

## Гомологи РНК-лигазы 2 бактериофага Т4 в метагеномах океанической микробиоты

Зимин А.А.<sup>\*1</sup>, Никулин Н.А.<sup>1</sup>, Назипова Н.Н.<sup>2</sup>

<sup>1</sup>Институт биохимии и физиологии микроорганизмов им. Г. К. Скрыбина РАН –  
обособленное подразделение ФИЦ ПНЦБИ РАН, Пущино, Россия

<sup>2</sup>Институт математических проблем биологии РАН – филиал Института прикладной  
математики им. М.В. Келдыша РАН, Пущино, Россия

**Аннотация.** РНК-лигаза 2 фага Т4 – это уникальный фермент, функционально сходный, в отличие от других РНК-лигаз, с ДНК-лигазами, а также родственный редактирующим РНК-лигазам паразитических трипаносом и лейшманий. РНК-лигазы 2 присутствуют в ограниченном (небольшом) количестве геномов, которые, к тому же, сильно разбросаны по древу жизни. В работе был проведен поиск гомологов РНК-лигазы 2 в базах данных пелагических океанических генетических данных GOS и глубоководной осадочной микробиоты LCGC14. В метагеномах пелагической и осадочной глубоководной микробиоты было найдено соответственно 6 и 15 гомологов РНК-лигазы 2 бактериофага Т4, пригодных для анализа. Филогенетический анализ обнаруженных аминокислотных последовательностей показал, что большинство из них проявляют сходство с гомологами РНК-лигазы 2 из бактерий и грибов. На ветви филогенетического дерева, общей для гомологов из подсемейства бактериофагов *Tevenvirinae* и протистов типа *Euglenozoa*, были найдены пять гомологов океанического происхождения. Этот результат говорит о наличии как в толще воды открытого океана, так и на его дне новых, еще неизвестных организмов, геномы которых кодируют этот редкий фермент.

**Ключевые слова:** океанические метагеномы, генетические исследования глубоководной осадочной и пелагической микробиоты, РНК-лигаза 2, геномика бактериофагов.

### ВВЕДЕНИЕ

РНК-лигазы относятся к суперсемейству нуклеотидилтрансфераз. Вместе с ними туда входят ферменты, кэпирующие мРНК, тРНК-лигазы и ДНК-лигазы [1]. Все ферменты этого суперсемейства катализируют образование фосфодиэфирной связи по консервативному трехступенчатому механизму, который использует в качестве высокоактивного кофактора АТФ, ГТФ либо НАД<sup>+</sup> [1–3].

РНК-лигазы (ЕС 6.5.1.3) – это ферменты, соединяющие концы РНК. Они участвуют в процессах репарации, сплайсинга и редактирования РНК. В отличие от широко распространенных ДНК-лигаз, РНК-лигазы имеют более узкое филогенетическое распределение. Поиск гомологичных последовательностей позволил обнаружить РНК-лигазы во всех трех филах клеточных форм жизни, но только в ограниченном подмножестве видов [4]. РНК-лигазы подразделяются на два больших семейства [3, 5].

Семейство Rnl1 включает одноименную РНК-лигазу 1 (Rnl1) бактериофага Т4 [2] и тРНК-лигазы из грибов, дрожжей и растений [3, 5, 6]. Эти ферменты восстанавливают разрывы, которые были введены в одноцепочечную РНК сайт-специфическими нуклеазами. Фермент РНК-лигаза 1 обнаруживается у вирусов, млекопитающих и у

<sup>\*</sup>zimin@ibpm.pushchino.ru

грибов [6]. Этот фермент у вирусов, например, у бактериофага T4, используется для защиты от противовирусных стратегий бактерий [2], но также участвует в сплайсинге интронов тРНК [5] и в нетрадиционном сплайсинге РНК, инициируемым неправильной сборкой белков эндоплазматического ретикулула.

Семейство Rnl2 содержит РНК-лигазу 2 бактериофага T4 и лигазы, редактирующие митохондриальные информационные РНК, которые встречаются у простейших, таких как трипаносомы и лейшмании. Эти ферменты, в первую очередь, восстанавливают одноцепочечные разрывы в двуцепочечной РНК благодаря наличию комплементарной цепи - мостика [4, 7, 8]. РНК-лигазы 2 имеют широкое, но пунктирное распределение по всему древу жизни [6]: они обнаруживаются главным образом у вирусов с архетипическим примером РНК-лигазы 2 бактериофага T4 [3], а также у бактерий, в то время как у архей и эукариот известно лишь несколько подобных белков. Биологическая роль бактериальных РНК-лигаз 2 неизвестна, за исключением участия в жизни кинетопластов трипаносом [9–12].

РНК-лигазы имеют шесть консервативных нуклеотидилтрансферных мотивов, общих с ДНК-лигазами, однако уровни сходства этих консенсусов достаточно низкие. Это делает классификацию РНК-лигаз сложной, а ее результаты противоречивыми.

Для полноты картины отметим, что были сообщения о двух неканонических РНК-лигазах из архей *Pyrococcus horikoshii*. Первой является предполагаемая 2'-5'-РНК-лигаза, структура которой была исследована [13]. Вторая, RtcB, представляет собой очень необычную ГТФ-зависимую лигазу, которая присоединяет 2',3'-циклофосфатный или 3'-фосфатный конец РНК к 5'-гидроксильному концу РНК. Структура этого фермента и механизм его взаимодействия с новым белком-кофактором (Archease) были охарактеризованы сравнительно недавно [14, 15].

Как и ДНК-лигазы, РНК-лигазы широко применяются в молекулярной биологии. РНК-лигазы 1 и 2 бактериофага T4 стали незаменимыми компонентами множества методов быстрой амплификации концов кДНК, мечения 3'-РНК и, что наиболее важно в настоящее время, получения библиотек для секвенирования микроРНК (miRNA) [16]. АТФ-зависимые РНК-лигазы, способные образовывать фосфодиэфирные связи между 5'-фосфатным и 3'-гидроксильным кончиками, еще долго будут находиться в центре внимания экспериментаторов.

## ЗАДАЧА ИССЛЕДОВАНИЯ

До сих пор считалось, что РНК-лигазы 2 редки и сильно разбросаны по древу жизни. Лишь два члена этого семейства хорошо изучены: митохондриальная РНК-редактирующая лигаза из паразитических трипаносом (Kinetoplastea), многообещающая лекарственная мишень, и РНК-лигаза 2 бактериофага T4, рабочий инструмент современной молекулярной биологии [17].

Установление происхождения, эволюции и биологической роли этого необычного фермента требует поисков его гомологов везде, где это возможно. Биота мирового океана представляют собой огромный и малоизученный источник генов. Целью данного исследования является поиск гомологов хорошо исследованных фаговых полинуклеотидлигаз в метагеномах океанической микробиоты.

Обширное метагеномное исследование морской планктонической микробиоты было проведено из поверхностных вод в Средней Атлантике и рядом с Панамским каналом в Тихом океане [18, 19]. Было определено 7,7 млн. последовательностей ДНК общей длиной 6,3 Gbp. База данных GOS содержит более 6 млн. аминокислотных последовательностей, транслированных из ДНК.

Для анализа последовательностей ДНК пелагической микробиоты была извлечена фракция (0,1–0,8 мкм) из 37 новых поверхностных морских, пресноводных и гиперсолевых проб, собранных во время первой фазы экспедиции глобального отбора

проб в океане (GOS) Sorcerer II [18, 19] с дополнительными данными с четырех станций, отобранных в рамках пилотного исследования Саргассового моря. Были предприняты попытки охарактеризовать вирусные последовательности из этой базы данных с точки зрения их встречаемости и распространения в разнообразных водных экосистемах. Для этого авторы использовали сравнительный геномный анализ для функциональной характеристики вирусных последовательностей посредством кластеризации сходства последовательностей и для выяснения важности захвата генов хозяина, кодирующих экологически значимые метаболические функции в водной среде, вирусами океана. Это важно с точки зрения возможности горизонтального переноса генов.

Другая большая база океанических генетических данных была создана при исследовании глубоководных осадков. В этой работе были получены данные о метагеномах глубоководной морской микробиоты донных отложений Северного Ледовитого океана в районе подводного Срединного арктического хребта имени Гаккеля (Arctic Mid-Ocean Ridge) [20]. Последовательности ДНК были получены методом глубокого метагеномного секвенирования ДНК образца осадка GC14, в результате чего был получен меньший по размеру набор генетических последовательностей (LCGC14, 8.6 Гбит). Далее была произведена амплификация ДНК (MDA) и получен большой набор метагеномных данных (LCGC14AMP, 56.6 Gbp) (<http://opensource.scilifelab.se/>) [20]. Одновременно были созданы базы данных аминокислотных последовательностей из генетических данных, они были получены путем трансляции стандартным генетическим кодом.

В данной работе базы данных GOS и LCGC14 были проанализированы на наличие в них последовательностей, сходных с РНК-лигазой 2 бактериофага T4. После создания набора гомологов РНК лигазы 2 бактериофага T4 в метагеномах морской микробиоты было проведено филогенетическое исследование, целью которого было выяснить, к каким известным на данный момент времени последовательностям они проявляют наибольшее сходство.

## МАТЕРИАЛЫ И МЕТОДЫ

### Характеристика последовательностей РНК-лигаз 2 из близкородственных организмов

Для характеристики аминокислотной последовательности РНК-лигазы 2 бактериофага T4 мы провели её филогенетическое сравнение с рядом гомологов из родственных бактериофагов. Для этого была создана выборка аминокислотных последовательностей продуктов гена *rnIB* из Genbank из хорошо аннотированных [21] геномов родственных бактериофагов (подсемейство *Tevenvirinae*). В ряде случаев в аннотациях последовательностей геномов в кластере РНК-лигаз 2 присутствовал не один, а два белка, в этом случае оба паралога включались в выборку. При сравнении филогенетического дерева *Tevenvirinae*, построенного на основе аминокислотных последовательностей гомологов белка *RnIB*, с используемой на данный момент классификацией вирусов по геномной идентичности [22, 23], было обнаружено, что в целом полученное дерево соответствует современному представлению об эволюции данных вирусов. Что интересно, аминокислотные последовательности РНК-лигаз 2 рода *Tequatrovirus* оказались разделенными на две клады, одна из которых также объединила в свой состав белки фагов рода *Mosigvirus*. Так как эти два рода являются достаточно близкими друг к другу, можно сделать предположение о рекомбинационных процессах, которые произошли между предками представителей данной объединенной клады в области исследуемого локуса, нуклеотидные последовательности которого кодируют гомологи белков *RnIB*. В остальном же

наблюдается четкая корреляция между сходством геномов и аминокислотных последовательностей РНК-лигаз 2 внутри рассмотренных родов подсемейства *Tevenvirinae*. Возможность рекомбинации между родами в один из которых включен Т4, широкий спектр хозяев вирусов данного подсемейства, самые различные экологические ниши, которые занимают эти бактерии-хозяева, а также широкое распространение гомологов гена, кодирующего РНК-лигазу 2 у родственных Т4 фагов, обосновывают использование аминокислотной последовательности именно гомолога из Т4 в качестве целевой последовательности для анализа морских метагеномов как с точки зрения филогении, так и экологии. Проведенный филогенетический анализ также указывает на распространение этого фермента у фагов данного подсемейства развивающихся в самых разнообразных экологических нишах, включая в себя и морские биомы.

В Таблице 1 перечислены 189 РНК-лигаз 2 (белков Rn1B) из близкородственных фагу Т4 бактериофагов, которые были найдены в GenBank [24] с помощью инструмента PSI-BLAST [25] и использованы для анализа степени консервативности аминокислотной последовательности этого фермента. Классификация Т-четных бактериофагов взята из GenBank [23], в начале таблицы идет неклассифицированная часть семейства. Рядом с названиями родов помещены цветные кружки. В последней колонке таблицы приведены идентификаторы GenBank для аминокислотных последовательностей РНК-лигаз 2 анализированных бактериофагов. В тех строках последней колонки, где вместо идентификатора GenBank стоят знак «\*», аминокислотная последовательность РНК-лигазы 2 была получена путем использования программного обеспечения для аннотирования геномов RASTtk [21].

**Таблица 1.** Бактериофаги, РНК-лигазы 2 которых были использованы для анализа степени консервативности аминокислотной последовательности, сгруппированные по родам

№	род	вид	Идентификатор GenBank
1	unclassified <i>Tevenvirinae</i>	Acinetobacter phage 133	YP_004300608.1
2		Acinetobacter phage AM101	AWY10415.1
3		Acinetobacter phage Ac42	YP_004009372.1
4		Acinetobacter phage Acj61	YP_004009627.1
5		Acinetobacter phage Acj9	YP_004010147.1
6		Acinetobacter phage KARL-1	AXY82640.1
7		Acinetobacter phage ZZ1	YP_006488989.1
8		Acinetobacter phage vB_ApiM_fHyAci03	AXF40578.1
9		Aeromonas phage 65.2	APU01459.1
10		Aeromonas phage 65	YP_004300909.1
11		Aeromonas phage AS-szw	ATI17438.1
12		Aeromonas phage AS-zj	ASU00157.1
13		Aeromonas virus Aeh1	NP_944126.1
14		Aeromonas phage AsFcp_2	QAX98490.1
15		Aeromonas phage Asswx_1	QAX97878.1
16		Aeromonas phage Aswh_1	QAY01272.1
17		Aeromonas phage CC2	YP_007010325.1
18		Aeromonas phage PX29	YP_009011662.1
19		Citrobacter phage IME-CF2	YP_009218766.1
20		Citrobacter phage Margaery	YP_009195056.1
21		Citrobacter phage Maroon	AYJ73100.1
22		Citrobacter phage Miller	YP_009097842.1
23		Citrobacter phage vB_CfrM_CfP1	YP_009285773.1
24		Enterobacteria phage RB16	YP_003858534.1
25		Enterobacteria phage RB43	YP_239225.1

26		Cronobacter phage vB_CsaM_GAP161	YP_006986511.1
27		Cronobacter phage vB_CsaM_leB	AOG16366.1
28		Cronobacter phage vB_CsaM_leN	AOG16651.1
29		Erwinia phage Cronus	AWD90329.1
30		Escherichia phage Lw1	YP_008060759.1
31		Klebsiella phage Marfa	QDB71837.1
32		Proteus phage phiP4-3	AUM58455.1
33		Proteus phage vB_PmiM_Pm5461	YP_009195583.1
34		Pseudomonas phage PspYZU05	ASD52094.1
35		Klebsiella phage vB_Kpn_F48	AUO78857.1
36		Morganella phage vB_MmoM_MP1	YP_009280031.1
37		Pectobacterium bacteriophage PM2	YP_009211620.1
38		Serratia phage PS2	AHY25432.1
39		Vibrio phage vB_VmeM-32	ALY07226.1
40	● <i>Moonvirus</i>	Citrobacter phage Merlin	YP_009203914.1
41		Citrobacter phage Moon	YP_009146629.1
42		Cronobacter phage Pet-CM3-4	SCN45872.1
43	● <i>Karamvirus</i>	Enterobacter phage PG7	YP_009005458.1
44		Enterobacter phage myPSH1140	AVR55365.1
45		Enterobacteria phage CC31	YP_004010038.1
46		Enterobacter phage phiEap-3	YP_009607157.1
47	● <i>Slopekvirus</i>	Klebsiella phage Matisse	YP_009194487.1
48		Klebsiella phage Miro	YP_009607435.1
49		Enterobacteria phage HX01	*
50		Enterobacteria phage RB69	NP_861881.1
51		Enterobacteria phage ATK47	ANZ51023.1
52		Enterobacteria phage ATK48	ANZ51366.1
53		Escherichia coli O157 typing phage 6	YP_009593176.1
54		Escherichia coli O157 typing phage 3	YP_009592780.1
55		Escherichia phage APCEc01	YP_009225077.1
56		Escherichia phage HP3	*
57		Escherichia phage OLB35	AYR04095.1
58		Escherichia phage SF	AWY07818.1
59		Escherichia phage ST0	YP_009608487.1
60		Escherichia phage p000v	AYN56237.1
61		Escherichia phage p000y	AYN56688.1
62		Escherichia phage vB_EcoM_G2285	QBO62608.1
63		Escherichia phage vB_EcoM_G2469	QBO62878.1
64	● <i>Mosigvirus</i>	Escherichia phage vB_EcoM_G53	QBO65324.1
65		Escherichia phage vB_EcoM_JS09	YP_009037590.1
66		Escherichia phage vB_EcoM_KAW3E185	QBQ78739.1
67		Escherichia phage vB_EcoM_MM02	QBQ79237.1
68		Escherichia phage vB_EcoM_WFK	QBQ77198.1
69		Escherichia phage vB_EcoM_WFL6982	QBQ76937.1
70		Escherichia phage vB_EcoM_WFbE185	QBQ77690.1
71		Escherichia phage vB_EcoM_PhAPEC2	YP_009056764.1
72		Escherichia virus vB_Eco_mar005P1	VCU44444.1
73		Escherichia virus vB_Eco_mar005P1 strain vB_Eco_mar006P2	VCU44449.1
74		Escherichia virus vB_Eco_mar005P1 strain vB_Eco_mar007P3	VCU43253.1
75		Escherichia_virus_vB_Eco_mar005P1_strain_vB_Eco_mar008P4	VCU43961.1
76		Escherichia_virus_vB_Eco_mar005P1_strain_vB_Eco_mar009P5	VCU44820.1
77		Shigella phage SHSML-52-1	YP_009289105.1
78	● <i>Dhakavirus</i>	Enterobacteria phage IME08	YP_003734320.1
79		Enterobacteria phage JS10	YP_002922522.1
80		Escherichia phage JS98	YP_001595305.1



81		Escherichia phage QL01	YP_009202912.1
82		Enterobacteria phage vB_EcoM_VR5	YP_009205871.1
83		Escherichia phage AnYang	QAU03652.1
84		Escherichia phage EcWhh-1	QAX99933.1
85		Escherichia phage MX01	YP_009324073.1
86		Escherichia phage WG01	YP_009323380.1
87		Phage NC-G	QBP35520.1
88	● <i>Gaprivervirus</i>	Escherichia phage vB_EcoM_VR7	YP_004063874.1
89		Escherichia phage vB_EcoM_VR20	YP_009207369.1
90		Escherichia phage vB_EcoM_VR25	YP_009209935.1
91		Escherichia phage vB_EcoM_VR26	YP_009214027.1
92		Shigella phage SP18	YP_003934814.1
93	● <i>Tequatrovirus</i>	Enterobacteria phage Aplg8	ANZ50774.1
94		Enterobacteria phage Bp7	YP_007004267.1
95		Enterobacteria phage GiZh	*
96		Escherichia phage AR1	YP_009167991.1
97		Enterobacteria phage Kha5h	ANZ51799.1
98		Enterobacteria phage RB10	AIT74268.1
99		Escherichia virus RB14	YP_002854510.1
100		Escherichia virus RB32	YP_803117.1
101		Enterobacteria phage RB18	AXF42485.1
102		Enterobacteria phage RB27	YP_009102379.1
103		Enterobacteria phage RB33	AIT74814.1
104		Escherichia phage RB3	YP_009098560.1
105		Enterobacteria phage RB51	ACP31095.1
106		Enterobacteria phage RB55	AIT75086.1
107		Enterobacteria phage RB59	AIT75360.1
108		Enterobacteria phage RB5	AIT73181.1
109		Enterobacteria phage RB68	AIT75637.1
110		Enterobacteria phage RB6	AIT73452.1
111		Enterobacteria phage RB7	AIT73723.1
112		Enterobacteria phage RB9	AIT73995.1
113		Enterobacteria phage T4T	ADJ39898.1
114		Escherichia virus T4	NP_049790.1
115		Enterobacteria phage T4 strain wild	*
116		Enterobacteria phage T6	AXN58211.1
117		Escherichia phage vB_EcoM_ACG-C40	YP_006986729.1
118		Enterobacteria phage vB_EcoM_IME339	AWD91505.1
119		Enterobacteria phage vB_EcoM_IME340	AWD91761.1
120		Escherichia phage CF2	*
121		Escherichia phage D5505	QBO61184.1
122		Escherichia phage ECML-134	YP_009102648.1
123		Escherichia phage HY01	YP_009148617.1
124		Escherichia phage HY03	YP_009284038.1
125		Escherichia phage HY03	YP_009284039.1
126		Escherichia phage KIT03	BBG28692.1
127		Escherichia phage PE37	ANH49691.1
128		Escherichia phage PP01	BBC14517.1
129		Escherichia phage T2	BBC14796.1
130	Escherichia phage UFV-AREG1	YP_009281510.1	
131	Escherichia phage vB_EcoM_112	YP_009030783.1	
132	Escherichia phage slur02	YP_009210095.1	
133	Escherichia phage slur03	YP_009625074.1	
134	Escherichia phage slur04	YP_009625392.1	
135	Escherichia phage slur07	YP_009197280.1	
136	Escherichia phage slur08	CUL02618.1	
137	Escherichia phage slur11	CUL02946.1	
138	Escherichia phage slur13	CUL03732.1	
139	Escherichia phage slur14	YP_009180676.1	

140		Escherichia phage vB_EcoM-G28	AVH86005.1
141		Escherichia phage vB_EcoM-Sa451w	QDF15200.1
142		Escherichia phage vB_EcoM-UFV13	YP_009290443.1
143		Escherichia phage vB_EcoM-ffEco06	AUV61038.1
144		Escherichia phage vB_EcoM-fHoEco02	AUV61311.1
145		Escherichia phage vB_EcoM_DalCa	AYP69772.1
146		Escherichia phage vB_EcoM_G10400	QBO63692.1
147		Escherichia phage vB_EcoM_G2540-3	QBO65583.1
148		Escherichia phage vB_EcoM_G2540	QBO63414.1
149		Escherichia phage vB_EcoM_G29	QBO64498.1
150		Escherichia phage vB_EcoM_G4498	QBO64219.1
151		Escherichia phage vB_EcoM_G4500	QBO65870.1
152		Escherichia phage vB_EcoM_G4507	QBO66145.1
153		Escherichia phage vB_EcoM_G50	QBO65046.1
154		Escherichia phage vB_EcoM_G8	QBQ80057.1
155		Escherichia phage vB_EcoM_G9062	QBQ77959.1
156		Escherichia phage vB_EcoM_JB75	AXC34018.1
157		Escherichia phage vB_EcoM_KAW1E185	QBQ78462.1
158		Escherichia phage vB_EcoM_OE5505	QBQ79505.1
159		Escherichia phage vB_EcoM_R5505	QBQ79789.1
160		Escherichia phage wV7	YP_007004917.1
161		Escherichia virus VEc20	QDK04303.1
162		Shigella phage SH7	APC45045.1
163		Shigella phage SHBML-50-1	YP_009288541.1
164		Shigella phage SHFML-11	YP_009277549.1
165		Shigella phage Sf21	YP_009618985.1
166		Shigella phage Sf22	YP_009614828.1
167		Shigella phage Sf23	ATE86434.1
168		Shigella phage Sf24	YP_009619109.1
169		Shigella phage Shf12	YP_004415069.1
170		Shigella phage pSs-1	YP_009110995.1
171		Yersinia phage PST	YP_009153775.1
172		Yersinia phage fPS-2	VEV89788.1
173		Yersinia phage phiD1	YP_009149419.1
174		Klebsiella phage KP1	AUV57527.1
175	● <i>Jiaodavirus</i>	Klebsiella phage KPV15	APD20579.1
176		Klebsiella phage Mineola	AWY07077.1
177		Klebsiella phage PKO111	YP_009289602.1
178		Klebsiella phage vB_KpnM_KpV477	YP_009288865.1
179		Salmonella phage Melville	YP_009615654.1
180	● <i>Gelderlandvirus</i>	Salmonella phage vB_SenMS16	YP_007501209.1
181		Salmonella phage STML-198	YP_009148159.1
182		Salmonella phage STP4-a	YP_009126374.1
183		Salmonella phage vB_SnwM_CGG4-1	YP_009286529.1
184		Vibrio phage KVP40	NP_899381.1
185		Vibrio phage ValB1 HC	QBX05956.1
186	● <i>Schizotequatrovirus</i>	Vibrio phage ValKK3	YP_009201230.1
187		Vibrio phage phi-Grn1	ALP47010.1
188		Vibrio phage phi-ST2	ALP47390.1
189		Vibriophage phi-pp2	AFN37364.1

На рисунке 1 показано филогенетическое дерево РНК-лигаз 2 подсемейства Т-четных бактериофагов, построенное с помощью метода максимального правдоподобия (Maximum Likelihood Method) с использованием модели матриц замен JTT [26]. Консенсусное дерево было получено после 1000 итераций bootstrap-анализа, предпринятых для выяснения эволюции анализируемых таксонов [27]. Ветви, образовавшиеся менее чем в 50% повторов статистического анализа, были убраны. Исходное дерево было получено методом максимальной экономии (Maximum

Parsimony Method). Все позиции, содержащие бреши или места, где данные отсутствовали, были убраны. Исследование проводили с помощью пакета программ MEGAX [28].

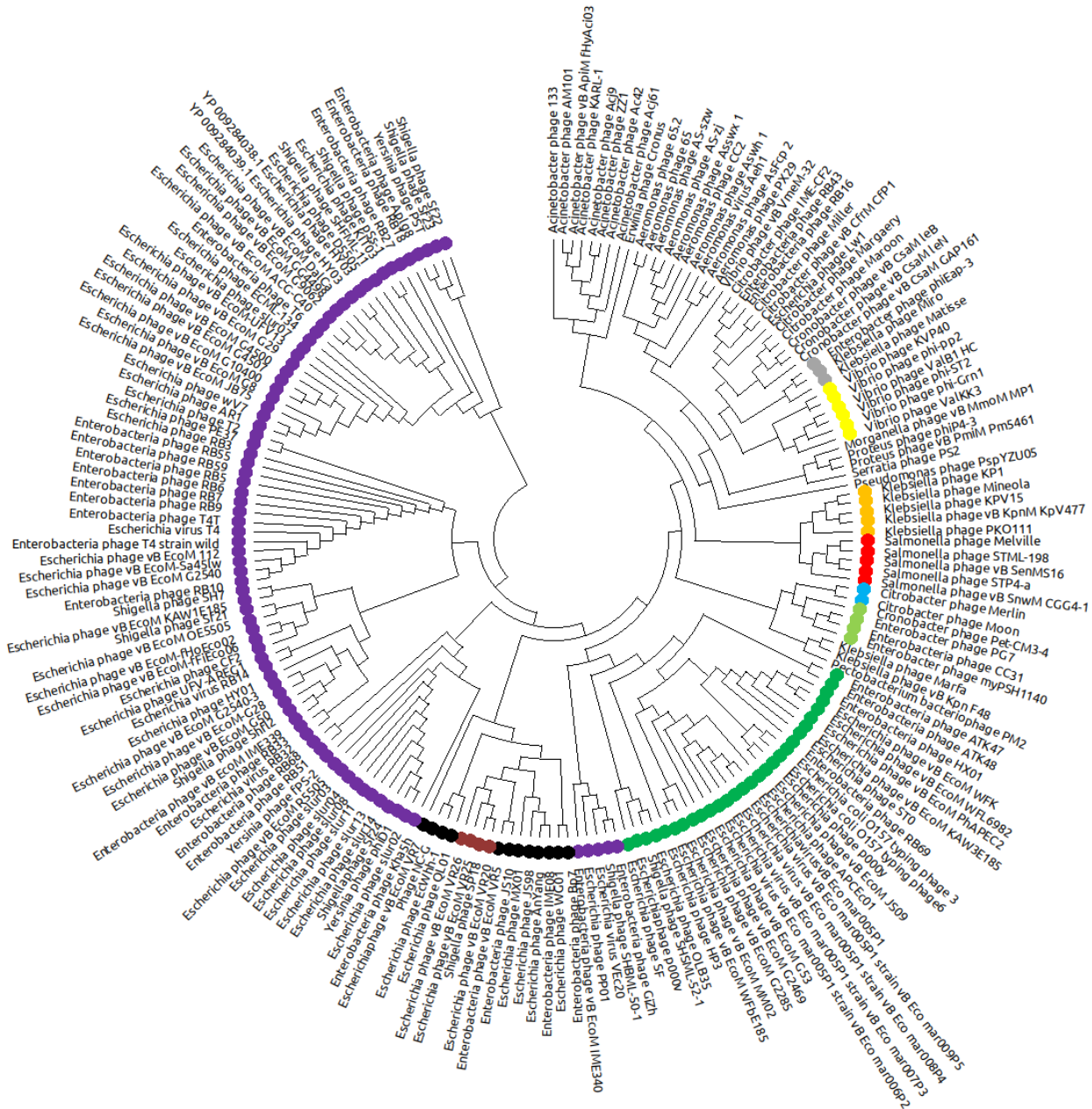


Рис. 1. Филогенетическое дерево, построенное для 189 РНК-лигаз 2 подсемейства Т-четных бактериофагов.

На дереве, представленном на рисунке 1, концевые вершины, соответствующие классифицированным последовательностям подсемейства, отмечены кружками, которые покрашены в цвета, соответствующие их роду. На рисунке видно, что концевые вершины одного цвета на дереве лежат рядом.

Проведенный анализ показал, что РНК-лигазы 2 представляют собой белки с консервативной первичной структурой, а также то, что аминокислотная последовательность РНК-лигазы 2 бактериофага Т4 является представительным



образцом исследуемого фермента, позволяющим не только достоверно находить ортологи в других геномах, но и использовать его для классификации.

### **Подготовка набора белков-гомологов RnlB бактериофага T4 из метагеномов океанических проб и геномов филогенетически далеких живых существей**

На первом этапе подготовки набора гомологов из метагеномов океанической микробиоты использовался алгоритм PSI-BLAST [25] с уровнем достоверности результатов  $E\text{-value} < 3e-21$  для поиска гомологов РНК-лигазы 2 бактериофага T4, продукта гена rnlB (идентификатор GenBank ADJ39898.1), в базах данных белковых последовательностей океанических метагеномов на сервере NCBI. Для получения набора аминокислотных последовательностей гомологов РНК-лигазы 2 бактериофага T4 итерационный поиск сходных последовательностей производился 5 раз, пока каждая последующая итерация обнаруживала новые локальные сходства. Было обнаружено 15 гомологичных последовательностей в метагеномах GOS и 21 в LCGC14.

На втором этапе данные были подвергнуты дополнительной фильтрации. Из набора были удалены последовательности, которые отличались по длине от РНК-лигазы 2 бактериофага T4 более, чем на 20 %. Это связано, в первую очередь, с тем, что до 70 % общего числа последовательностей в метагеноме любой пробы составляют короткие фрагменты последовательностей. Статистические ограничения, имеющиеся у программ любого пакета филогенетического анализа, не позволяют работать с короткими последовательностями. Длина РНК-лигазы 2 бактериофага T4 составляет 334 аминокислоты. В результате проведения редактирования в наборе данных осталась 21 белковая последовательность с диапазоном длин от 271 до 389 аминокислот.

Данная ситуация вполне адекватна для генетических наборов, полученных средствами метагеномики. В ряде других случаев найденные аминокислотные последовательности кодировались на концах фрагментов – контигов того или иного метагенома. Процесс подготовки пробы для метагеномного секвенирования морской микробиоты включает в себя несколько стадий фильтрации воды или экстракта морского бентоса. Это может сказываться на интактности живого материала. Уже на этой стадии происходит разрыв длинных геномных ДНК в самых различных местах. Дальнейшая обработка выделенной и очищенной от других веществ ДНК ультразвуком вносит массу дополнительных разрывов. В результате секвенирования большого разнообразия геномов трудно находятя протяженные перекрытия для прочтений, что часто приводит к появлению последовательностей фрагментов генов. Наиболее существенным моментом является большое разнообразие живых существей в океане, поэтому сложно безошибочно определять в прочтениях все последовательности из того или иного организма.

Поскольку перед нами стояла задача определения места найденных в океанических метагеномах последовательностей на филогенетическом дереве, построенном с использованием аминокислотной последовательности РНК-лигазы 2, встала задача создания максимально представительной таксономической выборки последовательностей гомологов РНК-лигазы 2. фага T4, имеющих в базах данных. Многочисленная представленность последовательностей RnlB у бактериофагов и у других вирусов, имеющих между собой высокий уровень гомологии, скрывает вывод результатов с меньшей схожестью из других таксонов, особенно простейших, при поиске сразу по всем организмам. Поэтому был проведен поиск по отдельным таксонам.

Можно было ожидать нахождение гомологов в таксонах простейших, но при этом с меньшим значением уровня достоверности. Для определения таксономической принадлежности находок из морских метагеномов было важно не упустить ни одной контрольной последовательности, что и было сделано путем кропотливого анализа,

проведенного по крупным таксонам одноклеточных и многоклеточных живых организмов. Мы провели исследование следующих таксонов: *Cercozoa*, *Ciliophora*, *Euglenozoa*, *Amoebozoa*, *Homidia*, *Crustacea*, *Alveolata*, *Haplosporea*, *Acantharia*, *Parabasalia*, *Mycetozoa*, *Dinophyta*, *Fungi*, *Placozoa*, *Viridiplantae*, *Rhodophyta*, *Rhizaria*, *Stramenopiles* и *Porifera*, пытаясь охватить все разнообразие генетически исследованных клеточных форм жизни на земле. Здесь и далее использована классификация живых существ Т. Кавалье-Смита [29]. Под названием *Giant viruses* объединены гигантские вирусы эукариот.

То, что мы приняли во внимание не только одноклеточных представителей прокариотической и эукариотической микробиоты, как моря, так и суши, связано с тем, что пелагические воды содержат многочисленные гаметы самых разнообразных многоклеточных эукариот. Концентрация гамет в планктоне может достигать  $10^7$  и более частиц на 1 мл. Осадочные породы при своем образовании захватывают эту существеннейшую часть планктона. Большинство гамет имеет достаточно большие размеры, но продукты их деструкции могут содержать ДНК самых разнообразных представителей морской биоты.

Названия таксонов даны согласно таксономии, принятой на данный момент Национальным центром биотехнологических исследований (NCBI, NIH, USA) [23]. Во второй колонке приведено число хитов, находящихся в достоверной области сравнения, согласно заданным параметрам сравнения ( $E\text{-value} < 0,001$ ). В третьей колонке приведено значение  $E\text{-value}_{\min}$  для «наилучшего» результата из недостоверной области сравнения, если таковой имелся. В четвертой колонке приведены «наихудшие» значения  $E\text{-value}_{\max}$  среди результатов из достоверной области сравнения. Другие данные о результатах в статистически недостоверной области сравнения не приводятся.

**Таблица 2.** Представленность гомологов РНК-лигазы 2 среди белковых последовательностей биологических существ различных таксонов

Название таксона	Число хитов в достоверной области	$E\text{-value}_{\min}$	$E\text{-value}_{\max}$
<i>Cercozoa</i>	0	0.48	
<i>Ciliophora</i>	1		2e-22
<i>Euglenozoa</i>	42	7e-11	
<i>Amoebozoa</i>	4		1e-04
<i>Homidia</i>	0		
<i>Crustacea</i>	0		
<i>Alveolata</i>	2		9e-10
<i>Haplosporea</i>	0	0.47	
<i>Acantharia</i>	0	3.3	
<i>Parabasalia</i>	0	4.8	
<i>Mycetozoa</i>	0	0.78	
<i>Fungi</i>	13		1e-05
<i>Placozoa</i>	0	0.25	
<i>Viridiplantae</i>	1		8e-38
<i>Rhodophyta</i>	0	2.2	
<i>Rhizaria</i>	2		0.003
<i>Stramenopiles</i>	2		0.002
<i>Porifera</i>	1		1e-07
<i>Giant viruses</i>	17	4e-9	

Результаты поиска контрольных аминокислотных последовательностей представлены в таблице 2. На фоне 42 достоверных результатов у *Euglenozoa*, небольшое число гомологичных последовательностей найдено у *Ciliophora*, *Amoebozoa*,

*Alveolata, Dinophyta*. 13 гомологов РНК лигазы 2 фага T4 были обнаружены у грибов в статистически достоверной области сравнения. У многоклеточных также достоверно найдены гомологи: по одному у растений и губок и по два у *Rhizaria* и *Stramenopiles*. Некоторые результаты, скорее всего, не являются последовательностями РНК лигаз, но и найденные в морских метагеномах, при использованных параметрах сравнения последовательности, также могут быть сходными, но не гомологичными. Например, для *Rhodophyta* наилучшим статистически недостоверным результатом был E3 убиквитин – белок лигаза HERC3. Значение *E-value* для этой последовательности составляло 2.2, покрытие 17 %, а процент идентичных аминокислот в парном выравнивании 31.67 %. У *Haplosporea* это был актин *Bonamia ostreae* при покрытии 7 %, *E-value* 0.47, идентичности 38.46 %. Интересно, что у достаточно близкого к этому классу простейших *Acantharia*, с которым ранее объединяли таксономическую группу *Haplosporea*, наилучшим недостоверным результатом была находка сократимого белка, тубулина *Astrolonche serrata* с низким покрытием в 10 %, плохим значением *E-value* 3.3, и очень высоким процентом идентичных аминокислот 40.91 % в парном выравнивании. Низкое покрытие этих двух белков может быть объяснено их большой длиной. Более того, нельзя исключить ошибок секвенирования и искусственного слияния белков. Нам удалось найти гомологи и для этих групп одноклеточных простейших. Можно ожидать, что в недалеком будущем появятся больше генетической информации об этих группах и можно будет провести более достоверный поиск гомологов РНК-лигазы 2 фага T4. У *Glaucophyta* был найден фитохром *Glaucocystis nostochinearum* (покрытие 26 %, *E-value* 0.23, идентичность 29.29 %). Для ряда таксономических групп это были гипотетические белки, например, G3 у *Trichomonas vaginalis* у простейших типа *Parabasalia*, с покрытием 39 %, недостоверным *E-value* 4.8, достаточно высокой идентичностью 23.24 %. У *Mycetozoa* был найден гипотетический белок *Heterostelium album* PN500 с недостоверными, но весьма интересными статистическими параметрами: 18 %, 0.78 и 30.77 %. Подобная ситуация была и с белком *Trichoplax adhaerens Placozoa*: 32 %, 0.25 и 26.27 %. Можно ожидать, что процесс редукции РНК путем вставки или делеции уридина достаточно широко распространен у одноклеточных эукариот и минорно встречается у простейших многоклеточных в качестве биохимического «атавизма». Для филогенетического анализа были отобраны все последовательности из тех таксонов, где результатов было менее 10. В случае, если их было более 10, отбирались последовательности с разными уровнями сходства в соответствии с найденными значениями *E-value*, равномерно распределенные в списке хитов. Также были взяты последовательности из 17 бактериофагов подсемейства *Tevenvirinae* для того, чтобы была видна связь наиболее близких последовательностей в таксономическом сравнительном анализе последовательностей из метагеномов океанической микробиоты и последовательностей из различных фил и таксонов клеточных форм жизни. С помощью этого подхода мы старались обеспечить максимально представительную выборку различных последовательностей из всех таксонов для проведения филогенетического анализа и построения дерева. Только единственная сходная последовательность представителя таксона *Viridiplantae* была исследована отдельно.

Контрольные последовательности и последовательности океанических гомологов РНК-лигазы 2 были объединены в один набор, который приведен в таблице 3. Для анализа использовалось 122 гомолога, представленных в GenBank (Release 234, 05.10.2019), включая саму РНК-лигазу бактериофага T4, а также 21 гомолог из океанических проб. Для построения множественного выравнивания последовательностей использовался алгоритм MUSCLE [30]. На основе полученного выравнивания была выведена эволюционная история, было построено филогенетическое дерево с помощью метода максимального правдоподобия (Maximum

Likelihood Method) с использованием модели матриц замен JTT [26]. Консенсусное дерево было построено после 3000 повторов статистического bootstrap-анализа [27] предпринятых для выяснения эволюции анализируемых таксонов. Ветви, образовавшиеся менее, чем в 50% повторах статистического анализа, были убраны. Исходное дерево было получено методом максимальной экономии (Maximum Parsimony Method). Все позиции, содержащие бреши или места, где данные отсутствовали, были убраны. Исследование проводили с помощью пакета программ MegaX [28].

**Таблица. 3.** Гомологи РНК-лигазы 2 бактериофага T4 из геномов живых существ разных таксонов, которые были использованы для филогенетического анализа и определения места гомологов из океанических метаженомов на древе жизни

№	Таксон	Название белка	Идентификатор GenBank
1	▲ <i>Bacteria</i>	Cand Thiosymbion oneisti	WP_089724394.1
2		Chryseolinea serpens	SHH97071.1
3		Bacteroidetes bacterium 4484_276	OQX79431.1
4		Chryseolinea serpens_II	WP_073143442.1
5		Desulfamplus magnetovallimortis	SLM31278.1
6		Cand Gottesmanbacteria bacterium	KKW10401.1
7		Cand Vecturithrix granuli	GAK56201.1
8		Cand Propionivibrio aalborgensis	SBT10709.1
9		Cand Contendobacter odensis	WP_051497828.1
10		Parabacteroides distasonis	WP_081033086.1
11		Porphyromonas sp. 31_2	KEJ87078.1
12		Cand Contendobacter odensis Run_B_J11	CDH45942.1
13		Bacteroides sp. CAG:633	CDB10874.1
13	● <i>Stramenopiles</i>	FNF27_00344 Cafeteria roenbergensis	CAA0178495.1
15		FNF31_06539 Cafeteria roenbergensis	CAA0153076.1
16		FNF29_01479 Cafeteria roenbergensis	CAA0156063.1
17		Thraustotheca clavata	OQR97834.1
18		57867_21281 Aphanomyces stellatus	VFT97953.1
19	● <i>Rhizaria</i>	RNA ed ligase 2 Reticulomyxa filosa	ETO29077.1
20		RNA ed lig 1 Reticulomyxa filosa	ETO09630.1
21	● <i>Alveolata</i>	PPERSA_08025_1 Pseudocohnilembus persalinus	KRX08714.1
22		RNA ed ligase 2 mt Symbiodinium microadriaticum	OLP91488.1
23	● <i>Amoebozoa</i>	RNA lig put Acanthamoeba castellanii Neff	XP_004341081.1
24		RNA ligase Acanthamoeba castellanii Neff	XP_004333230.1
25		RNA ed lig Acanthamoeba castellanii Neff	XP_004349586.1
26		EIN_146330 Entamoeba invadens IP1	XP_004254400.1
27	▼ <i>Archaea</i>	DRN27_08370 Thermoplasmata archaeon	RLF57127.1
28		hp Thermoplasmata archaeon	MAH42159.1
29		DRN58_06670 Thermococci archaeon	RLF98605.1
30		DRN26_00060 Thermoplasmata archaeon	RLF68204.1
31	● <i>Ciliophora</i>	PPERSA_08025 Pseudocohnilembus persalinus	KRX08714.1
32	● <i>Euglenozoa</i>	mt RNA lig2 Leishmania mexicana MHOM	XP_003875227.1
33		mt RNA lig2 Leishmania major str. Friedlin	XP_001682919.1
34		mt RNA lig2 Leishmania donovani	XP_003860510.1
35		mt RNA lig2 Leishmania tarentolae	AAN77726.1
36		mt RNA lig2 Leishmania infantum JPCM5	XP_001465293.1
37		mt RNA lig2 Leishmania braziliensis MHOM	XP_003723137.1
38		mt RNA lig2 Leishmania panamensis	XP_010698781.1
39		mt RNA lig2 Leishmania guyanensis	CCM15306.1
40		mt RNA lig2 Trypanosoma congolense IL3000	CCC92745.1
41		mt RNA lig2 Leptomonas pyrrocoris	XP_015653633.1



42		RNA ed lig Angomonas deanei	EPY40852.1
43		mt RNA lig1 Leishmania major strain Friedlin	XP_003721581.1
44		RNA lig2 Trypanosoma theileri	XP_028882973.1
45		mt RNA lig1 Leishmania major	AAR10824.1
46		RNA ed lig Trypanosoma vivax Y486	CCC46487.1
47		hp Phytomonas sp. isolate EM1	CCW64502.1
48		RNA ed lig Trypanosoma congolense IL3000	CCC89377.1
49		mt RNA lig2 Trypanosoma grayi	XP_009308197.1
50		mt RNA lig Trypanosoma brucei gambiense DAL972	XP_011776426.1
51		Euglenozoa RNA ed lig Trypanosoma conorhini	XP_029226606.1
52		mt RNA lig1 Trypanosoma conorhini	XP_029226428.1
53		RNA ed lig Trypanosoma rangeli	XP_029240616.1
54		RNA ed lig Trypanosoma vivax Y486	CCC50326.1
55		mt RNA lig1 Trypanosoma rangeli SC58	ESL07901.1
56		mt RNA lig1 Trypanosoma rangeli	XP_029237953.1
57		KREL2 Trypanosoma equiperdum	SCU71059.1
58		mt RNA lig1 Trypanosoma theileri	XP_028885257.1
59		mt RNA lig1 Leishmania panamensis	XP_010696701.1
60		hp Phytomonas sp. isolate EM1	CCW64180.1
61		mt KREL1 Leptomonas pyrrocoris	XP_015652175.1
62		RNA-editing ligase Angomonas deanei	EPY33068.1
63		mt RNA lig1 Trypanosoma grayi	XP_009311548.1
64		mt RNA lig1 Leptomonas seymouri	KPI83474.1
65		RNA ed lig Bodo saltans	CUE61637.1
66		REL1 Trypanosoma cruzi	RNF20713.1
67		RNA ed lig 1 Leptomonas seymouri	KPI86142.1
68		mt RNA lig1 Trypanosoma cruzi strain CL Brener	XP_820361.1
69		hp Phytomonas sp. isolate Hart1	CCW66362.1
70		RNA ed lig Bodo saltans	CUG93424.1
71		RNA ed lig Perkinsela sp. CCAP 1560/4	
72		RNA ed lig2 Perkinsela sp. CCAP 1560/4	KNH03941.1
73		RNA ed lig Angomonas deanei	EPY29391.1
74		BC937DRAFT_89812 Endogone sp. FLAS-F59071	RUS17558.1
75		C2G38_128065 Gigaspora rosea	RIB11192.1
76		C1645_817038 Glomus cerebriforme	RIA95193.1
77		C1645_795269 Glomus cerebriforme	RIA78863.1
78		RclHR1_02510016 Rhizophagus clarus	GBB95322.1
79		C2G38_1022383 Gigaspora rosea	RIB01290.1
80	● Fungi	BC938DRAFT_476924 Jimgerdemannia flammicorona	RUS31849.1
81		BC936DRAFT_147411 Jimgerdemannia flammicorona	RUP46053.1
82		DNA lig/mRNA capp Rhizophagus irregularis	PKK73150.1
83		GLOIN_2v1881488 Rhizophagus irregularis DAOM	XP_025171120.1
84		C1646_745321 Rhizophagus diaphanus	RGB29488.1
85		BC938DRAFT_470766 Jimgerdemannia flammicorona	RUS26442.1
86		CcCBS67573_g06065 Chytriomycetes confervae	TPX71677.1
87	● Porifera	RNA ed lig 2 mtl Amphimedon queenslandica	XP_011410182.1
88		RnlB Escherichia virus T4	2HVR_A Chain A
89		RnlB Staphylococcus phage	ATN93943.1
90		RnlB Escherichia phage ECML-134	YP_009102648.1
91		RnlB Escherichia phage slur14	YP_009180676.1
92		RnlB Enterobacteria phage ATK47	ANZ51023.1
93		RnlB Escherichia phage APCEc01	YP_009225077.1
94	◆ Viruses	RnlB Enterobacteria phage JS10	YP_002922522.1
95		RnlB Edwardsiella phage PEi26	BAQ23148.1
96		RnlB Citrobacter phage Merlin	YP_009203914.1
97		RnlB Klebsiella phage Marfa	QDB71837.1
98		RnlB Erwinia phage Cronus	AWD90329.1
99		RnlB Citrobacter phage Miller	YP_009097842.1
100		RnlB Aeromonas phage Ah1	AUE22774.1

101		RnlB Vibriophage phi-pp2	AFN37364.1
102		RNA ligase 2 Vibrio phage CHOED	YP_009021711.1
103		RNA ligase 2 Acinetobacter phage TAC1	AZF88522.1
104		RNA ligase 2 Streptomyces phage Annadreamy	AXG66278.1
105		RNA ligase 2 Tupanvirus deep ocean	AUL79719.1
106		RNA ligase 2 Pithovirus LCPAC401	QBK92682.1
107		RNA ligase 2 Mimivirus sp. SH	AZL89535.1
108		RNA ligase 2 Powai lake megavirus	ANB50426.1
109		RNA ligase 2 Megavirus vitis	AVL93643.1
110		RNA ligase 2 Megavirus chiliensis	YP_004894359.1
111		RNA ligase 2 Bandra megavirus	AUV58249.1
112		RNA ligase 2 Saudi moumouvirus	AQN68128.1
113		RNA ligase 2 Marseillevirus LCMAC101	QBK85972.1
114		RNA ligase 2 Acanthamoeba polyphaga moumouvirus	YP_007354215.1
115		RNA ligase 2 Moumouvirus Monve	AEX63037.1
116		RNA ligase 2 Moumouvirus australiensis	AVL94635.1
117		RNA ligase 2 Harvfovirus sp.	AYV80481.1
118		RNA ligase 2 Moumouvirus goulette	AGF85515.1
119		RNA ligase 2 Tupanvirus2 deep ocean	AUL79055.1
120		RNA ligase 2 Tupanvirus soda lake	AUL77767.1
121		RNA ligase 2 Powai lake megavirus	ANB50967.1
122		RNA ligase 2 Acanthamoeba polyphaga mimivirus	AVG46618.1
123	● Гомологи из океанической планктонической микробиоты	GOS_3224571	ECD49735.1
124		GOS_6108359	ECF47459.1
125		GOS_8459675	EBM12135.1
126		GOS_1057497	EDE85953.1
127		GOS_4814015	EBY11901.1
128		GAG86666.1	GAG86666.1
129	● Гомологи из микробиоты океанических осадочных отложений	GAJ13060.1MSM	GAJ13060.1
130		RNA ligaseMSM	EFK97083.1
131		LCGC14_1419600MSM	KKM72535.1
132		LCGC14_1490430MSM	KKM65524.1
133		LCGC14_0465520MSM	KKN66998.1
134		LCGC14_0236200MSM	KKN89774.1
135		LCGC14_1657970MSM	KKM19213.1
136		LCGC14_2451880MSM	KKL20798.1
137		LCGC14_2597070MSM	KKL06334.1
138		LCGC14_2115790MSM	KKL69358.1
139		LCGC14_0535600MSM	KKN60091.1
140		LCGC14_3030500MSM	KKK59824.1
141		LCGC14_2458620	KKL20123.1
142		LCGC14_1232570MSM	KKM91033.1
143		LCGC14_1282820MSM	KKM86052.1

## РЕЗУЛЬТАТЫ

### Филогенетический анализ РНК-лигаз 2 из метагеномов морской микробиоты и живых существ различных таксонов

Представительный набор гомологов РНК-лигазы 2 из различных таксонов живых существ, геномные данные которых имеются в распоряжении исследователей на сегодняшний день, был использован для построения филогенетического дерева, представленного на рисунке 2.



Рис. 2. Молекулярно-филогенетический анализ гомологов РНК-лигазы 2 бактериофага T4 из разных таксонов живых существей.

Филогенетическое дерево гомологов РНК-лигазы 2, представленное на рисунке 2, образует две ветви, причем на верхней ветви находятся белки вирусов, включая фаги, подобные T4, высокоомологичные им РНК-лигазы трипаносом, белки грибов, все гомологи из *Rhizaria*, *Porifera*, две архейные последовательности, и две последовательности из амёбы. На этой ветви находится лишь одна последовательность планктонического гомолога, три глубоководные последовательности и одна из других водных осадочных пород. большинство последовательностей океанических гомологов представлены в другой основной ветви данного дерева.

**Таблица 3.** Последовательности, филогенетически наиболее близкородственные морским гомологам

Название последовательности из океанического метагенома	Филогенетически наиболее близкие последовательности	
	1*	2**
LCGC14 1419600MSM	Fungi C1645 795269 <i>Glomus cerebriforme</i>	Fungi C1646 745321 <i>Rhizophagus diaphanus</i>
RNA ligaseMSM	Archaea DRN58 06670 <i>Thermococci archaeon</i>	Fungi DNA lig/mRNA capp <i>Rhizophagus irregularis</i>
GOS 3224571	Archaea DRN27 08370 <i>Thermoplasmata archaeon</i>	Viruses RnlB <i>Aeromonas phage Ah1</i>
LCGC14 3030500MSM	Amoebozoa RNA ed lig <i>Acanthamoeba castellanii</i> Neff	Viruses RNA ligase 2 <i>Harvfovirus sp.</i>
LCGC14 2458620	Stramenopiles <i>Thraustotheca clavata</i>	Viruses RNA ligase 2 <i>Acinetobacter phage TAC1</i>
LCGC14 1232570MSM	Stramenopiles <i>Thraustotheca clavata</i>	Viruses RNA ligase 2 <i>Acinetobacter phage TAC1</i>
LCGC14 2597070MSM	Amoebozoa EIN 146330 <i>Entamoeba invadens</i> IP1	Fungi BC938DRAFT 470766 <i>Jingerdemannia flammicorona</i>
LCGC14 0535600MSM	Bacteria Cand <i>Contendobacter odensis</i>	-**
GOS 1057497	Bacteria Cand <i>Vecturithrix granuli</i>	-
LCGC14 0236200MSM LCGC14 2115790MSM	Bacteria Cand <i>Vecturithrix granuli</i>	-
LCGC14 1657970MSM	Bacteria Cand <i>Gottesmanbacteria bacterium</i>	-
GOS 8459675	Archaea DRN26 00060 <i>Thermoplasmata archaeon</i>	Bacteria <i>Chryseolinea serpens</i>
LCGC14 0465520MSM	Bacteria <i>Chryseolinea serpens</i>	Archaea DRN26 00060 <i>Thermoplasmata archaeon</i>
LCGC14 1282820MSM	Bacteria <i>Chryseolinea serpens</i>	Archaea DRN26 00060 <i>Thermoplasmata archaeon</i>
GAJ13060.1MSM	Rhizaria RNA ed ligase 2 <i>Reticulomyxa filosa</i>	Viruses RNA ligase 2 <i>Acanthamoeba polyphaga mimivirus</i>
GAG866666.1	Viruses RNA ligase 2 <i>Tupanvirus2 deep ocean</i>	Viruses RNA ligase 2 <i>Tupanvirus soda lake</i>
GOS 6108359 LCGC14 1490430MSM	Stramenopiles 57867 21281 <i>Aphanomyces stellatus</i>	-
LCGC14 2451880MSM GOS 4814015	Bacteria <i>Bacteroidetes bacterium</i> 4484 276	Alveolata RNA ed ligase 2 mt <i>Symbiodinium microadriaticum</i>

\* - наиболее близкая последовательность на филогенетическом дереве;

\*\* - чуть более удаленная на дереве, но также близкая последовательность.

Основными последовательностями, представленными на нижней ветви, являются белки бактерий, ряда бактериофагов и мегавирусов, а также архей, грибов и некоторых других эукариот. На этой ветви находятся и все 14 последовательностей из



метагеномов дна океана. Их распределение на этой ветви неравномерное, нет объединений в крупные общие клады. Филогенетически наиболее близкородственные представители морских гомологов РНК-лигазы бактериофага T4 представлены в таблице 3.

Нижняя подветвь дерева, которая содержит три последовательности гигантских вирусов, одну бактериальную последовательность и два гомолога из глубоководных океанических осадков. Один из представленных в этой кладе вирусов *Tupanovirus* также выделен с глубины океана более 3000 м, как и пробы для глубоководного метагеномного анализа. Надо заметить, что в составе этой ветви планктонические и глубоководные гомологи РНК-лигазы 2 часто кластеризуются совместно на одних и тех же ветвях.

Из таблицы 3 видно, что большинство найденных последовательностей проявляет филогенетическое родство с последовательностями из бактерий, мегавирусов, грибов и архей. Во многих случаях наиболее близкородственные последовательности принадлежали бактериям, известным только по данным океанических метагеномов и определение которых в настоящему моменту времени еще не завершено. Две последовательности оказались родственны простейшим из таксона *Amoebozoa*, и по одной *Rhizaria* и *Stramenopiles*. Сравнительный анализ путем построения парных выравниваний (данные не приводятся) показал, что найденное сходство говорит лишь о родстве данных последовательностей, а не о их совпадении. Таким образом найденные гомологи относятся к новым видам организмов из данных групп.

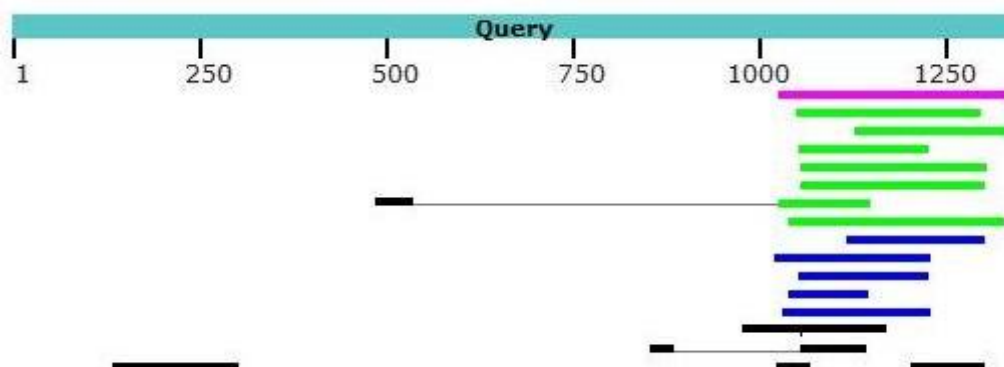
Неподалеку от РНК-лигазы 2 трипаносом обнаруживаются последовательности губки большого барьерного рифа *Amphimedon queenslandica* (*Porifera*) и одноклеточной пресноводной фораминиферы *Reticulomyxa filosa* (*Rhizaria*) относящейся к моноспецифическому роду. Как в первом, так и втором случае, наиболее вероятным видится происхождение этого гена в данных геномах за счет горизонтального переноса генов. С другой стороны, у одноклеточных животных, которых на данный момент времени относят к *Rhizaria*, за счет их малой изученности нельзя исключать изменений в классификации и возможного общего происхождения митохондриальной ДНК с митохондриальной ДНК *Euglenozoa*.

Среди *Euglenozoa* гомологи РНК-лигазы 2 встречаются только в митохондриях трипаносом (рис. 2.) и разделены на две родственные ветви. В верхней части дерева чаще представлены РНК-лигазы 2 первого типа в нижней чаще второго типа. Оба фермента близкородственны, и, вероятно, могут эволюционировать, заменяя друг друга функционально. Один из этих ферментов входит в комплекс белков по редакции первичных транскриптов митохондрий трипаносом путем делеции уридина, другой путем вставки этого нуклеотида. Функции обеих РНК-лигаз в обоих случаях одинаковы – лигирование одноцепочечного разрыва в РНК, остающегося после работы других ферментов.

Вирусные гомологи РНК-лигазы 2 встречаются в различных ветвях дерева. Это связано, вероятно, с несколькими причинами. Во-первых, эти ферменты могли исходно появиться именно у вирусов и эволюционно разойтись внутри разнообразных вирусных геномов, в дальнейшем они могли быть переданы различным клеточным формам жизни независимо. Другой причиной может быть использование природой в своих экспериментах вирусов в качестве векторов горизонтального переноса генов. Этот процесс может приводить к вторичному появлению этих генов у вирусов, имея происхождение из генома той или иной клеточной формы жизни.

Последовательность LCGC14 1419600MSM располагается среди грибных гомологов и примыкающих к ним ветвями последовательностей белков гигантских эукариотических вирусов.

Рядом с этой ветвью располагается последовательность RNA ligaseMSM из морских осадков, близко примыкающая к гомологу из археи *Thermococci archaeon*. Эта ветвь также находится в ветви грибных гомологов. GOS 3224571 является единственной последовательностью, располагающейся в ветви гомологов из бактериофагов Т4-типа, хотя непосредственно рядом с ней находится гомолог из царства архей - *Thermoplasmata archaeon*.



**Рис. 3.** Графическое представление множественного выравнивания белка RnlB из *Vigna angularis*, РНК-лигазы 2 фага Т4 (сиреневая линия) и океанических белков (зеленые, синие и черные линии) изображают по убыванию различные уровни достоверности выравнивания E-value, значения приведены в табл. 3). Порядок следования океанических белков на рисунке соответствует порядку следования в табл. 3).

**Таблица 4.** Результат выравнивания белка RnlB из *Viridiplantae* (*Vigna angularis*, 1338 а.к.) и фага Т4 с океаническими гомологами (обозначения те же, что и на рис. 2)

№	Название последовательности	Процент длины покрытия	E-value
	<i>Vigna angularis</i> 1338 а.к.	100%	0
	Enterobacteria phage T4	23%	3e-43
1	LCGC14_1419600MSM	18%	7e-14
2	GAJ13060.1MSM	16%	3e-13
3	LCGC14_0465520MSM	13%	1e-11
4	LCGC14_0236200MSM	18%	2e-11
5	LCGC14_2115790MSM	18%	9e-11
6	GOS_3224571	12%	5e-10
7	LCGC14_2597070MSM	21%	8e-10
8	GAG86666.1	13%	2e-09
9	LCGC14_1657970MSM	15%	3e-09
10	GOS_8459675	13%	8e-08
11	GOS_6108359	7%	1e-07
12	LCGC14_0535600MSM	14%	5e-07
13	RNA ligaseMSM	14%	1e-04
14	LCGC14_2451880MSM	8%	0.005
15	LCGC14_1490430MSM	23%	0.13

В царстве зеленых растений (*Viridiplantae*) гомолог РНК-лигазы 2 фага Т4 длиной 1338 а.к. был найден у фасоли угловатой, или адзуки. Как показало множественное выравнивание (рис. 3), сделанное программой ClustalX [31] все гомологичные фрагменты сконцентрировались рядом с С-концом этого белка. Распределение выборки океанических последовательностей по сходству с С-концом дано в таблице 4.

Океанические гомологи LCGC14\_1419600MSM, GAJ13060.1MSM, LCGC14\_0465520MSM, LCGC14\_0236200MSM, LCGC14\_2115790MSM, GOS\_3224571, LCGC14\_2597070MSM проявляют большое достоверное сходство с фрагментом этого

белка фасоли. Наличие гомолога РНК-лигазы 2 фага T4 только у одного зеленого растения и отсутствие подобных находок у каких-либо в той или иной степени сходных видов, а также большая разница в длинах полных аминокислотных последовательностей данных белков не позволили включить данный контроль в рассмотрение в пакете программ MegaX, и мы представили данный результат отдельно.

### ОБСУЖДЕНИЕ

Научная группа Крега Вентера провела поиск вирусных последовательностей в микробных метагеномах путем сравнения GOS-последовательностей с 27 полными геномами вирусов [18]. Таксономический анализ проводился следующим методом. Были взяты четыре лучшие находки из каждого поиска со значением  $E\text{-value} < 1e-10$ . Эти данные были объединены в общий пул результатов. Анализ метагеномов проб морской воды, собранных в рамках Глобальной океанической экспедиции Sorcerer II (GOS), выявил большое количество вирусных последовательностей, составляющих приблизительно 3% от общего количества предсказанных белков. Филогенетический анализ этих вирусных последовательностей, выявил большое число генов бактерий, сцепленных с вирусными последовательностями. Авторы исследования предположили, что приобретение вирусом экологически значимых генов бактерий-хозяев является более распространенным явлением, чем предполагалось ранее. Было показано, что в микробных фракциях преобладали вирусные последовательности, которые происходят из геномов хвостатых бактериофагов. Глобальное распределение последовательностей в зависимости от семейства вирусов было различным при исследовании проб воды, отобранных из разных мест. Сравнение найденных в таком анализе фрагментов вирусных геномов из метагенома GOS с 27 полными геномами вирусов, выделенных из воды, выявило, что с высокой частотой был представлен в данных пробах только один эталонный геном бактериофага. Это был геном цианомиовируса P-SSM4. Это доминирование совпадало с доминированием в пробах его хозяина, *Prochlorococcus marinus*. Найденные новые последовательности не имели полного совпадения с последовательностью генома ранее секвенированного фага P-SSM4, а были лишь родственны ему. На основе этих наблюдений исследователи сделали вывод, что данный бактериофаг, а также родственные фаги, могут оказывать существенное влияние на численность, распределение и разнообразие *Prochlorococcus marinus* – доминирующего компонента пикофитопланктона в олиготрофных водах океана [18].

Хотя вирусы бактерий, как правило, намного меньше по размеру, чем их бактерио-хозяева, существует ряд причин, по которым вирусные последовательности могут быть обнаружены в микробной фракции морской воды. Во-первых, вирусы, размер частиц которых превышает определенный размер фильтра (как правило, от 0,1 до 0,22 мкм), автоматически сохраняются из-за их геометрии; и такие вирусы были обнаружены, во все большем количестве, благодаря усилиям по отбору проб в океане [18]. Однако, несмотря на присутствие вирусов, способных инфицировать различные группы эукариотических хозяев, подавляющее большинство вирусов в морских экосистемах представляют собой бактериофаги, размер которых обычно составляет менее 0,2 мкм [18,19].

Группа К. Вентера проводила фильтрование через фильтр с диаметром проб 0,1 мкм. Длина частицы бактериофага P-SSM4, относящегося к группе T-четных (*Teevenvirinae*) составляет от 110 до 130 нм и более. Эта группа бактериофагов может эффективно задерживаться на 0.1 мкм мембранных фильтрах (Pall Life Sciences, East Hills, NY), которые использовались авторами. То, что нам удалось обнаружить лишь одного гомолога из океана, близкого к бактериофагам этой группы, говорит о том, что в обеих базах данных представлены бактериофаговые геномы, в том числе и родственные T4, но не содержащие данный ген, как несущественный для бактериофага T4 и его

родственников [2], либо данная аллель у морских бактериофагов сильно различается по последовательности аминокислот. Нами было найдено определенное число гомологов таксономически сходных с ферментами гигантских фагов, которые могли хорошо задерживаться при нескольких видах фильтрования через мембранные фильтры с порами 3.0 мкм, 0.8 мкм, и 0.1 мкм [18].

Интересными находками является обнаружение гомологов РНК-лигазы 2 у низших беспозвоночных животных, а именно у представителей таксонов *Alveolata*, *Stramenopiles*, *Rhizaria*, *Ciliophora*. У этих групп животных гомологи фермента RnlB обнаруживались весьма редко и, возможно, появились только за счет горизонтального переноса генов. Найденные аминокислотные последовательности гомологов РНК-лигазы 2 могут послужить для дальнейшего структурного и функционального исследования этих ферментов.

Большинство гомологов, сходных с белками архей и бактерий, относится к малоизученным группам видов, в том числе выделенным из океана. Это значит, что найденные последовательности являются новыми, и получение продуцентов этих ферментов может существенно обогатить исследования в области энзимологии бактерий и архей.

Было проведено определение таксономической принадлежности обнаруженных программой PSI-BLAST гомологов РНК-лигазы 2 бактериофага T4 в океанических метагеномах (6 аминокислотных последовательностей из базы данных GOS и 15 – из базы данных LCGC14). Было показано, что большинство находок родственно гомологам РНК-лигазы 2 бактериального происхождения и последовательностям из гигантских вирусов эукариот.

Показано, что ряд океанических последовательностей проявляет большое сходство с гомологами РНК-лигазы 2 из низших беспозвоночных животных, принадлежащих к царствам *Alveolata*, *Stramenopiles*, *Rhizaria*. В этих группах животных гомологи данного фермента обнаруживаются весьма редко и возможно появились только у ряда представителей за счет горизонтального переноса генов.

Единственный выявленный океанический гомолог наиболее близкий к РНК-лигазе 2 фага T4 и других фагов подсемейства *Teevenvirinae* имеет также больше сходство с представителем архей *Thermoplasmata archaeon*.

Среди целей исследования было изучение возникновения, эволюции и биоразнообразия таких редких в природе ферментов, как РНК-лигазы 2. Найденные последовательности гомологов этих белков океанического происхождения могут быть использованы для получения новых ферментов – компонентов биотехнологического инструментария – путем рутинного синтеза гена и суперпродукции соответствующего фермента в бактериальных системах. Обнаружение и исследование родственных РНК-лигазе 2 белков может расширить наши знания об этой группе ферментов, и создать фундаментальные основы для поиска новых ингибиторов этих ферментов, получения на их основе лекарственных препаратов для борьбы с опаснейшими болезнями: трипаносомозами и лейшманиозом.

Авторы искренне благодарят Микулинскую Г.В. за конструктивные замечания по тексту статьи. Работа Назиповой Н.Н. была поддержана Российским фондом фундаментальных исследований (грант РФФИ №19-07-00996).

## СПИСОК ЛИТЕРАТУРЫ

1. Shuman S., Schwer B. RNA capping enzyme and DNA ligase: A superfamily of covalent nucleotidyl transferases. *Molecular Microbiology*. 1995. V. 17. P. 405–410. doi: [10.1111/j.1365-2958.1995.mmi.17030405.x](https://doi.org/10.1111/j.1365-2958.1995.mmi.17030405.x).



2. Silber R., Malathi V.G., Hurwitz J. Purification and properties of bacteriophage T4-induced RNA ligase. *Proc. Natl. Acad. Sci. U S A.* 1972. V. 69. P. 3009–3013. doi: [10.1073/pnas.69.10.3009](https://doi.org/10.1073/pnas.69.10.3009).
3. Wang L. K., Shuman S. Structure-function analysis of yeast tRNA ligase. *RNA.* 2005. V. 11. № 6. P. 966–975. doi: [10.1261/rna.2170305](https://doi.org/10.1261/rna.2170305).
4. Ho C.K., Shuman S. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci. U S A.* 2002. V. 99. P. 12709–12714. doi: [10.1073/pnas.192184699](https://doi.org/10.1073/pnas.192184699).
5. Abelson J., Trotta C.R., Li H. tRNA splicing. *The Journal of Biological Chemistry.* 1998. V. 273. P. 12685–12688. doi: [10.1074/jbc.273.21.12685](https://doi.org/10.1074/jbc.273.21.12685).
6. Englert M., Beier H. Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Research.* 2005. V. 33. P. 388–399. doi: [10.1093/nar/gki174](https://doi.org/10.1093/nar/gki174).
7. Blanc V., Alfonzo J.D., Aphasizhev R., Simpson L. The mitochondrial RNA ligase from *Leishmania tarentolae* can join RNA molecules bridged by a complementary RNA. *Journal of Biological Chemistry.* 1999. V. 274. P. 24289–24296. doi: [10.1074/jbc.274.34.24289](https://doi.org/10.1074/jbc.274.34.24289).
8. Palazzo S.S., Panigrahi A.K., Igo R.P. Jr., Salavati R., Stuart K. Kinetoplastid RNA editing ligases: complex association, characterization, and substrate requirements. *Molecular and Biochemical Parasitology.* 2003. V. 127. P. 161–167. doi: [10.1016/s0166-6851\(02\)00333-x](https://doi.org/10.1016/s0166-6851(02)00333-x).
9. Stuart K., Brun R., Croft S., Fairlamb A., Gurtler R.E., McKerrow J., Reed S., Tarleton R. Kinetoplastids: related protozoan pathogens, different diseases. *J. Clin. Invest.* 2008. V. 118. P. 1301–1310. doi: [10.1172/JCI33945](https://doi.org/10.1172/JCI33945).
10. Simpson L., Da Silva A. Isolation and characterization of kinetoplast DNA from *Leishmania tarentolae*. *J. Mol. Biol.* 1971. V. 56. P. 443–473. doi: [10.1016/0022-2836\(71\)90394-9](https://doi.org/10.1016/0022-2836(71)90394-9).
11. Blum B., Bakalara N., Simpson L. A model for RNA editing in kinetoplastid mitochondria: RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell.* 1990. V. 60. P. 89–198. doi: [10.1016/0092-8674\(90\)90735-W](https://doi.org/10.1016/0092-8674(90)90735-W).
12. Sturm N.R., Simpson L. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell.* 1990. V. 61. P. 879–884. doi: [10.1016/0092-8674\(90\)90198-N](https://doi.org/10.1016/0092-8674(90)90198-N).
13. Rehse P.H., Tahirov T.H. Structure of a putative 2'-5' RNA ligase from *Pyrococcus horikoshii*. *Acta Crystallographica Section D: Biological Crystallography.* 2005. V. 61. P. 1207–1212. doi: [10.1107/s0907444905017841](https://doi.org/10.1107/s0907444905017841).
14. Desai K.K., Bingman C.A., Phillips G.N. Jr., Raines R.T. Structures of the Noncanonical RNA Ligase RtcB Reveal the Mechanism of Histidine Guanylylation. *Biochemistry.* 2013. V. 52. P. 2518–2525. doi: [10.1021/bi4002375](https://doi.org/10.1021/bi4002375).
15. Desai K.K., Cheng C.L., Bingman C.A., Phillips G.N. Jr., Raines R.T. A tRNA splicing operon: archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation. *Nucleic Acids Research.* 2014. V. 42. P. 3931–3942. doi: [10.1093/nar/gkt1375](https://doi.org/10.1093/nar/gkt1375).
16. Aphasizhev R., Aphasizheva I. Mitochondrial RNA editing in trypanosomes: small RNAs in control. *Biochimie.* 2014. V. 100. P. 125–131. doi: [10.1016/j.biochi.2014.01.003](https://doi.org/10.1016/j.biochi.2014.01.003).
17. Moreira S., Noutahi E., Lamoureux G., Burger G. Three-dimensional structure model and predicted ATP interaction rewiring of a deviant RNA ligase 2. *BMC Struct. Biol.* 2015. V. 15. Article No. 20. doi: [10.1186/s12900-015-0046-0](https://doi.org/10.1186/s12900-015-0046-0).
18. Williamson S.J., Rusch D.B., Yooseph S., Halpern A.L., Heidelberg K.B., Glass J.I., Andrews-Pfannkoch C., Fadrosh D., Miller C.S., Sutton G., Frazier M., Venter J.C..

- The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. *PLoS One*. 2008. V. 3. Article No. e1456. doi: [10.1371/journal.pone.0001456](https://doi.org/10.1371/journal.pone.0001456).
19. Yooseph S., Sutton G., Rusch D.B., Halpern A.L., Williamson S.J., Remington K., Eisen J.A., Heidelberg K.B., Manning G., Li W., et al. The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol.* 2007. V. 5. Article No. e16. doi: [10.1371/journal.pbio.0050016](https://doi.org/10.1371/journal.pbio.0050016).
  20. Jorgensen S.L., Hannisdal B., Lanzén A., Baumberg T., Flesland K., Fonseca R., Ovreås L., Steen I.H., Thorseth I.H., Pedersen R.B., Schleper C. Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc. Natl. Acad. Sci. U S A*. 2012. V. 109. P. E2846–E2855. doi: [10.1073/pnas.1207574109](https://doi.org/10.1073/pnas.1207574109).
  21. Brettin T., Davis J.J., Disz T., Edwards R.A., Gerdes S., Olsen G.J., Olson R., Overbeek R., Parrello B., Pusch G.D., et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 2015. V. 5. Article No. 8365. doi: [10.1038/srep08365](https://doi.org/10.1038/srep08365).
  22. King A.M.Q., Lefkowitz E.J., Mushegian A.R., Adams M.J., Dutilh B.E., Gorbalenya A.E., Harrach B., Harrison R.L., Junglen S., Knowles N.J., et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Arch. Virol.* 2018. V. 163. P. 2601–2631. doi: [10.1007/s00705-018-3847-1](https://doi.org/10.1007/s00705-018-3847-1).
  23. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res.* 2012. V. 40. P. D136–D143. doi: [10.1093/nar/gkr1178](https://doi.org/10.1093/nar/gkr1178).
  24. Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. GenBank. *Nucleic Acids Res.* 2013. V. 41. P. D36–D42. doi: [10.1093/nar/gks1195](https://doi.org/10.1093/nar/gks1195).
  25. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997. V. 25. P. 3389–3402. doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389).
  26. Jones D.T., Taylor W.R., Thornton J.M. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*. 1992. V. 8. P. 275–282. doi: [10.1093/bioinformatics/8.3.275](https://doi.org/10.1093/bioinformatics/8.3.275).
  27. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 1985. V. 39. P. 783–791. doi: [10.1111/j.1558-5646.1985.tb00420.x](https://doi.org/10.1111/j.1558-5646.1985.tb00420.x).
  28. Kumar S., Stecher G., Li M., Knyaz C., Tamura K.). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*. 2018. V. 35. P. 1547–1549. doi: [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096).
  29. Cavalier-Smith T. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma*. 2018. V. 255. P. 297–357. doi: [10.1007/s00709-017-1147-3](https://doi.org/10.1007/s00709-017-1147-3).
  30. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004. V. 32. P. 1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
  31. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. ClustalW and ClustalX version 2.0. *Bioinformatics*. 2007. V. 23. P. 2947–2948. doi: [10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404).

Рукопись поступила в редакцию 21.11.2019, переработанный вариант поступил 23.12.2019.  
Дата опубликования 31.12.2019.