

Перевод оригинальной статьи

2015 Skoblikov N.E., Zimin A.A. *Mathematical Biology and Bioinformatics*. 2015. V. 10. № 1. P. 116–130.

doi: [10.17537/2015.10.116](https://doi.org/10.17537/2015.10.116)

===== **БИОИНФОРМАТИКА** =====

УДК: 577.217.347

Поиск реликтовых рибонуклеотидных и аминокислотных последовательностей, игравших ключевую роль в формировании рибосомы и современного разнообразия белков

Скобликов Н.Э.¹, Зимин А.А.²

¹*Северо-Кавказский научно-исследовательский институт животноводства,
Краснодар, 350055, Россия*

²*Институт биохимии и физиологии микроорганизмов им. Г.К. Скрыбина, Российская
академия наук, Пушкино, Московская область, 142290, Россия*

Аннотация. Представлены результаты анализа базы данных белковых последовательностей прокариотических микроорганизмов, выявившего наличие элементов консервативной пептидной последовательности длиной 11 аминокислотных остатков в 20 локусах 16 функционально и филогенетически различных консервативных белков у представителей различных таксонов. Этот аминокислотный мотив IKAVRELGLER, возможно, является одним из последних общих предшественников матрично-синтезируемых пептидов (Last Universal Peptide Ancestors, LUPAs). Источником генетической информации такой последовательности, по всей вероятности, служит фрагмент рибосомной РНК (участок А-сайта, включающий ветви H92, H90 и H93 пептидил-трансферазного центра), транслированный с одной из возможных рамок считывания. Такой фрагмент рРНК, исполнявший функцию матрицы для синтеза LUPA, назван нами последним общим РНК-предшественником (Last Universal RiboNucleic Ancestor, LURNA). Предполагается, что LURNA и транслируемые с него пептидные продукты послужили источником формирования современного разнообразия пептидов.

Ключевые слова: рибосома, рибосомная РНК, трансляция, пептидил-трансферазный центр, универсальный пептидный мотив, Last Universal Peptide Ancestor, LUPA, Last Universal RiboNucleic Ancestor, LURNA.

ВВЕДЕНИЕ

Вопрос о монофилетическом (от одного общего предка) происхождении матрично-кодируемых белков, на первый взгляд, представляется неразрешимой задачей. Мало того, сама постановка задачи до сих пор казалась невозможной. Доминирующая точка зрения [1, 2] предполагает исходное разнообразие протеинов. Все классификации современных белков и специфических консервативных пептидных мотивов, скорее фиксируют их выраженные различия, нежели сходство [3]. Даже если предположить, что источник такого разнообразия (некий «общий предковый ген») когда-либо существовал, непонятно, где его искать (если он вообще сохранился). Какую из массива хранящихся в депозитариях нуклеотидных последовательностей можно рассматривать в качестве исходной?

В поисках такой последовательности стоит обратить внимание на молекулы РНК, находящиеся в центре современного аппарата трансляции – молекулы рибосомных РНК. Это крупные, относительно консервативные, общие для всех клеточных форм жизни полимеры, потенциально способные выполнять функцию носителей наследственной информации о структуре белков. Традиционно эти некодирующие молекулы не рассматривались как потенциальные носители генетической информации. Однако в последнее время появились работы, в которых рРНК рассматриваются как кодирующие – для ряда тРНК [4] и одновременно для тРНК и белков [5].

При применении такого подхода первым вопросом будет выбор конкретной молекулы/молекул рРНК и/или её части в качестве потенциальной матрицы трансляции, поскольку молекулы рРНК структурно и функционально гетерогенны.

Все типы молекул рРНК (для прокариот это 23S, 16S и 5S) имеют значительные различия в пределах рода и даже вида (что служит основой для филогенетической дифференциации). Причём эти различия затрагивают не только состав (замены нуклеотидов), но и структуру, что наиболее существенно при гипотетической трансляции с такой РНК. Зачастую утрачиваются и приобретаются целые ветви и группы ветвей. Наиболее выражены отличия у молекул рРНК эукариот, особенно высших животных и растений, которые отличаются более крупными размерами за счёт наличия многих дополнительных ветвей и петель. С другой стороны, молекулы рРНК эукариотических эндосимбионтов – митохондрий и пластид – отличаются утратой многих структурных элементов, что, очевидно, является результатом их редуктивной эволюции [6]. Значительная гетерогенность состава вряд ли позволяет рассматривать в качестве древнейшего источника генетической информации молекулы рРНК на всём их протяжении [5]. Достаточно часто встречающиеся замены в рРНК даже у близкородственных микроорганизмов, а также её структурные различия у прокариот разных таксонов приведут, в случае выбора рибосомной РНК какого-либо конкретного организма в качестве матричной, к вариабельным результатам гипотетической трансляции. Вероятно, в качестве источника наследственной информации следует рассматривать лишь наиболее консервативные участки молекул рРНК, не имеющие структурных различий и наименее различающиеся по составу нуклеотидов среди представителей разных, даже наиболее филогенетически далёких таксонов.

Существует точка зрения [7, 8], согласно которой современные 23S рРНК, при всём их разнообразии, являются продуктом последовательной эволюции относительно небольшого участка цепи РНК в сложную мультидоменную рибонуклеотидную структуру. В большинстве случаев таким древнейшим «ядром» рибосомы признаётся основная часть V домена 23S рРНК – пептидил-трансферазный центр (ПТЦ) [7, 8]. Именно эта часть рибосомы непосредственно отвечает за реализацию её важнейшей роли в процессе формирования транслируемой пептидной цепи.

Вторичная структура ПТЦ представляет собой специфическую структуру в виде пептидил-трансферазного кольца (ПТК) и отходящих от него ветвей (имеющих номерные обозначения от Н74 до Н93), делящих ПТК на 5 сегментов [9]. Первичная структура ПТЦ представляет собой не сплошной линейный участок молекулы 23S рРНК, а комбинацию из трёх фрагментов:

1. 5'-фрагмент (содержащий один сегмент ПТК и 5'-цепь ветвей Н74 и Н75);
2. небольшой средний фрагмент (содержащий 3'-цепь ветви Н75 и ветвь Н80);
3. наибольший 3'-фрагмент (содержащий четыре сегмента ПТК, 3'-цепь ветви Н74 и ветви Н89, Н90, Н91, Н92 и Н93).

Именно последний участок был нами выбран в качестве потенциального источника транслируемой информации. Мы предположили, что в составе этого участка следует выделить консервативный фрагмент, характеризующийся отсутствием структурных изменений (инсерций и делеций) при сравнении между различными микроорганизмами

и отличающийся достаточной длиной для предполагаемой трансляции с него пептида (пептидов) длиной в десятки аминокислотных остатков.

Для того чтобы быть репрезентативной, выборка должна включать не менее 20 микроорганизмов – представителей основных групп прокариот (цианобактерии, протеобактерии, термофильные грам-отрицательные бактерии, микоплазмы, фирмакуты, актинобактерии, археи разных таксонов). Использование только одного микроорганизма – источника рРНК (например, *E.coli*, [5]) может привести к ошибкам в установлении фрагмента рРНК, стабильного по структуре и составу. В эту группу не должны входить эукариоты ввиду их очевидно более позднего происхождения. Рибосомные РНК митохондрий и хлоропластов следует исключить вследствие их редуктивной эволюции [6].

Сравнение участков 23S рРНК, содержащих основную часть ПТЦ, между микроорганизмами различных таксономических групп позволит определить границы и размер структурно консервативной последовательности, а также установить общие для всех инвариантные нуклеотиды. В отношении участков высокой вариабельности следует опираться не только на частоту встречаемости нуклеотида в выборке микроорганизмов, но и учитывать общую филогению прокариот [10]. После определения консенсусной рибонуклеотидной последовательности такого РНК-фрагмента можно представить результаты трансляции.

Прежде чем приступить к поиску этих последовательностей в белках, сформулируем ряд ограничений, связанных с современными представлениями о предполагаемом пре-рибосомном кодировании.

Во-первых, следует учитывать отличный от современного набора аминокислот, которым могли оперировать древние системы трансляции. По этой причине среди полученных аминокислотных последовательностей для дальнейшего поиска в белках следует избегать аминокислот, являющихся (согласно распространённым представлениям [11]) эволюционно более поздними продуктами клеточного биосинтеза и обнаруживающихся в экспериментах абиогенного синтеза в минимальных количествах [12]. К таким аминокислотам следует отнести триптофан, тирозин, фенилаланин и гистидин, не имеющих очевидных абиогенно синтезируемых структурных гомологов. В отношении аргинина и лизина можно сделать допущение о наличии в структуре древних пептидов их метаболитического предшественника и структурного гомолога – орнитина [11], а для пролина – 5-оксипролина / пироглутамата. Обе эти аминокислоты встречаются в пептидах – продуктах нематричного синтеза.

Во-вторых, следует учитывать, что древний генетический код мог быть отличным от современного стандартного (не говоря уже о его более низкой специфичности, допускающей кодирование аминокислот-гомологов [11]). В нашей постановке задачи мы будем использовать стандартный код.

В-третьих, из рассмотрения следует исключить фрагменты, включающие стоп-кодона.

После отбора оставшихся (вероятно, относительно коротких) аминокислотных последовательностей, удовлетворяющих перечисленным требованиям, следует провести их поиск у прокариот различных таксономических групп (в первую очередь – в отобранных микроорганизмах). Наиболее перспективным результатом будет служить наиболее протяжённая последовательность из отобранных, которая транслируется с наименее вариабельного участка рРНК.

Маловероятно, что мы сможем найти полные точные копии искомой последовательности. Скорее всего, будут встречаться её хорошо сохранившиеся протяжённые фрагменты, особенно (если гипотеза верна) те, которые являются

результатом трансляции с участков рРНК, наиболее консервативных по нуклеотидной последовательности.

Для подтверждения того, что обнаруженные в белках последовательности, схожие с искомой, являются гомологами, а не продуктами случайных мутационно-рекомбинационных событий, необходимо построить множественное выравнивание гомологичных белков, содержащих элементы искомой последовательности и кодируемых геномами различных прокариот. В случае совпадения контекста (установления факта, что последовательность обнаруживается на одном и том же участке белков-гомологов), из таких фрагментов белков-гомологов следует составить список консервативных участков белков-гомологов, содержащих искомую последовательность. Критериями отбора найденных фрагментов белков могут служить:

- хорошая сохранность последовательности в одном и том же белке у разных прокариот,
- обнаружение последовательности в одном и том же участке белка у разных прокариот,
- присутствие последовательности в одном и том же участке белка у разных, филогенетически далёких микроорганизмов.

Дополнительным аргументом в пользу консервативности фрагментов может служить то, что содержащие их белки, скорее всего, должны быть важными, универсальными и эволюционно древними (в первую очередь – полимеразы нуклеиновых кислот, аминоацил-тРНК-синтазы, рибосомные белки и факторы трансляции).

Реликтовая природа искомых последовательностей могла стать основой их сохранения в структуре множества различных белков с самыми разнообразными функциями, положив начало их разнообразию. По этой причине не обязательно обнаружение элементов искомой последовательности в функциональной части белков. Скорее наоборот, их обнаружение будет более вероятно в участках белков с минимальной специфической активностью.

Обнаружение даже одного такого относительно протяжённого пептидного мотива во многих филогенетически и функционально различных консервативных белках у микроорганизмов – представителей широкого спектра таксонов прокариот – позволит рассматривать отобранный фрагмент рРНК как гипотетическую универсальную матричную РНК («РНК-ген»), предковую для всех современных клеточных организмов. Гипотетически она может рассматриваться в качестве последнего универсального рибонуклеотидного предшественника (Last Universal RiboNucleic Ancestor, LURNA).

Можно с большой долей вероятности предположить, что среди нескольких отобранных мотивов – продуктов трансляции одной нуклеотидной последовательности – один будет обнаруживаться в белках с наибольшей частотой. Однако, даже в случае, если такой мотив (или его фрагмент) встречается во многих основных белках филогенетически далёких прокариот чаще остальных, вряд ли можно говорить об одном универсальном пептидном предшественнике, общем для большинства современных организмов (Last Universal Peptide Ancestor, LUPA). Но вполне вероятно, что различные пептиды – продукты трансляции различных рамок считывания одной исходной молекулы РНК можно рассматривать как кластер пептидных предшественников, имеющий единый генетический источник.

МЕТОДЫ ИССЛЕДОВАНИЯ

Из базы данных Genbank [13] были отобраны 20 последовательностей генов 23S рРНК из различных прокариотических микроорганизмов, принадлежащих различным таксонам (таблица 1).

После множественного выравнивания участков 23S рРНК, содержащих основную часть ПТЦ, была выписана консенсусная последовательность консервативного фрагмента 23S рРНК. При выводе консенсуса в отношении гипервариабельных участков анализируемых рРНК принимали в расчёт близость микроорганизмов, источников генов рРНК, к основному филогенетическому стволу [3].

Расчёт результатов гипотетической трансляции (аминокислотных последовательностей) проводили с использованием стандартного генетического кода. В список установленных аминокислотных последовательностей, подлежащих поиску, отбирали только продукты гипотетической трансляции, не содержащие стоп-кодона и биогенные аминокислоты (W, Y, F, H).

Поиск аминокислотной последовательности у прокариот различных таксономических групп осуществляли с помощью программы protein BLAST [14] с использованием алгоритма blastp [15].

Таблица 1. Список микроорганизмов, отобранных в качестве источников генов 23S рРНК

№ п/п	Название микроорганизма	Таксономическая группа	Идентификатор гена	Длина гена
1	<i>Synechococcus sp. RCC307</i>	Cyanobacteria	NC_009482.1	2866
2	<i>Aquifex aeolicus VF5</i>	Aquificae	NC_000918.1	2956
3	<i>Thermus thermophilus JL-18</i>	Deinococcus-Thermus	NC_017587.1	2910
4	<i>Ehrlichia ruminantium str. Welgevonden</i>	Alphaproteobacteria	NC_006832.1	2765
5	<i>Escherichia coli str. K-12 substr. W3110</i>	Gammaproteobacteria	NC_007779.1	2904
6	<i>Treponema pallidum str. Fribourg-Blanc</i>	Spirochaetes	NC_021179.1	2951
7	<i>Chlamydia trachomatis A2497</i>	Chlamydiae	NC_017437.1	2940
8	<i>Mycoplasma pneumoniae M129-B7</i>	Mollicutes	NC_020076.1	2894
9	<i>Clostridium cf. saccharolyticum K10</i>	Firmicutes	NC_021047.1	2884
10	<i>Corynebacterium diphtheriae C7</i>	Actinobacteria	NC_016801.1	3083
11	<i>Thermoplasma volcanium GSS1</i>	Euryarchaeota	NC_002689.2	2907
12	<i>Pyrococcus furiosus COM1</i>	Euryarchaeota	NC_018092.1	2872
13	<i>Methanococcus maripaludis X1</i>	Euryarchaeota	NC_015847.1	2957
14	<i>Methanosarcina acetivorans C2A</i>	Euryarchaeota	NC_003552.1	2999
15	<i>Pyrolobus fumarii 1A</i>	Crenarchaeota	NC_015931.1	3116
16	<i>Desulfurococcus mucosus DSM 2162</i>	Crenarchaeota	NC_014961.1	3080
17	<i>Sulfolobus solfataricus 98/2</i>	Crenarchaeota	NC_017274.1	3038
18	<i>Pyrobaculum neutrophilum V24Sta</i>	Crenarchaeota	NC_010525.1	3647
19	<i>Korarchaeum cryptofilum OPF8</i>	Korarchaeota	NC_010482.1	3527
20	<i>Nitrososphaera gargensis Ga9.2</i>	Thaumarchaeota	NC_018719.1	2963

Множественное выравнивание гомологичных белков из различных прокариот с обнаруженными фрагментами последовательности проводили с помощью программы ClustalX [16].

РЕЗУЛЬТАТЫ

Установление структурных различий в участке V домена 23S рРНК прокариот различных таксонов

При сопоставлении последовательностей 23S rRNA различных микроорганизмов было установлено, что у *Clostridium saccharolyticum* имеется вставка одного нуклеотида в ветви Н88, у *Ehrlichia ruminantium* – вставка после ветви Н73.

Таким образом, в рамках группы отобранных микроорганизмов структурно неизменным оставался участок длиной 211 нуклеотидов, соответствующий на карте гена 23S rRNA *E.coli* участку от С2416 до С2626. Примечательно, что этот, самый длинный структурно инвариантный участок 23S rRNA, включает 4 из 5 сегментов основного структурного элемента ПТЦ – пептидил-трансферазного кольца, ПТК. На него же приходится 9 из 11 модифицированных оснований ПТЦ. Максимально возможная длина такого гипотетического пептида была бы равна 70 аминокислотных остатков.

Установление вариабельности нуклеотидного состава структурно стабильного фрагмента рРНК

При сопоставлении последовательностей 23S rRNA различных микроорганизмов были установлены участки различной вариабельности. Как и ожидалось, наименьшая вариабельность нуклеотидного состава отмечалась у участков, включающих сегменты ПТК, а наиболее вариабельные участки приходились на участки, максимально удалённые от ПТК (отдалённые от ПТК участки ветвей Н89, Н91 и Н73). При этом вариабельность фланкирующих участков выбранной последовательности (в ветвях перед 3'-Н74 и в 3'-Н73) была настолько велика, что мы сочли возможным пренебречь установлением в них консенсусных нуклеотидов и определением соответствующих аминокислот. Таким образом, подлежащий гипотетической трансляции фрагмент рРНК уменьшился до размера 185 нуклеотидов (соответствующий на карте гена 23S rRNA *E.coli* участку от А2432 до С2616), а максимальный размер предполагаемого пептидного продукта – до 62 аминокислот.

Конечно, выявленные различия в самой основной последовательности привели бы к различному результату трансляции. При этом в пользу каждого из вариантов можно было бы привести аргументы разной степени убедительности. В качестве основного аргумента для установления окончательного варианта последовательности в отношении выявленных гипервариабельных позиций, мы опирались на близость микроорганизма – источника рРНК основному филогенетическому стволу согласно схеме Т. Cavalier-Smith [10].

Итоговая последовательность фрагмента РНК имела следующий вид (рис. 1).

```

aaaagttaacccggggataaacaggctgatctcccccaggagttcacatcgacggggaggt
1.....10.....20.....30.....40.....50.....60
ttggcacctcgatgtcggtctctgcacacctggggtgaagcaggtcccaaggggtgggc
61.....70.....80.....90.....100.....110.....120
tgttcgccattaaagcggtaacgcagctgggtttagaacgtcctgagacagttcggtcctatc
121....130.....140.....150.....160.....170.....180.....

```

Рис. 1. Консенсусная последовательность фрагмента 23S рРНК, отобранного в качестве предполагаемого объекта трансляции. Выделено цветом: зелёным – инвариантные нуклеотиды (в выборке 20 из 20), сине-зелёным – имеющие 1 из 20 замен, светло-зелёным – от 2 до 4 замен, жёлтым – от 5 до 7 замен, красным – от 8 до 9 замен, тёмно-красным – 10 и более замен по сравнению с наиболее распространённым нуклеотидом.

Трансляция с гипотетических рамок считывания фрагмента рРНК

По результатам гипотетической трансляции (с использованием стандартного кода) с трёх возможных рамок считывания (reading frames – RF), нами получены следующие варианты аминокислотных последовательностей (рис. 2):

RF-1: kgtpgitg-spprvhidgevwhldvgssbpgaevgpkgwavrplkrvasv-nvvrqggryp
 RF-2: krvsgdnrlispkssrrgglaprcrlfaswg-srsqglgcspikavrelglerretvrsls
 RF-3: kvlgg-gadlpqefistgrfgtsmsahrilgk-vprvglfah-sgtragfrts-dssvpi

Рис. 2. Результаты гипотетической трансляции трёх возможных рамок считывания (reading frames, RF) фрагмента V домена 23S рРНК. Красным цветом выделены стоп-кодоны и «эволюционно поздние» аминокислоты (W, Y, F, H). Подчёркиванием обозначены продукты трансляции инвариантных участков рРНК.

Как видно из анализа результата трансляции, лишь 3 продукта (все – результат трансляции с одной рамки считывания) имели непрерываемый размер более 9 аминокислотных остатков (12, 11 и 29 соответственно):

1. sgdnrlispkss
2. rrgglaprcrl
3. srsqglgcspikavrelglerretvrsls

Очевидно, что нами был сделан выбор в пользу наиболее длинной последовательности. При этом для повышения специфичности поиска мы отбросили первые 4 аминокислоты, являющиеся продуктом гипотетической трансляции участка рРНК с высокой вариабельностью. Таким образом, мы выбрали для исследования последовательность длиной 25 аминокислот GLGCSPIKAVRELGLERRETVRSLS, являющуюся продуктом гипотетической трансляции с участка 23S рРНК, соответствующего на карте гена 23S rRNA *E.coli* участку от G2543 до C2616 (74 нуклеотида) и включающего фрагменты ветвей H91, H92, H90, H93, H73 и двух сегментов ПТК.

Поиск сходства наиболее протяжённой аминокислотной последовательности GLGCSPIKAVRELGLERRETVRSL с белками прокариот

С помощью программы поиска сходств в базе данных BLASTp мы провели поиск аминокислотной последовательности GLGCSPIKAVRELGLERRETVRSLS в белках различных микроорганизмов, прежде всего следующих:

1. *Synechococcus* spp.,
2. *Aquifex aeolicus*,
3. *Thermus thermophilus* ,
4. *Chlamydophila pneumoniae*,
5. *Treponema pallidum*,
6. *Escherichia coli*,
7. *Mycoplasma pneumoniae*,
8. *Bacillus subtilis*,
9. *Corynebacterium* spp.,
10. *Haloarcula marismortui*,
11. *Pyrobaculum neutrophilum*,
12. *Nitrososphaera gargensis*.

Как и предполагалось, полная 25-аминокислотная последовательность ни в одном из белков не обнаруживалась в неизменном виде. Обнаруживались только фрагменты большей или меньшей длины и структурной целостности.

При этом из анализа нами исключались фрагменты последовательности, обнаруженные в продуктах, определяемых как “hypothetical protein”, “probable conserved protein”, “unnamed protein product”, “putative ...”, “reading frame ...” etc (рис. 3).

hypothetical protein aq_2011 [Aquifex aeolicus VF5]
 Sequence ID: [ref|NP_214380.1|](#) Length: 938 Number of Matches: 8
[▶ See 1 more title\(s\)](#)

Range 1: 386 to 399 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
21.4 bits(43)	2.0	9/14(64%)	9/14(64%)	0/14(0%)

Query	6	PIKAVRELGLERRE	19
		P K VREL L RE	
Sbjct	386	PLKEVRELILKARE	399

Рис. 3. Пример обнаружения части искомой аминокислотной последовательности в первичной структуре гипотетического белка у бактерии *Aquifex aeolicus*.

Также не учитывались белки следующих групп:

- регуляторы транскрипции,
- белки модификации тРНК,
- транспозазы,
- CRISPR-ассоциированные белки,
- ABC транспортеры,
- РНКазы,
- протеазы.

Несмотря на обнаружение в таких белках последовательностей, обладающих значительным сходством с искомой (рис. 4), адекватное сопоставление представителей данных многочисленных семейств протеинов является задачей дальнейшего исследования.

LacI family transcription regulator [Thermus thermophilus HB8]
 Sequence ID: [ref|YP_145320.1|](#) Length: 330 Number of Matches: 8
[▶ See 1 more title\(s\)](#)

Range 1: 201 to 211 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
21.8 bits(44)	2.3	8/11(73%)	8/11(72%)	0/11(0%)

Query	8	KAVRELGLERR	18
		KA RE GLE R	
Sbjct	201	KAMREAGLEAR	211

Рис. 4. Пример обнаружения части искомой аминокислотной последовательности в первичной структуре регулятора транскрипции у бактерии *Thermus thermophilus*.

Для анализа отбирались только белки, удовлетворяющие следующим критериям:

- наличие не менее 5 инвариантных аминокислотных остатков от искомой последовательности в 25 аминокислотных остатков;
- отсутствие делеций и инсерций;
- детальная информация о самом белке, исключающая ошибку идентификации и предполагающая возможность сравнительного анализа с гомологами и неродственными белками.

После обнаружения таких белков и формирования выборки, проводилась оценка присутствия искомой последовательности в группах гомологичных белков у других

микроорганизмов. Зачастую, обнаруживаясь в одном из белков конкретного микроорганизма, фрагмент последовательности практически отсутствовал в аналогичном локусе белков-гомологов из других микроорганизмов. Такой результат мы расценивали либо как случайно возникшую комбинацию аминокислот в результате спонтанных мутационно-рекомбинационных событий, либо как быструю утрату исходной последовательности у большого числа других, филогенетически далёких (и, возможно, эволюционно более поздних) таксонов. В любом случае, такой фрагмент первичной последовательности белка не отбирали для последующего анализа.

Таблица 2. Пример обнаружения части искомой аминокислотной последовательности в первичной структуре фермента фосфоенолпируват синтазы в выборке филогенетически далёких прокариот. Красным цветом обозначены аминокислоты, совпадающие в последовательности белка и искомой последовательности

Микроорганизм	Идентификатор белка	Длина белка	Номер позиции	Фрагмент выравнивания
<i>Synechococcus sp. PCC 7502</i>	AFY72263.1	813	435	GRTCHAAIIARELGIPAIVGCGDAS
<i>Aquifex aeolicus VF5</i>	NP_214468.1	856	427	GRTSHAAIVARELGIPAVVGTGNAT
<i>Mucilaginibacter paludis DSM18603</i>	EHQ24539.1	819	429	GRTCHAAIVARELGVPAIVGCGNAT
<i>Escherichia coli str. K-12 W3110</i>	BAA15471.1	792	417	GRTCHAAIIARELGIPAVVGC GDAT
<i>Lactobacillus oryzae JCM 18671</i>	GAK48720.1	800	393	GMTCHAAIVSREMQIPCI VGTKSQH
<i>Corynebacterium glycinophilum</i>	AHW65035.1	812	425	GRTCHAAIIARELGIPAI VGTGDAT
<i>Haloarcula marismortui</i>	YP_136393.1	786	423	GMTSHAAIVSRELGVPA VVGAEDAT
<i>Pyrococcus yayanosii CH1</i>	YP_004624420.1	791	412	GRTSHAAIVSRELGI PCVVGT K VAT
Consensus				G C RELG+ +
RF-2				GLGCSPIKAVRELGLERRET VRSLS

Таблица 3. Пример обнаружения части искомой аминокислотной последовательности в первичной структуре фермента аланил-тРНК-синтазы в выборке филогенетически далёких прокариот. Красным цветом обозначены аминокислоты, совпадающие в последовательности белка и искомой последовательности

Микроорганизм	Идентификатор белка	Длина белка	Номер позиции	Фрагмент выравнивания
<i>Synechococcus PCC 6301</i>	BAD78854.1	892	231	LTALEKQNIDTGMGLERMAQVLQGV
<i>Aquifex aeolicus VF5</i>	NP_213887.1	867	210	LTPLPHPNIDTGMGLERIASVLQ GK
<i>Deinococcus proteolyticus</i>	ADY25918.1	892	244	LAPLPFKNIDTGMGLERVASVVDV
<i>Ehrlichia ruminantium</i>	CAH57866.1	887	221	LSVLPRKCIDTGMGLERIAAVMQGV
<i>Escherichia coli str. K-12 W3110</i>	BAA16559.1	876	228	MEPLPKPSVDTGMGLERIAAVLQHV
<i>Clostridium perfringens F262</i>	EIA16630.1	879	216	YNELAQKNIDTGMGLERIA TIMQGV
<i>Streptomyces coelicolor A3(2)</i>	NP_625781.1	890	219	LGELPSKNIDTGLGLERLAMILQGV
<i>Haloarcula marismortui</i>	WP_011224298.1	927	246	YSPMDTYIVDTGYGLERWTWMSQGT
Consensus				L P K + GLER V
RF-2				LGCSPIKAV-RELGLERRET VRSLS

Делая множественное выравнивание, мы устанавливали консенсусную последовательность аминокислот, характерную для данного участка белка (таблица 2, таблица 3). Фрагмент отбирался для последующего анализа в том случае, если он обнаруживался в аналогичном локусе белка у, по меньшей мере, пяти прокариот разных

таксонов – как правило, у двух групп грам-отрицательных бактерий (одна из которых – протеобактерии), у Firmacutes, Actinobacteria и Archea. В случае отсутствия сходства между бактериальными и архейными белками – не менее, чем у пяти бактерий разных таксонов, о чём делалось соответствующе примечание.

Был сформирован список белковых фрагментов, содержащих наибольшее количество аминокислотных остатков, идентичных и гомологичных искомой последовательности. Из-за большого объёма данных мы посчитали целесообразным отобрать только локусы, содержащие аминокислоты, совпадающие с основной частью последовательности длиной 15 аминокислотных остатков – **SPIKAVRELGLERRE**.

В результате первичного анализа было выявлено и отобрано 20 белковых фрагментов, содержащих элементы последовательности, предположительно являющихся производными одной предковой последовательности – продукта трансляции участка 23S рРНК (таблица 4).

Таблица 4. Список белковых локусов, содержащих элементы последовательности, предположительно являющейся производной от одной предковой последовательности – продукта трансляции участка 23S рРНК*

№ п/п	Белок	Длина белка	Номер начальной позиции	Консенсусная последовательность фрагмента
1	CPS_L	1028-1088	390-397	S...KALR.L...R..
2	RpL7/12*	121-131	66-73	..IKAVRE...LGL..
3	RpL18*	116-136	94-118	..A.AAREAGLE...
4	PEPs	786-856	393-435	..A.VARELGLI....
5	Acc_C	444-600	13-24	..IRA.RELGL....
6	DLD	449-499	255-274	...E...RNIGLEK..
7	PheRS_B	584-836	114-125	..I...L.ELGLE.K.
8	MetRS	497-717	53-60	...MKA.ELGI...E.
9	dnaE*	1128-1179	529-602	..M...V.ELGL.K.D
10	HdrB	265-331	19-35	...LLKELGIE...E
11	HisRS	403-466	144-154	ILR.LGL....
12	GluRS*	468-491	54-62	..L...LK.LGLD...E
13	GlyRS_A*	285-303	88-95	..L...LR.LGID...E
14	GlyRS_A*	285-303	33-42	.PA...LR.LG.E...
15	GlyRS_A*	285-303	148-157	.PV.V...YGLER..
16	RpL30*	59-62	10-12	...AVK.LGL.R..
17	HisRS*	403-466	293-331	...AV...IGLER..
18	AlaRS	867-927	40-61	...PLK...LGLE...
19	AlaRS	867-927	210-246	...I...MGLER..
20	LigA	647-691	475-489	..L.....LGLER..
	RF-2			SPIKAVRELGLERRE

* Звёздочкой обозначены белки, отобранные только у бактерий, но не у архей. Красным цветом обозначены аминокислоты, совпадающие в последовательностях большинства рассмотренных белков-гомологов и искомой последовательности. Точками обозначены негомологичные аминокислоты. Жёлтым цветом выделены возможные инсерции в «предковой» последовательности. Зелёным цветом выделены возможные транслоцированные фрагменты «предковой» последовательности.

Интересно, что источником этих фрагментов служили 16 белков, то есть некоторые белки содержали несколько фрагментов. Было найдено сходство исследуемого пептида с белками, которые можно охарактеризовать как важные, универсальные и эволюционно древние для всех клеточных форм жизни: 4 аминоксил-тРНК-синтетазы II класса (фенилаланиновая (β -субъединица – PheRS_B), глициновая (α -субъединица – GlyRS_A), аланиновая (AlaRS), гистидиновая (HisRS)), 2 аминоксил-тРНК-синтетазы I класса (метиониновая (MetRS), глутаминовая, GluRS), 3 рибосомных белка (L7/L12,

L18, L30), ДНК-полимераза III (α -субъединица – dnaE), НАД-зависимая ДНК-лигаза A (LigA) и 5 ферментов метаболизма (карбамоилфосфатсинтаза (большая субъединица – CPS_L), фосфоенолпируватсинтаза (PEPs), ацетил-CoA-карбоксилаза (субъединица A, биотинкарбоксилаза – Acc_C), дигидролипоамиддегидрогеназа (DLD), гетеродисульфидредуктаза, субъединица B (HdrB)).

Чаще всего отобранные фрагменты содержали аминокислоты из средней части «предковой» последовательности. Наиболее часто встречаются аминокислоты относящиеся к последовательности пептида из 11 аминокислот: IKAVRELGLER.

Итак, результатом нашего исследования стало выявление *аминокислотной последовательности, фрагментарно сохраняющейся в основе одних и тех же участков структурно и функционально различных белков у филогенетически далёких микроорганизмов.*

ОБСУЖДЕНИЕ

Очевидно, что полученный нами результат может иметь неоднозначную интерпретацию как в вопросе обоснования хотя бы фрагментарного происхождения генов определённых нами белков от некоего м/рРНК-ого предка, так и в вопросе ограничения поиска рамками только одной относительно короткой аминокислотной последовательности.

В первую очередь, следует отметить, что выделенные фрагменты белков, возможно, не являются филогенетически родственными потомками предковой аминокислотной последовательности. Скорее всего, выявленный мотив (A/V)RE(L/A)GL в рибосомном белке L18, фосфоенолпируватсинтазе и биотинкарбоксилазе имеет различное происхождение, являясь продуктом конвергентной эволюции. То же можно сказать и в отношении мотивов ELG(L/I) и (L/I/M)GLER. Это особенно важно, учитывая то, что аргинин и лейцин, кодируемые шестью триплетами, дважды встречаются в мотивах. Для прояснения этого следует, по меньшей мере, провести анализ нуклеотидных последовательностей генов отобранных белков.

Translation:	/S//P//I//K//A//V//R//E//L//G//L//E//R//R//
LURNA	124 tcgcccattaaagcggtagcgcGAGCTGGGTTTAGAAcgtcgt
GluRS [A.aeolicus]	169 atgctgatatagaggatttaaagtggctcggtagattgggac
GlyRS_A [A.aeolicus]	271 atttaccttgaaagccttgaaagactaggcataaatccctg
GlyRS_A [A.aeolicus]	103 aaccctgccactttcttgaaagttcttggttaaagccgtg
GlyRS_A [A.aeolicus]	451 gacgagatatccgttgagatcacttacggactggaaaggata
PheRS_B [A.aeolicus]	346 ggactccttctctccgctcaggaactcggacttgaagagaaa
CPS_L [A.aeolicus]	1176 agcactcctaaaggctgtaaggagtttagaactcgacaggta
RpL30 [A.aeolicus]	52 agacatatccaggcggtaaaagtccttaggtcttaaaaaaga
Acc_C [A.aeolicus]	46 agggctataaggcgtgtagggaaactgggattgaaactgtg
LigA [A.aeolicus]	1732 tacgacatgggttcagctcgaaagaactcggctctcctgaagatg
HdrB [A.aeolicus]	108 tccaccagaatagtagccaaagaacttggctctgaacttgac
Consensus	a.c..yat.r..g...t.aaggaaactygggtct.raa..gr.r
Translation:	/...//I//...//K//E//L//G//L//E//...//

Рис. 5. Множественное выравнивание 10 фрагментов 8 генов различных белков *Aquifex aeolicus*. Красным цветом обозначены нуклеотиды, совпадающие с последовательностью, производной от 23S рРНК. Синим цветом обозначены нуклеотиды, идентичные в большинстве сравниваемых фрагментов, но не идентичные указанной последовательности.

Например, при сравнении фрагментов генов глутаминовой аминокислот-тРНК-синтетазы, α -субъединицы глициновой аминокислот-тРНК-синтетазы (3 фрагмента), β -субъединицы фенилаланиновой аминокислот-тРНК-синтетазы, карбамоилфосфатсинтазы, рибосомного белка L30, биотинкарбоксилазы, α -субъединицы ДНК-полимеразы III и субъединицы B гетеродисульфидредуктазы бактерии *Aquifex aeolicus* можно увидеть,

что они обладают сходством. Даже в случае несовпадения отдельных нуклеотидов с последовательностью, производной от 23S рРНК, обнаруживались множественные взаимные нуклеотидные совпадения (рис. 5).

Полученные результаты позволяют определить выявленную аминокислотную последовательность IxhhxELGLE (или (I/L)xhh(K/R)ELG(L/I)E), где h – гидрофобные аминокислоты), как один из универсальных предковых пептидных мотивов. Вероятно, этот мотив является представителем группы ещё не детализированных универсальных пептидных мотивов, объединённых общностью происхождения кодирующей их нуклеотидной последовательности от одной предковой, гомологичной современной рибосомной РНК.

В дальнейшем потребуются более подробное и масштабное сравнительное изучение нуклеотидных последовательностей в участках генов, кодирующих фрагменты белков-гомологов у других микроорганизмов. Интересным результатом такого поиска может стать установление микроорганизма с наибольшим количеством неизменённых форм указанных универсальных пептидных мотивов в основных белках с ранними функциями, что может послужить аргументом в пользу его наибольшей близости к последнему универсальному клеточному предшественнику (Last Universal Cellular Ancestor, LUCA) [17]. Согласно нашим предварительным данным, вероятным кандидатом на эту роль являются бактерии рода *Aquifex*.

Такие мотивы, имея различный аминокислотный состав, но единое генетическое происхождение, можно будет определить как группу последних общих предшественников матрично-синтезируемых пептидов (Last Universal Peptide Ancestors, LUPAs), структурно лежащих в основании происхождения и эволюции современного разнообразия белков. Возможно, обнаруженный нами мотив, окажется не самым распространённым. Важно было показать саму возможность существования гипотетически м/рРНК-кодируемого пептидного мотива в гетерологичных белках филогенетически очень различающихся организмов.

Наш подход позволяет частичное построение филогенетически обусловленного универсального «куста белков»: структуры, имеющей не один, а несколько стволов по причине продукции от одной РНК-матрицы по меньшей мере 6 (с учётом комплементарной цепи РНК) вариантов аминокислотных последовательностей. Поиск аминокислотных последовательностей, полученных в результате гипотетической трансляции с цепи РНК, комплементарной установленной [5], кажется продуктивным подходом, учитывая, что для одновременного выполнения информационной и рибозимной функций требовалось одновременно несколько идентичных молекул РНК, наверняка синтезированных с исходной комплементарной цепи.

При этом следует учитывать, что гены многих современных белков могут быть потомками древних РНК-генов, появившихся раньше рассматриваемой м/рРНК и сосуществовавших параллельно с ней. По этой причине выбранный нами фрагмент рРНК следует рассматривать не как первый, но как последний общий предшественник для части (возможно, очень значительной или даже преобладающей) матричных РНК, предковых для современных белок-кодирующих генов. Более того, выбранный фрагмент рРНК, представляющийся нам наиболее вероятным кандидатом на роль последнего универсального мРНК-предшественника (LURNA), может оказаться не единственным.

Следует отметить, что сама рассчитанная нами последовательность вариабельна, т.е., вполне допускает некоторые отклонения в нуклеотидном составе, обусловленные невозможностью определить точную нуклеотидную последовательность исходного фрагмента рРНК. Поиск по другим рамкам считывания и снятие ограничений на наличие «поздних» аминокислот (что позволило значительно увеличить длину искомым последовательностей) дал интересные результаты, подтверждающие это

предположение (рис. 6-8). Проверка наличия данных последовательностей в гомологичных белках других микроорганизмов (на предмет устранения случайного совпадения) нами не осуществлялась.

AcrB/AcrD/AcrF family cation efflux protein [Aquifex aeolicus VF5]
Sequence ID: [ref|NP_213330.1|](#) Length: 1019 Number of Matches: 1
▶ [See 1 more title\(s\)](#)

Range 1: 159 to 179 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
23.1 bits(48)	0.64	Composition-based stats.	11/22(50%)	12/22(54%)	1/22(4%)

Query 39 W A V R P L K R Y A S W V X N V V R Q F G P 60
W R P L A V + V Q F G P
Sbjct 159 W V A R P L I M S A G G V E E I V - Q F G P 179

Рис. 6. Пример обнаружения части аминокислотной последовательности – продукта трансляции рамки считывания RF-1 в первичной структуре белка у бактерии *Aquifex aeolicus*.

Следует обратить внимание и на оставшиеся более мелкие фрагменты рРНК, принимающие участие в формировании пептидил-трансферазного центра, особенно отставленный нами высоко консервативный фрагмент, содержащий 3'-цепи ветвей H74 и H89, а также сегмент пептидил-трансферазного кольца, на котором непосредственно осуществляется этап транспептидации (тем более, что именно этот участок рассматривается в качестве древнейшего элемента 23S рРНК [18]).

Nitrate reductase chaperone (delta subunit) [Thermus thermophilus HB8]
Sequence ID: [emb|CAA71212.1|](#) Length: 171 Number of Matches: 1

Range 1: 77 to 103 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
21.6 bits(44)	4.0	Compositional matrix adjust.	11/27(41%)	18/27(66%)	1/27(3%)

Query 28 F A S W G X S R S Q G - L G C S P I K A V R E L G L E 53
F A + G S + G L + + + A R E + G L +
Sbjct 77 F A V Y G E S Y Q R G E L L A A L V R A F R E I G L D 103

Рис.7. Пример обнаружения части аминокислотной последовательности – продукта трансляции рамки считывания RF-2 в первичной структуре белка у бактерии *Thermus thermophilus*.

peptidoglycan-binding protein [Nostoc sp. PCC 7120]
Sequence ID: [ref|WP_010996898.1|](#) Length: 305 Number of Matches: 1
▶ [See 1 more title\(s\)](#)

Range 1: 209 to 226 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
22.7 bits(47)	4.1	Compositional matrix adjust.	10/18(56%)	13/18(72%)	0/18(0%)

Query 16 S T G R F G T S M S A H R I L G L K 33
S T R F G + S A R I G + +
Sbjct 209 S T P R F G P T P S A Q R I S G I Q 226

Рис. 8. Пример обнаружения части аминокислотной последовательности – продукта трансляции рамки считывания RF-3 в первичной структуре белка у цианобактерии *Nostoc*.

При этом следует учитывать, что менее консервативные участки 23S рРНК, вероятно, сами являются производными одной или двух исходных относительно небольших молекул рРНК (примерно соответствующих А-сайту и Р-сайту),

сформировавших раннюю структуру ПТЦ [7, 8, 19]. Такой взгляд на происхождение крупных молекул рибосомных 23S РНК современного типа приводит к мысли об отсутствии необходимости рассмотрения в качестве потенциальных кандидатов для LURNA других участков 23S рРНК.

Что касается молекулы 16S рРНК, можно предположить успешность рассмотрения участков 3'-домена, содержащего наиболее консервативные ветви h44 и h45 [20]. Однако очевидной проблемой будет большая структурная вариабельность этого участка (не говоря уже о вариабельности состава) у разных микроорганизмов по сравнению с ПТЦ-содержащими фрагментами 23S рРНК.

Важным аргументом в пользу реликтовой природы обнаруженного нами мотива является спектр содержащих его белков. Он обнаруживался в нескольких рибосомных белках (причём в наиболее сохранённом виде) и ферментах, обеспечивающих усложнение полинуклеотидных цепей – α -субъединице ДНК-полимеразы III и НАД-зависимой ДНК-лигазы А. Однако наибольшее распространение он имеет в ферментах, играющих главную роль в реализации генетического кода в современном аппарате трансляции – аминоацил-тРНК синтетазах обоих классов. Особенно интересным представляется множественное присутствие элементов искомой аминокислотной последовательности в молекуле α -субъединицы глициновой аминоацил-тРНК синтетазы (причём два фрагмента из трёх гомологичны), а также выраженное сходство одного из таких участков с участком иноклассовой глутаминовой аминоацил-тРНК синтетазы. Сравнительный анализ фрагментов генов, кодирующих данные участки, выявляет выраженное сходство, позволяющее, в рамках нашей гипотезы, сделать предположение о происхождении основы предковых частей генов данных белков от одной нуклеотидной последовательности (рис. 9).

Определённый интерес представляет спектр содержащих выделенный нами мотив белков, реализующих метаболические функции. Три из них (биотин карбоксилаза, фосфоенолпируватсинтаза и дигидролипоамиддегидрогеназа) играют ключевые роли в метаболизме углеводов с участием ацетил-КоА. Важным является присутствие мотива в гетеродисульфид редуктазе, учитывая роль металлосульфидов на самых ранних этапах биогенеза [21, 22].

```
GluRS [A. aeolicus]   169  atgctgataagaggatttaaagtggctcggtagagttgggac
GlyRS_A [A. aeolicus] 271  atttaccttgaaagccttgaaagactaggcataaatccctg
GlyRS_A [A. aeolicus] 103  aacccctgccactttcctgaaagtcttggtaaaagcgtgg
```

Рис. 9. Выравнивание участка гена глутаминовой аминоацил-тРНК синтетазы и 2 участков гена α -субъединицы глициновой аминоацил-тРНК синтетазы *Aquifex aeolicus*. Красным цветом обозначены нуклеотиды, совпадающие с последовательностью, производной от 23S рРНК. Синезелёным и светло-зелёным цветом выделены нуклеотиды, совпадающие у двух из трёх локусов; зелёным цветом выделен нуклеотид, общий для всех трёх локусов.

Показательным является присутствие мотива в составе карбоамилфосфатсинтазы – фермента, чей продукт – карбоамилфосфат является ключевым интермедиатом в орнитинном цикле и необходим для биосинтеза пиримидиновых нуклеотидов. Химически карбоамилфосфат является фосфорилированным производным формамида, рассматриваемого в качестве ключевой молекулы в происхождении как метаболизма, так и систем генетической наследственности [23, 24].

Несмотря на присутствие мотива в белках, выполняющих столь различные функции, остаётся неясным его функциональная роль. Предварительный анализ не обнаружил его присутствия в локусах белков, характеризующихся специфической активностью. Вероятно, это связано с химической природой самого мотива, содержащего много гидрофобных аминокислотных остатков. Однако, имеются наблюдения, что в

присутствии гидрофобных компонентов отмечается повышение эффективности многих рибозимов [25]. Не исключено, что вовлечение одного или нескольких таких пептидов в ранние стадии эволюции «материнской» прото-рибосомы способствовали повышению её функциональности и предопределили их фиксацию в структуре рибосомных белков. Интересно, что в рамках одного из сценариев происхождения системы трансляции важную роль играет гипотетический пептид, служащий своего рода кофактором для рибозима – предшественника рибосомы. Предполагается, что он обладает неспецифической способностью стимулировать и (или) стабилизировать рибозимы, содержит пару отрицательно заряженных аминокислот и образует комплекс с двухвалентным катионом [1]. Локализация двух остатков глутаминовой кислоты в обнаруженном нами мотиве удовлетворяет этому требованию.

Мы предполагаем, что основной первичной функцией (точнее, свойством) исходного короткого пептида была структурная. В дальнейшем при увеличении длины матрицы, скорее всего за счет различных механизмов модификации (прежде всего, дубликации) исходного гена этот пептид играл роль низко специфической универсальной основы, своего рода «стержня», «каркаса» или «рукоятки» для множества других, более специализированных пептидных доменов – частей новой более длинной молекулы. Такая «скелетная» роль данного предполагаемого предкового пептида могла стать основой его сохранения в структуре множества различных белков с самыми разнообразными функциями.

ЗАКЛЮЧЕНИЕ

Проведённые нами исследования позволили сделать два основных вывода.

В разных участках первичной структуры многих консервативных белков, даже структурно и функционально неродственных, стабильно обнаруживается сходство с элементами аминокислотной последовательности IKAVRELGLER, по-видимому, являющейся потомком древнего пептида с неустановленными функциями, который можно назвать одним из группы последних общих предшественников матрично-синтезируемых пептидов (Last Universal Peptide Ancestors, LUPAs).

Вероятно, этот пептид является продуктом трансляции одной из рамок считывания молекулы РНК, принимавшей участие в формировании значительной части основного (и, вероятно, древнейшего) функционального участка рибосомы – пептидил-трансферазного центра, ПТЦ. Будучи одновременно *двойной предковой формой*: структурно – для ПТЦ, информационно – для спектра древних пептидов, эта последовательность РНК (содержащая в центральной части подпоследовательность ATTAAAGCGGTACGCGAGCTGGGTTTAGAACGT) является, вероятно, последним универсальным РНК-предшественником (Last Universal RiboNucleic Ancestor, LURNA).

СПИСОК ЛИТЕРАТУРЫ

1. Koonin E.V. *The Logic of Chance: The Nature and Origin of Biological Evolution*. Upper Saddle River: FT Press, 2011. 528 p.
2. Caetano-Anolles G., Wang M., Caetano-Anolles D., Mittenthal J.E. The origin, evolution and structure of the protein world. *Biochem. J.* 2009. V.417. P. 621–637. doi: [10.1042/BJ20082063](https://doi.org/10.1042/BJ20082063)
3. Sobolevsky Y., Guimarães R.C., Trifonov E.N. Towards functional repertoire of the earliest proteins. *J. Biomol. Struct. Dyn.* 2013. V.31. № 11. P.1293–300. doi: [10.1080/07391102.2012.735623](https://doi.org/10.1080/07391102.2012.735623)
4. de Farias S.T., do Rêgo T.G., José M.V. Evolution of transfer RNA and the origin of the translation system. *Front. Genet.* 2014. V. 5. P.303. doi: [10.3389/fgene.2014.00303](https://doi.org/10.3389/fgene.2014.00303)

5. Root-Bernstein M., Root-Bernstein R. The ribosome as a missing link in the evolution of life. *Journal of Theoretical Biology*. 2015. V. 367. № 21. P. 130–158 doi: [10.1016/j.jtbi.2014.11.025](https://doi.org/10.1016/j.jtbi.2014.11.025)
6. Greber B.J., Boehringer D., Leitner A., Bieri P., Voigts-Hoffmann F., Erzberger J.P., Leibundgut M., Aebersold R., Ban N. Reductive evolution of the mitochondrial 16S rRNA. *Nature*. V. 505. P. 515–519 doi: [10.1038/nature12890](https://doi.org/10.1038/nature12890)
7. Bokov K., Steinberg S.V. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*. 2009. V. 457. P. 977–980. doi: [10.1038/nature07749](https://doi.org/10.1038/nature07749)
8. Hsiao Ch., Mohan S., Kalahar B.K., Williams L.D. Peeling the onion: ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* 2009. V. 26 № 11. P. 2415–25. doi: [10.1093/molbev/msp163](https://doi.org/10.1093/molbev/msp163)
9. Polacek N., Mankin AS. The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit. Rev. Biochem. Mol. Biol.* 2005. V. 40. № 5. P. 285–311. doi: [10.1080/10409230500326334](https://doi.org/10.1080/10409230500326334)
10. Cavalier-Smith T. Rooting the tree of life by transition analyses. *Biology Direct*. 2006. V. 1. P. 19. doi: [10.1186/1745-6150-1-19](https://doi.org/10.1186/1745-6150-1-19)
11. Copley S.D., Smith E., Morowitz H.J. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc Natl Acad Sci USA*. 2005. V. 102. № 12. P. 4442–4447. doi: [10.1073/pnas.0501049102](https://doi.org/10.1073/pnas.0501049102)
12. Miller S.L., Urey H.C. Organic compound synthesis on the primitive Earth. *Science*. 1959. V. 3370. №130. P.245–251.
13. GENBANK Entrez Nucleotide Database. URL: <http://www.ncbi.nlm.nih.gov/nucleotide/> (accessed 15.03.2015).
14. Basic Local Alignment Search Tool. URL: <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (accessed. 15.03.2015).
15. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic local alignment search tool. *J. Mol. Biol.* 1990. V. 215. № 3. P. 403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
16. Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G., Gibson T.J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 1998. V. 23. P. 403–405. doi: [10.1016/S0968-0004\(98\)01285-7](https://doi.org/10.1016/S0968-0004(98)01285-7)
17. Woese C. The universal ancestor. *PNAS*. 1998. V. 95. №. 12. P. 6854–6859. doi: [10.1073/pnas.95.12.6854](https://doi.org/10.1073/pnas.95.12.6854)
18. Petrov A.S., Bernier C.R., Hsiao Ch., Norris A.M., Kovacs N.A., Waterbury C.C., Stepanov V.G., Harvey S.C., Fox G.E., Wartell R.M., Hud N.V., Williams L.D. Evolution of the ribosome at atomic resolution. *PNAS*. 2014. V. 111. № 28. P. 10251–10256. doi: [10.1073/pnas.1407205111](https://doi.org/10.1073/pnas.1407205111)
19. Tamura K. Ribosome evolution: Emergence of peptide synthesis machinery. *Journal of Biosciences*. 2011. V. 36. № 5. P. 921–928. doi: [10.1007/s12038-011-9158-2](https://doi.org/10.1007/s12038-011-9158-2)
20. Harish A., Caetano-Anollés G. Ribosomal History Reveals Origins of Modern Protein Synthesis. *PLoS One*. 2012. V.7. № 3. e32776. doi: [10.1371/journal.pone.0032776](https://doi.org/10.1371/journal.pone.0032776)
21. Pascal R., Boiteau L., Forterre P., Gargaud M., Lazcano A., Lopez-Garcia P., Maurel M-C., Moreira D., Pereto J., Prieur D., Reisse J. Prebiotic Chemistry – Biochemistry - Emergence of Life (4.4-2 Ga). *Earth, Moon and Planets*. 2007. V. 98. P. 153–203. doi: [10.1007/s11038-006-9089-3](https://doi.org/10.1007/s11038-006-9089-3)
22. Mulikidjanian A.Y., Galperin M.Y. On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. *Biol. Direct*. 2009. V. 4. P. 27. doi: [10.1186/1745-6150-4-27](https://doi.org/10.1186/1745-6150-4-27)
23. Saladino R., Botta G., Pino S., Costanzo G., Di Mauro E. Genetics first or metabolism first? The formamide clue. *Chem. Soc. Rev.* 2012. V. 41. № 16. P. 5526–5565. doi: [10.1039/c2cs35066a](https://doi.org/10.1039/c2cs35066a).

24. Pino S., Sponer J.E., Costanzo G., Saladino R., Di Mauro E. From Formamide to RNA, the Path Is Tenuous but Continuous. *Life*. 2015. V. 5. P. 372–384. doi: [10.3390/life5010372](https://doi.org/10.3390/life5010372)
25. Müller U.F., Bartel D.P. Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA*. 2008. V. 14. № 3. P. 552–562. doi: [10.1261/rna.494508](https://doi.org/10.1261/rna.494508)

Материал поступил в редакцию 02.04.2015, опубликован 15.04.2015.