaziev¹, Oleg Zverkov¹, Alexandr Seliverstov¹, Vassily Lyubetsky¹

¹*Institute for Information Transmission Problems of the Russian Academy of Sciences
(Kharkevich Institute), Moscow, Russia*

Abstract. The aim of the work is to analyze changes during the evolution of those pre-microRNAs that are associated with long-term memory and learning in mammals. Large changes are observed in the pre-microRNAs mir-195 and mir-206 in mouse-like rodents (superfamily Muroidea) and in opossum. But the changes are small in all primates. Moreover, for mir-206, changes were shown to be related to the logarithm of maximum life span. We suggest a significant weakening of the effect of stabilizing selection on mir-206 in species with small maximum life span. In contrast, both mir-134 and mir-182 are more conservative. Their small changes are not related to maximum life span. Changes in the secondary structure of pre-microRNA are also considered. For this study, the authors developed an algorithm to truncate the input sequence to imperfect palindrome that is close to a perfect palindrome and is obtained by removing a prefix or suffix when possible. Such an imperfect palindrome roughly corresponds to an RNA hairpin. It is important that the palindrome is delimited during the optimization. Software implementation of the considered algorithm is created in Python. The programs are available at <http://lab6.iitp.ru/~pali> and <http://lab6.iitp.ru/~trimmers>. Searching for a perfect palindrome, close in the Levenshtein metric to the original sequence, often yields several optimal palindromes. In this way, for mir-195, a relationship was shown between the number of optimal palindromes and the logarithm of maximum life span. There was no such relationship for the mir-134, mir-182, and mir-206. Note that mir-195 and mir-206 differ from each other in the type of changes.

Key words: *microRNA, long-term memory, life span, alignment, palindrome, hairpin, edit distance, Mammalia, bioinformatics.*