

УДК 575.117.2:577.217.522:579.88

ЭКСПРЕССИЯ ГЕНОВ И ВТОРИЧНЫЕ СТРУКТУРЫ В мРНК В РАЗНЫХ ВИДАХ *MYCOPLASMA*

© 2013 г. В.С. Соколов¹, В.А. Лихошвай^{1, 2}, Ю.Г. Матушкин^{1, 2}

¹ Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: sokovlad1@bionet.nsc.ru;

² Новосибирский национальный исследовательский государственный университет,
Новосибирск, Россия

Поступила в редакцию 15 августа 2013 г. Принята к публикации 5 сентября 2013 г.

Определение эффективности экспрессии генов организма является актуальной и важной задачей современной биологии. Особый интерес представляют микроорганизмы, паразитирующие на человеке и домашних животных. В работе был проведен биоинформатический анализ геномов 62 штаммов бактерий, принадлежащих к роду *Mycoplasma*. Показано, что эффективность трансляции генов у этих организмов зависит от количества потенциальных вторичных структур в них и не зависит от кодонного состава. Обнаружены виды с пониженным содержанием локальных инвертированных повторов в генах. Анализ филогении показал возможную связь этой особенности со средой обитания данных организмов. Обнаружена не свойственная остальным микоплазмам высокая концентрация локальных инвертированных повторов в районе старт-кодона трансляции в генах *M. haemofelis*.

Ключевые слова: *Mycoplasma*, частоты кодонов, вторичные структуры, эффективность трансляции.

ВВЕДЕНИЕ

Определение эффективности экспрессии генов организма является актуальной и важной задачей современной биологии. Для ее решения были разработаны специальные экспериментальные методы, основными из которых на данный момент являются ДНК микрочипы и ПЦР в реальном времени. Однако данные методы требуют специального оборудования и реактивов. Поэтому разработка методов оценки предполагаемого уровня экспрессии генов на основе биоинформатического анализа их нуклеотидных последовательностей является актуальной и полезной.

Экспрессия гена – это процесс, в ходе которого наследственная информация из последовательности нуклеотидов ДНК преобразуется в функциональный продукт – РНК или белок. Процесс экспрессии генов состоит из нескольких стадий: транскрипция, трансляция и посттрансляционная модификация белков. Подробного изучения требуют все перечисленные

стадии. Данная работа посвящена изучению именно стадии трансляции. Большое количество времени и энергии в процессе трансляции затрачивается на стадию элонгации, движение комплекса рибосомных белков вдоль мРНК с одновременным синтезом закодированной в ней молекулы белка. Поэтому изучение особенностей нуклеотидных последовательностей генов, связанных со скоростью прохождения стадии элонгации, может помочь в определении итогового уровня их трансляции.

Во многих организмах была обнаружена неравномерность в использовании синонимичных кодонов при кодировании аминокислот в белках (Grantham *et al.*, 1980; Sharp, Li, 1987; Andersson, Kurland, 1990; Wada *et al.*, 1990; Stenico *et al.*, 1994). Установлено, что частоты кодонов коррелируют с концентрациями соответствующих им молекул тРНК (Bennetzen, Hall, 1982; Gouy, Gautier, 1982; Ikemura, 1985). Чем больше в мРНК наиболее часто используемых кодонов, тем быстрее проходит стадия элонгации трансляции для данного гена, так

как не происходит задержки рибосомы на кодонах, которым соответствуют тРНК с низкой концентрацией (Varenne *et al.*, 1984; Sorensen *et al.*, 1989).

Было показано, что кроме частот кодонов на скорость движения рибосомы по мРНК могут влиять вторичные структуры (шпильки), образующиеся перед ней (Jacks *et al.*, 1988; Dam *et al.*, 1990; Thanaraj, Argos, 1996; Lopinski *et al.*, 2000; Takyar *et al.*, 2005). Поэтому если в нуклеотидной последовательности гена встречается много локальных инвертированных повторов, которые потенциально могут образовать шпильки, то скорость трансляции такого гена может быть ниже, чем у других.

В опубликованных статьях есть данные о различных организмах, которые по-разному оптимизировали первичную структуру своих генов для повышения эффективности трансляции. Так, например, эффективность трансляции у *E. coli* и *S. cerevisiae* коррелирует с неравномерностью использования кодонов в их генах (Bennetzen, Hall, 1982; Gouy, Gautier, 1982; Li, Luo, 1996). В то же время для *H. pylori* такой корреляции не обнаружено, но обнаружена корреляция с количеством вторичных структур в мРНК (Vladimirov *et al.*, 2007). У *Mycoplasma gallisepticum* была обнаружена корреляция между количеством в клетке белка и количеством вторичных структур в кодирующем его гене (данных нет в широкой печати, но присутствуют в отчете по гранту РФФИ № 06-04-49556).

Основанием к проведению анализа организмов, принадлежащих именно к роду *Mycoplasma*, послужил размер их геномов. Поскольку большинство данных организмов являются паразитами, их геном значительно редуцирован, что позволяет упростить изучаемую систему и, возможно, обнаружить более явно присущие ей закономерности (табл. 1). А в свете того, что на данный момент процесс трансляции и механизмы его регуляции у микоплазм мало изучены, в том числе экспериментально, данная работа является весьма актуальной.

МАТЕРИАЛЫ И МЕТОДЫ

Анализируемые последовательности белок-кодирующих генов с фланкирующими районами длиной 600 нуклеотидов экстрагировались из файлов в формате **gbk**, содержащих полные геномы исследуемых организмов. Данные файлы были получены из базы данных GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Из итогового списка удалялись гены, отмеченные как псевдогены, а также гены длиной менее 30 кодонов. Дальнейшие расчеты проводились на оставшихся в списке генах.

Суть работы алгоритма заключается в оценке среднего времени, затрачиваемого рибосомой на стадию элонгации трансляции (Лихошвай, Матушкин, 2000). Для этого для каждого гена рассчитывается специальный индекс эффективности элонгации (EEI – elongation efficiency

Таблица 1

Организмы, на которых паразитируют различные виды *Mycoplasma*

Хозяин	Паразит	Поражаемые клетки
Человек	<i>M. genitalium</i>	Реснитчатый эпителий дыхательных и половых путей
	<i>M. pneumoniae</i>	Реснитчатый эпителий трахей
Кошки	<i>C.M. haemominutum</i>	Зрелые эритроциты
	<i>M. haemofelis</i>	
	<i>C.M. turicensis</i>	
Собаки	<i>C.M. haematoparvum</i>	
	<i>M. haemocanis</i>	
Овцы и козы	<i>M. ovis</i>	
Свиньи	<i>M. suis</i>	
Коровы	<i>M. wenyonii</i>	
Ламы и альпаки	<i>C.M. haemolamae</i>	

index), пропорциональный скорости элонгации (Likhoshvai, Matushkin, 2002). Чем выше значение EEI, тем быстрее рибосома движется по мРНК и тем быстрее синтезируется белок. Кодонный состав гена и количество (и «прочность») потенциальных вторичных структур в мРНК – два основных фактора, учитываемых при расчете индекса эффективности элонгации.

Существуют пять типов индекса эффективности элонгации:

1. EEI1 – зависит только от кодонного состава гена;
2. EEI2 – зависит от количества потенциальных вторичных структур в мРНК без учета их энергии;
3. EEI3 – зависит от количества и энергии (стабильности) потенциальных вторичных структур в мРНК;
4. EEI4 – зависит от кодонного состава и от количества потенциальных вторичных структур;
5. EEI5 – зависит от кодонного состава и от количества и энергии потенциальных вторичных структур (Vladimirov *et al.*, 2007).

Для определения, какой из 5 типов индекса адекватно определяет эффективность элонгации в конкретном организме, все его гены сортируются по EEI и рассчитываются параметры *M* в интервале $[-100; 100]$ и *R* в интервале $[0; 100]$. *M* имеет смысл среднего положения генов рибосомных белков в отсортированном списке, а *R* – стандартного отклонения от среднего значения. Если рибосомные гены оказываются близко к тому краю списка, где располагаются гены с наибольшей скоростью элонгации, тогда параметр *M* у такого организма и данного типа индекса будет близок к значению 100. Таким образом, для конкретного организма адекватен тот тип индекса, для которого значение параметра *M* наибольшее. Если у двух типов индекса значения параметра *M* совпадают, тогда выбирается индекс с меньшим значением параметра *R*, что говорит о более компактном расположении рибосомных генов.

Для более детального анализа последовательностей генов на наличие вторичных структур для каждого гена организма рассчитывались индексы локальной комплементарности (local complementary index – LCI, один из составляющих EEI) (Likhoshvai, Matushkin, 2002).

Данный индекс имеет смысл среднего числа локальных инвертированных повторов на один ген. Таким образом, чем больше таких повторов встречается в последовательности (т. е. чем больше потенциальных вторичных структур может в ней образоваться), тем выше значение индекса LCI. Далее последовательности всех генов одного организма выравнивались по старту (или стоп-кодону) трансляции и по ним рассчитывались средние значения в области $[-500; +500]$ относительно старта (стоп-кодона) трансляции. По полученным данным строились профили LCI для всех организмов.

Для проведения исследований была написана специальная программа, реализующая описанный выше алгоритм. Программа в ближайшее время будет доступна на сайте ИЦиГ СО РАН по адресу: <http://www.bionet.nsc.ru/razrabotki/prikladnyie-razrabotki/programmyi-dlya-evm.html>.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При помощи программы были проанализированы геномы 62 штаммов *Mycoplasma*. Часть результатов представлена в табл. 2 (полная версия табл. 2 представлена в Приложении).

Суммарное распределение штаммов по типам индекса представлено на рис. 1.

Как видно из рис. 1, в большинстве штаммов адекватно работает второй тип индекса (EEI2) – эффективность элонгации зависит только от количества вторичных структур в мРНК и не зависит от кодонного состава генов. Почти у всех видов рибосомные гены хорошо определяются как высокоэкспрессируемые, значения параметра смещения *M* для них высоки. Но есть виды со значительно более низкими значениями параметра *M*: *C.M. haemominutum*, *M. suis*, *M. pneumoniae* и особенно *M. haemocanis* и *M. haemofelis*. У данных видов организмов практически отсутствует смещение рибосомных генов в сторону высокоэкспрессирующихся.

Объяснить этот факт можно с помощью двух альтернативных предположений.

1. Рибосомные гены у данных видов не являются высокоэкспрессируемыми.

2. Все остальные гены у данных видов характеризуются повышенным уровнем экспрессии, что нивелирует уровень экспрессии рибосомных генов.

Таблица 2

Типы индексов для штаммов *Mycoplasma*

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
<i>Mycoplasma wenyonii</i> str. Massachusetts	1	70	–9	–21	69	25
<i>Candidatus Mycoplasma haemolamae</i> str. Purdue	1	30	5	20	25	29
<i>Mycoplasma capricolum</i> subsp. <i>capricolum</i> ATCC 27343	2	–60	78	–26	40	–52
<i>Mycoplasma fermentans</i> JER	2	–40	77	–42	71	–62
<i>Mycoplasma mycoides</i> subsp. <i>capri</i> LC str. 95010	2	–53	76	–27	25	–51
<i>Mycoplasma leachii</i> PG50	2	–56	75	–31	41	–56
<i>Mycoplasma cynos</i> C142	2	–39	74	–29	47	–47
<i>Mycoplasma synoviae</i> 53	2	–37	71	–8	50	–32
<i>Mycoplasma penetrans</i> HF 2	2	–1	71	–32	60	–37
<i>Mycoplasma hominis</i> ATCC 23114	2	–7	69	–28	67	–38
<i>Mycoplasma putrefaciens</i> KS1	2	–20	68	–21	59	–41
<i>Mycoplasma pulmonis</i> UAB CTIP	2	–13	68	5	48	–16
<i>Mycoplasma hyopneumoniae</i> 168	2	–67	68	–28	–33	–60
<i>Mycoplasma genitalium</i> G37	2	–66	59	–39	–40	–64
<i>Mycoplasma gallisepticum</i> CA06 2006 052-5-2p	2	23	49	–7	47	–1
<i>Mycoplasma suis</i> KI3806	2	18	27	0	23	3
<i>Mycoplasma suis</i> str. Illinois	2	6	26	0	17	2
<i>Mycoplasma pneumoniae</i> M129-B7	2	–10	25	–27	23	–30
<i>Mycoplasma pneumoniae</i> FH	2	–13	24	–32	21	–31
<i>Candidatus Mycoplasma haemominutum</i> Birmingham 1	2	–8	24	2	8	–9
<i>Mycoplasma haemocanis</i> str. Illinois	2	–15	–6 (63)	–6 (70)	–19	–24
<i>Mycoplasma haemofelis</i> Ohio2	3	–23	–21	7	–36	–3
<i>Mycoplasma haemofelis</i> str. Langford 1	3	–25	–16	3	–35	–12
<i>Mycoplasma agalactiae</i>	4	26	64	–24	66	–26
<i>Mycoplasma bovis</i> PG45	4	13	63	–40	66	–35
<i>Mycoplasma pneumoniae</i> 309	4	–9	23	–26	25	–29
<i>Mycoplasma pneumoniae</i> M129	4	–10	22	–25	25	–29

Примечание. Серым цветом в строке выделен тот тип индекса EEI, который работает в соответствующем штамме (наибольшее значение параметра М в строке). Темно-серым выделены штаммы с наименьшим смещением генов рибосомных белков в сторону высокоэкспрессирующихся генов ($M \leq 30$). Для штаммов с одинаковыми значениями параметра М для разных типов индекса в скобках дополнительно приведены значения параметра R.

Для выбора одного из двух объяснений были рассчитаны средние значения количества вторичных структур на один ген для исследуемых 62 представителей рода *Mycoplasma*. Полученные результаты были отсортированы по увеличению среднего количества вторичных структур на один нерибосомный ген и отображены на графике, представленном на рис. 2.

Из графика видно, что рибосомные гены у всех штаммов мало отличаются друг от друга по

количеству вторичных структур, что хорошо согласуется с предположением о высокой консервативности нуклеотидных последовательностей данных генов. С другой стороны, нерибосомные гены у разных штаммов могут значительно различаться по среднему количеству вторичных структур на один ген. В частности, интересующие нас «особые» штаммы с низкими значениями параметра М попали в самую крайнюю группу с наименьшими значениями количества

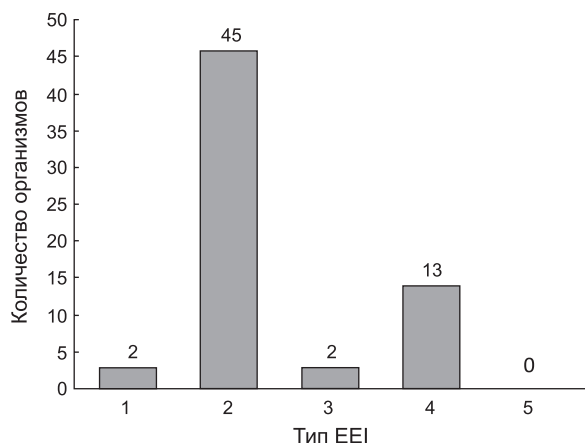


Рис. 1. Распределение 62 штаммов микоплазм по типам индекса EEI.

вторичных структур на один ген (ограничена пунктиром на рис. 2). Благодаря этому разница между рибосомными и нерибосомными генами у данных штаммов значительно меньше, чем у других, что и объясняет их особенность.

Установленный факт неравномерности по среднему числу локальных инвертированных повторов в генах разных штаммов *Mycoplasma* является новым и ранее неизученным, поэтому требует более подробного исследования.

У четырех штаммов (*M. haemofelis* Ohio2, *M. haemofelis* Langford1, *M. haemocanis* Illinois, *C.M. haemolamae* Purdue) с самым низким зна-

чением среднего количества локальных инвертированных повторов на один нерибосомный ген (крайние слева на рис. 2) были рассмотрены 100 генов с самым низким количеством локальных инвертированных повторов для определения их функций. Большинство из данных 100 генов отмечены «hypothetical protein» и их функции не известны. Но среди генов с известными функциями встречаются следующие: субъединицы рибосом, субъединицы ДНК полимераз, относящиеся к синтезу АТФ и синтезу тРНК, переносчики АТФ и др. Таким образом, самое низкое содержание локальных инвертированных повторов наблюдается в генах «домашнего хозяйства».

Чтобы понять, почему же именно у данных штаммов (ограничены пунктиром на рис. 2) наблюдается такое низкое значение среднего количества локальных инвертированных повторов на один ген, мы рассмотрели исследуемые штаммы *Mycoplasma* с точки зрения филогении.

Филогенетическое дерево *Mycoplasma*, построенное на основе анализа последовательностей 16S рРНК, было взято из статьи Peters с соавт. (2008) и представлено на рис. 3. На дереве серым цветом отмечены виды со значением параметра $M > 30$, а черным – с $M \leq 30$. Видно, что почти все виды, отмеченные черным (кроме *M. pneumoniae*), попадают в группу гемоплазм. Гемоплазмы – это гемотрофные

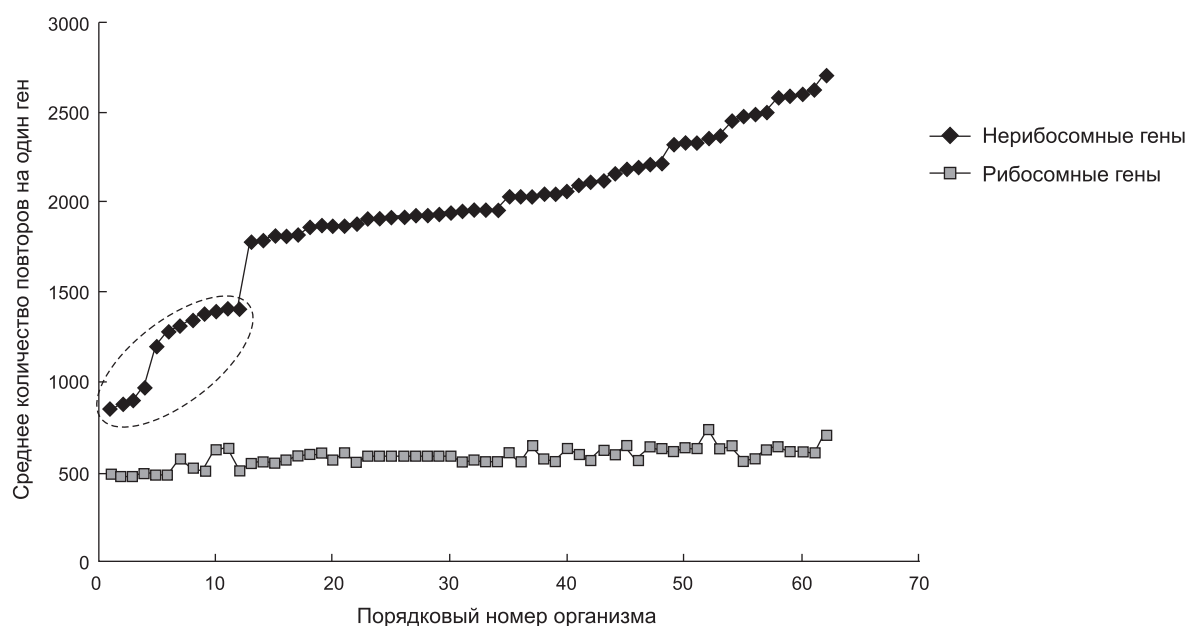


Рис. 2. Среднее число вторичных структур на один ген для каждого штамма.

Пунктиром выделены штаммы с параметром $M \leq 30$.

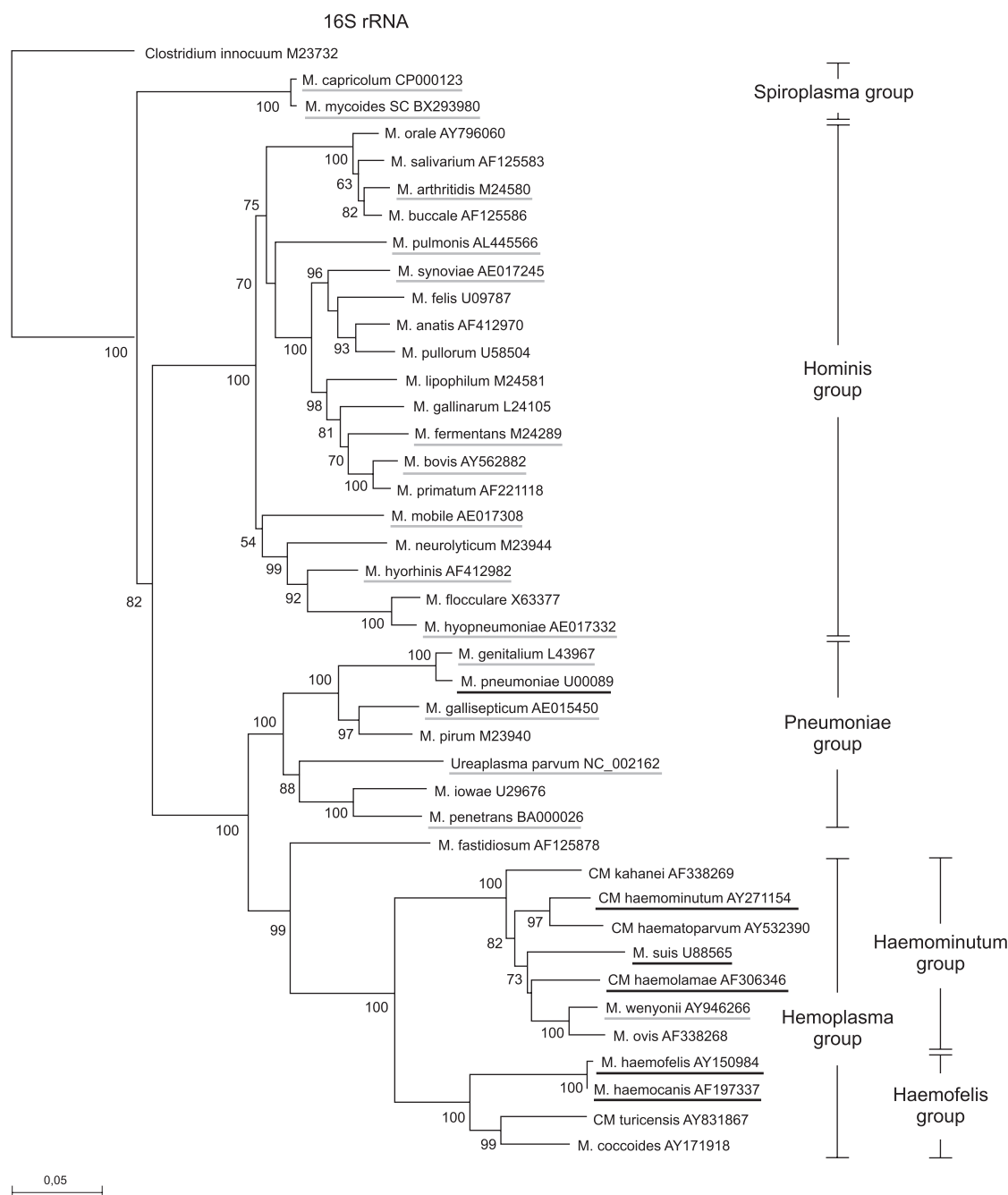


Рис. 3. Филогенетическое дерево *Mycoplasma*, построенное на основе анализа последовательностей 16S рРНК (из: (Peters *et al.*, 2008)).

Серым цветом отмечены штаммы с параметром $M > 30$, а черным – с $M \leq 30$.

организмы, их жизнь связана с красными кровяными тельцами (эритроцитами). Считается, что эти микоплазмы паразитируют на поверхности эритроцитов и даже могут проникать внутрь них.

На основании всего вышеизложенного мы предполагаем, что в особых условиях обита-

ния на поверхности или внутри эритроцитов данные штаммы эволюционировали в сторону уменьшения количества вторичных структур в своих генах. Возможно, таким образом они уменьшили энергетические затраты на процесс трансляции, чтобы повысить эффективность экспрессии.

Вместе с другими микоплазмами в группу гемоплазм попала *M. wenyonii*, паразит крупного рогатого скота. У данного организма тоже отмечается низкое значение среднего количества вторичных структур на один ген (также попадает в выделенную пунктиром область на рис. 2). Однако в нем работает первый тип индекса ЕЕІ и значение параметра $M = 70$, т. е. рибосомные гены располагаются в области высокоэкспрессирующихся генов. Возможно, данный организм эволюционировал в сторону оптимизации кодонного состава, а пониженное содержание вторичных структур в генах досталось ему от предка.

Как было отмечено выше, *M. pneumoniae* не попала в группу гемоплазм, но у данного организма также наблюдается низкое значение параметра M и сниженное среднее количество вторичных структур на один ген. *M. pneumoniae* является паразитом верхних и нижних дыхательных путей человека. Она может взаимодействовать с поверхностью клеток дыхательного эпителия (реснитчатый эпителий трахей и клетки, выстилающие подслизистые железы (Collier, Clyde, 1971; Powell *et al.*, 1976)) и, возможно, проникать внутрь них (Waites *et al.*,

2008). Возможно, условия обитания данного организма чем-то схожи с условиями обитания на поверхности эритроцитов, что и способствовало эволюции данного организма в сторону уменьшения количества вторичных структур в его генах. Для получения более точного ответа на данный вопрос необходимы дополнительные исследования.

Для более подробного изучения распределения вторичных структур в генах организмов рода *Mycoplasma* были рассчитаны специальные индексы локальной комплементарности для каждого нуклеотида в гене и на его флангах ($LCI(i, j)$, где i – номер гена, j – номер нуклеотида в гене). После расчетов все гены одного организма выравнивались по старт- (стоп-) кодону трансляции и рассчитывались средние значения индексов $LCI(i, j)$. Полученные результаты показаны на рис. 4–7.

У большинства представителей рода *Mycoplasma* средний профиль $LCI(i, j)$ для 5'- и 3'-районов гена имеет вид, как у *M. fermentans* JER (рис. 4, 5). В районе старт-кодона трансляции наблюдается характерный спад профиля, а в районе стоп-кодона – пик. Графики схожи с LCI профилями для других организмов

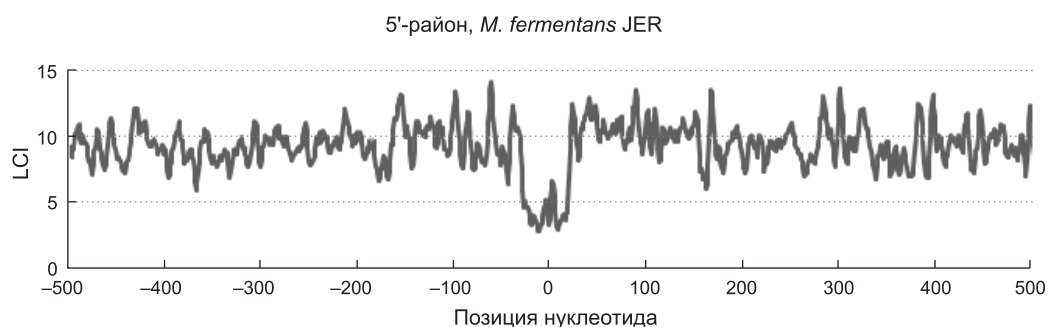


Рис. 4. Средний профиль $LCI(i, j)$ по всем генам *M. fermentans* JER (0 – старт-кодон).

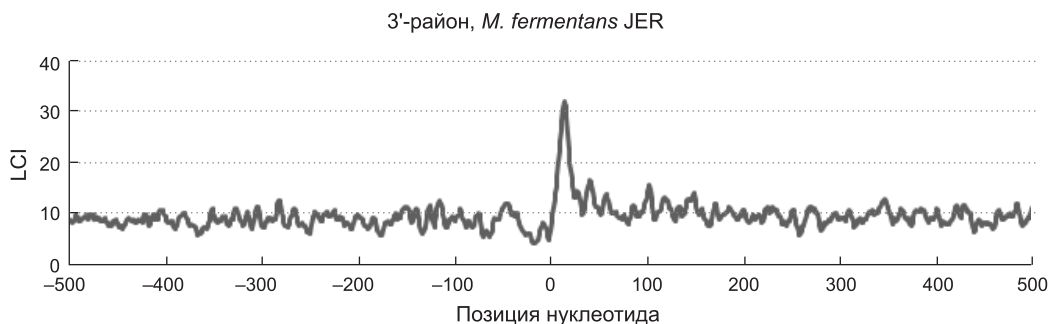


Рис. 5. Средний профиль $LCI(i, j)$ по всем генам *M. fermentans* JER (0 – стоп-кодон).

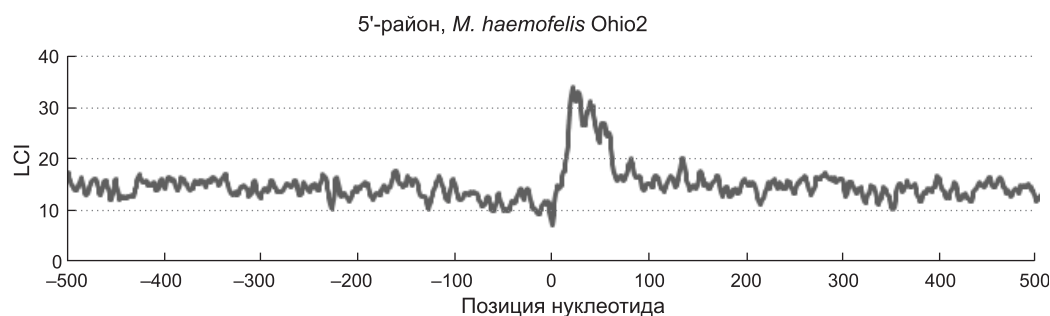


Рис. 6. Средний профиль $LCI(i, j)$ по всем генам *M. haemofelis* Ohio2 (0 – стоп-кодон).

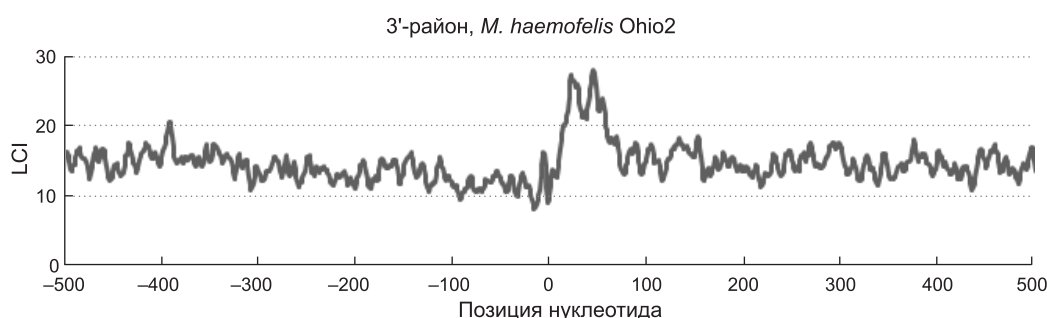


Рис. 7. Средний профиль $LCI(i, j)$ по всем генам *M. haemofelis* Ohio2 (0 – стоп-кодон).

(*E. coli*, *S. cerevisiae*, *H. sapiens* и др.) (Matushkin *et al.*, 2004).

Спад профиля в 5'-районе, т. е. пониженная вероятность образования шпилек, вероятнее всего, способствует сборке рибосомного комплекса и началу трансляции. Наличие пика в 3'-районе говорит о повышенной вероятности образования шпилек в данной области, которые могут отвечать за терминацию трансляции или, возможно, транскрипции.

Интересной особенностью обладают профили для 5'-районов у *M. haemofelis*. У данных организмов вместо спада профиля, наоборот, наблюдается его повышение (рис. 6), что говорит о повышенной вероятности образования шпилек в данной области. Пока сложно сказать что-либо о причинах данного явления и почему именно *M. haemofelis* обладает данной особенностью. Но, как уже было сказано выше, данный организм обитает в особых условиях (поверхность или внутриклеточное пространство эритроцитов), которые, возможно, способствовали эволюции первичной структуры его генов в сторону уменьшения количества вторичных структур. Если предположить, что скорость прохождения элонгации трансляции у данного штамма возросла, возможно, лимитирующей стадией стала

именно стадия начала трансляции. Поэтому мы предполагаем, что шпильки в 5'-районе могут отвечать именно за регуляцию начала трансляции у данного штамма. Возможно, у *M. haemofelis* есть особенный механизм регуляции начала трансляции, отличный от механизмов в других *Mycoplasma*. Для выяснения причин наличия пика в 5'-районе профиля $LCI(i, j)$ у *M. haemofelis* необходимы дополнительные исследования, в частности более детальное изучение распределения и строения вторичных структур в данном районе генов этого организма.

ЗАКЛЮЧЕНИЕ

Таким образом, проведен биоинформатический анализ геномов 62 штаммов, принадлежащих к роду *Mycoplasma*. Установлено, что эволюционный отбор по эффективности трансляции генов у данных организмов шел на минимизацию количества потенциальных вторичных структур в них, а не на оптимизацию кодонного состава. Получена новая информация о количестве локальных инвертированных повторов в генах разных штаммов *Mycoplasma*. Была обнаружена группа мико-

плазм (*C.M. haemolamae*, *C.M. haemominutum*, *M. haemocanis*, *M. haemofelis*, *M. pneumoniae*, *M. suis*) с пониженным содержанием в генах локальных инвертированных повторов. Было показано, что почти все они (кроме *M. pneumoniae*) относятся к одной филогенетической группе гемоплазм и характеризуются обитанием на поверхности или внутри эритроцитов. Также при построении профилей распределения локальных инвертированных повторов в районах старт- и стоп-кодонов трансляции у *M. haemofelis* обнаружен нехарактерный для остальных микоплазм пик в районе старт-кодона. Пик свидетельствует о повышенной вероятности образования шпилек в данном районе гена. По нашим предположениям, это может быть связано с каким-то альтернативным механизмом регуляции трансляции, отличным от механизмов у других видов микоплазм.

БЛАГОДАРНОСТИ

Работа выполнена при частичной поддержке программ Президиума РАН «Молекулярная и клеточная биология» (проект 6.6) и «Происхождение биосферы и эволюция гео-биологических систем» (№ 15), гранта НШ-5278.2012.4 и гранта РФФИ № 13-04-0062013.

ЛИТЕРАТУРА

- Лихошвай В.А., Матушкин Ю.Г. Предсказание эффективности экспрессии генов по их нуклеотидному составу // Молекуляр. биология. 2000. Т. 34. № 3. С. 406–412.
- Andersson S.G.E., Kurland C.G. Codon preferences in free-living microorganisms // Microbiol. Rev. 1990. V. 54. P. 198–210.
- Bennetzen J.L., Hall B.D. Codon selection in Yeast // J. Biol. Chem. 1982. V. 257. P. 3026–3031.
- Dam E.B., Pleij C.W., Bosch L. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs // Virus Genes. 1990. V. 4. P. 121–136.
- Collier A.M., Clyde W.A. Jr. Relationships between *Mycoplasma pneumoniae* and human respiratory epithelium // Infect. Immun. 1971. V. 3. No. 5. P. 694–701.
- Gouy M., Gautier C. Codon usage in bacteria: correlation with gene expressivity // Nucl. Acids Res. 1982. V. 10. P. 7055–7070.
- Grantham R., Gautier C., Gouy M. *et al.* Codon catalog usage and the genome hypothesis // Nucl. Acids Res. 1980. V. 8. P. 49–62.
- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms // Mol. Biol. Evol. 1985. V. 2. P. 13–34.
- Jacks T., Madhani H.D., Masiarz F.R., Varmus H.E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region // Cell. 1988. V. 55. P. 447–458.
- Li H., Luo L. The relation between codon usage, base correlation and gene expression level in *Escherichia coli* and Yeast // J. Theor. Biol. 1996. V. 181. Iss. 2. P. 111–124.
- Likhoshvai V.A., Matushkin Yu.G. Differentiation of single-cell organisms according to elongation stages crucial for gene expression efficacy // FEBS Lett. 2002. V. 516. P. 87–92.
- Lopinski J.D., Dinman J.D., Bruenn J.A. Kinetics of ribosomal pausing during programmed–1 translational frameshifting // Mol. Cell. Biol. 2000. V. 20. P. 1095–1103.
- Matushkin Yu.G., Likhoshvai V.A., Kochetov A.V. Local secondary structure may be a critical characteristic influencing translation of unicellular organisms mRNA // Bioinformatics of Genome Regulation and Structure. Boston a.o.: Kluwer Acad. Publ., 2004. P. 103–114.
- Peters I.R., Helps C.R., McAuliffe L. *et al.* RNase P RNA gene (*rnpB*) phylogeny of Hemoplasmas and other *Mycoplasma* species // J. Clin. Microbiol. 2008. V. 46. No. 5. P. 1873–1877.
- Powell D.A., Hu P.C., Wilson M. *et al.* Attachment of *Mycoplasma pneumoniae* to respiratory epithelium // Infect. Immun. 1976. V. 13 No. 3. P. 959–966.
- Sharp P.M., Li W.H. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications // Nucl. Acids Res. 1987. V. 15. P. 1281–1295.
- Sorensen M.A., Kurland C.G., Pedersen S. Codon usage determines translation rate in *Escherichia coli* // J. Mol. Biol. 1989. V. 207. P. 365–377.
- Stenico M., Lloyd A.T., Sharp P.M. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases // Nucl. Acids Res. 1994. V. 22. P. 2437–2446.
- Takyar S., Hickerson R.P., Noller H.F. mRNA helicase activity of the ribosome // Cell. 2005. V. 120. P. 49–58.
- Thanaraj T.A., Argos P. Ribosome-mediated translational pause and protein domain organization // Protein Sci. 1996. V. 5. P. 1594–1612.
- Varenne S., Buc J., Lloubes R., Lazdunski C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains // J. Mol. Biol. 1984. V. 180. P. 549–576.
- Vladimirov N.V., Likhoshvai V.A., Matushkin Yu.G. Correlation of codon biases and potential secondary structures with mRNA translation efficiency in unicellular organisms // Mol. Biol. 2007. V. 41. No. 5. P. 926–933.
- Wada K.S., Aota R., Tsuchiya F. *et al.* Codon usage tabulated from GenBank genetic sequence data // Nucl. Acids Res. 1990. V. 18. (Suppl.). P. 2367–2411.
- Waites K.B., Balish M.F., Atkinson T.P. New insights into the pathogenesis and detection of *Mycoplasma pneumoniae* infections // Future Microbiol. 2008. V. 3. No. 6. P. 635–648.

Приложение
Таблица

Типы индексов для 62 штаммов *Mycoplasma*

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
<i>Mycoplasma wenyonii</i> str. Massachusetts	1	70	-9	-21	69	25
<i>Candidatus Mycoplasma haemolamae</i> str. Purdue	1	30	5	20	25	29
<i>Mycoplasma fermentans</i> M64	2	-57	79	-50	67	-74
<i>Mycoplasma hyorhinis</i> MCLD	2	-20	79	-23	45	-42
<i>Mycoplasma fermentans</i> PG18	2	-53	78	-50	75	-66
<i>Mycoplasma hyorhinis</i> HUB 1	2	-9	78	-16	46	-29
<i>Mycoplasma capricolum</i> subsp. capricolum ATCC 27343	2	-60	78	-26	40	-52
<i>Mycoplasma fermentans</i> JER	2	-40	77	-42	71	-62
<i>Mycoplasma hyorhinis</i> SK76	2	-15	77	-24	48	-40
<i>Mycoplasma hyopneumoniae</i> 232	2	-61	77	-24	-17	-57
<i>Mycoplasma crocodyli</i> MP145	2	-44	76	-44	56	-57
<i>Mycoplasma mycoides</i> subsp. mycoides SC str. Gladysdale	2	-37	76	-24	61	-42
<i>Mycoplasma mycoides</i> subsp. mycoides SC str. PG1	2	-45	76	-21	65	-43
<i>Mycoplasma leachii</i> 990146	2	-59	76	-29	70	-49
<i>Mycoplasma mycoides</i> subsp. capri LC str. 95010	2	-53	76	-27	25	-51
<i>Mycoplasma leachii</i> PG50	2	-56	75	-31	41	-56
<i>Mycoplasma cynos</i> C142	2	-39	74	-29	47	-47
<i>Mycoplasma synoviae</i> 53	2	-37	71	-8	50	-32
<i>Mycoplasma penetrans</i> HF 2	2	-1	71	-32	60	-37
<i>Mycoplasma hyopneumoniae</i> 7448	2	-69	71	-29	-21	-63
<i>Mycoplasma putrefaciens</i> Mput9231	2	-17	70 (36)	-24	70 (49)	-37
<i>Mycoplasma hyorhinis</i> GDL 1	2	-20	70	-23	40	-36
<i>Mycoplasma hyopneumoniae</i> J	2	-70	70	-30	-23	-62
<i>Mycoplasma mobile</i> 163K	2	-48	70	-9	53	-30
<i>Mycoplasma hominis</i> ATCC 23114	2	-7	69	-28	67	-38

Продолжение таблицы

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
<i>Mycoplasma hyopneumoniae</i> 168 L	2	-64	69	-26	-27	-58
<i>Mycoplasma putrefaciens</i> KS1	2	-20	68	-21	59	-41
<i>Mycoplasma pulmonis</i> UAB CTIP	2	-13	68	5	48	-16
<i>Mycoplasma hyopneumoniae</i> 168	2	-67	68	-28	-33	-60
<i>Mycoplasma conjunctivae</i> HRC/581	2	-8	67	-18	50	-29
<i>Mycoplasma arthritis</i> 158L3-1	2	9	62	-23	52	-21
<i>Mycoplasma genitalium</i> G37	2	-66	59	-39	-40	-64
<i>Mycoplasma genitalium</i> M6282	2	-52	53	-24	-29	-48
<i>Mycoplasma gallisepticum</i> str. R (high)	2	18	53	2	46	-3
<i>Mycoplasma gallisepticum</i> str. R (low)	2	19	53	2	46	-2
<i>Mycoplasma genitalium</i> M2288	2	-53	52	-30	-30	-51
<i>Mycoplasma genitalium</i> M6320	2	-50	50	-26	-29	-47
<i>Mycoplasma genitalium</i> M2321	2	-51	50	-26	-30	-46
<i>Mycoplasma gallisepticum</i> NY01 2001.047-5-1P	2	24	49	-7	49	-2
<i>Mycoplasma gallisepticum</i> ca06 2006 052-5-2p	2	23	49	-7	47	-1
<i>Mycoplasma gallisepticum</i> NC96 1596-4-2P	2	24	48	-6	47	-2
<i>Mycoplasma suis</i> KI3806	2	18	27	0	23	3
<i>Mycoplasma suis</i> str. Illinois	2	6	26	0	17	2
<i>Mycoplasma pneumoniae</i> M129-B7	2	-10	25	-27	23	-30
<i>Mycoplasma pneumoniae</i> FH	2	-13	24	-32	21	-31
<i>Candidatus Mycoplasma haemominutum</i> Birmingham 1	2	-8	24	2	8	-9
<i>Mycoplasma haemocanis</i> str. Illinois	2	-15	-6 (63)	-6 (70)	-19	-24
<i>Mycoplasma haemofelis</i> Ohio2	3	-23	-21	7	-36	-3
<i>Mycoplasma haemofelis</i> str. Langford 1	3	-25	-16	3	-35	-12
<i>Mycoplasma bovis</i> Hubei-1	4	25	66	-44	80	-27
<i>Mycoplasma agalactiae</i> PG2	4	22	64	-27	73	-27

Окончание таблицы

Организм	Тип EEI	M1 (R1)	M2 (R2)	M3 (R3)	M4 (R4)	M5 (R5)
<i>Mycoplasma bovis</i> HB0801	4	15	67	-38	69	-34
<i>Mycoplasma agalactiae</i>	4	26	64	-24	66	-26
<i>Mycoplasma bovis</i> PG45	4	13	63	-40	66	-35
<i>Mycoplasma gallisepticum</i> str. F	4	31	54	1	62	3
<i>Mycoplasma gallisepticum</i> NC08 2008.031-4-3P	4	26	49	-7	53	-1
<i>Mycoplasma gallisepticum</i> WI01 2001.043-13-2P	4	26	50	-7	52	-1
<i>Mycoplasma gallisepticum</i> NC06 2006.080-5-2P	4	26	49	-7	51	-2
<i>Mycoplasma gallisepticum</i> NC95 13295-2-2P	4	26	49	-7	51	-2
<i>Mycoplasma gallisepticum</i> VA94 7994-1-7P	4	26	48	-7	51	-2
<i>Mycoplasma pneumoniae</i> 309	4	-9	23	-26	25	-29
<i>Mycoplasma pneumoniae</i> M129	4	-10	22	-25	25	-29

Примечание. Серым цветом в строке выделен тип индекса EEI, который работает в соответствующем штамме (наибольшее значение параметра М в строке). Темно-серым выделены штаммы с наименьшим смещением генов рибосомных белков в сторону высокоэкспрессирующихся генов ($M \leq 30$). Для штаммов с одинаковыми значениями параметра М для разных типов индекса в скобках дополнительно приведены значения параметра R.

GENE EXPRESSION AND mRNA SECONDARY STRUCTURES IN DIFFERENT *MYCOPLASMA* SPECIES

V.S. Sokolov¹, V.A. Likhoshvai^{1,2}, Yu.G. Matushkun^{1,2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: sokovlad1@bionet.nsc.ru;

² Novosibirsk National Research State University, Novosibirsk, Russia

Summary

Evaluation of gene expression efficiency in different organisms is a vital task of modern biology. Microorganisms that feed on humans and pets are particularly interesting. In this work, bioinformatical analysis of 62 *Mycoplasma* strains is performed. It has been shown that translation efficiency in these organisms depends on the number of potential secondary structures in genes and does not depend on codon compositions. Several species with low concentrations of local inverted repeats in genes have been found. Phylogenetic analysis shows that this feature may be associated with their environment. High concentrations of local inverted repeats, not typical of other *Mycoplasma* species, have been found in the translation start regions of *M. haemofelis* genes.

Key words: *Mycoplasma*, codon frequencies, secondary structures, translation efficiency.