Математическая биология и биоинформатика. 2012. Т. 7. № 2. С. 611–631.
URL: http://www.matbio.org/2012/Zimin_7_611.pdf

========== **БИОИНФОРМАТИКА** ==========

УДК: 578.81, 51-76

Сравнительный анализ аминокислотных последовательностей отдельных доменов белков Hoc бактериофагов подсемейства *Teequatrovirinae*

©2012 Зимин А.А.[1], Микулинская Г.В.[2], Нигматуллина Л.Ф.[1], Назипова Н.Н.[3]

[1]*Институт биохимии и физиологии микроорганизмов им. Г.К. Скрябина, Российская академия наук, Пущино, Московская область, 142290, Россия*
[2]*Филиал Института биоорганической химии им. академиков М. М. Шемякина и Ю. А. Овчинникова, Российская академия наук, Пущино, Московская область, 142290, Россия*
[3]*Институт математических проблем биологии, Российская академия наук, Пущино, Московская область, 142290, Россия*

В данной работе представлены результаты сравнительного исследования иммуноглобулин-подобных доменов в геномах бактериофагов, родственных бактериофагу T4. Предложено использовать белки Hoc для классификации подсемейства фагов *Teequatrovirinae*. Осуществлен филогенетический анализ отдельных доменов 31 белка Hoc бактериофагов подсемейства. Показано, что число доменов в белках Hoc у разных бактериофагов подсемейства варьирует от одного до пяти. На основе этого бактериофаги можно разделить на 6 подгрупп. Филогенетическое дерево доменов белков - продуктов генов hoc - бактериофагов, родственных T4, образует три основные ветви. Это ветвь C-концевых доменов, ветвь N-концевых доменов и ветвь промежуточных доменов. Обязательное присутствие C-концевого домена во всех белках Hoc указывает на его функциональную и структурную значимость для формирования белка и его прикрепления к капсиду фага. Сформулированы гипотетические схемы эволюционного происхождения повторяющихся аминокислотных последовательностей в белках Hoc.

Ключевые слова: *бактериофаг T4, белок Hoc, иммуноглобулин-подобные белки, домены, филогенетическое дерево, эволюция белков.*

ВВЕДЕНИЕ

Бактериофаги составляют наиболее многочисленную популяцию биологических существ на планете [3]. Исследования показывают, что разнообразие в этой популяции бактериальных вирусов огромно [22], это обеспечивается активной ролью горизонтального переноса генетического материала, который наблюдается у бактериофагов. Таксономическая классификация бактериофагов является непростой задачей. Долгое время единственной была система Международного комитета по вирусной таксономии (ICTV, International Committee on the Taxonomy of Viruses) [15], которая основывалась на ряде физиологических и морфологических признаков организмов и природе их генетического материала (двухцепочечная или одноцепочечная ДНК, двухцепочечная или одноцепочечная РНК). В начале 21-го века появились работы, в которых стали использоваться геномные последовательности, появились новые данные, характеризующие разнообразие бактериофагов [21, 22]. В отличие от других видов организмов для бактериофагов невозможно использование генов рибосомальных РНК, этих универсальных для большинства биологических

существ таксономических маркеров, потому что у них нет рибосомных РНК. Кроме того, было показано, что нет ни одного универсального белкового маркера, который можно было бы использовать для классификации бактериофагов. Геномная классификация фагов (Phage Proteomic Tree) была осуществлена на основе идеи о сходстве протеомов у родственных бактериофагов [22], она хорошо согласуется с таксономической системой ICTV. Авторы геномной классификации предположили, что для каждой ветви таксономического дерева фагов должны быть использованы разные белки в качестве маркеров.

Существует подход к классификации фагов, основанный на теории модульной эволюции фагов [25]. Согласно этой концепции, фаговые геномы состоят из генетических элементов, которыми они обмениваются в процессе эволюции.

Целью данной работы являлось уточнение классификации бактериофагов подсемейства *Teequatrovirinae* (родственных бактериофагу T4) с использованием синтетического подхода, который основан на использовании основных идей упомянутых методов. Подсемейство T4-подобных фагов *Teequatrovirinae* поделено по сходству аминокислотных последовательностей на 2 группы родов (фаги T4-типа и KVP40-подобные вирусы). Мы исследовали первую из этих групп, которая содержит 4 подгруппы (фаги T4-типа, 44RR2.8-типа, RB43-типа, RB49-типа) [20]. Мы изучали каждую подгруппу в отдельности.

Белки, содержащие иммуноглобулин-подобные домены, относят к суперсемейству иммуноглобулинов [4, 28, 27, 8]. Иммуноглобулин-подобные белки найдены у организмов всех типов животного мира [28, 27, 11], у бактерий [4] и вирусов [5]. Они выполняют самые различные функции от адгезии до проявления эластичности. Есть специальные программные средства, разработанные для поиска в аминокислотных последовательностях участков, характерных для иммуноглобулин-подобных доменов [8].

Бактериофаг T4 – большой вирус, инфицирующий *Escherichia coli* - относится к подсемейству *Teequatrovirinae* семейства *Myoviridae* отряда *Caudovirales* [20]. Он состоит из вытянутой икосаэдрической головки и сокращающегося отростка, заканчивающегося базальной пластинкой, к которой присоединены шесть длинных фибрилл (см. рис. 2). Головка (капсид), содержащая внутри себя геномную двухцепочечную ДНК длиной 171000 пар нуклеотидов (п.н.), построена из трех белков - продуктов генов 20, 23 и 24.

Шесть белков 23 образуют основные капсомеры, из которых сформированы 20 граней икосаэдра, 11 вершин капсида образованы пентамерами белка 24, а специальная портальная вершина, к которой присоединяется отросток, образована двенадцатью копиями белка 20 [10]. Существуют также два декорирующих белка: Hoc (от: head outer capsid protein) и Soc (от: small outer capsid protein), которые присоединяются уже к собранному капсиду [16-18]. Белки Hoc и Soc не требуются для сборки капсида. Делеция одного или обоих генов, их кодирующих, не приводит к снижению продукции фага, его жизнеспособности или инфекционности при стандартных лабораторных условиях.

Белок Hoc присоединяется к центру капсомера, образованного гексамерами белков 23; это не сказывается существенным образом на стабильности капсида [23]. С помощью криоэлектронной микроскопии было показано, что молекула белка Hoc имеет форму гантели. Вероятно, два шара гантели представляют собой два функциональных модуля. Один из них, консервативный, взаимодействует с поверхностью капсида [24], а удаленный от капсида вариабельный модуль взаимодействует с поверхностью бактериальной клетки [9]. Аминокислотная подпоследовательность ESRNG, ответственная за присоединение к капсиду, локализована в 25-ти C-концевых аминокислотах, которые содержат консервативную предсказываемую петлю, окруженную двумя β-структурами, ориентирующими петлю на взаимодействие с

основным капсидным белком [24]. Структурная организация консервативного модуля формирует точку опоры на поверхности вируса, а вариабельный модуль может быть приспособлен к взаимодействию с разнообразными поверхностями, включая поверхность бактериальной клетки.

Анализ последовательности белка Hoc бактериофага T4 показал, что он состоит из четырех тандемных иммуноглобулин-подобных (Ig-подобных) доменов [5]. Три первых домена имеют типичную Ig-подобную укладку, которая обычно состоит из семи β-цепей, уложенных в два антипараллельных β-листа, упакованных в β-сандвич [12, 6].

Все известные Ig-подобные домены объединены в четыре группы: Ig-домены (I-Set), фибронектин3-домены (FN3), бактериальные Ig-подобные домены (Big2), домены полицистита печени (PKD, polycystic kidney disease) [14].

Первые два домена белка Hoc сходны по аминокислотной последовательности с PKD-доменами, а третий домен сходен с I-Set-доменами [24]. Четвертый домен имеет иммуноглобулин-подобную упаковку, сходную с доменами телокина [4]. Роль белка Hoc для развития бактериофага и для его организации до сих пор не установлена.

Белки Hoc различных бактериофагов, родственных T4, гетерогенны по длине и могут содержать разное число доменов, сходных по аминокислотной последовательности и по структуре [9]. Природа гетерогенности белков Hoc, а также пути их формирования в нынешнем виде, могут быть исследованы путем сравнительного анализа аминокислотных последовательностей отдельных доменов. Такое исследование позволит провести классификацию фагов внутри подсемейства, а кроме того, пролить свет на пути эволюции белков Hoc и других белков, содержащих повторяющиеся иммуноглобулин-подобные домены. Поиск белков, имеющих последовательности, сходные с отдельными доменами Hoc, может помочь в установлении биологической функции этого белка.

МЕТОДЫ ИССЛЕДОВАНИЯ

Для выполнения поставленной задачи мы провели сравнение аминокислотных последовательностей отдельных доменов белков Hoc бактериофагов T4 и RB49 с базами данных белковых последовательностей на сервере NCBI [2] алгоритмом PSI-BLAST [1] с уровнем достоверности результатов E-value < 0.0001. При этом позиционный итерационный поиск аналогов отдельных доменов производился до тех пор, пока каждая последующая итерация обнаруживала в базе данных GenBank новые локальные сходства. Картина сходства последовательностей не изменялась, либо почти не изменялась, на четвертом и пятом этапах сравнения. Следовательно, пяти итераций работы PSI-BLAST оказалось достаточно для получения осмысленных результатов.

На основе полученных данных был осуществлен филогенетический анализ отдельных доменов 31 белка Hoc бактериофагов подсемейства *Teequatrovirinae*. Для этого предварительно были проведены выравнивания последовательностей с помощью программы ClustalX [13], филогенетическое дерево строилось с помощью пакета программ Mega4 [26].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Классификация белков - продуктов гена hoc

Аминокислотная последовательность белка Hoc бактериофага T4 имеет длину 376 аминокислот (а.к.) и состоит из четырех похожих участков, каждый из которых имеет выраженное сходство с иммуноглобулинами. Первые три участка имеют довольно сильное сходство между собой. Длина повторяющихся фрагментов - 94 а.к. Для трех первых подобных участков из родственного белка фага RB49 в результате рентгеноструктурного анализа было показано, что каждый из них образует отдельный

домен [9]. В этой работе также было показано, что белок Hoc способен присоединяться к поверхности клетки *E.coli*. Поэтому мы называем повторяющиеся фрагменты белка Hoc доменами. Гомологичные Hoc белки обнаружены у большинства фагов подсемейства *Teequatrovirinae*. Всего была проанализирована 31 аминокислотная последовательность.

Таблица 1. Классификация белков - продуктов гена hoc

№ п /п	Название фага	Длина белка	Число доме-нов	Идентификатор белка	Макс. идент. Hoc белку T4 (%)	Макс. идент. Hoc белку RB49 (%)
Однодоменные варианты						
1	Enterobacteria phage RB43	61	1	AAX78759.1	33	55
2	Klebsiella phage KP15_01	91	1	YP_003580072.1	36	49
3	Acinetobacter phage Acj61	106	1	YP_004009803.1	68	38
4	Enterobacteria phage RB16	167	1 (2?)	ADJ55528.1	40	48
Двухдоменные варианты						
5	Aeromonas phage 31	180	2	AAX63659.1	29	32
6	Aeromonas phage 44RR2.8t	180	2	AAQ81490.1	29	32
7	Aeromonas phage 25	177	2	ABF72722.1	24	31
8	Klebsiella phage KP15_02	177	2	ADE35027.1	45	41
9	Aeromonas phage phi AS4	178	2	YP_003969113.1	25	32
Трехдоменные варианты						
10	Aeromonas phage 65_01	264	3	YP_004300980.1	37	39
11	Shigella phage SP18	282	3	YP_003934817.1	59	
12	Enterobacteria phage Bp7	282	3	AEN93941.1	59	34
13	Enterobacteria phage JS98	282	3	YP_001595307.1	59	35
14	Enterobacteria phage IME08_01	288	3	YP_003734322.1	43	37
15	Enterobacteria phage JS10_01	286	3	YP_002922526.1	29	33
Четырехдоменные варианты						
16	Enterobacteria phage T4	376	4	AAD42581.1	100	35
17	Enterobacteria phage vB_EcoM-VR7	367	4	YP_004063879.1	45	35
18	Enterobacteria phage IME08_02	377	4	YP_003734323.1	54	33
19	Enterobacteria phage RB32	376	4	ABI95002.1	89	34
20	Enterobacteria phage RB14	376	4	YP_002854513.1	86	35
21	Enterobacteria phage RB51	376	4	YP_002854135.1	91	36
22	Shigella phage Shf12	376	4	YP_004415072.1	89	34
Четырехдоменные варианты с С-концевым удлинением 3 домена						
23	Enterobacteria phage RB49	404	4	AAQ15404.1	35	100
24	Enterobacteria phage Phi1	404	4	YP_001469514.1	36	97
25	Enterobacteria phage JSE	404	4	YP_002922245.1	36	93
Пятидоменные варианты						
26	Enterobacteria phage JS10_02	469	5	YP_002922527.1	50	34
27	Enterobacteria phage RB30	472	5	AAM52483.1	77	34
28	Enterobacteria phage AR1	474	5	BAI83192.1	77	37
29	Enterobacteria phage RB69	471	5	AAP76093.1	53	37
30	Enterobacteria phage wV7	474	5	AEM00840.1	77	37
31	Enterobacteria phage ime09	472	5	AEK12435.1	76	34

Классификация белков - продуктов гена hoc - представлена в таблице 1. По числу повторяющихся доменов мы подразделяем гомологичные Hoc белки на шесть

основных групп: однодоменные, двухдоменные, трехдоменные, четырехдоменные, четырехдоменные с С-концевым удлинением 3-го домена и пятидоменные.

Первая из групп отличается наибольшим разнообразием белков по длине (61–167 а.к.). Минимальный по длине белок фага RB43 (61 а.к.) не содержит последовательности ESRNG, участвующей в присоединении к капсиду. Белок - продукт гена hoc фага RB16 - содержит один обычный С-концевой домен, который имеет С-концевое удлинение в виде последовательности длиной в 71 а.к. Этот участок последовательности не имеет сходства с повторяющимися последовательностями белков Hoc. Как известно [17], белок Hoc является основным антигеном бактериофага T4. Возможно, дополнительная последовательность белка Hoc бактериофага RB16 ответственна за новые антигенные свойства этого фага.

Группа двухдоменных вариантов варьирует по длине от 177 до 180 а.к., трехдоменные варианты содержат от 264 до 282 а.к., четырехдоменные варианты - от 367 до 377 а.к. Пятидоменные варианты незначительно варьируют по длине и содержат от 469 до 474 а.к. Четырехдоменные варианты с С-концевым удлинением третьего домена гомогенны по длине - в них по 404 а.к.

С точки зрения современной классификации можно сделать следующие замечания. Фаги, принадлежащие к T4-группе подсемейства *Teequatrovirinae,* имеют трехдоменные, четырехдоменные и пятидоменные варианты этих белков. Фаги RB43-группы имеют однодоменные варианты этого белка. Единственным исключением является бактериофаг Kp15. Его геном кодирует как однодоменный, так и двухдоменный варианты. Фаги RB49-группы несут только четырехдоменные варианты с С-концевым удлинением третьего домена. Фаги 44RR2.8-группы несут двухдоменные варианты белка Hoc. Фаги бактерий родов *Aeromonas* и *Acinetobacter* имеют однодоменные, двухдоменные и трехдоменные варианты. Бактериофаги, инфицирующие *E. coli*, несут все варианты белков Hoc, кроме двухдоменных. Фаги бактерий рода *Shigella* имеют трехдоменный и четырехдоменный варианты.

Ряд геномов бактериофагов кодируют по два Hoc-белка. Это фаги JS10 и IME08, геномы которых кодируют трехдоменный и четырехдоменный варианты белка Hoc, а также фаг KP15, геном которого кодирует однодоменный и двухдоменный варианты данного белка.

Белки Hoc бактериофагов T4 и RB49 - наиболее изученные из всего набора. Белок Hoc бактериофага T4 интенсивно изучался биохимически и структурно [16-18, 24]. Для белка Hoc бактериофага RB49 определена пространственная структура первых трех доменов [9]. Эти два белка идентичны только на 20%. Они находятся на разных концах спектра гомогенности первичных структур Hoc белков. Поэтому было интересно искать в белковой базе данных сходства с аминокислотными последовательностями отдельных доменов именно этих белков. Сравнительное исследование аминокислотных последовательностей доменов известных на сегодняшний день белков Hoc может внести ясность в эволюционные пути происхождения разнообразия этих белков.

2. Сравнительное исследование аминокислотных последовательностей отдельных доменов белка Hoc бактериофага T4

Для подробного изучения природного разнообразия белка Hoc и эволюции его последовательностей мы применили следующий подход. Ранее было показано [4], что данный белок состоит из сходных между собой четырех последовательностей по 94 а.к. в каждой. Эти участки являются доменами белка. Мы провели позиционно-специфическое сравнение с белковой базой данных GenBank каждого домена в отдельности. Результаты работы программы PSI-BLAST приведены в дополнительных материалах к статье.

2.1. Анализ первого домена

Первый домен белка Hoc бактериофага T4 имеет наибольшую степень идентичности с доменами белков Hoc родственных бактериофагов RB32, RB14, RB30, RB51, JS10, RB69, JS98. Наилучшее сходство характерно для фага RB30 (95% идентичных позиций), а также для фагов RB32, RB51 (88%). Замены в выравнивании сохраняют химическую природу (полярность, гидрофобность) заменяемых аминокислот.

Вторая итерация позиционно-специфического поиска в базе данных выдала сходство доменов белка Hoc бактериофага T4 и бактериофагов Ae25, 44RR2, Ae31, JSE, RB49 и Phi1. Следует отметить наличие консервативных подпоследовательностей FTA (22-24 а.к.), TY**h**WxxD (36-42 а.к.), где **h** обозначает гидрофобную, x – любую аминокислоту.

На третьем этапе сравнения появляются белки-рецепторы цитокина(ов) млекопитающих (E-value составляет от 10^{-11} до 10^{-12}): cytokine receptor-like factor 1 precursor (*Mus musculus*), cytokine receptor-like factor 1 (*Rattus norvegicus*), а также cytokine receptor-like factor 1 (*Homo sapiens* и *Macaca mulatta*). Известно, что биологический эффект цитокинов реализуется через специфическое взаимодействие с рецептором клетки. Многие бактериофаги обнаруживаются в крови животных, с током крови они попадают в ткани и органы. Данный факт может указывать на то, что кровь является нормальной средой для существования в ней вируса. Предполагается, что мог происходить отбор бактериофагов в организме животного на наличие иммуноглобулинподобных доменов, обеспечивающих возможность присутствия фага в крови до тех пор, пока он не достигнет бактерии-хозяина. В связи с этим мы допускаем такой эволюционный путь развития белка Hoc, как его молекулярная мимикрия под компоненты иммунной системы. Следует отметить, что группа польских исследователей указывает на возможное эукариотическое происхождение основного антигенного белка бактериофага T4 [7].

Четвертый этап сравнения добавил к списку похожих белковых последовательностей еще 15 рецепторов цитокина и 2 белка с иммуноглобулин-подобными доменами. На пятом этапе сравнения появились следующие интересные объекты: основной компонент хвоста фага VP5 (major tail subunit Vibrio phage VP5), а также компонент хвоста, кодируемый профагом CP-933P в *E. coli*. Компоненты хвоста, кодируемые профагами, и компонент хвоста фага VP5 сходны с разными участками первого домена белка Hoc бактериофага T4.

Первый домен обнаруживает некоторое сходство с капсидным белком бактериофага VP2 семейства *Podoviridae*, а также сходство с белком хвоста фага T5 в области C-конца данного белка.

2.2. Анализ второго домена

Первый этап сравнения закономерно обнаружил белки Hoc семи родственных бактериофагов. Высокая степень идентичности (~80% для RB32, RB30, RB51, RB14, а для RB69 - 57%) указывает на эволюционную близость этих бактериофагов и их основных антигенов. Во всех белках совпадают первые шесть аминокислот. Наше внимание привлекла последовательность 169-176 а.к. фага T4 (TDYDALS). В белках Hoc фагов RB51 и RB32 данный участок заменен другим консервативным сайтом (ENYNEKE). Этот факт указывает на возможное расположение белков Hoc этих двух родственных бактериофагов на параллельных ветвях эволюционного развития белка.

Второй домен белка Hoc фага T4 обнаруживает большое сходство (E ~ 10^{-17}) с двумя участками длиной по 94 аминокислоты в белковой последовательности Hoc бактериофага RB30, представляющими собой второй и третий домены этого белка (рис. 1).

```
T4      1    TLAVTPASPAAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVGGETNSTFSYTPTTSG  60
             TLAVTPASPAAGVIGT V+FTAALASQP GASATYQWYVDDS V   T++TF+YTP TSG
RB30   98    TLAVTPASPAAGVIGTAVEFTAALASQPSGASATYQWYVDDSPVSEATSATFNYTPDTSG  157

T4      61    VKRIKCVAQVTATDYDALSVTSNEVSLTVNKKTM  94
              VK+IKC AQVTAT+YDALSVTSNEVSLTVNKKT
RB30   158    VKKIKCTAQVTATNYDALSVTSNEVSLTVNKKTQ  191
```

a)

```
T4     1    TLAVTPASPAAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVGGETNSTFSYTPTTSG  60
            TLAVTPASP+AGVIGTPVQFTAALASQPDGASATYQWYVDDSQV GETNSTF+YTPTT+G
RB30   194  TLAVTPASPSAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVSGETNSTFNYTPTTNG  253

T4      61    VKRIKCVAQVTATDYDALSVTSNEVSLTVNKKTM  94
              VKRIKCVAQVTA DY+A  VTSNEVSLTVNKKTM
RB30   254    VKRIKCVAQVTADDYNAKEVTSNEVSLTVNKKTM  287
```

b)

Рис. 1. a) - фрагмент выравнивания последовательностей белков Hoc фага T4 (второй домен) и RB30 (второй домен); **b)** - фрагмент выравнивания последовательностей белков Hoc фага T4 (второй домен) и RB30 (третий домен), синим цветом окрашены аминокислотные последовательности белка RB30, совпадающие с Hoc T4. Присутствуют характерные для основного антигенного белка мотивы (выделены серым цветом).

Этот факт дает основание полагать, что эволюционно белок фага RB30 стал пятидоменным в результате дупликации участка гена hoc. Белок приобрел дополнительный иммуноглобулин-подобный домен, более чем на 90% сходный со вторым доменом белка Hoc бактериофага T4. На основании этого можно предположить, что присоединение белка к поверхности капсида бактериофага T4 осуществляется не за счет первых трех N-концевых доменов, как предполагалось ранее [4], а за счет C-концевого домена. N-концевые домены Hoc - три у T4 и четыре у RB30 - в этом случае могут быть экспонированы на поверхности капсида и обеспечивают высокую антигенность белка Hoc.

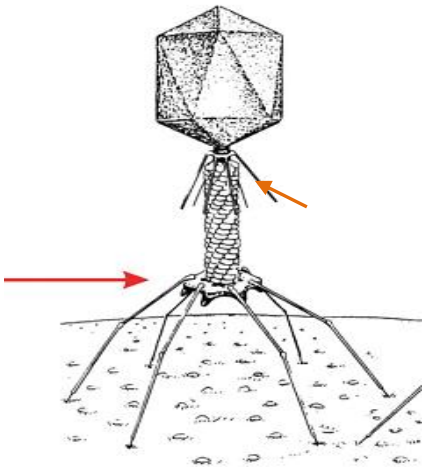

Рис. 2. Бактериофаг T4. Красная стрелка указывает на фибриллы отростков, оранжевая на воротничок фага с усиками, образованными белками Wac.

На втором этапе сравнения были обнаружены (E < 0.005) белки Hoc бактериофагов *Aeromonas* (25, 44RR2.8t, 31); *E.coli* (JS98 и JS10), в дополнительном списке (при E > 0.005) белки бактериофагов *E.coli* (Phi1, RB49 и JSE). Кроме того, второй домен Hoc T4 обнаруживает сходство с компонентом отростка различных бактериофагов (например, фага BP-4795 *Escherichia coli* O157:H7, криптического профага CP-933X *Escherichia coli* O157:H7 EDL933), а также других профагов различных штаммов *E.coli*. Следует также отметить сходство белка Hoc T4 и капсидного белка фага *Vibrio* VP2.

В ходе изучения особенностей второго домена на третьем этапе сравнения было обнаружено его сходство с белком Wac бактериофага RB43, который входит в состав воротничка фага. Мы предполагаем, что данные последовательности белков могут быть экспонированы на поверхности усиков воротничка (рис. 2). Надо заметить, что белок Hoc бактериофага RB43 состоит только из одного домена, гомологичного четвертому домену Hoc T4. Возможно, наличие последовательностей, сходных со вторым доменом Hoc, у белка Wac RB43 компенсирует их отсутствие у белка Hoc этого фага.

На третьем, четвертом и пятом этапах сравнения были найдены многие основные белки отростка профагов различных штаммов *Escherichia coli*, разные варианты рецептора лейкоцитов иммунного типа (leukocyte immune-type receptor) и белки, содержащие иммуноглобулин-подобные домены (immunoglobulin I-set domain-containing protein), а также рецепторы Fc-фрагмента иммуноглобулина G (IgG Fc receptor).

2.3. Анализ третьего домена

На первом этапе сравнения обнаруживаются белки бактериофагов RB32, RB14, RB51, RB30, JS10, RB69, JS98, JSE. Белки Hoc данных фагов проявляют разную степень сходства, однако все они сохраняют в своей последовательности следующий мотив: YSWKKDxS (рис. 3).

```
T4   189  KTMNP-QVTLTPPSINVQQDASATFTANVTGAPEEAQITYSWKKDSSPVEGSTNVYTVDT  247
          KTM+   VTLTP SINV +   A+F A V GAP  A   YSWKKD SPV G+T+   +DT
JS98 93   KTMSGVSVTLTPESINVPEGTPASFKATVAGAPAGATFAYSWKKDGSPVVGTTDTLAIDT  152
```

Рис. 3. Мотив (окрашен серым цветом), присутствующий в третьем домене Hoc T4; на рисунке представлен пример фрагмента выравнивания последовательностей третьего домена Hoc фага T4 и Hoc JS98.

Кроме того, белки Hoc бактериофагов RB32, RB14, RB51, RB30 полностью совпадают по первым 29 аминокислотам (рис 4.)

```
T4    188  KKTMNPQVTLTPPSINVQQDASATFTANVTGAPEEAQITYSWKKDSSPVEGSTNVYTVDT  247
           KKTMNPQVTLTPPSINVQQDASATFTANVT APEEAQI YSWKKDSSPVEGSTNVYTVDT
RB32  188  KKTMNPQVTLTPPSINVQQDASATFTANVTDAPEEAQIEYSWKKDSSPVEGSTNVYTVDT  247
                                       a)

T4    188  KKTMNPQVTLTPPSINVQQDASATFTANVTGAPEEAQITYSWKKDSSPVEGSTNVYTVDT  247
           KKTMNPQVTLTPPSINVQQDASATFTANVT APEEAQITYSWK+DSSPVEGSTNVYTVDT
RB30  284  KKTMNPQVTLTPPSINVQQDASATFTANVTDAPEEAQITYSWKRDSSPVEGSTNVYTVDT  343
                                       b)
```

Рис. 4. a) - фрагмент выравнивания последовательности третьего домена белка Hoc фага T4. Картина сходства (представлена не полностью) одинакова для RB32, RB14, RB51; **b)** - для белка Hoc бактериофага RB30 картина сходства (представлена не полностью) аналогична, но относится к последовательности аминокислот с 284 по 378 – участку четвертого домена.

Сходство с третьим доменом Hoc фага T4 проявляют второй, третий и четвертый домены белка Hoc фага JS10. Эта гомология доменов, как и сходство второго и третьего доменов Hoc бактериофага RB30, может говорить об их происхождении за счет внутригенной дупликации. В отличие от JS10, трехдоменный белок Hoc JS98 обнаруживает сходство с третьим доменом Hoc T4 только в области своего второго домена (93-187 а.к.). Четвертый домен Hoc фага RB30 идентичен третьему домену исследуемого белка T4 на 86%, что указывает на очевидное эволюционное родство белков данных бактериофагов и происхождение белка Hoc RB30 путем дупликации участка гена hoc фага T4.

На втором этапе сравнения в дополнительном списке мы обнаружили основной белок отростка фага VP5. Сходство последнего с последовательностью Hoc можно считать значительным, если учесть, что выравнивание последовательностей приведено

без разрывов, а многие отличающиеся аминокислоты по своим химическим характеристикам аналогичны таковым в последовательности Hoc фага T4. Кроме того, сохраняется консенсус FTA (22-24 а.к.), присутствующий во многих основных антигенных белках бактериофагов T4-типа (рис. 5.) Возможно, белки отростков, содержащие подобные домены, также как и Hoc, обладают повышенной антигенностью.

```
T4    193  PQVTLTPPSINVQQDASATFTANVTGAPEEAQITYSWKKDSSPVEGSTNVYTVDTSSVG  251
           P V+++P S +  + A  TFTA+V        +T  W  + + V+     YT   ++VG
VP5   393  PTVSISPVSASPLEPAPVTFTASVVDDGGAPPVTLKWYLNGNLVQNGGTTYTSPPTAVGQ  452
```

Рис. 5. Фрагмент выравнивания последовательностей белка hoc фага T4 и белка отростка фага VP5, серым окрашен мотив, желтым – аналогичные замены аминокислот.

Кроме того, обнаружился капсидный белок Hoc фага Phi1; его первый домен имеет сходный участок (55 а.к.), содержащий консервативную последовательность TYhWxKDxxP (226-235 а.к.), где h – обозначает гидрофильную аминокислоту. Аналогичный участок капсидного белка Hoc фага RB49 имеет 37% идентичности с его гомологом из фага T4 (рис. 6). Длина сходной области составляет 53 а.к.

```
T4     211  TFTANVTG-APEEAQITYSWKKDSSPVEGSTNVYTV-DTSSVGSQTIEVTATVTAA  264
            T TA V G  P  + +TY+W KD  P E +T   TV D +S  + + +VT   T
Rb49   21   TLTATVAGDEPLPSNLTYTWTKDDQPHENNTATLTVADATSENAGSYKVTVQDTDT  76
```

Рис. 6. Фрагмент выравнивания последовательностей белков Hoc бактериофагов T4 и RB49.

На третьем этапе сравнения была обнаружена последовательность белка отростка бактериофага T5, в которой следует отметить наличие несколько отличающейся от ранее указанной для фагов T4-типа консенсусной последовательности: TYhWxKDxSPh (226-236 а.к.). На данном этапе сравнения обнаруживаются белки отростка следующих бактериофагов: YYZ-2008, BP-4795, профага CP-933N; в них также присутствует вышеуказанный мотив.

На четвертом этапе сравнения программа выявила многочисленные белки, принадлежащие суперсемейству иммуноглобулинов, или фрагменты белков, содержащие иммуноглобулин-подобные домены. Это сходство исследуемого белка и данных доменов может объясняться сходством фолдинга белков, т.к. именно подобная иммуноглобулин-подобная укладка является энергетически выгодной и/или эволюционно древней. Также возможно предположить молекулярную мимикрию белка Hoc под компоненты иммунной системы во избежание элиминации в организме животного. Однако, вероятна и противоположная версия о возможном угнетении и преодолении системы иммунного ответа животного за счет молекулярного сходства с цитокинами, селектинами, фрагментами рецепторов иммуноглобулинов. Так, на данном этапе сравнения обнаруживается сходство 3 домена Hoc фага T4 с рецептором Fc-фрагмента IgG в области N-конца (344 а.к.).

Также на четвертом этапе выявлено сходство с последовательностью белка отростка профага *E.coli* O157:H7 str. Sakai (E = 0.045), сходные участки белков представлены без разрывов, имеется мотив YhWKKD (рис. 7).

```
T4     205  QQDASATFTANVTGAPEEAQITYSWKKDSSPVEGSTNVYTVDTSSVGSQTIEVTATVTAA  264
              +  A  T +V+         Y+WKKD  PV+G T        +  +   + T  VT +
Sakai  158  TVNTGALLTMSVSANGGTPPYKYAWKKDGQPVDGQTTDTFSKPGAQSADAGKYTCVVTDS  217
```

Рис. 7. Фрагмент выравнивания последовательностей Gp hoc фага T4 и белка отростка профага *E.coli* O157:H7 str. Sakai, серым цветом окрашен мотив YhWKKD.

Заключительный, пятый, этап сравнения не выявил новых белковых последовательностей, но в перечне подобных белков увеличилось количество

иммуноглобулин-подобных доменов, а также белков с незначительным сходством последовательностей (тайтин, паладин, обскурин, фибронектин и др.)

2.4. Анализ четвертого домена

На первом этапе сравнения было обнаружено значительное сходство четвертого домена Hoc фага T4 с белками Hoc 14-ти родственных бактериофагов T4-типа: RB69, RB14, RB30, RB32, RB51, JS10, JS98, Phi1, JSE, RB49, Ae31, 44RR2.8, Ae 25, RB43. Наибольшую идентичность (96% и более) обнаруживают белки Hoc бактериофагов RB69, RB14, RB30, RB32, RB51. Ранее мы указали на возможный факт внутригенной дупликации в гене hoc фага RB30, по причине которой этот белок приобрел дополнительный домен. Программа показала очевидное сходство четвертого домена белка Hoc фага T4 и пятого домена данного белка фага RB30 (98% идентичности). Наш интерес привлек и основной антигенный белок фага JS10. Его аминокислотная последовательность обнаруживает сходство с 4 доменом Hoc T4 также в районе пятого домена (87% идентичности; при этом замещающие аминокислоты в большинстве своем сохраняют химические характеристики: напр., E→D, K→R и т.п.). Аналогичная картина сходства (85% идентичности) представлена для Hoc фага JS98, однако у него последовательности четвертого домена белка Hoc T4 подобен третий домен. Действительно, в отличие от пятидоменных белков Hoc фагов RB30 и JS10, Hoc фага JS98 является трехдоменным. Белки Hoc остальных бактериофагов (Phi1, JSE, RB49, Ae31, 44RR2.8, Ae 25, RB43) обнаруживают гораздо меньшую степень сходства с исследуемым белком.

Для всех вышеуказанных основных антигенных белков характерен консервативный участок аминокислот ESRNG (355-359 а.к.) (табл. 2), который обеспечивает присоединение к капсиду [24].

Таблица 2. Множественное выравнивание аминокислотных последовательностей C-концевых фрагментов белков Hoc программой CLUSTALX. Красным цветом обозначен консервативный мотив. В первой колонке обозначены названия фагов, во второй колонке – координата консервативного элемента относительно начала фрагмента

Названия фагов	Номер позиции	Фрагмент множественного выравнивания
RB49	347	WRDREVYSTSK-YAKDLETIAAAEEKYSDCTCMESRNGFMYQSKELHKLDRETLERVLR
Phi1	347	WRDREVYSTSK-YAKDLETIAAAEEKYSDCTCMESRNGFMYHSKELHKLDRETLERVLR
44RR2.8t	135	WRDDPVNS--P-WPKVTYAIDKAVTDYGDCLMQESRNGYIYKASQFVKS---------
Ae31	135	WRDDPVNS--P-WPKVTYAIDKAVTDYGDCLMQESRNGYIYKASQFVKS--------
RB32	326	WKTDDPDS--K-YYLHRYTLQKMMKDYPEVDVQESRNGYIIHKTALETGIIYTYP---
T4	326	WKTDDPDS--K-YYLHRYTLQKMMKDYPEVDVQESRNGYIIHKTALETGIIYTYP---
JS98	230	WKTDDPDS--P-YYLHRYTLQKMITDYPEVDVQESRNGRIIHRTALEAGIIYDYVY--
RB69	420	WKTEDPDS--K-YYLHRYTLQKMMKDYPEVDVQESRNGYIIHKTALETGIIYTYP--
RB30	421	WKTDDPDS--K-YYLHRYTLQKMMKDYPEVDVQESRNGYIIHKTALETGIYYTYP
Ae25	133	WRQTPAES--P-WPIITFSIDRAVEEYGECLMQESRNGYIYKASAYTK---------
RB43	40	WRDYAEEK----YAKEFVTLKQAYID

Также следует отметить, что на первом этапе сравнения, помимо очевидного сходства капсидных белков бактериофагов, было выявлено некоторое сходство четвертого домена с белками иммунной системы – селектинами (E-selectin [Equus caballus], P-selectin precursor [Salmo salar]). Выравненные аминокислотные последовательности представлены с большим количеством разрывов. Однако, примечателен тот факт, что четвертый (C-концевой) домен Hoc подобен N-концевой последовательности селектинов. Известно, что селектины представляют собой адгезивные молекулы, N-конец (лектинподобный домен) которых обеспечивает

адгезию лейкоцитов к эндотелиальным клеткам. Таким образом, сходство с ними С-концевого домена Hoc – косвенное подтверждение участия именно этого домена в присоединении к поверхности капсида фага.

На втором этапе сравнения среди сходных с четвертым доменом последовательностей представлены белки-ингибиторы протеазы бактериофагов Т4-типа (RB43, Ae25, Ae 44RR2.8, JS10, JS98, RB32, RB14, RB30, RB51, RB69, JSE, Phi1, RB49). Степень сходства данных белков с исследуемой последовательностью невысока, однако следует отметить наличие консенсусной последовательности, которая была выявлена на первом этапе сравнения. В случае ингибиторов протеазы она выглядит так: **hh**ESRN (353-358 а.к.). Вероятно, эта последовательность, как и у белков Hoc, может обеспечивать присоединение к поверхности капсида.

Третий, четвертый и пятый этапы сравнения не выявили дополнительных белковых последовательностей, сходных с четвертым доменом белка Hoc фага Т4.

3. Сравнительное исследование аминокислотных последовательностей отдельных доменов белка Hoc бактериофага RB49

Белок Hoc бактериофага RB49 сильно отличается от белка Hoc фага Т4: он идентичен на 35% в зоне наилучшего наложения и идентичен на 20% по всей длине. Поэтому было интересным провести сравнение аминокислотных последовательностей его доменов с базами данных белковых последовательностей аналогично сравнению доменов Hoc белка бактериофага Т4. Первые три домена белка Hoc бактериофага RB49 были кристаллизованы, а их структура решена с помощью рентгеноструктурного анализа [10]. В ходе этого анализа было определено, что первый и второй домен содержат соответственно 90 и 91 аминокислоту, а третий домен значительно длиннее – в нем 123 аминокислоты. Было показано, что все три домена образуют линейную структуру. Кристаллизовать белок Hoc бактериофага RB49 целиком оказалось невозможно, так как между третьим и четвертым доменом есть гибкое сочленение.

3.1. Анализ первого домена

На первом этапе сравнения были найдены белки Hoc родственных RB49 бактериофагов Phi1 и JSE с высоким процентом идентичности - 98% и 93%, соответственно. Более низким процентом идентичности (36%) обладает белок Hoc бактериофага RB69. Белки отростка у бактериофагов *Escherichia coli* EcoS-CEV2 и T5C обладают бóльшим процентом максимальной идентичности (44% и 41%, соответственно). Также были обнаружены предполагаемые белки отростка профагов *E. coli* ED1a, E110019, DEC9D, *Salmonella enterica* и *Shigella boydii* Sb227, подобные белки профагов у других штаммов *E. coli*, предполагаемый белок капсида бактериофага энтеробактерий pb10 H8, белки Wac бактериофагов RB43 и RB16, белки Opitutus terrae, Desulfitobacterium hafniense, *E. coli* STEC_B2F1, *E. coli* OK1180 и bacterium Ellin514, содержащие иммуноглобулинподобный домен 1-набора (immunoglobulin I-set domain-containing protein), рецептор В-клеток CD22 *Cricetulus griseus*, белок тайтин *Alligator sinensis* и *Crocodylus siamensis*.

На втором этапе сравнения дополнительно были найдены белки отростка фага энтеробактерий EPS7 и Felix 01, фага *Yersinia* PY54, вибриофагов pVp-1, VP5 и фага BP-4795, белок Wac бактериофага *Klebsiella* KP15, бактериальный поверхностный белок *Klebsiella pneumonia*, содержащий Ig-подобные домены.

На третьем, четвертом и пятом этапах сравнения было дополнительно найдено большое число рецепторов Fc-фрагмента антител из различных эукариотических организмов. Эти находки, как и в случае белка Hoc фага Т4, могут говорить о молекулярной мимикрии бактериофагов под компоненты иммунной системы.

Первый домен белка Hoc бактериофага RB49 проявляет сходство со многими белками отростков бактериофагов. Вероятно, так же как и Hoc, эти белки несут иммуноглобулин-подобные домены. Сходство с подобными белками было обнаружено и при анализе первых трех доменов белка Hoc бактериофага T4.

3.2. Анализ второго и третьего доменов

При сравнении второго и третьего доменов с базой данных известных белковых последовательностей во всех пяти итерациях работы алгоритма PSI-BLAST в достоверной области ($E < 0.0001$) результатов было обнаружено лишь сходство с последовательностями соответствующих доменов белков Hoc близкородственных фагов Phi1 и JSE.

Отсутствие сходства второго и третьего доменов белка Hoc бактериофага RB49 с иммуноглобулинами и иммуноглобулин-подобными белками явилось неожиданным результатом. Это говорит о существенном отличии в антигенных и других свойствах белка Hoc фага RB49 (а также Phi1 и JSE) от соответствующего белка фага T4 и ряда других родственных бактериофагов.

3.3. Анализ четвертого домена

На первом этапе сравнения было обнаружено сходство с белками Hoc бактериофагов энтеробактерий Phi1, JSE, RB16, CC31, RB43, wV7, AR1, T4, ime09, RB51, RB30, RB32, vB_EcoM-VR7, T4T, RB14, JS10, IME08, JS98, Bp7, а также бактериофагов родов *Acinetobacter* (Acj61 и Acj9), *Aeromonas* (65, phiAS4, 44RR2.8t, 31), *Shigella* (SP18 и Shfl2) и *Klebsiella* (KP15).

На втором этапе сравнения дополнительно обнаружены белки Hoc фага *Aeromonas* 25 и второй вариант белка Hoc энтеробактериофага JS10.

На третьем этапе сравнения дополнительно в достоверной области сравнения был обнаружен белок Inh - ингибитор протеазы бактериофага RB43.

На четвертом этапе сравнения были найдены 32 ингибитора протеазы проголовки бактериофагов энтеробактерий RB16, CC31, Bp7, JS10, wV7, JS98, IME08, RB32, ime09, T4T, RB14, AR1, RB51, T4, RB30, vB_EcoM-VR7, RB69, Phi1, JSE, RB49, фагов Acinetobacter Acj9, 133, Acj61, Ac42, фагов *Aeromonas* 44RR2.8t, phiAS4, 65, 25, phiAS5, бактериофагов *Shigella* Shfl2 и SP18 и фага *Klebsiella* KP15.

На пятом этапе сравнения были обнаружены еще два белка Inh бактериофагов *Aeromonas* Aeh1 и PX29 и тяжелые цепи иммуноглобулинов различных животных. Белок Inh – ингибитор протеазы проголовки бактериофагов подсемейства Teequatrovirinae - имеет C-концевой домен, сходный с C-концевым доменом белка Hoc. Более того, этот белок содержит близко к C-концу подпоследовательность ESRN, схожую с последовательностью ESRNG, для которой у белка Hoc биохимически показано участие во взаимодействии с поверхностью фага [24]. C-концевой домен белков Hoc ответственен за присоединение к головке бактериофага. Вероятно, ингибитор протеазы проголовки, белок Inh, за счет своего C-концевого домена, может временно присоединяться к головке, ингибируя протеазу, участвующую в ее созревании.

Интересным результатом является сходство последнего домена Hoc RB49 с тяжелыми цепями иммуноглобулинов. Ранее считалось, что четвертый домен Hoc не имеет сходства с иммуноглобулинами [24], однако выяснилось, что такое сходство есть, что согласуется с данными группы Chotia [4].

Сравнение отдельных доменов бактериофага RB49 с базой данных белковых последовательностей существенным образом отличается от подобного анализа отдельных доменов фага T4. Все домены бактериофага T4 проявляли сходство с белками Hoc многих родственных бактериофагов. У бактериофага RB49 для первых

трех доменов программой PSI-BLAST обнаруживается сходство лишь с белками Hoc близкородственных бактериофагов Phi1 и JSE, относящихся к одной RB49-группе фагов. Единственным исключением является сходство первого домена с первым доменом белка Hoc бактериофага RB69, относящегося к T4-группе. Сходство белка Hoc бактериофага RB49 с белками Hoc других бактериофагов ограничивается четвертым доменом, ответственным за присоединение к поверхности капсида. Следовательно, на поверхности капсидов бактериофагов RB49, Phi1 и JSE экспонирован отличающийся от других бактериофагов антиген.

4. Филогенетическое дерево отдельных доменов белков Hoc бактериофагов подсемейства *Teequatrovirinae*

В рамках анализа роли отдельных доменов в обеспечении структуры и функции Hoc было построено дерево сходства отдельных доменов в белках - продуктах генов hoc бактериофагов, родственных T4; оно представлено на рисунке 8.

Белки Hoc различных бактериофагов обозначены аналогично самим бактериофагам, здесь приняты следующие сокращения: Ae31 –фаг *Aeromonas* 31; 44RR - фаг *Aeromonas* 44RR2.8t; phi AS4 - фаг *Aeromonas* phi AS4; JS98 - фаг энтеробактерий JS98; VR7 - фаг энтеробактерий vB_EcoM - VR7; RB30 - фаг энтеробактерий RB30; RB69 - фаг энтеробактерий RB69; RB49 - фаг энтеробактерий RB49; AR1 - фаг энтеробактерий AR1; RB16 - фаг энтеробактерий RB16; RB43 - фаг энтеробактерий RB43; Acj9 - фаг *Acinetobacter* Acj9; Acj61 - фаг *Acinetobacter* Acj61; T4 - фаг энтеробактерий T4; Ae25 - фаг *Aeromonas* 25; Ae65 – фаг *Aeromonas* 65; Phi1 - фаг энтеробактерий Phi1; JSE - фаг энтеробактерий JSE; RB32 - фаг энтеробактерий RB32; RB14 - фаг энтеробактерий RB14; RB51 - фаг энтеробактерий RB51; wV7 - фаг энтеробактерий wV7; ime09 - фаг энтеробактерий ime09. Ряд фагов имеют по два белка Hoc; они обозначены цифрами после сокращенного названия фага: IME01 - фаг энтеробактерий IME08 (трехдоменный вариант белка фага IME08); IME02 - фаг энтеробактерий IME08 (четырехдоменный вариант белка фага IME08); JS1001 - фаг энтеробактерий JS10 (трехдоменный вариант белка фага JS10); JS1002 - фаг энтеробактерий JS10 (четырехдоменный вариант белка фага JS10); KP1501 – фаг *Klebsiella* KP15 (однодоменный вариант белка фага KP15); KP1502 – фаг *Klebsiella* KP15 (двухдоменный вариант белка фага KP15). Цифра после сокращенного названия фага обозначает номер домена, считая с N-конца белка Hoc.

Дерево имеет три ветви: ветвь C-концевых доменов, ветвь N-концевых доменов и ветвь промежуточных доменов. Это соответствует результатам моделирования структуры белка Hoc фага T4, которое показало, что первый и последний домен сближены и являются организаторами структуры белка и участвуют в присоединении белка к поверхности фага [24].

Роль промежуточных доменов другая – они экспонированы на поверхности фага и, скорее всего, ответственны за взаимодействие с антителами и поверхностью клетки *E.coli* [9]. Результаты анализа последовательностей отдельных доменов косвенно подтверждают эти заключения.

Можно отметить, что однодоменные белки фагов RB43, RB16 и KP1501 собраны в одну подветвь внутри ветви C-концевых доменов и находятся на близком расстоянии друг от друга. Они образуют общую ветвь с C-концевыми доменами двудоменных белков фагов *Aeromonas*: 44RR2.8t, 31, 25 и phi AS4. Это может говорить о том, что они параллельно эволюционировали за счет делеций, происходивших в генах более длинных белков-предшественников. В то же время однодоменный белок фага Acj61 проявляет большее сходство с C-концевыми доменами белков, подобных Hoc фага T4.

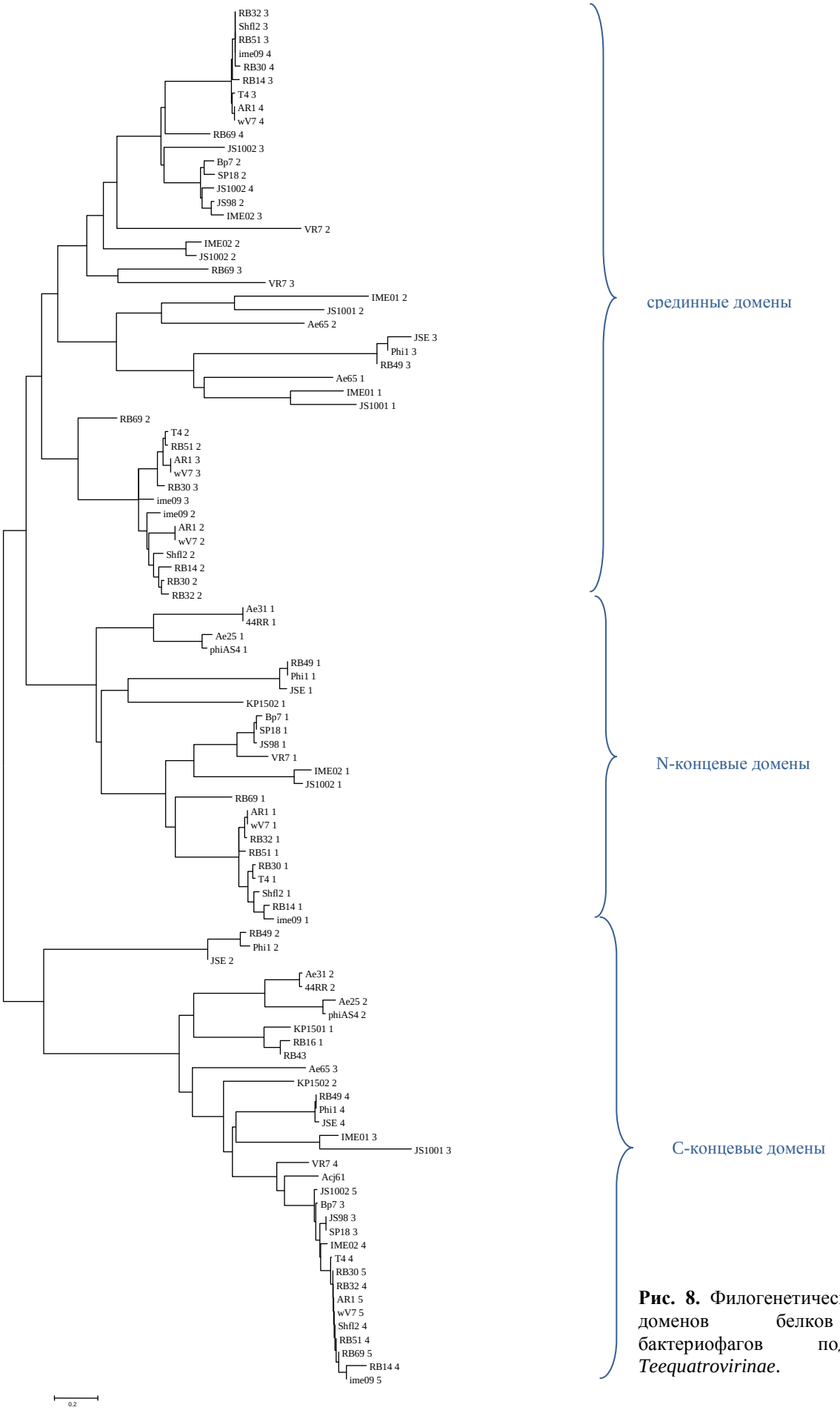

Рис. 8. Филогенетическое дерево доменов белков Hoc бактериофагов подсемейства *Teequatrovirinae*.

Однодоменные белки Hoc представляют собой отдельно существующие С-концевые домены. Это говорит о первоочередной важности для функционирования именно С-концевых доменов. Также можно предположить, что С-концевые домены более эволюционно древние, чем другие домены белков Hoc. Возможно, они были предшественниками других доменов, а сами эволюционировали в направлении совершенствования свойства присоединения к капсиду фага.

На ветви С-концевых доменов есть аномальная уникальная подветвь вторых доменов Hoc бактериофагов RB49, Phi1, JSE. Вероятно, сходство этих доменов с С-концевыми доменами связано с происхождением второго домена белков бактериофагов RB49, Phi1, JSE из С-концевого домена.

На ветви промежуточных доменов есть также уникальная подветвь третьих доменов белков Hoc бактериофагов RB49, Phi1, JSE и N-концевых доменов трехдоменных белков фагов Ae65, IME08, JS10. Это может указывать на эволюционную связь третьего домена белка Hoc бактериофагов RB49, Phi1, JSE и N- концевого домена.

Домены 2 и 3 пятидоменных белков бактериофагов RB30, AR1, wV7, ime09 находятся на одной подветви общего дерева, что может говорить об их происхождении за счет внутригенной дупликации. На этой же подветви находятся домены 2 четырехдоменных Hoc бактериофага T4 и близкородственных ему бактериофагов RB14, RB32, RB51. Вероятно, эти домены могли дуплицироваться в ходе эволюции, которая привела к появлению доменов 2 и 3 пятидоменных белков. Домены 2 и 3 фага RB30 различаются только по 17 а.к. из 94 (рис. 9), домены 2 и 3 фагов AR1 и ime09 различаются по 25 а.к. из 94 (рис. 1 и 2, доп. материалы), а соответствующие домены белка Hoc бактериофага wV7 различаются по 24 аминокислотам (рис. 3, доп. материалы). Аминокислотные замены различны у разных фагов, что говорит о независимой эволюции дуплицированных доменов.

Второй и третий домен пятидоменного белка Hoc бактериофага RB69 идентичны на 40% (38 из 96 аминокислот совпадают) (рис. 4, доп. материалы). Они также могли произойти за счет внутригенной дупликации, и, судя по числу отличий, это результат более давнего эволюционного процесса.

Можно заключить, что, во-первых, пятидоменные варианты образуются из четырехдоменных путем дупликации генетического материала в участке, кодирующем второй домен четырехдоменного варианта. В результате образуются два домена, сходных между собой и с родительским доменом. Во-вторых, дупликация в этих фагах происходила в разное время: в фаге RB30 относительно недавно, в фаге AR1 - немного ранее, а у фага RB69 значительно раньше. В-третьих, можно предположить, что, если дупликация может происходить в разное время в одном и том же месте, то существует генетический механизм, являющийся генератором дупликаций. Действительно, на относительно небольшом расстоянии от этого участка генома фага T4 находится сайт нуклеазы SegE, являющийся горячей точкой рекомбинации [19]. Неправильное разрешение рекомбинационных структур, часто возникающих в этой области генома, может служить механизмом появления дупликаций. Внутригенная дупликация привела к образованию пятидоменных вариантов последовательности белка Hoc.

```
RB30_3    191   QTTTLAVTPASPSAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVSGETNSTFNYTPT
RB30_2     95   QTTTLAVTPASPAAGVIGTAVEFTAALASQPSGASATYQWYVDDSPVSEATSATFNYTPD
T4_2       95   QTTTLAVTPASPAAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVGGETNSTFSYTPT
                ************:******.*:*********.************* *.  *.:**.***

RB30_3    251   TNGVKRIKCVAQVTADDYNAKEVTSNEVSLTVNKKT
RB30_2    155   TSGVKKIKCTAQVTATNYDALSVTSNEVSLTVNKKT
T4_2      155   TSGVKRIKCVAQVTATDYDALSVTSNEVSLTVNKKT
                *.***:***.***** :*:* .**************
```

Рис. 9. Множественное выравнивание аминокислотных последовательностей доменов 2 и 3 белка Hoc фага RB30 и второго домена белка Hoc фага T4.

```
RB69_2     94   ENNSTVAVTPASPAAVEIGTATTFTANVSNQPSGAAIAYTWKVDGVAVDGQKQSTFEYTP
T4_2       95   -QTTTLAVTPASPAAGVIGTPVQFTAALASQPDGASATYQWYVDDSQVGGETNSTFSYTP
RB69_3    190   ANSSTLKITPESPTT-VFGVPITLTANVSGAPSGATTSFQWSMDDSNILDATSATYKFTP
                 :.:*: :** **::  :*..  :** ::. *.**: :: * :*.  : . ..:*:.:**

RB69_2    154   TSEGTKSITCSVTVTATDYVDKTVESSAVSLTVNKK-
T4_2      154   TTSGVKRIKCVAQVTATDYDALSVTSNEVSLTVNKKT
RB69_3    259   TEVGSKTLKCTVSVSATNYVTKEISAEATVVTNNATF
                *  * * :.* . *:**:*    : :. . :* * .
```

Рис. 10. Множественное выравнивание аминокислотных последовательностей доменов 2 и 3 белка Hoc фага RB69 и второго домена белка Hoc фага T4

Второй и третий домены четырехдоменного белка фага vB_EcoM-VR7 идентичны на 32% (рис. 11), однако сходство между ними большее, чем с первым доменом (26%). Эта степень сходства достаточно велика, чтобы предположить, что их происхождение также является результатом дупликации последовательности домена-предшественника. Таким образом, можно также предполагать происхождение четырехдоменного варианта vB_EcoM-VR7 от трехдоменного путем внутригенной дупликации.

Таким образом, промежуточные домены белков всех фагов (за исключением фагов RB49, Phi1, JSE) располагаются на отдельной ветви дерева сравнения доменов, как мы писали ранее, на одной из трех основных ветвей. Это говорит о большем сходстве промежуточных доменов между собой, чем с концевыми доменами, и о возможном участии внутригенных дупликаций в их эволюции.

5. Эволюционные схемы образования вариантов аминокислотной последовательности белка Hoc

Трехдоменные варианты белка Hoc встречаются у бактериофагов разнообразных родов бактерий - *E. coli*, *Aeromonas*, *Shigella*. Трехдоменный вариант белка Hoc обладает всеми основными чертами предполагаемой структуры. Его C-концевой домен ответственен за присоединение к поверхности фага, N-концевой домен взаимодействует с C-концевым и образует структуру типа петли [24], промежуточный домен осуществляет связывание антител и присоединение к поверхности клетки [9]. Поскольку трехдоменный вариант обладает всеми основными чертами структуры, то мы предполагаем, что именно он лежит в основе разнообразия белков Hoc. Попытаемся смоделировать происхождение разнообразия последовательностей белка Hoc от трехдоменного варианта.

Схема эволюции за счет внутригенной дупликации могла бы включать в себя две стадии. Во-первых, дупликацию части гена, кодирующую промежуточный домен с образованием из трехдоменного варианта четырехдоменного, во-вторых, дальнейшую дупликацию одного из промежуточных доменов с образованием из четырехдоменного варианта пятидоменного (рис. 12).

```
VR7_2      94   MSATATLTTSTPTVKVGEEYNASADVTGEPGGATIAYLWS--TGEITKDITRTATVAGPV
VR7_3     189   ---VEISGPSTATVDV--PFNLTASVSPAIPGATLAYKWDDNSTEATRAIT--ESTEGLK
                   .    .**.**.*   :* :*.*:    ***:** *.  : * *: **   :. *

VR7_2     152   SLTCQITVSAADYDNQVINPEAVVVTVENN-TFPEFT----
VR7_3     142   SYTCEVTASQTGFTNSVKSGNKSVTVEEAEPVIPEECPLIY
                * **::*.* :.: *.* . :  *.. * : .:**
```

Рис. 11. Парное выравнивание аминокислотных последовательностей доменов 2 и 3 белка Hoc фага vB_EcoM-VR7.

Движущей силой эволюции путем увеличения числа промежуточных доменов могла бы быть их способность связывать поверхность бактериальной клетки, что может быть предварительным этапом взаимодействия фага и бактерии, предшествующим заражению. Также очень существенным для изменения последовательности в этом месте генома является наличие горячей точки рекомбинации, возможно, даже более

существенным, чем положительный отбор на варианты с дополнительными срединными доменами.

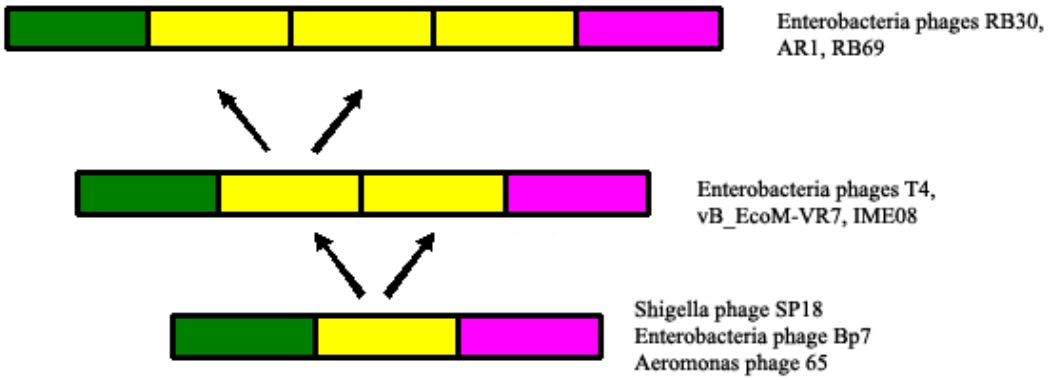

Рис. 12. Схема внутригенных дупликаций в процессе эволюции трехдоменного варианта белка Hoc в пятидоменный. Зеленым цветом обозначен N-концевой домен, желтым промежуточные домены и сиреневым C-концевой домен. Процесс дупликации обозначен стрелками. Справа приведены фаги, несущие те или иные варианты белка Hoc, соответствующие стадиям процесса.

Мутация по гену hoc почти никак не отражается на урожае фага или его устойчивости к внешним воздействиям [17, 18]. С другой стороны, отсутствие этого белка на поверхности капсида приводит к повышенной агрегации частиц [29], что может сказываться на инфекционности популяций фагов. С этой точки зрения, любой из вариантов белка Hoc, включая однодоменный и двухдоменный варианты, предпочтительнее полного отсутствия домена. Широкая вариабельность длин однодоменных белков позволяет предполагать, что они являются продуктом делеций. Возможно, рекомбинация между полидоменными вариантами могла продуцировать и делеционные варианты, состоящие из одного или двух доменов (рис. 13), при этом они отбирались по одному лишь свойству – возможности присоединения к поверхности фага. Агрегация фага может снижать его инфекционность, поэтому присутствие на поверхности головки хотя бы однодоменного варианта необходимо.

В связи с тем, что некоторые из бактериофагов содержат два варианта гена hoc, между ними может происходить рекомбинация с образованием новых вариантов этого гена (рис. 13).

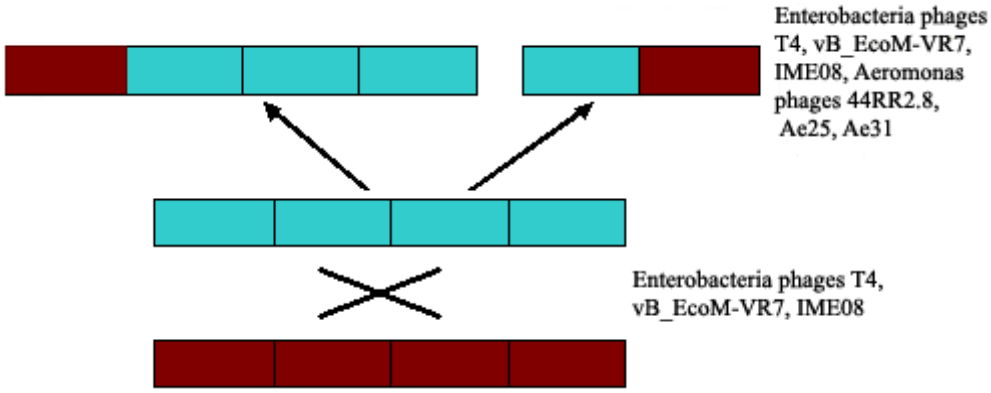

Рис. 13. Схема перемешивания доменов при рекомбинации между генами hoc различных близкородственных бактериофагов. Рассмотрен вариант замещения домена и делеции промежуточных доменов.

Для четырехдоменных белков Hoc бактериофагов RB49, Phi1, JSE можно предположить альтернативный механизм происхождения четырехдоменного варианта. Положение вторых доменов бактериофагов RB49, Phi1, JSE на общем филогенетическом дереве позволяет проследить сходство этих доменов с C-концевыми доменами как этих, так и других фагов. В свою очередь, третьи домены бактериофагов

RB49, Phi1, JSE относятся к одной ветви с первыми (N-концевыми) доменами белков Hoc фагов Ae65, IME08, JS10. Подобное сходство позволяет предположить другую модель происхождения белков Hoc бактериофагов RB49, Phi1, JSE - за счет дупликации гена двухдоменного предшественника (рис. 13).

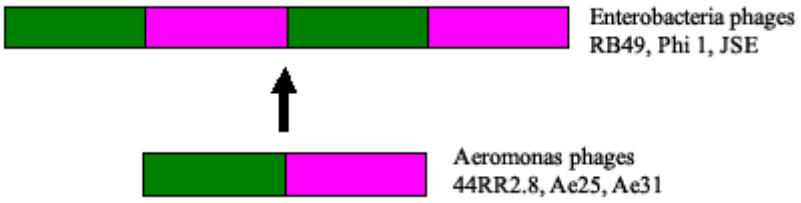

Рис. 14. Схема дупликации в процессе эволюции двухдоменного варианта белка Hoc в четырехдоменный. Зеленым цветом обозначен N-концевой домен и сиреневым C-концевой домен.

Таким образом, мы предположили три варианта эволюции аминокислотной последовательности белков Hoc: вариант внутригенных дупликаций части гена, кодирующей промежуточный домен, вариант перемешивания доменов при рекомбинации между генами hoc различных близкородственных бактериофагов и вариант происхождения четырехдоменного белка при дупликации гена двухдоменного белка. Скорее всего, в возникновении большого разнообразия аминокислотных последовательностей белка Hoc, основного фагового антигена, принимали участие все три варианта рекомбинационной эволюции их генов.

ЗАКЛЮЧЕНИЕ

Научная литература и базы данных изобилуют описаниями T4-подобных бактериофагов. Их выделяют из неуклонно растущего числа различных биологических объектов. Это позволяет утверждать, что имеющиеся в распоряжении исследователей отсеквенированные геномы T4-подобных бактериофагов – это лишь небольшая часть разнообразия этих вирусов, которое реально существует в природе. Разработка методов их классификации имеет особую важность. Разработанный здесь подход согласуется с известными методами классификации по геномным последовательностям [20] и основан на более естественных критериях. В качестве генетического маркера выбраны белки, содержащие иммуноглобулин-подобные домены. Эти маркеры существуют в геномах почти всех представителей подсемейства. Классификация, основанная на числе и качестве таких доменов, адекватна общепринятым взглядам, согласно которым эволюционная связь между фагами маскируется горизонтальным переносом генетического материала.

Число иммуноглобулин-подобных доменов в белке Hoc разных бактериофагов подсемейства *Teequatrovirinae* варьирует от одного до пяти и определяет антигенные свойства белка. Использование в качестве критерия для классификации бактериофагов количество доменов в белке Hoc приводит к разделению всех бактериофагов подсемейства на 6 групп (табл. 1):

- бактериофаги, имеющие однодоменные варианты белка,
- бактериофаги, имеющие двухдоменные варианты белка,
- бактериофаги, имеющие трехдоменные варианты белка,
- бактериофаги, имеющие четырехдоменные варианты белка,
- бактериофаги, имеющие четырехдоменные варианты белка с C-концевым удлинением третьего домена,
- бактериофаги, имеющие пятидоменные варианты белка.

Филогенетическое дерево бактериофагов, родственных T4, построенное по отдельным доменам белков - продуктов генов hoc - образует три основные ветви. Это

ветвь С-концевых доменов, ветвь N-концевых доменов и ветвь промежуточных доменов.

Обязательное присутствие С-концевого домена во всех белках Hoc указывает на его функциональную и структурную значимость для формирования белка и его прикрепления к капсиду фага. Сходство этого домена с адгезивными белками – селектинами и наличие в нем консервативной белковой последовательности ESRNG также косвенно свидетельствует о его участии в присоединении к капсиду.

При анализе первых трех доменов белка Hoc бактериофага T4 было обнаружено выраженное сходство с белками-компонентами иммунной системы млекопитающих. Это может указывать на то, что кровь является нормальной средой для существования в ней фага. В связи с этим возможна его молекулярная мимикрия под компоненты иммунной системы. Наличие подобных доменов в белках отростков других бактериофагов может говорить о широком распространении такой молекулярной мимикрии среди фагов.

Сравнительный анализ отдельных доменов белка Hoc бактериофага RB49 показал, что он вместе с белками Hoc бактериофагов Phi1 и JSE образует обособленную группу.

Для этой группы характерно отсутствие сходства второго и третьего доменов белка Hoc с иммуноглобулинами и иммуноглобулин-подобными белками, а также с соответствующими доменами белков Hoc фага T4 и ряда других родственных бактериофагов. Это говорит о существенном отличии в антигенных и других свойствах белка Hoc фага RB49 (а также Phi1 и JSE).

Сравнительный анализ доменной организации Hoc позволяет выделить три возможных пути эволюции его генов: а) внутригенная дупликация промежуточных доменов, б) перемешивание доменов при рекомбинации между генами hoc различных родственных фагов и в) дупликация гена двухдоменного варианта белка Hoc.

Работа была частично поддержана грантами РФФИ №12-07-00530, № 11-04-96575, № 12-04-90864.

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. *Basic Local Alignment Search Tool*. URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi . (дата обращения 12.11.2012).
3. Balter M. Virology. Evolution on life’s fringes. *Science*. 2000. V. 289. P. 1866-1867.
4. Bateman A., Eddy S.R., Chothia C. Members of the immunoglobulin superfamily in bacteria. *Protein Sci*. 1996. V. 5. № 9. P.1939-1941.
5. Bateman A., Eddy S.R., Mesyanzhinov V.V. A member of the immunoglobulin superfamily in bacteriophage T4. *Virus Genes*. 1997. V. 14. P. 163–165.
6. Bork P., Holm L., Sander C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol*. 1994. V. 242. P. 309–320.
7. Dabrowska K., Zembala M., Boratynski J., Switala-Jelen K., Wietrzyk J., Opolski A., Szczaurska K., Kujawa M., Godlewska J., Gorski A. Hoc protein regulates the biological effects of T4 phage in mammals. *Arch. Microbiol*. 2007. V. 187. № 6. P. 489-498.
8. De Bono B., Chothia C. Exegesis: a procedure to improve gene predictions and its use to find immunoglobulin superfamily proteins in the human and mouse genomes. *Nucleic Acids Res*. 2003. V. 31. № 21. P. 6096-6103.
9. Fokine A., Islam M.Z., Zhang Z., Bowman V.D., Rao V.B., Rossmann M.G.. Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. *J. Virol*. 2011. V. 85. № 16. P.8141-8148.

10. Fokine A., Leiman P.G., Shneider M.Мю, Ahvazi B., Boeshans K.M., Steven A.C., Black L.W., Mesyanzhinov V.V., Rossmann M.G. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Nat. Acad. Sci. USA*. 2005. V. 102. P. 7163–7168.
11. Fong S., Hamill S.J., Proctor M., Freund S.M., Benian G.M., Chothia C., Bycroft M., Clarke J. Structure and stability of an immunoglobulin superfamily domain from twitchin, a muscle protein of the nematode *Caenorhabditis elegans*. *J. Mol. Biol*. 1996. V. 264. № 3. P. 624-639.
12. Halaby D.M., Poupon A., Mornon J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng*. 1999. V. 12. P. 563–571.
13. 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.
14. Jing H., Takagi J., Liu J.H., Lindgren S., Zhang R.G., Joachimiak A., Wang J.H., Springer T.A. Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure*. 2002. V. 10. P. 1453–1464.
15. *ICTV Virus Taxonomy: 2011 Release (current)*. URL: http://ictvonline.org/virusTaxonomy.asp?version=2011 (дата обращения 12.11.2012).
16. Ishii T., Yamaguchi Y., Yanagida M. Binding of the structural protein Soc to the head shell of bacteriophage T4. *J. Mol. Biol*. 1978. V. 120. P. 533–544.
17. Ishii T., Yanagida M. Molecular organization of the shell of the T-even bacteriophage head. *J. Mol. Biol*. 1975. V. 97. P. 655–660.
18. Ishii T., Yanagida M. The two dispensable structural proteins (Soc and Hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads *in vitro*. *J. Mol. Biol*. 1977. V. 109. P. 487–514.
19. Kadyrov F.A., Shlyapnikov M.G., Kryukov V.M. A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. *FEBS Lett*. 1997. V. 415. № 1. P. 75-80.
20. Lavigne R., Darius P., Summer E.J., Seto D., Mahadevan P., Nilsson A.S., Ackermann H.W., Kropinski A.M. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiol*. 2009. V. 9. Article No. 224.
21. Lawrence J.G., Hatfull G.F., Hendrix R.W. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol*. 2002. V. 184. P. 4891-4905.
22. Rohwer F., Edwards R. The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol*. 2002. V. 184. № 16. P. 4529-4535.
23. Ross P.D., Black L.W., Bisher M.E., Steven A.C. Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. *J. Mol. Biol*. 1985. V. 183. P. 353–364.
24. Sathaliyawala T., Islam M.Z., Li Q., Fokine A., Rossmann M.G., Rao V.B. Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. *Mol. Microbiol*. 2010. V. 77. № 2. P. 444-455.
25. Susskind M.M., Botstein D. Molecular genetics of bacteriophage. *Microbiol. Rev*. 1978. V. 42. P. 385-413.
26. Tamura K., Dudley J., Nei M., Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol*. 2007. V. 24. P. 1596–1599.
27. Teichmann S.A., Chothia C. Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. *J. Mol. Biol*. 2000. V. 296. № 5. P. 1367-1383.
28. Vogel C., Teichmann S.A., Chothia C. The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development*. 2003. V. 130. № 25. P. 6317-6328.

29. Yamaguchi Y., Yanagida M. Head shell protein hoc alters the surface charge of bacteriophage T4. Composite slab gel electrophoresis of phage T4 and related particles. *J. Mol. Biol.* 1980. V. 141. P. 175–193.

Материал поступил в редакцию 11.10.2012, опубликован 26.11.2012.