

Федеральное государственное бюджетное учреждение науки
Институт биоорганической химии
им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук

На правах рукописи

Евсеев Петр Владимирович

**Биоинформатические подходы
к таксономической классификации бактериофагов**

1.5.8 – математическая биология, биоинформатика

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
доктор химических наук
член-корреспондент РАН
Мирошников Константин Анатольевич

Москва — 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
СПИСОК СОКРАЩЕНИЙ.....	15
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	17
1.1. Эволюция представлений о вирусах	17
1.2. Бактериофаги и их практическое значение	24
1.3. Систематика и таксономия вирусов	31
1.4. Таксономия бактериофагов – история развития и современная классификационная схема	40
1.5. Современные критерии таксономических рангов бактериофагов и специализированные биоинформатические методы и инструменты	45
1.5.1. Инструменты анализа межгеномного нуклеотидного сходства	46
1.5.2. Инструменты анализа протеомного сходства	49
1.5.3. Инструменты филогенетического анализа	54
1.5.4. Инструменты анализа сети родства фагов	56
1.6. Проблемы таксономии бактериофагов и задачи биоинформатики	57
ГЛАВА 2. ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ.....	59
2.1. Объекты исследования	59
2.2. Сбор данных	59
2.3. Методы исследования.....	59
2.3.1. Протокол таксономического анализа.....	59
2.3.2. Аннотирование фаговых геномов	60
2.3.3. Сравнение нуклеотидного сходства фаговых геномов	61
2.3.4. Сравнение расположения генов и геномных модулей.....	61
2.3.5. Выравнивания последовательностей	61

2.3.6. Филогенетический анализ	62
2.3.7. Анализ протеома	62
2.3.8. Предсказание третичной структуры белков.....	63
2.3.9. Сравнение структур белков.....	63
2.3.10. Поиск и аннотация профаговых последовательностей.....	63
2.3.11. Рабочие базы данных	64
ГЛАВА 3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ	65
3.1. Таксономический анализ бактериофагов <i>Pectobacterium</i>	65
3.1.1. Фаги семейства <i>Autographiviridae</i>	65
3.1.2. Фаги семейства <i>Schitoviridae</i>	74
3.1.3. Фаги, инфицирующие фитопатогенные бактерии <i>Pectobacterium</i> и <i>Dickeya</i>	80
3.2. Таксономический анализ бактериофагов <i>Curtobacterium</i>	85
3.2.1. Фаг <i>Curtobacterium</i> Аука.....	85
3.2.2. Профаги <i>Curtobacterium</i>	92
3.3. Таксономический анализ бактериофага <i>Pseudomonas</i> MD8.....	103
3.3.1. Геномный и филогенетический анализ.....	103
3.3.2. Анализ мозаичности генома.....	115
3.4. Использование предсказаний структуры фаговых белков для эволюционной таксономии.....	122
3.4.1. Эволюция чехольных белков	122
3.4.2. Использование структурных предсказаний для классификации таксонов высокого ранга.....	141
ЗАКЛЮЧЕНИЕ	169
ВЫВОДЫ	170

СПИСОК ЛИТЕРАТУРЫ.....	172
ПРИЛОЖЕНИЕ А	196
ПРИЛОЖЕНИЕ Б.....	200
ПРИЛОЖЕНИЕ В	201
ПРИЛОЖЕНИЕ Г	209
ПРИЛОЖЕНИЕ Д	212
ПРИЛОЖЕНИЕ Е.....	224

ВВЕДЕНИЕ

Актуальность темы исследования и степень ее разработанности. Бактериофаги (сокращённо – «фаги») – это вирусы, инфицирующие бактерии. Бактериофаги вездесущи – они обитают в воде, почве, во множестве живых организмов [1]. По некоторым оценкам, в течение всего лишь одной секунды фаги вызывают порядка 10^{18} успешных заражений бактериальных клеток [2]. Общее количество бактериофагов можно оценить в 10^{31} вирусных частиц, что больше количества клеток в 10–100 раз [3]. Общая масса этих частиц составляет порядка триллиона тонн [4]. Фаги также являются участниками микробиомов растений и животных, в том числе, человека – например, его желудочно-кишечный тракт содержит более 10^{12} фаговых вирионов [5].

Способность бактериофагов уничтожать клетки патогенных бактерий привлекла внимание учёных ещё в начале XX века. В последние десятилетия интерес к лечению бактериофагами стал расти, в первую очередь, из-за распространения устойчивости к антибиотикам. Фаговая терапия обладает важными преимуществами [6], в том числе устойчивым бактерицидным действием, «автодозированием», заключающимся в том, что количество фагов положительно коррелирует с количеством бактерий-хозяев. Фагам присуща низкая собственная токсичность, фаговую терапию отличает минимальное нарушение нормальной флоры и отсутствие перекрестной устойчивости с антибиотиками.

Рост интереса к изучению бактериофагов и их применению требует совершенствования методов характеристики фагов, включая таксономическую классификацию. Исторически фаги классифицировались в соответствии с морфологией частиц, но во времена первых классификационных схем, описывающих бактериофаги, ещё не существовало ПЦР, секвенирования и многих молекулярных методов, известных нам сегодня [7,8]. Первая распространённая схема таксономической классификации наиболее многочисленных изученных

представителей фагов, хвостатых фагов, относящихся к классу *Caudoviricetes*, была предложена Дэвидом Брэдли в 1967 году [9,10]. На основе данных электронной микроскопии бактериофаги классифицировались по трём морфотипам: морфотип А (фаги с сокращающимся хвостом), морфотип В (фаги с длинным, несокращающимся хвостом), морфотип С (фаги с коротким несокращающимся хвостом). Важную роль в развитии систем классификации вирусов, в том числе бактериальных, сыграла Балтиморская классификация – система, представленная Д. Балтимором в 1971 году [11]. Балтиморские классы были установлены на основе типа нуклеиновой кислоты, включенной в вирионы, и, следовательно, на типе репродукции вируса. Классификация бактериофагов долгое время строилась в соответствии с Балтиморской системой и морфологией фагов.

Совершенствование методов молекулярной биологии, секвенирование фаговых геномов и бурное развитие биоинформатики в начале 2000-х годов выявило гораздо более высокое геномное разнообразие, чем считалось ранее, особенно у хвостатых бактериофагов. Последние годы ознаменовались взрывным ростом количества новых таксонов бактериофагов. В 2021 году была принята новая схема таксономической классификации хвостатых бактериофагов, ликвидировавшая семейства, определения которых были основаны на морфологических характеристиках [12]. Все хвостатые фаги были отнесены к классу *Caudoviricetes*, в свою очередь, принадлежащего типу *Uroviricota* царства *Heunggongvirae* реалма *Duplodnaviria*. Было образовано три отряда, включающие вирусы архей; остальные хвостатые бактериофаги были отнесены к 35 семействам, 37 подсемействам и 493 родам, которые приписаны к классу *Caudoviricetes* напрямую, без отнесения к таксонам промежуточного ранга. Тем не менее, большинство из более чем двадцати тысяч известных науке бактериофагов, геномы которых находятся в базе данных NCBI GenBank, на настоящий момент не классифицировано, что отчасти объясняется трудностями обоснованного таксономического описания фагов.

Современная таксономическая классификация бактериофагов требует привлечения целого комплекса биоинформатических методов, включая анализ состава и организации генома, кластеризацию на основе средненуклеотидного геномного сходства, филогенетический анализ с использованием последовательностей консервативных генов и белков, анализ протеома. Определение возможной таксономической классификации может затрудняться из-за частых генетических обменов, которые могут включать как отдельные гены, так и их группы. Другая проблема биоинформатических классификационных подходов – это быстрая эволюция вирусных белков, затрудняющая создание иерархической кластеризации, особенно для высокоранговых таксонов.

Задача определения таксономического положения фагов имеет как практический смысл, позволяя предсказывать биологические свойства бактериофагов (в том числе, для целей фаговой терапии), так и теоретическое фундаментальное значение для эволюционной биологии и вирусологии. Эта задача особенно актуальна для малоисследованных и новых бактериофагов, инфицирующих микроорганизмы, которые являются патогенами человека и животных, либо представляющих другой интерес для экономики, например, в качестве перспективных средств биоконтроля патогенов сельскохозяйственных культур.

Цель работы:

Проанализировать применимость биоинформатических методов для таксономического описания бактериофагов на примере новых бактериофагов, инфицирующих патогены растений *Curtobacterium* и *Pectobacterium*, а также на примере мало изученного фага *Pseudomonas* MD8.

Задачи исследования:

1. Предложить обоснованную таксономическую классификацию на основе геномных данных для новых бактериофагов, инфицирующих бактерии родов *Curtobacterium* и *Pectobacterium*.
2. Оценить таксономическое разнообразие бактериофагов, инфицирующих бактерии семейства *Pectobacteriaceae*, которые являются патогенами сельскохозяйственных культур.
3. Проанализировать профаговые области фитопатогенных бактерий рода *Curtobacterium* с целью классификации потенциальных умеренных фагов и поиска генов белков, способных к разрушению клеточных оболочек этих бактерий.
4. Проанализировать применимость биоинформатических подходов для таксономической классификации умеренного бактериофага *Pseudomonas* MD8.
5. Проанализировать применимость биоинформатических подходов, основанных на сравнении структурного сходства белков, моделированных с помощью современных методов глубокого обучения, для выявления эволюционных взаимосвязей бактериофагов и целях построения высокоранговой таксономической иерархии.

Объектом исследования являлись литические и умеренные неклассифицированные бактериофаги, инфицирующие бактерии, относящиеся к таксонам *Curtobacterium*, *Pectobacterium* и *Pseudomonas*.

Предметом исследования являлись разнообразные биоинформатические методы, позволяющие классифицировать изучаемые бактериофаги.

Научная новизна работы. Впервые проведён таксономический и геномный анализ новых бактериофагов, инфицирующих грамположительные бактерии рода *Curtobacterium* (фаг Аука, геномные области профагового происхождения) и грамотрицательные бактерии рода *Pectobacterium* (Horatius, Possum, PP47, PP81,

Q19) с использованием биоинформатических методов. Проанализировано таксономическое разнообразие бактериофагов, инфицирующих бактерии семейства *Pectobacteriaceae*, вызывающих мягкую гниль сельскохозяйственных культур. На примере фага *Pseudomonas* MD8 впервые детально показан процесс мозаичного формирования геномов умеренных фагов псевдомонад, обсуждены трудности, возникающие при таксономической классификации, и предложено их возможное решение. С использованием предсказанных структур фаговых белков методами глубокого обучения проанализированы возможности улучшения описания функций фаговых белков и эволюционных взаимосвязей между бактериофагами.

Теоретическая и практическая значимость работы. Результаты диссертационной работы расширяют представления о таксономическом разнообразии бактериофагов, в том числе, фагов, инфицирующих бактерии, являющиеся опасными для человека и экономически важных сельскохозяйственных культур. Полученные в ходе работы аннотированные геномные последовательности депонированы в международной базе данных NCBI GenBank.

Предложенные подходы к биоинформатическому анализу бактериофагов и методы решения проблем, вызванных генетическим мозаицизмом и быстрой эволюцией вирусных белков, могут быть использованы для построения более точной таксономической иерархии.

Методология и методы исследования. Автором выполнены анализ отечественной и зарубежной литературы по теме исследования, осуществлён сбор данных и их обработка с использованием биоинформатических методов, включая сборку, аннотацию и анализ геномов, филогенетический анализ, анализ геномного сходства, предсказание структур и функций фаговых белков. Полученные результаты были проанализированы и систематизированы, написаны все главы диссертации, сформулированы выводы и практические рекомендации.

Положения, выносимые на защиту:

1. Биоинформатические методы позволяют уверенно классифицировать бактериофаги *Pectobacterium* PP47, PP81, Q19 и предложить классификационную схему на уровне рода, подсемейства, семейства.
2. Использование биоинформатических методов для таксономической классификации нового фага *Curtobacterium* Аука позволяет предложить классифицировать этот фаг как представителя нового вирусного семейства или подсемейства.
3. Использование биоинформатических методов для геномного анализа исследованных умеренных профагов может быть затруднено в связи с ярко выраженным генетическим мозаицизмом этих фагов.
4. Новые алгоритмы структурного моделирования белков могут быть использованы в целях таксономической классификации.

Степень достоверности и апробация результатов. Данные, представленные в работе, получены с использованием современных программ и программных пакетов. Обзор литературы и обсуждение подготовлены с использованием актуальных литературных источников. Достоверность полученных результатов определяется достаточным объемом проведенных исследований, использованием в работе современных экспериментальных и биоинформатических методов. Достоверность результатов также подтверждается публикациями в рецензируемых отечественных и международных журналах, депонированием генетических последовательностей в международную базу данных NCBI GenBank. Результаты диссертации были представлены на следующих международных и российских научных конференциях: «The Future Applications of Bacteriophages» (Зевейл, Египет, 2021), «Bioinformatics: from algorithms to applications (BIATA)» (С.-Петербург, Россия, 2021), «IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB)» (Новосибирск, Россия, 2021), «24th Evergreen Phage Meeting» (Олимпия,

США, 2021), III Всероссийской конференции «Высокопроизводительное секвенирование в геномике» (Новосибирск, Россия, 2022).

Личный вклад автора заключается в постановке цели исследования и задач, анализе литературных данных, получении данных и обработке полученных результатов, подготовке публикаций и научных докладов. Соискателем был проведен всесторонний биоинформатический анализ геномных данных, включая детальную аннотацию фаговых геномов. Все этапы биоинформатического анализа от сборки геномов до обработки результатов вычислений и выбора задействованных биоинформатических алгоритмов осуществлялись соискателем лично. Все биоинформатические иллюстрации в диссертации были сделаны соискателем. Соискатель также принимал участие в постановке задач и планировании экспериментов (в частности, в экспериментах по индукции профагов и характеристике фагов). Соискателем написаны все главы диссертации, сформулированы выводы и практические рекомендации. Электронная микроскопия выполнена Е.Е. Куликовым (Московский физико-технический институт, Институт микробиологии им. С.Н. Виноградского РАН) и В.А. Кадыковым (Институт биоорганической химии им. академиком М.М. Шемякина и Ю.А. Овчинникова).

Соответствие диссертации паспорту специальности. Представленные в диссертации результаты принадлежат областям исследования «исследование эволюции живой природы с помощью средств информатики и математики» и «компьютерная генетика». Диссертация соответствует паспорту специальности 1.5.8 – математическая биология, биоинформатика.

Объем работы. Работа состоит из следующих разделов: «Введение», «Список сокращений», «Обзор литературы», «Объекты и методы исследования», «Результаты и их обсуждение», «Заключение», «Выводы», и «Список литературы». Работа изложена на 246 страницах, содержит 6 таблиц, 103 рисунка, 6 приложений.

Список литературы включает 278 источников, из которых 3 на русском языке, 275 на иностранных языках и 6 интернет-ресурсов.

Публикации. По теме диссертации опубликовано 9 научных работ, среди них 8 статей в рецензируемых журналах, включённых в системы цитирования Scopus, Web of Science и RSCI, рекомендованных для защиты в диссертационном совете МГУ имени М.В. Ломоносова, и 1 тезис конференций. В статьях, опубликованных в соавторстве, автором сделаны все биоинформатические исследования.

Статьи в рецензируемых журналах Scopus, Web of Science и RSCI¹:

1. **Evseev P.V.**, Lukianova A.A., Shneider M.M., Korzhenkov A.A., Bugaeva E.N., Kabanova A.P., Miroshnikov K.K., Kulikov E.E., Toshchakov S.V., Ignatov A.N., Miroshnikov K.A. Origin and Evolution of *Studiervirinae* Bacteriophages Infecting *Pectobacterium*: Horizontal Transfer Assists Adaptation to New Niches // *Microorganisms*, 2020, Vol. 8, No. 11, P. 1707. IF 4,782, (1,64/1,15).
2. Miroshnikov K.A., **Evseev P.V.**, Lukianova A.A., Ignatov A.N. Tailed Lytic Bacteriophages of Soft Rot *Pectobacteriaceae*. *Microorganisms* // *Microorganisms*, 2021, Vol. 9, No. 9, P. 1819. IF 4,782, (2,26/0,90).
3. **Evseev P.**, Lukianova A., Sykilinda N., Gorshkova A., Bondar A., Shneider M., Kabilov M., Drucker V., Miroshnikov K. *Pseudomonas* Phage MD8: Genetic Mosaicism and Challenges of Taxonomic Classification of Lambdoid Bacteriophages // *International Journal of Molecular Sciences*, 2021, Vol. 22, No. 19, P. 10350. IF 6,009, (2,03/1,42).
4. **Evseev P.**, Shneider M., Miroshnikov K. Evolution of Phage Tail Sheath Protein // *Viruses*, 2022, Vol. 14, No. 6, P. 1148. IF 5,712, (2,29/1,61).
5. Lukianova A.A., **Evseev P.V.**, Shneider M.M., Dvoryakova E.A., Tokmakova A.D., Shpirt A.M., Kabilov M.R., Obraztsova E.A., Shashkov A.S., Ignatov A.N., Knirel Y.A., Dzhililov F.S.-U., Miroshnikov K.A. *Pectobacterium* versatile Bacteriophage

¹ В скобках приведен объем публикации в условных печатных листах и вклад автора в условных печатных листах

- Possum: A Complex Polysaccharide-Deacetylating Tail Fiber as a Tool for Host Recognition in Pectobacterial *Schitoviridae* // International Journal of Molecular Sciences, 2022, Vol. 23, No. 19, P. 11043. IF 6,009, (1,12/0,34).
6. Tarakanov R.I., Lukianova A.A., **Evseev P.V.**, Pilik R.I., Tokmakova A.D., Kulikov E.E., Toshchakov S.V., Ignatov A.N., Dzhililov F.S.-U., Miroshnikov K.A. Ayka, a Novel Curtobacterium Bacteriophage, Provides Protection against Soybean Bacterial Wilt and Tan Spot // International Journal of Molecular Sciences, 2022, Vol. 23, No. 18, P. 10913. IF 6,009, (1,11/0,33).
 7. **Evseev P.**, Lukianova A., Tarakanov R., Tokmakova A., Popova A., Kulikov E., Shneider M., Ignatov A., Miroshnikov K. Prophage-Derived Regions in Curtobacterium Genomes: Good Things, Small Packages // International Journal of Molecular Sciences 2023, Vol. 24, No. 2, P. 1586. IF 6,009, (1,78/1,25).
 8. **Evseev P.**, Gutnik D., Shneider M., Miroshnikov K. Use of an Integrated Approach Involving AlphaFold Predictions for the Evolutionary Taxonomy of Duplodnaviria Viruses // Biomolecules 2023, Vol. 13, No. 1, P. 110. IF 5,880, (1,97/1,38).

В материалах конференций:

1. **Евсеев П.В.**, Лукьянова А.В., Токмакова А.Д., Шнейдер М.М., Игнатов А.Н., Попова А.В., Мирошников К.А. Профаговые области в геномах *Curtobacterium* spp. и *Curtobacterium flaccumfaciens* pv. *flaccumfaciens*: геномика и белки, разрушающие клеточную стенку//Сборник тезисов III Всероссийской конференции «Высокопроизводительное секвенирование в геномике (HSG-2022)» Новосибирск, 2022. С. 48. (0,10/0,07).

Благодарности. Автор выражает искреннюю благодарность и глубокую признательность за неоценимую помощь и всестороннюю поддержку при выполнении работы своему научному руководителю – доктору химических наук, члену-корреспонденту Российской академии наук Константину Анатольевичу Мирошникову.

Автор благодарит Нину Николаевну Сыкилинду, Александра Николаевича Игнатова, Михаила Марковича Шнейдера, Сергея Константиновича Комаревцева,

Дарью Игоревну Гутник, Кирилла Константиновича Мирошникова, Анастасию Владимировну Попову, Андрея Викторовича Летарова, за консультации и помощь в подготовке диссертации, Анну Александровну Лукьянову, Анну Дмитриевну Токмакову, Рашита Ислямовича Тараканова, Анну Сергеевну Горшкову, Евгения Евгеньевича Куликова, Марсея Расимовича Кабилова, Анастасию Петровну Кабанову за экспериментальную работу, результаты которой использованы для биоинформатических исследований, а также всех сотрудников лаборатории молекулярной биоинженерии Института биоорганической химии РАН и сотрудников лаборатории водной микробиологии Лимнологического института СО РАН.

СПИСОК СОКРАЩЕНИЙ

а.о. – аминокислотные остатки

вРНКП – вирионная ДНК-зависимая РНК-полимераза

джамбо-фаг – фаг с размером генома более 200 тыс. н.п.

ДНКП – ДНК-полимераза

дцДНК – двухцепочечная ДНК

ГДБ – гликополимер-деградирующих белки

ГКБ – главный капсидный белок

н.п. – нуклеотидные пары

оцДНК – одноцепочечная ДНК

РНКП – ДНК-зависимая РНК-полимераза

РСБ – рецептор-связывающий белок

ЧБ – чехольные белки хвоста фагов миовирусной морфологии

AF2 – AlphaFold 2

AFP – anti-feeding prophage system

ANI – average nucleotide identity, средненегеномное нуклеотидное сходство

ATPD – АТФазный домен большой субъединицы терминазы

BLAST – Basic Local Alignment Search Tool

CDS – coding sequence, кодирующая последовательность

CIS – contractile injection system, сократительная система инъекции

ORF – open reading frame, открытая рамка считывания

HMM – hidden Markov model, скрытая марковская модель

ICTV – International Committee on Taxonomy of Viruses, Международный комитет по таксономии вирусов

IDDT – local Distance Difference Test

NCBI – National Center for Biotechnology Information, Национальный центр биотехнологической информации США

PD – фаги инфицирующие фитопатогенные бактерии *Pectobacterium* и *Dickeya*

PDB – protein databank, база данных белковых структур

pRNA – packaging RNA, упаковочная РНК

RMSD – root-mean-square deviation

SRP – soft rot Pectobacteriaceae, мягкая гниль Pectobacteriaceae

TerL – large subunit of terminase, большая субъединицы терминазы

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Эволюция представлений о вирусах

Вирус – это инфекционный агент, способный размножаться (реплицироваться) только в организме хозяина. Вирусы могут инфицировать различные живые организмы, включая прокариот, растения и животных. По сравнению с клеточными организмами вирусы имеют более простую структуру. В свободном состоянии вирусная частица состоит из вирусного генетического материала (генома), заключенного в белковую оболочку, называемую капсидом. У некоторых вирусов белковая оболочка окружена липидной мембраной. Вирусные геномы очень разнообразны, они могут быть представлены в виде ДНК или РНК, в одноцепочечной или двухцепочечной форме. Геном вирусов может быть как линейным, так и кольцевым, а также различаться по длине и количеству молекул ДНК или РНК [13].

Процесс репликации вируса начинается, когда вирус инфицирует своего хозяина, прикрепляясь к клетке и проникая через клеточную стенку или мембрану. После этого геном вируса покидает капсид и происходит перехват контроля механизмов жизнедеятельности клетки-хозяина, в результате чего молекулярные механизмы клетки начинают реплицировать вирусный геном и производить вирусные белки для создания новых капсидов. Затем происходит сборка дочернего поколения вирусов. Новые вирусы покидают клетку-хозяин путем экзоцитоза или за счет процесса, называемого лизисом, который разрушает клетку-хозяин. Некоторые вирусы, называемые оболочечными, приобретают часть мембраны хозяина, которая формирует оболочку вокруг капсида. Новые вирусы могут заражать новых хозяев.

В открытие вирусов внесли свой вклад многие учёные, но первооткрывателем вирусов считается Дмитрий Иосифович Ивановский [14]. В

1887 году, будучи еще студентом, он начал работу над болезнью табачной мозаики (рис. 1), которая привела к первому открытию вируса. Болезнь, поражала растения табака на ранних этапах их цикла роста и была распространена на растениях в Голландии, вызывая зелено-коричневый мозаичный узор на пораженных листьях. Д. И. Ивановский начал свои исследования с повторения и проверки экспериментов, проведенных Адольфом Майером, в которых он брал больные листья, отбирал их сок, после чего вводил его в здоровые растения [15]. В результате этих опытов заражалось до 80% здоровых растений.

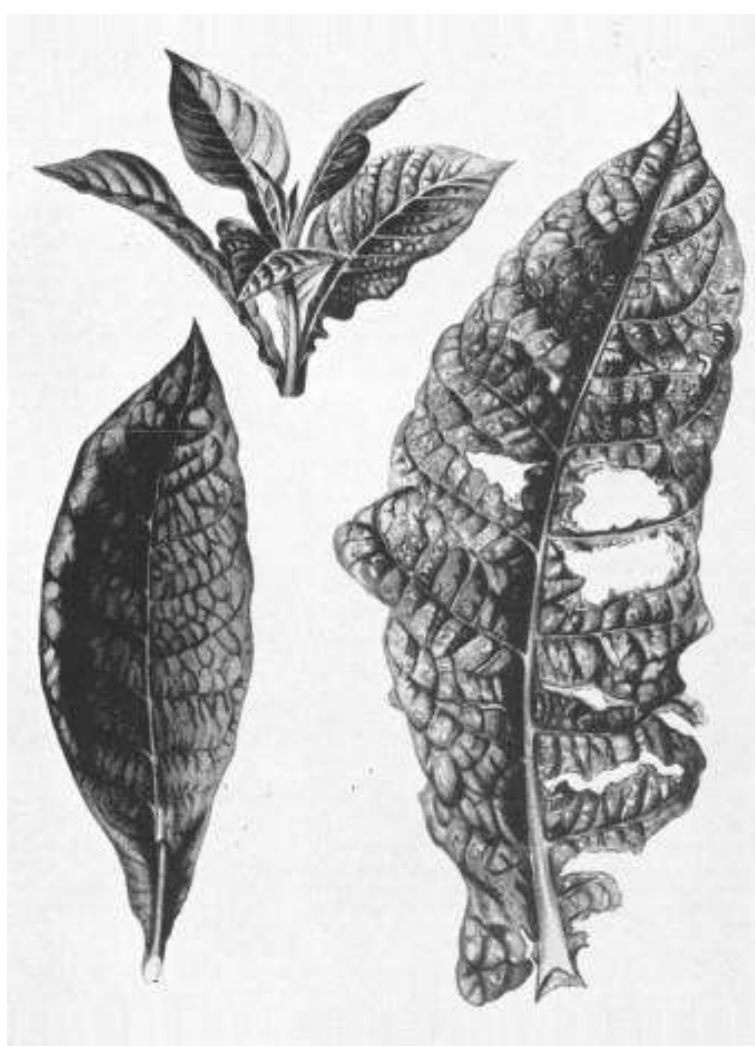


Рисунок 1. Листья табака, поражённые вирусом табачной мозаики. Оригинальное изображение 1842 г., опубликованное в [15].

Ивановский начал расширять этот эксперимент. Он обнаружил, что табачные растения заражаются болезнью при внесении суспензии зараженных мертвых

листьев в почву, но не в том случае, если больные растения выращиваются рядом со здоровыми. Первоначально Ивановский думал, что болезнь является бактериальной, и поэтому разработал эксперимент, в котором сок больного табака фильтровали через свечу Шамберлена (трубка под давлением с керамическим/фарфоровым фильтром, разработанным для фильтрации бактерий). Профильтровав сок, Ивановский впрыскивал его в здоровые листья табака. Как только здоровые растения начали проявлять признаки инфекции, Ивановский увидел, что бактериальная фильтрация болезнетворного сока не предотвращает болезнь, и, следовательно, инфекционный агент должен был отличаться от любых бактерий, которые были известны раньше. Кроме того, пытаясь вырастить таинственный агент на бактериальном пищевом геле (агаре), Ивановский получил доказательства того, что агент, заражающий растения табака, был скорее частицей, чем жидкостью. Он пришел к выводу, что болезнь, скорее всего, была вызвана либо живым существом, либо большой молекулой, которая не проходила через почву или легко проникала в клетки [15].

Ивановский представил исследования болезни табачной мозаики в своей диссертации «О мозаичной болезни табачного растения» (1902 г.). Им были сделаны следующие выводы:

1. Сок больных растений был заразен.
2. При нагревании зараженный сок перестает быть заразным (тепло денатурирует РНК и белки вирусов, убивая их).
3. Учитывая отсутствие грибов и других паразитов, заболевание может быть вызвано какими-то бактериями.

То есть, Ивановский считал, что болезнь вызывается какой-то микробактерией или токсином. В 1898 году Мартин Бейеринк (1851-1931), независимо от Ивановского, повторил его эксперименты и пришёл к похожим выводам. Бейеринк утверждал, что вирус был в некоторой степени жидким по

своей природе, назвав его «*contagium vivum Fluidum*» (заразная живая жидкость) [16].

Представления о том, что такое вирусы, менялось по мере накопления знаний о них. Поначалу, после открытия вируса табачной мозаики, вирусы считались просто крошечными микроорганизмами, которые не могут расти на искусственных средах. При этом понимания, что вирусы разных групп организмов представляют собой один тип живой материи, не существовало ещё достаточно продолжительное время, а разделы вирусологии, описывающие вирусы бактерий, животных и растений, развивались независимо друг от друга. Представление о вирусах разных групп организмов, как об одном виде живой материи сложилось в конце 1920-х – начале 1930-х годов, когда эти болезнетворные агенты стали называть «фильтрующимися вирусами», «ультравирუსами», а затем просто «вирусами». Развитие вирусологии привело к пониманию того, что вирусы, как и клеточные организмы, способны размножаться, эволюционировать, как и живые клетки, обладая наследственностью и изменчивостью [17].

В 1936 году удалось кристаллизовать вирус табачной мозаики [18], что позволило получить его детальную структуру с помощью рентгеновских исследований (рис. 2) [19]. В течение нескольких последующих десятилетий стали развиваться и превалировать представления о вирусах как об организмах, а о вирионе, содержащем вирусный геном, – как об индивидууме со своими особенностями, нехарактерными для клеточных организмов. К числу особенностей вирусов, отличавших их от других организмов, считалось, в том числе, наличие только одного из двух типов нуклеиновых кислот – ДНК либо РНК, что, как стало ясно впоследствии, было ошибкой. Во время своего жизненного цикла вирусы, в том числе, с геномом, состоящим из ДНК, задействуют клеточный аппарат синтеза белков, использующий информационные (матричные) рибонуклеиновые кислоты (мРНК), которые содержат информацию об аминокислотной последовательности синтезируемого белка. Интересно, что РНК может использоваться ДНК-содержащими вирусами не только в качестве информационной. У некоторых

небольших бактериофагов, принадлежащих к семейству *Tectiviridae* реалма *Varidnaviria* с геномом в виде двухцепочечной ДНК, в составе малой субъединицы фермента терминазы, который участвует в упаковке генома фага в капсид, тоже присутствует РНК [20]. Также было показано, что некоторые вирусы с геномом, кодируемым ДНК, например, вирус простого герпеса, содержат в своём вирионе небольшое количество РНК [21], что тоже не согласуется с определением вирусов как организмов, использующим только один тип нуклеиновых кислот. Были открыты ретровирусы, к числу которых относят вирус человеческого иммунодефицита и многие онковирусы, содержащие в своём составе геном, кодируемый одноцепочечной РНК, при этом в процессе репликации этот геном копируется в ДНК с помощью фермента обратной транскриптазы [22].

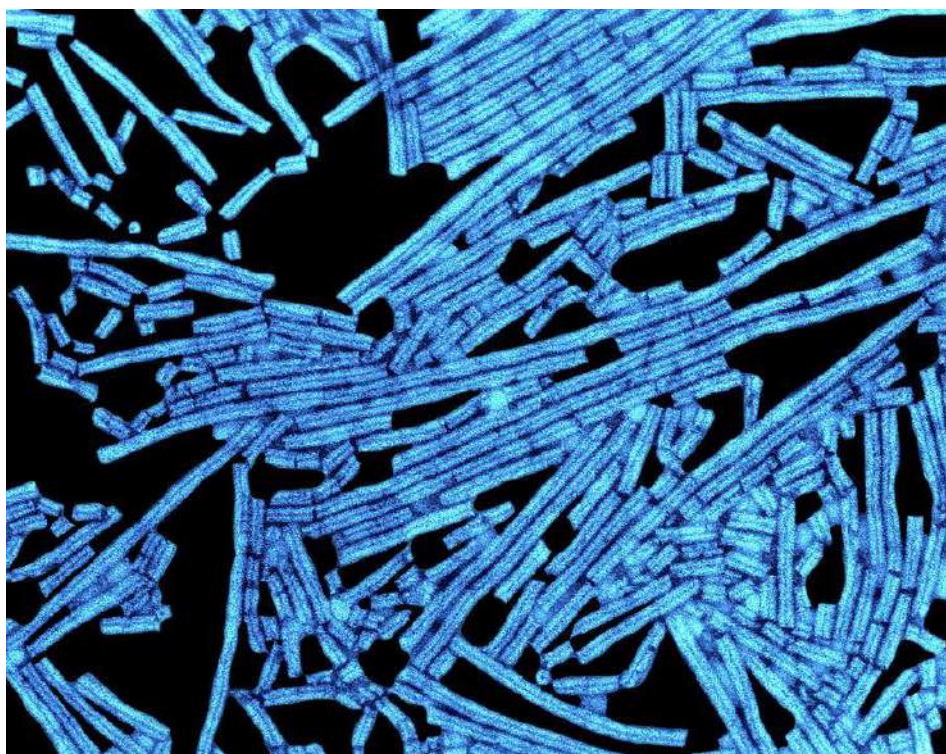


Рисунок 2. Вирус табачной мозаики, микрофотография Денниса Кункеля [23].

Тем не менее, между вирусами и клеточными организмами есть принципиальное отличие – это отсутствие собственного действующего аппарата синтеза белка (трансляции), который декодирует информацию, содержащуюся в РНК, синтезируя белки. Хотя некоторые крупные небактериальные вирусы и

бактериофаги кодируют отдельные белки, участвующие в трансляции [24,25], ни один вирусный геном не содержит всех генов, необходимых для полноценного функционирования трансляционного аппарата.

Другое принципиальное различие между вирусами и клеточными организмами – это способ размножения. В основе всех известных способов размножения клеточных организмов, включая деление прокариотических клеток, amitoz, митоз и мейоз у эукариот, размножение спорами, почкование, вегетативное размножение и фрагментация, лежит деление клетки. В отличие от клеток, белковые вирионы формируются путём самосборки. Этот процесс может быть более или менее сложным. Например, вирионы хвостатых фагов миовирусного морфотипа образуются из собирающихся по отдельности капсида и хвоста, состоящих, в свою очередь, из белковых субъединиц капсидных белков, белков базальной пластинки, белков хвостовой трубки и рецептор-связывающих белков хвостовых фибрилл [26,27], а вирионы нитчатых фагов формируются с помощью самосборки белковых субъединиц вокруг одноцепочечной геномной ДНК. В отличие от вирусов, клетка к самосборке не способна.

Некоторые вирусы обладают такой интересной характеристикой и особенностью, как мультипартитность. Мультипартитность – это размещение сегментов вирусного генома в разных частицах. При этом вирус становится полнофункциональным только при инфицировании клетки всеми частями. К числу мультипартитных вирусов относятся РНК-вирусы растений, принадлежащие к семействам *Bromoviridae*, *Tombusviridae*, *Virgaviridae* и другим таксонам, геном которых размещён в двух-трёх частицах [28]. Некоторые вирусы насекомых переносят свой геном в составе ещё большего количества вирионов, вплоть до двадцати восьми [17].

По мере изучения структуры вирусных частиц и механизмов, с помощью которых они производятся в клетках организма-хозяина, появлялись все более

точные определения вирусов. Современные свойства вирусов формулируются следующим образом:

- Вирус – это инфекционный облигатный внутриклеточный паразит.
- Вирусный геном состоит из ДНК или РНК.
- Вирусный геном направляет синтез вирусных компонентов клеточными системами в соответствующей клетке-хозяине.
- Инфекционные потомки вирусных частиц, называемые вирионами, образуются путем самосборки *de novo* из вновь синтезированных компонентов.
- Вирион потомства, собранный во время инфекционного цикла, является средством передачи вирусного генома в следующую клетку-хозяин, где происходит инициация следующего инфекционного цикла.

При том, что у вирусов отсутствуют сложные энергогенерирующие и биосинтетические системы, необходимые для самостоятельного существования, они, тем не менее, не являются самыми простыми биологически активными агентами: вирионы, которые инфицируют растения, состоят лишь из одной молекулы РНК.

Принадлежность вирусов к живой или неживой природе – причина для длительных дискуссий. Вирусы можно рассматривать как микробы, которые существуют в двух фазах: неживой фазе, вирионе, и фазе размножения в инфицированной клетке. Большинство исследователей считают, что вирусы – это организмы, а «неодушевленные» вирионы можно рассматривать как «споры», которые переходят в «живое» состояние в инфицированных клетках.

1.2. Бактериофаги и их практическое значение

Бактериофаги (сокращённо – «фаги») – это вирусы, инфицирующие бактерии. В отличие от бактерий и некоторых гигантских эукариотических вирусов [29], даже самые большие бактериофаги слишком малы, чтобы увидеть их с помощью обычного светового микроскопа, для изучения морфологии фагов используются различные виды электронной микроскопии (рис. 3, 4) и рентгеноструктурный анализ (рис. 5).

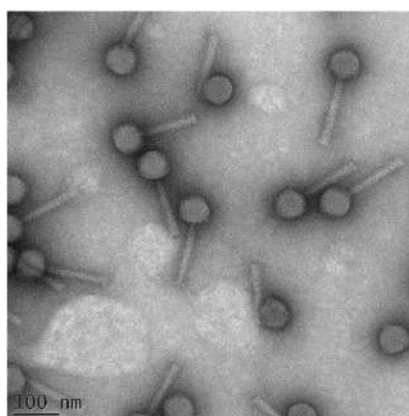


Рисунок 3. Электронная микрофотография хвостатого бактериофага семейства *Chaseviridae Pectobacterium PP101*. Шкала на микрофотографии соответствует 100 нм [30].

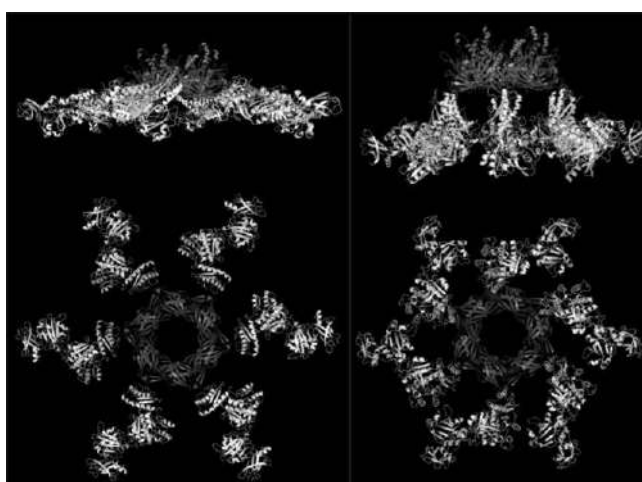


Рисунок 4. Реконструкция части хвоста бактериофага *Escherichia T4*, полученная методом электронной криомикроскопии в сокращённом (слева) и несокращённом (справа) состояниях [27].

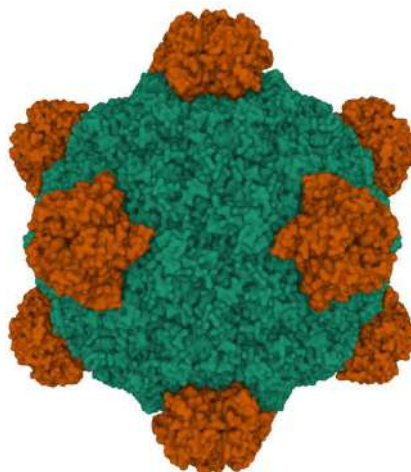


Рисунок 5. Структура прокапсида бактериофага *Escherichia* G4, полученная методом рентгеноструктурного анализа [31].

По состоянию на начало 2023 г., база данных Genome NCBI [32] содержала информацию о геномах более чем тридцати тысяч бактериофагов, большинство которых инфицирует всего лишь один или несколько штаммов одного вида бактерий. Бактериофаги обычно прикрепляются к клетке (адсорбируются с помощью взаимодействия с клеточным рецептором) и разрушают (лизуют) её после успешного размножения генетического материала и синтеза биомолекул, собирающихся в вирион (рис. 6), хотя некоторые из них (нитчатые фаги семейства *Inoviridae*), могут покидать инфицированную клетку без её разрушения [33]. Некоторые бактериофаги, такие как, например, фаг λ [34] способны к встраиванию своего генетического материала в бактериальную хромосому и называются лизогенными или умеренными. В отличие от умеренных, вирулентные, или литические фаги не способны встраивать свой генетический материал в хромосому хозяина, и после размножения внутри клетки разрушают её без промежуточного этапа лизогенного состояния, в дальнейшем заражая новые клетки [35].

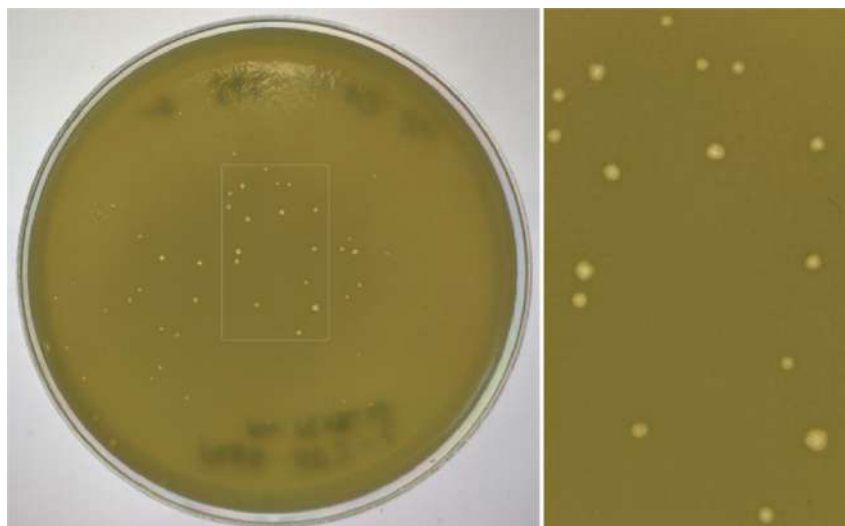


Рисунок 6. Бляшки бактериальных колоний, инфицированных фагом *Pseudomonas* PaBG на газоне *Pseudomonas aeruginosa* PAO1 [36].

Одним из первых опубликованных в научной печати случаев литического действия бактериофагов стало наблюдение Эрнеста Хэнбери Ханкина, английского бактериолога, работавшего в Индии. Ханкин показал, что воды из индийских рек Ганг и Джамна содержат некий биологический компонент, разрушающий холерные вибрионы и проходящий через фарфоровые фильтры, которые задерживают бактерии [37]. В 1914 г. директор Брауновского ветеринарного института и профессор бактериологии Лондонского университета Фредерик Туорт обнаружил, что «чистые» культуры бактерий могут быть как-то связаны с прозрачным материалом, проходящим через фильтры и разрушающим бактерии в гранулы [38]. Он продемонстрировал, что этот действующий агент был способен инфицировать чистые культуры стафилококков, разрушать бактерии и передаваться другим стафилококкам. Туорт описал этот прозрачный материал как фермент, секретируемый бактериями для не вполне понятных целей.

В 1917 г. французский микробиолог Феликс д'Эрелль получил похожие результаты при изучении бактериальной дизентерии. Д'Эрелль смог выделить из стула выздоравливающих больных шигеллезом так называемый «антишигамикроб», который не задерживался фарфоровым фильтром, задерживающим шигеллы. При добавлении этого активного фильтрата к бактериальной культуре

или к эмульсии шигелл рост культуры прекращался, сменяясь лизисом клеток [39]. Д'Эрелль описал действующий агент как микроб, который был «настоящим» микробом иммунитета и облигатным бактериофагом. Он также продемонстрировал активность этого «антишига-микроба», инокулируя лабораторных животных для лечения шигеллеза. Эти эксперименты показали клиническую значимость его открытия.

Вскоре после открытия д'Эрелля, эксперименты с бактериофагами и попытками их применения в качестве антибактериальных агентов, сопровождавшиеся насыщенной дискуссией с д'Эреллем (и, возможно, интригами), проводились бельгийским иммунологом, нобелевским лауреатом и учеником И. И. Мечникова, Жюлем Борде и его коллегой Андре Грация [40,41], однако в ходе этих экспериментов были получены противоречивые результаты. В 1921 г. бельгийские учёные Р. Брюиног и Ж. Майзен описали чёткие доказательства клинического улучшения у пациентов, которые получили инъекции стафилококкового специфического бактериофага для лечения кожных фурункулов и карбункулов, с уменьшением боли, отека и лихорадки в течение 48 часов после введения [42]. В 1931 г. после публикации большого количества нерандомизированных испытаний [40], д'Эрелль обобщил результаты своих многолетних исследований и экспериментов, включая лечение инфекций путём внутривенных инъекций препаратами бактериофагов [43].

После победы Октябрьской революции и окончания Гражданской войны был проведён ряд реформ, направленных на преодоление технологического и научного отставания страны от промышленно развитых государств. В числе мер, направленных на рост научного и технического потенциала, было создание в марте 1921 г. Бюро иностранной науки и техники, а также организация зарубежных командировок советских ученых, многие из которых в дальнейшем стали организаторами и выдающимися деятелями советской науки, в их числе были советский физик Абрам Фёдорович Иоффе и микробиолог Георгий Григорьевич Элиава [44]. Под руководством А.Ф. Иоффе и с помощью видных европейских

учёных в Москве, Ленинграде, Днепропетровске, Харькове и других городах в дальнейшем была создана сеть научно-исследовательских институтов [44]. Г. Г. Элиава, познакомившийся во время работы в парижском Институте Пастера с Ф. д'Эреллем, основал бактериологический институт в Тбилиси в 1923 г., а в дальнейшем, уже совместно с ним, продолжил работу над исследованием бактериофагов и их применением для лечения (фаговой терапии) в этом Институте, до сих пор остающимся единственным в мире, специализирующимся на исследовании бактериофагов [45,46] (рис. 7). Исследования бактериофагов в Советском Союзе, в том числе с целью практического применения, приобрели большой размах [46,47] (рис. 8).

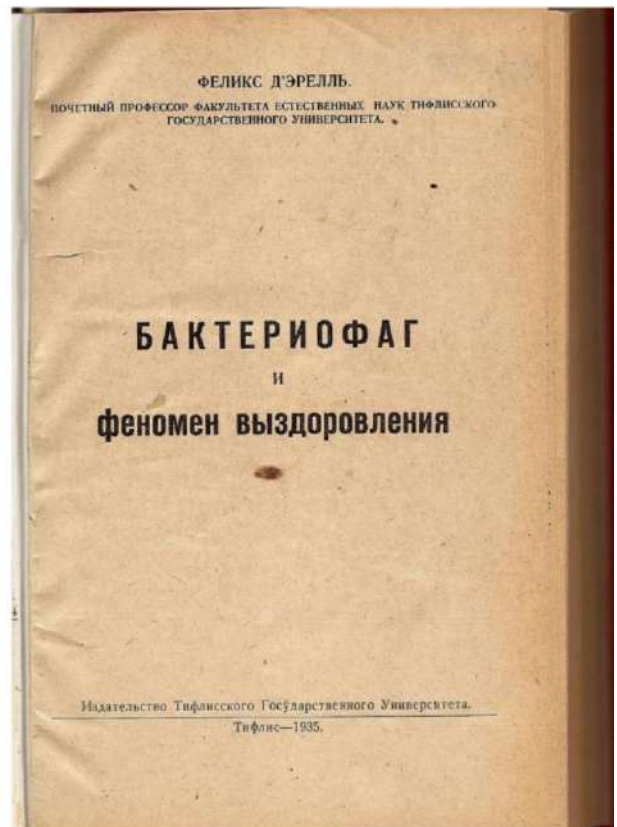


Рисунок 7. Первая страница публикации Ф. д'Эрелля в Журнале Канадской Медицинской Ассоциации и титульная страница книги, опубликованной в 1935 г. в Тбилиси [48].

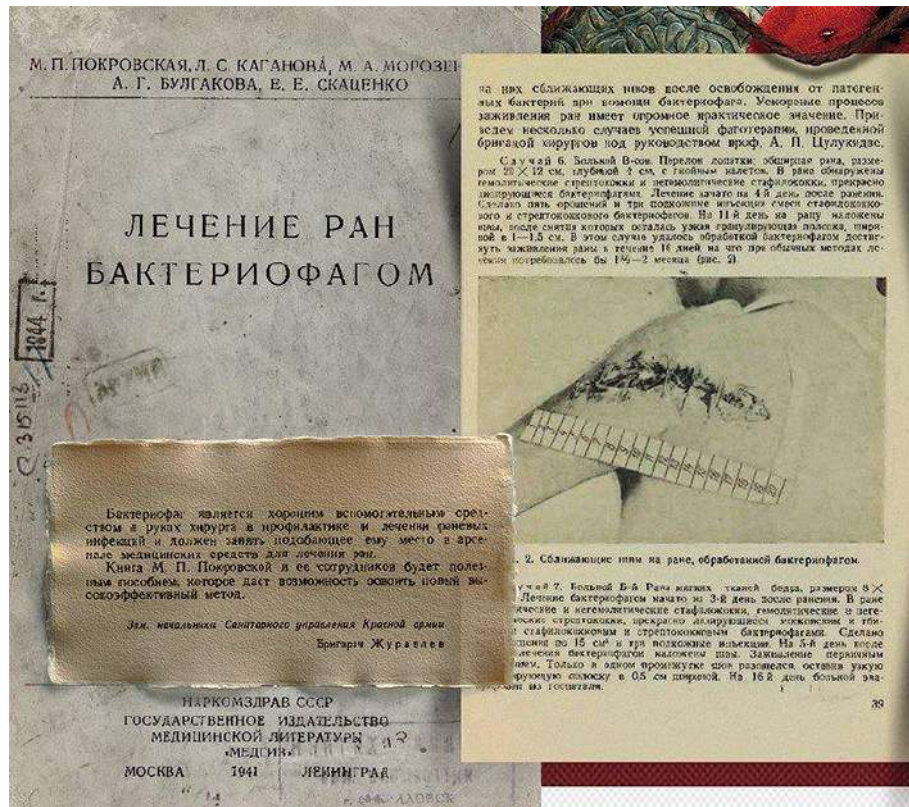


Рисунок 8. Учебное пособие по лечению ран бактериофагами и отзыв на него заместителя начальника Санитарного управления Красной армии [45].

Разработки советских учёных в области фаговой терапии были успешно использованы во время Советско-финской войны 1939-1940 гг. и Великой Отечественной войны 1941-1945 гг., уменьшив потери в живой силе от вызываемых бактериями осложнений после ранений. Во время войны бактериофаги также использовались для лечения дизентерии и профилактики холеры [47–49]. Использование бактериофагов сыграло важную роль во время Сталинградской битвы (1942-1943 гг.), предотвратив начавшуюся на оккупированных немецкими войсками эпидемию холеры [45]. В качестве материала для выделения бактериофагов использовались, в том числе, трупы немецких военнослужащих, похищенные советскими разведчиками [45].

Фаговая терапия, развившаяся и получившая широкое распространение, в первую очередь, в Советском Союзе, не была популярна на Западе, в основном, вследствие политической конфронтации с СССР и, а также из-за неприятия

западными учёными отторжения менделевской генетики школой Т. Д. Лысенко и дискредитировавших себя методов лечения рака Н. Г. Ключевой и И. Г. Роскина [50].

После открытия антибиотиков Александром Флемингом в 1929 г. [51] и их последующего широкого распространения, бактериофаги стали чаще использоваться в качестве моделей в молекулярной биологии [48]. Тем не менее, использование фагов в медицинских целях в Советском Союзе продолжилось [49].

Распространение устойчивости к антибиотикам и успехи молекулярной биологии и вирусологии стали снова привлекать внимание к лечению бактериофагами в последние десятилетия. Фаговая терапия обладает важными преимуществами [6], среди которых:

- **«Автодозирование».** Количество фагов положительно коррелирует с количеством бактерий-хозяев этих фагов.
- **Низкая собственная токсичность.** Поскольку фаги состоят в основном из нуклеиновых кислот и белков, они нетоксичны по своей природе.
- **Минимальное нарушение нормальной флоры.** В отличие от антибиотиков, фаги в минимальной степени воздействуют на полезные для здоровья бактерии нормальной флоры.
- **Отсутствие перекрестной устойчивости с антибиотиками.** Поскольку фаги заражают и убивают патогены, используя механизмы, отличные от механизмов антибиотикорезистентности, специфические механизмы устойчивости к антибиотикам не превращаются в механизмы устойчивости к фагам.
- **Простота выделения.** Фаги против многих патогенных бактерий легко обнаруживаются, часто в сточных водах и других отходах, содержащих высокие концентрации бактерий.

– **Удаление биопленок.** Фаги могут очищать по крайней мере некоторые биопленки, поочередно лизируя бактериальные слои.

Среди недостатков фаговой терапии – узкий диапазон специфичности [52], не всегда высокая эффективность, потенциальное воздействие на иммунную систему [6], дорогостоящее исследование с детальным изучением генома и морфологии [53].

1.3. Систематика и таксономия вирусов

Современная биология немыслима без таксономии, занимающейся систематизацией и классификацией групп биологических объектов на основе общих характеристик. Слово «таксономия» произошло от древнегреческих слов «τάξις» («расположение») и «νομία» («метод»). Основным понятием биологической таксономии является таксон, представляющий собой группу организмов. Таксону присваивается таксономический ранг, объединение таксонов данного ранга формирует группу более высокого ранга. Порядок таксономических групп образует таксономическую иерархию. Основной структурной единицей таксономии является вид, близкие виды объединяются в роды.

Основоположником современной таксономии является шведский ботаник Карл Линней (1707-1778), разработавший ранжированную систему, известную как линнеевская таксономия, классифицировавшую организмы с применением т. н. «биномиальной номенклатуры». Биномиальная биологическая номенклатура использует двухсловное название на латыни, в котором первое слово обозначает род, а второе – вид (или видовой эпитет в ботанике). Оригинальная система Линнея также включала ранги отрядов и классов (рис. 9, [54,55]). Среди учёных, внёсших существенный вклад в биологическую таксономию, можно отметить эволюциониста Жана Батиста Ламарка (1744-1829), Антуана Лорана де Жюссье (1748-1836), предложившего категорию семейства, Жоржа Леопольда Кювье (1769-1832), предложившего категорию типа, Эрнста Генриха Геккеля (1834-1919), предложившего использовать генеалогические деревья для отображения связей

между таксонами и, конечно, Чарлза Роберта Дарвина (1809-1882), предположившего, что иерархия таксонов связана с их происхождением [56].

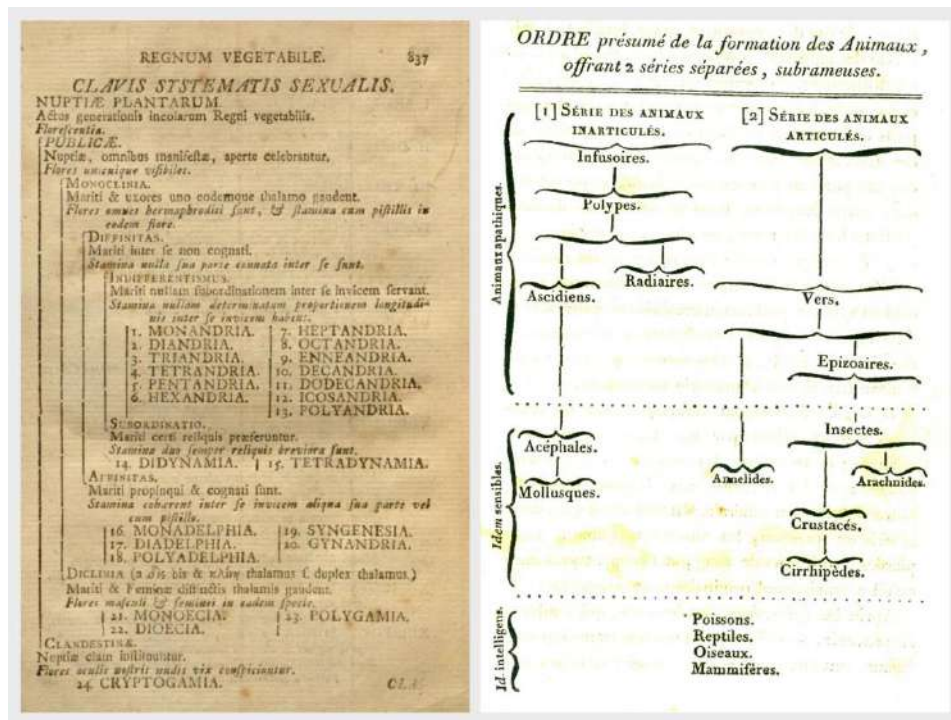


Рисунок 9. Страницы из печатных трудов К. Линнея (слева) и Ж. Ламарка (справа), посвящённые таксономии живых организмов [54,55].

С развитием биологии таксономическая система Линнея усложнялась и постепенно трансформировалась в современную биологическую таксономию. Главная цель биологической таксономии – создание такой системы классификации, которая отражает эволюционные связи между организмами (биологическими объектами). С формированием с начала 1940-х годов синтетической теории эволюции и современного понимания эволюции основных групп термины «эволюционная таксономия» и «линнеевская таксономия» стали в значительной степени взаимозаменяемы [57,58].

Вместе с промежуточными рангами современная таксономия включает более тридцати рангов. Основные таксономические ранги, по мере движения сверху вниз по иерархической структуре, это:

1. Биота (Жизнь)

2. Домен/Реалм
3. Царство
4. Тип/Отдел
5. Класс
6. Порядок/Отряд
7. Семейство
8. Род
9. Вид
10. Штамм

Важными направлениями биологической систематики являются филогенетика, кладистика и фенетика. Филогенетика (от греч. φυλή – «племя», «клан», «раса», γένετικός – «происхождение», «источник», «рождение») занимается изучением эволюционной истории и взаимосвязей внутри группы организмов или между разными группами, используя такие наследуемые признаки и характеристики, как морфология, последовательности нуклеотидных оснований в ДНК или аминокислотных остатков в белках [59]. Результатом филогенетического анализа является филогенетическое дерево – диаграмма, содержащая предположения об эволюционных взаимосвязях изучаемых организмов.

Кладистика (от греч. κλάδος, «ветвь») – это подход к биологической классификации, при котором организмы классифицируются по группам (кладам) на основе гипотез о самом близком общем предке. Свидетельством предполагаемых отношений обычно являются общие производные характеристики (синапоморфии), которых нет у более отдаленных групп и предков [60]. Теоретически, общий предок и все его потомки являются частью клады. С эмпирической точки зрения общие предки – это гипотетические организмы, существование которых предполагается исходя из кладистической гипотезы взаимоотношений таксонов и состояния анализируемых признаков. Для

кладистики важна монофилетичность группы – группа должна включать всех потомков общего предка.

Фенетика (от греч. φαίνω «являю», «обнаруживаю») – это подход к биологической классификации организмов, основанный на анализе общего сходства, независимо от происхождения объектов [61]. Гипотезы, сделанные на основании общего сходства, могут быть полезны, например, в случае слишком большой вычислительной сложности более надёжных филогенетических методов. Методы кладистики позволяют разработать более проверяемые гипотезы [62].

Таксономия и классификация вирусов исторически были предметом научных дискуссий. По мере обнаружения новых вирусов и их изучения с помощью электронной микроскопии разнообразие размеров, форм и составов стремительно росло. Поэтому новые таксономические системы и разные подходы, описывающие классификацию и номенклатуру вирусов, появлялись регулярно. С одной стороны, утверждалось, что было бы преждевременным делать выводы из известных свойств вирусов, не имея информации об их эволюционном происхождении или отношении друг к другу — главной цели классической таксономии. С другой стороны, отмечалось, что в группировании вирусов со сходными свойствами имеются значительные практические преимущества. Основным камнем преткновения было достижение согласия, какие свойства следует считать наиболее важными в построении схемы классификации вирусов.

Первая систематическая классификация вирусов была предложена в 1939 году К. Беннеттом от имени только что созданного Комитета по номенклатуре вирусов Совета Американского фитопатологического общества [63]. В этой системе вирусы были сгруппированы по типу и морфологическим/цитологическим проявлениям симптомов, которые они вызывают у своих хозяев, их тропизму к хозяину, способам их передачи через переносчиков болезней, а также антигенным, химическим и физическим свойствам их частиц. В 1941 г. Ф. Боуден предложил классификацию, основанную исключительно на химических, антигенных и

морфологических свойствах вирусных частиц [64]. В 1948 г. Ф. Холмс предложил первую всеобъемлющую таксономическую систему, чем-то напоминающую ту, что используется для животных и растений. Он предложил таксон высшего ранга, отряд *Virales*, который будет включать три подотряда вирусов, поражающих бактерии (*Phaginae*), растения (*Phytophaginae*) или животных (*Zoophaginae*). Далее Холмс создал 13 семейств, 32 рода и 248 видов. Однако эта система не получила поддержки сообщества, потому что Холмс разделял вирусы на основе тропизма хозяина и игнорировал значительное морфологическое сходство частиц, продуцируемых всеми вирусами, которые он разделил на 3 подотряда.

В 1953–1957 годах Андре Мишель Львофф (1902–1994) создал и расширил набор критериев для определения вируса [65,66]. Это определение легло в основу иерархической таксономии вирусов, предложенной в 1961 году Питером Купером, который разделил вирусы животных по характеристикам частиц содержащие либо ДНК (дезоксивирусы), либо РНК (рибовирусы). Дальнейшее разделение каждой группы зависело от чувствительности к эфиру, т. е. от того, были ли частицы завернуты в липидные оболочки (липовирусы) или устойчивы к эфиру (клатровирусы). В 1963–1967 годах эта система была усовершенствована, чтобы включить в качестве критериев классификации чувствительность вирусных частиц к низкому рН, а также их форму и симметрию. В 1962 г. Андре Львофф, Роберт Хорн и Поль Турнье предложили комплексную схему классификации всех известных на тот момент вирусов, включая бактериальные, растительные и животные вирусы (система LNT) [67]. Авторы использовали классическую линнеевскую иерархическую систему, состоящую из типа, класса, отряда, семейства, рода и вида. Хотя созданный международный комитет по номенклатуре вирусов не принял эту систему в целом, в дальнейшем для классификации вирусов стали использовать обозначения семейств, родов и видов. Один из важнейших принципов, заложенных в системе Львоффа и коллег, заключался в группировании вирусов в соответствии с их общими свойствами, а не по клеткам или организмам,

которые они заражают. Вторым принципом заключался в учете типа нуклеиновых кислот, формирующих вирусный геном.

Всего в системе ЛНТ использовались в качестве основных четыре характеристики:

1. Природа нуклеиновой кислоты вириона (ДНК или РНК)
2. Симметрия белковой оболочки (капсида)
3. Наличие или отсутствие липидной оболочки (оболочки)
4. Размеры вириона и капсида

Система ЛНТ группировала вирусы в один тип (*Vira*) с двумя подтипами. на основе присутствия РНК или ДНК в вирионах (*Deoxyvira* и *Ribovira*). Каждый подтип включал классы, основанные на симметрии капсидов вирионов (например, классы *Deoxyhelica* для спиральных ДНК-вирусов и *Ribocubica* для кубических РНК-вирусов). Эти классы далее делились на отряды («...virales») с подотрядами («...viridales»), семейства («...viridae»), подсемейства («...virinae»), роды («...virus»), подроды («... virus») [68]. Также были описаны типовые виды и общие названия членов таксона. Хотя ни один из высших таксонов не выдержал испытания временем, несколько названий таксонов, используемых в системе ЛНТ, все еще употребляются сегодня (например, стали нынешними названием современных семейств *Poxviridae*, *Adenoviridae* и *Paramyxoviridae*).

Важным этапом в развитии систем классификации вирусов стала система, представленная Дэвидом Балтимором в 1971 году (Балтиморская классификация) [11]. В 1971 году Балтимор опубликовал рабочую классификацию вирусов, которая до сих пор используется параллельно с официальной таксономией. Вместо того, чтобы относить вирусы к таксонам, Балтимор сгруппировал все вирусы в шесть разрозненных групп без каких-либо подразделений, которые теперь обычно называются балтиморскими классами (БК). Балтиморские классы были установлены на основе типа нуклеиновой кислоты, включенной в вирионы, и, следовательно, на типе репродукции вируса.

Таким образом, хотя классические обозначения используются для описания отдельных семейств вирусов, они также группируются в соответствии с типом генома. Балтиморская система выделяет семь типов генома для всех семейств вирусов:

- БК I. Вирусы с геномами двухцепочечной ДНК, которые имеют стратегию репликации-экспрессии, очень напоминающую стратегию клеточных организмов (например, *Adenoviridae*);
- БК II. Вирусы с геномами одноцепочечной ДНК (например, *Geminiviridae*);
- БК III. Вирусы с геномами двухцепочечной РНК (например, *Reoviridae*);
- БК IV. Вирусы с положительными РНК-геномами (например, *Picornaviridae*);
- БК V. Вирусы с отрицательными РНК-геномами (например, *Paramyxoviridae*);
- БК VI. Вирусы с положительными РНК-геномами, которые реплицируются через промежуточные ДНК, продуцируемые обратной транскрипцией генома (например, *Retroviridae*);
- БК VII. Вирусы с геномами двухцепочечной ДНК, которые упаковывают двухцепочечную ДНК или гибрид РНК-ДНК в вирионы и реплицируются с помощью обратной транскрипции (например, *Hepadnaviridae*).

До недавнего времени отдельные классы Балтиморской системы считались монофилетическими. Однако дальнейшие исследования поставили это предположение под сомнение [69,70].

В 1966 г. на Международном конгрессе по микробиологии в Москве Исполнительный комитет Международной ассоциации микробиологических обществ (IAMS) учредил Международный комитет по номенклатуре вирусов (ICNV) для разработки всемирно признанной и применимой таксономии для всех типов вирусов. ICNV, переименованный в 1974 году в Международный комитет по таксономии вирусов (ICTV, International Committee on Taxonomy of Viruses) при Международном союзе микробиологических обществ (IUMS, преемник IAMS), опубликовал свой первый отчет в 1971 году. С тех пор и по настоящее время

таксономия вирусов (включая не только вирусы, но также вириды и сателлиты) находится под юрисдикцией этого комитета.

В отличие от системы ЛНТ, которая подходила к классификации вирусов от высшего таксономического ранга вниз, ICNV/ICTV установила иерархическую таксономию вирусов от самого низкого таксономического ранга вверх. Такой подход был выбран потому, что отличительные свойства, общие для всех вирусов были неизвестны (в отличие, например, от 16S рРНК бактерий или митохондриальной ДНК животных). Информация о последовательности генома для большинства вирусов отсутствовала, и, следовательно, филогенетический анализ был невозможен. Кроме того, высокая частота мутаций и горизонтальный перенос генов делали сложным анализ эволюционного родства между отдалёнными группами вирусов.

Уже во время, соответствующее началу работы ICTV (сайт <https://talk.ictvonline.org/>, [71]), вирусы подозревались в полифилетическом происхождении, а вероятность того, что объединить большие группы вирусов в таксоны более высокого ранга, используя доказуемые эволюционные отношения, будет возможным, считалась низкой. Поэтому вирусы были сгруппированы по морфологическим, биофизическим и биохимическим свойствам их вирионов и фенотипу инфекций. После появления более эффективных методологий секвенирования генома вирусы были сгруппированы в соответствии с вирусными геномными последовательностями и филогенетическими отношениями друг к другу в пределах относительно близкородственных наборов вирусов. Результатом этой деятельности стали многочисленные индивидуальные таксономические иерархии вирусов, как правило, до уровня семейства, которые не были связаны друг с другом на высшем уровне.

Историческое развитие официальной таксономии вирусов ICTV можно просмотреть в опубликованных онлайн-отчетах и промежуточных обновлениях на сайте ICTV, которые обычно публикуются в архивах новостей отдела вирусологии

(VDN). Последний (10-й) отчет был опубликован в открытом доступе в Интернете и постоянно обновляется не реже одного раза в год. В первом отчете перечислялись только 2 семейства, 27 родов, 10 подродов и 18 групп вирусов, но с тех пор схема классификации значительно расширилась. Например, первый отряд (*Mononegavirales*) был создан в 1991 г., а в 2019 г. появились таксоны более высокого ранга, включающие отряд *Mononegavirales*, тип *Negarnaviricota* и царство *Riboviria*.

Несмотря на такое расширение рангов, многие семейства в настоящее не связаны с другими семействами более высокого таксономического ранга. Тем не менее, большинству РНК-вирусов был присвоен наивысший доступный ранг – царство. Это отнесение было основано на признании того, что все эти РНК-вирусы имеют геномы, которые имеют общий вирусный отличительный ген – РНК-зависимую РНК-полимеразу (RdRp), образующую монофилетическую группу.

В 2019 году решением рабочей группы ICTV количество рангов таксономической классификации вирусов было увеличено до пятнадцати [72], включая восемь основных (первичных) и семь производных (или вторичных) (рис. 10). Восемь основных рангов включают четыре, которые уже использовались раньше (отряд, семейство, род и вид) и четыре новых ранга (реалм, царство, тип и класс), которые выше ранга отряда. Базовый ранг в вирусной таксономии предложено называть реалмом, а не доменом (как в таксономических системах клеточных организмов), чтобы показать сложную взаимосвязь между таксономией вирусов и таксономией клеточных организмов.

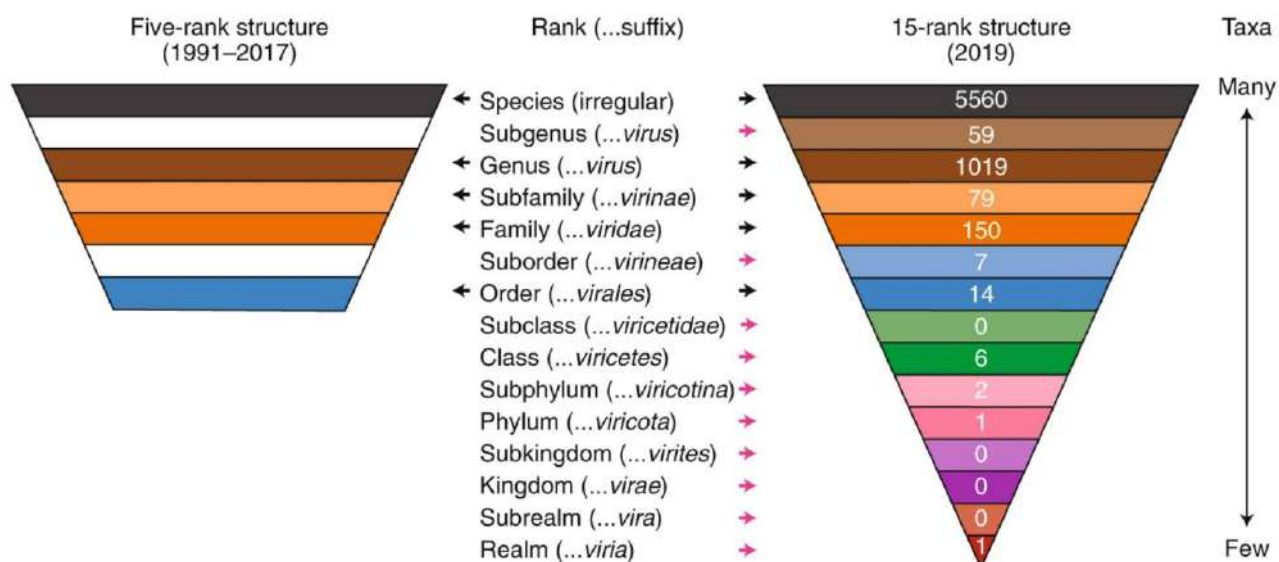


Рисунок 10. Сравнение иерархии таксономических рангов ICTV в 1991–2017 и 2020 гг. Количество таксонов, присвоенных каждому рангу (согласно текущему Основному списку видов ICTV, выпуск 2018b) показано белым цветом [72].

1.4. Таксономия бактериофагов – история развития и современная классификационная схема

Исторически фаги классифицировались в соответствии с их морфологией. Но во времена первых классификационных схем, описывающих бактериофаги, ещё не существовало ПЦР, секвенирования и многих молекулярных методов, известных нам сегодня [8,73]. Первая распространённая схема таксономической классификации наиболее многочисленных изученных представителей фагов, – хвостатых фагов, относящихся к классу *Caudoviricetes*, – была предложена Дэвидом Брэдли в 1967 году, расширена Анной Сергеевной Тихоненко в 1968 году, и усовершенствована Гансом-Вольфгангом Аккерманном и Абрахамом Айзенстарком в 1974 году [9,10]. На основе данных электронной микроскопии бактериофаги классифицировались по трём морфотипам: морфотип А (фаги с сокращающимся хвостом), морфотип В (фаги с длинным, несокращающимся хвостом), морфотип С (фаги с коротким несокращающимся хвостом). В 1971 году номенклатура была принята ICTV, а в 1981 и 1984 годах этим морфотипам, ставшим семействами, были присвоены названия *Myoviridae* (морфотип А),

Podoviridae (морфотип С) и *Siphoviridae* (морфотип В). В результате продолжительного обсуждения были также сформированы семейства фагов, не относящимся к хвостатым: *Inoviridae*, *Microviridae*, *Tectiviridae*, *Corticoviridae*, *Plasmaviridae*, *Leviviridae* и *Cystoviridae* (рис. 11, [7,8]).

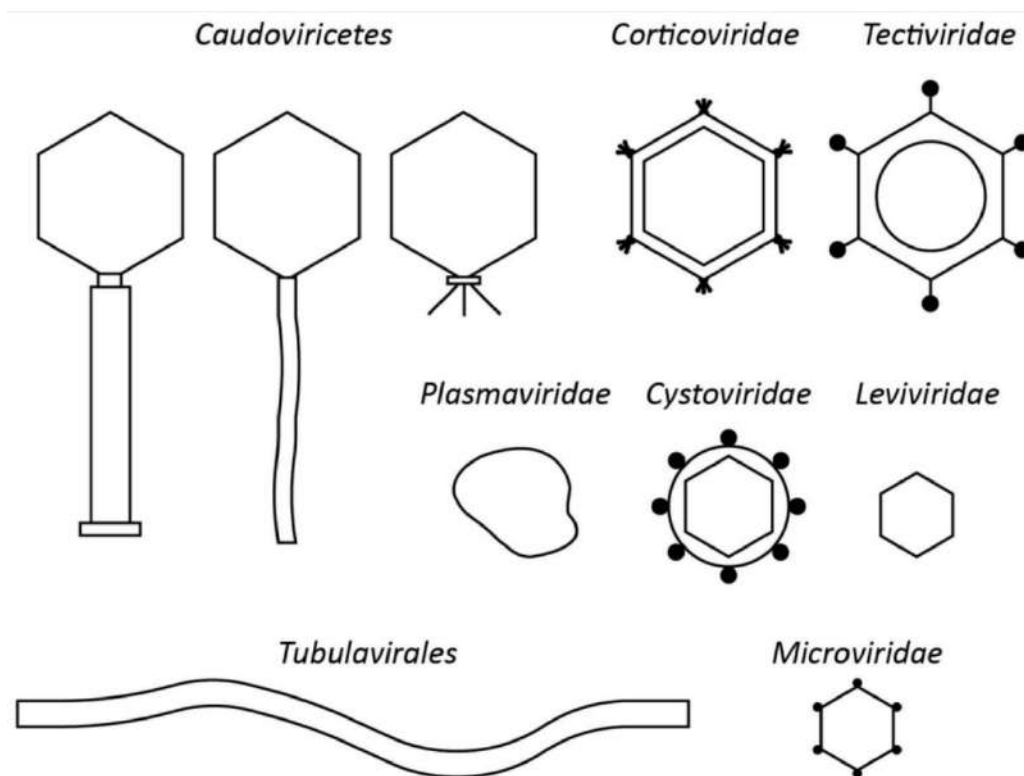


Рисунок 11. Схематическое изображение морфотипов бактериофагов. По Аккерману, 2004 [7,8].

Совершенствование методов молекулярной биологии, секвенирование фаговых геномов и бурное развитие биоинформатики в начале 2000-х годов выявило гораздо более высокое геномное разнообразие, чем считалось ранее, особенно у бактериофагов, принадлежащих к отряду *Caudovirales*, что привело к созданию первых подсемейств внутри трёх существующих семейств. По мере увеличения количества секвенированных геномов бактериофагов стало понятным, что три семейства хвостатых фагов, – *Myoviridae*, *Podoviridae* и *Siphoviridae* – не являются монофилетичными. Парафилетичность этих семейств была показана биоинформатическими методами, благодаря которым было выделено несколько новых семейств. Биоинформатические подходы позволили уточнить

классификацию семейства бактериофагов с одноцепочечной ДНК *Microviridae*, значительно увеличив количество подсемейств, в том числе, за счёт подсемейств *Alpavirinae* (2011 г.) [74], *Pichovirinae* (2015 г.) [75] и других.

В последнее десятилетие всё большую роль в описании новых организмов стала играть метагеномика. Метагеномный анализ позволил получить последовательности геномов некультивируемых организмов, которых, как представляется, существенно больше, чем культивируемых. На основе метагеномных данных были выделены фаги группы crAss (2014 г.) [76] и многие гигантские бактериофаги (2019-2020 гг.) [77,78].

Важное значение для вирусологии имело выделение нового фага с одноцепочечной ДНК, классифицированного в составе нового семейства *Finnlakeviridae*. Анализ генома и структурных белков этого фага позволили выявить эволюционные связи между фагами с икосаэдрическими капсидами с одноцепочечной ДНК и бесхвостыми бактериофагами с двухцепочечной ДНК (2017 г.) [79]. Изменения коснулись и нитчатых фагов с геномом, представленным одноцепочечной ДНК, которые, как было отмечено выше, не лизируют клетку хозяина, покидая её. Анализ бактериальных геномов, имеющихся в открытых базах данных, позволил обнаружить так называемые «криптические» нитчатые бактериофаги, которые не были выделены, но могут быть идентифицированы по наличию характерных генов (2019 г.) [80]. Семейство нитчатых фагов *Inoviridae* было разделено на два семейства, *Inoviridae* и *Plectroviridae*, объединённые в отряд *Tubulavirales* (2020 г.) [81], с потенциальным дальнейшим увеличением за счет пяти новых семейств [8]. Анализ геномных данных позволил выявить большое количество одноцепочечных РНК-вирусов семейства *Leviviridae*, сначала 158 (2016 г.) [82], а затем – около одной тысячи полных и пятнадцати тысяч неполных геномов (2020 г.) [83].

Последние годы ознаменовались взрывным ростом количества новых таксонов бактериофагов. В 2018-2020 годах ICTV официально ратифицировал три

новых семейства миовирусов (*Ackermannviridae* [84,85], *Chaseviridae* [81], *Herelleviridae* [86]), два новых семейства сифовирусов (*Demerecviridae* [81] и *Drexelviridae* [81]), а также два семейства подовирусов (*Autographiviridae* [81] и *Schitoviridae* [87,88]). В 2021 году была принята новая схема таксономической классификации хвостатых бактериофагов, ратифицированная в марте 2022 года. Новая классификация ликвидировала семейства *Myoviridae*, *Podoviridae* и *Siphoviridae*, основанные в 1980-х годах на основании морфологических характеристик и до 2021 года включавшие в себя хвостатые фаги, не включённые в новообразованные в дальнейшем таксоны [12]. Все хвостатые фаги были отнесены к классу *Caudoviricetes*, в свою очередь, принадлежащего типу *Uroviricota* царства *Heunggongvirae* реалма *Duplodnaviria*. Отряд *Caudovirales*, включавший ранее в себя все хвостатые фаги, был ликвидирован, взамен были образованы новые таксоны, в том числе, высокоранговые. В новообразованный отряд *Crassvirales* были включены фаги группы crAss, ранее обнаруженные в метагеноме человеческого пищеварительного тракта и инфицирующие бактерии типа *Bacteroidetes* [76,89,90]. По состоянию на середину 2022 года этот отряд включал четыре семейства. Также было образовано три отряда, включающие вирусы архей, ранее относившиеся к ликвидированному семейству *Myoviridae* – *Kirjokansivirales* (в составе четыре семейства), *Methanobavirales* (в составе два семейства) и *Thumleimavirales* (в составе четыре семейства). Остальные хвостатые бактериофаги в таксономической системе 2022 года отнесены к 35 семействам, 37 подсемействам и 493 родам, которые приписаны к классу *Caudoviricetes* напрямую, без отнесения к таксонам промежуточного ранга. Общее количество таксонов, отнесённых к классу *Caudoviricetes*, составляет 4 отряда (включая 3 отряда вирусов архей), 47 семейств, 98 подсемейств и 1197 родов. Общая информация о таксономии бактериофагов в соответствии с классификацией ICTV, ратифицированной в начале 2022 года [91], представлена в таблице 1.

Таблица 1. Основные таксономические группы бактериофагов в соответствии с классификацией ICTV, ратифицированной в марте 2022 года. «—» означает, что для более низкоранговых таксонов в столбцах справа более высокоранговые таксоны, кроме указанных, не предписаны.

Реалм	Царство	Тип	Класс	Отряд	Таксоны более низкого ранга
<i>Duplodnaviria</i>	<i>Heunggongvirae</i>	<i>Uroviricota</i>	<i>Caudoviricetes</i>	<i>Crassvirales</i>	4 семейства фагов группы crAss
<i>Duplodnaviria</i>	<i>Heunggongvirae</i>	<i>Uroviricota</i>	<i>Caudoviricetes</i>	—	35 семейств хвостатых фагов, отнесённых напрямую к классу <i>Caudoviricetes</i>
<i>Duplodnaviria</i>	<i>Heunggongvirae</i>	<i>Uroviricota</i>	<i>Caudoviricetes</i>	—	37 подсемейств хвостатых фагов, отнесённых напрямую к классу <i>Caudoviricetes</i>
<i>Duplodnaviria</i>	<i>Heunggongvirae</i>	<i>Uroviricota</i>	<i>Caudoviricetes</i>	—	1197 родов хвостатых фагов, отнесённых напрямую к классу <i>Caudoviricetes</i>
<i>Monodnaviria</i>	<i>Loebvirae</i>	<i>Hofneiviricota</i>	<i>Faserviricetes</i>	<i>Tubulavirales</i>	Нитчатые фаги, 3 семейства, включающие 51 род
<i>Monodnaviria</i>	<i>Sangervirae</i>	<i>Phixviricota</i>	<i>Malgrandaviricetes</i>	<i>Petitvirales</i>	Семейство фагов <i>Microviridae</i> , включающее 2 подсемейство с 7 родами
<i>Riboviria</i>	<i>Orthornavirae</i>	<i>Duplomaviricota</i>	<i>Vidaverviricetes</i>	<i>Mindivirales</i>	Семейство бесхвостых фагов <i>Cystoviridae</i> , содержит 5 родов
<i>Riboviria</i>	<i>Orthornavirae</i>	<i>Lenarviricota</i>	<i>Leviviricetes</i>	<i>Norzivirales</i>	4 семейства (+)РНК фагов, включающие 271 род
<i>Riboviria</i>	<i>Orthornavirae</i>	<i>Lenarviricota</i>	<i>Leviviricetes</i>	<i>Timlovirales</i>	2 семейства (+)РНК фагов, включающие 148 родов
<i>Riboviria</i>	<i>Orthornavirae</i>	<i>Pisuviricota</i>	<i>Duplopiviricetes</i>	<i>Durnavirales</i>	Семейство <i>Picobirnaviridae</i> , предположительно инфицирующее прокарнот, содержит 1 род
<i>Varidnaviria</i>	<i>Bamfordvirae</i>	<i>Preplasmiviricota</i>	<i>Tectiliviricetes</i>	<i>Kalamavirales</i>	Семейство бесхвостых фагов <i>Tectiviridae</i> , содержит 5 родов
<i>Varidnaviria</i>	<i>Bamfordvirae</i>	<i>Preplasmiviricota</i>	<i>Tectiliviricetes</i>	<i>Vinavirales</i>	Семейство бесхвостых фагов <i>Corticoviridae</i> , содержит 1 род
<i>Varidnaviria</i>	<i>Helvetiavirae</i>	<i>Dividoviricota</i>	<i>Laserviricetes</i>	<i>Halopanivirales</i>	Семейство <i>Matshushitaviridae</i> содержит 1 род <i>Hukuchivirus</i> (фаги, инфицирующие термофильных бактерий)
—	—	—	—	—	Семейство <i>Finnlakeviridae</i> , 1 род и один вид, оцДНК(+) вирус, инфицирующий <i>Flavobacterium</i> sp. [92]
—	—	—	—	—	Семейство <i>Plasmaviridae</i> , 1 род дцДНК фагов, инфицирующих представителей Mollicutes [93]

1.5. Современные критерии таксономических рангов бактериофагов и специализированные биоинформатические методы и инструменты

Для таксономической классификации бактериофагов ICTV предлагает критерии и методы, оценивающие геномные характеристики бактериофагов [8]. Состав применяемых методов определяется иерархией классифицируемого таксона. В качестве критериев предлагаются следующие параметры и особенности:

1. В 2012 году была предложена граница взаимного средненегеномного нуклеотидного сходства (ANI, average nucleotide identity) фаговых геномов 95%, в пересчёте на общую длину генома [94], которая является основанием для причисления их к одному виду. Эффективность применения критерия 95%-го сходства была показана в более поздних работах [95–97]. Критерий 95% рекомендован сейчас к использованию ведущими специалистами ICTV.
2. Общегеномное нуклеотидное сходство предлагается также использовать и в качестве критерия причисления бактериофагов к одному роду. Подкомитет ICTV по вирусам бактерий и архей предложил 70%-ную границу ANI в качестве критерия разделения фаговых родов, причём принадлежность к одному роду должна удостоверяться наличием характерных генов и детальным филогенетическим анализом [8].
3. ICTV пока не разработал демаркационную границу семейств и подсемейств. В качестве опционного критерия выделения группы фагов в подсемейство предложено наличие двух и более родов, которые недостаточно отдалённые, чтобы причислять их к разным семействам. Филогенетический анализ с использованием маркерных генов должен показывать принадлежность родов одного подсемейства к отдельной кладе. Также фаги, принадлежащие к одному семейству, могут обладать обнаруживаемым нуклеотидным сходством. В качестве критериев, необходимых для формирования семейства, предложено считать:

- группа фагов составляет одно семейство, если она является монофилетической на основании результатов кластеризации по протеомному сходству;
 - группа фагов, причисленных к одному семейству, должна обладать значительным количеством ортологичных генов;
 - в случае, если другие семейства разделяют некоторые ортологичные гены, характерные для данного семейства, филогенетический анализ должен показывать монофилию группы генов данного семейства.
4. Для таксономического ранга «отряд» также не предусмотрено каких-то количественных демаркационных критериев. В порядок включают два и более семейства, которые образуют монофилетическую группу по результатам кластеризации общепринятыми методами.

То есть, для классификации разных таксономических рангов используются разные наборы биоинформатических методов, которые можно разделить на методы анализа среднегеномного нуклеотидного сходства, протеомного сходства, сходства порядка генов, филогенетические методы. Отдельно можно выделить методы, демонстрирующие сетевой характер родственных взаимоотношений между фагами.

1.5.1. Инструменты анализа межгеномного нуклеотидного сходства

Возможно, самый традиционный и понятный инструмент сравнения нуклеотидного сходства – это BLASTn [98]. Значение межгеномного нуклеотидного сходства можно получить умножением попарной идентичности на покрытие (долю участков в геноме, которые программа смогла выровнять). Алгоритмы BLAST используются и в других инструментах определения межгеномного нуклеотидного сходства, включая несколько калькуляторов ANI (например, Kostas lab ANI calculator [99]), в вариантах алгоритмов, использующих BLASTn, реализованных сервером EzBioCloud [100], в программах FastANI [101],

CD-HIT [102], Gegenees [103], JSpecies [104], VIRIDIC [105] и других. Большинство этих инструментов перед сравнением делят геномы на короткие отрезки размером несколько сотен или тысяч нуклеотидов. Все эти инструменты позволяют рассчитать попарные значения ANI и степень покрытия. Кроме того, инструменты Kostas lab, CD-HIT, Gegenees, JSpecies и VIRIDIC позволяют выполнить иерархическую кластеризацию данных с получением матрицы, в которой близкие последовательности сгруппированы в соответствующих кластерах (рис. 12). Матрицы ANI могут быть использованы для построения филогенетических деревьев, основанных на значениях ANI и общегеномного сходства (ANI с учётом покрытия).

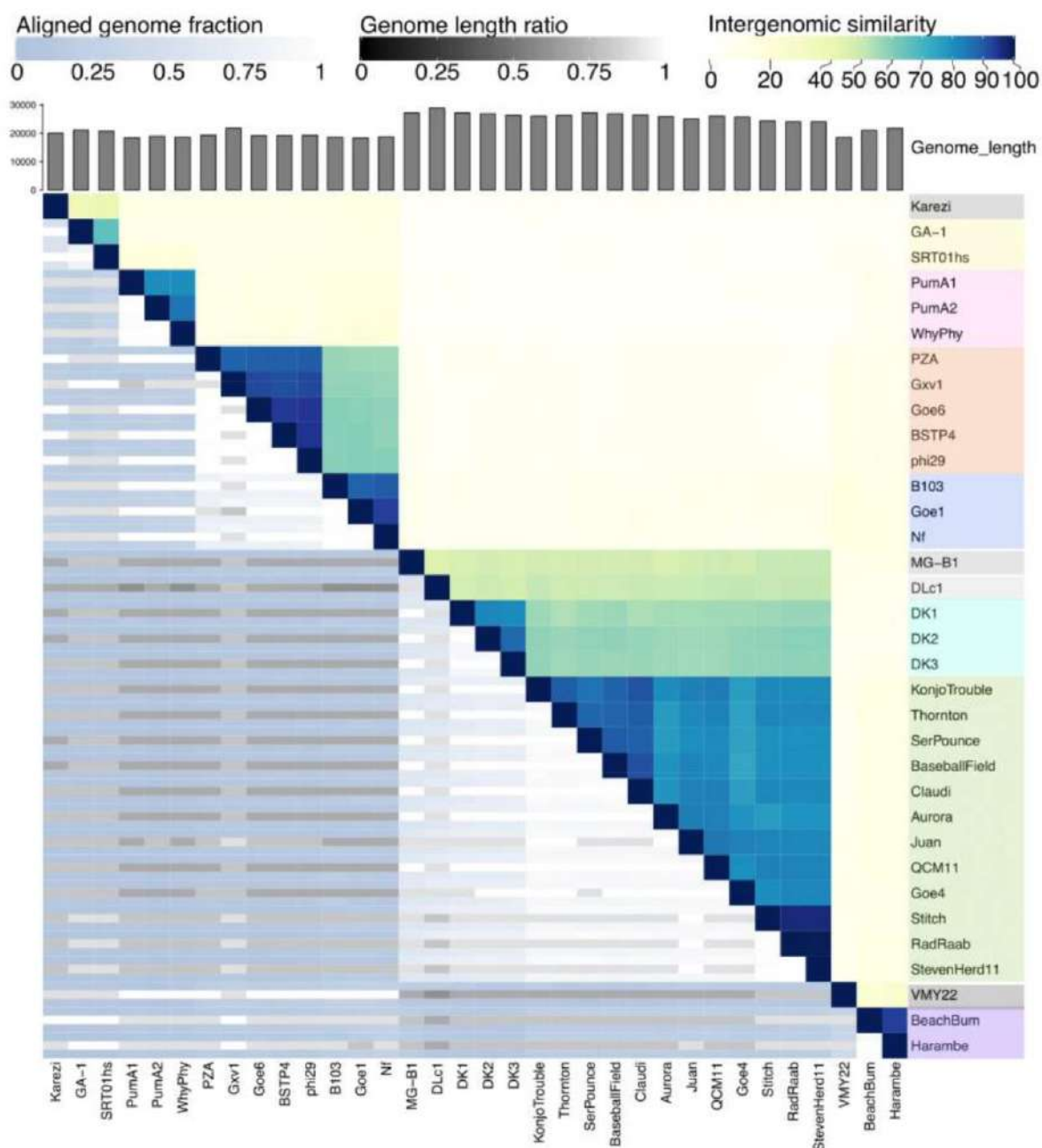


Рисунок 12. Тепловая карта межгеномного сходства бактериофагов, принадлежащих к роду *Bundooravirus* семейства *Salasmaviridae*, полученная с помощью программы VIRIDIC [106].

Алгоритмика, используемая программой VIRIDIC, проста и соответствует способам «ручного» анализа межгеномного сходства, используемого ICTV. Межгеномное сходство рассчитывается следующим образом:

$$sim(AB) = ((idAB + idBA) * 100) / (lA + lB)$$

$$dist(AB) = 100 - sim(AB)$$

где $sim(AB)$ – межгеномное сходство геномов A и B , $dist(AB)$ – межгеномная дистанция между геномами A и B , $idAB$ – количество идентичных оснований, при выравнивании генома A к геному B , $idBA$ – количество идентичных оснований при выравнивании генома B к геному A , lA – длина генома A , lB – длина генома B .

1.5.2. Инструменты анализа протеомного сходства

Инструменты анализа сходства протеома работают по тем же принципам, что и инструменты анализа нуклеотидного сходства, но вместо нуклеотидных последовательностей при сравнении используют аминокислотные. Соответственно, для сравнения могут быть использованы варианты BLAST и других алгоритмов, рассчитанных для работы с белками (BLASTP и др.). Как и в случае нуклеотидного сходства, ряд инструментов, например, Kostas lab [99] и CD-NIT [102], способны выполнять иерархическую кластеризацию.

Группа инструментов анализа протеомного сходства, включая «The Phage Proteomic Tree» [107] и «ViPTree» [108], нацелена на филогенетический анализ и осуществляет иерархическую кластеризацию конкатенированных последовательностей фаговых белков с представлением результата в виде дендрограммы (рис. 13, 14).

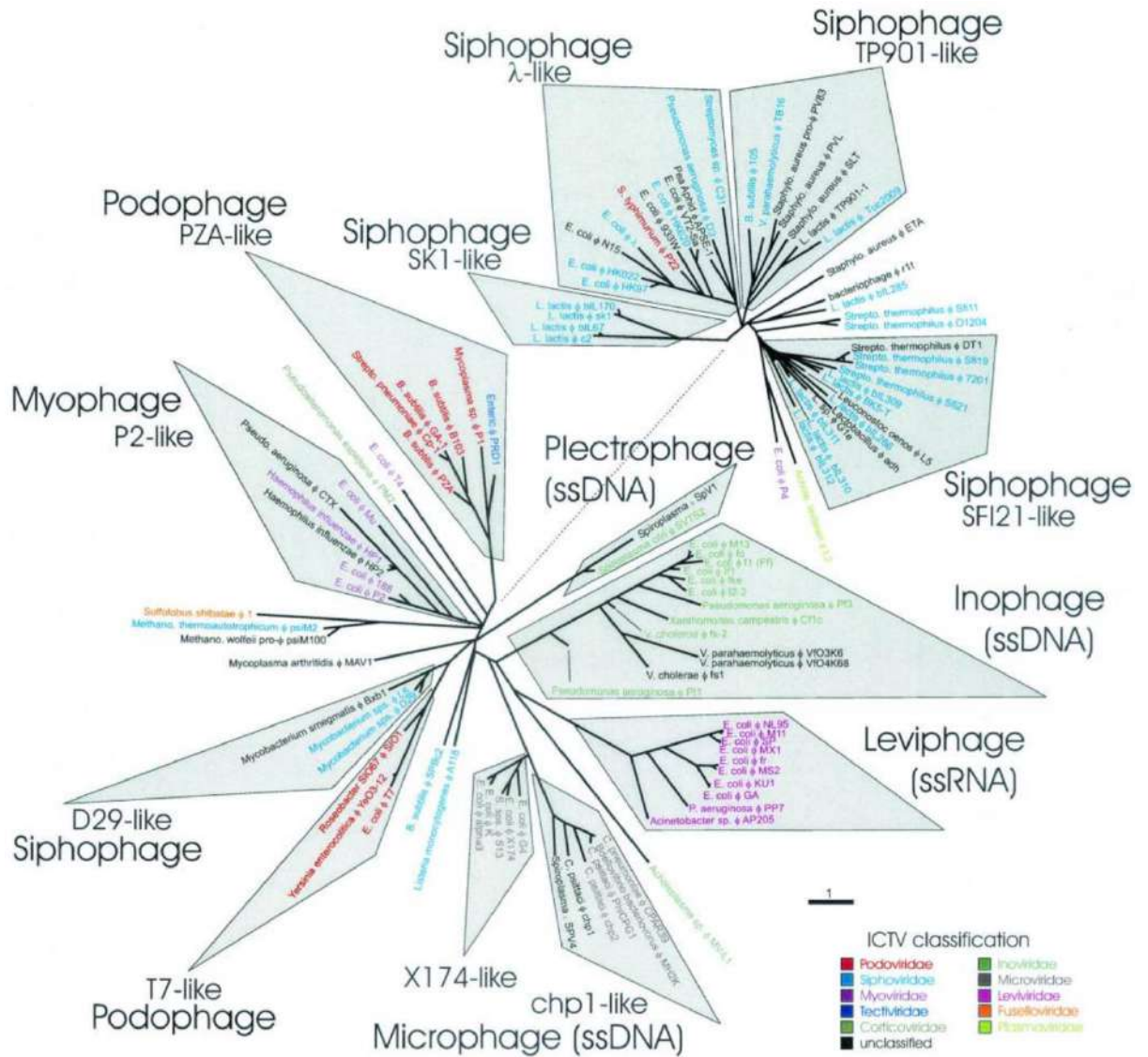


Рисунок 13. Фаговое протеомное дерево, построенное с использованием 105 полностью секвенированных фаговых геномов с помощью пайплайна The Phage Proteomic Tree. Дерево было создано на основе оценки аминокислотной дистанции между белками. Название каждого фага окрашено в соответствии с его классификацией ICTV, как показано в подписях. Пунктирная линия показывает место соединения двух групп [107].

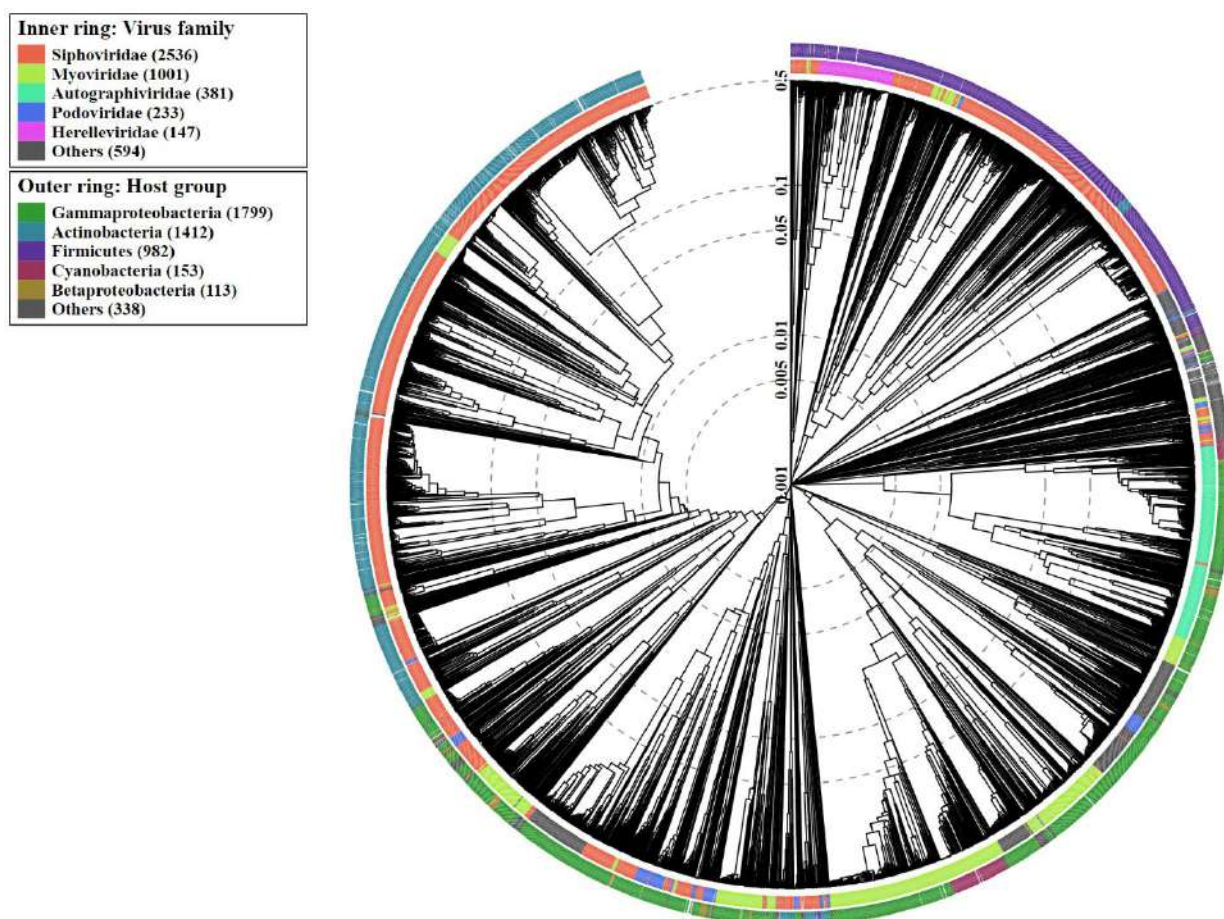


Рисунок 14. Круговое фаговое протеомное дерево, построенное с использованием 4919 полностью секвенированных фаговых геномов с помощью пайплайна ViPTree. Дерево было создано на основе оценки аминокислотной дистанции между белками. Названия таксономических групп и количество принадлежащих им фагов в дереве указано в легендах в верхнем левом углу [108].

Пайплайн «Genome Relationships Applied to Virus Taxonomy» (GRAViTy) [109–111] реализован как в виде отдельной программы, так и в виде онлайн-сервера Университета Глазго. Его алгоритмика которого задействует составную меру различия Жаккара, вычисленную с использованием сигнатур белкового профиля скрытой марковской модели (Protein profile hidden Markov model, PPHMM). GRAViTy использует как оценку наличия гомологичных белков, так и их расположение в геноме, позволяет сравнить не только межгеномное сходство, но и геномную организацию. Пример дендрограмм, полученных с помощью GRAViTy приведён на рис. 15.

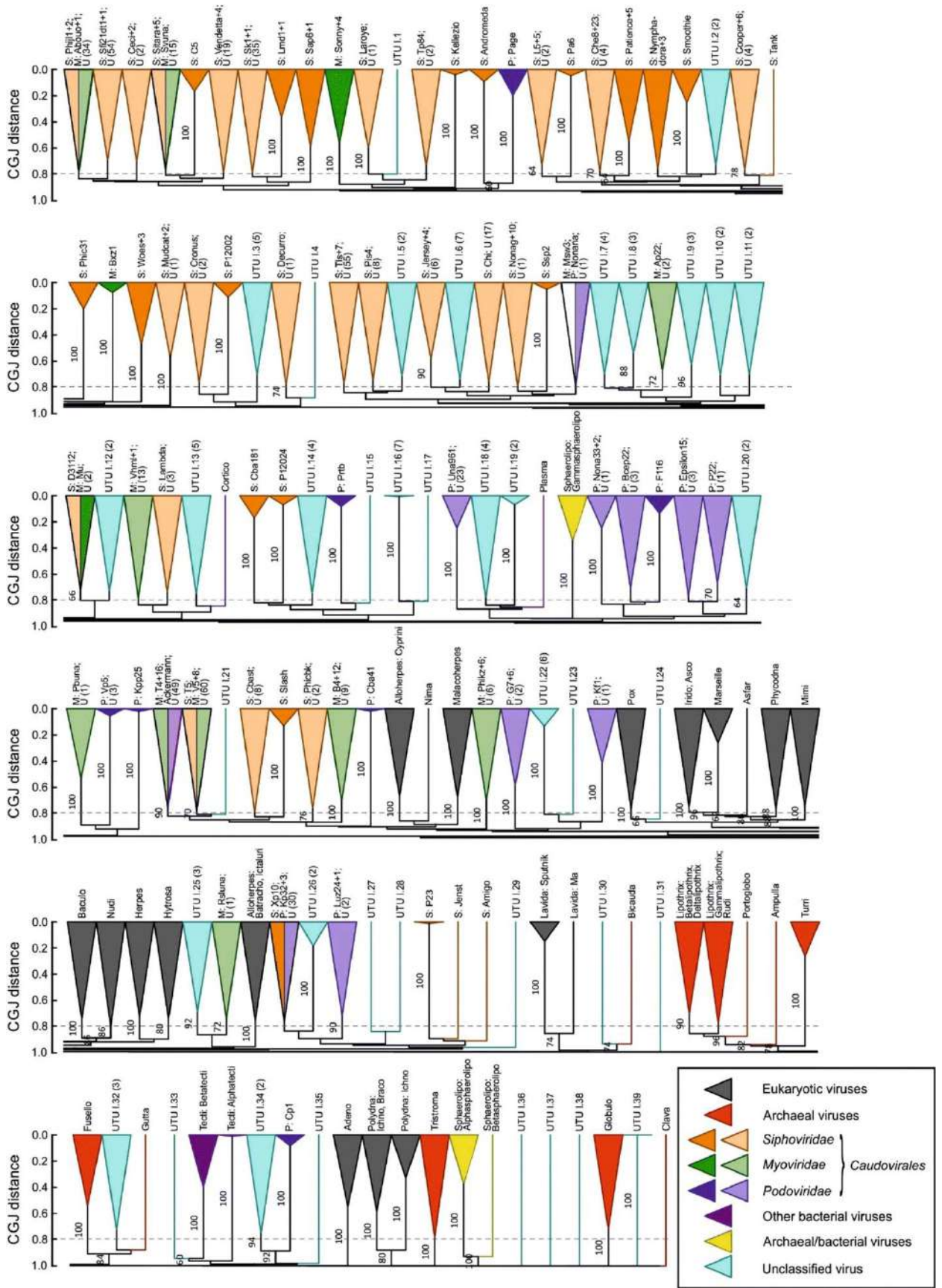


Рисунок 15. Дендрогаммы, основанные на вычисленной пайплайном GRAViTy составной мере различия Жаккара (composite generalised Jaccard distance, CGJ distance), для разных групп вирусов [111].

Сравнение белковых профилей на основе скрытых марковских моделей используется и в таксономическом пайплайне ClassiPhage [112]. ClassiPhage классифицирует фаги на основе данных о последовательности генома и может также применяться при анализе профаговых областей в геномах бактериальных хозяев и классификации профагов. Инструмент ClassiPhage тоже нацелен на анализ метагеномных данных и эффективен при разделении на уровнях семейств и подсемейств (в таксономии 2019 года). Пайплайн представляет результаты анализа в виде тепловой карты, с областями, окрашенными в соответствии с бит-шкалой НММ и анализируемым фагом (рис. 16).

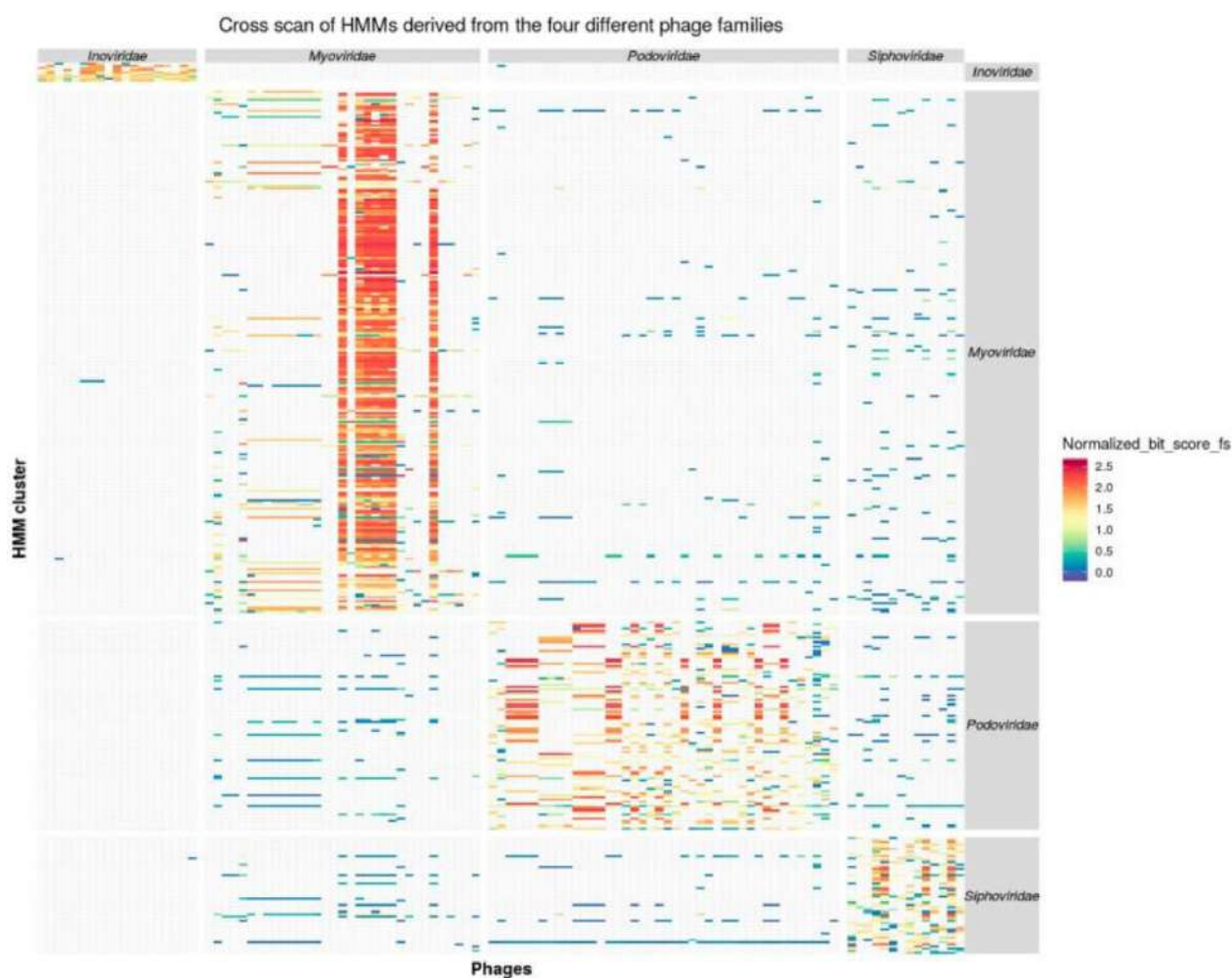


Рисунок 16. Тепловая карта НММ-сходства, полученная с помощью инструмента ClassiPhage [112].

1.5.3. Инструменты филогенетического анализа

Принадлежность фагов к одному роду, подсемейству, семейству и отряду требует монофилетичности группы, подтверждённой филогенетическим анализом. Подкомитет ICTV по вирусам бактерий и архей рекомендует использовать филогенетический анализ, выполненный методом максимального правдоподобия (ML) с помощью инструментов IQ-Tree, ModelFinder (для определения модели замещения) и UFBOOT (для бутстрэп-анализа) [113,114,115, p. 2], но допустимо использование и других программ, реализующих алгоритмы максимального правдоподобия [8].

Среди других широко распространённых филогенетических инструментов можно отметить RAxML (Randomized Axelerated Maximum Likelihood) [116]. RAxML обладает широчайшим функционалом, большим выбором алгоритмов и способностью определять лучшую модель замещения. Как и IQ-Tree, RAxML может использовать технику поиска восхождением к вершине (hill climbing) [117]. Новая имплементация RAxML-NG, по данным авторов программы [118], предлагает повышенную точность, гибкость, скорость, масштабируемость и удобство использования. Для наборов данных, отличающихся большим разнообразием таксонов, RAxML-NG обычно находит деревья с более высокой оценкой надёжности, чем IQ-Tree. RAxML-NG также представляет несколько новых функций, таких как обнаружение террас в пространстве деревьев [119] и новую метрику переноса ожидаемых значений бутстрэп-анализа (transfer bootstrap expectation branch support metric) [120]. Как альтернатива ModelFinder, в качестве инструмента нахождения модели замещения может использоваться программа ModelTest-NG [121], разработанная авторами RAxML-NG.

Из других распространённых продвинутых программных инструментов филогенетического анализа можно выделить MrBayes [122], использующий байесовские методы для построения филогенетических связей. Особенность MrBayes – это возможность одновременного использования в анализе разных типов

данных, например, нуклеотидных и аминокислотных последовательностей геномов разных типов (например, ядерного и митохондриального). Как и RAxML, MrBayes поддерживает разные модели нуклеотидного и аминокислотного замещения и предлагает продвинутое настраиваемое меню анализа.

Важную роль для определения эволюционных взаимосвязей с помощью филогенетического анализа играет выбор последовательностей. В последнее время большую популярность в филогенетическом анализе приобрело соединение (конкатенирование) генов в один «суперген» [123]. Использование конкатенированных последовательностей генов или кодируемых ими белков может существенно увеличить отношение сигнал/шум при условии, что эти гены (белки) имеют одинаковую эволюционную историю [124,125]. Поэтому использование конкатенированных последовательностей консервативных генов является важной частью характеристики таксонов бактериофагов [8]. Сигнатурные гены могут быть получены анализом всех генов и кодируемых их белков с помощью BLAST и HMM поиска. Более быстрый способ – это использование инструментов, созданных для пангеномного анализа. К популярным пангеномным программным пакетам можно отнести BPGA (Bacterial Pan Genome Analysis) [126], CoreGenes [127], PIRATE [128], PGAP [129,130], Proteinortho [131], Roary [132], Panaroo [133] и др. Эти инструменты первоначально были созданы для анализа геномов прокариот, но могут быть также успешно задействованы и для изучения фаговой геномики. Учитывая более высокую степень дивергенции фаговых последовательностей, специалисты ICTV рекомендуют использовать менее строгие критерии для создания пангеномов фагов, например, >30% идентичности для фаговых белков (вместо часто используемого предела >50% для прокариотических белков) и >50% покрытия. Применение инструментов пангеномного анализа также даёт возможность сравнительно быстрого анализа с помощью иерархической кластеризации фагов на основе состава генного пула.

1.5.4. Инструменты анализа сети родства фагов

Эволюция вирусов сопровождается регулярными событиями горизонтальных переносов генов и рекомбинации генетического материала, из-за чего сеть может быть более правильной и наглядной моделью эволюционных взаимосвязей между ними [134–136]. Сеть, построенная с помощью кластеризации белков, гомология которых была обнаружена BLASTP, изображена на рис. 17. Для таксономической классификации бактериофагов часто используется инструмент vConTACT, при сравнительном анализе продемонстрировавший свою эффективность для подавляющего большинства случаев [137].

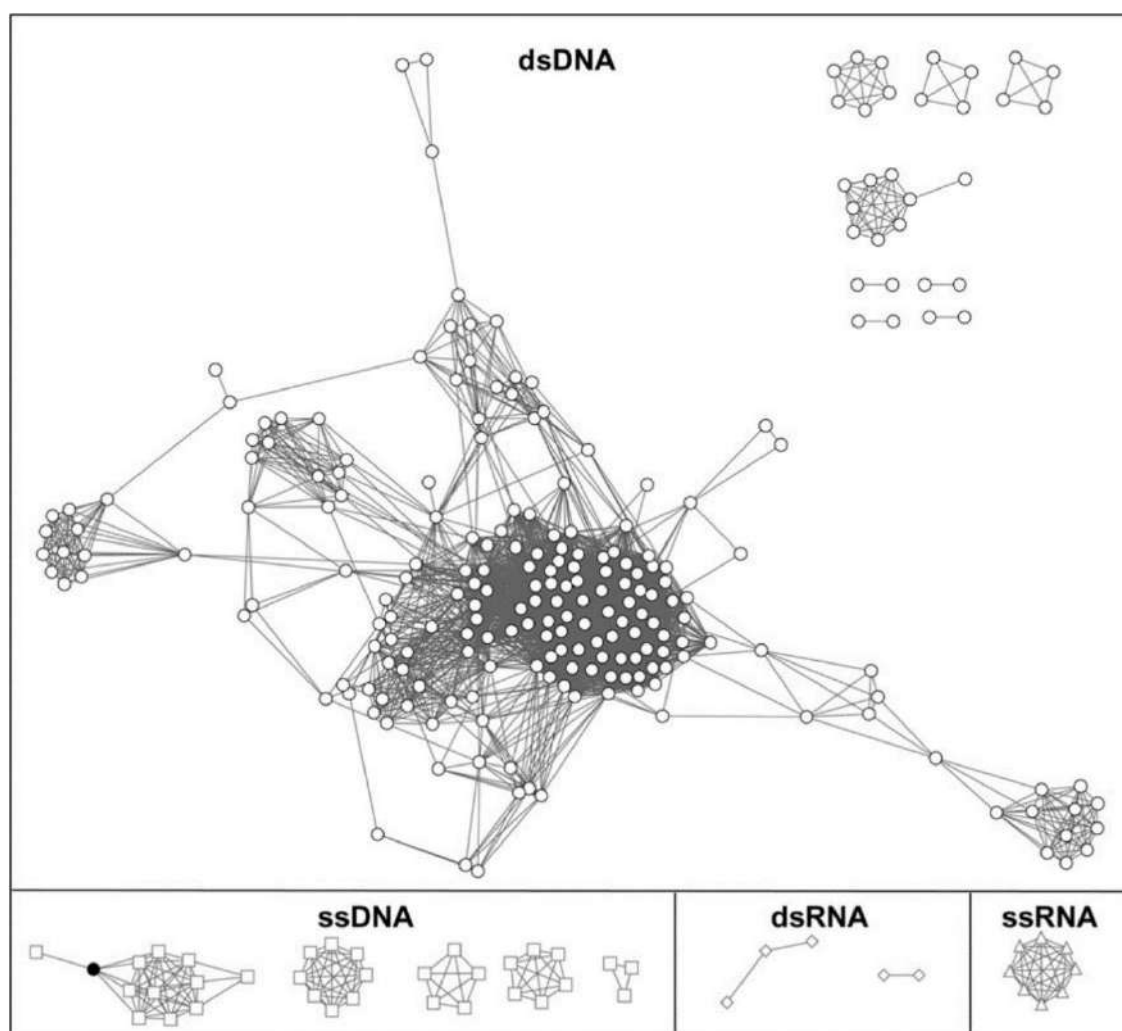


Рисунок 17. Сеть, демонстрирующая родственные связи 277 фагов разных таксономических групп. Узлы представляют фаги, а ребра представляют их статистически значимое попарное сходство [138].

1.6. Проблемы таксономии бактериофагов и задачи биоинформатики

Долгое время одной из наиболее серьёзных проблем в таксономии бактериофагов являлась парафилетичность семейств хвостатых фагов, составляющих большинство выделенных и, как предполагается, представленных в метагеномах и виромах вирусов [8]. С последними изменениями в классификации бактериофагов, утверждёнными ICTV, эта проблема частично решена. Тем не менее, необходима дальнейшая работа и по классификации родов, не приписанных к семействам и подсемействам, и по классификации на уровне отрядов. Эта работа требует выявления глубоких эволюционных взаимосвязей между давно дивергировавшими группами бактериофагов.

Определение глубоких эволюционных взаимосвязей между давно дивергировавшими семействами – это большая проблема эволюционной вирусологии. Предполагается, что вирусы имеют древнее происхождение, и, возможно, являются самыми древними созданиями на Земле [139]. Продолжительная эволюция после расхождения вирусных групп приводит к большому количеству мутаций в генах, что затрудняет выявление надёжных филогенетических связей. Особенно это актуально для РНК-вирусов, в которых скорость мутаций, как правило, существенно выше, чем у ДНК-вирусов. Проблема может быть частично решена за счёт использования консервативных генов [140], конкатенированных последовательностей консервативных генов [125] или анализа укладки белков и сходства их структур [141].

Возможно, основные группы вирусов имеют химерное происхождение [142]. В соответствии с химерным сценарием различные типы первичных эгоистичных репликонов дали начало вирусам, рекрутируя белки хозяев для образования вирионов, и новые группы вирусов неоднократно возникали на всех этапах эволюции жизни, часто за счет замещения предковых структурных генов и генов репликации.

Химерное происхождение имеют не только давно дивергировавшие группы вирусов. Так называемая «мозаичная» структура генома была продемонстрирована у более близких вирусов, например у λ -подобных бактериофагов [143]. Теория модульной эволюции бактериофагов, утверждающая, что различные геномные модули умеренных и вирулентных фагов могут иметь разное происхождение [144], находит свои экспериментальные подтверждения. Понятно, что при использовании конкатенированных последовательностей белков, имеющих разное эволюционное происхождение, полученные результаты филогенетического анализа могут показать не вполне достоверную эволюционную историю [145]. Использование сходства геномной организации, которое во многом обуславливает жизненный цикл фагов [146], порядка и списка генов, может быть важной характеристикой, способствующей более точной классификации фагов с химерными и мозаичными геномами.

Современная биоинформатика предлагает широкий набор инструментов для анализа эволюционных взаимосвязей между бактериофагами и их группами. Задачи таксономии требуют применения комплекса этих инструментов для обоснованной иерархической классификации. Несмотря на некоторые сложности классификационных схем и алгоритмов, биоинформатические методы показывают свою эффективность и безальтернативность при построении таксономии бактериофагов, основанной на геномных данных.

ГЛАВА 2. ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Объекты исследования

Объектами исследования являлись бактериофаги *Curtobacterium* Аука, *Pectobacterium* PP47, PP81, Q19, Possum и Horatius, впервые выделенные в Лаборатории молекулярной биоинженерии Института биоорганической химии РАН и фаг *Pseudomonas* MD8, выделенный и в Лаборатории водной микробиологии Лимнологического института СО РАН. Для обнаружения профаговых бактериальных областей *Curtobacterium* и анализа таксономического разнообразия фагов, инфицирующих фитопатогенные бактерии семейства *Pectobacteriaceae*, использовались геномные данные, опубликованные в открытых базах данных.

2.2. Сбор данных

Геномные последовательности новых бактериофагов были получены сборкой *de novo* данных секвенирования. Геномные последовательности выделенных ранее бактериофагов загружали из базы данных NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) и, при наличии исходных данных секвенирования, проверяли новой сборкой. Экспериментально определённые структуры белков загружали из базы данных RCSB Protein Data Bank (<https://www.rcsb.org/>).

2.3. Методы исследования

2.3.1. Протокол таксономического анализа

Протокол таксономического анализа включал следующие этапы характеристики фагового генома с использованием методов, рекомендованных ICTV:

- анализ организации генома, с учётом состава геномных модулей и их расположения;

- поиск родственных фагов, включая фаги, близкие по средненуклеотидному геномному составу и содержащие гомологичные гены;
- получение кластерной тепловой карты межгеномного сходства исследуемого и родственных фагов;
- филогенетический анализ с использованием последовательностей консервативных белков;
- анализ протеома;
- анализ характерных генов и белков.

2.3.2. Аннотирование фаговых геномов

Геномы собирали с помощью программ SPAdes версий 3.10 – 3.15 [147] и CLC Genomic Workbench версий 7.5 – 21 (<https://digitalinsights.qiagen.com/>). Аннотирование геномов проводили с помощью программы Prokka версий 1.12 – 1.14 [148] с использованием пользовательских баз данных, базы данных RVDB v19.0 [149] и встроенных баз данных Prokka. Открытые рамки считывания (open reading frames, ORF) предсказывали с помощью Prodigal 2.6.1 [150], Glimmer 3.02b [151] и Geneious Prime версий 2019-2022 (<https://www.geneious.com/prime/>), а также проверяли вручную поиском последовательностей Шайна-Дальгарно и более вероятных стартовых кодонов.

Функции кодируемых в геноме белков предсказывали поиском гомологичных последовательностей и сравнением мотивов HMM. Поиск гомологичных последовательностей выполняли с помощью BLAST [98] с использованием баз данных NCBI и пользовательских баз данных BLAST с использованием фаговых последовательностей GenBank. Поиск с использованием HMM-мотивов проводили с помощью Phyre2 [152] и HHpred [153] с использованием встроенных баз данных ECOD_F70, PDB_mmCIF70, Pfam_A, SCOPe70, и UniProt-SwissProt-viral70, а функции присваивали сравнением с

аналогичными белками, используя порог 95% достоверности Phyre2 или вероятности HHpred. Наличие генов тРНК проверяли с помощью tRNAscan-SE [154] и ARAGORN [155]. Промоторы транскрипции искали с помощью PhagePromoter [156]. Генетические карты визуализировали в Geneious Prime.

2.3.3. Сравнение нуклеотидного сходства фаговых геномов

Вычисления среднегеномного нуклеотидного сходства (average nucleotide identity, ANI) фаговых геномов проводили с помощью orthoANIu [157] с использованием алгоритма «USEARCH» для поиска сходства со всеми полными и частичными последовательностями геномов бактериофагов, размещёнными в базе данных NCBI Genome на момент проведения анализа (примерно 24 тысячи последовательностей в 2019 г. и 31 тыс. последовательностей в 2022 г.). Кластерную тепловую карту на основе значений ANI получали с помощью BIONJ [158]. Кластерную тепловую карту межгеномного сходства с учётом доли выровненных участков геномов, рассчитанных программой VIRIDIC [105], получали с помощью встроенных инструментов VIRIDIC с использованием настроек по умолчанию.

2.3.4. Сравнение расположения генов и геномных модулей

Визуальное сравнение геномов бактериофагов выполняли с помощью программ EasyFig [159], использующей для анализа сходства отдельных генов TBLASTX.

2.3.5. Выравнивания последовательностей

Выравнивания аминокислотных и нуклеотидных последовательностей для построения филогенетических деревьев делали с помощью Clustal Omega 1.2.3 [160] с настройками «ten refinement iterations, evaluating full distance matrix for initial and guide trees», MAFFT 7.48 [161,162] с настройками по умолчанию и с использованием алгоритма L-INS-i и MUSCLE 3.8 [163] с настройками по умолчанию.

2.3.6. Филогенетический анализ

Филогенетические деревья, основанные на выравнивании аминокислотных и нуклеотидных последовательностей, строили с помощью нескольких алгоритмов, а именно RAxML-NG [118], RAxML [116], MrBayes [122], FastTree [164]. RAxML-NG использовали вместе с графическим интерфейсом raxmlGUI 2 [165] и настройками «--tree rand{10} --bs-trees 1000», применяя весовую матрицу, выбранную с помощью ModelTest-NG [121]. Надежность деревьев RAxML-NG оценивали с помощью бутстрэп-анализа (количество реплик 1000) и вычисления значений бутстрэпа или ожидаемых результатов бутстрэп-анализа (transfer bootstrap expectation, TBE). Дендрограммы RAxML получали с использованием алгоритма «Rapid Bootstrapping and search for best-scoring tree» и весовых матриц BLOSUM62 и LG. Оценку надёжности деревьев RAxML проводили с помощью бутстрэп-анализа (количество реплик 1000). Дендрограммы MrBayes получали с использованием длины цепи 2200000–5500000 и длины «прожига» 200000–500000. Деревья FastTree строили с применением стандартных настроек и весовой матрицы Whelan and Goldman's 2001. Полученные дендрограммы визуализировали в iTOL [166] и Geneious Prime. Сравнение топологии полученных деревьев проводили с помощью ETE 3.1.2 [167] используя настройки «unrooted tree». Для определения топологических дистанций между деревьями и построения тепловых карт использовали нормализованные дистанции Робинсона-Фаулдса [168]. Полученные тепловые карты визуализировали с помощью Python Plotly Express module 5.11.0.

2.3.7. Анализ протеома

Дендрограммы, основанные на сходстве протеома и геномной организации фагов строили с помощью сервера GRAViTy [109,111] с использованием базы данных DB-B: Baltimore Group Ib - Prokaryotic and archaeal dsDNA viruses (VMRv34). Дендрограммы, основанные на сходстве протеома, строили с помощью ViPtree [108] с использованием встроенной базы данных прокариотических вирусов. Карты генетической сети получали с помощью ConTACT.2.0 [137] с

использованием встроенной базы данных ProkaryoticViralRefSeq database v94 и визуализировали в Cytoscape 3.8 [169].

2.3.8. Предсказание третичной структуры белков

Предсказание структур белков делали с помощью алгоритмов глубокого обучения AlphaFold версий 2.0 – 2.2 [170,171] и RoseTTAFold [172,173]. Для моделирования AlphaFold 2 использовали полные базы данных и параметры командной строки «--monomer_casp14», «--monomer» и «--multimer». Моделирование с помощью RoseTTAFold проводили, используя стандартные настройки сервера Robetta (<http://robeta.bakerlab.org/>).

2.3.9. Сравнение структур белков

Сравнение структур белков проводили с помощью программ структурного выравнивания DALI [174,175] и mTM-align [176], используя стандартные настройки. Структурное сходство оценивали по шкалам DALI Z-score [174] и по TM [177]. Матрица структурного подобия, определённого с помощью DALI Z-score, была получена с использованием сервера DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>). Филогенетические деревья, основанные на структурном сходстве, получали с помощью встроенных инструментов DALI и пакета PHYLIP Phylogeny Inference Package 3.6 (<https://evolution.genetics.washington.edu/phylip/>) с использованием метода кластеризации с соседним соединением. Множественное выравнивание последовательностей на основе структурного сходства получали с помощью mTM-align. Суперимпозицию и визуализацию структур осуществляли с помощью PyMOL 2 (<http://www.pymol.org/pymol>).

2.3.10. Поиск и аннотация профаговых последовательностей

Поиск профаговых участков в геномах проводили с использованием пайплайна PhiSpy [178] и сервера Phaster [179] с настройками по умолчанию. ORF проверяли с использованием Glimmer 3.02b, предсказанные последовательности

профагов были аннотированы с использованием Prokka, серверов HHpred и Phyre 2 и поиска BLAST по базам данных NCBI, а также по пользовательским базам данных с использованием фаговых геномных последовательностей.

2.3.11. Рабочие базы данных

Базы данных нуклеотидных последовательностей геномов, генов и аминокислотных последовательностей белков организовывали в среде Geneious Prime. Для создания поисковых баз данных BLAST использовали инструменты BLAST. Для создания баз данных мотивов HMM использовали HMMER 3 [180,181].

ГЛАВА 3. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ²

3.1. Таксономический анализ бактериофагов *Pectobacterium*

3.1.1. Фаги семейства *Autographiviridae*

Фаги PP47, PP89 и Q19 инфицируют фитопатогенные бактерии, относящиеся к роду *Pectobacterium*, включая фитопатогенные штаммы *Pectobacterium brasiliense*, встречающиеся в Центрально-европейской части России [182]. Эти фаги показывают эффективное антибактериальное действие против похожего круга (спектра) хозяев и обладают типичной подовирусной морфологией, схожей с морфологией фага *Escherichia* T7, относящегося к семейству *Autographiviridae* (рис. 18). Размер линейных дцДНК-геномов фагов составляет 40995 н.п. (нуклеотидных пар) (фаг PP47), 40751 н.п. (фаг PP81) и 40227 н.п. (фаг Q19). ГЦ-состав геномов PP47 и PP81 составил 48,9%, генома Q19 – 49.7%, что близко к ГЦ-составу бактериальных хозяев фагов.

² Основные результаты, изложенные в данной главе, опубликованы в следующих научных статьях автора в журналах, индексируемых в базах данных WoS, Scopus и RSCI, рекомендованных для защиты в диссертационном совете МГУ имени М.В. Ломоносова:

1. **Evseev P.V.**, Lukianova A.A., Shneider M.M., Korzhenkov A.A., Bugaeva E.N., Kabanova A.P., Miroshnikov K.K., Kulikov E.E., Toshchakov S.V., Ignatov A.N., Miroshnikov K.A. Origin and Evolution of *Studiervirinae* Bacteriophages Infecting *Pectobacterium*: Horizontal Transfer Assists Adaptation to New Niches // *Microorganisms*, 2020, Vol. 8, No. 11, P. 1707
2. Miroshnikov K.A., **Evseev P.V.**, Lukianova A.A., Ignatov A.N. Tailed Lytic Bacteriophages of Soft Rot *Pectobacteriaceae* // *Microorganisms*, 2021, Vol. 9, No. 9, P. 1819.
3. **Evseev P.**, Lukianova A., Sykilinda N., Gorshkova A., Bondar A., Shneider M., Kabilov M., Drucker V., Miroshnikov K. *Pseudomonas* Phage MD8: Genetic Mosaicism and Challenges of Taxonomic Classification of Lambdoid Bacteriophages // *International Journal of Molecular Sciences*, 2021, Vol. 22, No. 19, P. 10350.
4. **Evseev P.**, Shneider M., Miroshnikov K. Evolution of Phage Tail Sheath Protein // *Viruses*, 2022, Vol. 14, No. 6, P. 1148.
5. Lukianova A.A., **Evseev P.V.**, Shneider M.M., Dvoryakova E.A., Tokmakova A.D., Shpirt A.M., Kabilov M.R., Obraztsova E.A., Shashkov A.S., Ignatov A.N., Knirel Y.A., Dzhililov F.S.-U., Miroshnikov K.A. *Pectobacterium* versatile Bacteriophage Possum: A Complex Polysaccharide-Deacetylating Tail Fiber as a Tool for Host Recognition in Pectobacterial *Schitoviridae* // *International Journal of Molecular Sciences*, 2022, Vol. 23, No. 19, P. 11043.
6. Tarakanov R.I., Lukianova A.A., **Evseev P.V.**, Pilik R.I., Tokmakova A.D., Kulikov E.E., Toshchakov S.V., Ignatov A.N., Dzhililov F.S.-U., Miroshnikov K.A. Ayka, a Novel *Curtobacterium* Bacteriophage, Provides Protection against Soybean Bacterial Wilt and Tan Spot // *International Journal of Molecular Sciences*, 2022, Vol. 23, No. 18, P. 10913.
7. **Evseev P.**; Lukianova, A.; Tarakanov R.; Tokmakova A.; Popova A.; Kulikov E.; Shneider, M.; Ignatov, A.; Miroshnikov, K. Prophage-Derived Regions in *Curtobacterium* Genomes: Good Things, Small Packages // *International Journal of Molecular Sciences* 2023, Vol. 24, No. 2, P. 1586.
8. **Evseev P.**; Gutnik D.; Shneider M.; Miroshnikov K. Use of an Integrated Approach Involving AlphaFold Predictions for the Evolutionary Taxonomy of *Duplodnaviria* Viruses // *Biomolecules* 2023, Vol. 13, No. 1, P. 110.

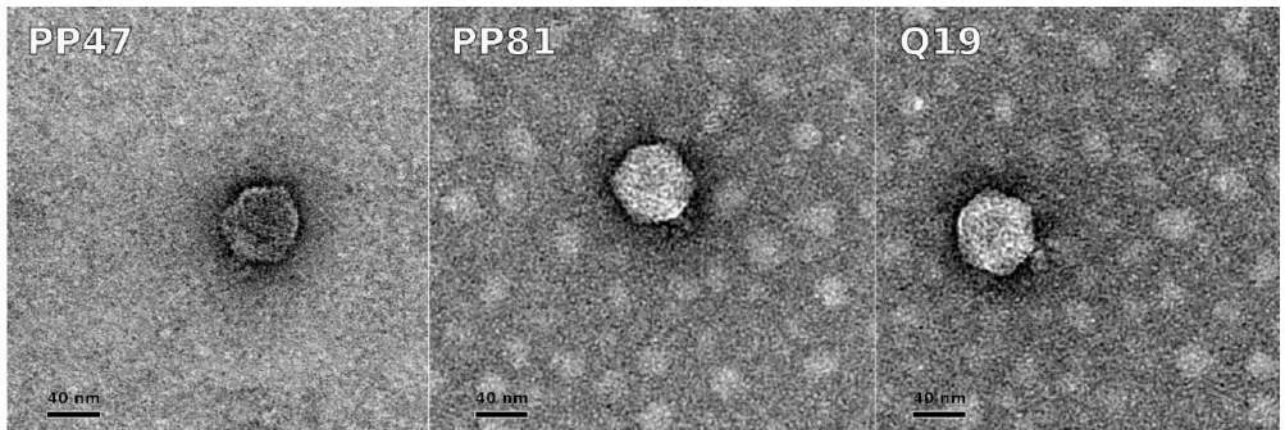


Рисунок 18. Электронная микрофотография фагов *Pectobacterium* PP47, PP81 и Q19. Масштаб шкалы — 40 нм. Электронная микроскопия выполнена Е.Е. Куликовым, микрофотографии опубликованы в статье, написанной совместно с автором диссертации [183].

Геномы фагов PP47, PP89 очень похожи друг на друга за исключением того, что геном PP47 содержит ген №22 (gene 22, g22), кодирующий гипотетический белок. Геномная организация фагов PP47, PP89 и Q19 близка к таковой для фагов семейства *Autographiviridae* (рис. 19, 20). В целом организация генома фагов PP47, PP81 и Q19 типична для *sensu lato* T7-подобных фагов, в настоящее время включающих подсемейство *Studiervirinae* в составе семейства *Autographiviridae*. Однонаправленные открытые рамки считывания можно разделить на три основные функциональные области, связанные с конверсией хозяина, репликацией и морфогенезом. Последние, морфогенетические блоки всех трех фагов очень похожи и кодируют 12 белков, гомологичных структурным белкам других фагов подсемейства *Studiervirinae* за исключением предполагаемого минорного капсидного белка, ген которого расположен сразу после гена ГКБ (главного капсидного белка). Расположение предсказанных промоторов в геномах фагов PP47/81 и Q19 практически идентично (рис. 19) и близко к расположению соответствующих промоторов в геноме фага *Escherichia* T7 [184].

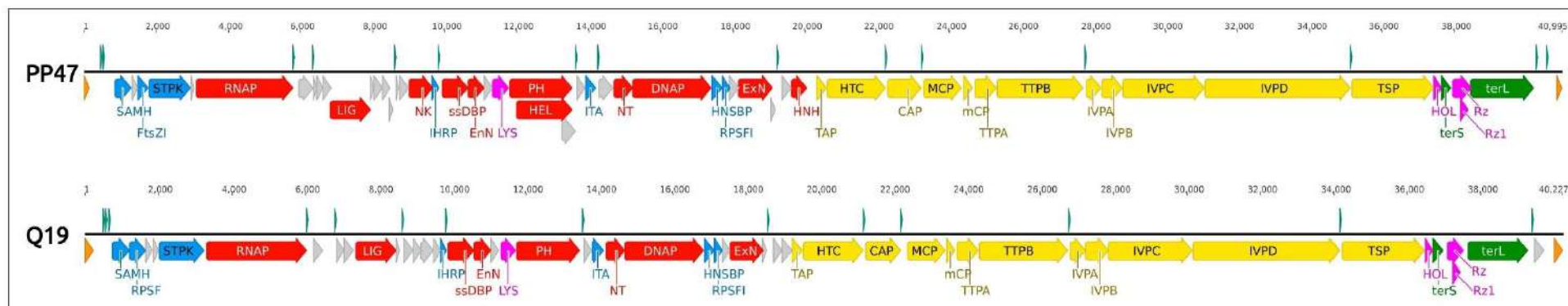


Рисунок 19. Генетическая карта фагов PP47 и Q19. Цвета различных геномных модулей соответствуют их функциональности: желтый – морфогенез; красный – репликация, транскрипция и процессинг нуклеиновых кислот; зеленый – упаковка; фиолетовый – лизирующие белки; синий – регуляция защитных и метаболических процессов хозяина; оранжевый – терминальные повторы; серый – гипотетические белки. Предполагаемые промоторы транскрипции показаны в виде зелёно-голубых стрелок над каждой последовательностью, представленной черной линией. Числа над последовательностями показывают расстояние в н.п. от начала генома. Аббревиатуры генов: SAMH – S-аденозил-L-метионингидролаза; RPSF – фактор $\sigma 54$ РНК-полимеразы; FtsZI – ингибитор FtsZ клеточного деления; STPK – серил-треонилпротеинкиназа; РНКП – ДНК-зависимая РНК-полимераза; LIG – ДНК-лигаза; NK – нуклеотидкиназа; IHRP – ингибитор бактериальной РНК-полимеразы хозяина; ssDBP – оцДНК-связывающий белок; EnN – эндонуклеаза I; LYS, лизоцим – N-ацетилмурамоил-L-аланинамидаза; PH – ДНК-примаза/хеликаза; HEL – геликаза ДНК; ITA – ингибитор системы токсин/антитоксин; NT – нуклеотидилтрансфераза; ДНКП – ДНК-полимераза; HNSBP – H-NS и тРНК-связывающий белок; RPSFI – ингибитор фактора $\sigma 70$ РНК-полимеразы хозяина; ExN – 5'-3'-экзонуклеаза; HNH – эндонуклеаза HNH; TAP – белок сборки хвоста; HTC – соединительный белок головки и хвоста фага; CAP – белок сборки капсида; MCP – главный капсидный белок; mCP – минорный капсидный белок; TTPA – белок хвостовой трубки А; TTPB – белок хвостовой трубки В; IVPA – внутренний белок вириона А; IVPB – внутренний белок вириона В; IVPC – внутренний белок вириона С; IVPD – внутренний белок вириона D; TSP – белок хвостового шипа, содержащий домен гидролазы SGNH; HOL – холин II класса; terS – малая субъединица терминазы; Rz – белок лизиса Rz; Rz1 – белок лизиса Rz1; terL – большая субъединица терминазы.

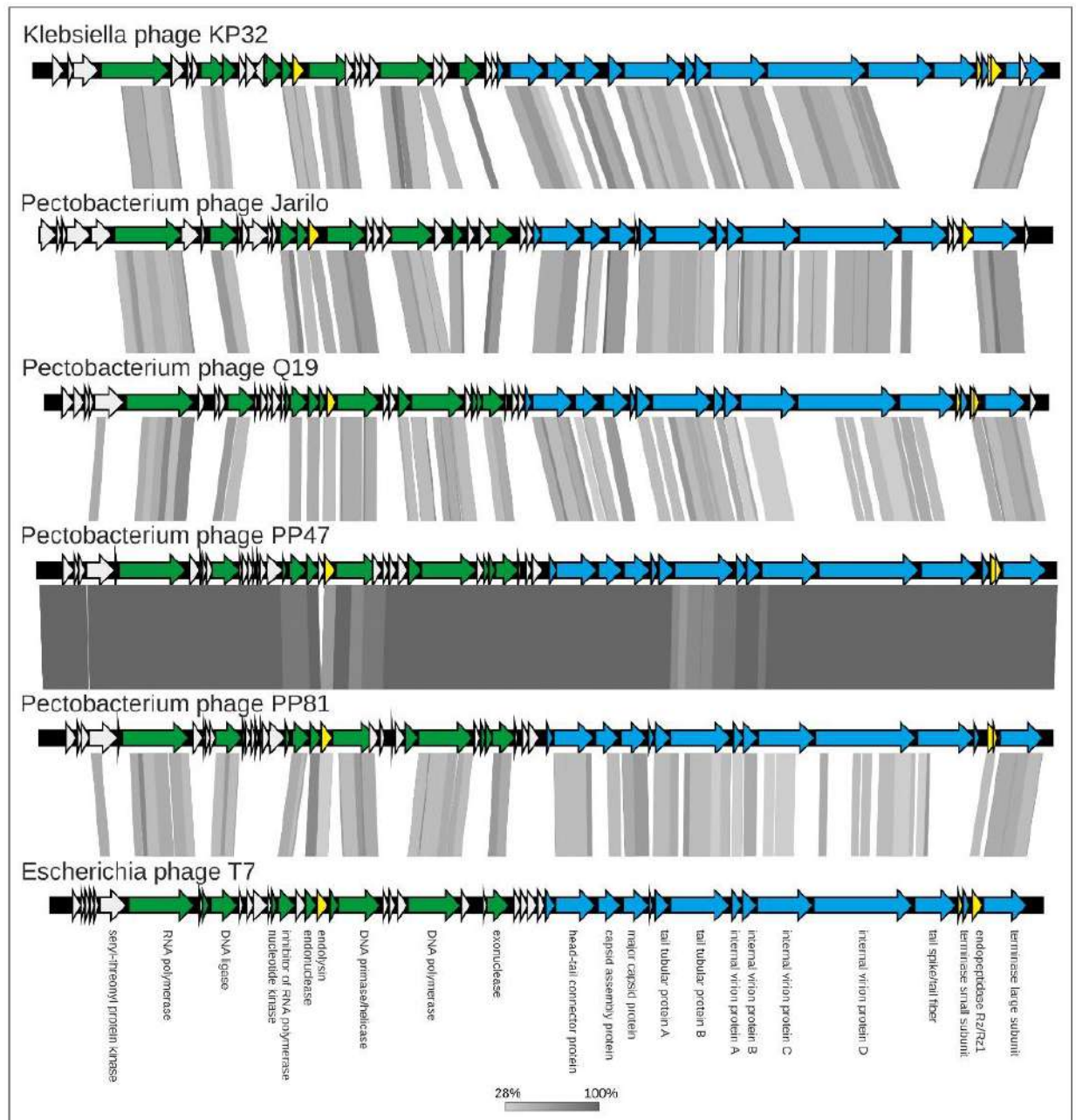


Рисунок 20. Сравнение организации геномов фагов PP47, PP89 и Q19 и других фагов семейства *Autographiviridae*. Гены белков, участвующих в транскрипции и репликации окрашены в зелёный цвет, гены структурных и упаковочных белков окрашены в голубой цвет гены белков лизиса окрашены в жёлтый цвет.

Сравнение нуклеотидного геномного при помощи ортоANI с использованием всех известных полных геномных последовательностей фагов *Autographiviridae* показало значительное сходство между фагами PP47 и PP81 (ANI около 98%), а также меньшее сходство этих фагов с Q19 и другими фагами

Autographiviridae. Учитывая 95%-ный видовой порог нуклеотидного сходства, близость геномов PP47 и PP81 позволяет отнести их к одному виду. Кластеризация фаговых геномов с помощью рекомендованной ICTV программы VIRIDIC сгруппировала фаги PP47 и PP81 в один кластер вместе с фагами рода *Pektosvirus* подсемейства *Studiervirinae* (рис. 21).

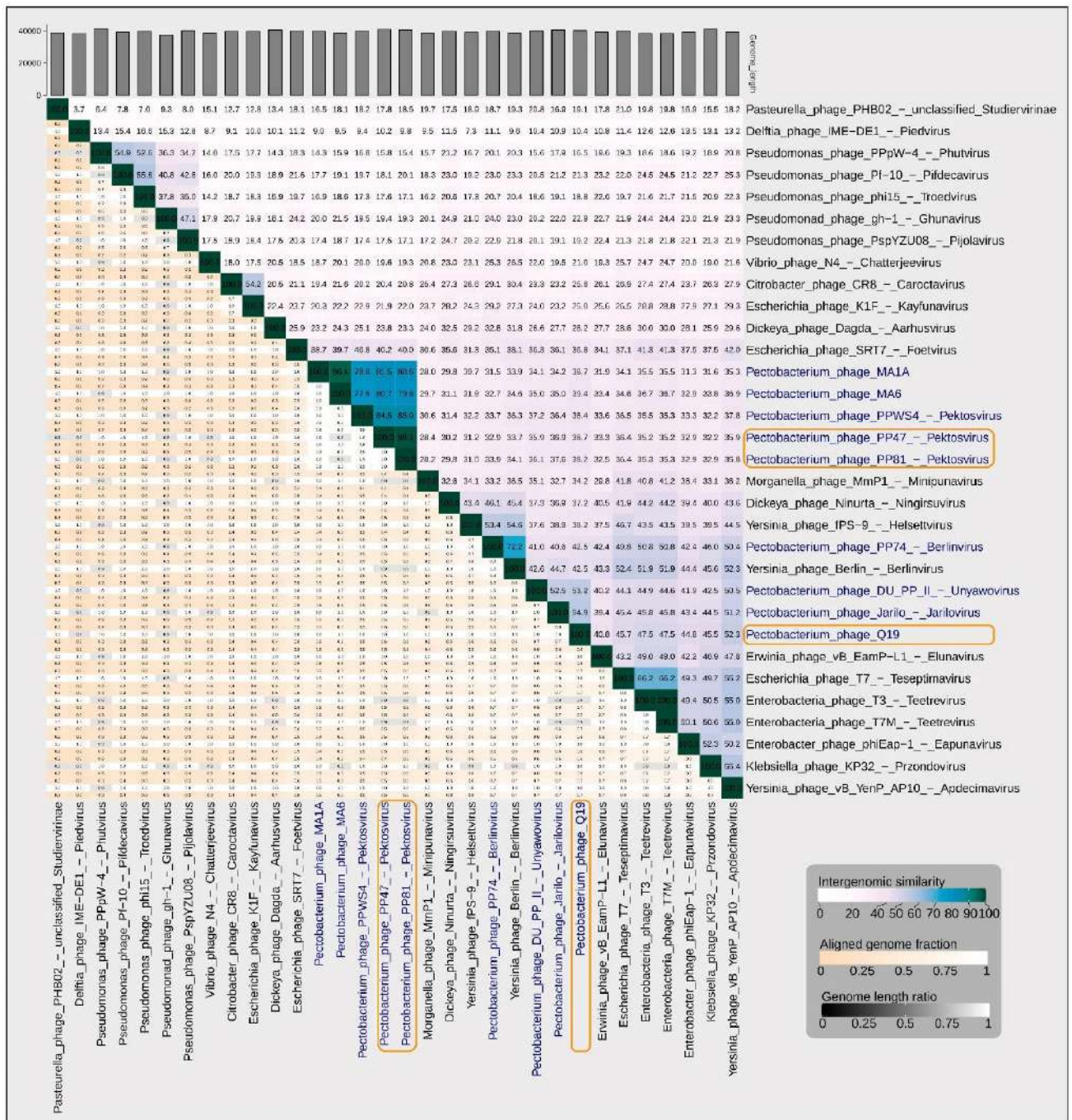


Рисунок 21. Кластерная тепловая карта фагов PP47, PP89 и Q19 и родственных, рассчитанного VIRIDIC на основе межгеномного сходства.

Согласно результатам вычислений VIRIDIC, фаг *Pectobacterium* Q19 ближе всего к фагам родов *Jarilovirus* и *Unyawovirus*, но значения межгеномного сходства Q19 с представителями *Jarilovirus* и *Unyawovirus* составили около 55% и 53% соответственно, что ниже, чем 65-70%-ная граница разграничения родов. Это не позволяет отнести Q19 к двум вышеупомянутым или другим родам. Межгеномное сходство фагов PP47, PP81 и других фагов внутри этого кластера составило более 70%, что позволяет отнести фаги PP47, PP81, а также PPWS4, MA1A и MA6 [185] к роду *Pektosvirus*.

Филогенетический анализ с использованием конкатенированных аминокислотных последовательностей консервативных генов (рис. 22) подтверждает родство фагов PP47, PP81 и фагов рода *Pektosvirus*, а также указывает на близость фага Q19 и фагов родов *Jarilovirus* и *Unyawovirus*. Филогения, основанная на анализе фагового протеома (рис. 23), поместила фаги PP47, PP81 и Q19 в кладу с представителями фагов подсемейства *Studiervirinae*, а также сгруппировала фаги PP47, PP81 вместе с фагами рода *Pektosvirus* (PPWS4, MA6 и MA1A) и отдельно от Q19.

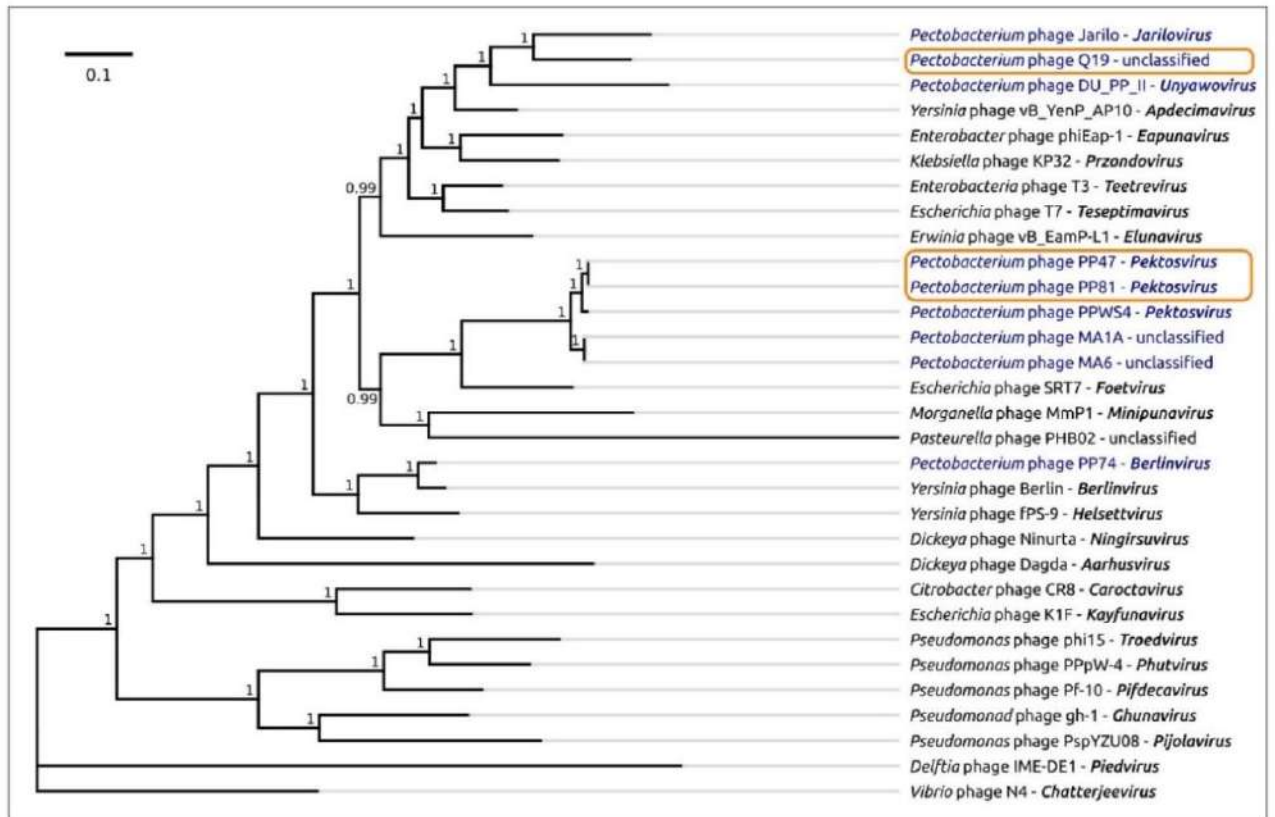


Рисунок 22. Филогенетическое дерево, построенное с использованием конкатенированных аминокислотных последовательностей пяти консервативных белков. Родовая принадлежность указана справа от названия фага. Значения апостериорной вероятности указаны около соответствующих веток.

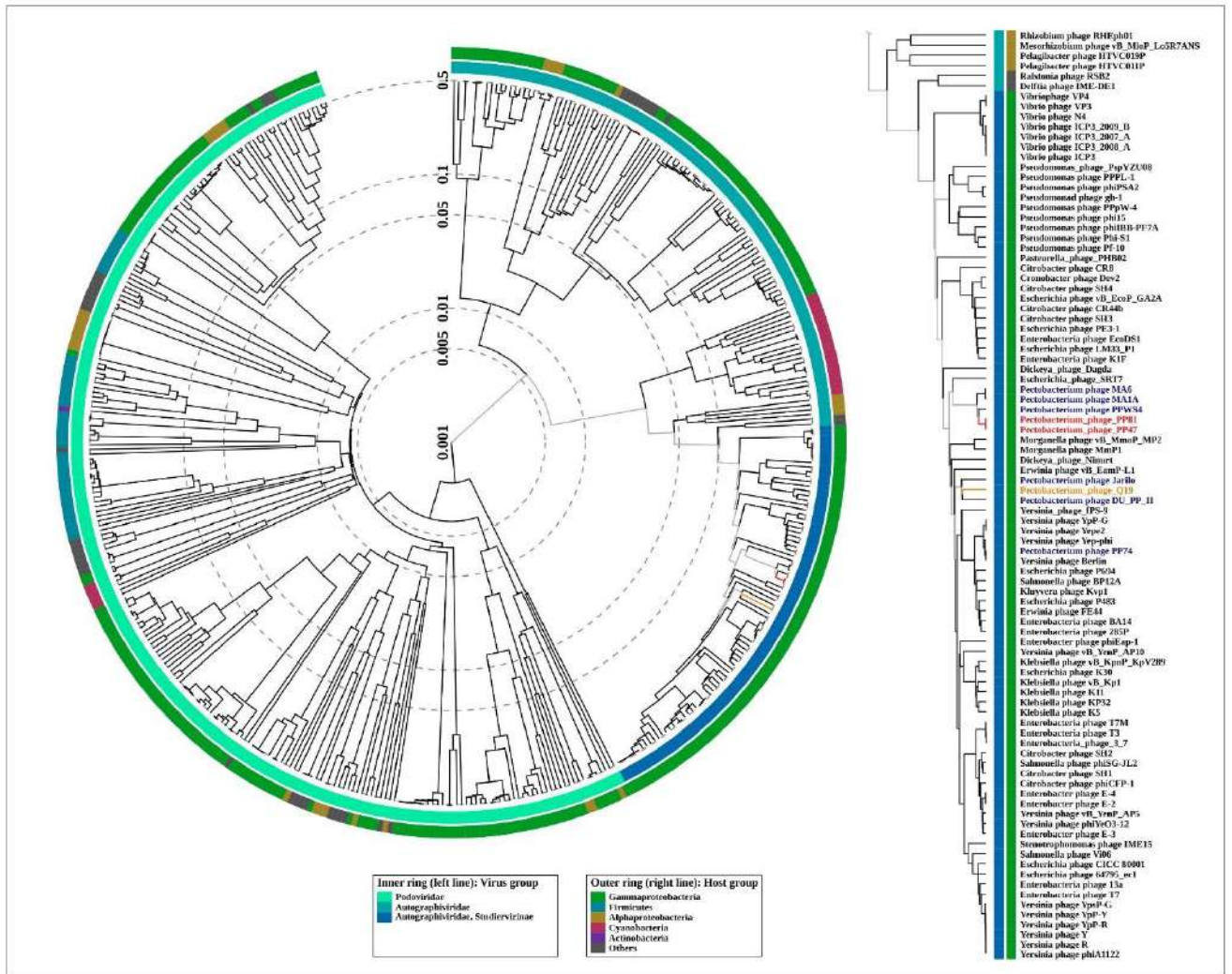


Рисунок 23. Дендрограмма, основанная на сходстве предсказанного протеома, полученная с помощью ViPtree.

Данные межгеномного сравнения и филогенетического анализа указывают на то, что согласно критериям ICTV, фаг Q19 может быть классифицирован как представитель нового рода подсемейства *Studiervirinae*. Фаги PP47 и PP81 можно классифицировать как представителей рода *Pektosvirus* подсемейства *Studiervirinae* и рассматривать как расы фагов одного вида.

Филогенетический и структурный анализы также показали интересную особенность белка хвостового шипа, участвующего в адсорбции фага на клетку хозяина, которая может объяснить близкий спектр хозяев фагов PP47, PP89 и Q19. Ранее было отмечено, что рецептор-связывающие белки (РСБ) разных фагов, включая фаг T7, характеризуются модульной структурной архитектурой [186,187]. При этом С-концевой домен хвостовых шипов, отвечающий за специфичность

фагов, и N-концевой домен, отвечающий за крепление шипа к вириону, могут иметь разную эволюционную историю, включающую горизонтальные переносы с участием бактерий-хозяев. С-концевые домены хвостовых шипов фагов PP47 и PP89 филогенетически ближе к С-концевому домену хвостового шипа фага Q19, чем к С-концевым доменам хвостовых фибрилл фагов *Pektosvirus*, что может объяснить сходный спектр хозяев PP47, PP89 и Q19 (рис. 24, 25).

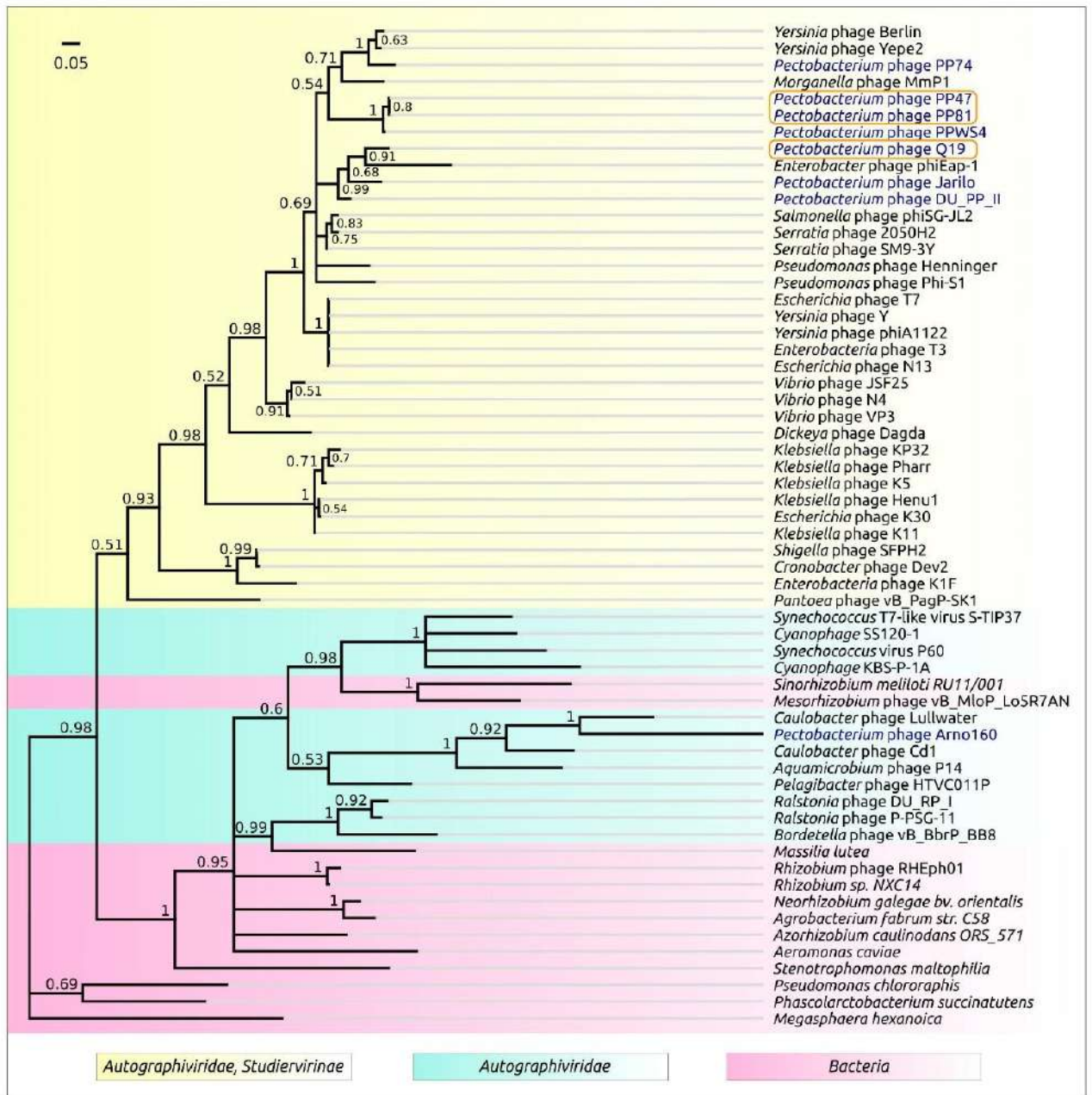


Рисунок 24. Филогенетическое дерево, построенное с использованием аминокислотных последовательностей N-концевого домена РСБ.

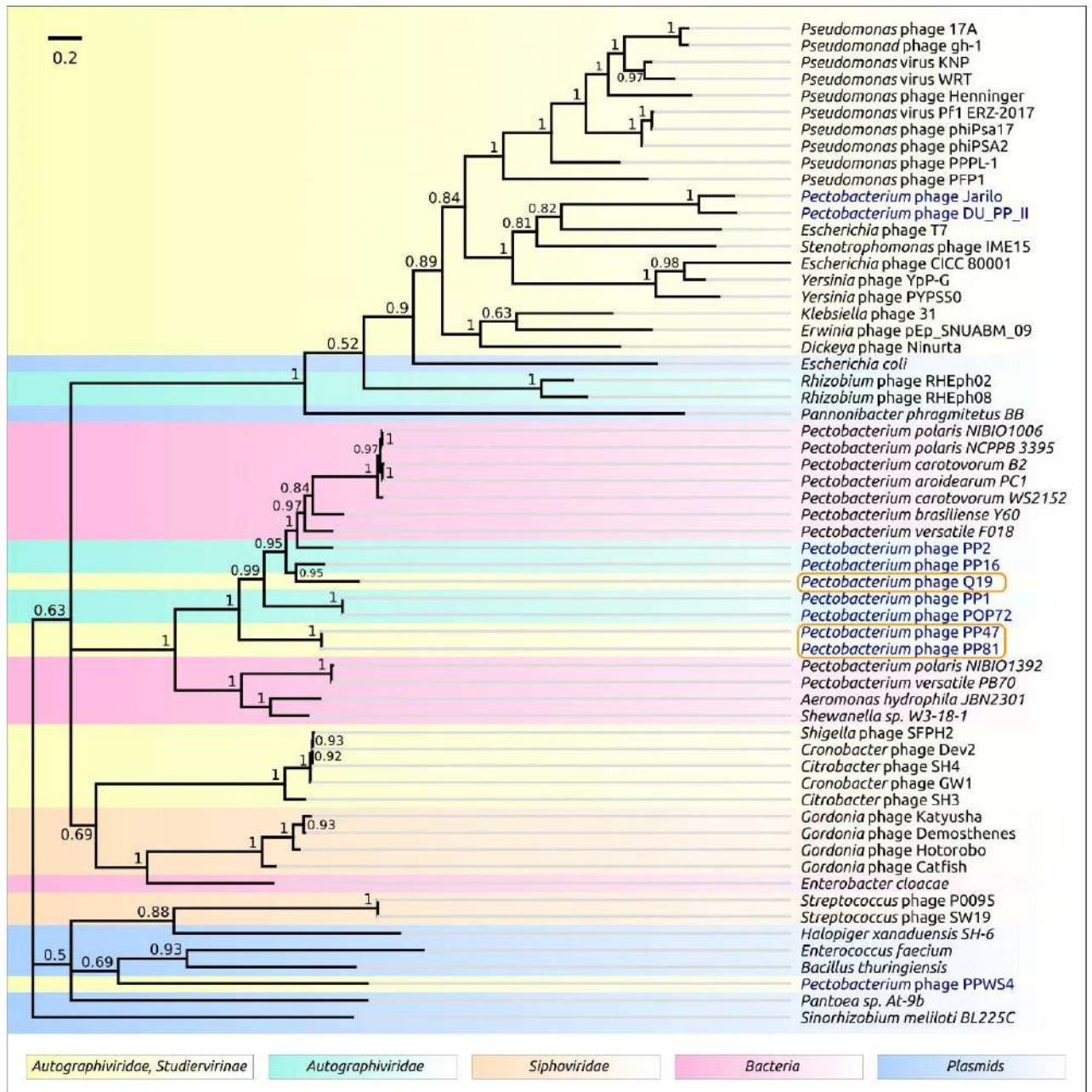


Рисунок 25. Филогенетическое дерево, построенное с использованием аминокислотных последовательностей С-концевого домена РСБ.

3.1.2. Фаги семейства *Schitoviridae*

Фаги семейства *Schitoviridae* *Pectobacterium* Possum и *Noratus* инфицируют сходный спектр хозяев и морфологически неразличимы (рис. 26). дцДНК линейные геномы этих двух фагов размером 73752 н.п. (Possum) и 73737 (Noratus) почти идентичны, за исключением вставки длиной 15 п.н. в гене *rIIB* фага Possum и

нескольких однонуклеотидных замен, не влияющих на аминокислотные последовательности кодируемых белков.

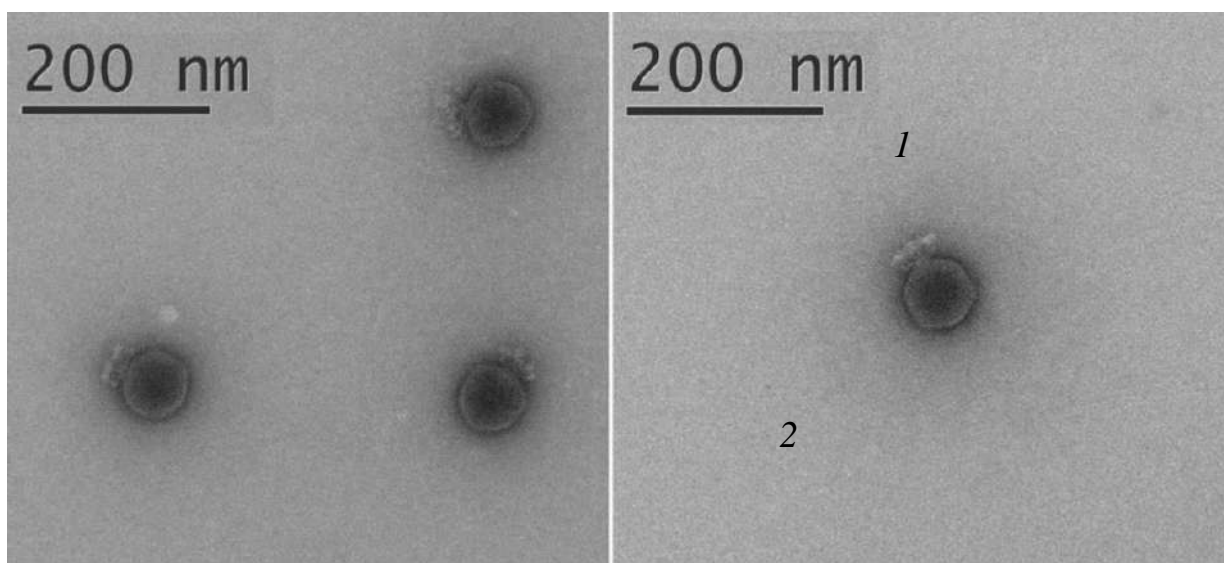


Рисунок 26. Электронная микрофотография фагов *Pectobacterium Possum* (слева) и *Horatius* (справа). Масштаб шкалы — 200 нм. Электронная микроскопия выполнена Е.Е. Куликовым, микрофотографии опубликованы в статье, написанной совместно с автором диссертации [188].

Размер и организация геномов *Possum* (рис. 27) и *Horatius* близки к таковым у фагов *Pectobacterium* CB1, CB3 и CB4 [189], *Nepra* [190], фА38 и фА41 [191], принадлежащих к группе N4-подобных фагов, сравнительно недавно выделенных в новое семейство *Schitoviridae* (ратифицировано в марте 2021 года) и отнесённых к роду *Sbunavirus* (рис. 28). Гены этих фагов сгруппированы в несколько блоков, ориентированных в разных направлениях. Первые 60 генов фага *Possum* (ранняя область и большинство генов репликации и транскрипции) ориентированы в прямом направлении, следующие за ними семь генов (кодирующие отростки хвоста и кассету лизиса) – в обратном направлении, следующие 15 генов (в основном принадлежащие к группе генов, кодирующих белки, изменяющие топологию ДНК) снова ориентированы в прямом направлении, и большинство оставшихся генов (кодирующих структурные белки) – обратном. ГЦ-состав генома *Possum* составляет 48,5% и распределён неравномерно, составляя примерно 44% в ранней области и 51% в области структурных генов. Среднее значение ГЦ-состава

меньше, чем ГЦ-состав в бактериальных хозяевах фага. Анализ ГЦ-состава других геномов *Cbnavirus* (47,7–48,6%) также показал значения ниже, чем ГЦ-состав геномов их хозяев. Эта особенность, по-видимому, является общей для некоторых других N4-подобных фагов, инфицирующих *Enterobacterales* (например, ГЦ-состав генома фага *Klebsiella* vB_KpP_FBKp27 – 44,2%, фага *Escherichia* N4 – 41,3%, фага *Escherichia* vB_EcoP_G7C – 43,3% и др.).

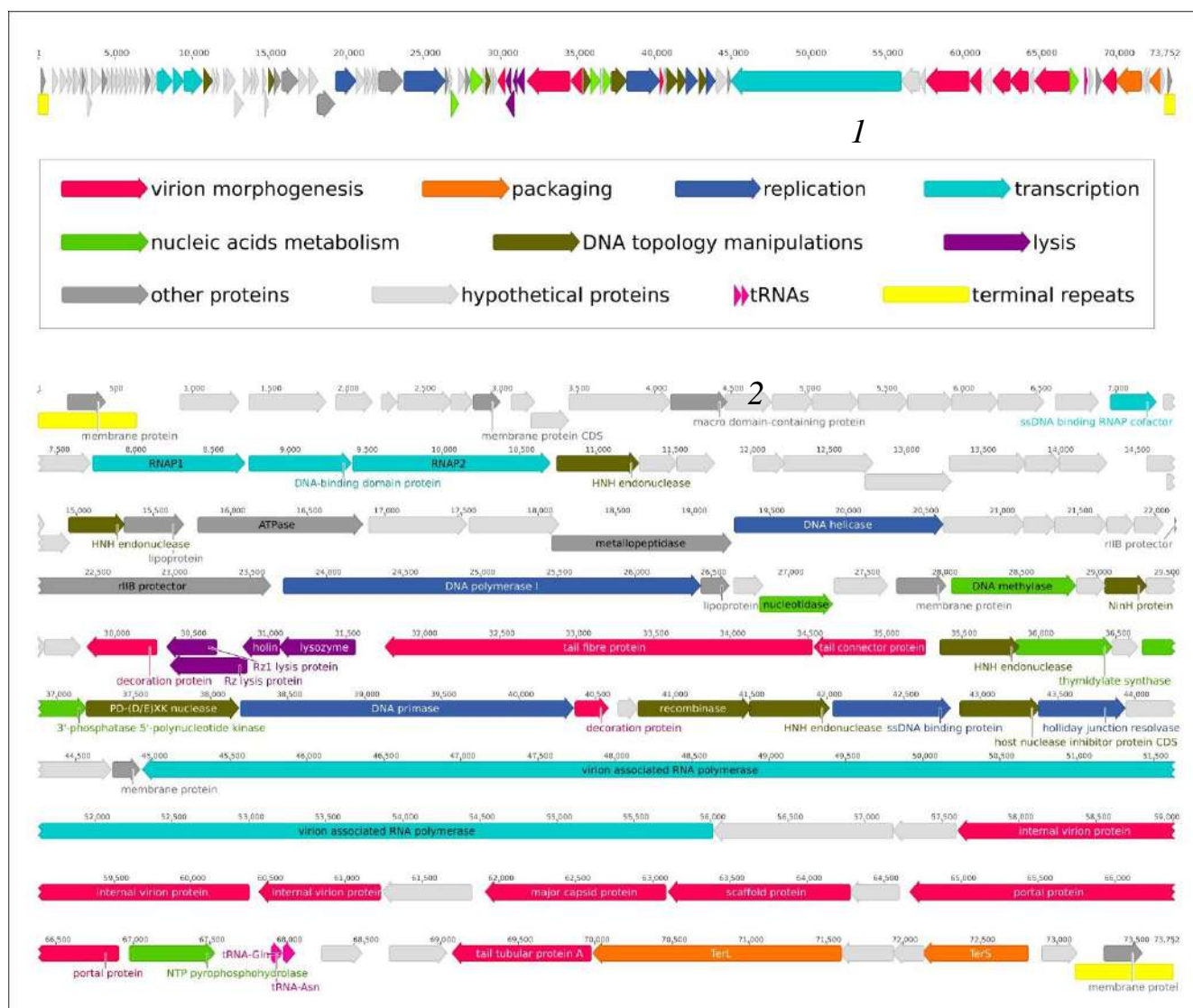


Рисунок 27. Генетическая карта фага *Pectobacterium Rossum*. Цвета различных геномных модулей соответствуют их функциональности: красный – морфогенез; оранжевый – упаковка; синий – репликация; циановый – транскрипция; зелёный – метаболизм нуклеиновых кислот; коричневый – процессы, изменяющие топологию ДНК; пурпурный – лизис; тёмно-серый – другие функции; светло-серый – гипотетические белки; жёлтый – терминальные повторы. Числа над последовательностью показывают расстояние в н.п. от начала генома

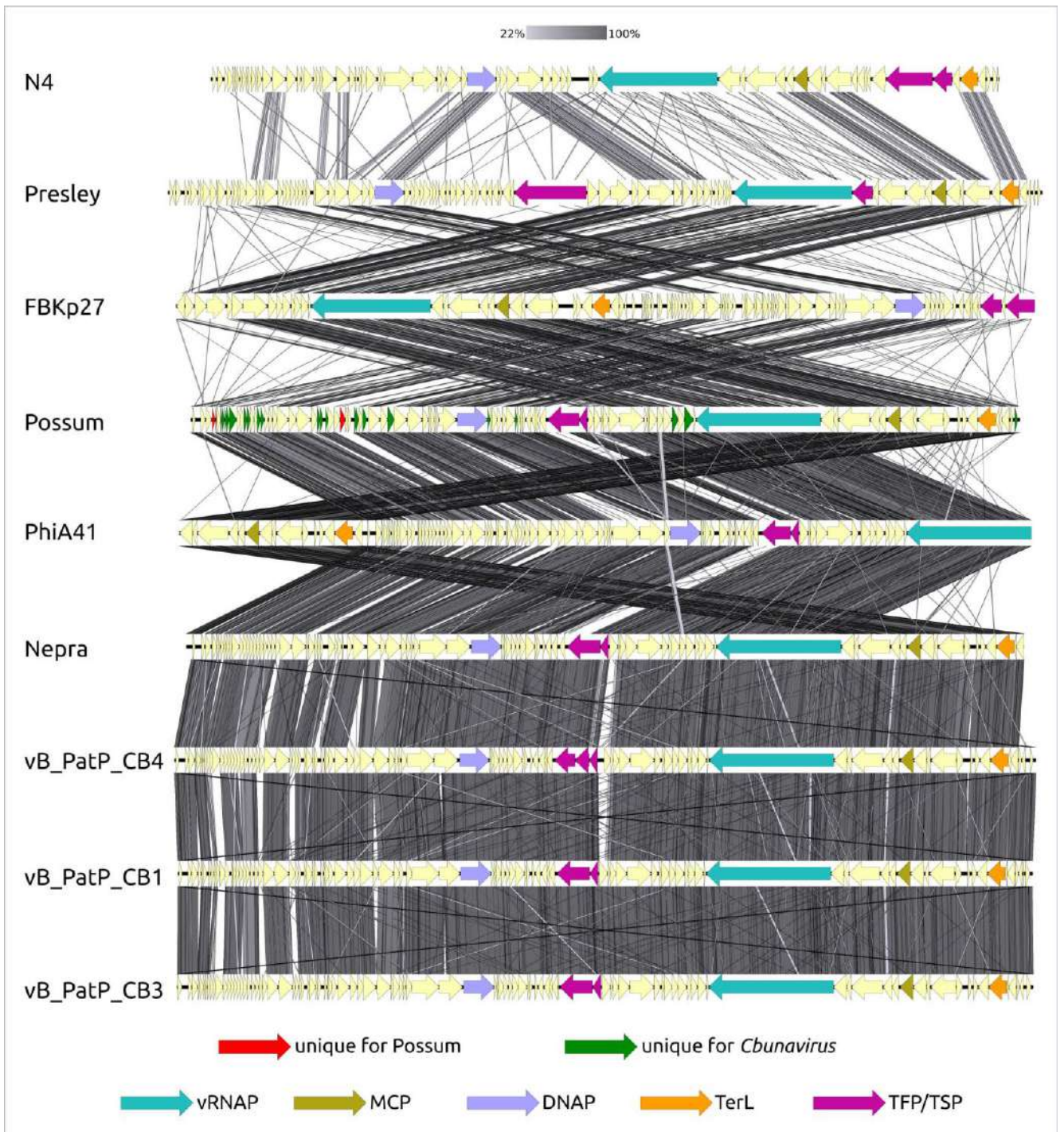


Рисунок 28. Сравнение геномов фага *Pectobacterium Possum* и родственных фагов.

Красным цветом обозначены области генов, в соответствии с поиском BLAST уникальные для фага Possum; зелёным – уникальные для фагов рода *Cbunavirus*.

Вирионная РНКП обозначена циановым цветом, главный капсидный белок – коричневым; ДНК-полимераза – голубым; терминаза – оранжевым; хвостовые фибриллы и хвостовые шипы - пурпурным. Степень сходства геномных участков обозначена градиентной шкалой сверху.

Геномы фагов *Possum* и *Horatius*, как и геномы других фагов семейства *Schitoviridae*, содержат три гена ДНК-зависимых РНК-полимераз (РНКП). Отличительной чертой этого семейства фагов является ген большой инкапсулированной РНКП (вирионной РНКП, вРНКП). вРНКП вводится в инфицируемую бактериальную клетку вместе с ДНК фага, что способствует быстрому началу транскрипции ранних генов.

Для уточнения таксономической классификации фагов *Pectobacterium* *Possum* и *Horatius* были проведены полногеномные сравнения и филогенетические исследования. Расчёты ANI показали, что фаги *Pectobacterium* vB_PatP_CB1, CD3, CB4, фА38, фА41 и Nepra, выделенные в род *Cbunavirus*, являются ближайшими родственниками *Possum* с близкими значениями ANI около 94%. Результаты анализа VIRIDIC (см. приложение А, рис. А.1) продемонстрировали, что все перечисленные выше фаги *Pectobacterium* группируются в один кластер с межгеномным сходством более 70%. Результаты филогенетического анализа с использованием последовательностей главного капсидного белка, порталного белка и терминазы (рис. 29), как и анализ протеома (аминокислотных последовательностей фаговых белков) (рис. 30) показали, что фаги *Possum* и *Horatius*, а также классифицированные фаги рода *Cbunavirus* группируются в одну монофилетическую ветку. Таким образом, по результатам комплексного анализа, включающего сравнение геномных характеристик, организации геномов, межгеномного сходства и филогенетического анализа, фаги *Pectobacterium* *Possum* и *Horatius* можно считать новыми представителями одного рода *Cbunavirus* семейства *Schitoviridae*.

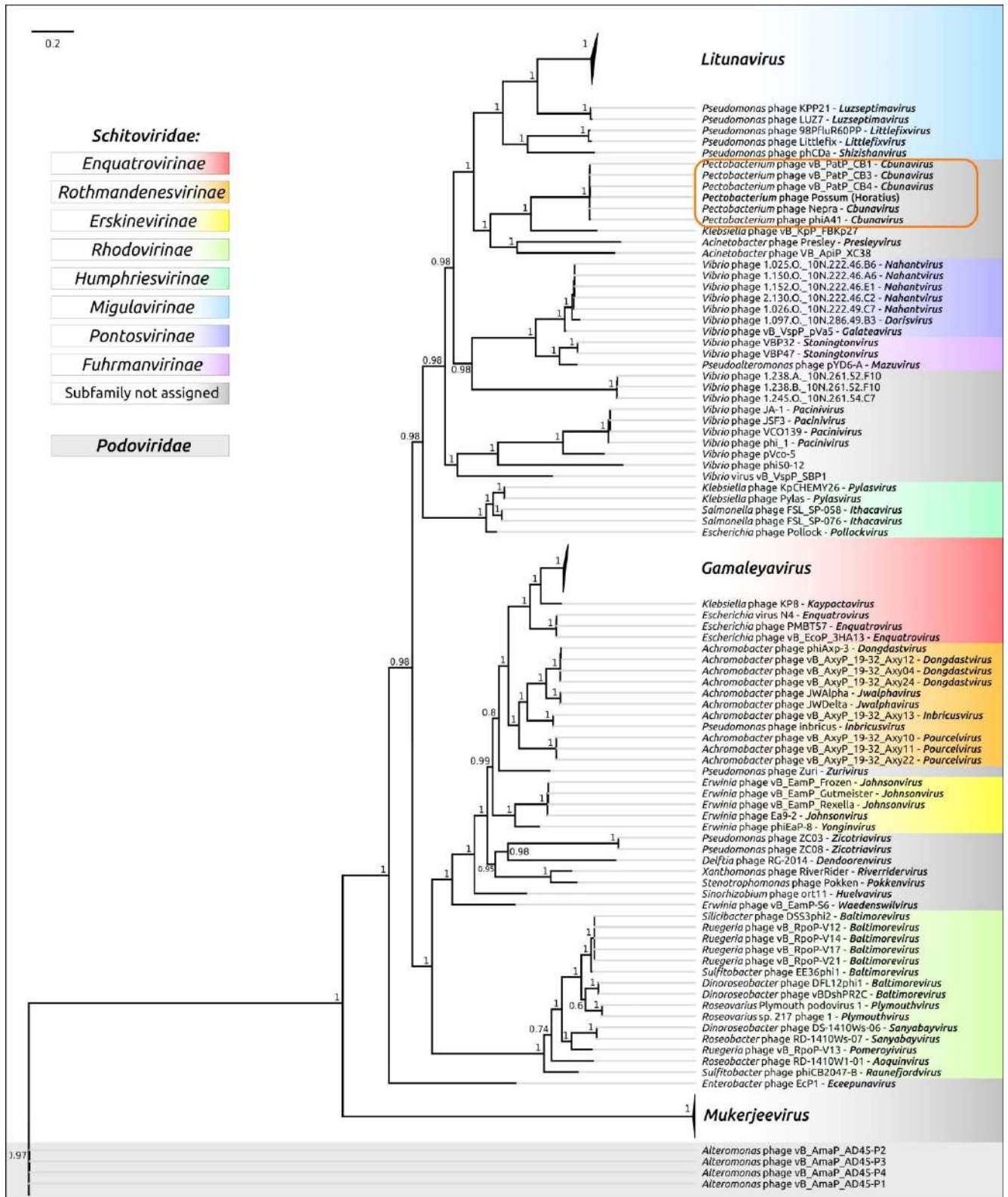


Рисунок 29. Филогенетическое дерево, построенное с использованием конкатенированных аминокислотных последовательностей пяти консервативных белков. Родовая принадлежность указана справа от названия фага. Значения апостериорной вероятности указаны около соответствующих веток. Справа от названий указана таксономическая принадлежность фагов.

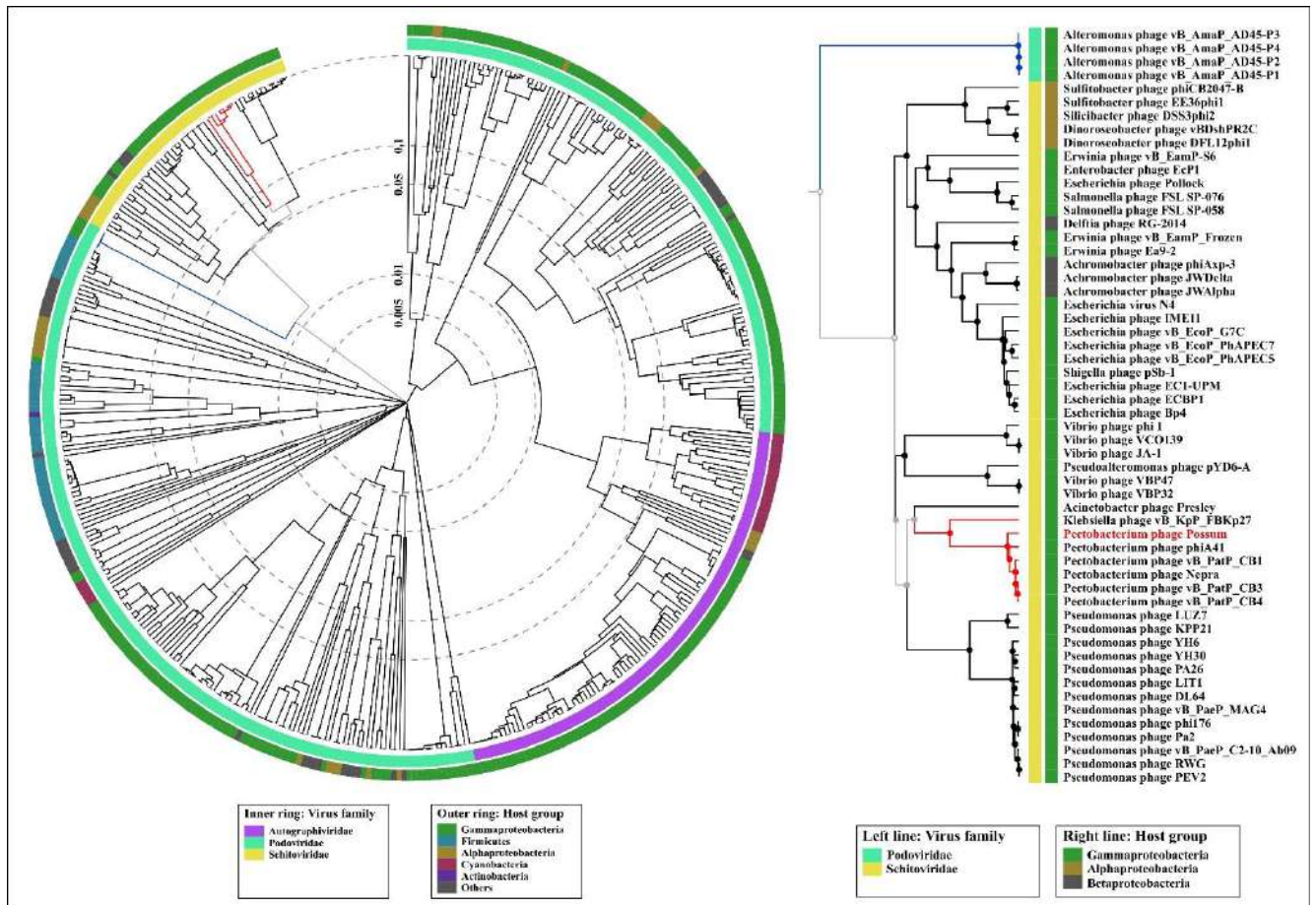


Рисунок 30. Круговая дендрограмма (слева) и её фрагмент (справа), построенная на основании сходства протеомов фагов. В легендах указана таксономическая принадлежность фагов.

3.1.3. Фаги, инфицирующие фитопатогенные бактерии *Pectobacterium* и *Dickeya*

Актуальность изучения фагов, инфицирующих фитопатогенные бактерии *Pectobacterium* и *Dickeya*, во многом связана с тем, что современное сельское хозяйство испытывает значительные трудности с лечением и профилактикой основных болезней растений, вызываемых фитопатогенными бактериями [192]. Фитопатогенные бактерии родов *Pectobacterium* и *Dickeya*, относимые к группе «Soft Rot Pectobacteriaceae» (SRP), являются возбудителями воздушных гнилей, мягкой гнили и черной ножки картофеля, капусты, кукурузы и других культур, а также декоративных растений [193]. Использование защитных химикатов часто малоэффективно. Передовым подходом в биоконтроле является использование

специфических бактериофагов, которые регулируют популяции бактерий в окружающей среде [194,195]. Таксономическое разнообразие SRP значительно: роды *Pectobacterium* и *Dickeya*, ранее считавшиеся «пектолитическими *Erwinia*», в настоящее время включают около 30 отдельных видов [196].

Таксономическое разнообразие фагов, инфицирующих фитопатогенные бактерии *Pectobacterium* и *Dickeya* (PD-фаги), было изучено с использованием геномных последовательностей 108 PD-фагов, размещённых в базе данных NCBI Genome по состоянию на начало 2021 г. В соответствии с номенклатурой на начало 2021 г. эти фаги относились к семи семействам отряда *Caudovirales* и включали разнообразные фаги с различными геномными характеристиками, в том числе хвостатый бактериофаг DU_PP_III с самым маленьким известным геномом размером 11,5 тыс. н.п. и гигантский, т. н. «джамбо-фаг» vB_PsaM_CBV с геномом размером 378 тыс. н.п. Большинство (68 из 108) фагов представляют фаги подовирусной морфологии; 30 фагов инфицируют представителей рода *Dickeya* и 78 фагов заражают бактерии, относящиеся к роду *Pectobacterium*.

Проведённый анализ указал на высокое таксономическое и биологическое разнообразие PD-фагов, включающих представителей 7 семейств и 9 подсемейств (по номенклатуре ICTV на начало 2021 г.) Большую часть PD-фагов (58 геномов) представляют литические фаги семейства *Autographiviridae*, интересные с точки зрения применения в фаговой терапии. Кластеризация фагов, основанная на сравнении ANI (рис. 31) и результаты оценки межгеномного сходства с помощью VIRIDIC указывают на существование нескольких кластеров внутри некоторых классифицированных родов с межгеномным сходством менее 70%. Филогенетический анализ, проведённый с использованием конкатенированных последовательностей главного капсидного белка и большой субъединицы терминазы (рис. 32), также показал наличие отдельных групп внутри этих подсемейств, что может указывать на ещё большее таксономическое разнообразие и возможности дальнейшего усовершенствования таксономической классификации PD-фагов.

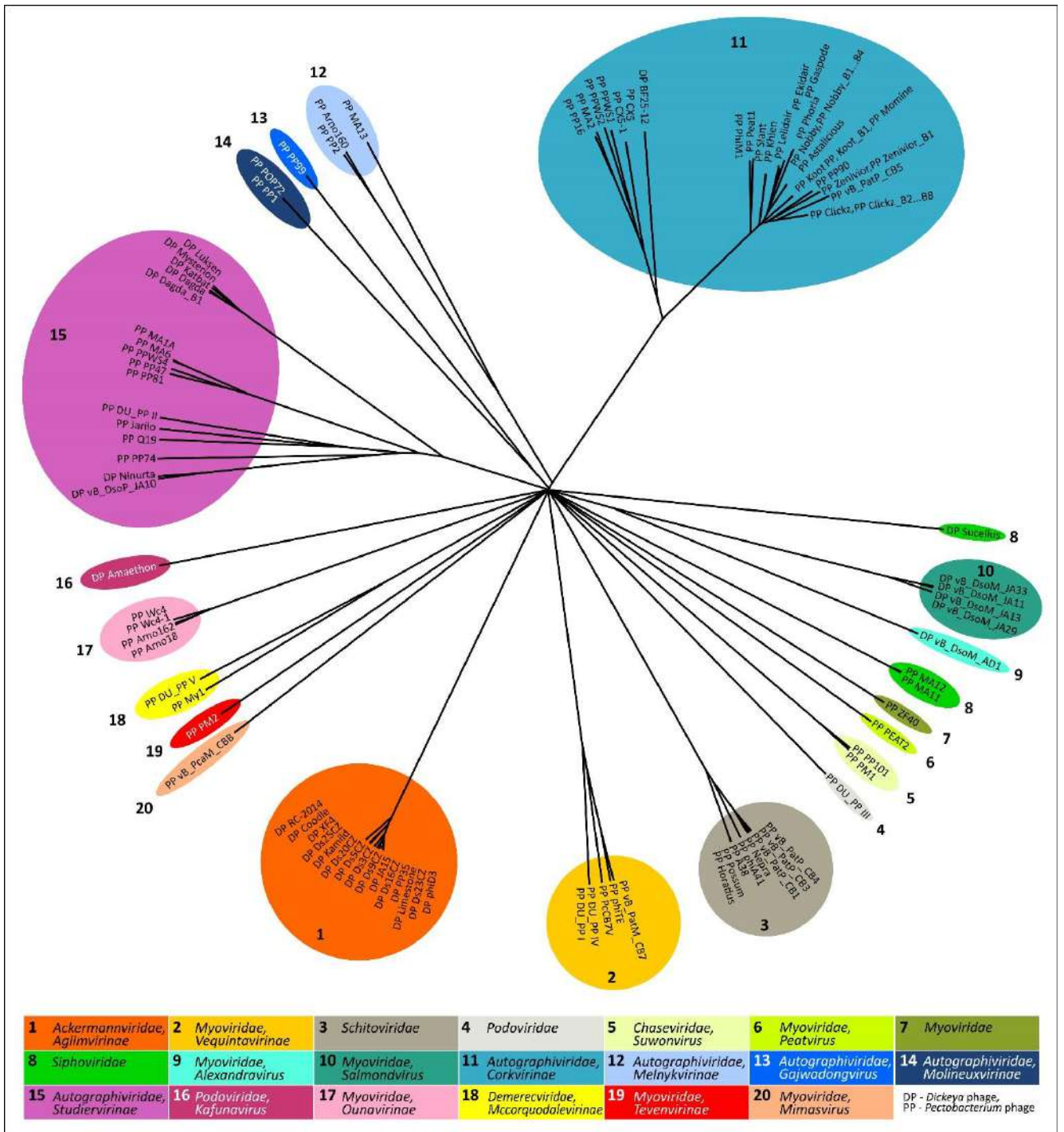


Рисунок 31. Дендрограмма, построенная с использованием значений среднегеномного нуклеотидного сходства ANI. В легендах указана таксономическая принадлежность фагов по состоянию на начало 2021 г.

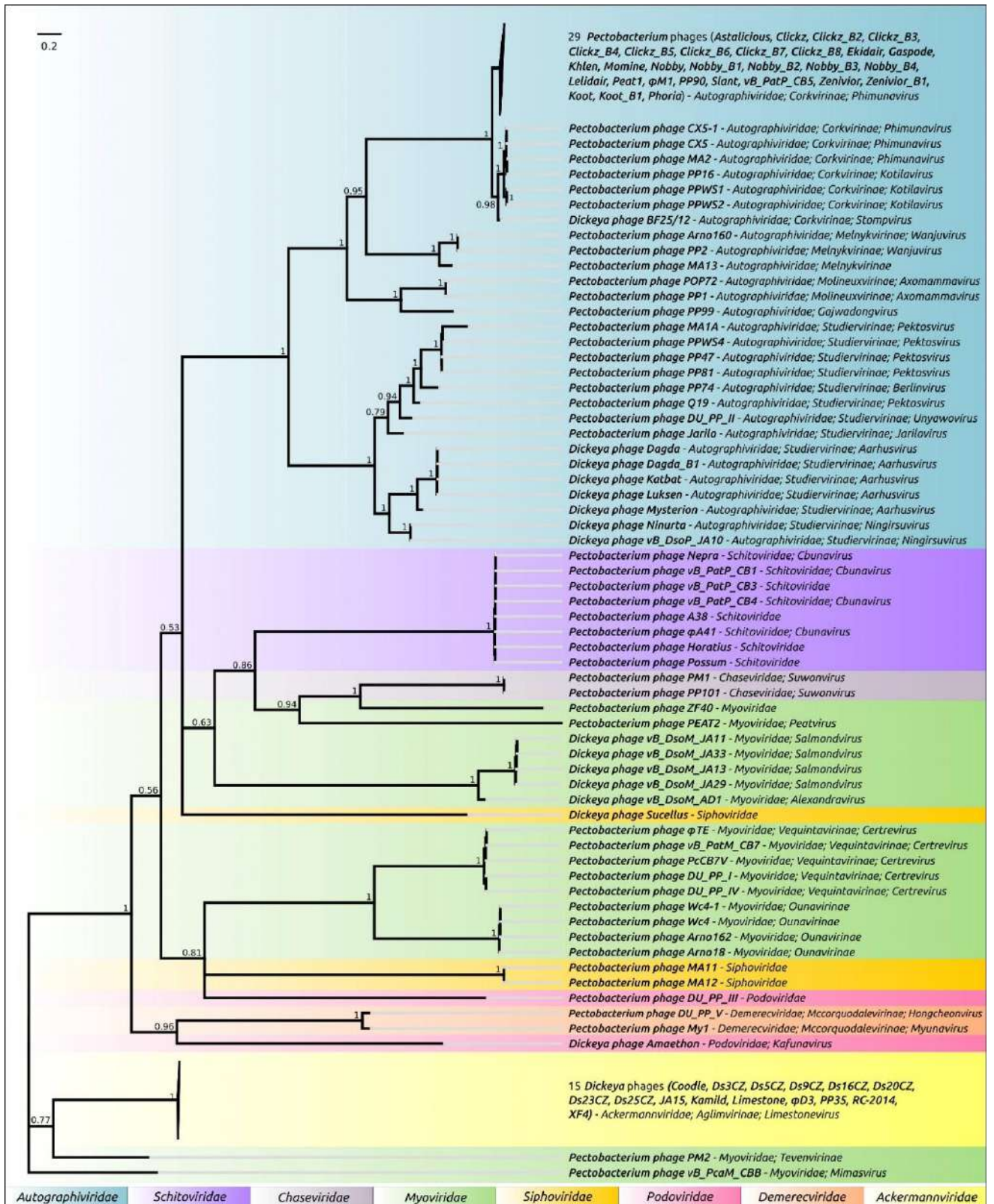


Рисунок 32. Филогенетическое дерево, построенное с использованием конкатенированных аминокислотных последовательностей главного капсидного белка и терминазы. Таксономическая принадлежность по состоянию на начало 2021 г. указана справа от названия фага. Значения апостериорной вероятности указаны около соответствующих веток.

Анализ ранее не аннотированного генома фага *Pectobacterium* DU_PP_III (рис. 33) показал наличие генов, кодирующих белки, более характерные для фагов грамположительных бактерий, родственных фагу *Bacillus* φ29 [197], который относится сейчас к семейству *Salasmaviridae* (ранее – роду *Salasmavirus*). Это может указывать на родственные эволюционные связи между небольшими фагами подовирусной морфологии, инфицирующими грамположительные и грамотрицательные бактерии. Среди белков, характерных и для DU_PP_III и φ29-подобных фагов, – ДНК-полимераза, использующая белковый праймер, репликационный белок и терминаза, схожая с терминазой, использующей для упаковки генома специальную РНК, pRNA (packaging RNA). Филогенетический анализ с использованием главного капсидного белка, порталного белка и ДНКП указал на близкое родство фага DU_PP_III и небольших фагов сальмонелл, а также более дальнее родство этой группы с неклассифицированным фагом *Clostridium* NM2 (приложение А, рис. А.2–А.4).

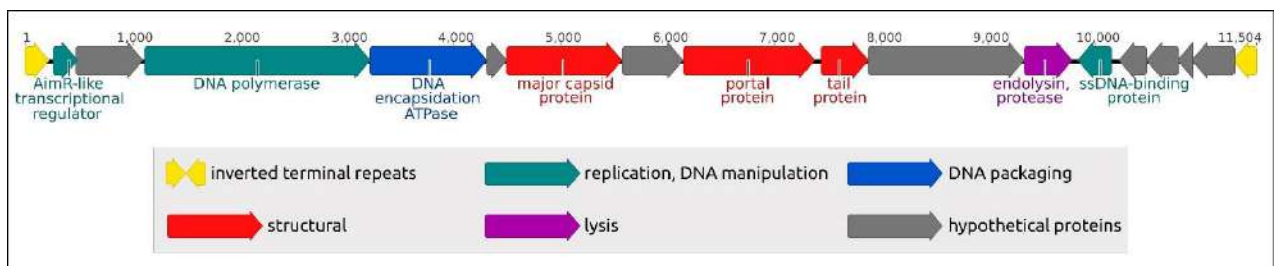


Рисунок 33. Генетическая карта фага *Pectobacterium* DU_PP_III. Цвета различных геномных модулей соответствуют их функциональности: зелёный – процессинг ДНК; синий – упаковка; красный – морфогенез; пурпурный – лизис; серый – гипотетические белки; жёлтый – терминальные повторы. Числа над последовательностью показывают расстояние в н.п. от начала генома.

3.2. Таксономический анализ бактериофагов *Curtobacterium*

3.2.1. Фаг *Curtobacterium* Аука

Фаг Аука (рис. 34) является первым описанным представителем бактериофагов, инфицирующих фитопатогенные бактерии, принадлежащие к роду *Curtobacterium*. Фаг *Curtobacterium* Аука – это дцДНК-вирус с размером генома 18400 пар оснований. ГЦ-состав генома составляет 52,6%, что заметно ниже типичного ГЦ-состава геномов инфицируемых штаммов *Curtobacterium flaccumfaciens* (около 71%). В геноме Аука предсказано 22 открытых рамки считывания (ORF, open reading frame), расположенных на прямой цепи. Предполагаемые функции были присвоены 16 генам, шесть генов были аннотированы как последовательности, кодирующие гипотетические белки.

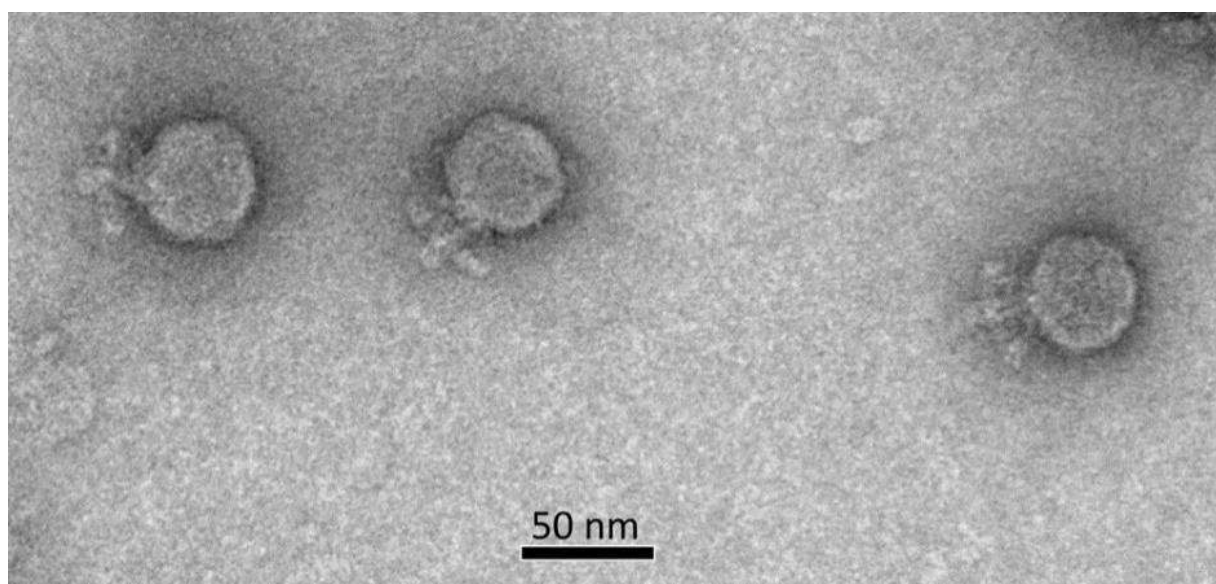


Рисунок 34. Электронная микрофотография фага *Curtobacterium* Аука. Масштаб шкалы – 50 нм. Электронная микроскопия выполнена Е.Е. Куликовым, микрофотография опубликована в статье, написанной совместно с автором диссертации [198].

Геном (рис. 35) содержит блок генов процессинга ДНК, включая ген, кодирующий терминальный белок. Наличие этого гена указывает на белок-праймированный механизм репликации [199], типичный для ряда фагов, родственных упомянутому выше фагу *Bacillus* φ29 и другим фагам *Salasmaviridae*.

Однако, структура генома фага Аука отличается от структуры генома φ29, в том числе, отсутствием ORF, ориентированных в противоположном направлении (рис. 36). Согласно результатам сравнений НММ-НММ, список структурных белков включает φ29-подобный главный капсидный белок и несколько других белков, сходных с φ29-белками. Интересно, что ген g16 кодирует белок хвостовой иглы gp16 (генный продукт 10, gene product 16), который больше напоминает белок хвостовой иглы gp8 фага P68, инфицирующего *Staphylococcus aureus*, чем белки φ29-подобных фагов. Белок хвостовой иглы фага P68 может участвовать в формировании поры в бактериальной мембране для доставки ДНК фага в цитоплазму бактерий [200]. Адсорбционный аппарат фага Аука также включает хвостовые шипы с предсказанной ацетилглюкозаминидазной активностью (gp10). Кассета белков лизиса включает эндолизин и два холина, а упаковочный аппарат включает φ29-подобную терминазу (АТФазу инкапсидации ДНК, gp13).

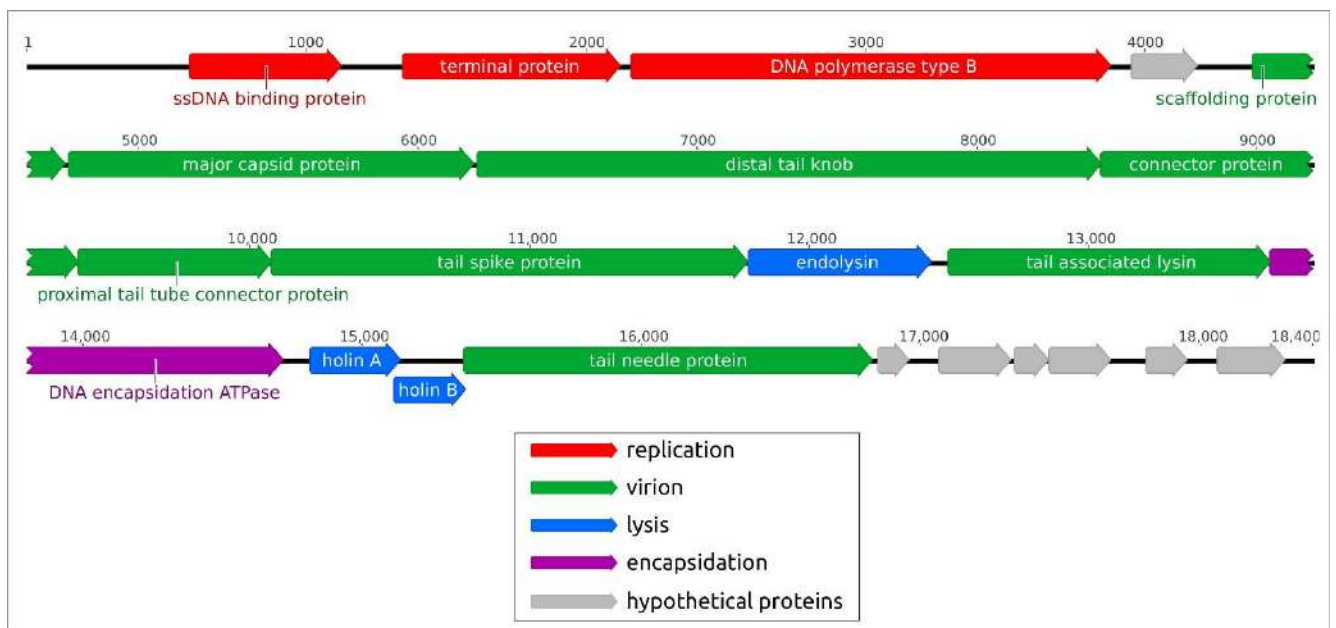


Рисунок 35. Генетическая карта фага *Curtobacterium* Аука. Цвета различных функциональных модулей соответствуют их функциональности: красный – репликация; зелёный – морфогенез; голубой – лизис; пурпурный – упаковка; серый – гипотетические белки. Числа над последовательностью показывают расстояние в н.п. от начала генома.

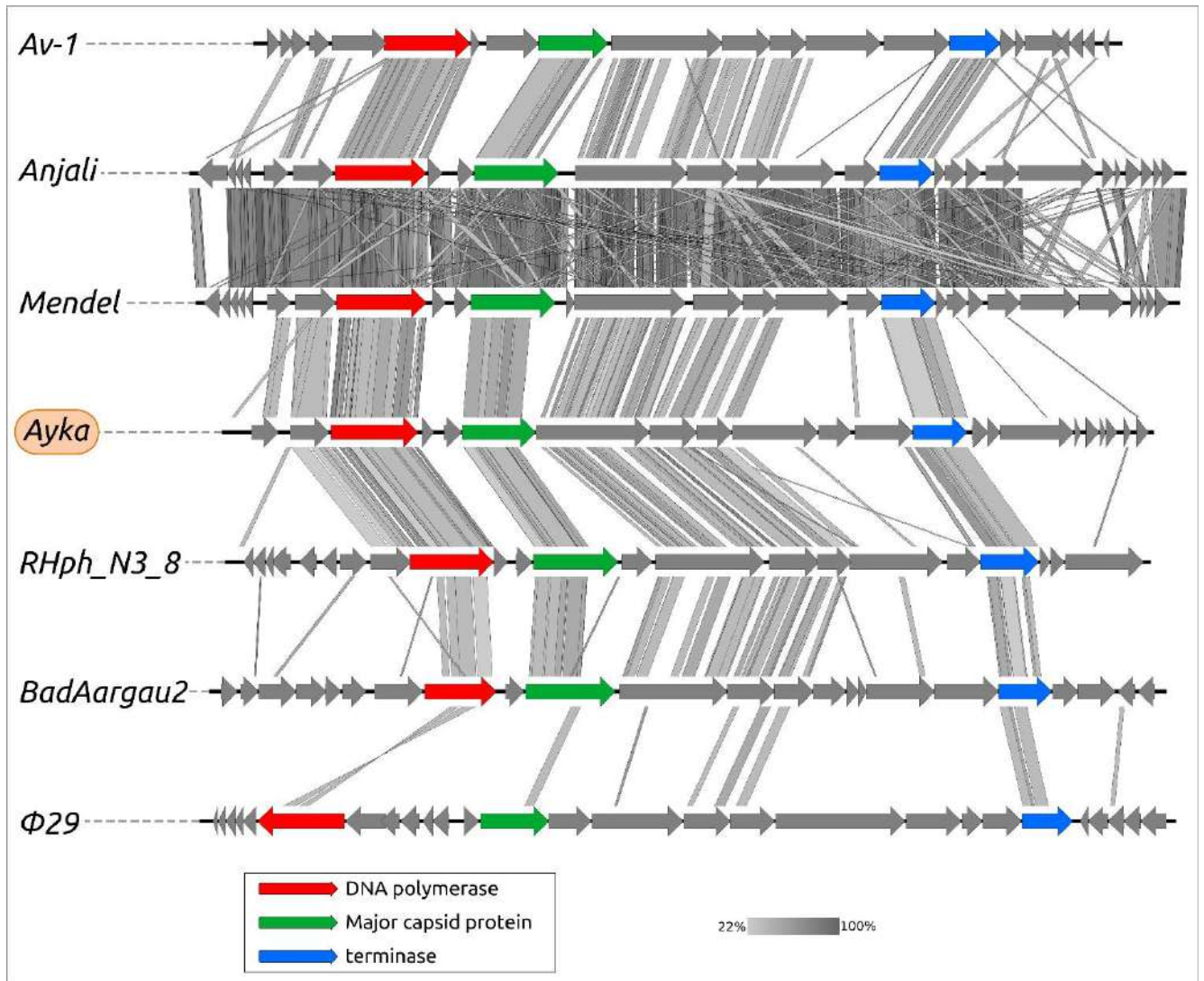


Рисунок 36. Сравнение геномов фага *Curtobacterium* Айка и родственных фагов. Красным цветом обозначены гены ДНКП, зелёным – ГКП, голубым – терминазы. Степень сходства геномных участков обозначена градиентной шкалой внизу рисунка.

Чтобы найти фаги, родственные фагу Айка, был проведен поиск BLAST с использованием последовательностей предсказанных белков и базы данных фагов NCBI GenBank PHG (около 30 тысяч всех депонированных в базе последовательностей на момент поиска). Поиск выявил несколько десятков фагов грамположительных бактерий и фаг *Rhizobium* (RHph_N3_8), который обладал некоторыми локальными гомологиями с фагом Айка (рис. 36). Главный капсидный белок, терминаза и ДНК-полимераза, а также ряд других структурных и лизисных

белков фага Айка показали наибольшее сходство с актинофагами, относящимися к роду *Anjalivirus*. Главный белок капсида имеет каноническую укладку НК97.

Сравнение межгеномных расстояний с использованием геномных последовательностей фага Аука и родственных фагов показало невысокий уровень межгеномного сходства по сравнению с другими фагами, найденными с помощью поиска BLAST (около 5% и менее), что значительно ниже, чем родовой порог 70% (рис. 37). Это не позволяет отнести фаг Аука к какому-либо роду, признанному Международным комитетом по таксономии вирусов (ICTV). Анализ генетической сети, включающей фаг Аука и геномы прокариотических вирусов базы данных пайплайна ConTACT.2.0 (рис. 38) показал наличие родственных связей между фагом Аука, фагами семейства *Salasmaviridae* и другими небольшими фагами, инфицирующими грамположительные бактерии.

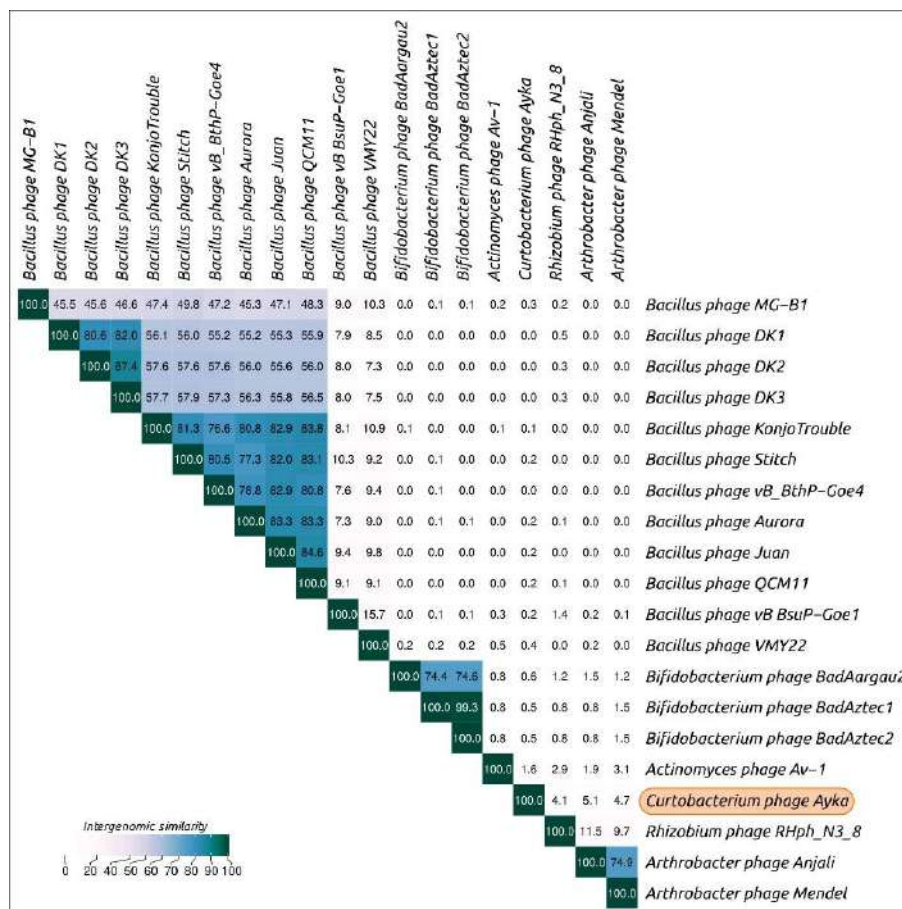


Рисунок 37. Кластерная тепловая карта межгеномного сходства фага Аука и родственных фагов, полученная с помощью VIRIDIC на основании межгеномного сходства.

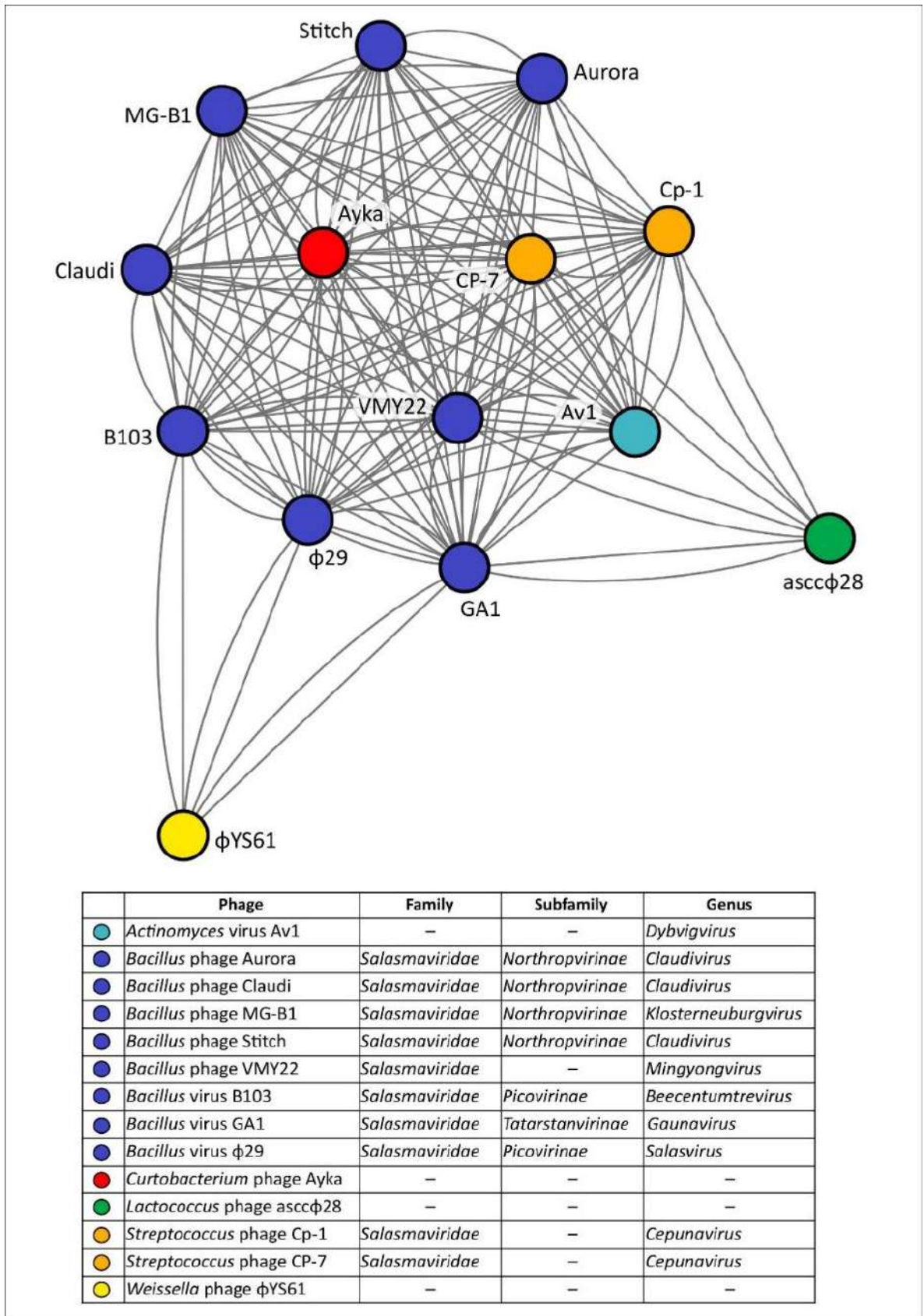


Рисунок 38. Карта генетической сети фага Аука и родственных фагов, полученная с использованием встроенной базы данных ProkaryoticViralRefSeq database v94 пайплайна ConTACT.2.0. В легендах указана таксономическая принадлежность фагов по состоянию на 2022 год.

Филогенетический анализ с использованием аминокислотных последовательностей пяти консервативных белков (главного капсидного белка, терминазы, ДНК-полимеразы, белка-коннектора и белка-соединителя проксимальной части хвостовой трубы) (рис. 39) поместил фаг Аука в кладу с небольшими подовирусами актиномицетов *Actinomyces* Av-1 рода *Dybvigvirus*, фагами *Arthobacter* рода *Anjalivirus* и неклассифицированным фагом *Rhizobium* RHph_N3_8. Эта кладка является сестринской по отношению к фагам *Vifidobacterium*, относящимся к роду *Badaztecivirus*. Генетические расстояния между упомянутыми выше родственными группами фагов и фагом Аука примерно того же порядка, что и генетические расстояния между классифицируемыми подсемействами семейства *Salasmaviridae*, которое также родственно фагу Аука, но более отдаленно.

Протеомная филогения GRAViTy показала схожую картину, поместив фаг Аука и фаги *Salasmaviridae* в сестринские группы, составляющие вместе с другими родственными фагами общую кладку (приложение Б, рис. Б.1). По совокупности результатов таксономического анализа можно сделать вывод о том, что фаг Аука может быть классифицирован как представитель нового семейства или подсемейства.

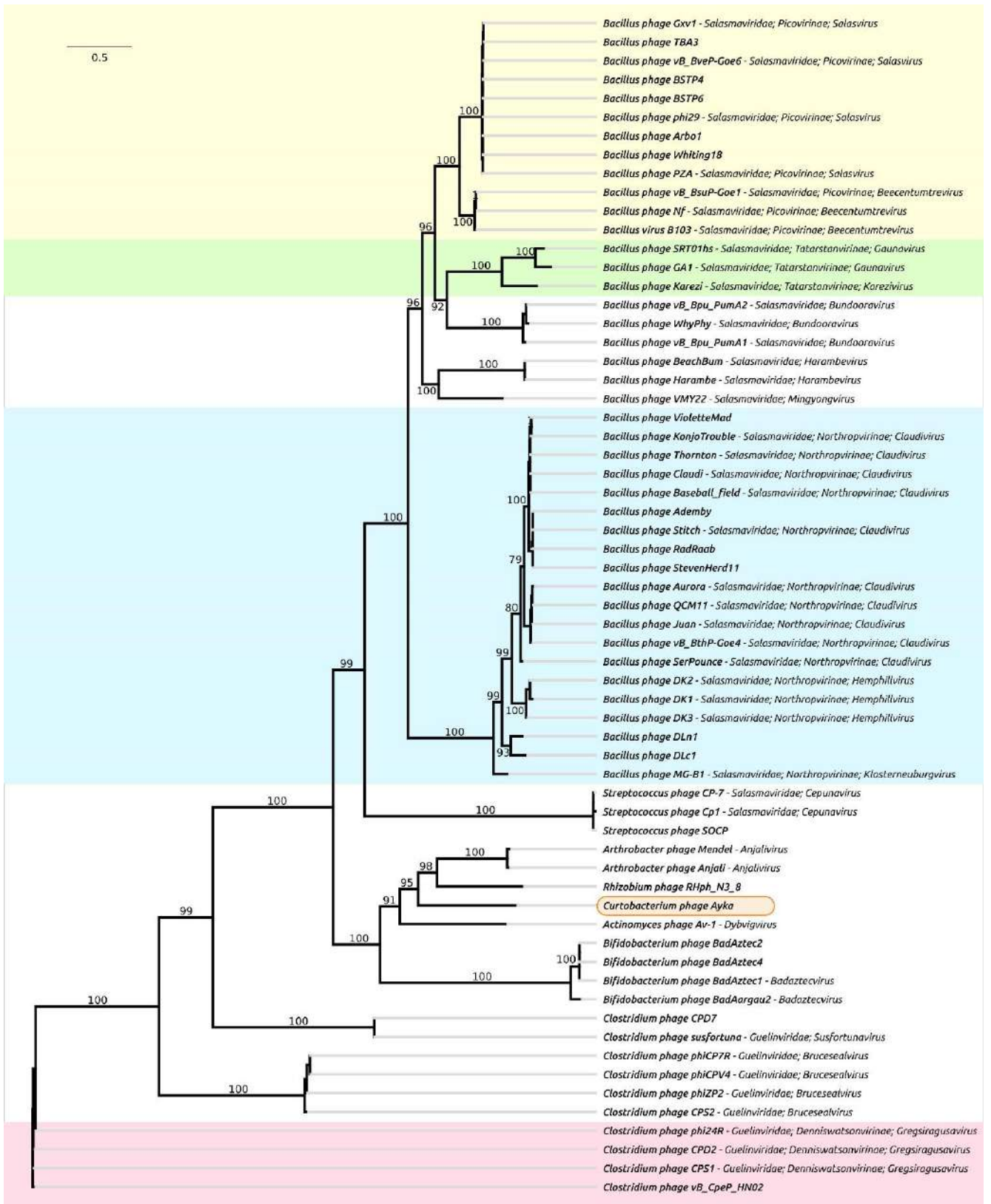


Рисунок 39. Филогенетическое дерево, построенное с использованием конкатенированных аминокислотных последовательностей 5 консервативных белков. Таксономическая принадлежность по состоянию на начало 2022 г. указана справа от названия фага. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

3.2.2. Профаги *Curtobacterium*

Поиск профаговых областей проводили с использованием всех 197 геномов, содержащихся в базе данных NCBI Genome, отнесенных к роду *Curtobacterium*. Области профагового происхождения искали с помощью двух популярных инструментов, пайплайна PHASTER и сервера PhiSpy, результаты предсказаний которых, тем не менее, приводили к несовпадающим результатам. Поиск с помощью PHASTER обнаружил 161 профаговую область, причём только одна из них была определена как интактный профаг, поиск с помощью PhiSpy обнаружил 75 профаговых областей. Большая часть предсказанных PHASTER профаговых областей относилась к 32 геномным локусам, содержащим гены, кодирующие УДФ-N-ацетилглюкозаминтрансферазу, N-ацетилглюкозаминил-дифосфодакапренол-L-рамнозилтрансферазу, белок F биосинтеза тейхоевой кислоты, dTDP-4-дегидрорамнозо-3,5-эпимеразу, гликозилтрансферазу семейства PMT, ундекапренил-фосфат-4-дезокси-4-формамидо-L-арабинозоттрансферазу, 4-амино-4-дезокси-L-арабинозоттрансферазу (или гликозилтрансферазу), dTDP-глюкозо-4,6-дегидратазу (RmlB) и dTDP-4-дегидрорамнозоредуктазу. Эти ферменты участвуют в биосинтезе полисахаридов клеточной стенки и экзополисахаридов у грамположительных бактерий, липополисахаридов у грамотрицательных бактерий [201,202] и, возможно, были определены программой как профаговые области так как могли быть получены путем горизонтального переноса генов мобильными элементами [203], в том числе, фагами. Некоторые предсказанные профаги не содержат структурных генов фага, но, как показал поиск BLAST и сравнение НММ-НММ, обладают генами, типичными для плазмид и других мобильных элементов, таких как транспозазы и гены конъюгативного переноса [204]. Таким образом, результаты использования указанных выше биоинформационных инструментов нуждались в дополнительной постобработке.

Постобработка результатов поиска профаговых областей включала тщательную проверку наличия гомологичных и схожих генов в известных последовательностях выделенных секвенированных фагах, а также присутствия в

предсказанных профаговых областях генов, кодирующих структурные белки фагов. Блок-схема поиска и проверки профаговых областей показана на рис. 40. Ручная проверка предсказанных кодирующих рамок и функциональная аннотация позволила определить границы 64 областей, содержащих гены ГКБ, терминазы и других генов фагового происхождения (см. приложение В, табл. В.1) (рис. 41).

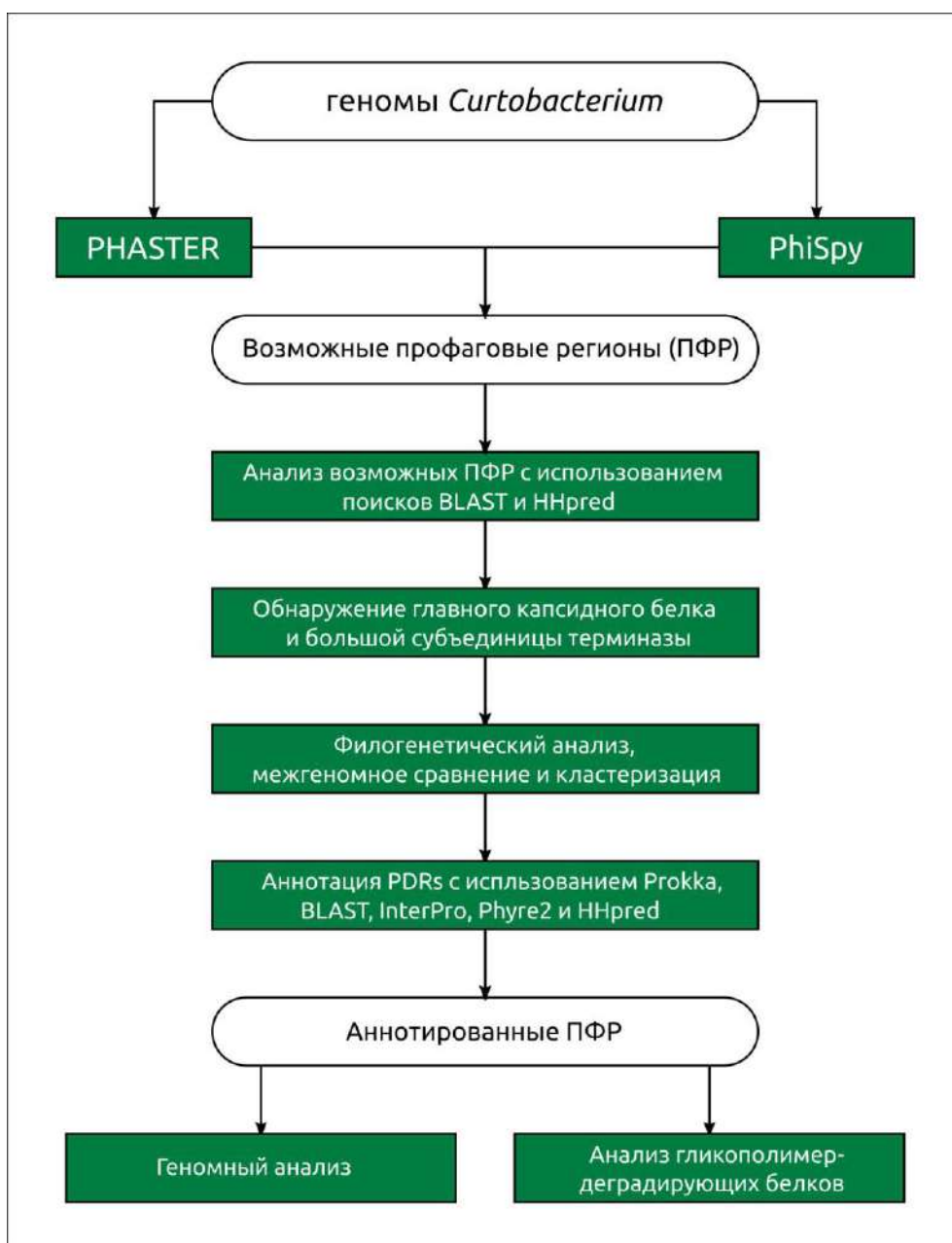


Рисунок 40. Блок-схема идентификации профаговых областей *Curtobacterium*. Показаны инструменты, методы и используемые данные.

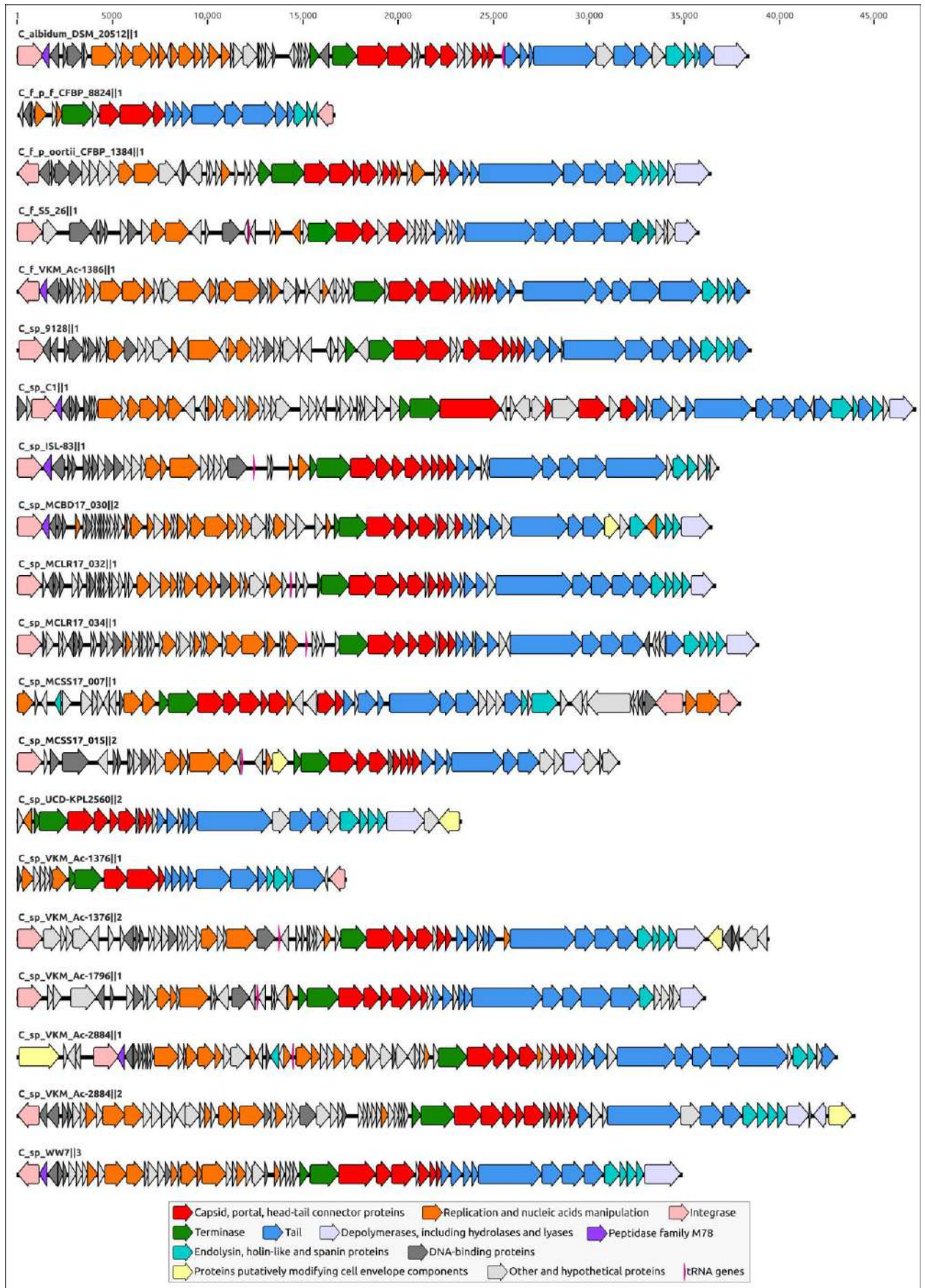


Рисунок 41. Генетические карты 20 профаговых областей в геномах *Curtobacterium*. Предсказанные функции генов указаны в легендах.

Эксперименты по индукции профагов, проведённые с 6 штаммами *Curtobacterium* сотрудниками лаборатории молекулярной биоинженерии А.А. Лукьяновой, Р.И. Таракановым и А.Д. Токмаковой, а также электронная микроскопия, выполненная Е.Е. Куликовым, позволили обнаружить фаговую частицу, индуцированную из *Curtobacterium* sp. VKM Ac-2884, которая с большой вероятностью представляет собой предсказанный профаг C_sp_VKM_Ac-2884|2 [205] (рис. 42). Межгеномные сравнения (рис. 43) показали сложную картину родственных взаимоотношений предсказанных профагов. Большинство из 64 профагов (46 профагов) можно было сгруппировать в несколько кластеров, причём только два из этих 46 профагов показали межгеномное сходство более 70%. Таким образом, большинство из предсказанных профагов в случае, если они представляют функциональные профаги, способные к индукции, могут рассматриваться как представители разных родов. Также можно предположить, что предсказанные профаговые области представляют собой сильно мутировавшие нефункциональные, дефектные профаги.

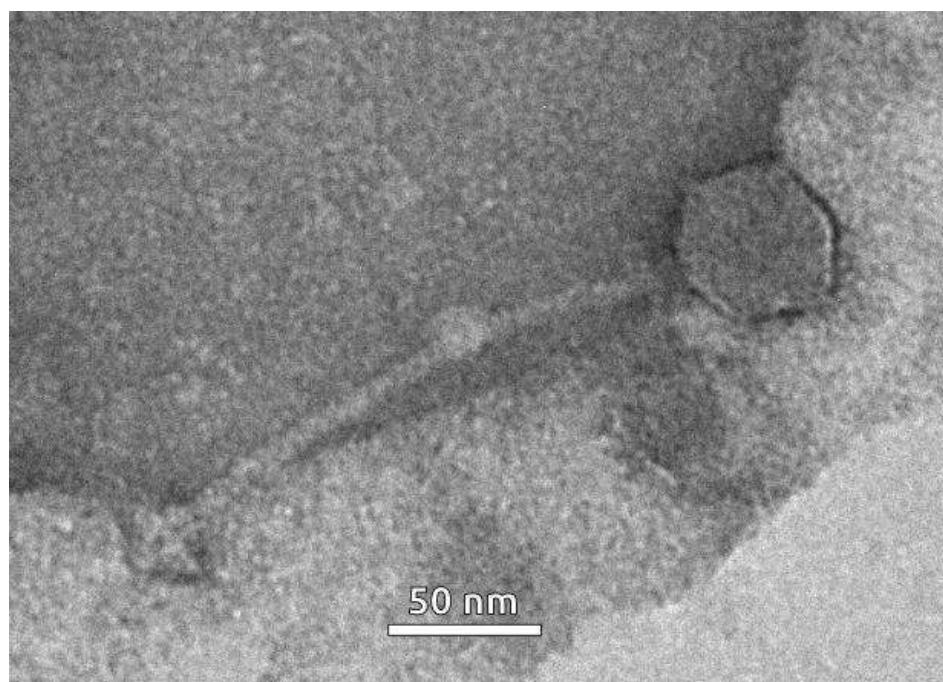


Рисунок 42. Электронная микрофотография профага, индуцированного из *Curtobacterium* sp. VKM Ac-2884. Масштаб шкалы — 50 нм. Электронная микроскопия выполнена Е.Е. Куликовым, микрофотографии опубликованы в статье, написанной совместно с автором диссертации [198].

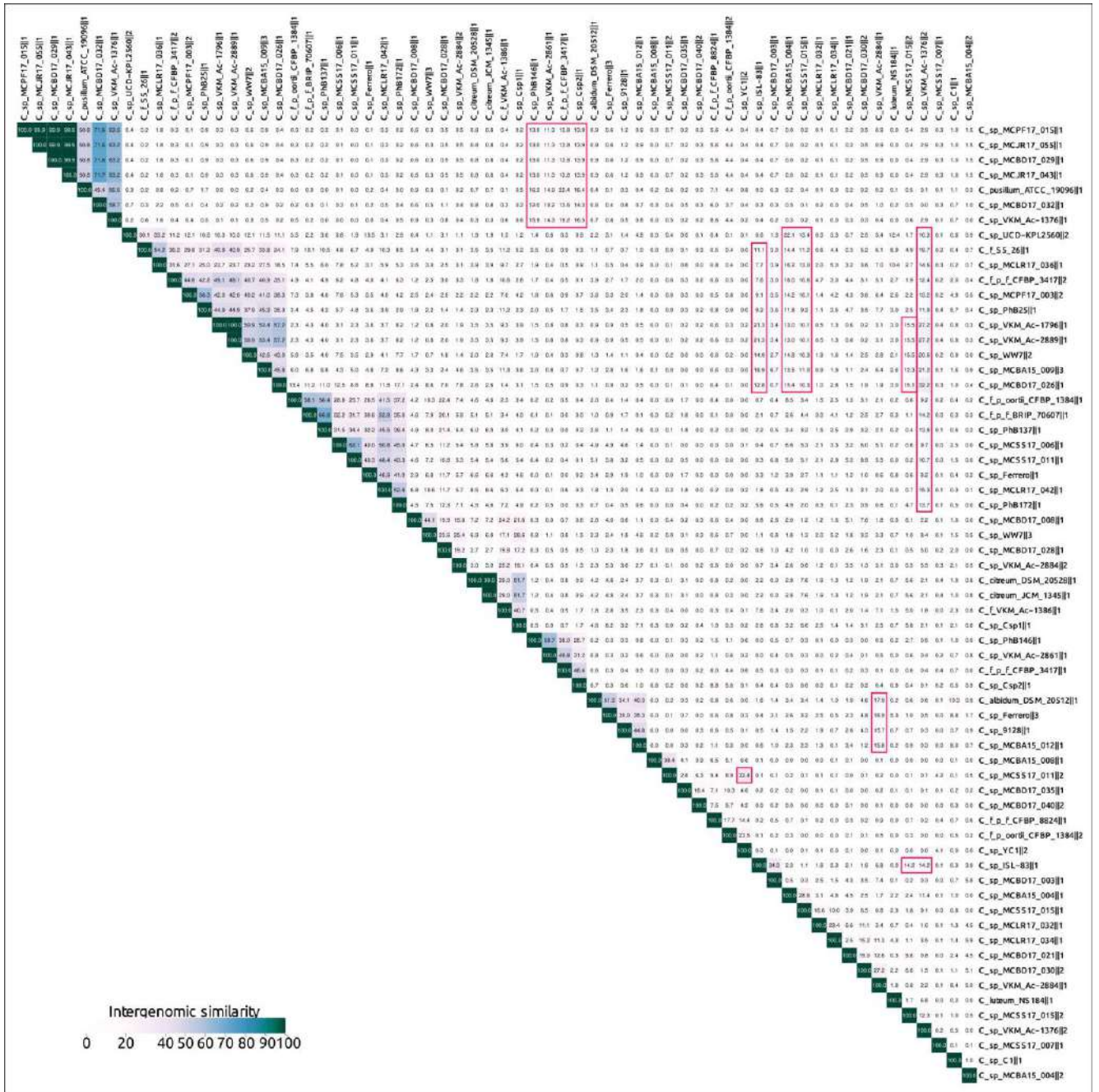
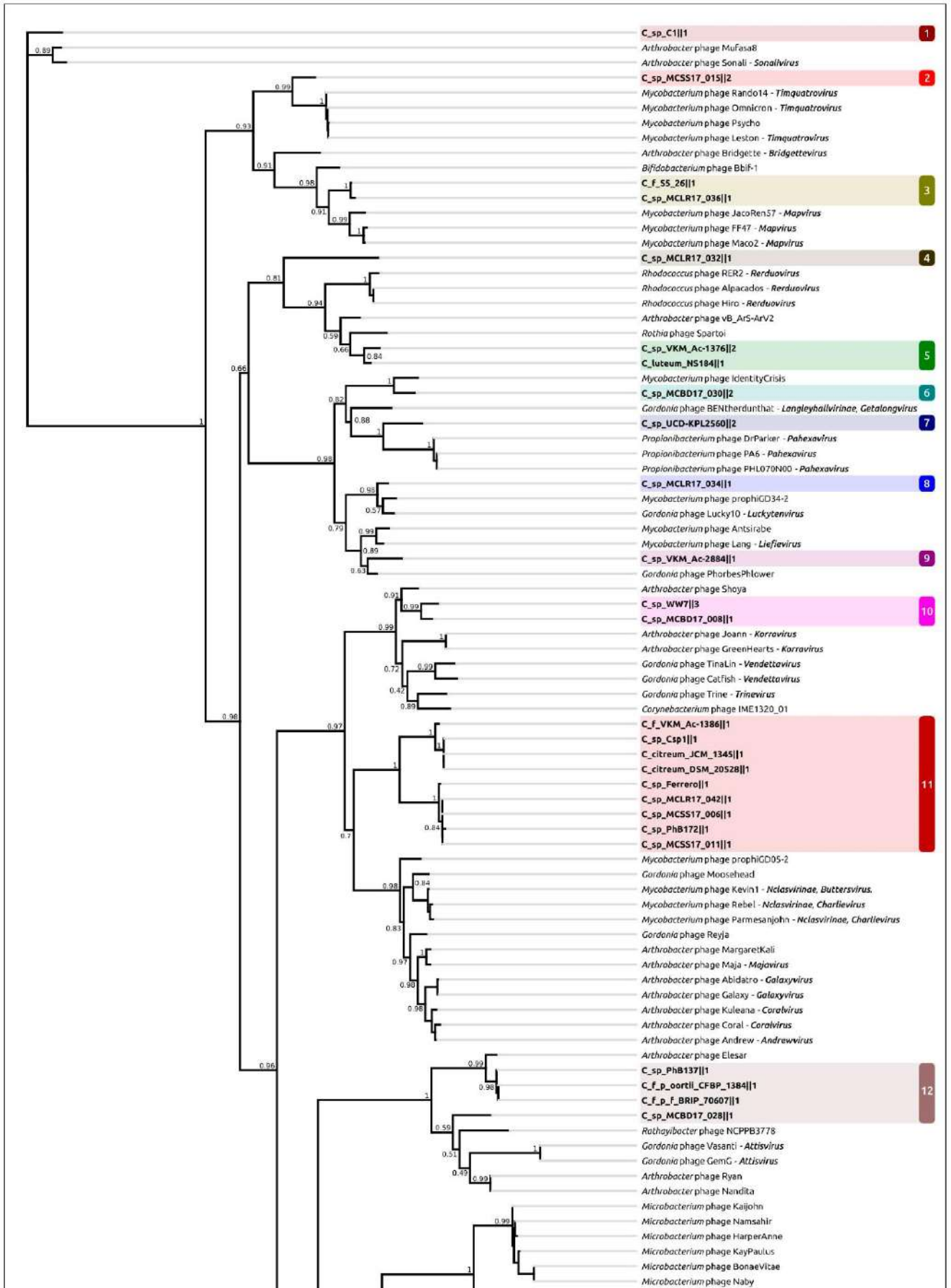


Рисунок 43. Кластерная тепловая карта 84 предсказанных профагов *Curtobacterium*, полученная с помощью VIRIDIC на основании межгеномного сходства.

Интересно, что некоторые области профагов показали сходство с профагами, принадлежащими к разным кластерам одновременно (рис. 43, области выделены красной рамкой), что может быть результатом генетического мозаицизма, характеризующего эволюцию фагов, особенно умеренных, в том числе актинофагов [206,207]. Таким образом, кластеризация предсказанных профагов с

использованием межгеномного сходства не позволяет предложить полностью непротиворечивую классификацию.

Филогенетический анализ, проведенный с использованием аминокислотных последовательностей ГКБ профагов, использовавшихся для кластеризации (рис. 44), и терминазы, закодированных в выбранных профаговых областях показал несовпадающую топологию деревьев, хотя состав клад был похожим (приложение В, рис. В.1, В.2). Эти различия также могут являться результатом мозаичной эволюции фаговых геномов. Для определения близкородственных таксономических групп фагов с помощью инструмента orthoANI был проведён поиск близких по ANI фагов с использованием представителей групп, объединенных по филогенетическим признакам с использованием последовательностей ГКБ, и всех полных геномов фагов *Caudoviricetes*, доступных в базе данных NCBI Genome по состоянию на июль 2022 года. Далее рассчитывали межгеномное сходство для предсказанных профагов и известных фагов с использованием инструмента VIRIDIC. Расчёты межгеномного сходства не выявили значимого сходства между профаговыми регионами *Curtobacterium* и известными полными фаговыми геномами. Однако небольшое межгеномное сходство порядка 10% было выявлено у некоторых фагов, поражающих бактерии *Microbacterium*, эволюционно близкие к роду *Curtobacterium*. В частности, небольшие фаги *Microbacterium* с размером генома менее 20 тыс. н.п., недавно отнесенные к новым таксонам (семейство *Orlajensenviridae*, подсемейство *Pelczarvirinae*, род *Paopuvirus*), показали некоторое сходство с профагами *Curtobacterium* с геномами близкого размера.



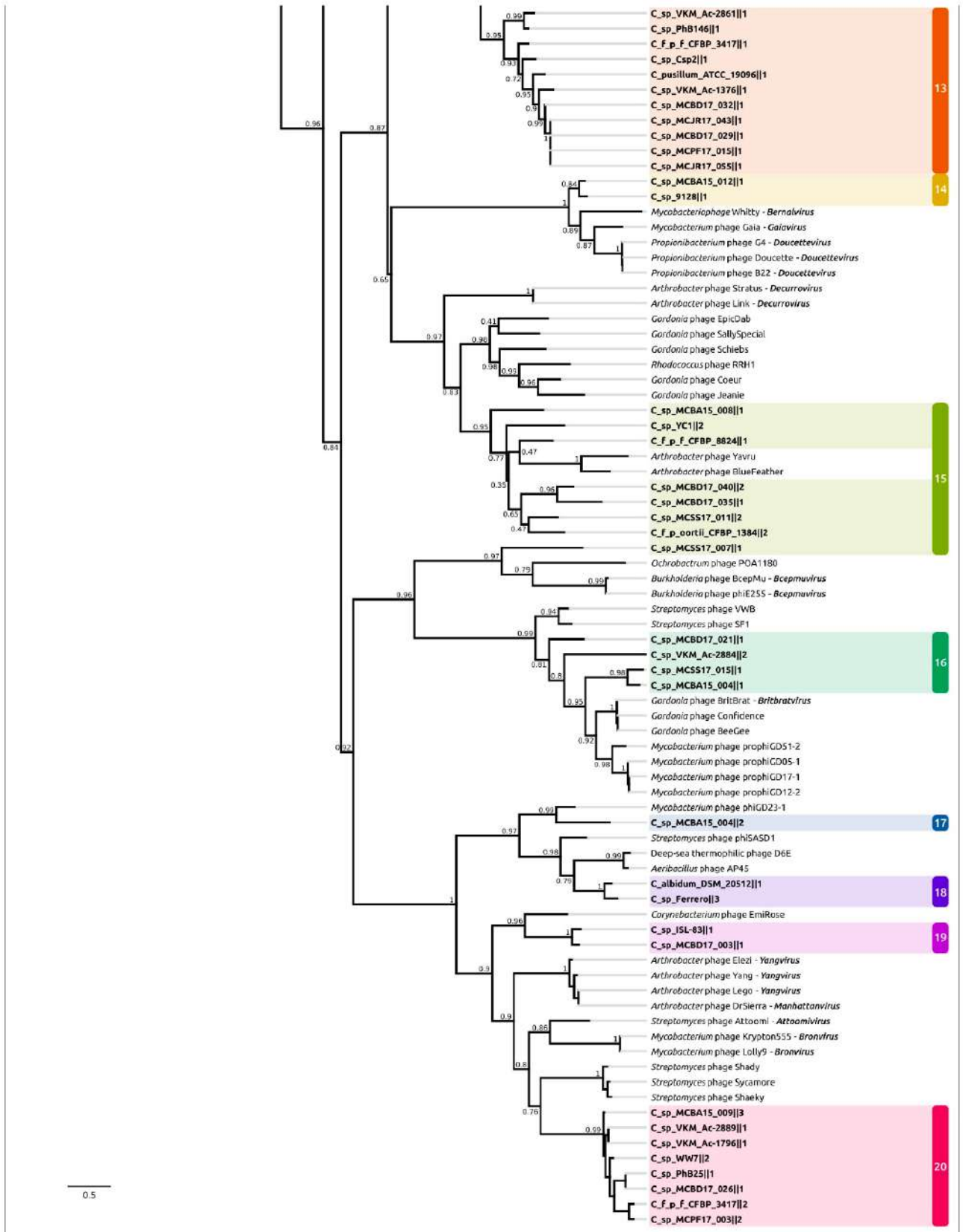


Рисунок 44. Филогенетическое дерево, полученное на основе аминокислотных последовательностей ГКБ и кластеры профагов (указаны справа). Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

Филогенетический анализ с использованием аминокислотных последовательностей консервативных белков может выявить более отдаленные эволюционные и таксономические отношения. Список родственных таксономических групп фагов, выявленных с использованием филогений ГКБ (рис. 44) и терминазы (приложение В, рис. В.3), в основном включает как неклассифицированные, так и классифицированные умеренные актинофаги сифовирусной морфологии, отнесенные к подсемейству *Nclasvirinae* и нескольким десяткам родов, не выделенных в подсемейство или семейство, в том числе *Bridgettevirus*, *Britbratvirus*, *Bronvirus*, *Coralvirus*, *Decurrovirus*, *Fromanvirus*, *Mapvirus*, *Timquatrovirus* и др. Родственные фаги инфицируют бактерии, принадлежащие к родам *Arthrobacter*, *Attisvirus*, *Bifidobacterium*, *Corynebacterium*, *Gordonia*, *Microbacterium*, *Mycobacterium*, *Streptomyces*, *Propionibacterium*, *Rathayibacter* и *Rhodococcus*. Поиск BLAST с помощью последовательностей спейсеров CRISPR, найденных в геномах *Curtobacterium*, обнаружил сходные участки, также, в основном, в геномах актинофагов сифовирусной морфологии, в том числе фагов, принадлежащих к подсемействам *Arquatrovirinae*, *Bclasvirinae*, *Guernseyvirinae*, *Mclasvirinae*, *Nclasvirinae*, *Nymbaxtervirinae*, *Weiservirinae*, неклассифицированных фагов и фагов, принадлежащих к разным родам, не отнесенных к подсемействам или семействам. Результаты исследования свидетельствуют о возможности существования обширных групп неклассифицированных функциональных актинофагов.

В ходе работы также было выявлено около 100 генов, кодирующих эндолизины и другие гликополимер-деградирующие белки (ГДБ), которые могут рассматриваться в качестве кандидатов для использования в качестве антибактериальных агентов. Фаговые эндолизины были кластеризованы с помощью филогенетического анализа (приложение В, рис. В.4), а их функции были предсказаны с помощью сравнения мотивов НММ, предсказания и сравнения структур (приложение В, рис. В.5). Интересно, что кластеризация не эндолизиновых ГДБ, корректно отражающая функции этих белков, не могла быть

проведена с использованием филогении, задействующей аминокислотные последовательности, из-за большого различия аминокислотных последовательностей белков разного происхождения. Тем не менее, анализ структурного сходства моделей, предсказанных AlphaFold 2 (AF2), с использованием mTM-align, позволил сгруппировать эти ГДБ в соответствии с их возможной функцией (рис. 45).

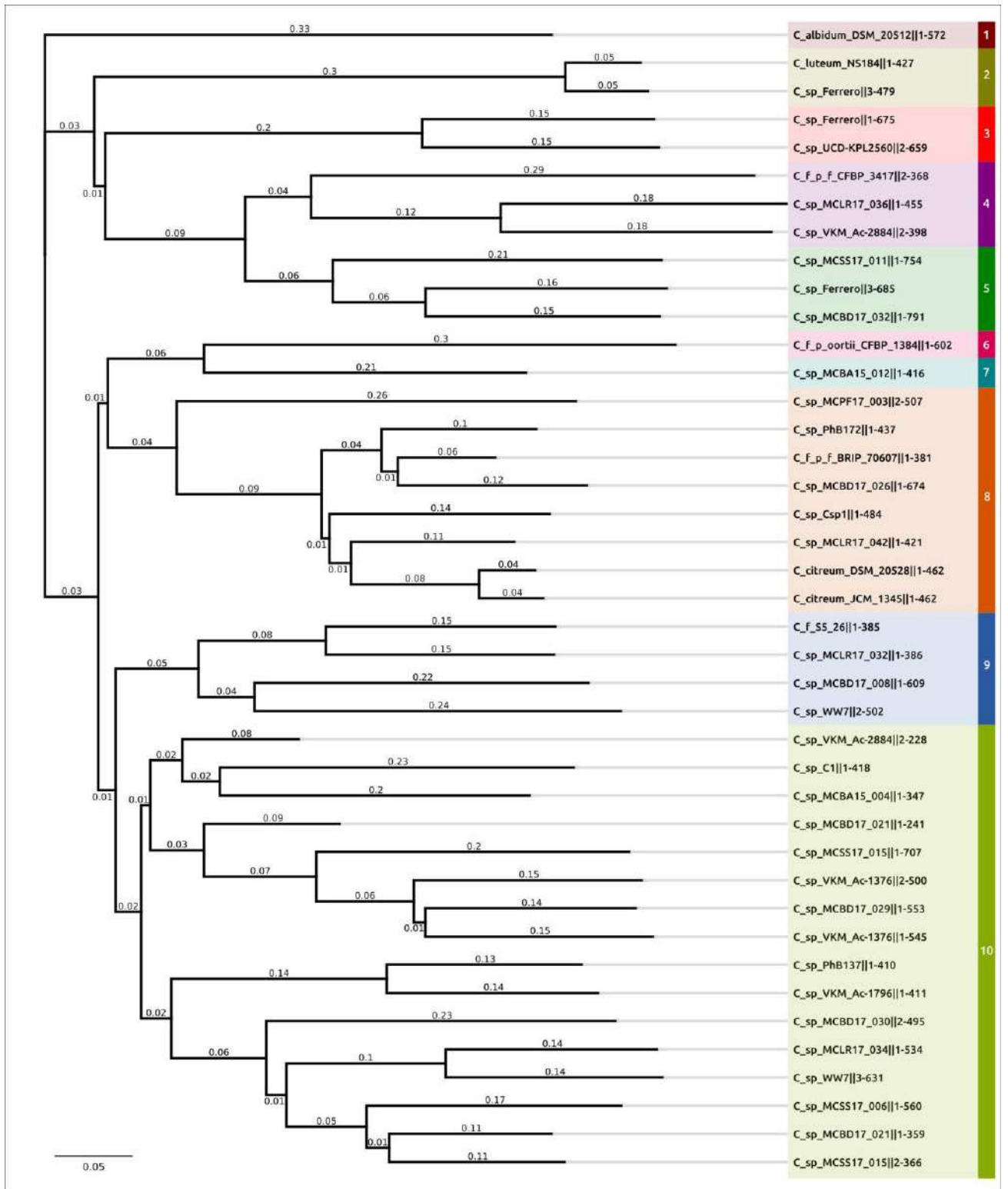


Рисунок 45. Кластеризация гликополимер-деградирующих белков профагов *Curtobacterium* с использованием структурного сходства предсказанных моделей, определённого mTM-align. Дистанции, определённые с помощью TM-score, показаны около соответствующих веток.

3.3. Таксономический анализ бактериофага

Pseudomonas MD8

3.3.1. Геномный и филогенетический анализ

Фаг *Pseudomonas* MD8 — умеренный фаг, выделенный из пресных вод озера Байкал (рис. 46). Это дцДНК-вирус сифовирусной морфологии с геномом размером 43277 н.п. ГЦ-состав генома составляет 61,1%, что несколько ниже ГЦ-состава у *P. aeruginosa* (около 66%). Геном MD8 содержит 67 предсказанных ORF и не содержит генов тРНК.

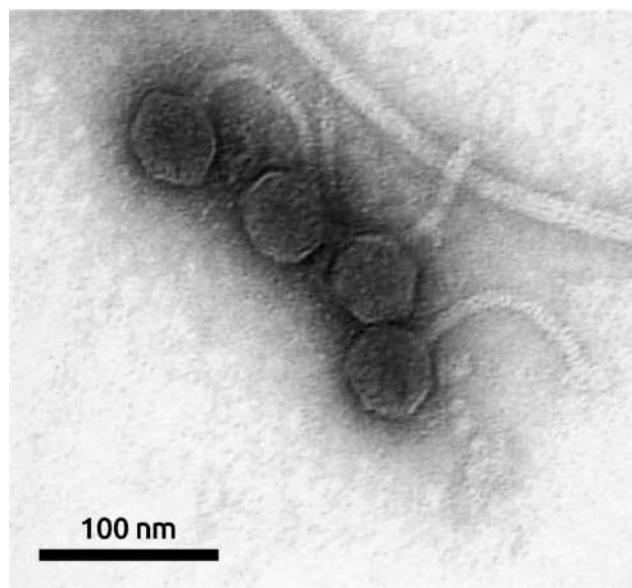


Рисунок 46. Электронная микрофотография фага *Pseudomonas* MD8. Масштаб шкалы – 100 нм [208]. Электронная микроскопия выполнена В.А. Кадыковым.

Структура генома (рис. 47) напоминает фаг *Escherichia* λ и другие лямбдоидные бактериофаги. Начиная с 5'-конца, геном содержит блок из 30 генов упаковки и вириона, ориентированных в прямом направлении, блок из 23 генов, включая интегразу, ориентированных в обратном направлении, и блок из 14 генов, в том числе генов репликации и лизиса, ориентированных снова в прямом направлении. Интересно, что главный капсидный белок, по всей видимости, кодируется вместе с протеазой и белком скаффолда одним геном g5.

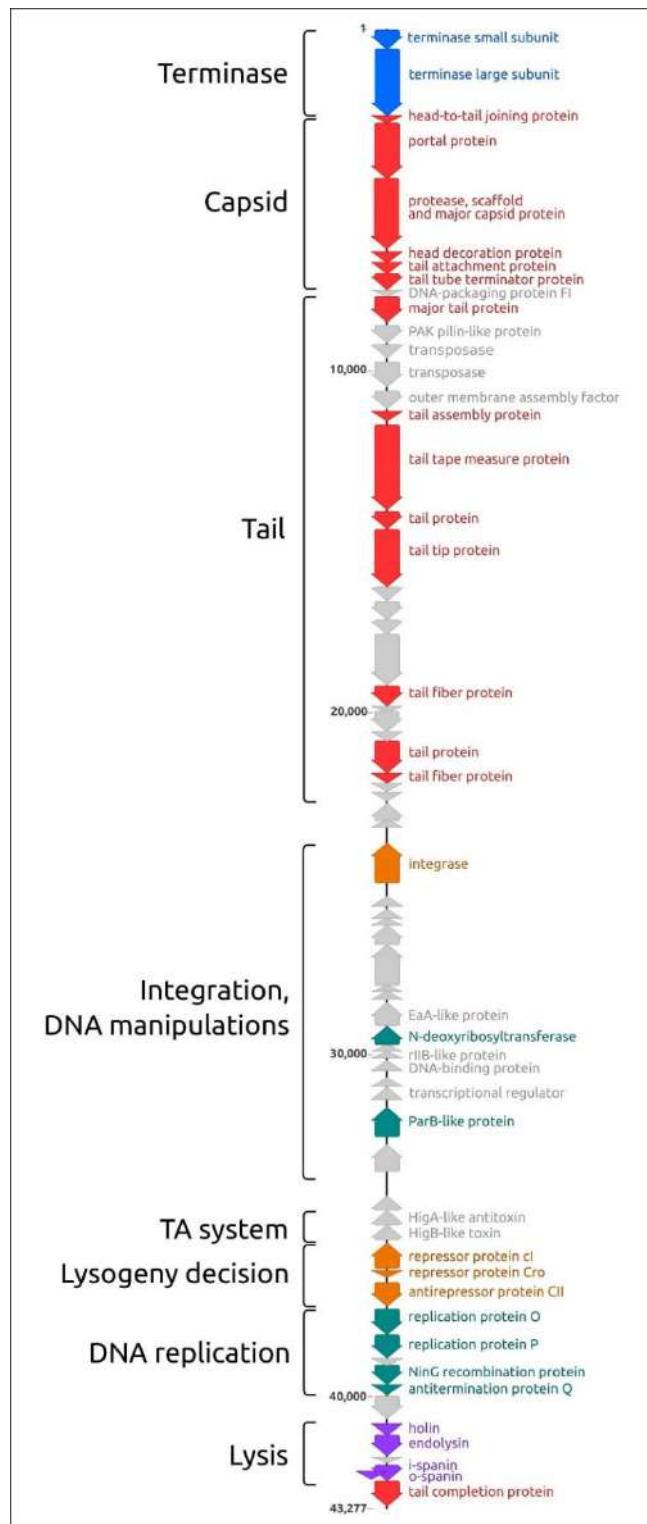


Рисунок 47. Генетическая карта фага *Pseudomonas* MD8. Terminase – блок генов терминазы, Capsid – блок генов капсида, Tail – блок генов хвоста, Integration, DNA manipulation – блок генов, отвечающих за интеграцию и гены белков, взаимодействующих с ДНК, TA system – гены системы “токсин–антитоксин”, Lysogeny decision – блок генов, регулирующих переход в литическую фазу, DNA replication – блок генов репликации, Lysis – блок генов лизиса. Числа слева от последовательности показывают расстояние в н.п. от начала генома.

Поиск BLAST с использованием последовательностей белков фага MD8 обнаружил гомологичные последовательности всех 67 белков в различных бактериальных геномах, Гомологи 13 белков были найдены исключительно в геномах *Pseudomonas*, гомологи 54 белков были обнаружены также и в других бактериальных геномах, что может указывать на долгую историю эволюции фага *Pseudomonas* MD8 и его связь с умеренным образом жизни.

Для нахождения ближайших родственников фагов была использована база данных BLAST, состоящая из всех 14 983 геномов фагов, содержащихся в GenBank по состоянию на май 2021 года. Поиск, проведённый с использованием последовательностей всех белков фага MD8, обнаружил 21 сифовирус *Pseudomonas*, которые часто группируются вместе с MD8, но отдалены от остальных фагов. Один из этих фагов, PAN70, был выделен экспериментально и показал биологическую активность в экспериментах по фаговой терапии на животных моделях [209], а большинство было обнаружено метагеномным анализом при исследовании хронических инфекций лёгких, вызванных *Pseudomonas aeruginosa* [210]. Несколько фагов, включая фаги *Pseudomonas* φ2 и TC7, показали высокий уровень сходства с большей частью генома MD8, за исключением группы генных продуктов 1–9 и некоторых других белков. Гомолог только одного белка MD8, gp32, не был обнаружен в геномах других фагов. Результаты поиска BLAST и вычисления ANI с использованием всех известных фаговых геномов позволили определить группу из 26 фагов (в том числе два фага с неполным геномом), включая MD8, которые обладают сходством, существенно превышающим сходство с другими фагами (ANI около 89–96% при длине выравниваемого участка более 10% от длины генома). Далее мы будем называть эту группу группой MD8-подобных фагов, список этих фагов приведён в приложении Г (табл. Г.1). Анализ межгеномного сходства с помощью VIRIDIC, проведённый с различными наборами фагов, включая все сифовирусы *Pseudomonas*, также выявил группу из 25 фагов *Pseudomonas*, связанных с фагом MD8, подтвердив приведённые выше результаты, и показал, что часть фагов этой

группы с межгеномным сходством более 95% можно рассматривать как расы одного вида (рис. 48).

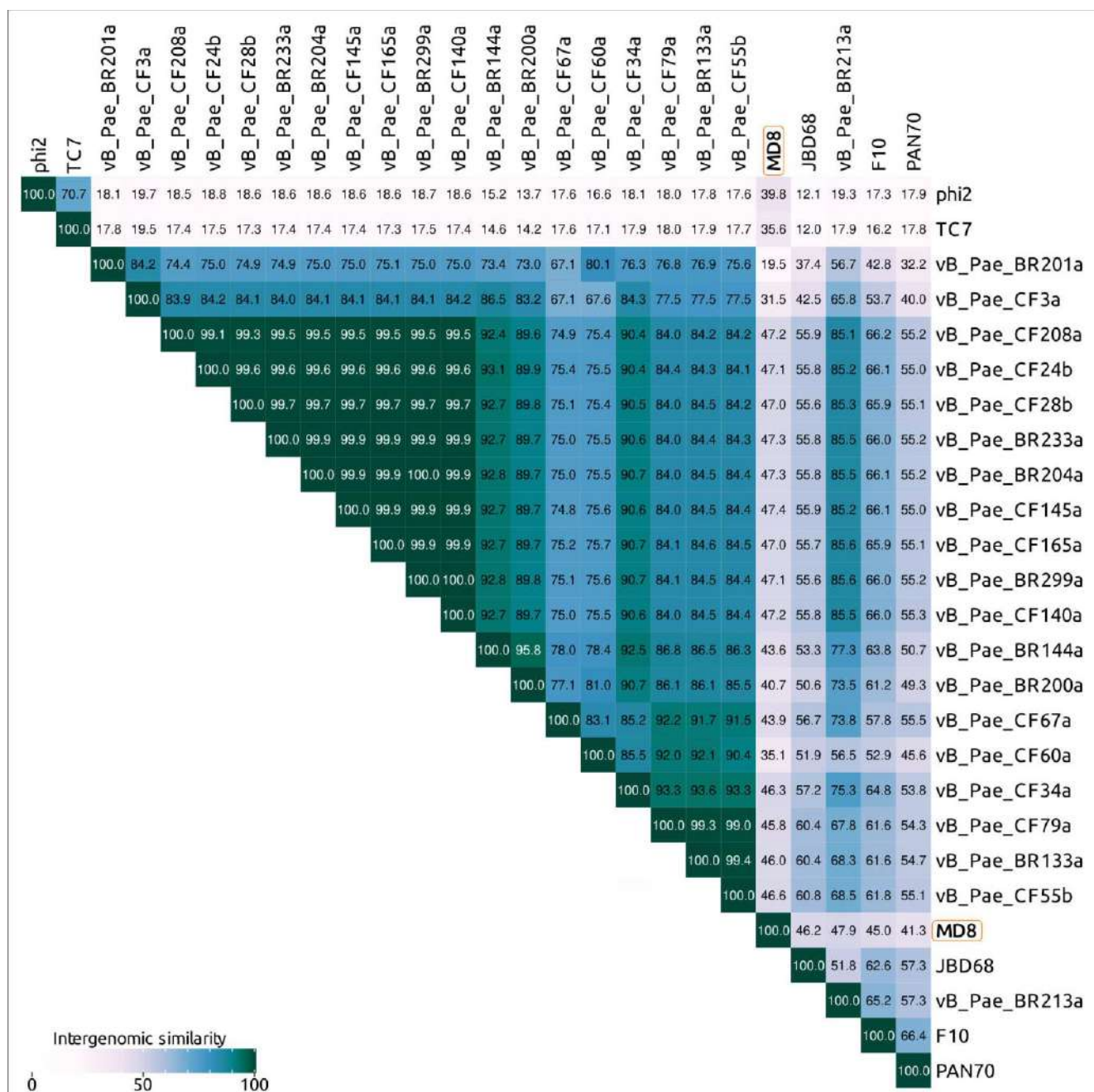


Рисунок 48. Кластерная тепловая карта межгеномного сходства MD8-подобных фагов, полученная с помощью VIRIDIC.

Хотя межгеномные сравнения не выявили сколь-либо значительного сходства MD8-подобных фагов и фага λ , сравнение мотивов НММ белков фага MD8 показало сходство этих белков с белками фага λ . В том числе, геном фага MD8, видимо, кодирует репликационные белки O и P, участвующие у фага λ в

образовании и расхождении репликационных вилок [211]. Как и у фага λ , задействующего механизмы репликации ДНК хозяина [212], в геноме MD8 отсутствуют гены ДНК-полимеразы и геликазы-праймазы (хотя эти гены присутствуют в некоторых других MD8-подобных лямбдоидных фагах [213]). Как и последовательность гена репликационного белка O фага λ , ген O-подобного белка MD8 содержит повторы, служащие в случае фага λ ориджином репликации $ori\lambda$, с которым связываются четыре димера белка O. По-видимому, механизм репликации фага MD8, в основном, аналогичен механизму фага λ . Геном MD8 также кодирует гомолог белка NinG фага λ (gp59 у MD8). NinG фага λ участвует в путях рекомбинации RecBCD *E. coli* [214]. Интересно, что последовательности белков репликации фага MD8 демонстрируют высокий уровень сходства с белками репликации MD8-подобных фагов $\phi 2$ и TC7, но не показывают заметной гомологии с белками всех остальных представителей группы MD8. Кроме того, геномы 19 MD8-подобных фагов, не показывающих сходства белков репликации с белками фага MD8, содержат гены ДНК праймазы/геликазы. Это может указывать на события горизонтального переноса, происходившие в эволюционной истории MD8-подобных фагов, которые привели к замене у части фагов всего модуля генов репликационных белков.

Как показал поиск BLASTP, геномы MD8-подобных фагов содержат гены гомологов белка фага λ CI (gp53 у MD8), который может участвовать в поддержке лизогенного состояния фагового генома [215]. Геномы MD8 и части фагов группы также содержат гомологичные гены конкурента CI белка Cro (gp54 у MD8) [216]. Гомологи другого белка, связанного с поддержкой лизогенного состояния, CII [217], тоже были обнаружены у всех фагов группы MD8 (gp55 у MD8). Сравнение последовательностей белков CI, CII и Cro, а также филогенетический анализ показали, что белки фагов MD8 и $\phi 2$ кластеризуются отдельно от гомологов, присутствующих у всех остальных MD8-подобных фагов, где были обнаружены гомологичные последовательности.

Филогенетический анализ с использованием аминокислотных последовательностей большой субъединицы терминазы указал на близкое родство и общее происхождение терминаз 22 фагов группы MD8 и λ -подобных фагов *Escherichia*, включая фаги родов *Lambdavirus* и *Ravinivirus* (рис. 49). Примечательно, что терминазы фагов ϕ 2 и TC7 расположены на филогенетическом дереве ближе к терминазе фага *Thermus phiFA1*, чем к терминазе фагов *Lambdavirus* и MD8. Также интересно, что P2-подобная терминаза фага *Vibrio VD1*, обладающего миовирусной морфологией, не сильно отличается от терминазы MD8 и *Lambdavirus*. В этом дереве миовирусы и сифовирусы не являются монофилетическими группами, что может свидетельствовать о горизонтальных переносах генов терминаз, сопровождающих сложное формирование геномов бактериофагов с различной морфологией. Филогенетический анализ с использованием аминокислотных последовательностей пропептида, содержащего главный капсидный белок, белок скаффолда и протеазу, проводили аналогично филогенетическому анализу большой субъединицы терминазы, но отдельно для доменов протеазы и ГКБ (рис. 50, 51). Как и в случае с терминазой, филогения протеазы и ГКБ показывает различную эволюционную историю для ϕ 2 и TC7, с одной стороны, и MD8 и других членов группы MD8, с другой. Также примечательно, что филогенетический анализ доменов протеаз и ГКБ не показал тесных эволюционных связей между белками MD8-подобных фагов (включая TC7 и ϕ 2) и фага λ . Однако оба дерева указывают на схожую относительно недавнюю эволюционную историю для обоих доменов во всех проанализированных MD8-подобных фагах. Эти данные могут свидетельствовать о том, что рекомбинационный обмен с участием генов терминазы не обязательно требует участия соседних генов протеаз и ГКБ, в то время как гены (или части генов) ГКБ и протеазы могут чаще передаваться вместе.

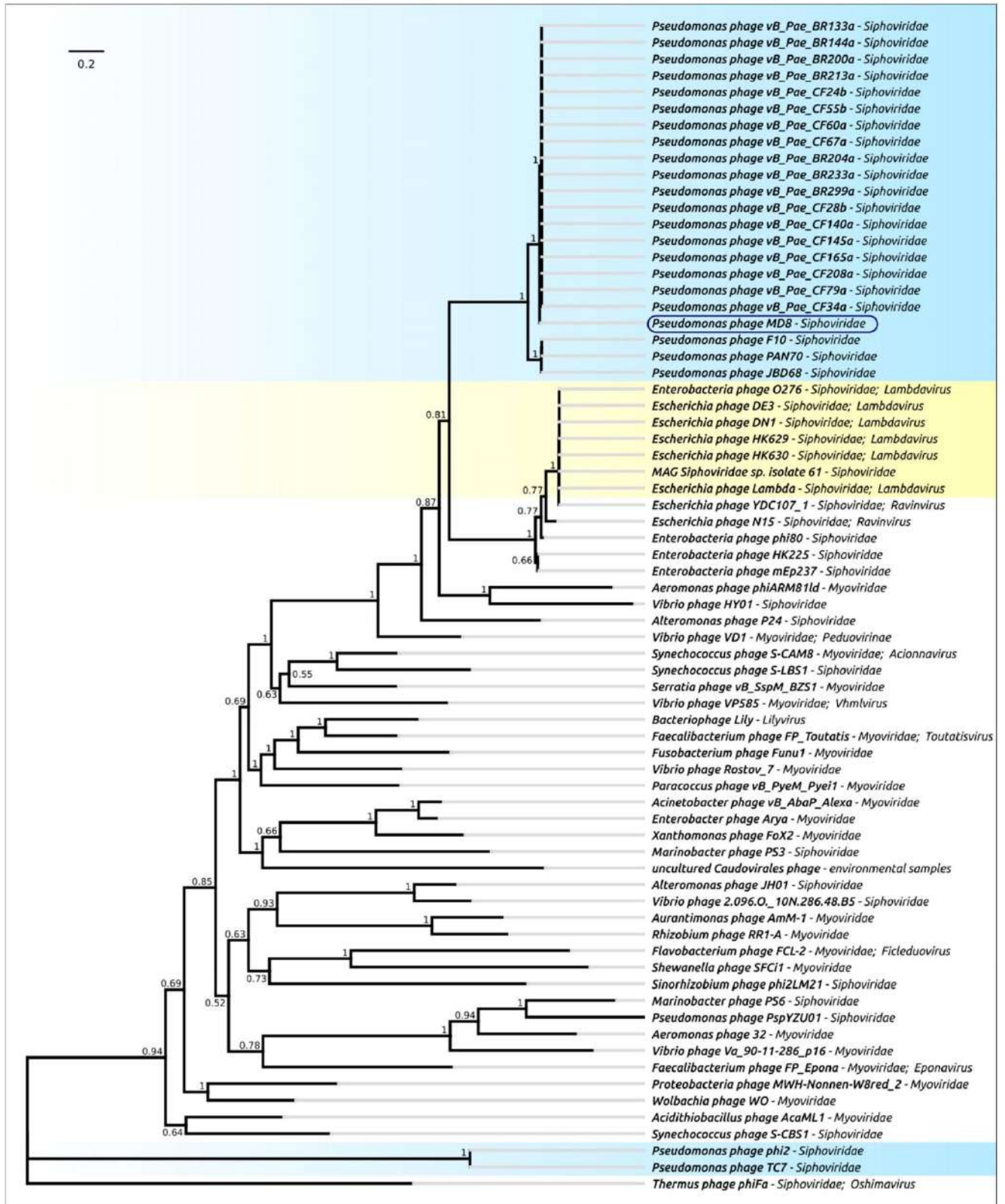


Рисунок 49. Филогенетическое дерево, полученное на основе аминокислотных последовательностей большой субъединицы терминазы MD8-подобных и других фагов. Значения апостериорной вероятности указаны около соответствующих веток.

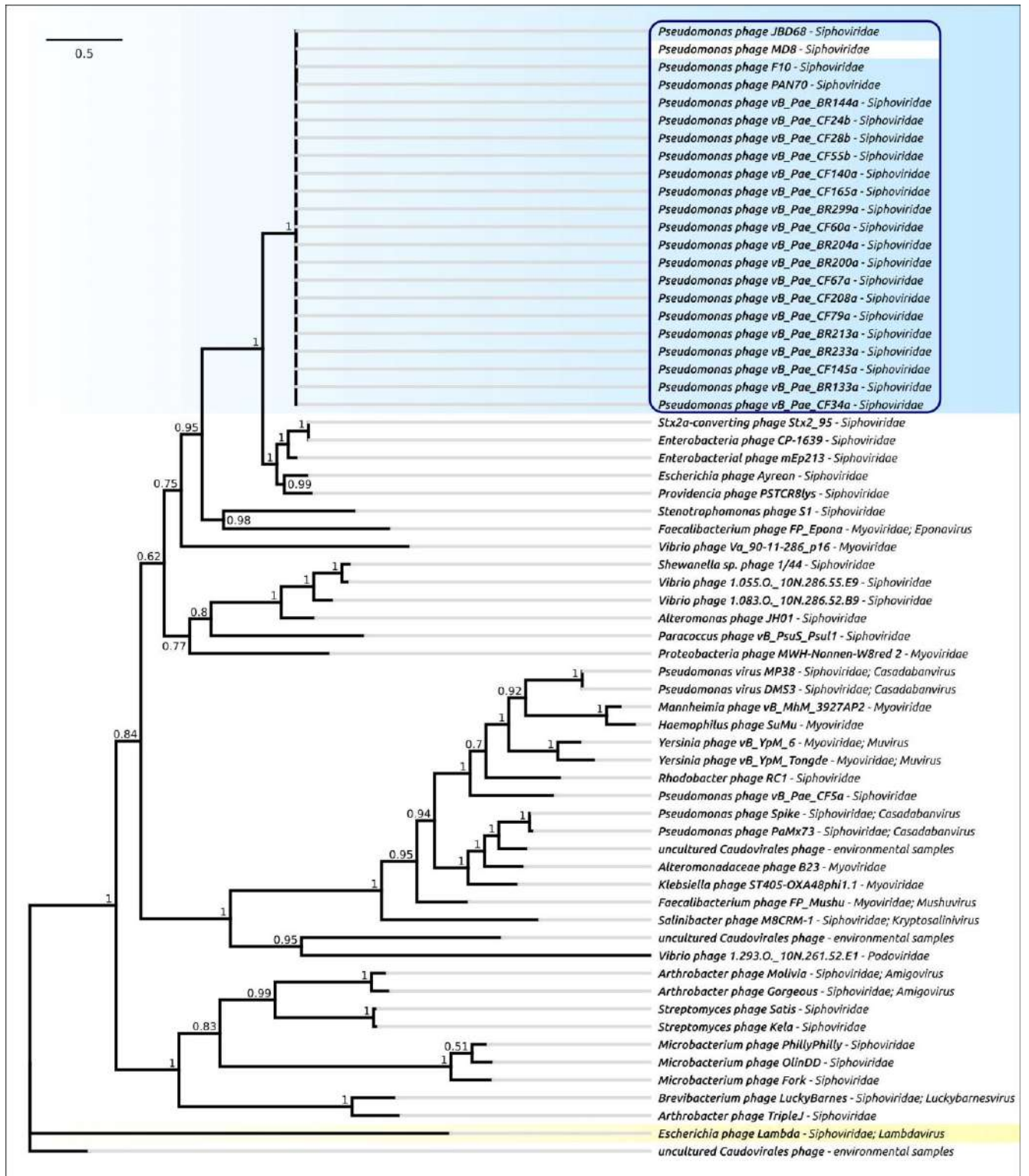


Рисунок 50. Филогенетическое дерево, полученное на основе аминокислотных последовательностей главного капсидного белка MD8-подобных и других фагов. Значения апостериорной вероятности указаны около соответствующих веток.

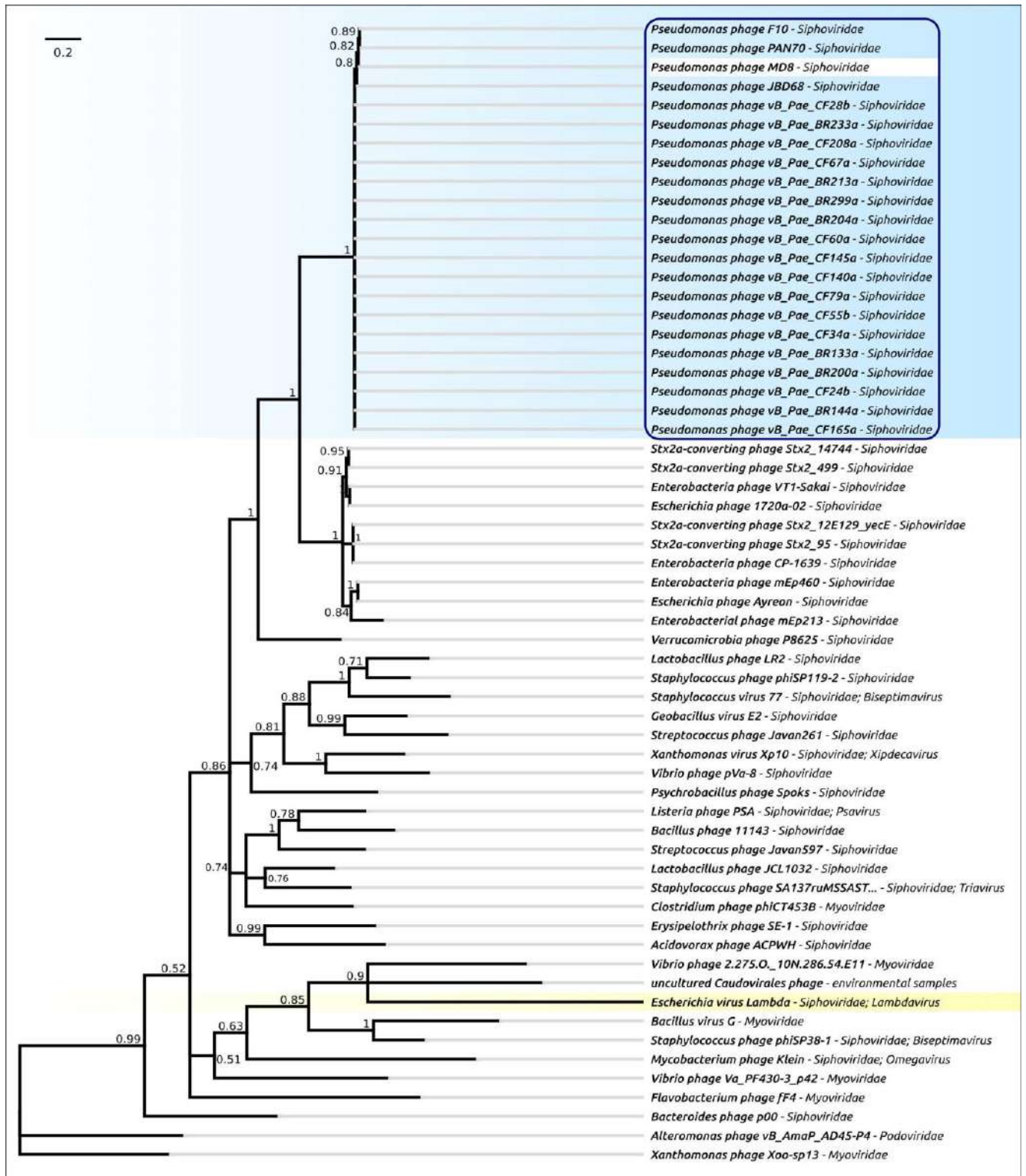
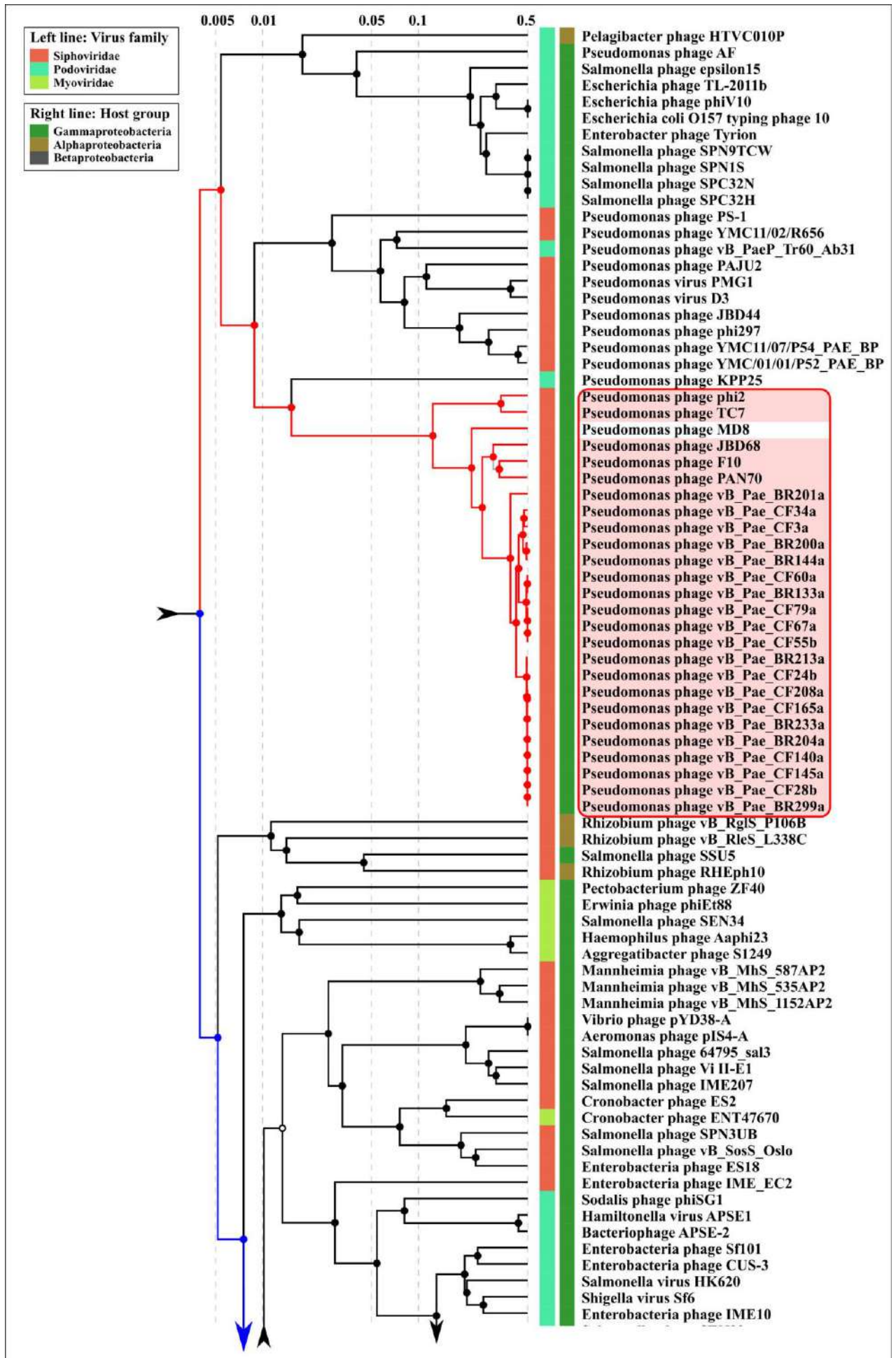


Рисунок 51. Филогенетическое дерево, полученное на основе аминокислотных последовательностей протеазы MD8-подобных и других фагов. Значения апостериорной вероятности указаны около соответствующих веток.

Анализ гена, кодирующего, в том числе, главный капсидный белок, показал его мультидоменную структуру. В вирусах *Duplodnaviria Heunggongvirae*,

включающих хвостатые бактериофаги и вирусы герпеса, капсид строится путем сборки главного капсидного белка с белком скаффолда с образованием промежуточных прокапсидов. После завершения строительства белки скаффолда покидают капсид [218,219]. Многие фаги, в том числе модельные фаги *Escherichia* НК97 [220], λ [221], фаги Mu [222], P2 [223] и T4 [224], используют протеазу созревания капсида, которая расщепляет части прокапсида, помогая белку скаффолда покинуть капсид. В некоторых случаях (например, фаги T7 [225], P22 [226] и ϕ 29 [227]) фаги не нуждаются в помощи протеазы для высвобождения белков скаффолда. Некоторые фаги, в том числе НК97, не имеют отдельных белков скаффолда – у фага НК97 102 N-концевых аминокислотных остатка образуют так называемый δ -домен, который выполняет роль белка скаффолда и далее отщепляется протеазой [220], причём ген g5, кодирующий ГКБ фага НК97, расположен сразу после гена g4, кодирующего протеазу созревания капсида. У некоторых умеренных фагов, в том числе P2 [223] и λ [228], протеаза и белок скаффолда кодируются одним геном. По-видимому, у MD8-подобных фагов все три белка – протеаза созревания капсида, белок скаффолда и ГКБ – кодируются одним геном. Информацию о других подобных случаях на момент публикации результатов в литературе найти не удалось.

Протеомное дерево, построенное с помощью ViPtree, показано на рис. 52, соответствующая круговая диаграмма приведена в приложении Г (рис. Г.1). Дендрограмма показывает общее сходство 26 MD8-подобных протеомов, помещая эти фаги в одну кладу. Она также указывает на небольшое сходство между протеомами группы MD8 и фагов *Lambdavirus*. Исходя из анализа топологии дерева, можно было бы предположить, что фаги TC7 и ϕ 2 являются родственными фагами, которые дивергировали раньше других членов группы MD8. Кластеризация, сделанная с помощью ViPtree, согласуется с кластеризацией VIRIDIC (приложение Г, рис. Г.2).



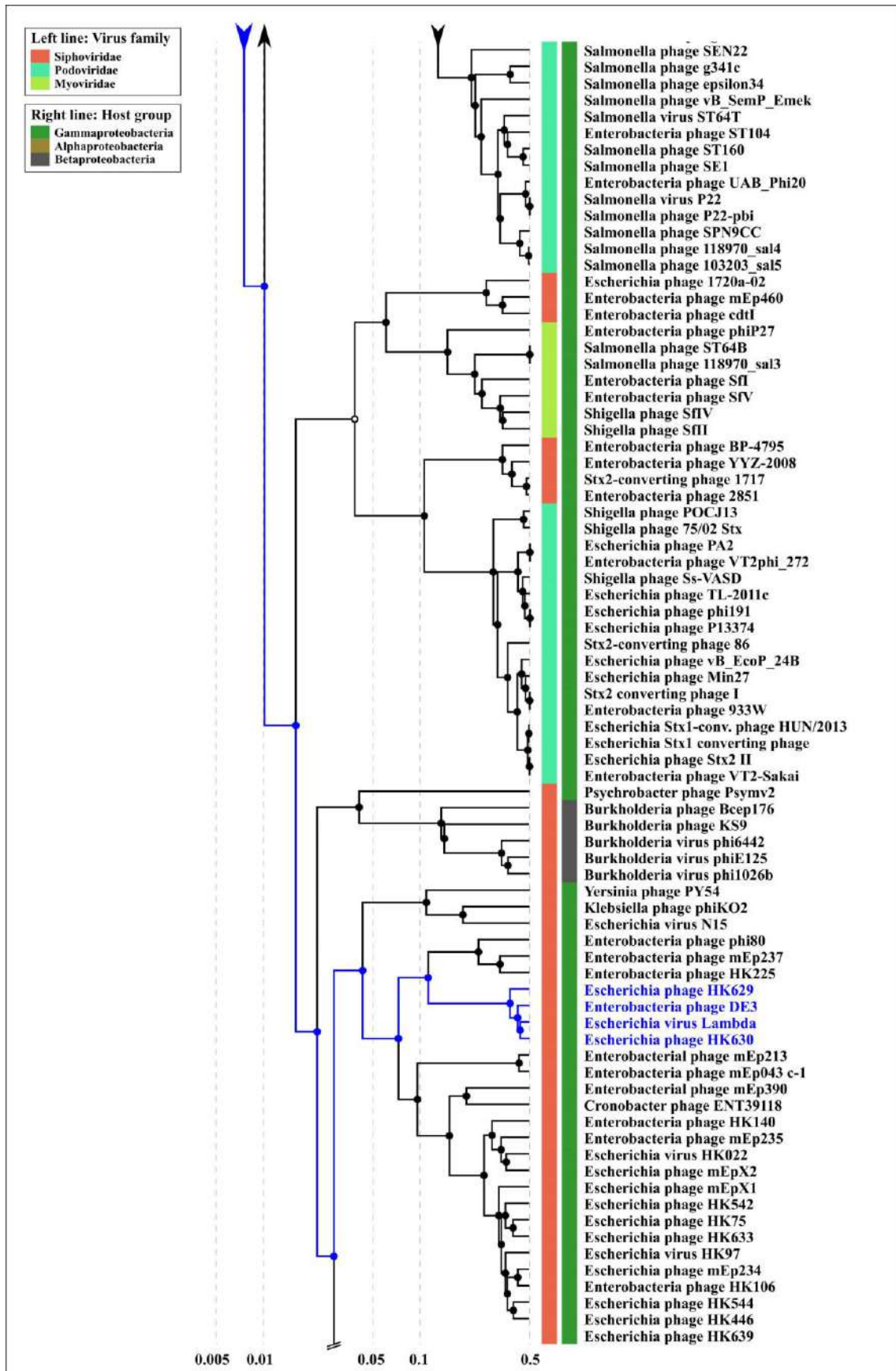


Рисунок 52. Филогенетическое дерево, полученное на основе сходства протеомов MD8-подобных (обведены красным) и других фагов. Таксономия фагов (Virus family) по состоянию на 2020 г. и хозяина (Host group) указаны в легендах.

3.3.2. Анализ мозаичности генома

Чтобы проверить предположение о мозаичном строении и химерном происхождении генома фага MD8, было проведено сравнение геномов и генов MD8-подобных и других фагов. Сравнение геномов MD8 с MD8-подобными фагами (рис. 53, верхняя схема) показывает, что формирование генома MD8 может быть связано с потерей 5'-концевых генов, представленных в геномах TC7 и φ2. Похожая ситуация наблюдается с блоками генов MD8-подобных фагов JBD68 и vB_Pae_BR133a, только она затрагивает гены, находящиеся в 5'-концевой, правой части генома MD8, как это показано на рисунке. Похоже, что это обмен генами с той же функциональностью, но разного происхождения. Аналогичная ситуация наблюдается для фагов λ и N15 (рис. 53), но рекомбинация произошла в другом месте. Ранее сообщалось о рекомбинации, ведущей к мозаицизму генома, с участием подовирусного фага Ab31 с другим λ-подобным фагом *Pseudomonas* [229]. Возможно, новые лямбдоидные бактериофаги могут возникать в результате обмена частями геномов фагов. Геномные сравнения MD8 с лямбдоидным фагом vB_Pae_BR161b, не принадлежащим к группе MD8, сифовирусом φ297, миовирусом vB_PaeM_PA01_Ab04 и подовирусом KPP25 показывают, что в генетических обменах могут участвовать даже отдаленно родственные фаги.

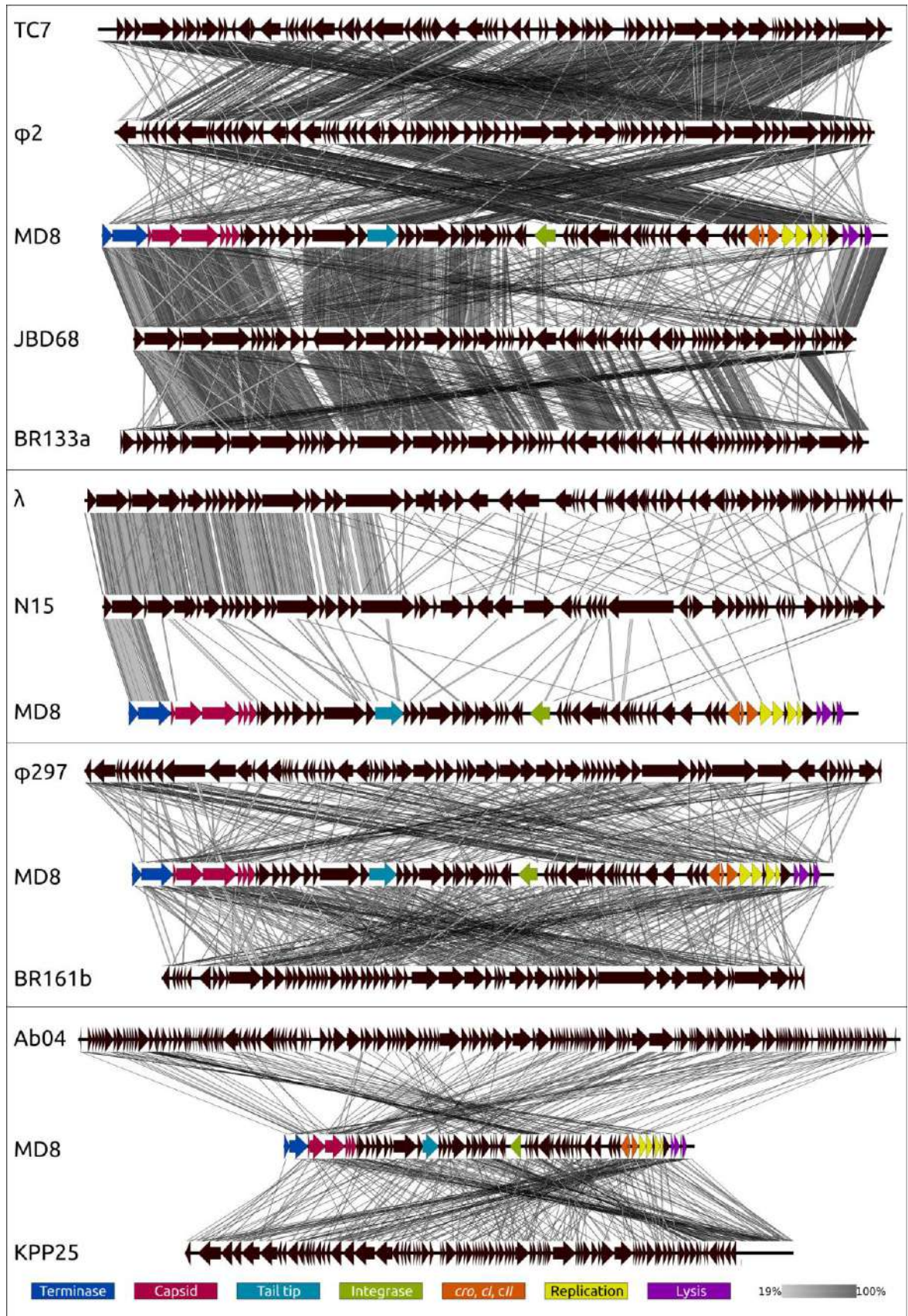


Рисунок 53. Сравнение геномов фагов MD8-подобных и других фагов. Гены белков окрашены в соответствии с легендой. Степень сходства указана на градиентной шкале внизу рисунка.

Структуру и возможное происхождение генома MD8 также изучали с помощью филогенетического анализа и сравнения последовательностей всех белков MD8. При изучении отдельных белков, предсказанные последовательности белков фага MD8 использовали для поиска BLASTP по базе фаговых последовательностей NCBI GenBank PHG. Если количество найденных гомологов было три или меньше, то использовали значения попарной идентичности, в противном случае для нахождения родственных последовательностей использовались филогенетические деревья. Результаты анализа согласуются с результатами геномного сравнения и указывают на то, что 5'-часть генома MD8, включая гены терминазы и капсида 1–10, имеет происхождение, отличное от большинства других генов. Большинство белков этого блока больше похожи на белки группы фагов, включающей MD8-подобный фаг F10 и другие, чем на белки фагов φ2 и TC7. Последние два фага содержат сходные гены терминазы и капсида, которые филогенетически далеки от других MD8-подобных фагов, кроме того, в отличие от остальных MD8-подобных фагов, протеаза и капсид фагов φ2 и TC7 кодируются отдельными генами. При этом анализ деревьев и выравниваний показал, что в части генома MD8, кодирующей гены поддержания лизогении и репликации, преобладают гены, более сходные с φ2 и TC7. Блок белков хвоста ближе к MD8-подобному фагу JBD68, а белки лизиса больше похожи на белки фага BR133a и других. Таким образом, геном MD8 можно условно разделить на несколько частей или модулей, включая F10-подобный модуль, JBD68-подобный модуль и φ2-подобный модуль (рис. 54). Топологии филогенетических деревьев генов, принадлежащих к одним и тем же функциональным модулям, часто схожи, что иллюстрирует теорию модульной эволюции [144]. Например, два гена терминазы, кодирующие её малую и большую субъединицы, по-видимому, обладают общей недавней эволюционной историей, так же, как и структурные гены.

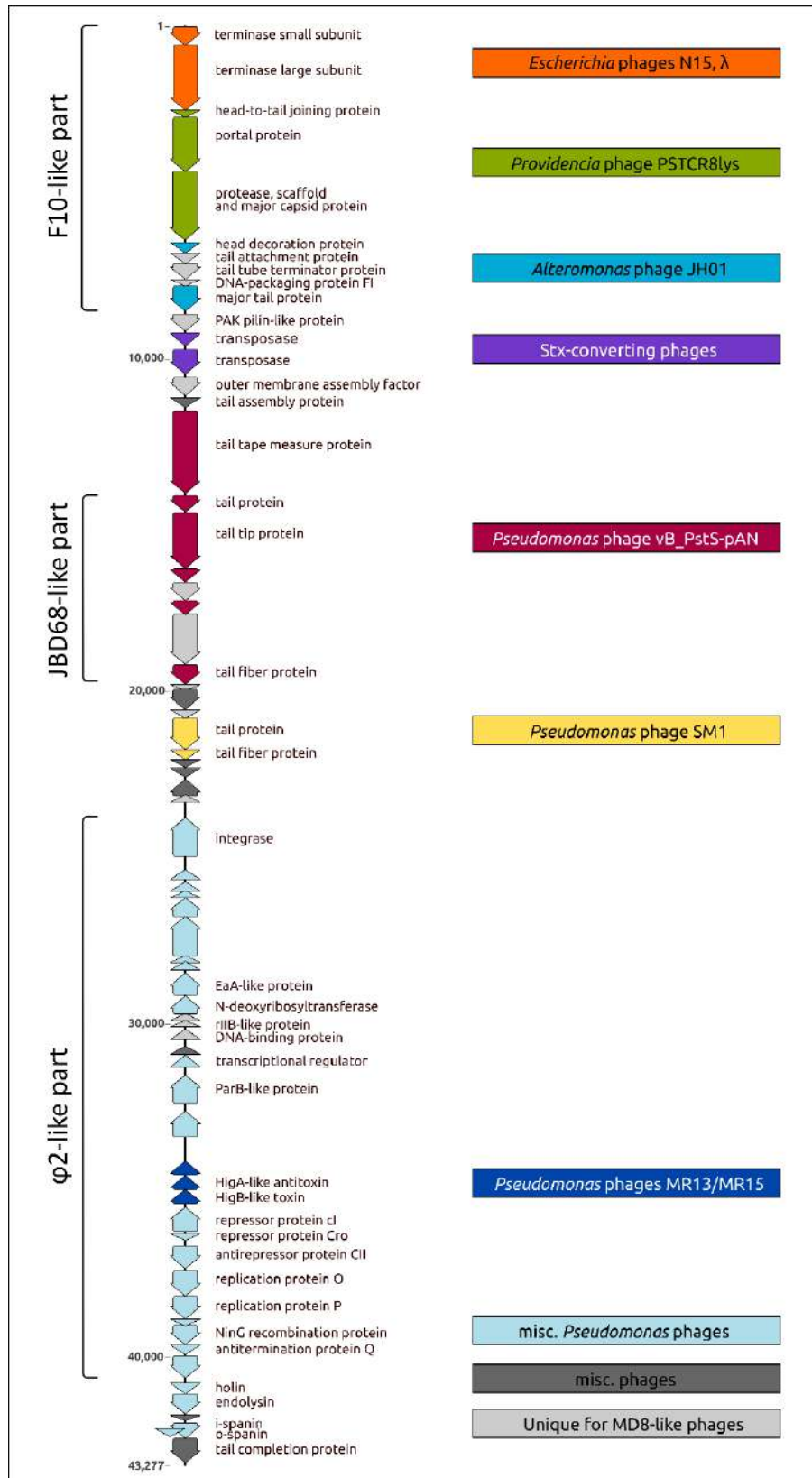


Рисунок 54. Генетическая карта фага *Pseudomonas* MD8. Слева указаны геномные модули в соответствии с названием MD8-подобного фага с самыми похожими генами. Справа указаны родственные фаги не из группы MD8.

Полученные результаты указывают на то, что организация генома MD8 в целом схожа с организацией генома фага λ , однако лишь немногие белки обнаруживают заметную гомологию с белками лямбдоидных фагов. Более того, белки MD8-подобных фагов демонстрируют разную эволюционную историю. Это ставит вопросы о самой возможности построения таксономической классификации, отражающей эволюционную историю лямбдоидных и других умеренных фагов.

Тем не менее, наличие осмысленной таксономической классификации необходимо как для фундаментальных, так и прикладных исследований. Возможно предположить, что такая классификация может быть предложена хотя бы для некоторых λ -подобных фагов. Можно отметить следующие особенности методов таксономической классификации с точки зрения их применения в анализе фагов, подверженных высокому уровню генетического мозаицизма:

1. Биологические свойства, геномная организация и классификация на уровне семейства и выше. Таксономическое описание фагов на этом уровне, основанное на их биологических свойствах, таких как образ жизни, условия роста, устойчивость к УФ-излучению и др., не всегда представляется возможным. Недавние открытия показали, что образ жизни фага может быть разным даже в пределах такой давно установленной таксономической группы, как семейство *Autographiviridae*, где были обнаружены умеренные фаги [230–232]. Для общего описания фагового генома часто используют геномную организацию, включающую перечень и порядок генов, регуляторных элементов, пространственную структуру ДНК [233,234]. Хотя ярко выраженный генетический мозаицизм делает невозможным установление непротиворечивых эволюционных отношений, используя отдельные гены лямбдоидных фагов с разной эволюционной историей, лямбдоиды показывают сходство в организации генома. Возможно, было бы целесообразно классифицировать все λ -подобные фаги на уровень семейства или более высокорангового таксона, используя организацию

генома как основной критерий принадлежности к группе независимо от происхождения отдельных генов.

2. Межгеномное сравнение и протеомный анализ. Результаты межгеномного сравнения и протеомного анализа MD8-подобных фагов частично согласуются с результатами филогенетического анализа отдельных генов. Тем не менее, в некоторых случаях алгоритмы кластеризации не могут определить единственную группу, к которой принадлежит изучаемый фаг из-за его сходства сразу с несколькими группами по причине модульного строения его генома. Обычный порог для разграничения видов бактериофагов на основе геномного сходства составляет 95%, а для разграничения родов – 70%. В случае, когда сразу несколько групп фагов показывают существенное взаимное сходство (десятки процентов межгеномного сходства по VIRIDIC) между собой, возможно, имеет смысл уменьшить порог для разграничения родов, либо определять принадлежность к роду на основании общей эволюционной истории маркерных белков, например, терминазы и ГКБ.

3. Филогенетический анализ отдельных генов и геномных модулей. Филогенетические исследования продемонстрировали различную эволюционную историю белков фага MD8. Это в разной степени характерно не только для λ -подобных и умеренных фагов, но и для других хвостатых бактериофагов [143]. Выяснение истории эволюции генетических модулей может помочь различить разные группы фагов, но задача построения последовательной иерархической таксономии является более сложной.

Если бы геномная организация могла быть основным критерием для классификации всех лямбдоидных фагов на уровне семейства, то разграничение подсемейств могло бы быть основано на таксономии инфицированных хозяев. Гены профагов лямбдоидных фагов входят в состав хромосом хозяина или существуют в виде плазмиды. Интересно, что расчеты ANI показывают около 87% идентичности нуклеотидов при высоком покрытии фаговой последовательности

для фага MD8 и *Pseudomonas aeruginosa* (штамм ВН9), 95% для фага λ и *Escherichia coli* и 73% для фага N15 и *Escherichia coli* (штамм ТНО-003). Все белки фага MD8 имеют гомологов в геномах *Pseudomonas*. Поэтому, вероятно, было бы целесообразно отнести все лямбдоидные фаги *Pseudomonas* к одному таксону.

В соответствии с результатами сделанных анализов и учитывая вышеприведённые соображения, 22 бактериофага *Pseudomonas*, включая MD8, F10, JBD68, PAN70, BR133a, BR144a, BR200a, BR204a, BR213a, BR233a, BR299a, CF24b, CF28b, CF34a, CF55b, CF60a, CF67a, CF79a, CF140a, CF145a, CF165a и CF208a можно было бы отнести к представителям одного рода. Межгеномное сходство внутри кластера VIRIDIC этих фагов ниже родового порога в 70%, но эти фаги обладают сходными генами терминазы и ГКБ. Ген ГКБ этих фагов кодирует пропептид, включающий протеазу, белок скаффолда и ГКБ, а большинство генов этих фагов имеют близкую недавнюю эволюционную историю. Межгеномное сходство фагов BR201a и CF3a, по сравнению с большинством вышеперечисленных, превышает 70%, и хотя основная часть белков этих двух фагов очень сходны с белками 22 перечисленных выше фагов, в их неполных геномных последовательности отсутствует секвенированный участок, содержащий гены терминазы и ГКБ, поэтому точная классификация пока невозможна. Фаги BR204a, BR233a, BR299a, CF24b, vB_Pae_CF28b, CF140a, CF145a, CF165a и CF208a, а также фаги BR133a, CF55b и CF79a представляют собой расы одних и тех же двух видов из-за высокого уровня сходства геномов (межгеномное сходство, определяемое VIRIDIC, составляет выше 95%). Фаги ϕ 2 и TC7 могут составлять еще один род подсемейства λ -подобных фагов, инфицирующих *Pseudomonas*, их межгеномное сходство, определяемое VIRIDIC, составляет 70,7%. Помимо этого, данные бактериофаги обладают сходными терминазами и ГКБ, лишь отдаленно родственными терминазам и ГКБ остальных MD8-подобных фагов.

3.4. Использование предсказаний структуры фаговых белков для эволюционной таксономии

3.4.1. Эволюция чехольных белков

Чехольные белки (ЧБ) хвоста фагов являются частью молекулярного механизма сокращения, присутствующего у бактериофагов с миовирусной морфологией, в бактериальной сократительной системе инъекции (CIS, contractile injection system) и системой секреции VI типа (T6SS, type VI secretion system), обнаруженной у многих грамотрицательных бактерий. Первым объектом изучения чехольных белков был ЧБ фага T4 (gp18), у которого была определена структура фрагмента (аминокислоты 83–365), устойчивого к протеазам, а также делеционного мутанта этого белка (остатки 1–510) [27]. Хотя в дальнейшем был экспериментально определён ещё ряд структур, общее количество известных структур не соответствует разнообразию вирусных и бактериальных сократительных систем.

По состоянию на середину 2022 г. в базе данных RCSB Protein Data Bank [23, 43] присутствовали данные о структурах 15 чехольных белков, определенных со сравнительно высоким разрешением в 1,8–4,2 Å, принадлежащих фагам, профаговым областям, бактериоцинам, так называемой «антипищевой профаговой системе» (AFP, anti-feeding prophage system), относящейся к CIS, и системе секреции VI типа (приложение Д, табл. Д.1). Кроме того, некоторые белковые структуры и реконструкции систем были определены с более низким разрешением. Чехольные белки, относящиеся к AFP *Serratia entomophila* и CIS *Photorhabdus asymbiotica*, кодируются тремя соседними генами, расположенными в кластере генов сократительной молекулярной машины. Выравнивание аминокислотных последовательностей чехольных белков, кодируемых разными генами, показало заметную гомологию между белками, принадлежащими этим двум видам.

Наложение экспериментально определённых структур в PyMOL показало сходство чехольных белков, принадлежащих фаговым хвостам, T6SS, AFP и CIS.

Структурное выравнивание экспериментально полученных структур с помощью mTM-align продемонстрировало наличие законсервированного ядра, общего для всех выровненных белков (рис. 55). К сожалению, в нескольких экспериментально определенных структурах отсутствовали некоторые фрагменты, присутствующие в аминокислотных последовательностях соответствующих генных продуктов. Использование моделей полных белков, полученных с помощью AlphaFold 2, позволило определить границы законсервированной части более точно (рис. 55). Анализ структурного выравнивания показал, что законсервированная часть белка из фагов *Escherichia* T4 и *Staphylococcus* 812 прерывается длинными вставками и включает остатки, расположенные как в N-концевой, так и в C-концевой частях (приложение Д, рис. Д.1). Визуальный анализ экспериментально определённых структур и предсказанной модели ЧБ фага T4 показал, что общее консервативное ядро расположено к хвостовой трубке ближе, чем остальные части белка (рис. 56). N-конец ЧБ расположен дальше от белков хвостовой трубки, чем C-конец, но домены вне общего ядра расположены еще дальше от хвостовой трубки. Это указывает на возможные взаимодействия между белками хвостовой трубки и чехольными белками, а также на то, что законсервированная часть ЧБ может быть важна для правильной сборки фагового хвоста. Эти наблюдения согласуются с наблюдениями и предположениями авторов реконструкции фага *Staphylococcus* 812, полученной с помощью криоэлектронной микроскопии [235].

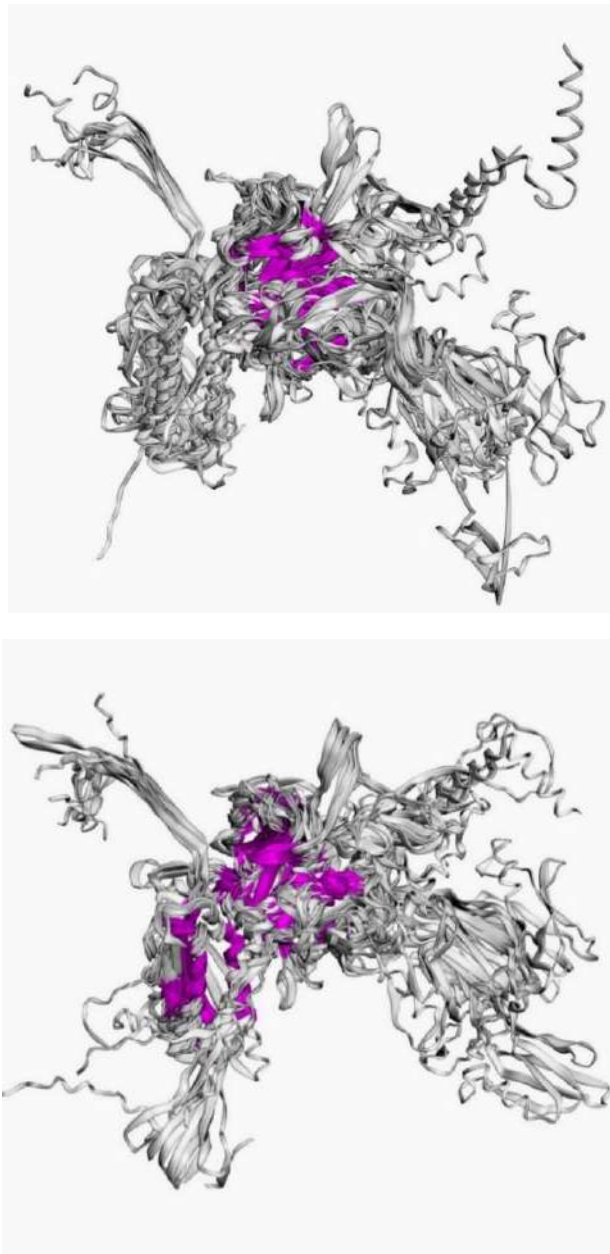


Рисунок 55. Визуализация структурного выравнивания, выполненного с помощью mTM-align для 15 экспериментально определенных ЧБ: фага *Escherichia* T4; профага *Desulfitobacterium hafniense*, профага *Listeria innocua*, фага *Staphylococcus* 812, бактериоцина *Peptoclostridium difficile*, пиоцина R2 *Pseudomonas aeruginosa*, T6SS *Francisella tularensis* subsp. *novicida*, T6SS *Pseudomonas aeruginosa*, T6SS *Vibrio cholerae*, AFP *Serratia entomophila* и AFP *Photorhabdus asymbiotica*). Части с максимальным расстоянием между парами остатков менее 4 Å окрашены в пурпурный цвет. Верхнее выравнивание содержит экспериментально определённые структуры, нижнее – предсказанные AlphaFold 2.

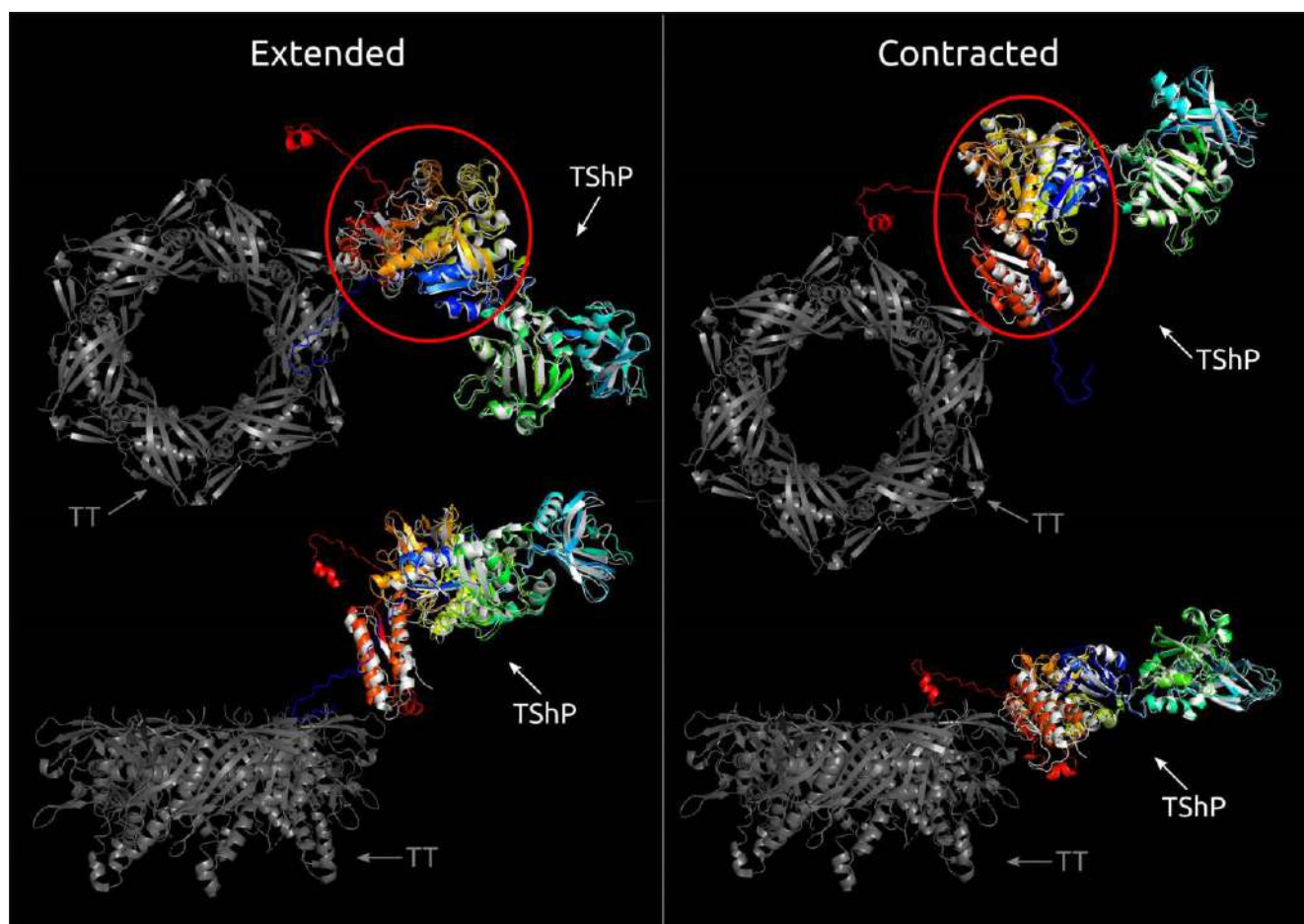


Рисунок 56. Структура гексамера хвостовой трубки (ТТ, изображена серым) и чехольного белка (TShP, изображён белым), соответствующие крио-ЭМ-реконструкции хвоста Т4 в растянутом (код PDB 3J2M) и сокращенном (код PDB 3J2N) состояниях [236] с наложенной предсказанной структурой ЧБ, изображённой радужным градиентом, где N-конец полипептидной цепи окрашен в синий цвет, а С-конец – в красный. Законсервированное ядро обведено красным.

Структурное моделирование с помощью AlphaFold проводили с использованием ЧБ представителей четырёх (на момент проведения исследования) ратифицированных ICTV семейств бактериофагов с миовирусной морфологией (*Myoviridae*, *Ackermannviridae*, *Chaseviridae* и *Herelleviridae*) и чехольного белка, предсказанного для фага *Paenibacillus Lily*, принадлежащего к роду *Lilyvirus*, не отнесенному ни к одному из фаговых семейств [91,237]. Фаг Lily был вначале описан как сифовирус, но его геном содержит ген ЧБ и показывает высокий уровень сходства с миовирусом *Paenibacillus ERIC V* [238] (ANI 81,6%). Отдельное внимание было уделено хвостатым вирусам архей (архефагам) из-за их большого

значения для систематики и эволюционной биологии. Многие архейные вирусы морфологически неотличимы от хвостатых бактериофагов [69,142], а геномный анализ архейных миовирусов показал наличие чехольных белков, напоминающих некоторые бактериальные миовирусы.

В январе 2022 года база данных NCBI GenBank PHG содержала 43 полных генома архейных миовирусов. Ниже перечислены архефаги, геномы которых содержат предсказанные гены чехольных белков:

- Вирусы *Haloarcula*, принадлежащие к роду *Haloferacalesvirus*: десять полных геномов (HCTV-6, -7, -8, -9, -10, -11, -15 и HJTV-1, -2, -3) обладают сходными размерами и организацией генома. Предсказанные ЧБ содержат от 431 до 438 аминокислотных остатков (а.о.). ЧБ HCTV-6 и HCTV-15 идентичны и отличаются по своей аминокислотной последовательности от восьми других ЧБ (% идентичности составляет около 45–47%). Последние белки очень похожи или идентичны, показывая попарное сходство 84% и выше.
- Вирусы *Halobacterium*, принадлежащие к роду *Myohalovirus*: два полных генома (phiH и ChaoS9) содержат чехольные белки длиной около 430 а.о., которые показывают 52% сходство и дают одинаковые результаты при сравнении мотивов ННМ с помощью HHPred.
- Вирус *Haloferax* HF1, принадлежащий к роду *Haloferacalesvirus*: аминокислотная последовательность чехольного белка HF1 аналогична последовательностям архефагов *Haloarcula* HCTV-7, -8, -9, -10, -11 и HJTV 1, -2, -3 и идентична последовательностям архефагов *Haloarcula* примерно на 90%.
- Вирусы *Halorubrum*, принадлежащие к роду *Haloferacalesvirus*: 28 геномов кодируют ЧБ длиной около 430 а.о. Из них 23 последовательности ЧБ показывают взаимное сходство 83–99%. Эти ЧБ принадлежат архефагам *Halobrum* HF2, *Halorubrum* 5 и 8, *Hardycor2*, *Halorubrum* HRTV-9, -10, -13, -14, -15, -16, -17, -18, -19, -20, -21, -22, -23, -24, -26, *Halorurum* HSTV-4, *Serpecor1* и VOLN27B. Чехольные белки *Halovirus* HSTV-2, *Halorubrum* Tailed Virus 7 и *Halorubrum* HRTV-2 и HRTV-11 показывают 97–98% сходство друг с другом и

менее 50% идентичности со всеми другими фагами *Halobrum*, составляя отдельную группу чехольных белков архефага *Halorurum*. Чехольные белки архефагов HRTV-25 и HRTV-27 демонстрируют менее чем 40–50% идентичности аминокислотной последовательности друг с другом и со всеми другими ЧБ архефагов *Halobrum*.

– Вирус *Natrialba* φCh1 рода *Myohalovirus*, в котором был обнаружен неаннотированный ген ЧБ. Белок имеет длину 426 а.о. и демонстрирует около 40% идентичности с фагами *Halobacterium*.

Для моделирования использовались представители всех перечисленных выше групп. Поиск BLAST с использованием базы данных NCBI GenBank BAC, содержащей архейные и бактериальные хромосомные и плазмидные последовательности, показал, что предполагаемые ЧБ кодируются в геномах штаммов *Natronorubrum bangense* JCM10635, *Methanolacinia Petrolearia* DSM 11571 и других представителей типа Euryarchaeota. Интересно, что гомологи чехольных белков также могут быть обнаружены в геномах архей, которые являются частью группы Asgard (Lokiarchaeota, Thorarchaeota), а также типов Crenarchaeota, Bathyarchaeota и Rasearchaeota. Большая часть этих последовательностей была получена с помощью метагеномного анализа. На настоящий момент хвостатые вирусы архей, не относящиеся к типу Euryarchaeota, выделены не были.

Дополнительно с помощью FastTree было кластеризовано около 2000 последовательностей ЧБ, проверенных поиском BLAST и сравнением мотивов НММ, полученных из аннотированных или реаннотированных вирусных и прокариотических геномов. В тех случаях, когда геном фага кодировал более одного ЧБ, для дальнейшего анализа использовали только один. Из кластеризованных последовательностей было отобрано 109 представителей фаговых последовательностей, представляющих различные клады дерева и таксоны, к которым были добавлены 54 последовательности, представляющие отдельный интерес (приложение Д, табл. Д.2). В число добавленных 54

последовательностей входили различные представители архейных и бактериальных вирусных семейств *Myoviridae*, *Ackermannviridae*, *Chaseviridae* и *Herelleviridae*, а также рода *Lilyvirus*. Также в это число вошли последовательности ЧБ из геномов джамбо-фагов рекордного размера, найденные с помощью метагеномного анализа [77] и несколько последовательностей, найденных в геномах прокариот. Последовательности экспериментально определённых структур ЧБ также использовались для дальнейшего анализа. Итоговое общее число отобранных последовательностей составило 153. В это количество вошли 112 чехольных белков прокариотических вирусов, 25 предсказанных ЧБ из прокариотных хромосом и плазмид, 8 ЧБ системы секреции VI типа, 6 ЧБ системы AFP и 2 ЧБ бактериоцинов (пиоцина и диффоцина). Функциональные принадлежности всех выбранных белков были подтверждены поиском BLAST и сравнением мотивов ННМ.

Визуальный анализ структурных моделей, полученных с помощью AlphaFold, показал различную структурную архитектуру. Все структуры имели общий консервативный домен, состоящий из N-концевой и C-концевой частей, некоторые модели имели дополнительные домены (с точки зрения домена как компактной структуры [239]). В части случаев смоделированные структуры не имели четко различимых границ доменов, но в большинстве случаев количество доменов можно было оценить (см. приложение Д, табл. Д.2). Самый маленький по размеру аминокислотной последовательности смоделированный однодоменный чехольный белок бактериофага принадлежал фагу *Burkholderia* ВЕК – представителю рода *Tigrvirus* подсемейства *Peduovirinae* семейства *Myoviridae*. Он состоял из 341 а.о., а большая часть белка показывала структурное сходство с законсервированным ядром экспериментально определенных структур, обнаруженных с помощью структурного выравнивания и показанных на рис. 55 (рис. 57). Структурную архитектуру чехольных белков бактериоцинов, T6SS и AFP также можно описать как однодоменные структуры.

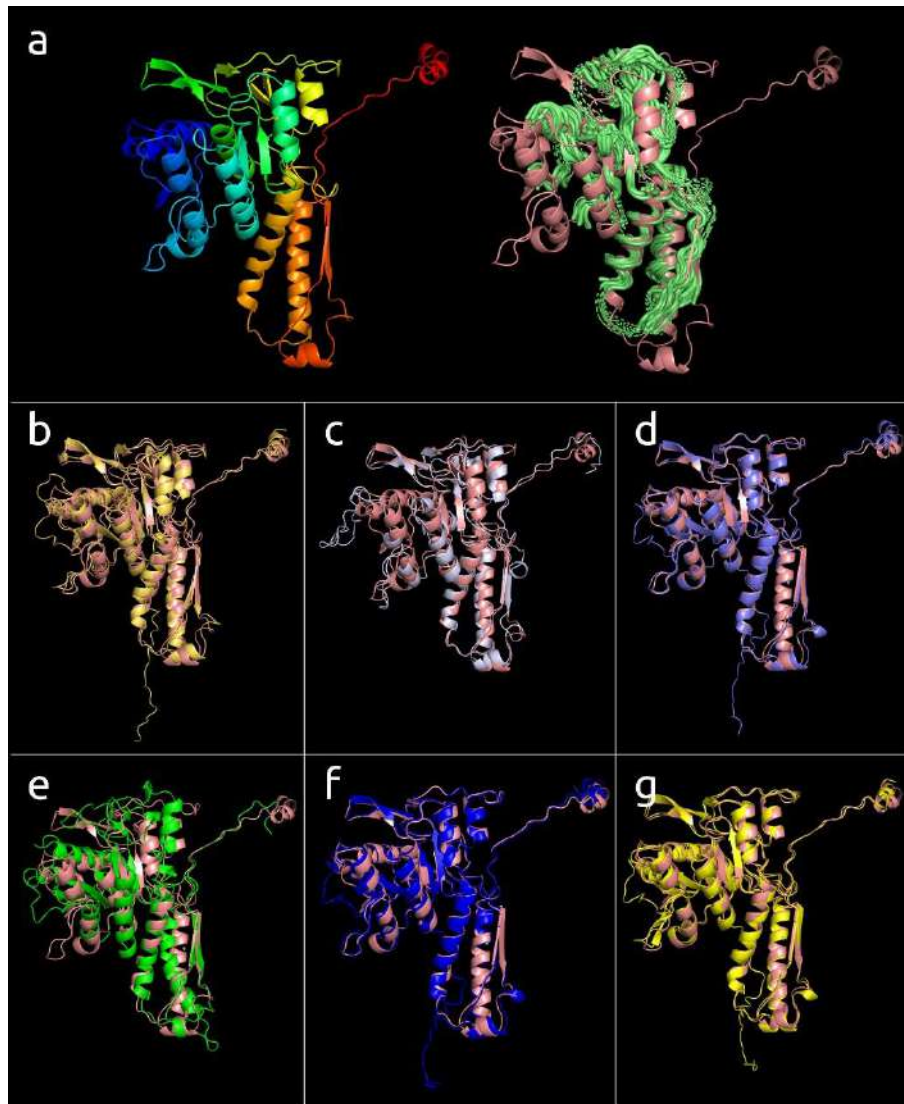


Рисунок 57. (а) – структурная модель ЧБ фага *Burkholderia* ВЕК (слева), изображённая радужным градиентом, где N-конец полипептидной цепи окрашен в синий цвет, С-конец – в красный, и та же модель справа (показанная светло-красным цветом) с наложенным законсервированным ядром (изображённым зелёным), показанном на рис. 55. (b-g) Модели однодоменных ЧБ (показаны разным цветом) с наложенной моделью ЧБ фага *Burkholderia* ВЕК (показанной светло-красным цветом), по порядку: ЧБ фага *Halomonas* НАР1, предсказанный ЧБ, обнаруженный в сборке генома, атрибутированного как изолят археи *Candidatus* Bathyarchaeota, ЧБ фагов *Erwinia* ENT90, *Flavobacterium* FPSV-S1, *Ralstonia* RSY1, *Vibrio* vB_VpaM_MAR .

Остальные смоделированные чехольные белки содержали не только часть, структурно сходную с однодоменными ЧБ, которую мы далее будем называть «основным доменом», но также обладали дополнительными доменами. Для некоторых смоделированных структур не удалось четко определить, можно ли

считать часть белка, не принадлежащую к основному домену, одним доменом. Это могло быть как связано с характерной сложной структурой части, не принадлежащей к основному домену, так вызвано неточностью моделирования. Структурная архитектура большинства белков, содержащих домен(ы) кроме основного, характеризовалась наличием двух доменов, включая основной (рис. 58). Дополнительный домен, как правило, содержал мотивы, включающие β -листы, и часто содержал иммуноглобулиноподобные (Ig-подобные) β -сэндвичи и короткие α -спиральные участки. Все выделенные вирусы архей содержали двухдоменные ЧБ. Кроме того, двухдоменные ЧБ были обнаружены у фагов, относящихся к семейству *Chaseviridae*, различных родов *Myoviridae*, фага *Paenibacillus* Lily, в последовательностях, закодированных в прокариотных хромосомах и содержащихся в геномных сборках грамположительных и грамотрицательных бактерий, а также архей, относящихся к типам Crenarchaeota, Euryarchaeota, Thaumarchaeota, и Thorarchaeota (см. приложение Д, табл. Д.2). Двухдоменные ЧБ имели размер от 426 до 713 аминокислотных остатков, самые большие из них в основном принадлежали джамбо-фагам, инфицирующим γ -протеобактерии. Двухдоменные ЧБ выделенных вирусов архей были меньше по размеру, чем большинство других двухдоменных ЧБ, и содержали дополнительный домен, который, в основном, состоял из β -листов, образующих β -сэндвич.

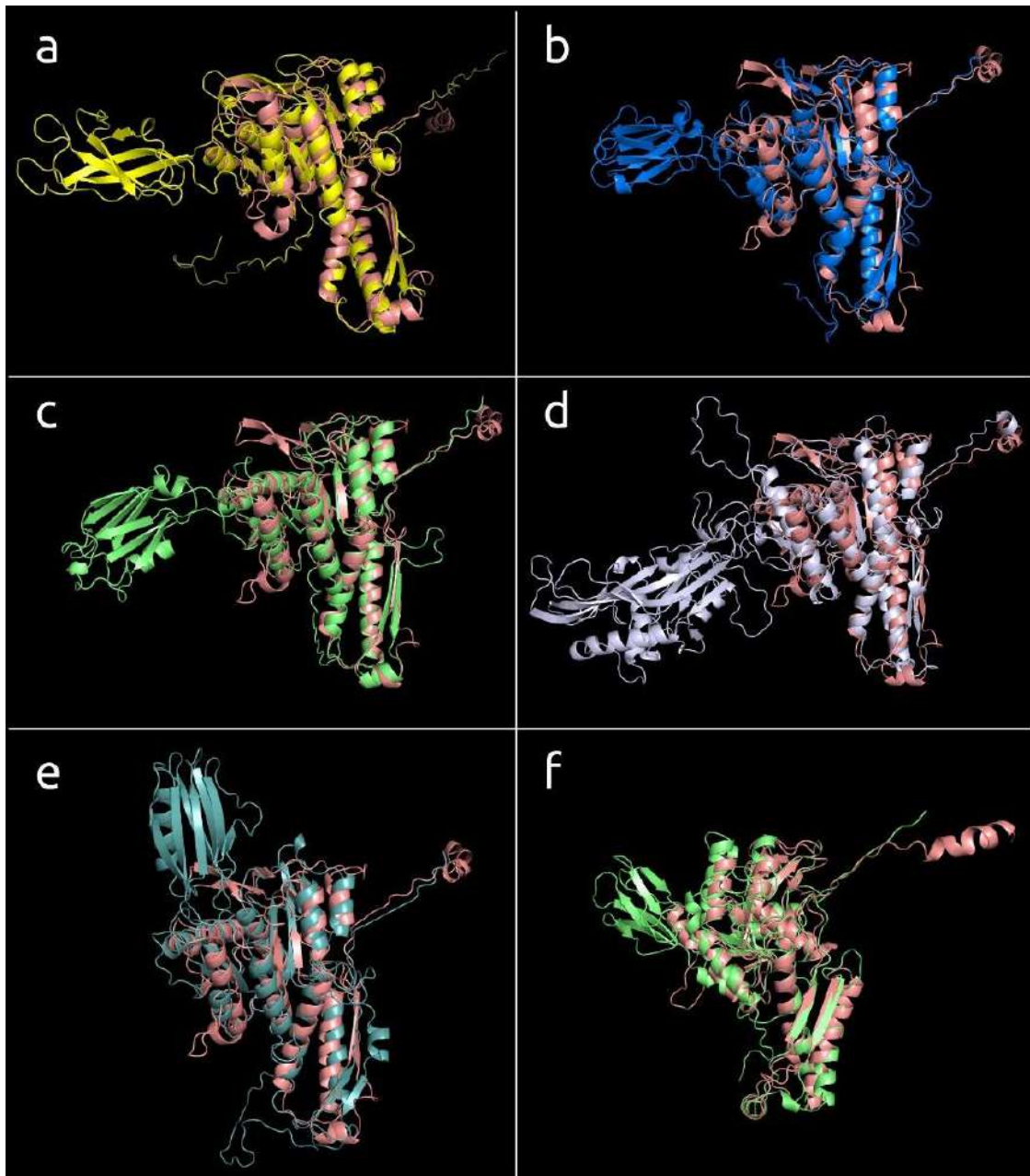


Рисунок 58. Модели двухдоменных чехольных белков (показаны разным цветом) с наложенной моделью ЧБ фага *Burkholderia* ВЕК (показанной светло-красным цветом), по порядку: (a) фаг *Mycobacterium Phabba*; (b) фаг *Brevibacillus Jimmer2*; (c) предсказанный ЧБ, обнаруженный в сборке генома, атрибутированного как *Thermoprotei archaeon B19_G17*; (d) фаг *Erwinia phage vB_EamM_RisingSun*; (e) фаг *Escherichia phage Mu*; (f) архейный вирус *Halobacterium ChaoS9*.

Около трети смоделированных структур показали более сложную архитектуру. Структурная архитектура этих белков, по-видимому, является дальнейшим эволюционным развитием двухдоменной структуры, описанной выше, отличаясь присутствием трех и более доменов (рис. 59). Как и в случае

двухдоменных чехольных белков, у мультидоменных ЧБ дополнительные домены часто содержали β -сэндвичи. В отличие от большинства моделей мультидоменных ЧБ, модели чехольных белков двух родственных (ANI 99,0%) фагов *Bacillus*, AR9 и PBS1, включали дополнительные домены, состоящие в основном из α -спиралей (приложение Д, рис. Д.2). Как и в случае двухдоменных ЧБ и экспериментально определенных структур, дополнительные домены располагались в стороне от основного домена, контактирующего с хвостовой трубкой. Большинство фаговых геномов размером более 100 тыс. н.п. кодировали ЧБ с тремя или более доменами. Наибольшее количество доменов, пять и более, было обнаружено у фагов *Ackermannviridae* (размер генома примерно 140–170 тыс. н.п.), а также джамбо-фагов с размером генома 200 тыс. н.п. и более. Варианты структурной архитектуры этих белков включали дополнительные домены, образованные одним непрерывным участком полипептидной цепи, или двумя участками, один из которых находился ближе к N-концу белка, а другой принадлежал возвратной части полипептидной цепи, расположенной ближе к C-концу белка. Некоторые типы структурной архитектуры смоделированных белков показаны на рис. 60. Как правило, число дополнительных доменов и тип структурной архитектуры был одинаковым для близких таксонов.

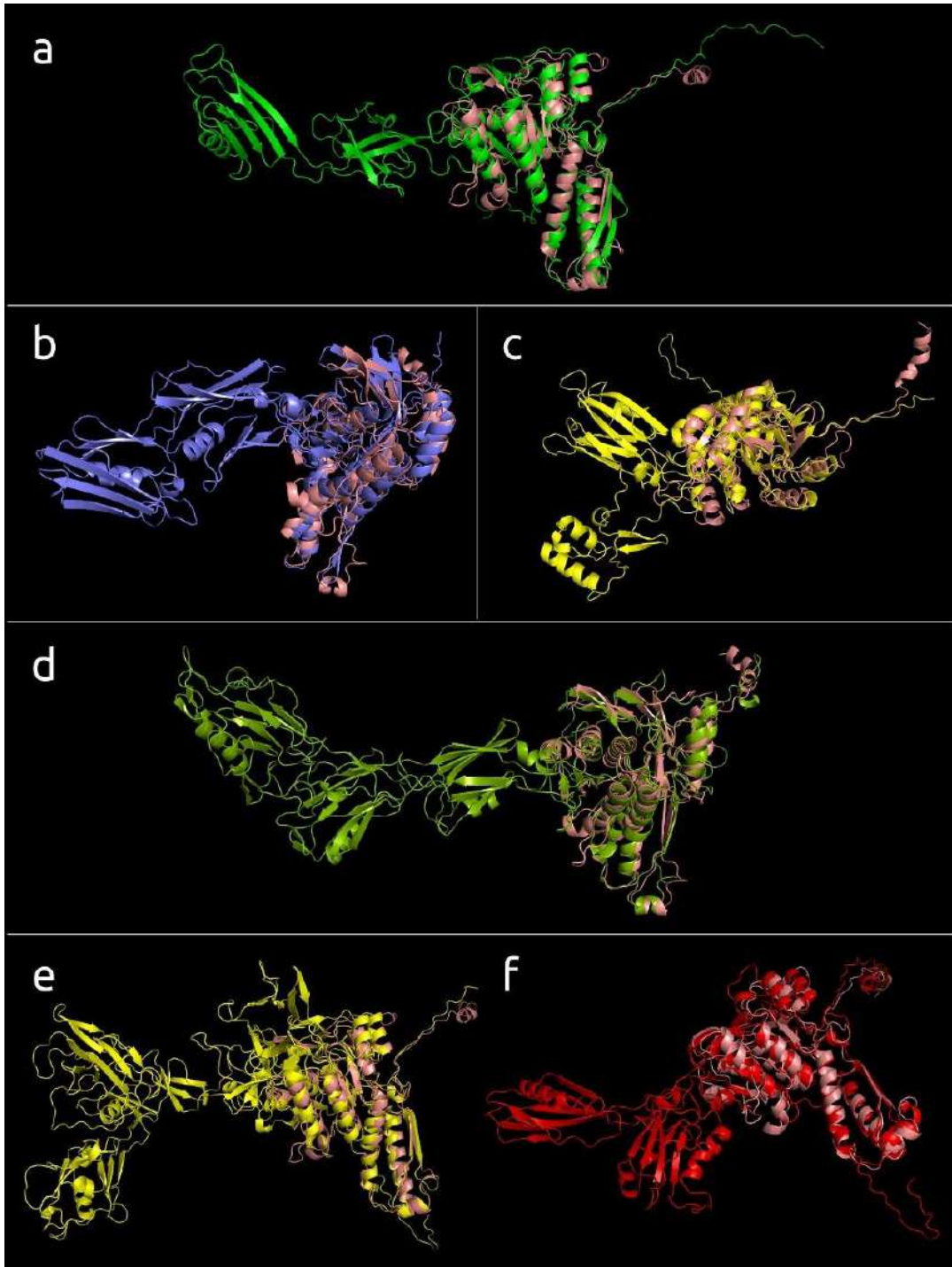


Рисунок 59. Модели мультидоменных чехольных белков (показаны разным цветом) с наложенной моделью ЧБ фага *Burkholderia* ВЕК (показанной светло-красным цветом), по порядку: (а) ЧБ фага *Agrobacterium* Atu_ph07; (б) ЧБ фага *Bacillus* BC01; (с) ЧБ профага, обнаруженного в геноме *Halovivax ruber* ХН-70; (d) ЧБ, обнаруженный в сборке генома *Crenarchaeota archaeon* isolate__LB_CRA_1; (е) ЧБ фага *Salicola* SCTP-2; (f) ЧБ фага *Serratia* phiMAM1.

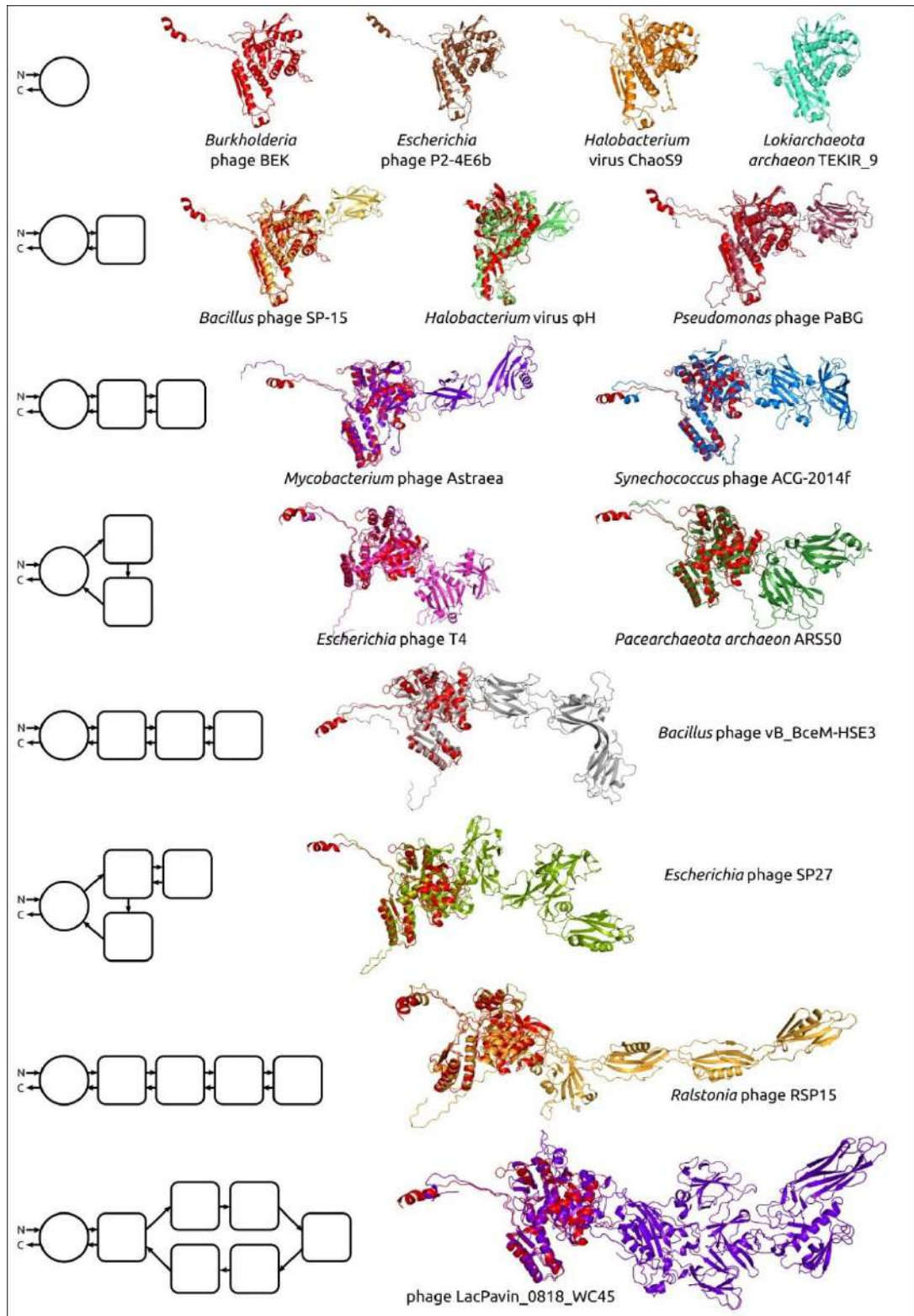


Рисунок 60. Примеры структурной архитектуры смоделированных сократительных чехольных белков. На ЧБ, состоящие из двух и более доменов, наложена смоделированная структура ЧБ фага *Burkholderia* ВЕК, показанная красным цветом. Схемы слева показывают структурную архитектуру белков.

Основной домен изображен в виде круга, а дополнительные домены представлены в виде квадратов со скругленными углами. Направление полипептидной цепи от N-конца к C-концу показано стрелками.

Кластеризацию смоделированных ЧБ проводили с использованием множественного структурного выравнивания с помощью программы mTM-align, оценивая сходство структур с помощью TM-score. Матрица, содержащая парные значения TM-score, использовалась для кластеризации с помощью BioNJ и для построения дендрограммы на основе структурного сходства (приложение Д, рис. Д.3).

Полученное дерево поместило чехольные белки фагов и другие чехольные белки отдельно от чехольных белков T6SS и показало их некоторое сходство с ЧБ гигантских фагов Mad1_20_16 и LacPavin_0818_WC45. Чехольные белки архей были сгруппированы в несколько клад, галоархейные миовирусы, принадлежащие к родам *Haloferacalesvirus* и *Myohalovirus*, группировались в две монофилетические ветви в соответствии с их таксономической классификацией. Интересно, что два ЧБ галоархейных профагов, по-видимому, структурно сходны с белками фагов *Haloferacalesvirus* и *Myohalovirus*. Примечательно, что это дерево поместило большую часть архейных последовательностей в ветки, примыкающие к веткам джамбо-фагов, инфицирующих грамположительные бактерии. Также интересно, что кластеризация указала на близость ЧБ диффоцина из *Peptoclostridium difficile* с чехольным белком фага *Clostridium* phiCDHM13 (род *Sherbrookevirus* семейства *Myoviridae*). Все белки из семейств *Herelleviridae* и *Chaseviridae* были кластеризованы в соответствии с таксономической принадлежностью, но два из восьми фагов *Ackermannviridae* группировались отдельно от остальных шести фагов *Ackermannviridae*. Возможно, это частично вызвано ошибками, связанными с проблемами предсказания мультидоменных структур и сложностями их структурного выравнивания. Пять из шести чехольных белков АФР группировались вместе в отдельную кладу, смежную с веткой, содержащей ЧБ фага *Bacillus* vB_VseM-HSE3 и гомологичные белки, обнаруженные в геномных последовательностях архей и бактерий, однако оставшийся чехольный белок АФР находился в другой ветке.

Филогенетический анализ также был выполнен на основе выравнивания аминокислотных последовательностей частей ЧБ, которые включали только консервативный домен. Всего для построения дерева, показанного на рис. 61, было использовано 90 последовательностей ЧБ. Эта дендрограмма продемонстрировала бóльшую согласованность с таксономией. Интересно, что это дерево тоже часто размещало фагов, инфицирующих грамположительные бактерии, в соседние ветки с фагами архей. Как и дерево, основанное на общем структурном сходстве, это дерево разделило архейные вирусы родов *Haloferacalesvirus* и *Myohalovirus*. Хотя это дерево тоже указывало на связь структурной архитектуры и таксономии, эта связь не являлась абсолютной. Например, количество доменов всех смоделированных ЧБ *Peduovirinae* было одинаковым и равнялось единице, а количество доменов мультидоменных чехольных белков *Ackermannviridae* отличалось для разных фагов. Важно отметить, что топология филогенетического дерева главных капсидных белков, построенного с использованием не структурного выравнивания аминокислотных последовательностей с помощью MAFFT (рис. 62), показывала сходный, но не идентичный состав клад, дерево также поместило архейные вирусы *Haloferacalesvirus* и *Myohalovirus* недалеко от фагов, инфицирующих грамположительные бактерии. Возможно, различия в топологии отражают как проблемы анализа сильно дивергировавших белков, так и результаты горизонтальных переносов. Филогенетический анализ, использующий аминокислотные последовательности большой субъединицы терминазы и белка хвостовой трубки (приложение Д, рис. Д.4, Д.5), показал аналогичную ситуацию с частичным сходством состава клад и топологией дерева, неидентичной топологии деревьев, полученных структурным выравниванием. Филогенетический анализ белка хвостовой трубки показал меньшую статистическую поддержку надёжности клад, что может быть связано с меньшим количеством сайтов для анализа в более коротких, по сравнению с другими использовавшимися последовательностями, последовательностях белков хвостовой трубки.

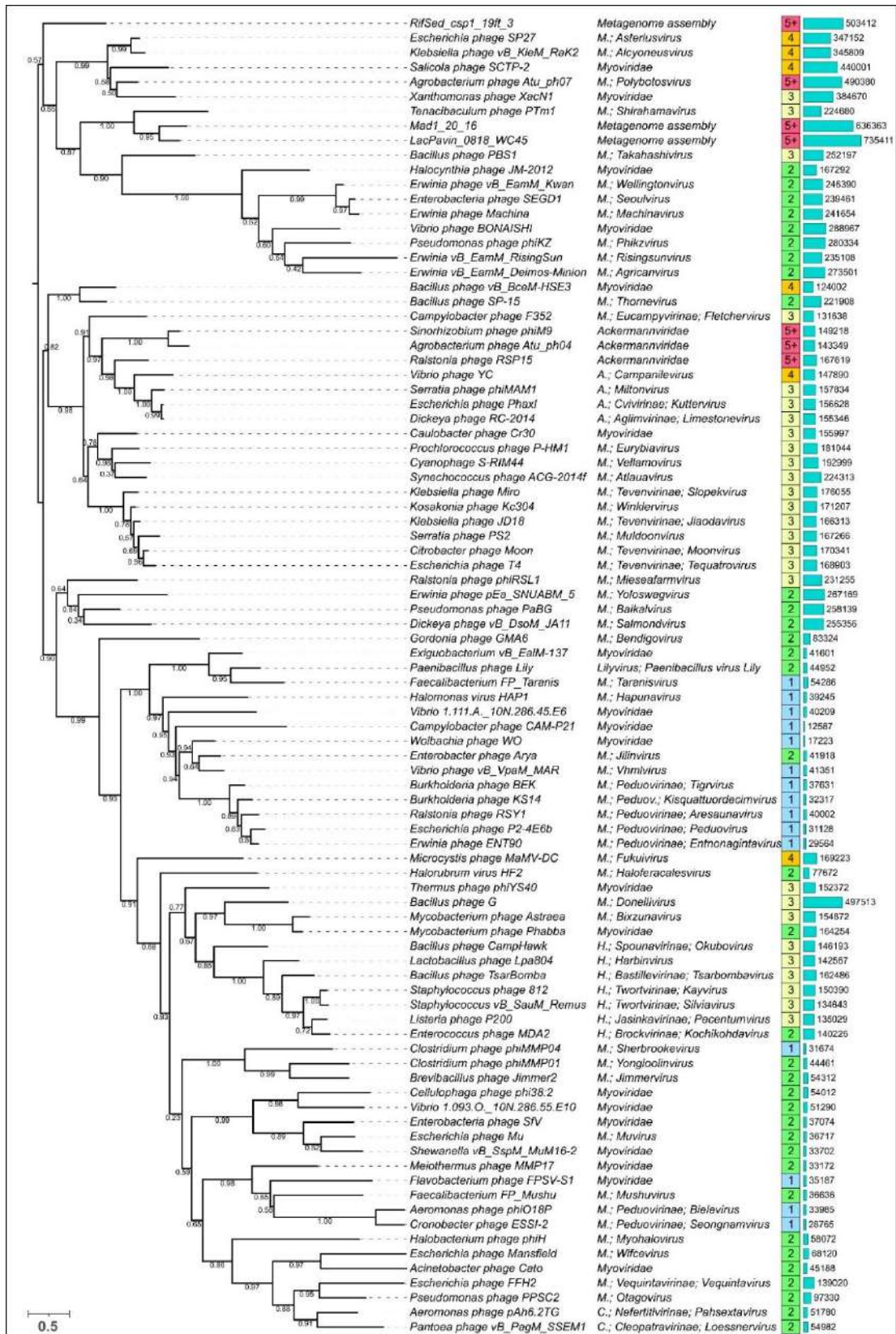


Рисунок 61. Филогенетическое дерево, полученное с использованием выравнивания 90 аминокислотных последовательностей чехольных белков с помощью mTM-align. Таксономия указана справа от названия, число доменов ЧБ обозначено цифрой в квадрате, размер генома в н.п. – гистограммой и числом справа.

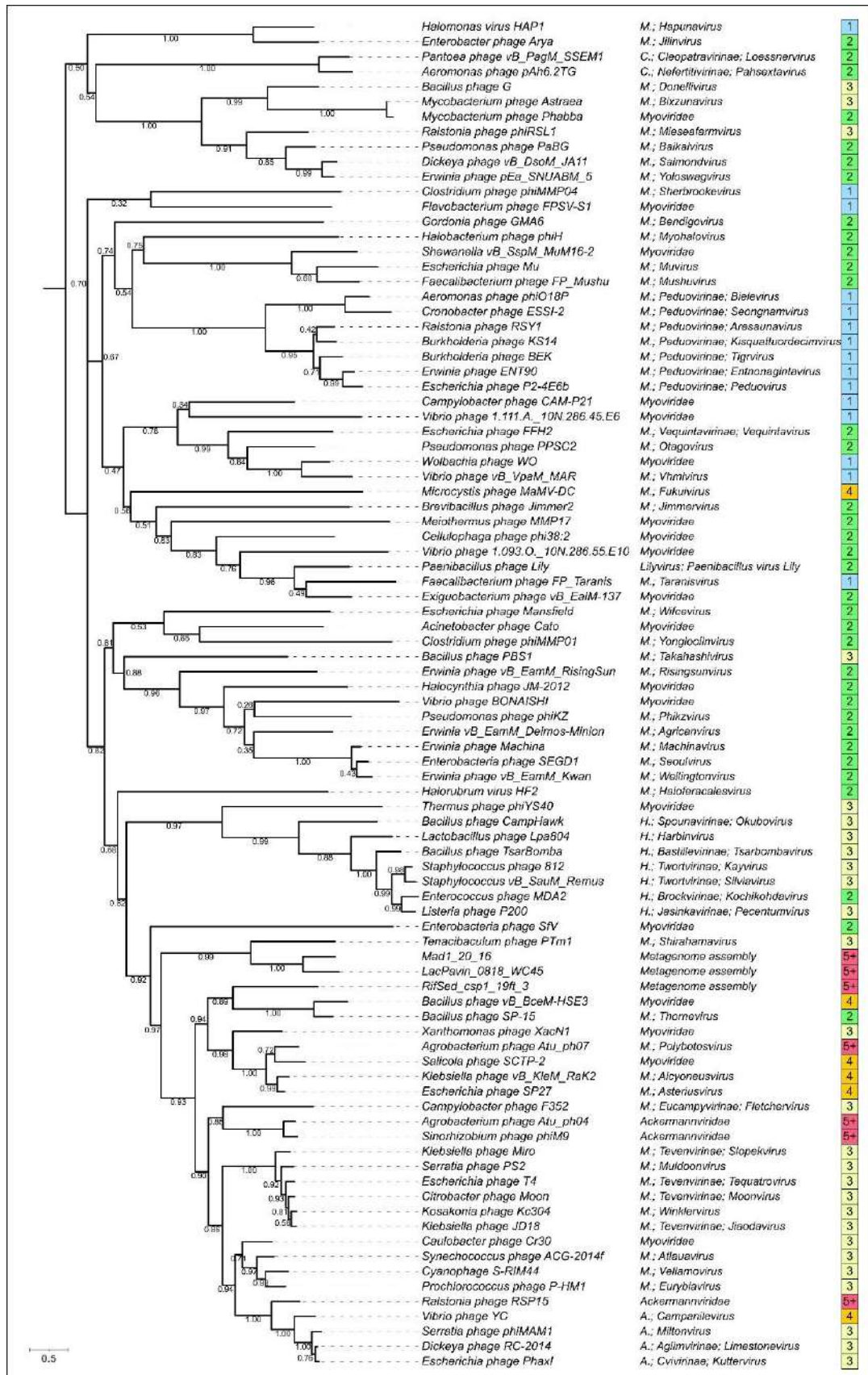


Рисунок 62. Филогенетическое дерево, полученное с использованием выравнивания 90 аминокислотных последовательностей главного капсидного белка, полученного с помощью MAFFT. Таксономия указана справа от названия, число доменов ЧБ обозначено цифрой в квадрате.

Важным выводом из результатов реконструкции и анализа эволюционной истории чехольных белков является сама возможность получения правдоподобных результатов, которые можно использовать для решения задач эволюционной вирусологии и таксономии, с помощью структурного моделирования белков методами глубокого обучения. Интересно, что сравнение структурного сходства моделей может указывать на родство чехольных белков Т6SS и джамбо-фагов, что может свидетельствовать о ранней дивергенции этих белков. Кроме того, сходство между чехольными белками системы АФР, грамположительных бактерий и предсказанными ЧБ провирусов архей может свидетельствовать о ранней дивергенции и специализации АФР. Возможно, чехольные белки бактериоцинов (диффоцинов и пиоцинов) дивергировали позже и полифилетичны, о чём может свидетельствовать то, что ЧБ диффоцина *Peptoclostridium difficile* структурно ближе к фагам *Clostridium*, чем ко всем другим чехольным белкам, а ЧБ пиоцина *Pseudomonas aeruginosa* структурно близок к ЧБ фагов, инфицирующих бета- и гаммапротеобактерии. Однако, более обоснованные выводы требуют дополнительных исследований.

Вероятное присутствие миовирусных последовательностей в архейных геномах не эвриархей может указывать на более широкое разнообразие хвостатых вирусов архей, чем предполагается в настоящее время. Присутствие нескольких генов чехольного белка в некоторых предположительно архейных геномных последовательностях может свидетельствовать об их провирусном происхождении. Возможное существование миовирусных провирусов в неэвриархейных геномах требует тщательного анализа. Кроме того, результаты филогенетического анализа могут свидетельствовать о полифилетичности чехольных белков архей, способствуя возникновению гипотез о возможности существования разных групп миовирусов до дивергенции основных архейных и бактериальных групп, либо о генных обменах с участием архейных вирусов и бактериофагов.

Видимая корреляция размера чехольных белков и количества дополнительных доменов с размером фагового генома может быть аргументом в пользу предположения о том, что предковая форма фаговых ЧБ имела один, основной домен, и в ходе дальнейшей эволюции, сопровождавшейся для ряда фагов увеличением размера генома, фаговые чехольные белки приобретали дополнительные домены. Это предположение кажется разумным, так как дополнительные домены, вероятно, не являются обязательными для сборки и работы сократительного механизма, но потребляют ресурсы для переноса дополнительного генетического материала и синтеза белка во время инфекции. Необходимость расходования дополнительных ресурсов должна быть оправдана конкурентными преимуществами, обеспечиваемыми дополнительными доменами. Многие дополнительные домены в исследованных ЧБ имели Ig-подобную укладку. Ранее было предположено, что наличие дополнительных Ig-подобных доменов может играть вспомогательную роль в фаговой инфекции, способствуя адсорбции с помощью слабого взаимодействия с углеводами на поверхности бактериальной клетки, а сами эти домены являются объектом горизонтальных переносов между различными классами как литических, так и умеренных фагов [240]. Также можно гипотетически предположить то, что наличие дополнительных доменов способствует повышению стабильности вириона, используя взаимодействия между дополнительными доменами.

Интересно, что число доменов часто является характерным признаком таксономической группы. При этом одинаковое количество доменов может быть присуще и эволюционно далёким группам. Поэтому структурная архитектура предсказанных чехольных белков хотя и может быть в некоторых случаях одним из таксономических маркеров, надёжная классификационная схема, очевидно, требует более продвинутого комплекса биоинформатических методов.

3.4.2. Использование структурных предсказаний для классификации таксонов высокого ранга

Важной задачей таксономической классификации вирусов реалма *Duplodnaviria*, в который входят хвостатые бактериофаги, хвостатые вирусы архей и герпесвирусы, является классификация высокоранговых таксонов, включая семейства и отряды. По состоянию на конец 2022 г. из 50 семейств реалма *Duplodnaviria* только 17 было отнесено к пяти утверждённым ICTV отрядам. Одна из основных проблем установления эволюционных взаимоотношений между вирусами – это высокий уровень дивергенции вирусных белков. Использование сравнения третичных структур белков часто способно выявить глубокие эволюционные взаимосвязи между белками, и появление новых высокоточных алгоритмов предсказания их структуры, в том числе с помощью программы глубокого обучения AlphaFold, могло бы способствовать выявлению эволюционных взаимосвязей между сильно дивергировавшими белками, на основании результатов структурного моделирования.

В ходе работы эволюционные отношения двух консервативных вирусных белков, главного капсидного белка и большой субъединицы терминазы, кодируемых геномами различных вирусов, включая представителей всех классифицированных семейств реалма *Duplodnaviria*, были проанализированы с учётом предсказаний их третичной структуры. Анализ проводили как с помощью структурных сравнений, использовавших предсказания AlphaFold, так и с помощью различных филогенетических методов.

На настоящий момент (начало 2023 года) классификация, утвержденная ICTV, относит все известные вирусы реалма *Duplodnaviria* к царству *Heunggongvirae*. Вирусы герпеса относят к типу *Peploviricota* и классу *Herviviricetes*, содержащему единственный отряд *Herpesvirales*, включающий все три классифицированных семейства герпесвирусов, а бактериофаги и вирусы хвостатых архей – к типу *Uroviricota* и классу *Caudoviricetes*. Однако

классификация внутри класса *Caudoviricetes* на уровне других таксономических рангов еще полностью не формализована и находится в стадии уточнения. В настоящее время класс *Caudoviricetes* включает 4 отряда, 47 семейств, 98 подсемейств, 1907 родов и 3301 вид. Четыре утверждённых отряда *Caudoviricetes* содержат 14 (из утверждённых 47) семейств прокариотических вирусов.

Структурное сходство широко используется для оценки эволюционных взаимосвязей между белками, уровень гомологии аминокислотных последовательностей которых невелик или, вообще, не может быть определён [141,241,242]. Структурное сходство между двумя белками можно оценить с помощью среднеквадратичного отклонения расстояния между атомными координатами наложенных белков (root-mean-square deviation, RMSD) или с помощью других метрик, таких как TM-score (template modeling score) и DALI Z-score, две последние метрики имеют ряд преимуществ перед RMSD (root-mean-square deviation) [177,243,244]. Ранее кластеризация экспериментально определенных структур главных капсидных белков с применением DALI Z-score уже использовалась для иллюстрации общего происхождения некоторых вирусных групп и для кластеризации прокариотических вирусов [141,245]. Однако для большинства семейств *Caudoviricetes* анализ с использованием экспериментально определенных структур невозможен, так как структуры консервативных белков для большинства вирусов экспериментально не определены.

Формирование представительской группы для исследования проводили следующим образом. Из списка классифицированных вирусов, опубликованного на официальном сайте ICTV (<https://ictv.global/taxonomy>), выбирали по одному представителю каждого семейства, отдавая предпочтение лучше описанным вирусам с полностью отсеквенированными и аннотированными геномами. Кроме того, в список для анализа (приложение E, табл. E.1) были дополнительно включены несколько бактериофагов, которые не были отнесены к утвержденным ICTV семействам, но являются важными (модельными) для вирусологии, а также другие бактериофаги, упомянутые в настоящей работе. В число дополнительных

бактериофагов вошли три джамбо-фага, которые могут представлять собой древние, рано дивергировавшие группы, два фага (λ и НК97), сыгравших важную роль в исследованиях вирусов, и два фага (*Curtobacterium* Аука и *Pseudomonas* MD8), которые не были классифицированы ранее. Джамбо-фаги были представлены первым выделенным джамбо-фагом, названным по новой номенклатуре *Phikzvirus* phiKZ (прежнее название «фаг *Pseudomonas* phiKZ») [246,247], фагом с самым большим известным геномом *Donellivirus* gee (фаг *Bacillus* G) [248] и фагом с самым большим известным геномом, обнаруженным с помощью метагеномного анализа, LacPavin_0818_WC45 [77].

Далее экспериментально определенные и биоинформатически предсказанные последовательности 57 генов, кодирующих главные капсидные белки, транслировали и моделировали с помощью AlphaFold 2 (AF). Для анализа брали структуры, наиболее высоко ранжированные программой. В качестве внешней группы для филогенетического анализа использовали структуру энкапсулина *Thermotoga maritima* (определена экспериментально, код PDB 7K5W), которая обладает заметным структурным сходством с ГКБ *Heunggongvirae* [206], структура которого была предсказана с помощью AF. Для оценки и сравнения качества предсказаний также использовали другую программу глубокого обучения, RoseTTAFold, с помощью которой предсказали структуры 54 ГКБ, принадлежащих вирусам, относящимся к типу *Uroviricota* и классу *Caudoviricetes* (бактериофаги и архейные вирусы) и энкапсулина.

Анализ доступных аннотаций вирусных геномов показал отсутствие аннотаций ГКБ для вирусов ряда классифицированных групп. Кроме того, проверка аннотаций с использованием сравнения НММ-мотивов выявила ошибки в аннотациях для представителей семейств *Duneviridae* и *Helgolandviridae*. Даже использование сравнения НММ-мотивов не позволило предположить обнаружить гены ГКБ в геномах архефага *Hacavirus* НСТV1 (*Haloarcula californiae* tailed virus 1, порядок *Thumleimavirales*, семейство *Soleiviridae*) и бактериофага *Phikzvirus* phiKZ. Функции предполагаемых главных капсидных белков этих вирусов

предсказывали моделированием третичных структур белков и сравнением с известными структурами (*Hacavirus* HCTV1), а также используя экспериментальные данные (*Phikzvirus* phiKZ).

Все модели (рис. 63) содержали домен с характерной укладкой НК97, впервые обнаруженной в ГКБ фага *Byrnievirus* НК97 (фаг *Escherichia* НК97), включая её консервативные элементы: А-домен (аксиальный домен), Р-домен (периферический домен), Е-петлю (расширенная петля) и N-плечо (N-концевое плечо). Большинство моделей содержали дополнительные элементы, обнаруженные в различных НК97-подобных капсидных белках, такие как G-петля (рис. 64). Предсказанные структуры часто содержали другие домены или субдомены, что можно объяснить присутствием в транслируемых последовательностях доменов протеаз и белков скаффолда. Суперимпозиция предсказанных с помощью AlphaFold структур ГКБ фагов НК97, λ , Т4 и Т7 и соответствующих экспериментально определенных структур 1OHG (фаг НК97, зрелый капсид) [249], 7SJ5 (фаг λ , мутант ГКБ в конформации перед сборкой капсида) [250], 5VF3 (фаг Т4, мутант ГКБ изометрическом капсиде) [251], 7VS5 (фаг Т4, ГКБ в вытянутом капсиде) [252] и 3J7V (фаг Т7, ГКБ в прокапсиде без ДНК) [225], показала значения RMSD 0,968 Å, 0,874 Å, 3,437 Å, 0,763 Å и 2,708 Å соответственно (рис. 65). Эти значения ниже соответствующих значений экспериментального разрешения (3,45 Å, 2,69 Å, 3,45 Å, 3,40 Å и 4,60 Å соответственно). Наиболее сложная структурная архитектура была обнаружена у моделей ГКБ герпесвирусов и джамбо-фага phiKZ. Интересно, что модель ГКБ фага *Pseudomonas* MD8 также имела сравнительно сложную архитектуру. Сравнение предсказанной структуры ГКП фага *Pseudomonas* MD8 и белков других фагов подтвердило предположение, что в этом фаге один ген кодирует главный капсидный белок, протеазу и белок скаффолда как один пропептид.

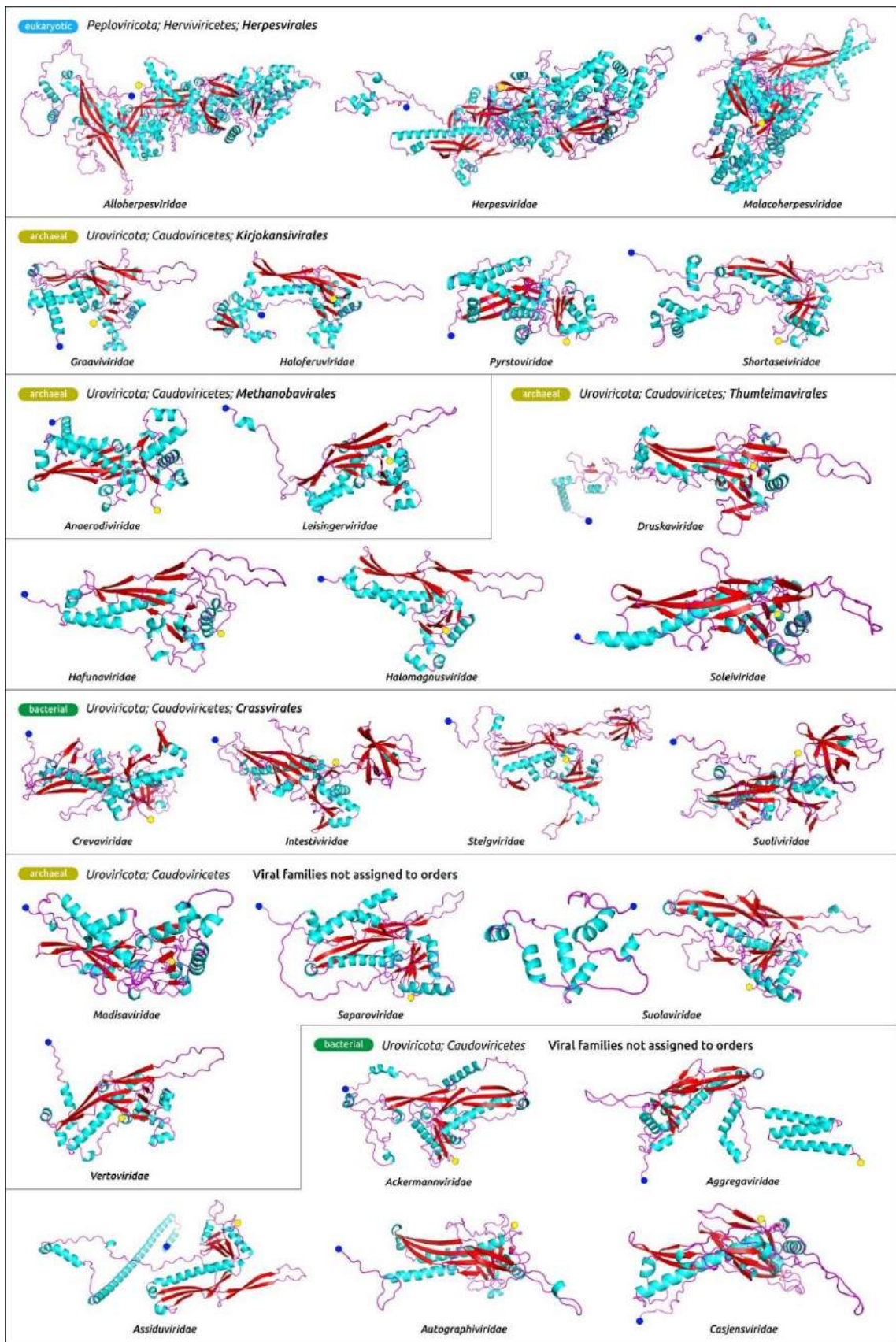


Рисунок 63. Структурные модели белков, кодируемых генами главного капсидного белка, полученные с помощью AlphaFold 2. N-конец полипептидной цепи отмечен синей меткой, а C-конец – жёлтой. Остальные модели показаны в приложении Е, рис. Е.1.

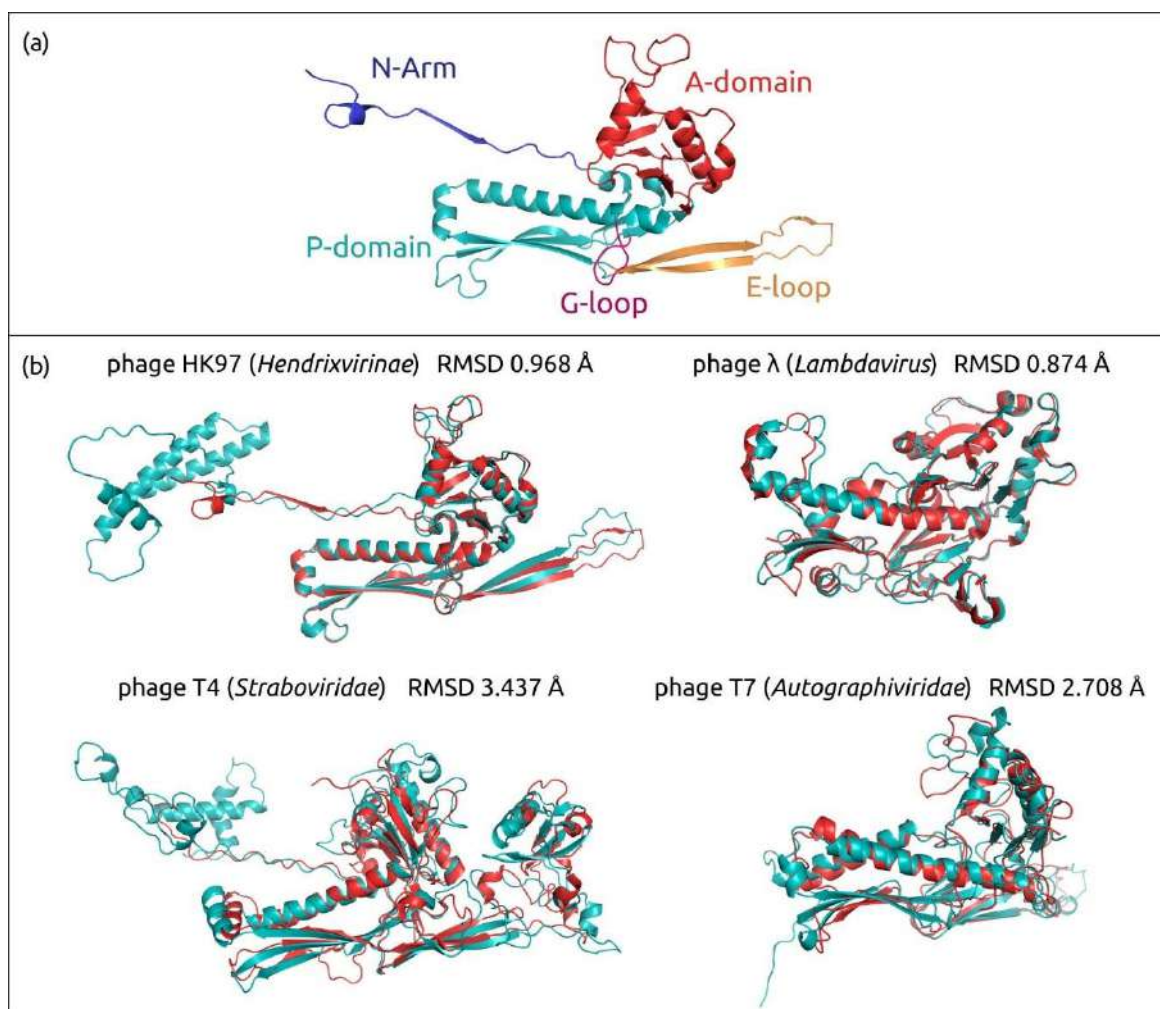


Рисунок 64. (а) Укладка НК97 и ее характерные особенности на примере структуры ГКБ зрелого капсида фага НК97 (код учётной записи PDB 1OHG_A). (b) Наложение предсказанных моделей ГКБ (изображены бирюзовым цветом) и соответствующих экспериментально определённых структур (изображены красным): фаг НК97, фаг λ, фаг Т4, фаг Т7.

Аналогичным образом моделировали большую субъединицу терминазы, содержащую АТФазный домен (далее – терминаза, TerL). Гены терминаз были извлечены из аннотаций репрезентативных геномов (приложение E, табл. E.1) и транслированы. Терминаза (ген IVa2) аденовируса Human adenovirus C, использующая механизм упаковки генома, сходный с герпесвирусами и хвостатыми фагами [253], также была смоделирована с помощью AF.

Структурная архитектура большой субъединицы терминазы *Duplodnaviria* отражает функцию этого белка. Типичные большие субъединицы терминазы

включают N-концевой аденозинтрифосфатазный (АТФазный) домен, который управляет транслокацией ДНК, и С-концевой эндонуклеазный домен, который расщепляет конкатемерный геном как при инициации, так и при завершении упаковки генома (рис. 65) [254]. АТФазный домен (АТDP) содержит пятицепочечный параллельный β -лист в канонической укладке ASCE, находящийся между несколькими α -спиралями, что легко распознаётся на моделях (рис. 66), а также дополнительные β -структуры, уникальные для вирусных терминаз [255].

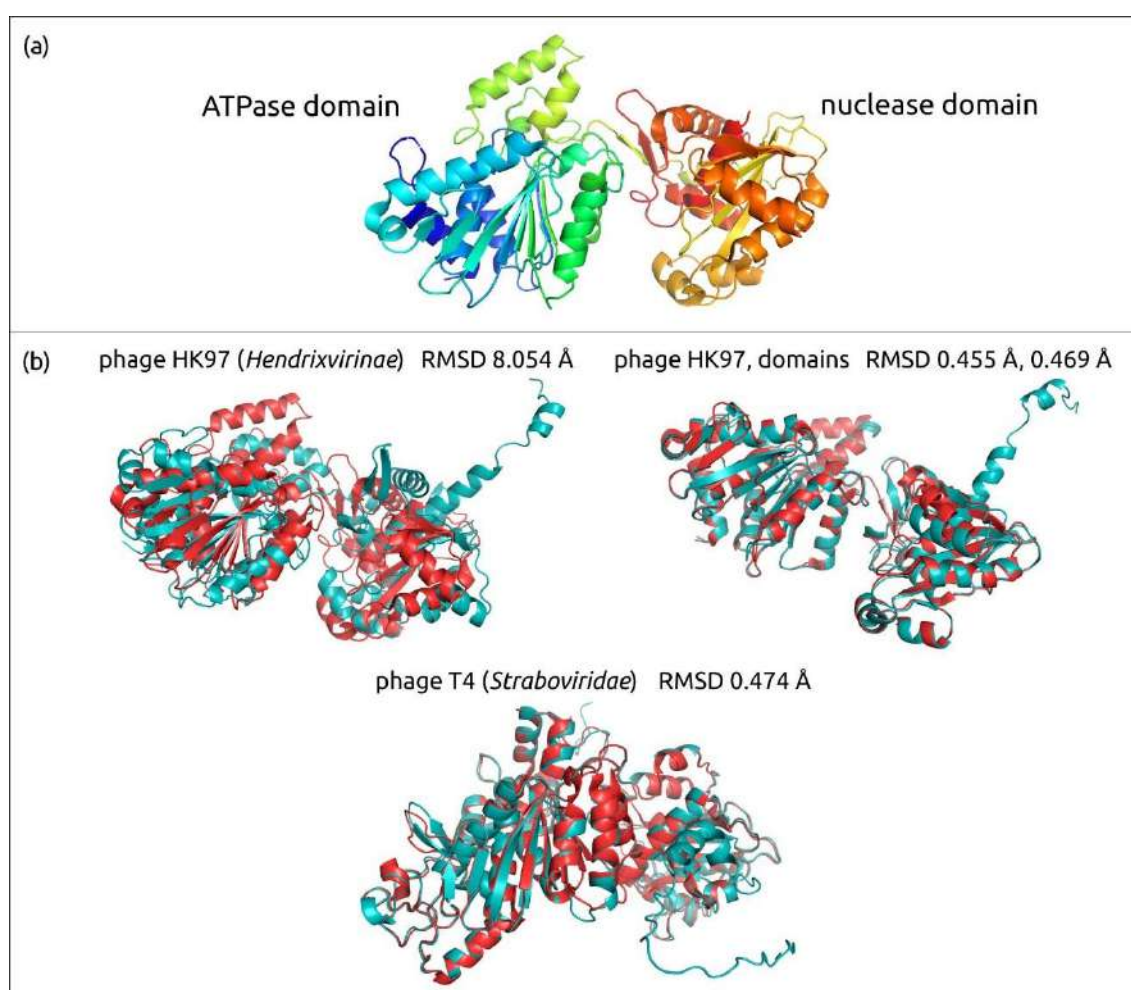


Рисунок 65. (а) Экспериментально определённая структура большой субъединицы терминазы фага НК97 (код учётной записи PDB 6Z6D). (б) Наложение предсказанных моделей терминазы (изображены бирюзовым цветом) и соответствующих экспериментально определённых структур (изображены красным) фага НК97 (целиком и подоменно) и фага Т4.

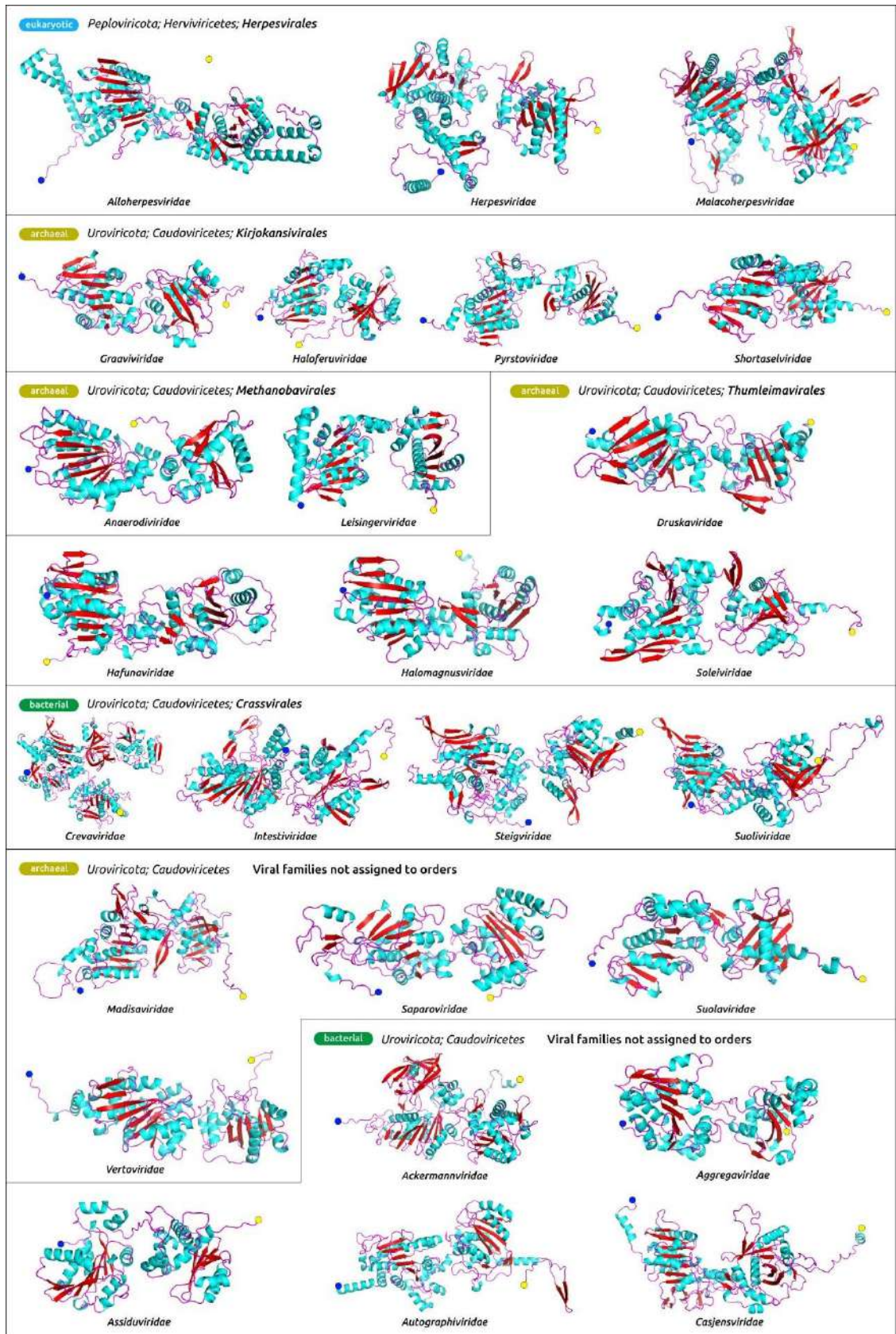


Рисунок 66. Структурные модели белков, кодируемых генами TerL, полученные с помощью AlphaFold 2. N-конец полипептидной цепи отмечен синей меткой, а C-конец – жёлтой. Остальные модели показаны в приложении Е, рис. Е.2.

Суперимпозиция модели TerL фага НК97 и соответствующей экспериментально определенной структуры 6Z6D дало значение RMSD 8,054 Å (экспериментальное разрешение 2,20 Å), а наложение модели TerL фага T4 и экспериментально определенной структуры 3CPE дало значение RMSD 0,474 Å (экспериментальное разрешение 2,80 Å). Анализ модели НК97 и экспериментальной структуры показал, что сравнительно высокое значение RMSD было связано с неверное предсказанной ориентацией АТФазного и нуклеазного доменов относительно друг друга. Наложение с использованием разделенных доменов без линкерной части дало значения RMSD 0,455 Å для АТФазного домена и 0,469 Å для нуклеазного домена.

Точность структурных предсказаний оценивались с помощью теста на разницу локальных расстояний (IDDT) с использованием предиктора точности DeepAccNet. Анализ средней оценки IDDT для 54 моделей главного капсидного белка и терминазы, предсказанных AlphaFold, показал в основном высокий уровень точности моделирования, а также то, что предсказание структуры TerL было более точным, чем для ГКБ (среднее значение IDDT TerL: 0,988, медиана: 0,996, q1: 0,991, q3: 0,999; среднее значение IDDT ГКБ: 0,907, медиана: 0,929, q1: 0,822, q3: 0,970). Среднее значение IDDT АТФазных доменов терминазы было еще выше (среднее: 0,998, медиана: 0,999, q1: 0,998, q3: 0,9997). Оценка моделей RoseTTAFold для тех же 54 ГКБ показала более низкую точность предсказаний (среднее значение IDDT: 0,634, медиана: 0,649, q1: 0,582, q3: 0,685), чем у моделей AlphaFold 2 (рис. 67).

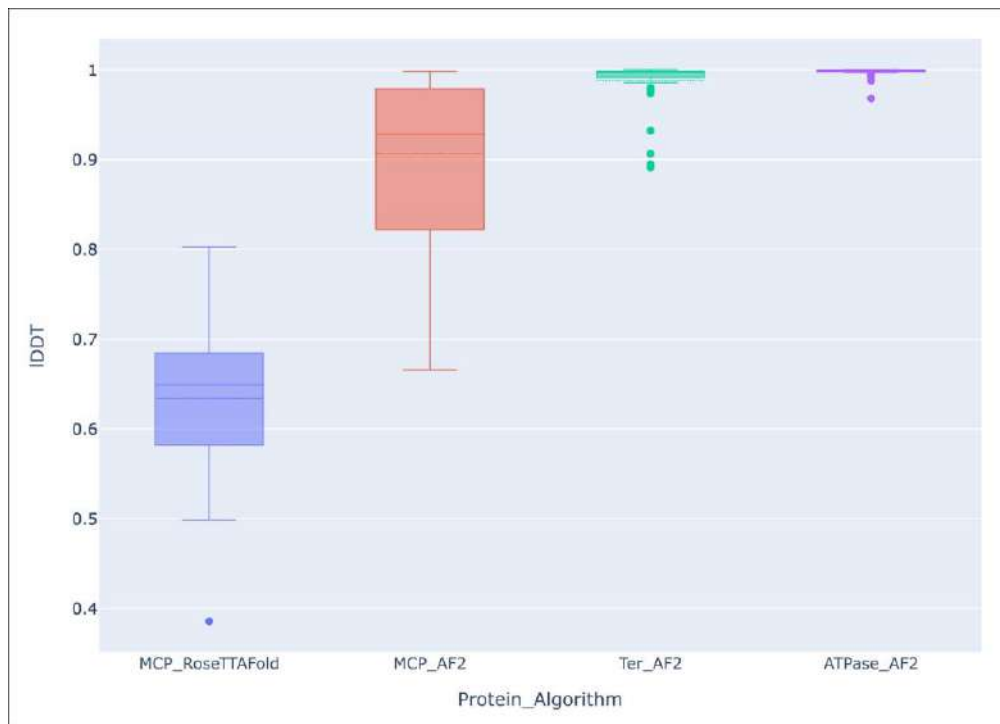
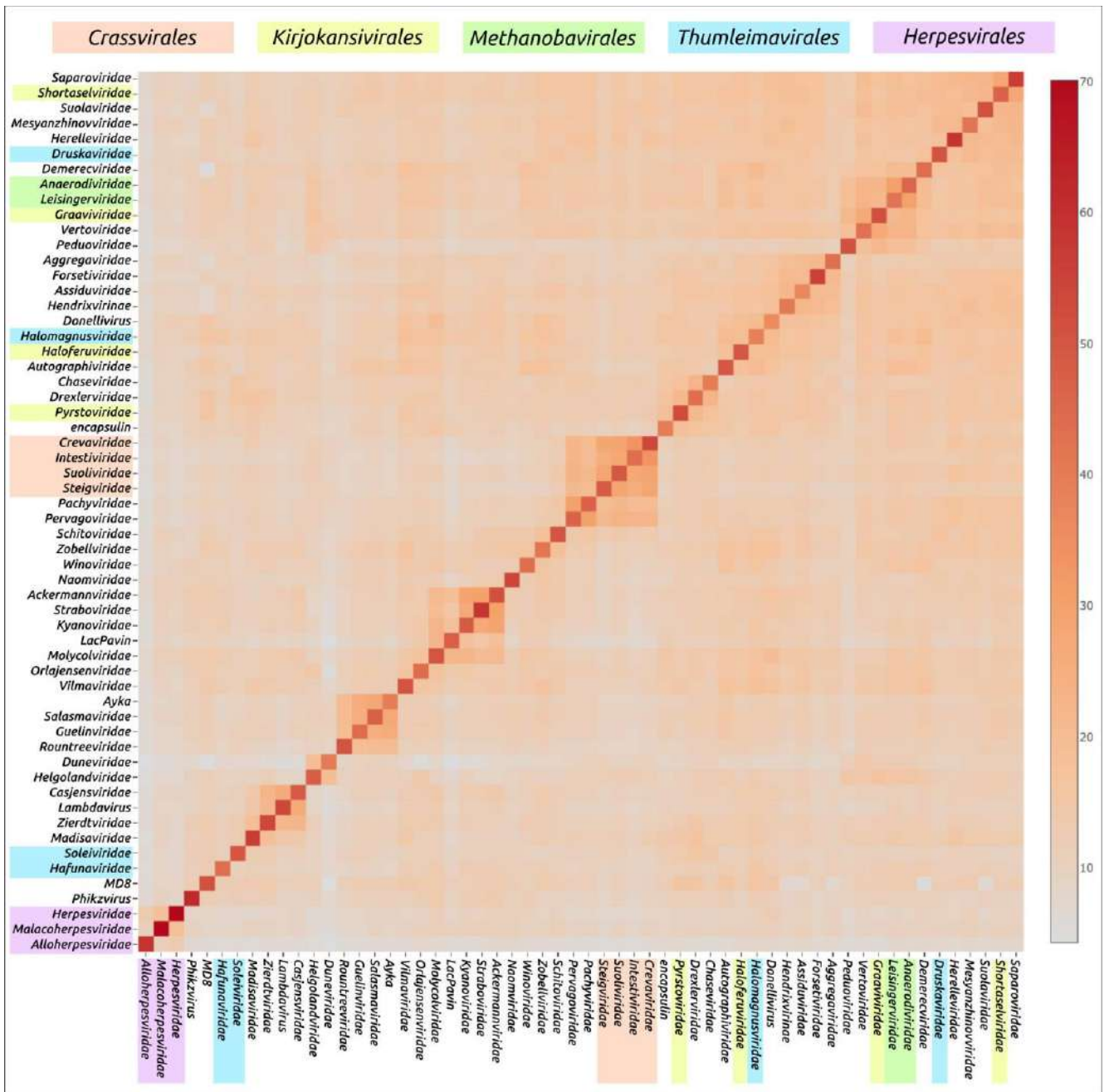


Рисунок 67. Оценка точности моделирования, сделанных с помощью теста на разницу локальных расстояний (IDDT). MCP_RoseTTAFold – оценка для моделей ГКБ, предсказанных RoseTTAFold, MCP_AF2 – оценка для моделей ГКБ, предсказанных AlphaFold 2, Ter_AF2 – оценка для моделей TerL, предсказанных AlphaFold 2, ATPase_AF2 – оценка для АТФазного домена терминазы, предсказанных с помощью AlphaFold 2.

Попарная оценка структурного сходства белков с помощью сервера DALI с использованием AF2-моделей главных капсидных белков (рис. 68) и последующей кластеризацией сгруппировала все четыре семейства бактериофагов порядка *Crassvirales* (*Crevaviridae*, *Intestiviridae*, *Steigviridae* и *Suoliviridae*) и обоих семейств вирусов архей отряда *Methanobavirales* (*Anaerodiviridae* и *Leisingerviridae*) в отдельные кластеры. В то же время ГКБ вирусов архей отрядов *Kirjokansivirales* и *Thumleimavirales* не образовывали отчетливых кластеров. Модели ГКБ представителей отряда *Herpesvirales* (семейства *Alloherpesviridae*, *Herpesviridae* и *Malacoherpesviridae*) были сгруппированы вместе, но не показали такой же степени сходства друг с другом, как ГКБ отрядов *Crassvirales* и *Methanobavirales*. Результаты кластеризации DALI с использованием структур с удаленными частями, примерно соответствующих доменам протеазы и белка

скаффолда, были практически идентичны результатам кластеризации полноразмерных моделей (приложение Е, рис. Е.3).



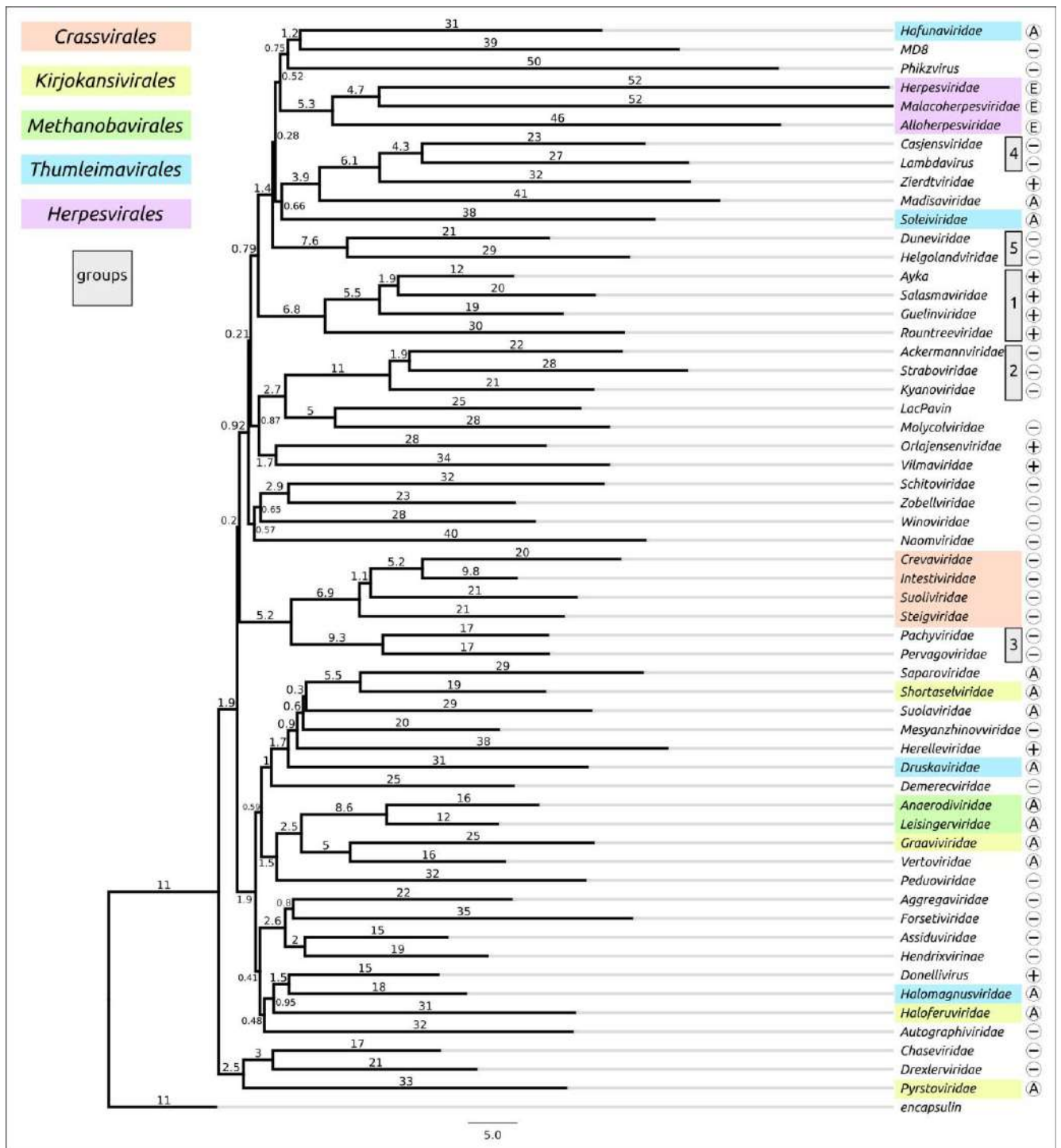


Рисунок 68. Кластерная тепловая карта и дендрограмма, основанные на попарном сравнении DALI Z-score 57 моделей главных капсидных белков и энкапсулина:
 «A» – вирусы архей; «E» – вирусы эукариот; «+» – фаги, инфицирующие грамположительные бактерии; «-» – фаги, инфицирующие грамотрицательные бактерии.

Кроме того, сравнение структурного сходства указало на заметное сходство моделей ГКБ других семейств, в том числе:

1. Бактериофагов семейств *Guelinviridae*, *Rountreeviridae* и *Salasmaviridae*, а также фага *Curtobacterium* Аука; (далее в настоящей работе объединены в группу 1);
2. Бактериофагов семейств *Ackermannviridae*, *Kyanoviridae* и *Straboviridae* (группа 2);
3. Бактериофагов семейств *Pachyviridae* и *Pervagoviridae* (группа 3), образующих субкластер более крупного кластера, в который входят фаги *Crassvirales*;
4. Бактериофагов семейства *Casjensviridae* и *Lambdavirus lambda* (группа 4);
5. Бактериофагов семейств *Duneviridae* и *Helgolandviridae* (группа 5).

В некоторых случаях отмеченное сходство может быть биологически осмысленным и связанным с общим происхождением и образом жизни вирусов. Группа 1 включает так называемые ф29-подобные литические фаги с подовирусной морфологией и сходным размером генома (около 16–20 тыс. н.п.), инфицирующие грамположительные бактерии [197,198,256,257]. Семейства *Guelinviridae*, *Rountreeviridae* и *Salasmaviridae* были образованы в 2020 г. для уточнения таксономической классификации ф29-подобных вирусов. Группа 2 включает фаги с большими геномами (около 150–200 тыс. н.п.), которые ранее были описаны как «Т4-подобные» фаги [258–260]. Интересно, что модели ГКБ неклассифицированного джамбо-фага LacPavin (размер генома 735 тыс. н.п.) и фага *Maribacter* Colly_1 (отряд *Molycolviridae*, размер генома 735 тыс. н.п.) также показали некоторое сходство с моделями группы 2.

Вирусы группы 3 [261,262] инфицируют флавобактерии (тип *Bacteroidota*) и имеют близкие по размеру (около 73 т.п.н.) геномы. Однако поиск BLAST с использованием базы данных NCBI GenBank PHG показал родство ГКБ репрезентативных фагов *Pachyviridae* и *Pervagoviridae*, но не выявил заметного

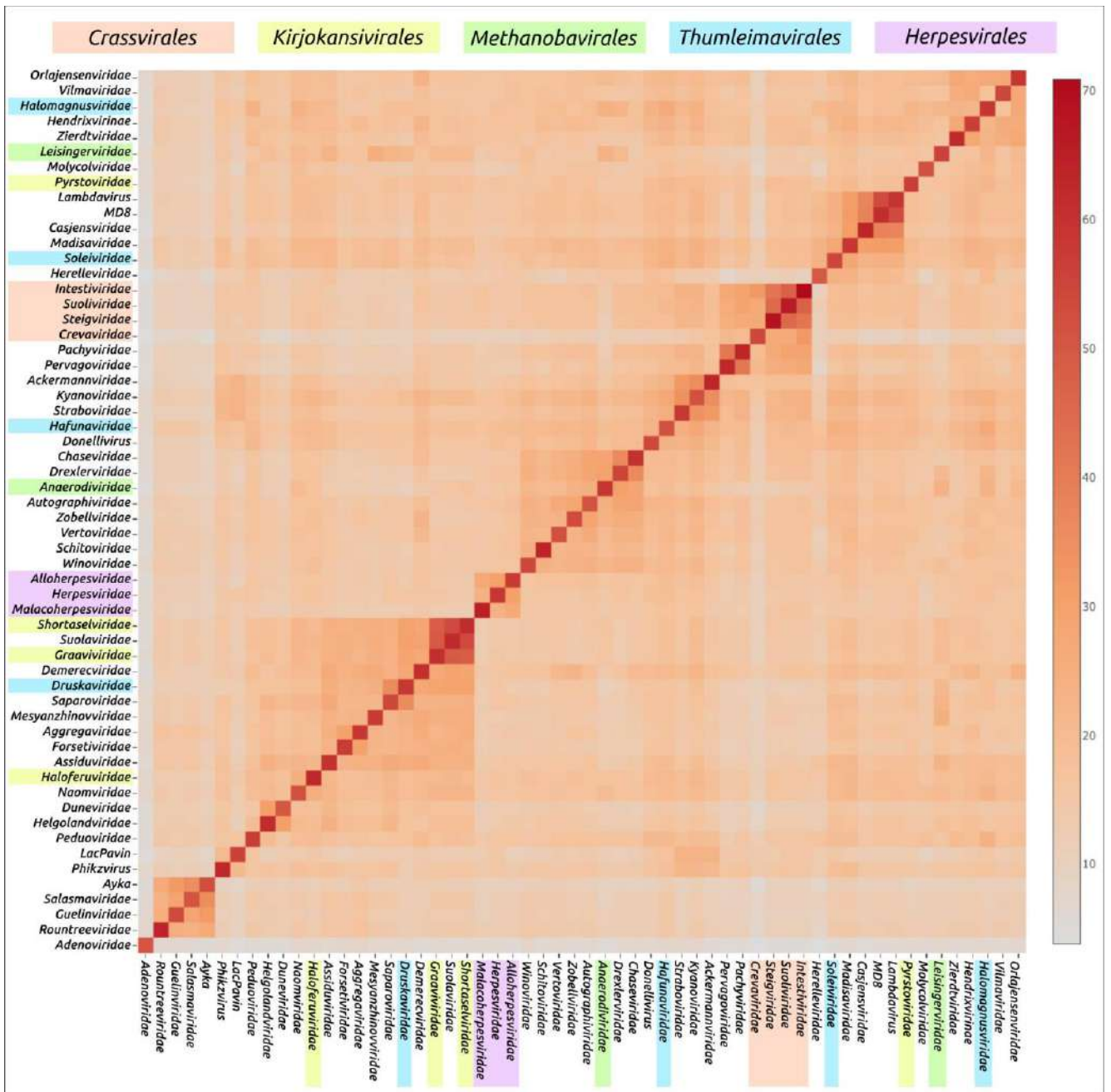
сходства с фагами *Crassvirales*, которые инфицируют кишечных симбионтов *Bacteroides* [76,90]. Как фаги группы 3, так и фаги *Crassvirales* характеризуются подовирусной морфологией.

Фаги группы 4 имеют геномы 49–67 тыс. н.п. и инфицируют представителей порядка *Enterobacterales* [34,263]. В группу 5 входят фаги *Flavobacterium* 1H (семейство *Duneviridae*) [264] и *Polaribacter* Leef_1 (семейство *Helgolandviridae*) [262], инфицирующие флавобактерии. Фаги имеют геномы близкого размера (около 38–39 тыс. н.п.) и показывают черты сифовирусной морфологии.

Примечательно, что, согласно сравнениям структурного сходства ГКБ с помощью DALI, представители различных семейств архейных хвостатых вирусов не образуют один или два отчетливых кластера. Они часто группируются с бактериофагами или не имеют сходства с какими-либо другими семействами.

Сравнения с помощью DALI, сделанные с использованием смоделированных структур большой субъединицы терминазы (TerL) и ее АТФазного домена (ATPD), продемонстрировали близкие друг с другом результаты (рис. 69; приложение E, рис. E.4). Однако оба этих сравнения показали картину, отличную от результатов сравнения моделей ГКБ, также сделанных с помощью DALI. В целом, структурное сходство репрезентативных моделей TerL и ATPD было больше, чем для ГКБ, что указывает на бóльшую консервативность терминазы. Структурные сравнения моделей терминазы также кластеризовали вирусы групп 1, 2 и 3, упомянутых выше, и показывали сходство моделей терминаз *Pachyviridae* и *Pervagoviridae* группы 3 с терминазами *Crassvirales*. Интересно, что модели терминазы фага λ (группа 4) и фага *Pseudomonas* MD8, отдаленно связанного с лямбдоидными фагами, фага *Salmonella* χ семейства *Casjensviridae* (группа 4) и архейного вируса *Haloarcula* hispanicatailed 1 (семейство *Madisaviridae*) показали видимое структурное сходство. Эти вирусы имеют сифовирусную морфологию и размер генома 48–59 тыс. н.п. Как и в случае сравнения структурного сходства с помощью DALI, анализ структурного сходства моделей терминазы показал сложный характер

ЭВОЛЮЦИОННЫХ ВЗАИМОТНОШЕНИЙ МЕЖДУ АРХЕЙНЫМИ ВИРУСАМИ, КОТОРЫЙ НЕ ВПОЛНЕ СООТВЕТСТВУЕТ СУЩЕСТВУЮЩЕЙ КЛАССИФИКАЦИИ ICTV.



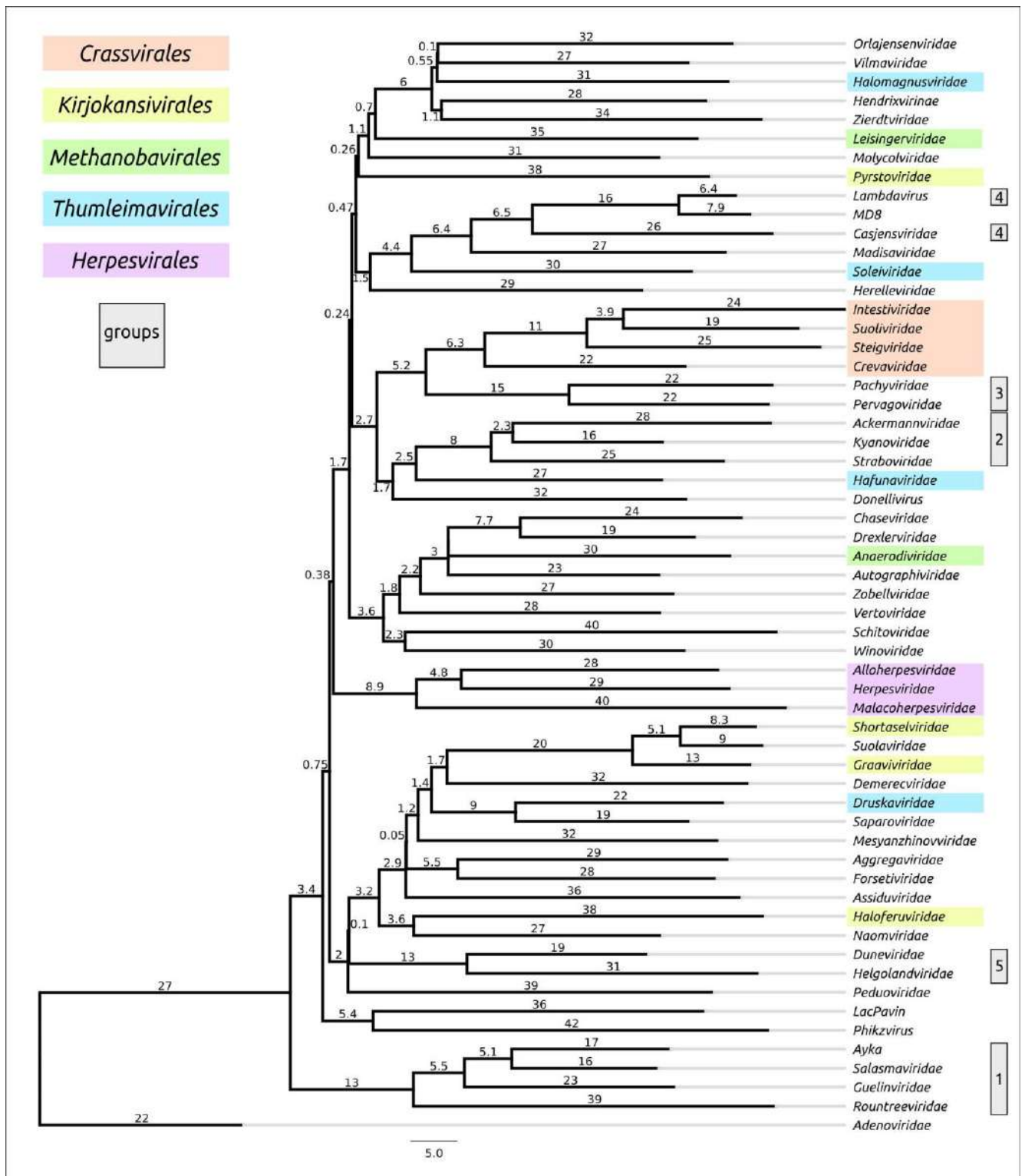


Рисунок 69. Кластерная тепловая карта и дендрограмма, основанные на попарном сравнении DALI Z-score 58 моделей большой субъединицы терминазы.

Результаты структурных сравнений с помощью mTM-align (приложение E, рис. E.5) не вполне совпали с выводами исследований DALI, что может

объясняться проблемами структурного выравнивания мультидоменных белков с помощью mTM-align. Однако некоторые особенности кластеризации оказались похожи. Например, главные капсидные белки вирусов группы 1, группы 2 и группы 3, кластеризуются вместе, как и терминазы. Сравнение с помощью mTM-align также показало сложные взаимосвязи между белками архейных вирусов. Как и дерево ГКБ, полученное с помощью DALI, дерево mTM-align поместило семейства порядка *Methanobavirales* в монофилетическую ветвь и показало сходство моделей ГКБ *Lambdavirus* и *Casjensviridae*, а также моделей *Hafunaviridae* и MD8. Дерево mTM-align, построенное с использованием полных моделей большой субъединицы терминазы (см. приложение E, рис. E.6), не группировало все три семейства отряда *Herpesvirales* в одну кладу, однако деревья ГКБ и ATPD поместили эти семейства в кладу (см. приложение E, рис. E.5, E.7). Последние упомянутые деревья, как и деревья DALI, поместили структуры *Pachyviridae* и *Pervagoviridae* в кладу, содержащую представителей *Crassvirales*.

В дополнение к сравнению структурного сходства был проведён филогенетический анализ, основанный на выравниваниях аминокислотных последовательностей главного капсидного белка с помощью алгоритмов Clustal Omega, MAFFT, MUSCLE и mTM-align (рис. 70 и приложение E, рис. E.8–E.10). Результаты анализа показали неконгруэнтность этих филогений (см. приложение E, рис. E.11), но при этом было также выявлено сходство состава ряда ветвей, особенно, дивергировавших позже. Из этих деревьев только дерево, построенное на основе выравнивания аминокислотных последовательностей с помощью структурного сравнения, полученное с использованием mTM-align, поместило всех представителей *Herpesvirales* в отдельную монофилетическую ветку. Кроме этого дерева, ни одно из деревьев не поместило представителей отряда *Crassvirales* и представителей группы 3 (семейства *Pachyviridae* и *Pervagoviridae*) в монофилетическую или в смежные ветви. За исключением дерева MUSCLE, ни одно из деревьев не поместило представителей группы 1 (φ29-подобные фаги семейств *Guelinviridae*, *Rountreeviridae* и *Salasmaviridae*, а также фаг

Curtobacterium Аука) в монофилетическую ветвь. Однако для последовательностей, принадлежащих к группе 2 (семейства T4-подобных фагов *Ackermannviridae*, *Kyanoviridae* и *Straboviridae*), а также последовательностей, принадлежащих к семействам порядка *Methanobavirales*, филогенетический анализ, основанный на различных выравниваниях аминокислотных последовательностей, показал сходство с результатами, полученными при структурных сравнениях. По-видимому, низкий уровень консервативности последовательности ГКБ (попарная идентичность MAFFT 6,0%) препятствует использованию филогенетического анализа, основанного на выравниваниях аминокислотных последовательностей без учёта структурного сходства, для надёжного выявления эволюционных взаимосвязей.

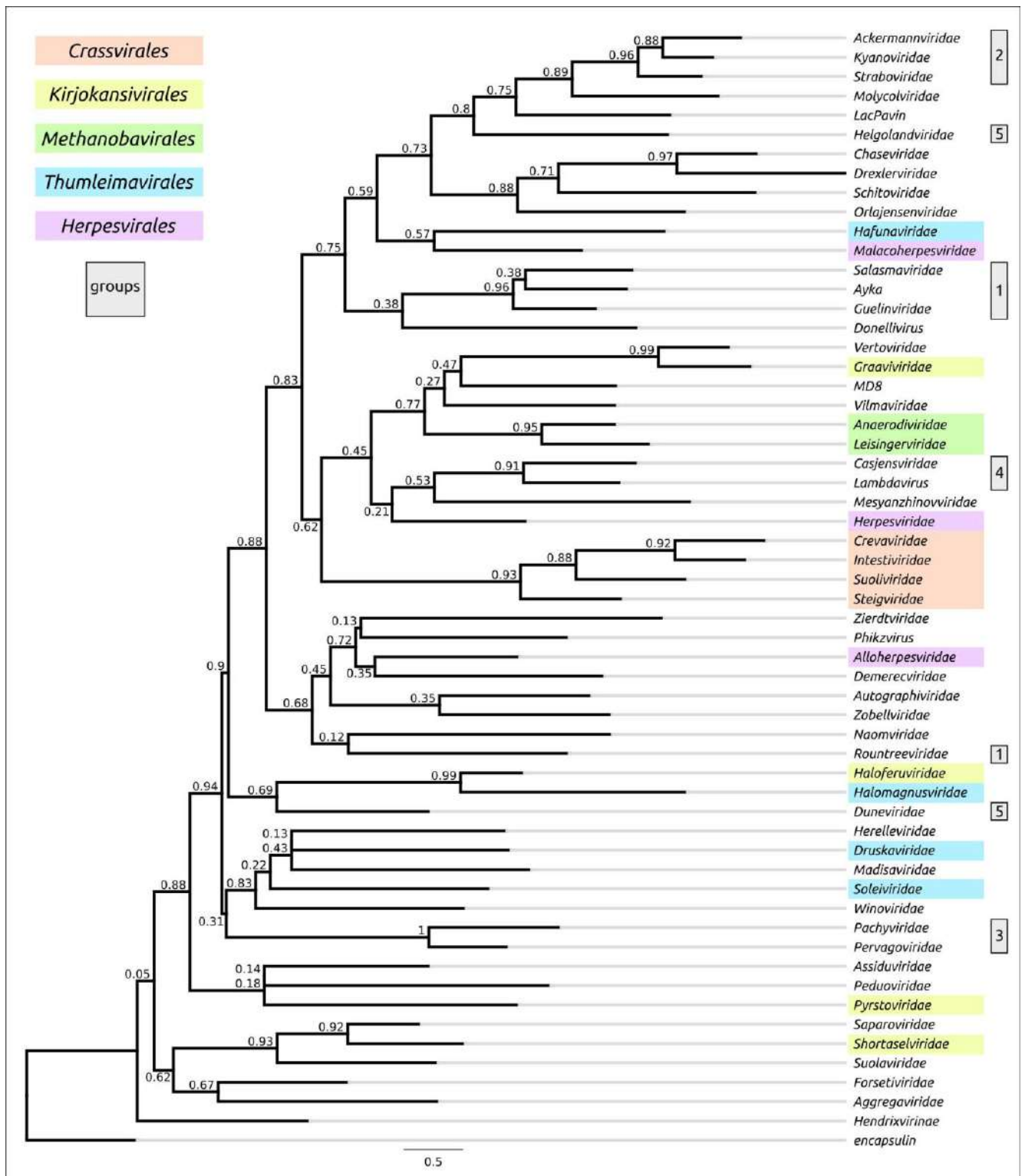


Рисунок 70. Филогенетическое дерево, построенное с использованием выравнивания 57 аминокислотных последовательностей главного капсидного белка и энкапсулина с помощью Clustal Omega. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

Филогенетический анализ, основанный на выравниваниях аминокислотных последовательностей большой субъединицы терминазы TerL и её АТФазного домена, был также проведён с использованием алгоритмов Clustal Omega, MAFFT, MUSCLE и mTM-align (рис. 71 и приложение E, рис. E.12–E.14). Уровень консервативности TerL был несколько выше, чем у ГКБ, попарная идентичность выравнивания TerL с помощью MAFFT составила 7,5%. За исключением деревьев MUSCLE, деревья сгруппировали представителей *Herpesvirales* и группы 2 в отдельные клады. Большинство деревьев расположило представителей как группы 3, так и отряда *Crassvirales* в монофилетическую ветвь, и все деревья разместили белки группы 1 (принадлежащие T4-подобным вирусам) в кладу. Однако, как и в описанных выше случаях, деревья показывали разную топологию. Как и остальные расчёты, включая структурное сравнение и филогенетический анализ, филогении терминаз показали сложные отношения между терминазами архейных хвостатых вирусов, которые не вполне соответствовали текущей классификации ICTV.

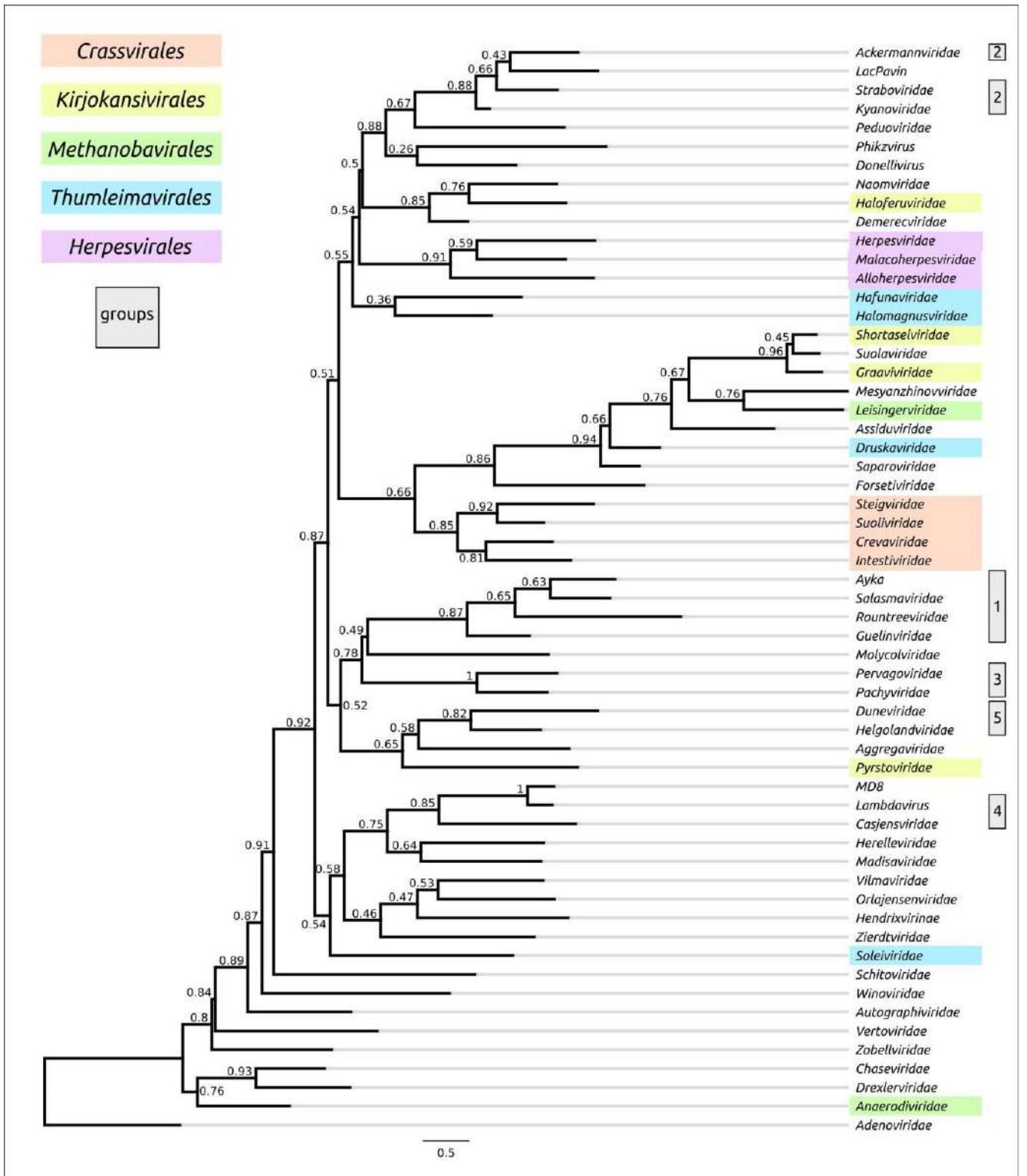


Рисунок 71. Филогенетическое дерево, построенное с использованием выравнивания 58 аминокислотных последовательностей терминазы с помощью Clustal Omega. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

Эволюционные отношения между вирусами также были оценены с помощью пайплайна GRAViTy, рекомендованного ICTV для разграничения высокоранговых таксонов [8] и классифицирующего вирусные группы в соответствии с гомологией между вирусными генами и сходством в организации геномов. Дерево GRAViTy (приложение E, рис. E.15) показывает отличия от дендрограммы структурного сходства DALI и других деревьев, объединяя два семейства герпесвирусов *Herpesvirales* вместе с четырьмя семействами архейных вирусов *Thumleimavirales*, и помещая представителей группы 3 (*Pachyviridae* и *Pervagoviridae*) и группы 4 (*Casjensviridae* и *Zierdtviridae*) в отдаленные группы. Следует отметить, что различия в топологиях также могут быть вызваны разным составом вирусов, участвующих в анализе. Дендрограмма GRAViTy сгруппировала всех представителей отряда *Crassvirales*, группы 1 (представители семейств *Guelinviridae*, *Rountreeviridae* и *Salasmaviridae*, а также фага *Curtobacterium* Айка), группы 2 (*Ackermannviridae*, *Kyanoviridae* и *Straboviridae*) и группы 5 (*Duneviridae* и *Helgolandviridae*) в отдельные кластеры, также содержащие другие вирусы, не представленные в списке 57 репрезентативных вирусов. Тем не менее, не все представители отрядов *Kirjokansivirales* и *Methanobavirales* были сгруппированы в соответствии с их таксономической классификацией. Интересно, что дендрограмма GRAViTy поместила представителей *Plasmaviridae* и *Helgolandviridae* в одну ветвь, хотя *Plasmaviridae* — семейство плеоморфных оболочечных вирусов, инфицирующих *Acholeplasma*, не принадлежащих к классу *Caudoviricetes* [93].

Результаты анализа эволюционных и таксономических взаимосвязей с использованием предсказаний структур главного капсидного белка и АТФазной субъединицы терминазы, продемонстрировали высокую точность предсказаний программой AlphaFold 2. Эта точность превышала точность RoseTTAFold, другого алгоритма глубокого обучения. Однако, используя только предсказанную точность, трудно судить о степени соответствия моделей реальным структурам. Следует также отметить, что нативное состояние вирусных белков может

различаться в зависимости от состояния вирусных частиц (например, пустые, полные, расширенные капсиды) и стадии их сборки [251,252,265,266]. Корреляция между структурным сходством и идентичностью последовательностей не является абсолютной из-за конформационной пластичности, эффектов растворителя и связывания лиганда [267]. Большинство этих ограничений относится и к исследованиям, в которых задействуются экспериментально определенные структуры, но, гипотетически, они могут усугубляться ошибками структурного предсказания. Поэтому прогнозирование эффективности использования AlphaFold для анализа структурного сходства и эволюционной истории, основываясь только на сходстве предсказанных структур, представляется непростой задачей. Кроме того, как показано в данном исследовании, разные алгоритмы структурного сравнения могут приводить к разным результатам.

К сожалению, различные ограничения присущи не только структурному, но и филогенетическому анализу, основанному на выравнивании аминокислотных последовательностей белков. Реконструкция эволюционной истории с помощью такого анализа затруднена различными проблемами, в том числе, связанными с высокой частотой мутаций, деталями молекулярной эволюции, неконгруэнтностью деревьев, полученных разными алгоритмами и методами [123,268–270], как это было видно из результатов приведённых выше филогенетических исследований. Представляется, что только комплексный подход, включающий оценку и сравнение различных методов, в том числе филогенетического анализа с использованием структур и последовательностей белков, организации генома, а также с учётом биологических данных, может привести к более обоснованным выводам.

Филогенетический анализ, основанный на аминокислотных последовательностях белков с применением одного и того же алгоритма построения деревьев RAxML-ng, привёл к разным топологиям деревьев при использовании различных алгоритмов выравнивания. Значения бутстрэп-поддержки также часто были низкими, что можно объяснить невысоким уровнем

сходства последовательностей из-за ранней дивергенции основных групп *Duplodnaviria* и большой скоростью мутаций. Уровень консервативности последовательности был выше для терминазы, чем для ГКБ. Также следует отметить, что из-за эффектов модульной эволюции, присущей различным вирусным группам [144,271], использование одного, пусть даже самого консервативного белка, не может раскрыть эволюционную историю группы. Анализ фага *Pseudomonas* MD8, рассмотренный в этой работе, показывает пример модульной эволюции, когда основные консервативные белки – ГКБ и терминаза – характеризуются разной эволюционной историей, что можно наблюдать в ходе сравнений предсказанных структурных моделей и филогенитических деревьев, построенных на основании выравнивания аминокислотных последовательностей, выполненных в последнем анализе.

Результаты большинства структурных сравнений и филогений, основанных на аминокислотных последовательностях, по-видимому, лучше соответствуют друг другу для вирусных групп, которые тесно связаны эволюционно либо дивергировали сравнительно недавно, таких как T4-подобные фаги. Как анализ структурного сходства, так и филогения на основе последовательностей, группировали T4- и φ29-подобные вирусы в монофилетические группы, и эти результаты кажутся биологически обоснованными. Следует также отметить, что более крупные белки эволюционно родственных представителей трех семейств *Herpesvirales* [272,273] были лучше кластеризованы с использованием алгоритмов, использующих структурное выравнивание предсказанных моделей. Это важный аргумент в пользу использования структурного сравнения, который может свидетельствовать о точности моделирования и эффективности использования алгоритмов, основанных на структурном сходстве, для анализа рано дивергировавших белков. Алгоритм анализа структурного сходства DALI и, в меньшей степени, mTM-align, показали результаты, схожие с результатами филогенетического анализа, задействующего аминокислотные последовательности, для относительно небольших эволюционных дистанций

(например, T4-подобные вирусные семейства), а также лучше кластеризовали значительно более отдаленных (если оценивать по длине соответствующих ветвей) представителей *Hepevirales*. Поэтому кажется обоснованным предположить, что структурные сравнения, основанные на предсказаниях AlphaFold, также могут быть достоверными и на промежуточных расстояниях. Вероятно, изучение изменения архитектуры структуры белков с использованием моделирования AlphaFold (например, появление новых доменов и субдоменов) может быть также полезным для эволюционного анализа, что было показано выше на примере анализа эволюционной истории чехольных белков. Например, в случае главного капсидного белка, выводы о родстве фагов *Crassvirales* и фагов группы 3 (представителей *Pachyviridae* и *Pervagoviridae*) также подтверждаются наличием сходного по структуре дополнительного домена.

Похоже, выбор единственного белка для выполнения реконструкции эволюционной истории вирусных групп и использование полученных результатов в целях таксономической классификации вряд ли может привести к обоснованным выводам, однако использование набора консервативных маркерных генов, имеющих общую эволюционную историю, может увеличить обоснованность эволюционных выводов для таксономии вирусов. Такой подход широко используется для различных групп ДНК-вирусов [274–276], хотя в случае с вирусами *Duplodnaviria* его применение осложняется ограниченным числом общих генов, а также событиями их рекомбинации в процессе модульной эволюции. Кроме этого, у многих фагов отсутствуют некоторые гены белков репликации [277], что делает невозможным их использование в филогенетическом анализе.

Использование в реконструкции эволюционной истории на уровне семейств инструментов кластеризации (таких как инструменты кластеризации на основе сходства протеома [8]), гипотетически может привести к ситуации, когда вновь приобретенные гены, подверженные частым генетическим обменам между разными группами вирусов, могут маскировать сильно мутировавшие консервативные гены, приводя к ошибочным выводам о происхождении этих

групп. Возможно, похожая ситуация привела к тому, что протеомное дерево группирует вирусы архей в более близкие группы, чем структурное сравнение консервативных белков и их филогения.

Вирусная эволюция характеризуется сетевым характером [135,278], и основанные на геномике оценки эволюционной истории с использованием кластеризации генных сетей хорошо показывают связи между родственными группами, однако, они не так очевидно раскрывают историю возникновения этих связей. Вероятно, для оценки эволюции *Duplodnaviria* необходимы какие-то маркеры и ориентиры, и они должны включать белки общего происхождения для всех представленных вирусов. Недавняя кропотливая работа по классификации архейных вирусов [245] заложила основу для классификации этой важной группы. Анализ всех доступных полных геномов архейных хвостатых вирусов позволил сделать предположение о древней дивергенции архейных и бактериальных хвостатых вирусов, а также интенсивном обмене генами, участвующими в метаболизме ДНК и работе механизмов, противодействующих защитным системам прокариотических хозяев. Результаты настоящего исследования, включая структурные сравнения и филогению на основе последовательностей, указывают на то, что такие предположения могут нуждаться в некоторых уточнениях, например, объясняющих немонафилетичность архейных ГКБ и терминаз. Тот факт, что простой поиск BLAST с использованием аминокислотных последовательностей ГКБ находит наиболее близкие родственные белки не среди представителей того же отряда архейных вирусов, а среди представителей других таксонов вирусов архей и бактерий, а также результаты наших филогенетических исследований, могут указывать на более сложную картину ранней эволюции вирусов *Caudoviricetes*.

По-видимому, теория и практика эволюционной таксономии вирусов *Duplodnaviria* нуждается в дальнейших уточнениях, которые должны быть основаны на решениях о приоритетах различных геномных данных, принципиальных решениях, имеющих несколько аспектов, включая философские

вопросы о соотношении холистического и редукционистского подходов в эволюционной биологии. К сожалению, на данный момент не все вопросы таксономии *Duplodnaviria* остаются отвеченными полностью.

Учитывая отсутствие экспериментальных данных о вирусных белках, предсказания AlphaFold могут быть полезны для реконструкции эволюции белков, делая его важным инструментом в эволюционной таксономии. Моделирование с помощью AlphaFold также можно использовать и для предсказания функций белков, когда обычно используемые поиски BLAST и HMM терпят неудачу, как было показано на примере представителей некоторых анализируемых групп.

Анализ данных, полученных в ходе этого исследования, включая предсказания AlphaFold и филогенетический анализ с использованием выравниваний аминокислотных последовательностей, позволил сформулировать следующие предположения относительно дальнейших исследований для уточнений возможности усовершенствования существующей таксономической классификации класса *Caudoviricetes*:

1. Бактериофаги семейств *Guelinviridae*, *Rountreeviridae* и *Salasmaviridae*, фаг *Curtobacterium* Аука (группа 1) и родственные фаги могут рассматриваться как кандидаты для выделения в новый отряд.
2. Семейства *Ackermannviridae*, *Kyanoviridae* и *Straboviridae* (группа 2) и родственные им фаги могут быть отнесены к новому таксону более высокого ранга.
3. Бактериофаги семейств *Pachyviridae* и *Pervagoviridae* (группа 3) родственны фагам *Crassvirales*. Эти и родственные им фаги могут рассматриваться как кандидаты на формирование нового отряда.
4. Бактериофаги семейства *Casjensviridae* и *Lambdavirus lambda* (группа 4) эволюционно родственны. Таксономия этих и родственных им групп требует дополнительных исследований и уточнений с учетом специфики эволюции

умеренных фагов, в значительной степени предрасположенных к генетическим обменам.

5. Бактериофаги семейств *Duneviridae* и *Helgolandviridae* (группа 5) эволюционно родственны и вместе с родственными фагами могут рассматриваться как кандидаты на выделение в новый отряд.

Результаты структурного сравнения и филогенетического анализа указывают на то, что эволюционная история и таксономическая классификация хвостатых вирусов архей также требуют дополнительных исследований и дальнейшего уточнения.

Результаты настоящей работы указывают на существование важных проблем и вопросов, требующих дополнительного осмысления. Это, в первую очередь, проблемы таксономической классификации бактериофагов с мозаичным геномом, а также необходимость совершенствования критериев и методов классификации таксонов высокого ранга. В целом, следует отметить, что развитие биоинформатических методов таксономической классификации бактериофагов не только позволило решить задачи, поставленные таксономией перед биоинформатикой, но и позволило обнаружить неизвестные ранее детали формирования вирусных геномов, поставив в свою очередь, новые важные задачи перед эволюционной вирусологией, систематикой и таксономией вирусов.

ЗАКЛЮЧЕНИЕ

В результате работы был впервые проведён таксономический и геномный анализ новых бактериофагов, инфицирующих грамположительные бактерии рода *Curtobacterium* (фаг Аука, геномные области профагового происхождения) и грамотрицательные бактерии рода *Pectobacterium* (Horatius, Possum, PP47, PP81, Q19). Проанализировано таксономическое разнообразие бактериофагов, инфицирующих бактерии семейства *Pectobacteriaceae*, вызывающие мягкую гниль сельскохозяйственных культур. На примере фага *Pseudomonas* MD8 впервые детально показан процесс мозаичного формирования геномов умеренных фагов псевдомонад, обсуждены трудности, возникающие при таксономической классификации, и предложено их возможное решение. С использованием предсказанных структур фаговых белков методами глубокого обучения проанализированы возможности улучшения описания функций фаговых белков и эволюционных взаимосвязей между бактериофагами.

Результаты диссертационной работы расширяют представления о таксономическом разнообразии бактериофагов, в том числе, фагов, инфицирующих бактерии, являющиеся опасными для человека и экономически важных сельскохозяйственных культур. Полученные в ходе работы аннотированные геномные последовательности депонированы в международную базу данных NCBI GenBank.

Предложенные подходы к биоинформатическому анализу бактериофагов и методы решения проблем, вызванных генетическим мозаицизмом и быстрой эволюцией вирусных белков, могут быть использованы для построения более точной таксономической иерархии.

ВЫВОДЫ

1. Таксономический анализ свидетельствует о принадлежности фагов *Pectobacterium* PP47, PP81, Q19 к семейству *Autographiviridae* и подсемейству *Studiervirinae*. Несмотря на схожую морфологию, спектр хозяев и организацию геномов, фаги PP47 и PP81, с одной стороны, и фаг Q19, с другой, относятся к разным близким таксономическим группам. Фаг Q19 должен быть отнесён к новому роду.
2. Фаги *Pectobacterium* Possum и Horatius являются представителями семейства *Schitoviridae* и рода *Cbunaviruses*.
3. Биоинформатический анализ указывает на высокое таксономическое разнообразие бактериофагов, инфицирующих бактерии рода *Pectobacterium* и другие фитопатогены, вызывающие болезнь мягкой гнили растений.
4. Фаг *Curtobacterium* Аука может быть классифицирован как представитель нового семейства или подсемейства.
5. Профаговые области геномов бактерий рода *Curtobacterium* представляют собой либо интактные профаги, родственные умеренным актинофагам, требующими классификации в новые таксоны ранга рода и выше, либо являются необратимо интегрированными нефункциональными профагами родственными умеренным актинофагам.
6. Геномный и филогенетический анализ указывают на высокую интенсивность горизонтальных переносов между различными умеренными фагами, инфицирующими *Pseudomonas aeruginosa*, вызвавшую ярко выраженный генетический мозаицизм фага MD8, что затрудняет его таксономическую классификацию. Высокий уровень генетического мозаицизма требует уточнения критериев для классификации подверженных ему бактериофагов.
7. Предсказание структуры белков с помощью новых алгоритмов глубокого обучения может способствовать построению эволюционно-биологически осмысленной иерархической классификации. Кластеризация и использование

структурного сходства таких моделей, вместе с другими данными, могут быть использованы для выявления глубоких эволюционных связей и построения классификационной системы таксонов высокого ранга.

СПИСОК ЛИТЕРАТУРЫ

1. Wommack K.E., Colwell R.R. Virioplankton: viruses in aquatic ecosystems // *Microbiol. Mol. Biol. Rev. MMBR.* 2000. Vol. 64, № 1. P. 69–114.
2. Rohwer F. et al. Life in Our Phage World: A Centennial Field Guide to the Earth's Most Diverse Inhabitants. Wholon, 2014. 408 p.
3. Simmonds P. et al. Consensus statement: Virus taxonomy in the age of metagenomics // *Nat. Rev. Microbiol.* 2017. Vol. 15, № 3. P. 161–168.
4. Hendrix R.W. et al. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 1999. Vol. 96, № 5. P. 2192–2197.
5. Shkoporov A.N., Hill C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome // *Cell Host Microbe.* 2019. Vol. 25, № 2. P. 195–209.
6. Loc-Carrillo C., Abedon S.T. Pros and cons of phage therapy // *Bacteriophage.* 2011. Vol. 1, № 2. P. 111–114.
7. Ackermann H.W. Bacteriophages // *Virologie.* 2004. Vol. 8, № 6. P. 409–412.
8. Turner D., Kropinski A.M., Adriaenssens E.M. A Roadmap for Genome-Based Phage Taxonomy // *Viruses.* 2021. Vol. 13, № 3. P. 506.
9. Ackermann H.W., Eisenstark A. The present state of phage taxonomy // *Intervirology.* 1974. Vol. 3, № 4. P. 201–219.
10. Bradley D.E. Ultrastructure of bacteriophage and bacteriocins. // *Bacteriol. Rev.* 1967. Vol. 31, № 4. P. 230–314.
11. Baltimore D. Expression of animal virus genomes. // *Bacteriol. Rev.* 1971. Vol. 35, № 3. P. 235–241.
12. Walker P.J. et al. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021) // *Arch. Virol.* 2021. Vol. 166, № 9. P. 2633–2648.
13. Modrow S. et al. Viruses: Definition, Structure, Classification // *Mol. Virol.* 2013. P. 17–30.
14. Lustig A., Levine A.J. One hundred years of virology. // *J. Virol.* 1992. Vol. 66, № 8. P. 4629–4631.

15. Artenstein A.W. The discovery of viruses: advancing science and medicine by challenging dogma // *Int. J. Infect. Dis.* 2012. Vol. 16, № 7. P. e470–e473.
16. Creager A.N.H. *The Life of a Virus: Tobacco Mosaic Virus as an Experimental Model, 1930-1965*. 1st edition. Chicago: University of Chicago Press, 2001. 352 p.
17. Жданов, Виктор Михайлович. *Эволюция вирусов. Медицина. Москва, 1990.* 376 p.
18. Stanley W.M., Loring H.S. The Isolation of Crystalline Tobacco Mosaic Virus Protein from Diseased Tomato Plants // *Science*. 1936. Vol. 83, № 2143. P. 85.
19. Creager A.N.H., Morgan G.J. After the double helix: Rosalind Franklin’s research on Tobacco mosaic virus // *Isis Int. Rev. Devoted Hist. Sci. Its Cult. Influ.* 2008. Vol. 99, № 2. P. 239–272.
20. Aksyuk A.A., Rossmann M.G. Bacteriophage Assembly // *Viruses*. 2011. Vol. 3, № 3. P. 172–203.
21. Sciortino M.-T. et al. RNAs Extracted from Herpes Simplex Virus 1 Virions: Apparent Selectivity of Viral but Not Cellular RNAs Packaged in Virions // *J. Virol.* 2001. Vol. 75, № 17. P. 8105–8116.
22. Nisole S., Saïb A. Early steps of retrovirus replicative cycle // *Retrovirology*. 2004. Vol. 1, № 1. P. 9.
23. Tobacco Mosaic Virus by Dennis Kunkel Microscopy/science Photo Library [Electronic resource] // Fine Art America. URL: <https://fineartamerica.com/featured/3-tobacco-mosaic-virus-dennis-kunkel-microscopyscience-photo-library.html> (accessed: 13.04.2022).
24. Saini H.K., Fischer D. Structural and functional insights into Mimivirus ORFans // *BMC Genomics*. 2007. Vol. 8, № 1. P. 115.
25. Yuan Y., Gao M. Jumbo Bacteriophages: An Overview // *Front. Microbiol.* 2017. Vol. 8.
26. Sanz-Gaitero M., Seoane-Blanco M., van Raaij M.J. Structure and Function of Bacteriophages // *Bacteriophages: Biology, Technology, Therapy* / ed. Harper D.R. et al. Cham: Springer International Publishing, 2019. P. 1–73.

27. Aksyuk A.A. et al. The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria // *EMBO J.* 2009. Vol. 28, № 7. P. 821–829.
28. Michalakis Y., Blanc S. The Curious Strategy of Multipartite Viruses // *Annu. Rev. Virol.* 2020. Vol. 7, № 1. P. 203–218.
29. Sharma V. et al. Mimivirus inaugurated in the 21st century the beginning of a reclassification of viruses // *Curr. Opin. Microbiol.* 2016. Vol. 31. P. 16–24.
30. Miroshnikov K.A. et al. Tailed Lytic Bacteriophages of Soft Rot Pectobacteriaceae // *Microorganisms.* 2021. Vol. 9, № 9. P. 1819.
31. McKenna R. et al. Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications // *J. Mol. Biol.* 1996. Vol. 256, № 4. P. 736–750.
32. Home - Genome - NCBI [Electronic resource]. URL: <https://www.ncbi.nlm.nih.gov/genome> (accessed: 11.11.2021).
33. Hay I.D., Lithgow T. Filamentous phages: masters of a microbial sharing economy // *EMBO Rep.* 2019. Vol. 20, № 6. P. e47427.
34. Lederberg E.M., Lederberg J. Genetic Studies of Lysogenicity in *Escherichia Coli* // *Genetics.* 1953. Vol. 38, № 1. P. 51–64.
35. Kutter E., Sulakvelidze A. *Bacteriophages: Biology and Applications.* CRC Press, 2004. 527 p.
36. Evseev P. et al. Pseudomonas Phage PaBG—A Jumbo Member of an Old Parasite Family: 7 // *Viruses. Multidisciplinary Digital Publishing Institute*, 2020. Vol. 12, № 7. P. 721.
37. Hankin ME. The bactericidal action of the waters of the Jamuna and Ganges rivers on Cholera microbes. *Ann. Inst. Pasteur* 10:511–523 (1896) // *Bacteriophage. Taylor & Francis*, 2011. Vol. 1, № 3. P. 117–126.
38. Twort F.W. An Investigation on the Nature of Ultra-Microscopic Viruses // *The Lancet. Elsevier*, 1915. Vol. 186, № 4814. P. 1241–1243.
39. D’Herelle F. On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D’Herelle, presented by Mr. Roux. 1917 // *Res. Microbiol.* 2007. Vol. 158, № 7. P. 553–554.

40. Wittebole X., De Roock S., Opal S.M. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens // *Virulence*. 2014. Vol. 5, № 1. P. 226–235.
41. Summers W.C. In the beginning... // *Bacteriophage*. 2011. Vol. 1, № 1. P. 50–51.
42. Essais de thérapeutique au moyen du bacteriophage – ScienceOpen [Electronic resource]. URL: <https://www.scienceopen.com/document?vid=f9178fff-aba9-440f-a4dd-1316136e86a7> (accessed: 10.04.2022).
43. d’Herelle F. Bacteriophage as a Treatment in Acute Medical and Surgical Infections // *Bull. N. Y. Acad. Med.* 1931. Vol. 7, № 5. P. 329–348.
44. Доклад НТС Гамбург: Наука и образование в России накануне и после 1917 года [Electronic resource]. URL: <http://nts-hamburg.de/vortraginfo.php?v=279> (accessed: 11.04.2022).
45. Фаги атакуют [Electronic resource] // Наука из первых рук. URL: <http://scfh.ru/papers/fagi-atakuyut/> (accessed: 11.04.2022).
46. Летаров А. Современные концепции биологии бактериофагов. Москва: ДеЛи, 2019. 384 p.
47. Myelnikov D. An Alternative Cure: The Adoption and Survival of Bacteriophage Therapy in the USSR, 1922–1955 // *J. Hist. Med. Allied Sci.* 2018. Vol. 73, № 4. P. 385–411.
48. Reuter M., Kruger D.H. Approaches to optimize therapeutic bacteriophage and bacteriophage-derived products to combat bacterial infections // *Virus Genes*. 2020. Vol. 56, № 2. P. 136–149.
49. Morozova V.V., Vlassov V.V., Tikunova N.V. Applications of Bacteriophages in the Treatment of Localized Infections in Humans // *Front. Microbiol.* 2018. Vol. 9.
50. Summers W.C. The strange history of phage therapy // *Bacteriophage*. 2012. Vol. 2, № 2. P. 130–133.
51. Fleming A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae // *Br. J. Exp. Pathol.* 1929. Vol. 10, № 3. P. 226–236.

52. Hyman P., Abedon S.T. Bacteriophage host range and bacterial resistance // *Adv. Appl. Microbiol.* 2010. Vol. 70. P. 217–248.
53. Krylov V.N. [Phagotherapy in terms of bacteriophage genetics: hopes, perspectives, safety, limitations] // *Genetika.* 2001. Vol. 37, № 7. P. 869–887.
54. Linné C. von, Salvius L. *Caroli Linnaei...Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* Holmiae : Impensis Direct. Laurentii Salvii, 1758. Vol. 1. P. 1–881881 p.
55. Lamarck J.-B. English: Branching diagram depicting two series of animal origins, from the *Histoire naturelle des animaux sans vertèbres* of Jean-Baptiste Lamarck (1815). 2013.
56. Schuh R.T., Brower A.V.Z. *Biological Systematics: Principles and Applications* // *Biological Systematics.* Cornell University Press, 2011.
57. Godfray H.C.J., Knapp S. Introduction. Taxonomy for the twenty-first century. // *Philos. Trans. R. Soc. B Biol. Sci.* 2004. Vol. 359, № 1444. P. 559–569.
58. Paterlini M. There shall be order. The legacy of Linnaeus in the age of molecular biology // *EMBO Rep.* 2007. Vol. 8, № 9. P. 814–816.
59. Phylogeny [Electronic resource] // *Biology Articles, Tutorials & Dictionary Online.* 2019. URL: <https://www.biologyonline.com/dictionary/phylogeny> (accessed: 12.04.2022).
60. Introduction to Cladistics [Electronic resource]. URL: <https://ucmp.berkeley.edu/clad/clad1.html> (accessed: 12.04.2022).
61. Шаталкин А. Таксономия. Основания, принципы и правила. Москва: Товарищество научных изданий КМК, 2012. 601 p.
62. Lipscomb D.L. Methods of systematic analysis: the relative superiority of phylogenetic systematics // *Orig. Life.* 1984. Vol. 13, № 3–4. P. 235–248.
63. Bennett C.W. The nomenclature of plant viruses // *Phytopathology.* 1939. Vol. 29. P. 422–430.
64. Bawden F.C. *Nomina ad infinitum* // *Chron Bot.* 1941. Vol. 6. P. 385–390.

65. Lwoff A. The concept of virus // *J. Gen. Microbiol.* 1957. Vol. 17, № 2. P. 239–253.
66. Lwoff A. Lysogeny // *Bacteriol. Rev.* 1953. Vol. 17, № 4. P. 269–337.
67. Lwoff A., Horne R., Tournier P. A system of viruses // *Cold Spring Harb. Symp. Quant. Biol.* 1962. Vol. 27. P. 51–55.
68. Kuhn J.H. Virus Taxonomy // *Encycl. Virol.* 2021. P. 28–37.
69. Koonin E.V. et al. Global Organization and Proposed Megataxonomy of the Virus World // *Microbiol. Mol. Biol. Rev. MMBR.* 2020. Vol. 84, № 2. P. e00061-19.
70. Wolf Y.I. et al. Origins and Evolution of the Global RNA Virome // *mBio.* 2018. Vol. 9, № 6. P. e02329-18.
71. Taxonomy [Electronic resource]. URL: <https://talk.ictvonline.org/taxonomy/> (accessed: 15.03.2022).
72. Gorbalenya A.E. et al. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks: 5 // *Nat. Microbiol.* Nature Publishing Group, 2020. Vol. 5, № 5. P. 668–674.
73. Ackermann H.-W., DuBow M.S. *Viruses of prokaryotes.* CRC press, 1987.
74. Krupovic M., Forterre P. Microviridae Goes Temperate: Microvirus-Related Proviruses Reside in the Genomes of Bacteroidetes // *PLoS ONE.* 2011. Vol. 6, № 5. P. e19893.
75. Quaiser A. et al. Diversity and comparative genomics of Microviridae in Sphagnum- dominated peatlands // *Front. Microbiol.* 2015. Vol. 6. P. 375.
76. Dutilh B.E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes // *Nat. Commun.* 2014. Vol. 5. P. 4498.
77. Al-Shayeb B. et al. Clades of huge phages from across Earth’s ecosystems: 7795 // *Nature.* Nature Publishing Group, 2020. Vol. 578, № 7795. P. 425–431.
78. Devoto A.E. et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes // *Nat. Microbiol.* 2019. Vol. 4, № 4. P. 693–700.
79. Laanto E. et al. Virus found in a boreal lake links ssDNA and dsDNA viruses // *Proc. Natl. Acad. Sci. U. S. A.* 2017. Vol. 114, № 31. P. 8378–8383.

80. Roux S. et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes // *Nat. Microbiol.* 2019. Vol. 4, № 11. P. 1895–1906.
81. Adriaenssens E.M. et al. Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee // *Arch. Virol.* 2020. Vol. 165, № 5. P. 1253–1260.
82. Krishnamurthy S.R. et al. Hyperexpansion of RNA Bacteriophage Diversity // *PLoS Biol.* 2016. Vol. 14, № 3. P. e1002409.
83. Callanan J. et al. Expansion of known ssRNA phage genomes: From tens to over a thousand // *Sci. Adv.* 2020. Vol. 6, № 6. P. eaay5981.
84. Adriaenssens E.M. et al. Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee // *Arch. Virol.* 2018. Vol. 163, № 4. P. 1125–1129.
85. Walker P.J. et al. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019) // *Arch. Virol.* 2019. Vol. 164, № 9. P. 2417–2429.
86. Barylski J. et al. ICTV Virus Taxonomy Profile: Herelleviridae // *J. Gen. Virol.* 2020. Vol. 101, № 4. P. 362–363.
87. Wittmann J. et al. From Orphan Phage to a Proposed New Family-the Diversity of N4-Like Viruses // *Antibiot. Basel Switz.* 2020. Vol. 9, № 10. P. E663.
88. Walker P.J. et al. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020) // *Arch. Virol.* 2020. Vol. 165, № 11. P. 2737–2748.
89. Guerin E. et al. Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut // *Cell Host Microbe.* 2018. Vol. 24, № 5. P. 653–664.e6.
90. Yutin N. et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut // *Nat. Microbiol.* 2018. Vol. 3, № 1. P. 38–46.
91. Current ICTV Taxonomy Release | ICTV [Electronic resource]. URL: <https://ictv.global/taxonomy> (accessed: 09.11.2022).

92. Mäntynen S. et al. ICTV Virus Taxonomy Profile: Finnlakeviridae // *J. Gen. Virol. Microbiology Society*,. Vol. 101, № 9. P. 894–895.
93. Krupovic M., ICTV Report Consortium YR 2018. ICTV Virus Taxonomy Profile: Plasmaviridae // *J. Gen. Virol. Microbiology Society*,. Vol. 99, № 5. P. 617–618.
94. Adriaenssens E.M. et al. Integration of genomic and proteomic analyses in the classification of the Siphoviridae family // *Virology*. 2015. Vol. 477. P. 144–154.
95. Gregory A.C. et al. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer // *BMC Genomics*. 2016. Vol. 17. P. 930.
96. Gregory A.C. et al. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut // *Cell Host Microbe*. 2020. Vol. 28, № 5. P. 724-740.e8.
97. Roux S. et al. Minimum Information about an Uncultivated Virus Genome (MIUViG) // *Nat. Biotechnol.* 2019. Vol. 37, № 1. P. 29–37.
98. Altschul S.F. et al. Basic local alignment search tool // *J. Mol. Biol.* 1990. Vol. 215, № 3. P. 403–410.
99. Rodriguez-R L.M., Konstantinidis K.T. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes: e1900v1. PeerJ Inc., 2016.
100. Yoon S.-H. et al. A large-scale evaluation of algorithms to calculate average nucleotide identity // *Antonie Van Leeuwenhoek*. 2017. Vol. 110, № 10. P. 1281–1286.
101. Jain C. et al. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries // *Nat. Commun.* 2018. Vol. 9, № 1. P. 5114.
102. Fu L. et al. CD-HIT: accelerated for clustering the next-generation sequencing data // *Bioinformatics*. 2012. Vol. 28, № 23. P. 3150–3152.
103. Ågren J. et al. Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups // *PLOS ONE. Public Library of Science*, 2012. Vol. 7, № 6. P. e39107.
104. Richter M. et al. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison // *Bioinformatics*. 2016. Vol. 32, № 6. P. 929–931.

105. Moraru C., Varsani A., Kropinski A.M. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses: 11 // *Viruses*. Multidisciplinary Digital Publishing Institute, 2020. Vol. 12, № 11. P. 1268.
106. Stanton C.R. et al. Isolation and Characterisation of the Bundooravirus Genus and Phylogenetic Investigation of the Salasmaviridae Bacteriophages // *Viruses*. 2021. Vol. 13, № 8. P. 1557.
107. Rohwer F., Edwards R. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage // *J. Bacteriol.* 2002. Vol. 184, № 16. P. 4529–4535.
108. Nishimura Y. et al. ViPTree: the viral proteomic tree server // *Bioinforma. Oxf. Engl.* 2017. Vol. 33, № 15. P. 2379–2380.
109. Aiewsakun P. et al. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy // *J. Gen. Virol.* 2018. Vol. 99, № 9. P. 1331–1343.
110. Aiewsakun P., Simmonds P. The genomic underpinnings of eukaryotic virus taxonomy: creating a sequence-based framework for family-level virus classification // *Microbiome*. 2018. Vol. 6, № 1. P. 38.
111. Simmonds P., Aiewsakun P. Virus classification – where do you draw the line? // *Arch. Virol.* 2018. Vol. 163, № 8. P. 2037–2046.
112. Chibani C.M. et al. Classifying the Unclassified: A Phage Classification Method: 2 // *Viruses*. Multidisciplinary Digital Publishing Institute, 2019. Vol. 11, № 2. P. 195.
113. Nguyen L.-T. et al. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies // *Mol. Biol. Evol.* 2015. Vol. 32, № 1. P. 268–274.
114. Kalyaanamoorthy S. et al. ModelFinder: fast model selection for accurate phylogenetic estimates: 6 // *Nat. Methods*. Nature Publishing Group, 2017. Vol. 14, № 6. P. 587–589.
115. Hoang D.T. et al. UFBoot2: Improving the Ultrafast Bootstrap Approximation // *Mol. Biol. Evol.* 2018. Vol. 35, № 2. P. 518–522.
116. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies // *Bioinforma. Oxf. Engl.* 2014. Vol. 30, № 9. P. 1312–1313.

117. Salter L.A., Pearl D.K. Stochastic search strategy for estimation of maximum likelihood phylogenetic trees // *Syst. Biol.* 2001. Vol. 50, № 1. P. 7–17.
118. Kozlov A.M. et al. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference // *Bioinformatics.* 2019. Vol. 35, № 21. P. 4453–4455.
119. Sanderson M.J., McMahon M.M., Steel M. Terraces in phylogenetic tree space // *Science.* 2011. Vol. 333, № 6041. P. 448–450.
120. Lutteropp S., Kozlov A.M., Stamatakis A. A fast and memory-efficient implementation of the transfer bootstrap // *Bioinformatics.* 2020. Vol. 36, № 7. P. 2280–2281.
121. Darriba D. et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models // *Mol. Biol. Evol.* 2020. Vol. 37, № 1. P. 291–294.
122. Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models // *Bioinformatics.* 2003. Vol. 19, № 12. P. 1572–1574.
123. Kubatko L.S., Degnan J.H. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence // *Syst. Biol.* 2007. Vol. 56, № 1. P. 17–24.
124. Andrade-Martínez J.S., Moreno-Gallego J.L., Reyes A. Defining a Core Genome for the Herpesvirales and Exploring their Evolutionary Relationship with the Caudovirales // *Sci. Rep.* 2019. Vol. 9. P. 11342.
125. Low S.J. et al. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales // *Nat. Microbiol.* 2019. Vol. 4, № 8. P. 1306–1315.
126. Chaudhari N.M., Gupta V.K., Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline // *Sci. Rep.* 2016. Vol. 6, № 1. P. 24373.
127. Turner D. et al. CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes // *BMC Res. Notes.* 2013. Vol. 6. P. 140.
128. Bayliss S.C. et al. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria // *GigaScience.* 2019. Vol. 8, № 10. P. giz119.

129. Chen X. et al. PGAweb: A Web Server for Bacterial Pan-Genome Analysis // *Front. Microbiol.* 2018. Vol. 9.
130. Zhao Y. et al. PGAP: pan-genomes analysis pipeline // *Bioinformatics.* 2012. Vol. 28, № 3. P. 416–418.
131. Lechner M. et al. Proteinortho: detection of (co-)orthologs in large-scale analysis // *BMC Bioinformatics.* 2011. Vol. 12. P. 124.
132. Page A.J. et al. Roary: rapid large-scale prokaryote pan genome analysis // *Bioinformatics.* 2015. Vol. 31, № 22. P. 3691–3693.
133. Tonkin-Hill G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline // *Genome Biol.* 2020. Vol. 21, № 1. P. 180.
134. Birin H. et al. Inferring Models of Rearrangements, Recombinations, and Horizontal Transfers by the Minimum Evolution Criterion // *Algorithms in Bioinformatics* / ed. Giancarlo R., Hannenhalli S. Berlin, Heidelberg: Springer, 2007. P. 111–123.
135. Koonin E.V., Dolja V.V. Virus World as an Evolutionary Network of Viruses and Capsidless Selfish Elements // *Microbiol. Mol. Biol. Rev. MMBR.* 2014. Vol. 78, № 2. P. 278–303.
136. Iranzo J., Krupovic M., Koonin E.V. The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing // *mBio.* 2016. Vol. 7, № 4. P. e00978-16.
137. Bolduc B. et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria // *PeerJ.* 2017. Vol. 5. P. e3243.
138. Lima-Mendez G. et al. Reticulate representation of evolutionary and functional relationships between phage genomes // *Mol. Biol. Evol.* 2008. Vol. 25, № 4. P. 762–777.
139. Forterre P., Prangishvili D. The origin of viruses // *Res. Microbiol.* 2009. Vol. 160, № 7. P. 466–472.
140. Yutin N., Wolf Y.I., Koonin E.V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life // *Virology.* 2014. Vol. 0. P. 38–52.

141. Krupovic M., Koonin E.V. Multiple origins of viral capsid proteins from cellular ancestors // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 2017. Vol. 114, № 12. P. E2401–E2410.
142. Krupovic M., Dolja V.V., Koonin E.V. Origin of viruses: primordial replicators recruiting capsids from hosts: 7 // *Nat. Rev. Microbiol.* Nature Publishing Group, 2019. Vol. 17, № 7. P. 449–458.
143. Kang H.S. et al. Prophage genomics reveals patterns in phage genome organization and replication // *bioRxiv*. Cold Spring Harbor Laboratory, 2017.
144. Botstein D. A theory of modular evolution for bacteriophages // *Ann. N. Y. Acad. Sci.* 1980. Vol. 354. P. 484–490.
145. Gontcharov A.A., Marin B., Melkonian M. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta) // *Mol. Biol. Evol.* 2004. Vol. 21, № 3. P. 612–624.
146. Mai-Prochnow A. et al. ‘Big things in small packages: the genetics of filamentous phage and effects on fitness of their host’ // *FEMS Microbiol. Rev.* 2015. Vol. 39, № 4. P. 465–487.
147. Bankevich A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing // *J. Comput. Biol. J. Comput. Mol. Cell Biol.* 2012. Vol. 19, № 5. P. 455–477.
148. Seemann T. Prokka: rapid prokaryotic genome annotation // *Bioinforma. Oxf. Engl.* 2014. Vol. 30, № 14. P. 2068–2069.
149. Goodacre N. et al. A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection // *mSphere*. American Society for Microbiology, 2018. Vol. 3, № 2. P. e00069-18.
150. Hyatt D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification // *BMC Bioinformatics*. 2010. Vol. 11, № 1. P. 119.
151. Delcher A.L. et al. Identifying bacterial genes and endosymbiont DNA with Glimmer // *Bioinforma. Oxf. Engl.* 2007. Vol. 23, № 6. P. 673–679.

152. Kelley L.A. et al. The Phyre2 web portal for protein modeling, prediction and analysis // *Nat. Protoc.* 2015. Vol. 10, № 6. P. 845–858.
153. Gabler F. et al. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit // *Curr. Protoc. Bioinforma.* 2020. Vol. 72, № 1. P. e108.
154. Schattner P., Brooks A.N., Lowe T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs // *Nucleic Acids Res.* 2005. Vol. 33, № Web Server issue. P. W686–W689.
155. Laslett D., Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences // *Nucleic Acids Res.* 2004. Vol. 32, № 1. P. 11–16.
156. Sampaio M. et al. Predicting promoters in phage genomes using PhagePromoter // *Bioinformatics.* 2019. Vol. 35, № 24. P. 5301–5302.
157. Lee I. et al. OrthoANI: An improved algorithm and software for calculating average nucleotide identity // *Int. J. Syst. Evol. Microbiol.* 2016. Vol. 66, № 2. P. 1100–1103.
158. Gascuel O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data // *Mol. Biol. Evol.* 1997. Vol. 14, № 7. P. 685–695.
159. Sullivan M.J., Petty N.K., Beatson S.A. Easyfig: a genome comparison visualizer // *Bioinformatics.* 2011. Vol. 27, № 7. P. 1009–1010.
160. Sievers F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega // *Mol. Syst. Biol.* 2011. Vol. 7. P. 539.
161. Katoh K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform // *Nucleic Acids Res.* 2002. Vol. 30, № 14. P. 3059–3066.
162. Katoh K., Standley D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability // *Mol. Biol. Evol.* 2013. Vol. 30, № 4. P. 772–780.
163. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Res.* 2004. Vol. 32, № 5. P. 1792–1797.

164. Price M.N., Dehal P.S., Arkin A.P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments // PLOS ONE. Public Library of Science, 2010. Vol. 5, № 3. P. e9490.
165. Edler D. et al. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML // Methods Ecol. Evol. 2021. Vol. 12, № 2. P. 373–377.
166. Letunic I., Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation // Nucleic Acids Res. 2021. Vol. 49, № W1. P. W293–W296.
167. Huerta-Cepas J., Serra F., Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data // Mol. Biol. Evol. 2016. Vol. 33, № 6. P. 1635–1638.
168. Robinson D.F., Foulds L.R. Comparison of phylogenetic trees // Math. Biosci. 1981. Vol. 53, № 1. P. 131–147.
169. Shannon P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks // Genome Res. 2003. Vol. 13, № 11. P. 2498–2504.
170. Jumper J. et al. Highly accurate protein structure prediction with AlphaFold // Nature. 2021. Vol. 596, № 7873. P. 583–589.
171. Evans R. et al. Protein complex prediction with AlphaFold-Multimer. 2021. P. 2021.10.04.463034.
172. Raman S. et al. Structure prediction for CASP8 with all-atom refinement using Rosetta // Proteins. 2009. Vol. 77, № 0 9. P. 89–99.
173. Baek M. et al. Accurate prediction of protein structures and interactions using a three-track neural network // Science. American Association for the Advancement of Science, 2021. Vol. 373, № 6557. P. 871–876.
174. Holm L. Using Dali for Protein Structure Comparison // Methods Mol. Biol. Clifton NJ. 2020. Vol. 2112. P. 29–42.
175. Holm L. Dali server: structural unification of protein families // Nucleic Acids Res. 2022. Vol. 50, № W1. P. W210–W215.

176. Dong R. et al. mTM-align: an algorithm for fast and accurate multiple protein structure alignment // *Bioinforma. Oxf. Engl.* 2018. Vol. 34, № 10. P. 1719–1725.
177. Zhang Y., Skolnick J. Scoring function for automated assessment of protein structure template quality // *Proteins.* 2004. Vol. 57, № 4. P. 702–710.
178. Akhter S., Aziz R.K., Edwards R.A. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies // *Nucleic Acids Res.* 2012. Vol. 40, № 16. P. e126.
179. Arndt D. et al. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes // *Brief. Bioinform.* 2017. Vol. 20, № 4. P. 1560–1567.
180. Eddy S.R. Accelerated Profile HMM Searches // *PLOS Comput. Biol. Public Library of Science*, 2011. Vol. 7, № 10. P. e1002195.
181. Finn R.D., Clements J., Eddy S.R. HMMER web server: interactive sequence similarity searching // *Nucleic Acids Res. Oxford University Press*, 2011. Vol. 39, № Web Server issue. P. W29.
182. Voronina M.V. et al. First Report of *Pectobacterium carotovorum* subsp. *brasiliense* Causing Blackleg and Stem Rot Disease of Potato in Russia // *Plant Dis. Scientific Societies*, 2019. Vol. 103, № 2. P. 364–364.
183. Evseev P.V. et al. Origin and Evolution of Studiervirinae Bacteriophages Infecting *Pectobacterium*: Horizontal Transfer Assists Adaptation to New Niches: 11 // *Microorganisms. Multidisciplinary Digital Publishing Institute*, 2020. Vol. 8, № 11. P. 1707.
184. Chen Z., Schneider T.D. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases // *Nucleic Acids Res.* 2005. Vol. 33, № 19. P. 6172–6187.
185. Zaczek-Moczydłowska M.A. et al. Phage cocktail containing Podoviridae and Myoviridae bacteriophages inhibits the growth of *Pectobacterium* spp. under in vitro and in vivo conditions // *PloS One.* 2020. Vol. 15, № 4. P. e0230842.
186. Bartual S.G. et al. Structure of the bacteriophage T4 long tail fiber receptor-binding tip // *Proc. Natl. Acad. Sci. U. S. A.* 2010. Vol. 107, № 47. P. 20287–20292.

187. Garcia-Doval C., van Raaij M.J. Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers // Proc. Natl. Acad. Sci. U. S. A. 2012. Vol. 109, № 24. P. 9390–9395.
188. Lukianova A.A. et al. Pectobacterium versatile Bacteriophage Possum: A Complex Polysaccharide-Deacetylating Tail Fiber as a Tool for Host Recognition in Pectobacterial Schitoviridae: 19 // Int. J. Mol. Sci. Multidisciplinary Digital Publishing Institute, 2022. Vol. 23, № 19. P. 11043.
189. Buttimer C. et al. Novel N4-Like Bacteriophages of Pectobacterium atrosepticum: 2 // Pharmaceuticals. Multidisciplinary Digital Publishing Institute, 2018. Vol. 11, № 2. P. 45.
190. Carstens A.B. et al. A novel six-phage cocktail reduces Pectobacterium atrosepticum soft rot infection in potato tubers under simulated storage conditions // FEMS Microbiol. Lett. 2019. Vol. 366, № 9. P. fnz101.
191. Smolarska A. et al. Isolation and phenotypic and morphological characterization of the first Podoviridae lytic bacteriophages ϕ A38 and ϕ A41 infecting Pectobacterium parmentieri (former Pectobacterium wasabiae) // Eur. J. Plant Pathol. 2018. Vol. 150, № 2. P. 413–425.
192. van der Wolf J.M. et al. Management of Diseases Caused by Pectobacterium and Dickeya Species // Plant Diseases Caused by Dickeya and Pectobacterium Species / ed. Van Gijsegem F., van der Wolf J.M., Toth I.K. Cham: Springer International Publishing, 2021. P. 175–214.
193. Toth I.K. et al. Pectobacterium and Dickeya: Environment to Disease Development // Plant Diseases Caused by Dickeya and Pectobacterium Species / ed. Van Gijsegem F., van der Wolf J.M., Toth I.K. Cham: Springer International Publishing, 2021. P. 39–84.
194. Holtappels D. et al. The future of phage biocontrol in integrated plant protection for sustainable crop production // Curr. Opin. Biotechnol. 2021. Vol. 68. P. 60–71.
195. Svircev A., Roach D., Castle A. Framing the Future with Bacteriophages in Agriculture // Viruses. 2018. Vol. 10, № 5. P. 218.

196. Zhang Y., Fan Q., Loria R. A re-evaluation of the taxonomy of phytopathogenic genera *Dickeya* and *Pectobacterium* using whole-genome sequencing data // *Syst. Appl. Microbiol.* 2016. Vol. 39, № 4. P. 252–259.
197. Meijer W.J.J., Horcajadas J.A., Salas M. ϕ 29 Family of Phages // *Microbiol. Mol. Biol. Rev.* 2001. Vol. 65, № 2. P. 261–287.
198. Tarakanov R.I. et al. Ayka, a Novel *Curtobacterium* Bacteriophage, Provides Protection against Soybean Bacterial Wilt and Tan Spot // *Int. J. Mol. Sci.* 2022. Vol. 23, № 18. P. 10913.
199. Mendez J., Blanco L., Salas M. Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. // *EMBO J.* 1997. Vol. 16, № 9. P. 2519–2527.
200. Hrebík D. et al. Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68 // *Sci. Adv.* 2019. Vol. 5, № 10. P. eaaw7414.
201. Gokey T. et al. Structure of the *Bacillus anthracis* dTDP-L-rhamnose biosynthetic pathway enzyme: dTDP- α -D-glucose 4,6-dehydratase, RfbB // *J. Struct. Biol.* 2018. Vol. 202, № 2. P. 175–181.
202. Theodorou I. et al. A dual-chain assembly pathway generates the high structural diversity of cell-wall polysaccharides in *Lactococcus lactis* // *J. Biol. Chem. Elsevier*, 2019. Vol. 294, № 46. P. 17612–17625.
203. Balzaretto S. et al. A Novel Rhamnose-Rich Hetero-exopolysaccharide Isolated from *Lactobacillus paracasei* DG Activates THP-1 Human Monocytic Cells // *Appl. Environ. Microbiol.* 2017. Vol. 83, № 3. P. e02702-16.
204. Zatyka M., Thomas C.M. Control of genes for conjugative transfer of plasmids and other mobile elements // *FEMS Microbiol. Rev.* 1998. Vol. 21, № 4. P. 291–319.
205. Evseev P. et al. Prophage-Derived Regions in *Curtobacterium* Genomes: Good Things, Small Packages: 2 // *Int. J. Mol. Sci. Multidisciplinary Digital Publishing Institute*, 2023. Vol. 24, № 2. P. 1586.
206. Juhala R.J. et al. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages // *J. Mol. Biol.* 2000. Vol. 299, № 1. P. 27–51.

207. Pedulla M.L. et al. Origins of Highly Mosaic Mycobacteriophage Genomes // Cell. Elsevier, 2003. Vol. 113, № 2. P. 171–182.
208. Evseev P. et al. Pseudomonas Phage MD8: Genetic Mosaicism and Challenges of Taxonomic Classification of Lambdoid Bacteriophages: 19 // Int. J. Mol. Sci. Multidisciplinary Digital Publishing Institute, 2021. Vol. 22, № 19. P. 10350.
209. Holguín A.V. et al. Phage Φ Pan70, a Putative Temperate Phage, Controls *Pseudomonas aeruginosa* in Planktonic, Biofilm and Burn Mouse Model Assays // Viruses. 2015. Vol. 7, № 8. P. 4602–4623.
210. Tariq M.A. et al. Temperate Bacteriophages from Chronic *Pseudomonas aeruginosa* Lung Infections Show Disease-Specific Changes in Host Range and Modulate Antimicrobial Susceptibility // mSystems. 2019. Vol. 4, № 4. P. e00191-18.
211. Zylicz M. et al. Formation of the preprimosome protects λ O from RNA transcription-dependent proteolysis by ClpP/ClpX // Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences, 1998. Vol. 95, № 26. P. 15259–15263.
212. Campbell A. Comparative Molecular Biology of Lambdoid Phages // Annu. Rev. Microbiol. 1994. Vol. 48, № 1. P. 193–222.
213. Ravin V. et al. Genomic sequence and analysis of the atypical temperate bacteriophage N15 // J. Mol. Biol. 2000. Vol. 299, № 1. P. 53–73.
214. Tarkowski T.A. et al. Gene products encoded in the ninR region of phage λ participate in Red-mediated recombination // Genes Cells. 2002. Vol. 7, № 4. P. 351–363.
215. Kaiser A.D. Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli* // Virology. 1957. Vol. 3, № 1. P. 42–61.
216. Johnson A., Meyer B.J., Ptashne M. Mechanism of action of the cro protein of bacteriophage lambda // Proc. Natl. Acad. Sci. U. S. A. 1978. Vol. 75, № 4. P. 1783–1787.
217. Casjens S.R., Hendrix R.W. Bacteriophage lambda: Early pioneer and still relevant // Virology. 2015. Vol. 479–480. P. 310–330.
218. King J., Casjens S. Catalytic head assembling protein in virus morphogenesis // Nature. 1974. Vol. 251, № 5471. P. 112–119.

219. Duda R.L., Oh B., Hendrix R.W. Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation // *J. Mol. Biol.* 2013. Vol. 425, № 15. P. 2765–2781.
220. Duda R.L. et al. Structural transitions during bacteriophage HK97 head assembly // *J. Mol. Biol.* 1995. Vol. 247, № 4. P. 618–635.
221. Medina E. et al. Assembly and Maturation of the Bacteriophage Lambda Procapsid: gpC Is the Viral Protease // *J. Mol. Biol.* 2010. Vol. 401, № 5. P. 813–830.
222. Fokine A., Rossmann M.G. Common Evolutionary Origin of Procapsid Proteases, Phage Tail Tubes, and Tubes of Bacterial Type VI Secretion Systems // *Struct. Lond. Engl.* 1993. 2016. Vol. 24, № 11. P. 1928–1935.
223. Chang J.R. et al. Functional domains of the bacteriophage P2 scaffolding protein: identification of residues involved in assembly and protease activity // *Virology.* 2009. Vol. 384, № 1. P. 144–150.
224. Chen Z. et al. Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 2017. Vol. 114, № 39. P. E8184–E8193.
225. Guo F. et al. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions // *Proc. Natl. Acad. Sci. U. S. A.* 2014. Vol. 111, № 43. P. E4606–E4614.
226. Casjens S., King J. P22 morphogenesis I: Catalytic scaffolding protein in capsid assembly // *J. Supramol. Struct.* 1974. Vol. 2, № 2–4. P. 202–224.
227. Nelson R.A., Reilly B.E., Anderson D.L. Morphogenesis of bacteriophage phi 29 of *Bacillus subtilis*: preliminary isolation and characterization of intermediate particles of the assembly pathway // *J. Virol.* 1976. Vol. 19, № 2. P. 518–532.
228. Shaw J.E., Murialdo H. Morphogenetic genes C and Nu3 overlap in bacteriophage lambda // *Nature.* 1980. Vol. 283, № 5742. P. 30–35.
229. Latino L. et al. A novel *Pseudomonas aeruginosa* Bacteriophage, Ab31, a Chimera Formed from Temperate Phage PAJU2 and *P. putida* Lytic Phage AF: Characteristics

- and Mechanism of Bacterial Resistance // PLOS ONE. Public Library of Science, 2014. Vol. 9, № 4. P. e93777.
230. Labrie S.J. et al. Genomes of marine cyanopodoviruses reveal multiple origins of diversity // Environ. Microbiol. 2013. Vol. 15, № 5. P. 1356–1376.
231. Huang S. et al. Comparative Genomic and Phylogenomic Analyses Reveal a Conserved Core Genome Shared by Estuarine and Oceanic Cyanopodoviruses // PLoS One. 2015. Vol. 10, № 11. P. e0142962.
232. Mizuno C.M. et al. Expanding the Marine Viroisphere Using Metagenomics // PLOS Genet. Public Library of Science, 2013. Vol. 9, № 12. P. e1003987.
233. Misteli T. Beyond the Sequence: Cellular Organization of Genome Function // Cell. Elsevier, 2007. Vol. 128, № 4. P. 787–800.
234. Zheng H., Xie W. The role of 3D genome organization in development and cell differentiation // Nat. Rev. Mol. Cell Biol. 2019. Vol. 20, № 9. P. 535–550.
235. Nováček J. et al. Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate // Proc. Natl. Acad. Sci. U. S. A. 2016. Vol. 113, № 33. P. 9351–9356.
236. Fokine A. et al. The molecular architecture of the bacteriophage T4 neck // J. Mol. Biol. 2013. Vol. 425, № 10. P. 1731–1744.
237. Carson S. et al. Genome Sequences of Six Paenibacillus larvae Siphoviridae Phages // Genome Announc. 2015. Vol. 3, № 3. P. e00101-15.
238. Beims H. et al. Discovery of Paenibacillus larvae ERIC V: Phenotypic and genomic comparison to genotypes ERIC I-IV reveal different inventories of virulence factors which correlate with epidemiological prevalences of American Foulbrood // Int. J. Med. Microbiol. IJMM. 2020. Vol. 310, № 2. P. 151394.
239. Richardson J.S. The Anatomy and Taxonomy of Protein Structure // Advances in Protein Chemistry / ed. Anfinsen C.B., Edsall J.T., Richards F.M. Academic Press, 1981. Vol. 34. P. 167–339.
240. Fraser J.S. et al. Ig-Like Domains on Bacteriophages: A Tale of Promiscuity and Deceit // J. Mol. Biol. 2006. Vol. 359, № 2. P. 496–507.

241. Salemme F.R., Miller M.D., Jordan S.R. Structural Convergence during Protein Evolution // Proc. Natl. Acad. Sci. U. S. A. National Academy of Sciences, 1977. Vol. 74, № 7. P. 2820–2824.
242. Wood T.C., Pearson W.R. Evolution of protein sequences and structures // J. Mol. Biol. 1999. Vol. 291, № 4. P. 977–995.
243. Holm L. et al. Searching protein structure databases with DaliLite v.3 // Bioinformatics. 2008. Vol. 24, № 23. P. 2780–2781.
244. Zhou X., Chou J., Wong S.T. Protein structure similarity from principle component correlation analysis // BMC Bioinformatics. 2006. Vol. 7, № 1. P. 40.
245. Liu Y. et al. Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes // PLOS Biol. Public Library of Science, 2021. Vol. 19, № 11. P. e3001442.
246. Krylov V.N., Zhazykov I.Z. Pseudomonas bacteriophage phiKZ--possible model for studying the genetic control of morphogenesis // Genetika. 1978. Vol. 14, № 4. P. 678–685.
247. Mesyanzhinov V.V. et al. The genome of bacteriophage ϕ KZ of *Pseudomonas aeruginosa* 1 Edited by M. Gottesman // J. Mol. Biol. 2002. Vol. 317, № 1. P. 1–19.
248. Kristensen D.M., Cai X., Mushegian A. Evolutionarily Conserved Orthologous Families in Phages Are Relatively Rare in Their Prokaryotic Hosts ∇ // J. Bacteriol. 2011. Vol. 193, № 8. P. 1806–1814.
249. Helgstrand C. et al. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution // J. Mol. Biol. 2003. Vol. 334, № 5. P. 885–899.
250. Davis C.R. et al. Characterization of a Primordial Major Capsid-Scaffolding Protein Complex in Icosahedral Virus Shell Assembly // J. Mol. Biol. 2022. Vol. 434, № 19. P. 167719.
251. Fokine A. et al. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry // Proc. Natl. Acad. Sci. U. S. A. 2005. Vol. 102, № 20. P. 7163–7168.

252. Fang Q. et al. Structures of a large prolate virus capsid in unexpanded and expanded states generate insights into the icosahedral virus assembly // *Proc. Natl. Acad. Sci. U. S. A.* 2022. Vol. 119, № 40. P. e2203272119.
253. Ahi Y.S. et al. Adenoviral L4 33K forms ring-like oligomers and stimulates ATPase activity of IVa2: implications in viral genome packaging // *Front. Microbiol.* 2015. Vol. 6. P. 318.
254. Hilbert B.J. et al. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain // *Nucleic Acids Res.* 2017. Vol. 45, № 6. P. 3591–3605.
255. Zhao H. et al. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 2013. Vol. 110, № 20. P. 8075–8080.
256. Kwan T. et al. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages // *Proc. Natl. Acad. Sci. Proceedings of the National Academy of Sciences*, 2005. Vol. 102, № 14. P. 5174–5179.
257. Ha E., Son B., Ryu S. *Clostridium perfringens* Virulent Bacteriophage CPS2 and Its Thermostable Endolysin LysCPS2: 5 // *Viruses. Multidisciplinary Digital Publishing Institute*, 2018. Vol. 10, № 5. P. 251.
258. Tétart F. et al. Phylogeny of the Major Head and Tail Genes of the Wide-Ranging T4-Type Bacteriophages // *J. Bacteriol.* 2001. Vol. 183, № 1. P. 358–366.
259. Adriaenssens E.M. et al. T4-Related Bacteriophage LIMEstone Isolates for the Control of Soft Rot on Potato Caused by ‘*Dickeya solani*’ // *PLOS ONE. Public Library of Science*, 2012. Vol. 7, № 3. P. e33227.
260. Sullivan M.B. et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments // *Environ. Microbiol.* 2010. Vol. 12, № 11. P. 3035–3056.
261. Green J. et al. Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in southeastern Virginia, USA // *Aquat. Microb. Ecol.* 2015. Vol. 75, № 2. P. 117–128.

262. Bartlau N. et al. Highly diverse flavobacterial phages isolated from North Sea spring blooms: 2 // *ISME J. Nature Publishing Group*, 2022. Vol. 16, № 2. P. 555–568.
263. Phothaworn P. et al. Characterization of Flagellotropic, Chi-Like Salmonella Phages Isolated from Thai Poultry Farms // *Viruses*. 2019. Vol. 11, № 6. P. 520.
264. Castillo D., Middelboe M. Genomic diversity of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum* // *FEMS Microbiol. Lett.* 2016. Vol. 363, № 24. P. fnw272.
265. Steven A.C. et al. Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, co-operativity, and super-stabilization by soc binding // *J. Mol. Biol.* 1992. Vol. 228, № 3. P. 870–884.
266. Bowman B.R. et al. Structure of the herpesvirus major capsid protein // *EMBO J. John Wiley & Sons, Ltd*, 2003. Vol. 22, № 4. P. 757–765.
267. Hark Gan H. et al. Analysis of Protein Sequence/Structure Similarity Relationships // *Biophys. J.* 2002. Vol. 83, № 5. P. 2781–2791.
268. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap // *Evolution*. [Society for the Study of Evolution, Wiley], 1985. Vol. 39, № 4. P. 783–791.
269. Som A. Causes, consequences and solutions of phylogenetic incongruence // *Brief. Bioinform.* 2015. Vol. 16, № 3. P. 536–548.
270. Thiergart T., Landan G., Martin W.F. Concatenated alignments and the case of the disappearing tree // *BMC Evol. Biol.* 2014. Vol. 14, № 1. P. 266.
271. Vakulenko Y. et al. Modular Evolution of Coronavirus Genomes: 7 // *Viruses. Multidisciplinary Digital Publishing Institute*, 2021. Vol. 13, № 7. P. 1270.
272. McGeoch D.J., Rixon F.J., Davison A.J. Topics in herpesvirus genomics and evolution // *Virus Res.* 2006. Vol. 117, № 1. P. 90–104.
273. Davison A.J. Evolution of the herpesviruses // *Vet. Microbiol.* 2002. Vol. 86, № 1–2. P. 69–88.
274. McGeoch D.J., Dolan A., Ralph A.C. Toward a Comprehensive Phylogeny for Mammalian and Avian Herpesviruses // *J. Virol.* 2000. Vol. 74, № 22. P. 10401–10406.

275. Koonin E.V., Yutin N. Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism // *Adv. Virus Res.* 2019. Vol. 103. P. 167–202.
276. Subramaniam K. et al. A New Family of DNA Viruses Causing Disease in Crustaceans from Diverse Aquatic Biomes // *mBio.* 2020. Vol. 11, № 1. P. e02938-19.
277. Weigel C., Seitz H. Bacteriophage replication modules // *FEMS Microbiol. Rev.* 2006. Vol. 30, № 3. P. 321–381.
278. Yutin N., Raoult D., Koonin E.V. Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies // *Viol. J.* 2013. Vol. 10, № 1. P. 158.

ПРИЛОЖЕНИЕ А

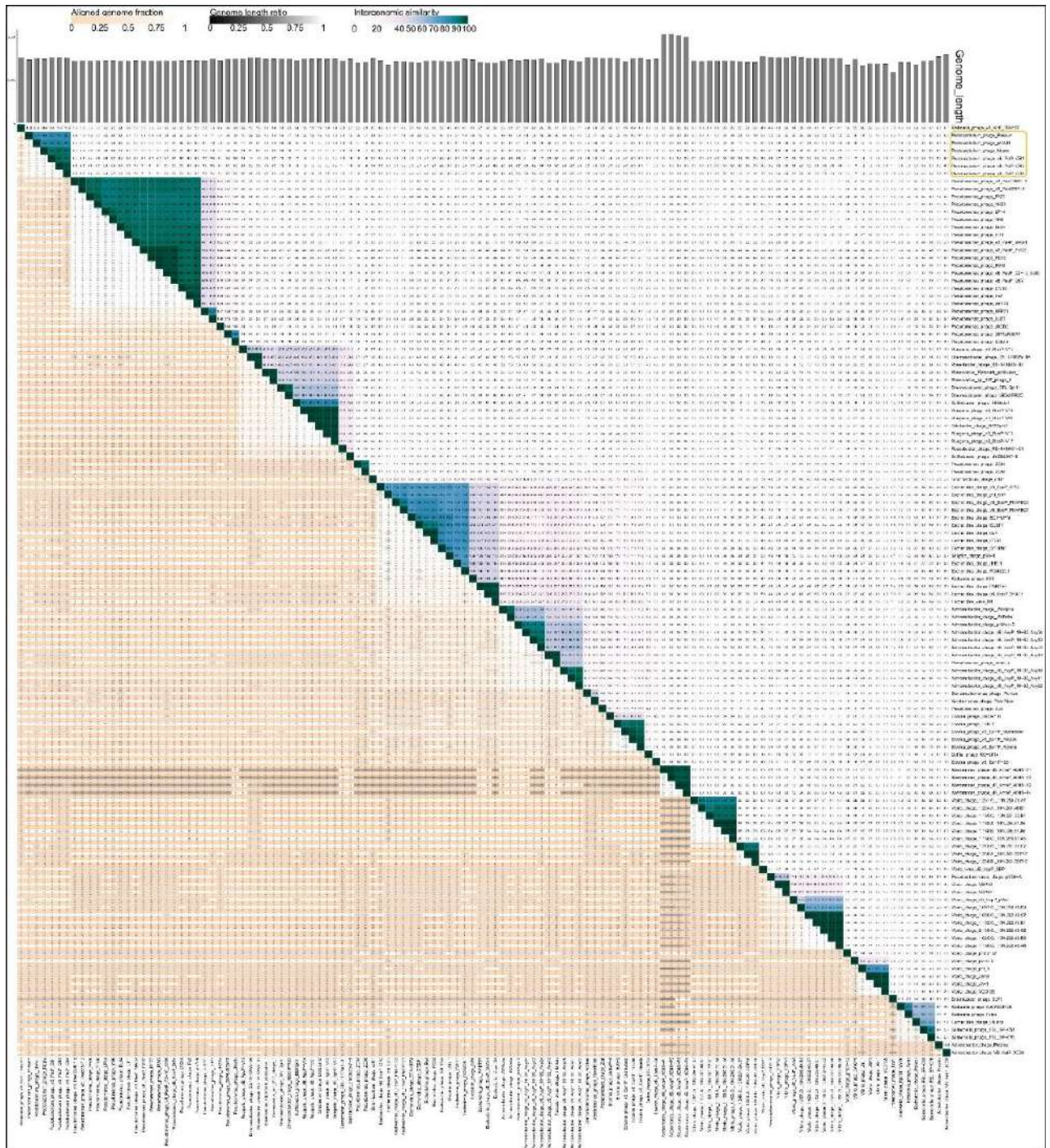
Дополнительные материалы по характеристике фагов *Pectobacterium*

Рисунок А.1. Общий вид кластерной тепловой карты фага *Pectobacterium* Possum и родственных фагов, полученной на основании межгеномного сходства. Карта в высоком разрешении доступна в сети Интернет по адресу <https://www.mdpi.com/article/10.3390/ijms231911043/s1>.



Рисунок А.2. Филогенетическое дерево, построенное с использованием аминокислотных последовательностей ДНКП фага *Pectobacterium DU_PP_III* и других фагов. Таксономическая принадлежность по состоянию на начало 2021 г. указана справа от названия фага. Значения бутстрэпа (количество реплик 1000) указаны около соответствующих веток.

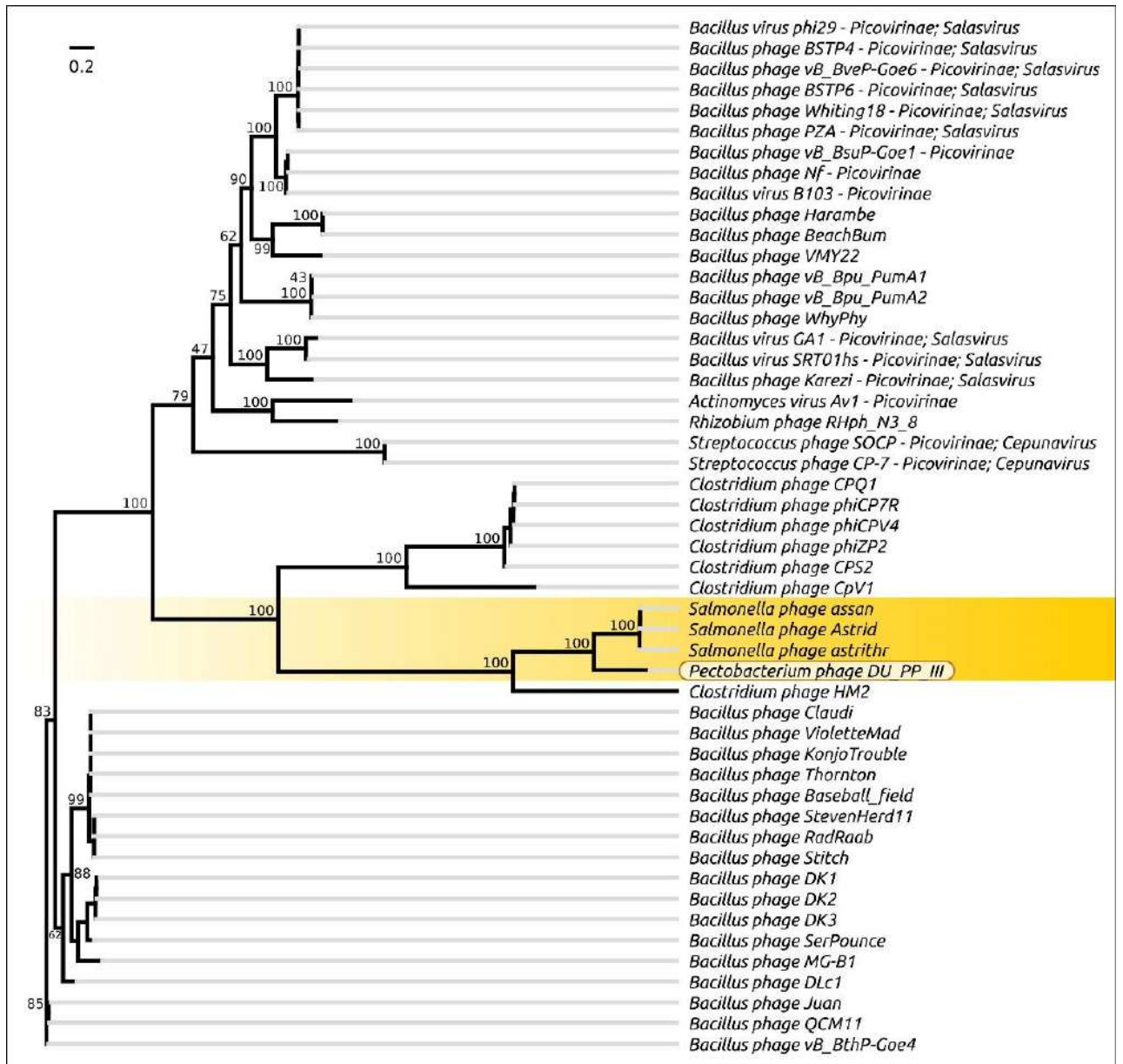


Рисунок А.3. Филогенетическое дерево, построенное с использованием аминокислотных последовательностей портового белка фага *Pectobacterium* DU_PP_III и других фагов. Таксономическая принадлежность по состоянию на начало 2021 г. указана справа от названия фага. Значения бутстрэпа (количество реплик 1000) указаны около соответствующих веток



Рисунок А.4. Филогенетическое дерево, построенное с использованием аминокислотных последовательностей терминазы фага *Pectobacterium* DU_PP_III и других фагов. Таксономическая принадлежность по состоянию на начало 2021 г. указана справа от названия фага. Значения бутстрэпа (количество реплик 1000) указаны около соответствующих веток.

ПРИЛОЖЕНИЕ Б

Филогенетическое дерево фага *Curtobacterium* Аука, полученное с использованием сервера GRAViTy

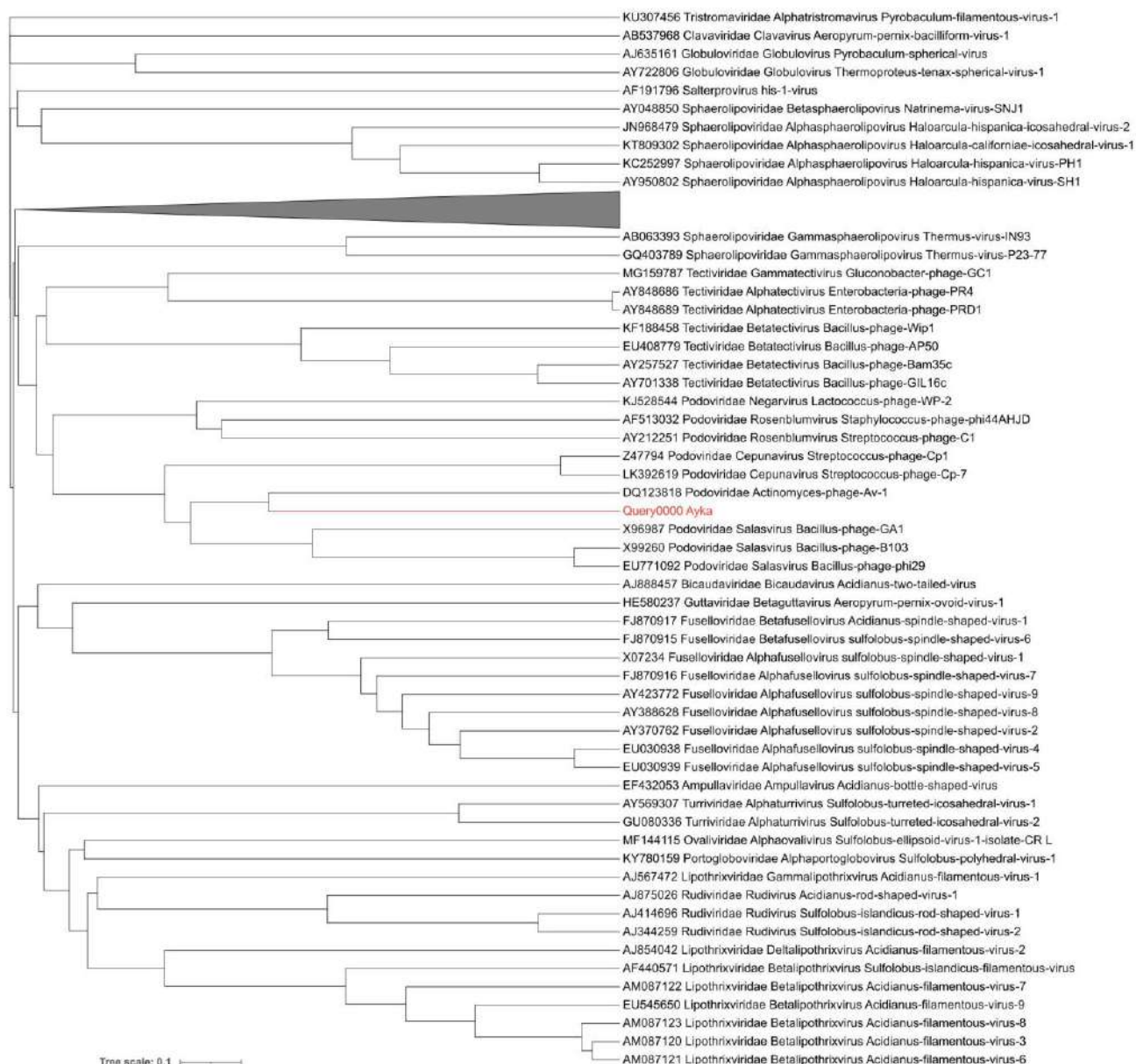


Рисунок Б.1. Филогенетическое дерево, полученное при помощи сервера GRAViTy с использованием геномных последовательностей фага *Curtobacterium* Аука и встроенной базы фаговых последовательностей. Сжатая клада содержит фаги, не включённые в остальные ветки. Дерево в высоком разрешении доступно в сети Интернет по адресу <https://www.mdpi.com/1422-0067/23/18/10913/s1?version=1663492425>.

ПРИЛОЖЕНИЕ В

Профаговые области, обнаруженные в геномах *Curtobacterium*Таблица В.1. Список профаговых областей, обнаруженных в геномах *Curtobacterium*.

Кластер	Профаг	Размер, тыс. н.п.	ГЦ- состав, %	Бактериальный штамм
1	C_sp_C111	47,1	69,5	<i>Curtobacterium</i> sp. C1
2	C_sp_MCSS17_01512	31,5	67,4	<i>Curtobacterium</i> sp. MCSS17_015
3	C_f_S5_2611	27,2	68,3	<i>Curtobacterium flaccumfaciens</i> S5_26
	C_sp_MCLR17_03611	26,6	68,6	<i>Curtobacterium</i> sp. MCLR17_036
4	C_sp_MCLR17_03211	38,3	62,5	<i>Curtobacterium</i> sp. MCLR17_032
5	C_luteum_NS18411	22,5	62,3	<i>Curtobacterium luteum</i> NS184
	C_sp_VKM_Ac-137612	39,4	64,1	<i>Curtobacterium</i> sp. VKM Ac-1376
6	C_sp_MCBD17_03012	39,4	65,6	<i>Curtobacterium</i> sp. MCBD17_030
7	C_sp_UCD-KPL256012	23,2	66,6	<i>Curtobacterium</i> sp. UCD-KPL2560
8	C_sp_MCLR17_03411	40	61,3	<i>Curtobacterium</i> sp. MCLR17_034
9	C_sp_VKM_Ac-288411	43	69	<i>Curtobacterium</i> sp. VKM_Ac-2884
10	C_sp_MCBD17_00811	35,7	69,7	<i>Curtobacterium</i> sp. MCBD17_008
	C_sp_WW713	34,8	69,7	<i>Curtobacterium</i> sp. WW7
11	C_citreum_DSM_2052811	36	69,6	<i>Curtobacterium citreum</i> DSM 20528
	C_citreum_JCM_134511	35,9	69,6	<i>Curtobacterium citreum</i> JCM 1345
	C_f_VKM_Ac-138611	38,3	69,3	<i>Curtobacterium flaccumfaciens</i> VKM Ac-1386
	C_sp_Csp111	37,5	70,1	<i>Curtobacterium</i> sp. Csp1
	C_sp_Ferrero11	34,9	65,4	<i>Curtobacterium</i> sp. Ferrero
	C_sp_MCLR17_04211	35,5	65,8	<i>Curtobacterium</i> sp. MCLR17_042
	C_sp_MCSS17_00611	35,6	65,7	<i>Curtobacterium</i> sp. MCSS17_006
	C_sp_MCSS17_01111	36,7	65,4	<i>Curtobacterium</i> sp. MCSS17_011
	C_sp_PhB17211	35,9	65,3	<i>Curtobacterium</i> sp. PhB172
12	C_f_p_f_BRIP_7060711	34,6	65,8	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> BRIP_70607
	C_f_p_oortii_CFBP_138411	36,3	65,6	<i>Curtobacterium flaccumfaciens</i> pv. <i>oortii</i> CFBP 1384
	C_sp_MCBD17_02811	39,3	69,4	<i>Curtobacterium</i> sp. MCBD17_028
	C_sp_PhB13711	34,1	65,2	<i>Curtobacterium</i> sp. PhB137
13	C_f_p_f_CFBP_341711	17,5	67,8	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> CFBP 3417
	C_pusillum_ATCC_1909611	17	68,5	<i>Curtobacterium pusillum</i> ATCC 19096
	C_sp_Csp211	16,8	68,9	<i>Curtobacterium</i> sp. Csp211
	C_sp_MCBD17_02911	17,2	69,1	<i>Curtobacterium</i> sp. MCBD17_02911
	C_sp_MCBD17_03211	17,7	69,3	<i>Curtobacterium</i> sp. MCBD17_03211
	C_sp_MCJR17_04311	17,2	69,2	<i>Curtobacterium</i> sp. MCJR17_04311
	C_sp_MCJR17_05511	17,2	69,2	<i>Curtobacterium</i> sp. MCJR17_05511
	C_sp_MCPF17_01511	17,2	69,2	<i>Curtobacterium</i> sp. MCPF17_01511
	C_sp_PhB14611	17,8	69,1	<i>Curtobacterium</i> sp. PhB14611
	C_sp_VKM_Ac-137611	17,2	69,5	<i>Curtobacterium</i> sp. VKM Ac-1376
	C_sp_VKM_Ac-286111	17,5	69,2	<i>Curtobacterium</i> sp. VKM Ac-2861
14	C_sp_912811	38,4	69,2	<i>Curtobacterium</i> sp. 9128

	C_sp_MCBA15_012 1	41,2	70,8	<i>Curtobacterium</i> sp. MCBA15_012
15	C_f_p_f_CFBP_8824 1	16,5	69,5	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> CFBP 8824
	C_f_p_oortii_CFBP_1384 2	16,5	68,3	<i>Curtobacterium flaccumfaciens</i> pv. <i>oortii</i> _CFBP 1384
	C_sp_MCBA15_008 1	15,2	64,6	<i>Curtobacterium</i> sp. MCBA15_008
	C_sp_MCBD17_035 1	16,5	66	<i>Curtobacterium</i> sp. MCBD17_035
	C_sp_MCBD17_040 2	16,2	66,5	<i>Curtobacterium</i> sp. MCBD17_040
	C_sp_MCSS17_011 2	16,5	65,3	<i>Curtobacterium</i> sp. MCSS17_011
	C_sp_YC1 2	17,2	65,2	<i>Curtobacterium</i> sp. YC1
16	C_sp_MCSS17_007 1	37,9	62,1	<i>Curtobacterium</i> sp. MCSS17_007
17	C_sp_MCBA15_004 1	46,3	65,4	<i>Curtobacterium</i> sp. MCBA15_004
	C_sp_MCBD17_021 1	47,1	67,8	<i>Curtobacterium</i> sp. MCBD17_021
	C_sp_MCSS17_015 1	45	63,8	<i>Curtobacterium</i> sp. MCSS17_015
	C_sp_VKM_Ac-2884 2	43,9	69,2	<i>Curtobacterium</i> sp. VKM Ac-2884
18	C_albidum_DSM_20512 1	38,2	70,2	<i>Curtobacterium albidum</i> DSM 20512
	C_sp_Ferrero 3	41	70	<i>Curtobacterium</i> sp. Ferrero
	C_sp_MCBA15_004 2	14	69,8	<i>Curtobacterium</i> sp. MCBA15_004
19	C_sp_ISL-83 1	36,7	66,9	<i>Curtobacterium</i> sp. ISL-83
	C_sp_MCBD17_003 1	24,7	68,5	<i>Curtobacterium</i> sp. MCBD17_003
20	C_f_p_f_CFBP_3417 2	37	67,3	<i>Curtobacterium flaccumfaciens</i> pv. <i>flaccumfaciens</i> CFBP 3417
	C_sp_MCBA15_009 3	36,8	67,8	<i>Curtobacterium</i> sp. MCBA15_009
	C_sp_MCBD17_026 1	36,8	67	<i>Curtobacterium</i> sp. MCBD17_026
	C_sp_MCPF17_003 2	37,3	68,5	<i>Curtobacterium</i> sp. MCPF17_003
	C_sp_PhB25 1	36,6	69	<i>Curtobacterium</i> sp. PhB25
	C_sp_VKM_Ac-1796 1	37,2	67,3	<i>Curtobacterium</i> sp. VKM Ac-1796
	C_sp_VKM_Ac-2889 1	37,2	67,3	<i>Curtobacterium</i> sp. VKM Ac-2889
	C_sp_WW7 2	36,4	66,6	<i>Curtobacterium</i> sp. WW7

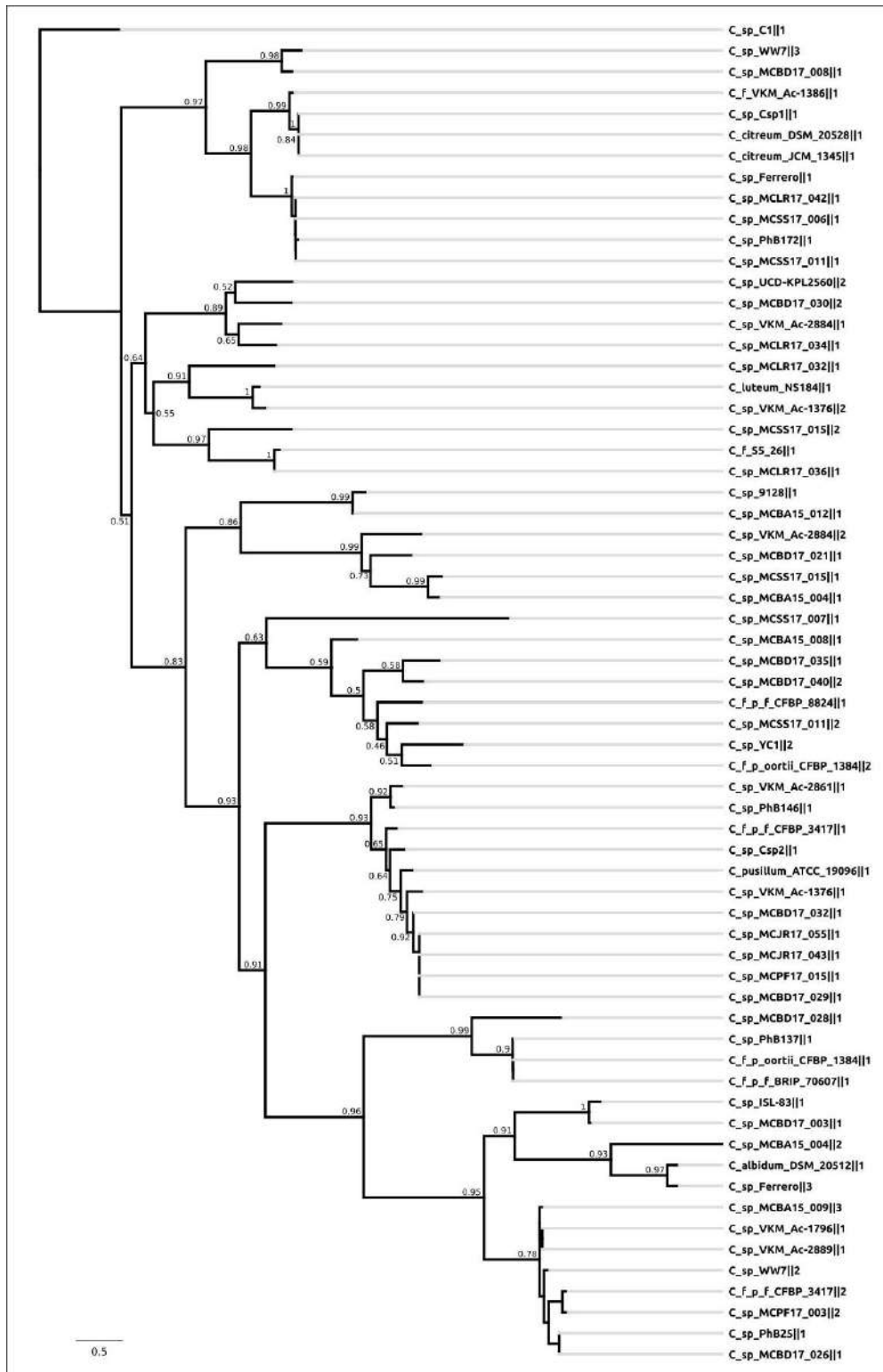


Рисунок В.1. Филогенетическое дерево, полученное на основе аминокислотных последовательностей главного капсидного белка, обнаруженного в профаговых областях *Curtobacterium*. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

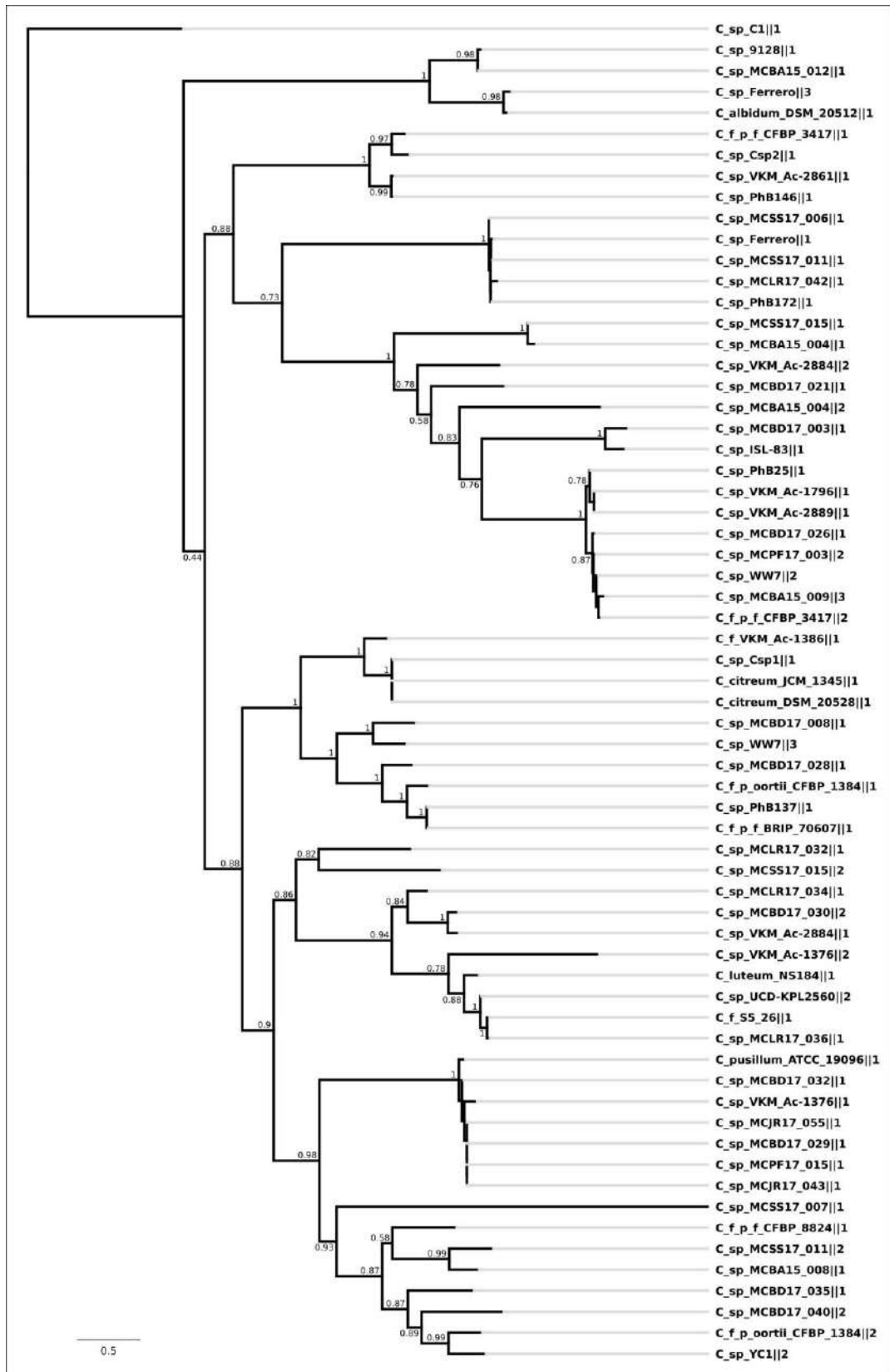
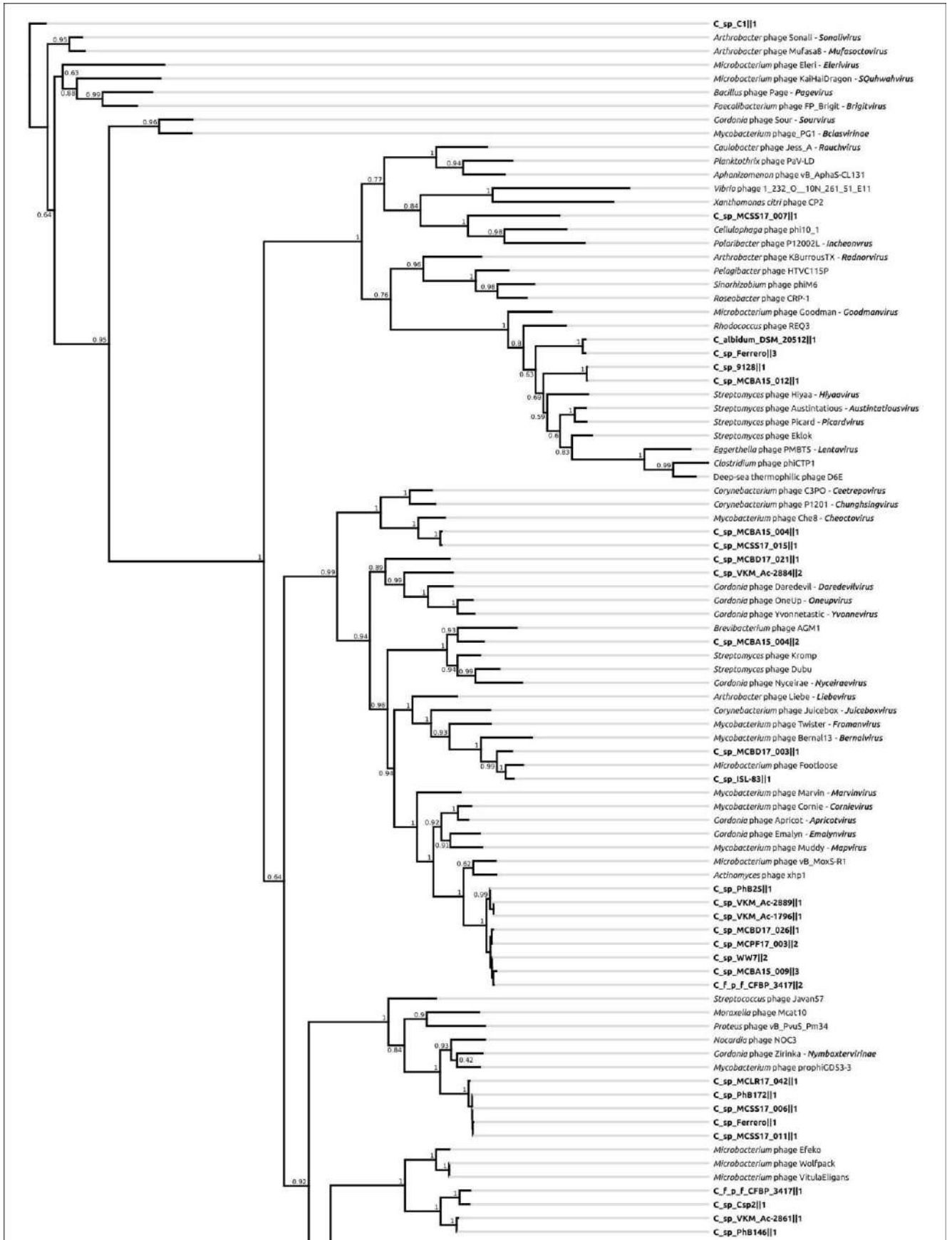


Рисунок В.2. Филогенетическое дерево, полученное на основе аминокислотных последовательностей большой субъединицы терминазы, обнаруженной в профаговых областях *Curtobacterium*. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.



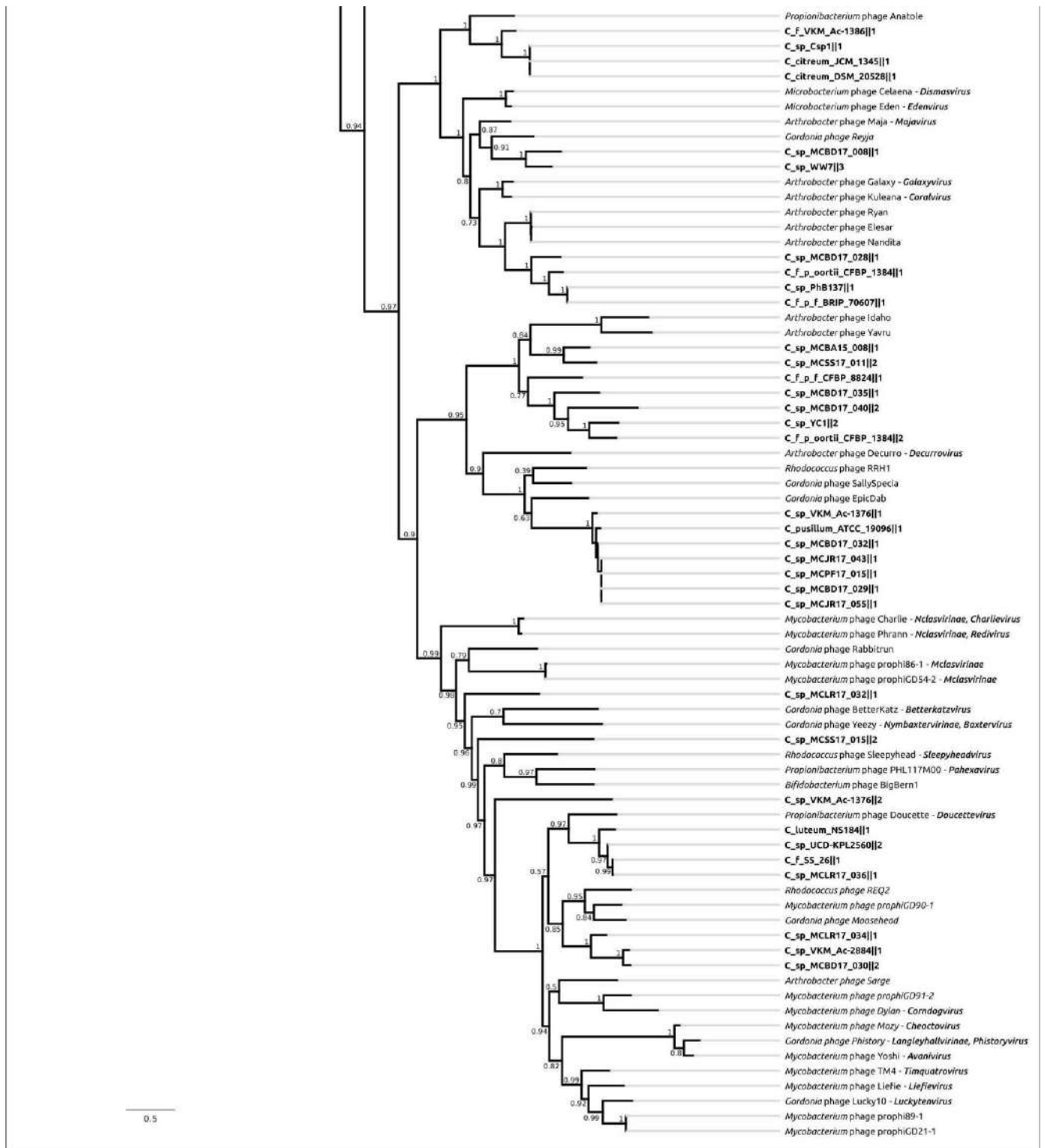


Рисунок В.3. Филогенетическое дерево, полученное на основе аминокислотных последовательностей терминазы, закодированной в профаговых областях *Curtobacterium* и в геномах родственных фагов. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

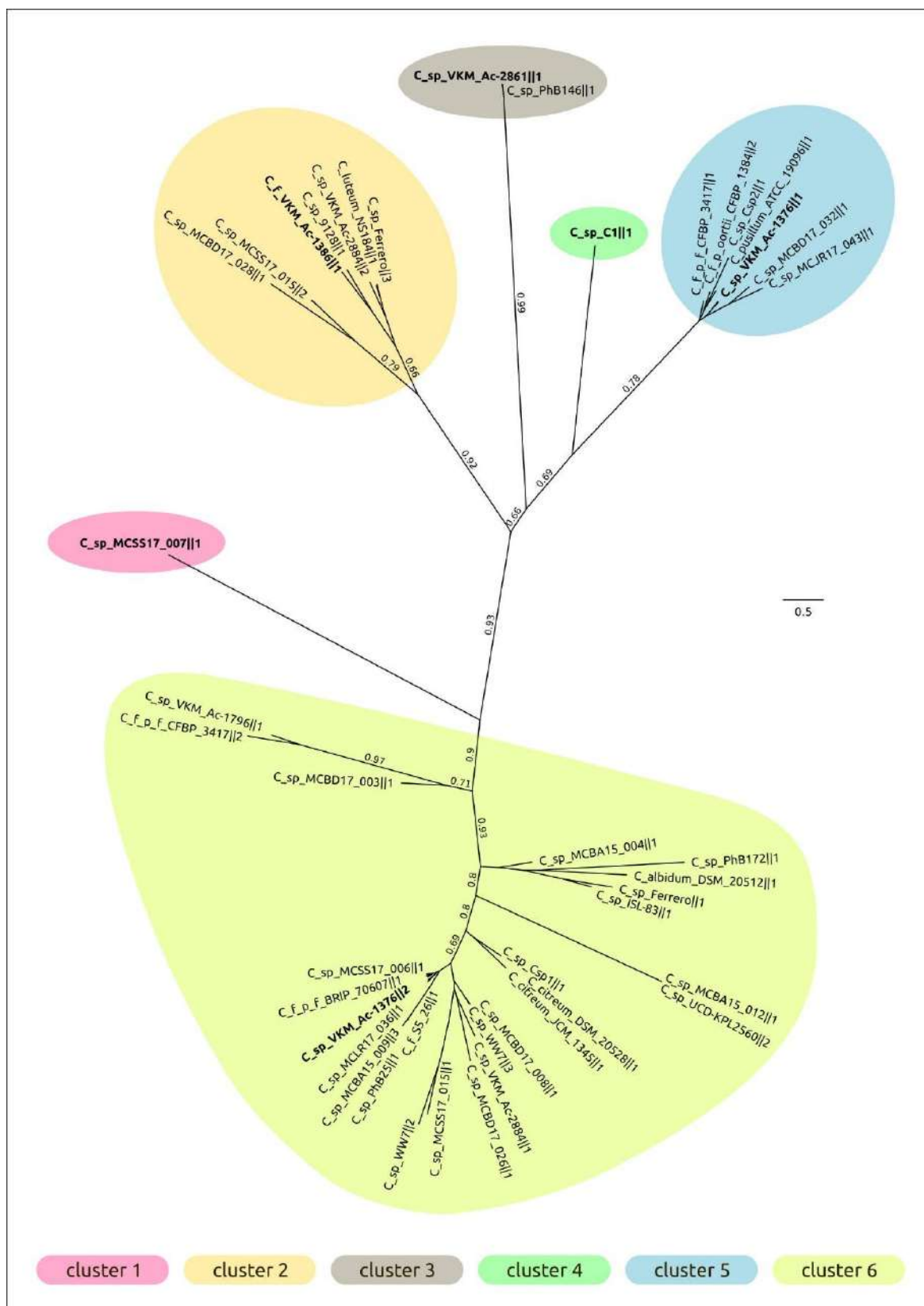


Рисунок В.4. Филогенетическое дерево, полученное на основе аминокислотных последовательностей эндוליцинов, обнаруженных в профаговых областях *Curtobacterium*. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

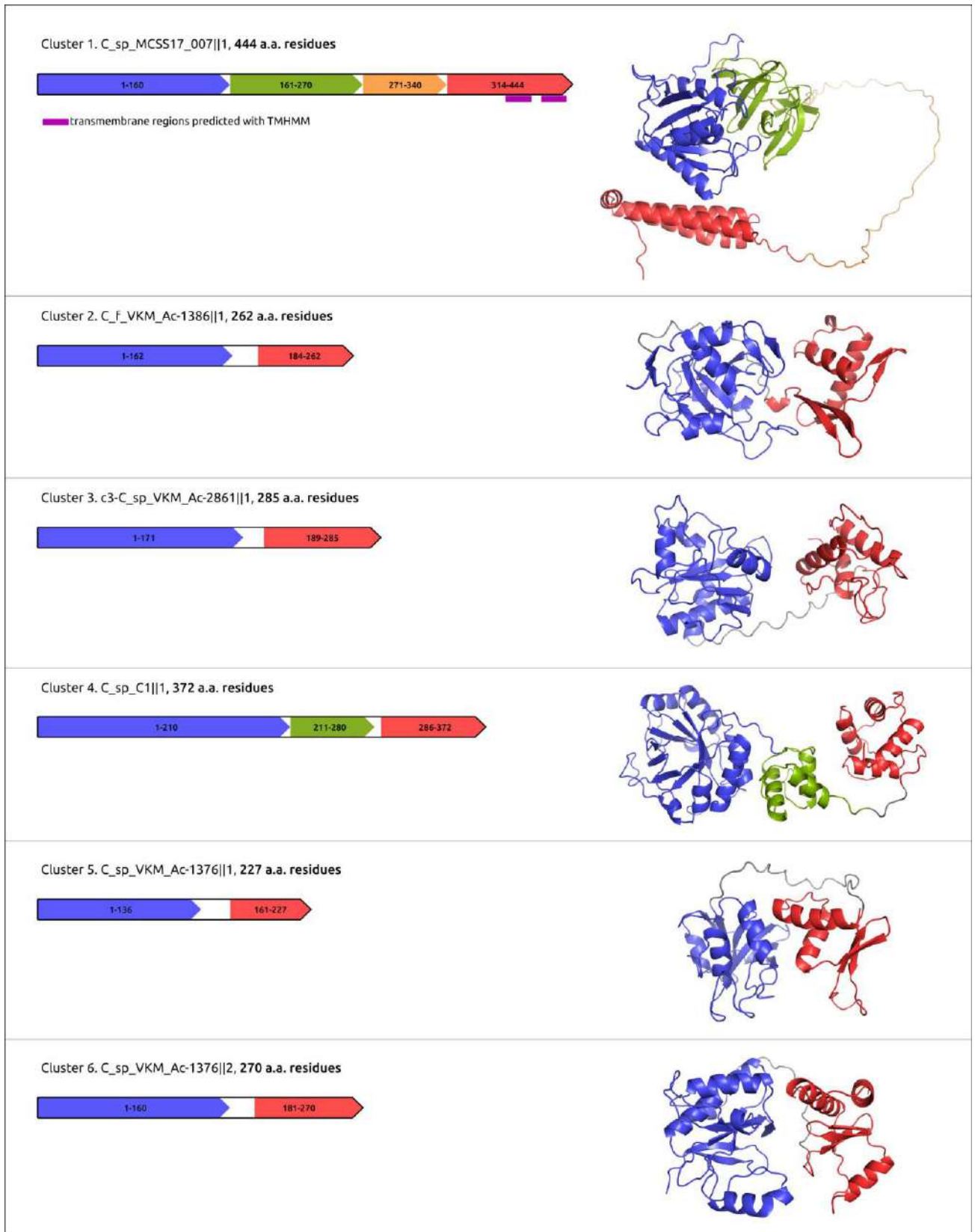


Рисунок В.5. Структурная архитектура эндолизиннов, обнаруженных в профаговых областях *Curtobacterium* spp., на основе предсказания структур с помощью AlphaFold и сравнения с экспериментально определёнными структурами.

ПРИЛОЖЕНИЕ Г

Фаг *Pseudomonas* MD8

Таблица Г.1. Фаги, составляющие группу MD8-подобных фагов.

Код NCBI GenBank	ANI	Размер выравнивания, н.п.	Размер генома, н.п.	Фаг
KX198612	100	42840	42840	<i>Pseudomonas</i> phage MD8
MK510975	96.19	10023	39780	<i>Pseudomonas</i> phage vB_Pae_CF145a
MK510976	95.86	11062	39780	<i>Pseudomonas</i> phage vB_Pae_CF165a
MK510964	95.08	10158	39780	<i>Pseudomonas</i> phage vB_Pae_CF28b
MK510968	94.95	13160	41820	<i>Pseudomonas</i> phage vB_Pae_CF55b
MK510969	94.51	9288	34680	<i>Pseudomonas</i> phage vB_Pae_CF60a
MK510986	94.08	15886	30600	<i>Pseudomonas</i> phage vB_Pae_BR213a
MK510988	93.96	14820	39780	<i>Pseudomonas</i> phage vB_Pae_BR299a
MK510962	93.92	7927	33660	<i>Pseudomonas</i> phage vB_Pae_CF3a
MK510990	93.91	11921	40800	<i>Pseudomonas</i> phage vB_Pae_BR133a
MK510963	93.88	16786	39780	<i>Pseudomonas</i> phage vB_Pae_CF24b
MK510970	93.68	14779	34680	<i>Pseudomonas</i> phage vB_Pae_CF67a
MK510974	93.68	17260	39780	<i>Pseudomonas</i> phage vB_Pae_CF140a
MK510987	93.58	17309	39780	<i>Pseudomonas</i> phage vB_Pae_BR233a
MK510971	93.57	14897	40800	<i>Pseudomonas</i> phage vB_Pae_CF79a
MK510985	93.56	17306	39780	<i>Pseudomonas</i> phage vB_Pae_BR204a
MK510978	93.25	15465	39780	<i>Pseudomonas</i> phage vB_Pae_CF208a
KT887558	93.11	11613	41820	<i>Pseudomonas</i> phage ϕ 2
MK510965	93.07	15680	40800	<i>Pseudomonas</i> phage vB_Pae_CF34a
MK510984	92.96	10307	37740	<i>Pseudomonas</i> phage vB_Pae_BR200a
MK510992	92.93	12157	39780	<i>Pseudomonas</i> phage vB_Pae_BR144a
MG707188	92.39	10879	42840	<i>Pseudomonas</i> phage TC7
KY707339	92.35	17040	39780	<i>Pseudomonas</i> phage JBD68
KJ959591	91.6	15111	38760	<i>Pseudomonas</i> phage PAN70
DQ163912	91.59	16514	38760	<i>Pseudomonas</i> phage F10
MK510993	89.11	5095	37740	<i>Pseudomonas</i> phage vB_Pae_BR201a

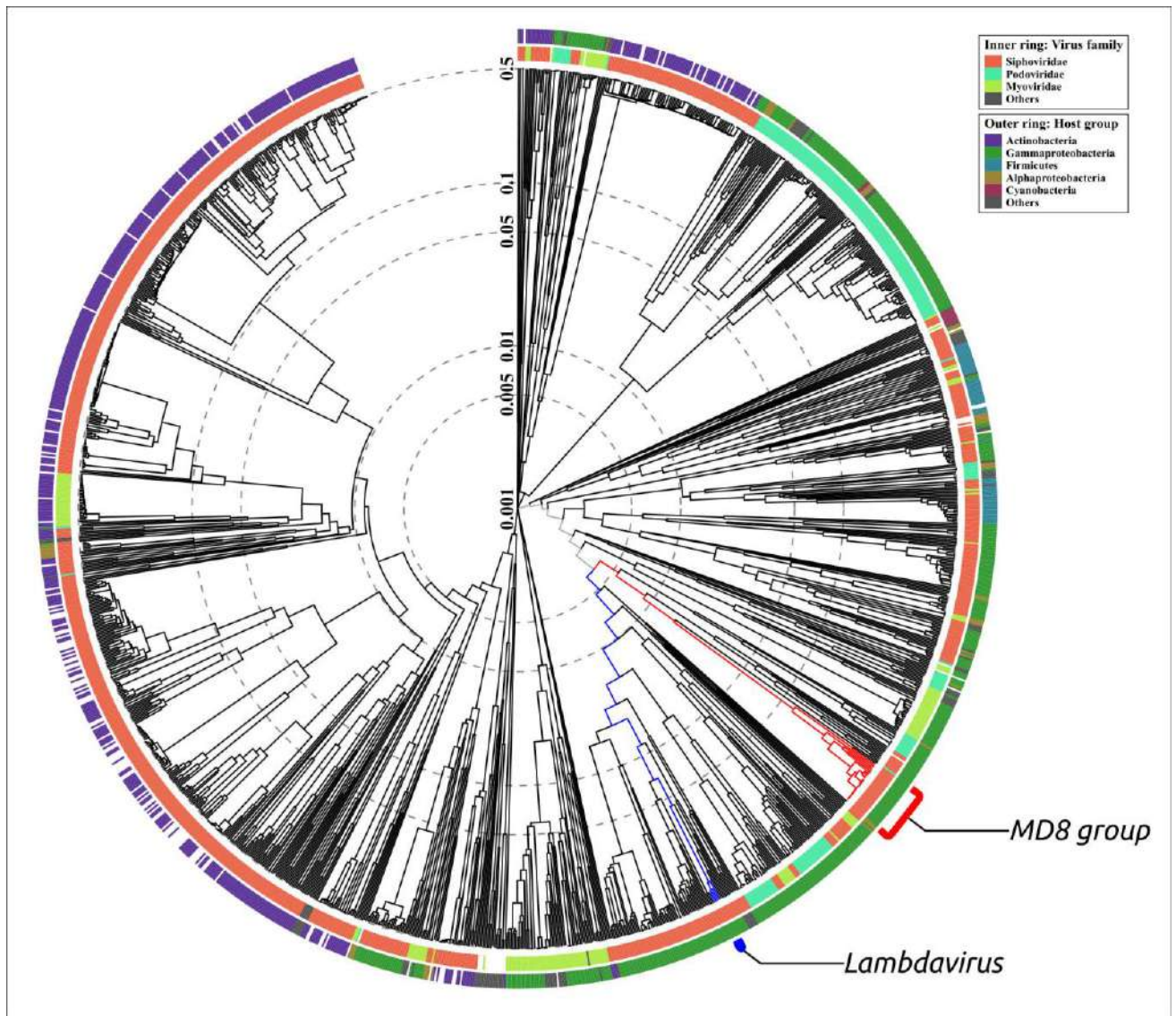


Рисунок Г.1. Филогенетическое дерево, полученное на основе сходства протеомов MD8-подобных (обведены красным) и других фагов. Таксономия фагов (по состоянию на 2020 г.) и хозяина указаны в легендах.

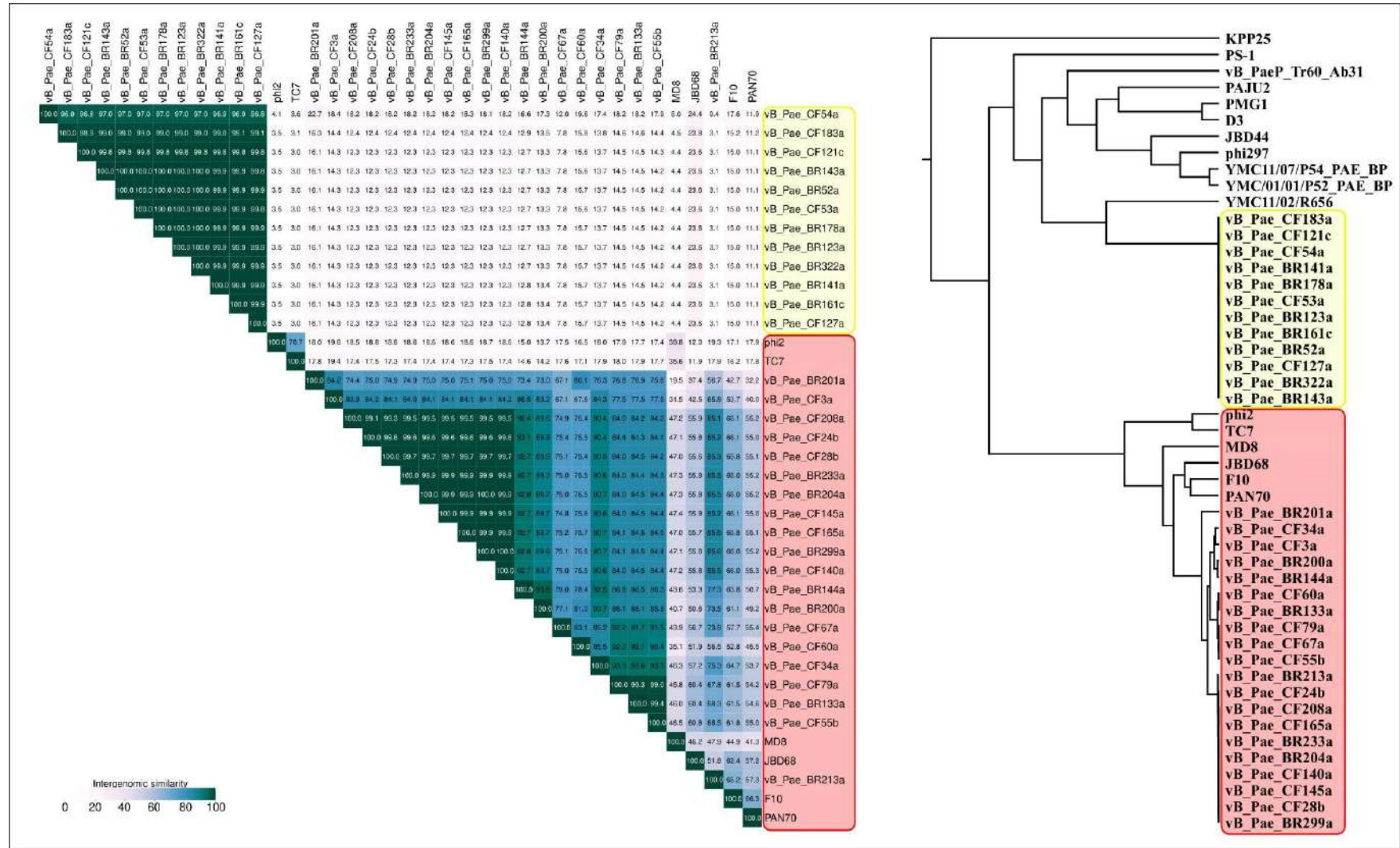


Рисунок Г.2. Тепловая карта межгеномного нуклеотидного сходства и протеомное дерево, полученное на основе сходства протеомов MD8-подобных (обведены красным) и других фагов.

ПРИЛОЖЕНИЕ Д

Эволюционная история чехольных белков

Таблица Д.1. Список экспериментально определённых структур чехольных белков. РД – рентгеновская дифракция, ЭМ – электронная микроскопия.

Код PDB	Описание	Организм, фаг	Разрешение	Метод
3FO8	Crystal structure of the bacteriophage T4 tail sheath protein, protease-resistant fragment gp18PR	фаг <i>Escherichia</i> T4	1.8 Å	РД
3FOA	Crystal structure of the bacteriophage T4 tail sheath protein, deletion mutant gp18M	фаг <i>Escherichia</i> T4	3.5 Å	РД
3HXL	Crystal structure of the sheath tail protein (DSY3957) from <i>Desulfitobacterium hafniense</i>	<i>Desulfitobacterium hafniense</i>	1.90 Å	РД
3LML	Crystal structure of the sheath tail protein Lin1278 from <i>Listeria innocua</i> , Northeast Structural Genomics Consortium Target LkR115	<i>Listeria innocua</i>	3.3 Å	РД
3SPE	Crystal structure of the tail sheath protein protease-resistant fragment from bacteriophage phiKZ	фаг <i>Pseudomonas</i> phiKZ	2.4 Å	РД
5LI4	Bacteriophage phi812K1-420 (<i>Staphylococcus</i> phage 812) tail sheath protein after contraction. This structure is related to 5LI2, 5LII, 5LIJ	фаг <i>Staphylococcus</i> 812	4.2 Å	ЭМ
6GKW	Crystal structure of the R-type bacteriocin (diffocin) sheath protein CD1363 from <i>Clostridium difficile</i> 630 in the pre-assembled state	<i>Clostridium difficile</i>	1.9 Å	РД
6PYT	CryoEM structure of precontracted pyocin R2 trunk from <i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>	2.9 Å	ЭМ
3J9O	CryoEM structure of a type VI secretion system from <i>Francisella tularensis</i> subsp. <i>novicida</i> U112	<i>Francisella tularensis</i> subsp. <i>novicida</i> a	3.70 Å	ЭМ
5N8N	CryoEM structure of contracted sheath of a <i>Pseudomonas aeruginosa</i> type VI secretion system consisting of TssB1 and TssC	<i>Pseudomonas aeruginosa</i>	3.28 Å	ЭМ
3J9G	Atomic model of the VipA/VipB, the type VI secretion system contractile sheath of <i>Vibrio cholerae</i>	<i>Vibrio cholerae</i>	3.5 Å	ЭМ

6RAO	Cryo-EM structure of the anti-feeding prophage (AFP) baseplate for <i>Serratia entomophila</i> . This structure is related to 6RAP, 6RBK, 6RBN, 6RC8, 6RGL	<i>Serratia entomophila</i>	3.1 Å	ЭМ
6J0B	Cryo-EM structure of an extracellular contractile injection system (CIS), PVC sheath-tube complex in extended state from <i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949	<i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949	2.9 Å	ЭМ
7AE0	Cryo-EM structure of an extracellular contractile injection system from the marine bacterium <i>Algoriphagus machipongonensis</i> with the sheath-tube module in its extended state. This structure is related to 7ADZ, 7AE0, 7AEB, 7AEF, 7AEK	<i>Algoriphagus machipongonensis</i>	2.4 Å	ЭМ
7B5I	Cryo-EM structure of the contractile injection system cap complex from <i>Anabaena</i> PCC7120	<i>Nostoc</i> sp.	2.8 Å	ЭМ



Рисунок Д.1. Выравнивание, полученное с использованием mTM-align и моделей ЧБ, предсказанных с помощью AlphaFold 2. Рисунок в высоком разрешении доступен в сети Интернет по адресу <https://www.mdpi.com/1999-4915/14/6/1148/s1?version=1653559708>.

Таблица Д.2. Список чехольных белков, отобранных для сравнительного структурного анализа.

№	Организм, тип системы для нефаговых ЧБ (AFP, Anti-Feeding Prophage; BCN, Bacteriocin; CHR, прокариотный геном; PMD, плаزمид; T6SS, система секреции VI типа)	Таксономия, в соответствии с NCBI Таксоному	Размер белка, а.о.	Кол-во доменов в модели
1	<i>Acinetobacter</i> phage Cato	<i>Myoviridae</i>	487	2
2	<i>Aeromonas</i> phage pAh6_2TG	<i>Chaseviridae; Nefertitivirinae; Pahsextavirus</i>	472	2
3	<i>Aeromonas</i> phage phiO18P	<i>Myoviridae; Peduovirinae; Bielevirus</i>	370	1
4	AFP-6JOB-SP1 <i>Photorhabdus asymbiotica</i> ATCC43949	Gammaproteobacteria; Enterobacterales; Morganellaceae	355	1
5	AFP-6JON-SP2 <i>Photorhabdus asymbiotica</i> ATCC43949	Gammaproteobacteria; Enterobacterales; Morganellaceae	440	1
6	AFP-6JON-SP3 <i>Photorhabdus asymbiotica</i> ATCC43949	Gammaproteobacteria; Enterobacterales; Morganellaceae	424	1
7	AFP-6RAO-SP3 <i>Serratia entomophila</i>	Gammaproteobacteria; Enterobacterales; Yersiniaceae	417	1
8	AFP-6RBN-SP1 <i>Serratia entomophila</i>	Gammaproteobacteria; Enterobacterales; Yersiniaceae	354	1
9	AFP-6RBN-SP2 <i>Serratia entomophila</i>	Gammaproteobacteria; Enterobacterales; Yersiniaceae	451	1
10	<i>Agrobacterium</i> phage Atu_ph04	<i>Ackermannviridae</i>	838	5+
11	<i>Agrobacterium</i> phage Atu_ph07	<i>Myoviridae; Polybotosvirus</i>	1086	5+
12	<i>Bacillus</i> phage AR9	<i>Myoviridae</i>	987	3
13	<i>Bacillus</i> phage BC01	<i>Herelleviridae; Bastillevirinae; Tsarbombavirus</i>	568	3
14	<i>Bacillus</i> phage CampHawk	<i>Herelleviridae; Spounavirinae; Okubovirus</i>	571	3
15	<i>Bacillus</i> phage G	<i>Myoviridae; Donellivirus</i>	579	3
16	<i>Bacillus</i> phage PBS1	<i>Myoviridae; Takahashivirus</i>	987	3
17	<i>Bacillus</i> phage phi18	<i>Herelleviridae; Spounavirinae; Okubovirus</i>	571	3
18	<i>Bacillus</i> phage SP-15	<i>Myoviridae; Thornevirus</i>	494	2
19	<i>Bacillus</i> phage TsarBomba	<i>Herelleviridae; Bastillevirinae; Tsarbombavirus</i>	568	3
20	<i>Bacillus</i> phage vB_BceM-HSE3	<i>Myoviridae</i>	727	4
21	BCN-6GKW- <i>Peptoclostridium difficile</i>	Firmicutes; Clostridia; Clostridiales; Peptostreptococcaceae; Clostridioides	356	1
22	BCN-6PYT- <i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; <i>Pseudomonas</i>	386	1
23	<i>Brevibacillus</i> phage Jimmer2	<i>Myoviridae; Jimmervirus</i>	437	2
24	<i>Brevibacillus</i> phage Osiris	<i>Myoviridae; Jimmervirus</i>	437	2
25	<i>Brevibacillus</i> phage Powder	<i>Myoviridae; Jimmervirus</i>	437	2

26	<i>Burkholderia</i> phage BEK	<i>Myoviridae; Peduovirinae; Tigrvirus</i>	342	1
27	<i>Burkholderia</i> phage KS14	<i>Myoviridae; Peduovirinae; Kisquattuordecimvirus</i>	391	1
28	<i>Burkholderia</i> phage phiE255	<i>Myoviridae; Bcepmyovirus</i>	477	2
29	<i>Campylobacter</i> phage CAM-P21	<i>Myoviridae</i>	397	1
30	<i>Campylobacter</i> phage F352	<i>Myoviridae; Eucampyvirinae; Fletchervirus</i>	636	3
31	<i>Caulobacter</i> phage Cr30	<i>Myoviridae</i>	688	3
32	<i>Cellulophaga</i> phage phi38:2	<i>Myoviridae</i>	508	2
33	<i>CHR-3HXL-Desulfitobacterium hafniense</i>	Firmicutes; Clostridia; Clostridiales; Peptococcales; <i>Desulfitobacterium</i>	446	2
34	<i>CHR-3LML-Listeria innocua</i>	Firmicutes; Bacilli; Bacillales; Listeriaceae	460	2
35	<i>CHR-Acidobacteria bacterium Mor1</i>	Acidobacteria	410	1
36	<i>CHR-Candidatus Bathyarchaeota archaeon isolate Bin-L-2</i>	Candidatus Bathyarchaeota	321	1
37	<i>CHR-Candidatus Lokiarchaeota archaeon isolate TEKIR_9</i>	Asgard group; Candidatus Lokiarchaeota	369	1
38	<i>CHR-Candidatus Nitrososphaera evergladensis SR1</i>	Thaumarchaeota; Nitrososphaeria; Nitrososphaerales; Nitrososphaeraceae	521	2
39	<i>CHR-Candidatus Pacearchaeota archaeon isolate ARS50</i>	Candidatus Pacearchaeota	634	3
40	<i>CHR-Candidatus Thorarchaeota archaeon isolate 2_13</i>	Asgard group; Candidatus Thorarchaeota	577	2
41	<i>CHR-Cellulomonas fimi</i> ATCC 484	Actinobacteria; Micrococcales; Cellulomonadaceae	523	2
42	<i>CHR-Crenarchaeota archaeon isolate LB_CRA_1</i>	Crenarchaeota	805	4
43	<i>CHR-Halovivax ruber</i> XH-70	Euryarchaeota; Stenosarchaea group; HaloNatrialbales; Natrialbaceae	574	3
44	<i>CHR-Methanolacinia petrolearia</i> DSM 11571	Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanolacinia	343	1
45	<i>CHR-Methanomethylovorans hollandica</i> DSM 15978	Euryarchaeota; Stenosarchaea group; Methanomicrobia; Methanosarcinales; Methanosarcinaceae	540	2
46	<i>CHR-Methanotherix</i> sp. isolate bin.308 Contig_420493	Euryarchaeota; Stenosarchaea group; Methanomicrobia; Methanosarcinales; Methanosarcinaceae	509	2
47	<i>CHR-Microbacterium wangchenii</i> strain dk512	Actinobacteria; Micrococcales; Microbacteriaceae	520	2
48	<i>CHR-Mycolicibacterium insubricum</i> JCM 16366	Actinobacteria; Corynebacteriales; Mycobacteriaceae	508	2
49	<i>CHR-Natronorubrum bangense</i> strain JCM 10635	Euryarchaeota; Stenosarchaea group; HaloNatrialbales; Natrialbaceae	348	1

50	<i>CHR-Nitrobacter hamburgensis</i> X14	Alphaproteobacteria; Rhizobiales; Bradyrhizobiac eae	478	2
51	<i>CHR-Nostoc</i> sp. <i>TCL26-01</i>	Cyanobacteria; Nostocales; Nostocaceae	474	2
52	<i>CHR-Ralstonia solanacearum</i> strain UW774	Betaproteobacteria; Burkholderiales; Burkholderiaceae	476	2
53	<i>CHR-Synechococcus elongatus</i> PCC 6301	Cyanobacteria; Synechococcales; Synechococcaceae	474	2
54	<i>CHR-Thermoplasmata</i> archaeon isolate B28_G1	Euryarchaeota; Diaforarchaea group; Thermoplasmata	436	2
55	<i>CHR-Thermoprotei</i> archaeon B19_G17	Archaea; Crenarchaeota; Thermoprotei	452	2
56	<i>Citrobacter</i> phage Moon	<i>Myoviridae</i> ; <i>Tevenvirinae</i> ; <i>Moonvirus</i>	658	3
57	<i>Clostridium</i> phage phiCDHM13	<i>Myoviridae</i> ; <i>Sherbrookevirus</i>	355	1
58	<i>Clostridium</i> phage phiMMP01	<i>Myoviridae</i> ; <i>Yongloolivirus</i>	436	2
59	<i>Clostridium</i> phage phiMMP04	<i>Myoviridae</i> ; <i>Sherbrookevirus</i>	355	1
60	<i>Cronobacter</i> phage ESSI-2	<i>Myoviridae</i> ; <i>Peduovirinae</i> ; <i>Seongnamvirus</i>	375	1
61	<i>Cyanophage</i> S-RIM44	<i>Myoviridae</i> ; <i>Vellamovirus</i>	635	3
62	<i>Dickeya</i> phage Kamild	<i>Ackermannviridae</i> ; <i>Aglimvirinae</i> ; <i>Limestonevirus</i>	632	3
63	<i>Dickeya</i> phage RC-2014	<i>Ackermannviridae</i> ; <i>Aglimvirinae</i> ; <i>Limestonevirus</i>	632	3
64	<i>Dickeya</i> phage vB_DsoM_JA11	<i>Myoviridae</i> ; <i>Salmondvirus</i>	558	2
65	<i>Enterobacter</i> phage Arya	<i>Myoviridae</i> ; <i>Jilivirus</i>	477	2
66	<i>Enterobacteria</i> phage SEGDI	<i>Myoviridae</i> ; <i>Seoulvirus</i>	681	2
67	<i>Enterobacteria</i> phage SfV	<i>Myoviridae</i>	498	2
68	<i>Enterococcus</i> phage MDA2	<i>Herelleviridae</i> ; <i>Brockvirinae</i> ; <i>Kochikohdavirus</i>	569	2
69	<i>Enterococcus</i> phage PBEF129	<i>Herelleviridae</i> ; <i>Brockvirinae</i> ; <i>Kochikohdavirus</i>	569	3
70	<i>Erwinia</i> phage ENT90	<i>Myoviridae</i> ; <i>Peduovirinae</i> ; <i>Entnonagintavirus</i>	389	1
71	<i>Erwinia</i> phage Machina	<i>Myoviridae</i> ; <i>Machinavirus</i>	680	2
72	<i>Erwinia</i> phage pEa_SNUABM_5	<i>Myoviridae</i> ; <i>Yoloswagvirus</i>	563	2
73	<i>Erwinia</i> phage vB_EamM_Deimos-Minion	<i>Myoviridae</i> ; <i>Agricanvirus</i>	695	2
74	<i>Erwinia</i> phage vB_EamM_Kwan	<i>Myoviridae</i> ; <i>Wellingtonvirus</i>	681	2
75	<i>Erwinia</i> phage vB_EamM_RisingSun	<i>Myoviridae</i> ; <i>Risingsunvirus</i>	713	2
76	<i>Escherichia</i> phage FFH2	<i>Myoviridae</i> ; <i>Vequintavirinae</i>	458	2
77	<i>Escherichia</i> phage Mansfield	<i>Myoviridae</i> ; <i>Wifcevirus</i>	512	2
78	<i>Escherichia</i> phage Mu	<i>Myoviridae</i> ; <i>Muvirus</i>	495	2
79	<i>Escherichia</i> phage P2-4E6b	<i>Myoviridae</i> ; <i>Peduovirinae</i> ; <i>Peduovirus</i>	396	1
80	<i>Escherichia</i> phage PhaxI	<i>Ackermannviridae</i> ; <i>Cvivirinae</i> ; <i>Kuttervirus</i>	631	3
81	<i>Escherichia</i> phage SP27	<i>Myoviridae</i> ; <i>Asteriusvirus</i>	887	4
82	<i>Escherichia</i> phage T4	<i>Myoviridae</i> ; <i>Tevenvirinae</i> ; <i>Tequatrovirus</i>	659	3

83	<i>Escherichia</i> phage vB_EcoM_112	<i>Myoviridae; Tevenvirinae; Tequatrovirus</i>	659	3
84	<i>Escherichia</i> phage vB_EcoM-ep3	<i>Myoviridae; Jilinvirus</i>	475	2
85	<i>Exiguobacterium</i> phage vB_EalM-137	<i>Myoviridae</i>	482	2
86	<i>Faecalibacterium</i> phage FP_Mushu	<i>Myoviridae; Mushuvirus</i>	481	2
87	<i>Faecalibacterium</i> phage FP_Taranis	<i>Myoviridae; Taranisvirus</i>	384	1
88	<i>Flavobacterium</i> phage FPSV-S1	<i>Myoviridae</i>	390	1
89	<i>Gordonia</i> phage GMA6	<i>Myoviridae; Bendigovirus</i>	482	2
90	<i>Haloarcula</i> virus HCTV-6	<i>Myoviridae; Haloferacalesvirus</i>	437	2
91	<i>Haloarcula</i> virus HJTV-1	<i>Myoviridae; Haloferacalesvirus</i>	430	2
92	<i>Halobacterium</i> phage phiH	<i>Myoviridae; Myohalovirus</i>	432	2
93	<i>Halobacterium</i> virus ChaoS9	<i>Myoviridae; Myohalovirus</i>	434	2
94	<i>Halocynthia</i> phage JM-2012	<i>Myoviridae</i>	681	2
95	<i>Haloferax</i> virus HF1	<i>Myoviridae; Haloferacalesvirus</i>	430	2
96	<i>Halomonas</i> virus HAP1	<i>Myoviridae; Hapunavirus</i>	388	1
97	<i>Halorubrum</i> phage HF2	<i>Myoviridae; Haloferacalesvirus</i>	430	2
98	<i>Halorubrum</i> Tailed Virus 5	<i>Myoviridae; Haloferacalesvirus</i>	430	2
99	<i>Halorubrum</i> virus HRTV-25	<i>Myoviridae; Haloferacalesvirus</i>	431	2
100	<i>Halorubrum</i> virus HRTV-27	<i>Myoviridae; Haloferacalesvirus</i>	430	2
101	<i>Klebsiella</i> phage JD18	<i>Myoviridae; Tevenvirinae; Jiaodavirus</i>	657	3
102	<i>Klebsiella</i> phage Miro	<i>Myoviridae; Tevenvirinae; Slopekvirus</i>	663	3
103	<i>Klebsiella</i> phage vB_KleM_RaK2	<i>Myoviridae; Alcyoneusvirus</i>	888	4
104	<i>Kosakonia</i> phage Kc304	<i>Myoviridae; Winklervirus</i>	660	3
105	<i>Lactobacillus</i> phage Lpa804	<i>Herelleviridae; Harbinvirus</i>	612	3
106	<i>Listeria</i> phage LP-124	<i>Herelleviridae; Jasinkavirinae; Pecentumvirus</i>	562	3
107	<i>Listeria</i> phage P200	<i>Herelleviridae; Jasinkavirinae; Pecentumvirus</i>	562	3
108	<i>Meiothermus</i> phage MMP17	<i>Myoviridae</i>	472	2
109	<i>Microcystis</i> phage MaMV-DC	<i>Myoviridae; Fukuivirus</i>	774	4
110	<i>Mycobacterium</i> phage Astraea	<i>Myoviridae; Bixzunavirus</i>	581	3
111	<i>Mycobacterium</i> phage DTDevon	<i>Myoviridae; Bixzunavirus</i>	581	3
112	<i>Mycobacterium</i> phage Phabba	<i>Myoviridae</i>	482	2
113	<i>Natrialba</i> phage PhiCh1	<i>Myoviridae; Myohalovirus</i>	426	2
114	<i>Paenibacillus</i> phage Lily	<i>Lilyvirus; Paenibacillus virus Lily</i>	478	2
115	<i>Pantoea</i> phage vB_PagM_SSEM1	<i>Chaseviridae; Cleopatravirinae; Loessnervirus</i>	483	2
116	phage LacPavin_0818_WC45	<i>metagenome assembly</i>	1283	5+
117	phage Mad1_20_16	<i>metagenome assembly</i>	1248	5+
118	phage RifSed_csp1_19ft_3	<i>metagenome assembly</i>	881	5+
119	<i>PMD-Clostridium baratii</i> str. Sullivan plasmid pCBJ	Firmicutes; Clostridia; Clostridiales; Clostridiaceae	814	4

120	<i>PMD-Deinococcus grandis</i> ATCC 43672 plasmid pDEGR-1	Deinococcus- Thermus; Deinococci; Deinococcales; Deinococcaceae	539	2
121	<i>Prochlorococcus</i> phage P-HM1	<i>Myoviridae</i> ; <i>Eurybiavirus</i>	669	3
122	<i>Pseudomonas</i> phage PaBG	<i>Myoviridae</i> ; <i>Baikalvirus</i>	547	2
123	<i>Pseudomonas</i> phage phiKZ	<i>Myoviridae</i> ; <i>Phikzvirus</i>	695	2
124	<i>Pseudomonas</i> phage PPSC2	<i>Myoviridae</i> ; <i>Otagovirus</i>	427	2
125	<i>Ralstonia</i> phage phiRSL1	<i>Myoviridae</i> ; <i>Mieseafarmvirus</i>	648	3
126	<i>Ralstonia</i> phage RSP15	<i>Ackermannviridae</i>	826	5+
127	<i>Ralstonia</i> phage RSY1	<i>Myoviridae</i> ; <i>Peduvirinae</i> ; <i>Aresaunavirus</i>	391	1
128	<i>Salicola</i> phage SCTP-2	<i>Myoviridae</i>	955	4
129	<i>Serratia</i> phage phiMAM1	<i>Ackermannviridae</i> ; <i>Miltonvirus</i>	636	3
130	<i>Serratia</i> phage PS2	<i>Myoviridae</i> ; <i>Muldoonvirus</i>	663	3
131	<i>Shewanella</i> phage vB_SspM_MuM16-2	<i>Myoviridae</i>	493	2
132	<i>Sinorhizobium</i> phage phiM9	<i>Ackermannviridae</i>	838	5+
133	<i>Staphylococcus</i> phage 812	<i>Herelleviridae</i> ; <i>Twortvirinae</i> ; <i>Kayvirus</i>	587	3
134	<i>Staphylococcus</i> phage A3R	<i>Herelleviridae</i> ; <i>Twortvirinae</i> ; <i>Kayvirus</i>	587	3
135	<i>Staphylococcus</i> phage vB_SauM_Remus	<i>Herelleviridae</i> ; <i>Twortvirinae</i> ; <i>Silviavirus</i>	586	3
136	<i>Synechococcus</i> phage ACG-2014f	<i>Myoviridae</i> ; <i>Atlauavirus</i>	731	3
137	<i>T6SS-3J9G-Vibrio cholerae</i>	Gammaproteobacteria; Vibrionales; Vibrionaceae	432	1
138	<i>T6SS-3J9O-Francisella tularensis</i> subsp. <i>novicida</i>	Gammaproteobacteria; Thiotrichales; Francisellaceae	506	1
139	<i>T6SS-5N8N-Pseudomonas aeruginosa</i>	Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae	498	1
140	<i>T6SS-Agrobacterium fabrum</i> C58	Alphaproteobacteria; Hyphomicrobiales; Rhizobiaceae	493	1
141	<i>T6SS-Burkholderia</i> sp. MSMB0852	Betaproteobacteria; Burkholderiales; Burkholderiaceae; Burkholderia	499	1
142	<i>T6SS-Limnoglobus roseus</i> strain PX52	Planctomycetes; Planctomycetia; Gemmatales; Gemmataceae	491	1
143	<i>T6SS-Luteitalea pratensis</i> DSM 100886	Acidobacteria; Vicinamibacteria; Vicinamibacteraceae	493	1
144	<i>T6SS-Rhodothermaceae</i> bacterium RA	Bacteroidetes; Bacteroidetes Order II. Incertae sedis; Rhodothermaceae	509	1
145	<i>Tenacibaculum</i> phage PTm1	<i>Myoviridae</i> ; <i>Shirahamavirus</i>	1032	3
146	<i>Thermus</i> phage phiYS40	<i>Myoviridae</i>	648	3
147	<i>Vibrio</i> phage 1.093.O._10N.286.55.E10	<i>Myoviridae</i>	486	2
148	<i>Vibrio</i> phage 1.111.A._10N.286.45.E6	<i>Myoviridae</i>	378	1
149	<i>Vibrio</i> phage BONAISHI	<i>Myoviridae</i>	682	2
150	<i>Vibrio</i> phage vB_VpaM_MAR	<i>Myoviridae</i> ; <i>Vhmlvirus</i>	386	1
151	<i>Vibrio</i> phage YC	<i>Ackermannviridae</i> ; <i>Campanilevirus</i>	756	4
152	<i>Wolbachia</i> phage WO	<i>Myoviridae</i>	383	1

153	<i>Xanthomonas</i> phage XacN1	<i>Myoviridae</i>	714	3
-----	--------------------------------	-------------------	-----	---



Рисунок Д.2. Модель чехольного белка фага *Bacillus* phages PBS1 (изображена пурпурно-голубым) с наложенной моделью ЧБ фага *Burkholderia* БЕК (показанной светло-красным цветом).

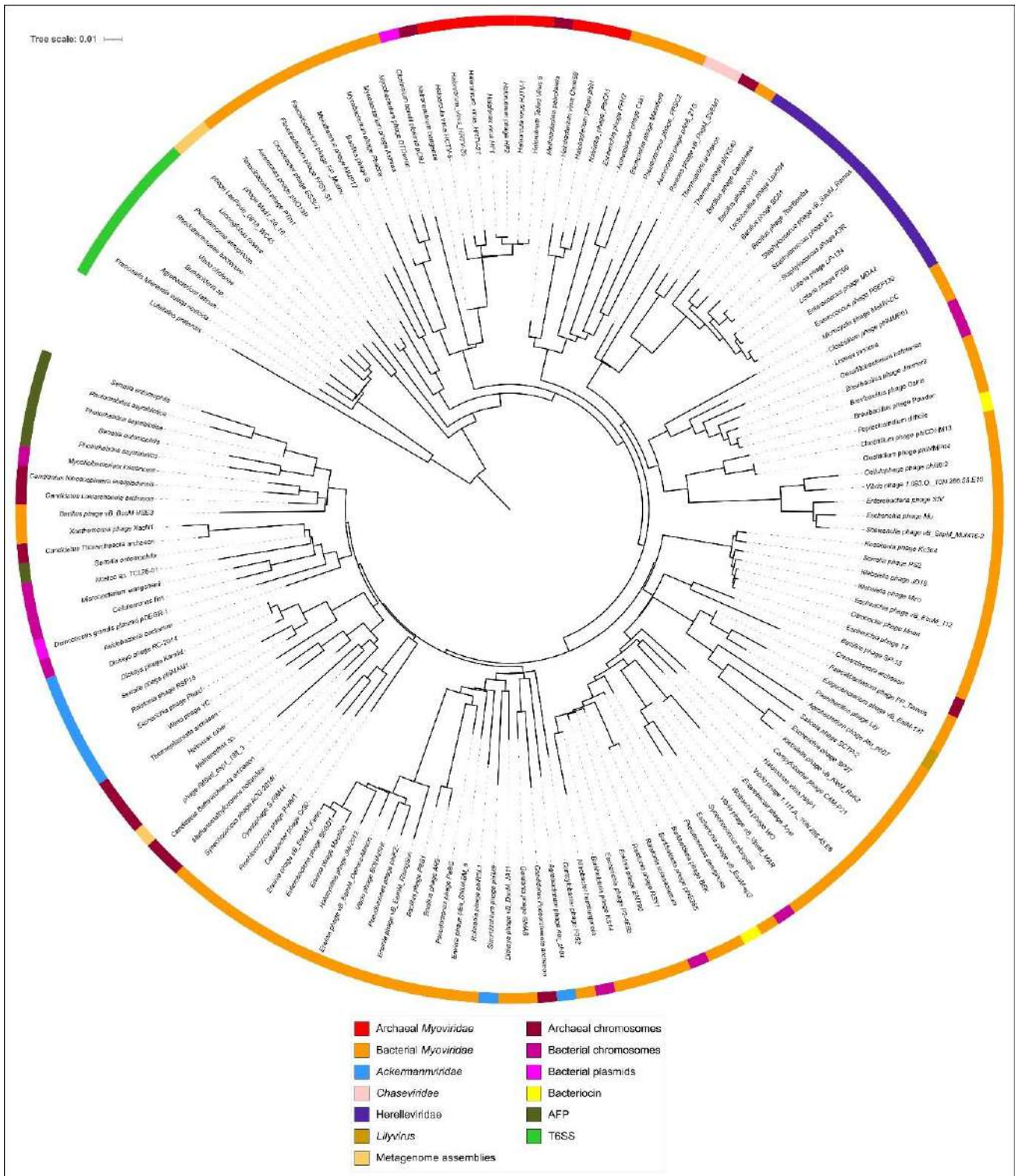


Рисунок Д.3. Круговая дендрограмма, построенная на основании структурного сходства 153 чехольных белков. Рисунок в высоком разрешении доступен в сети Интернет в высоком разрешении по адресу <https://www.mdpi.com/1999-4915/14/6/1148/s1?version=1653559708>.

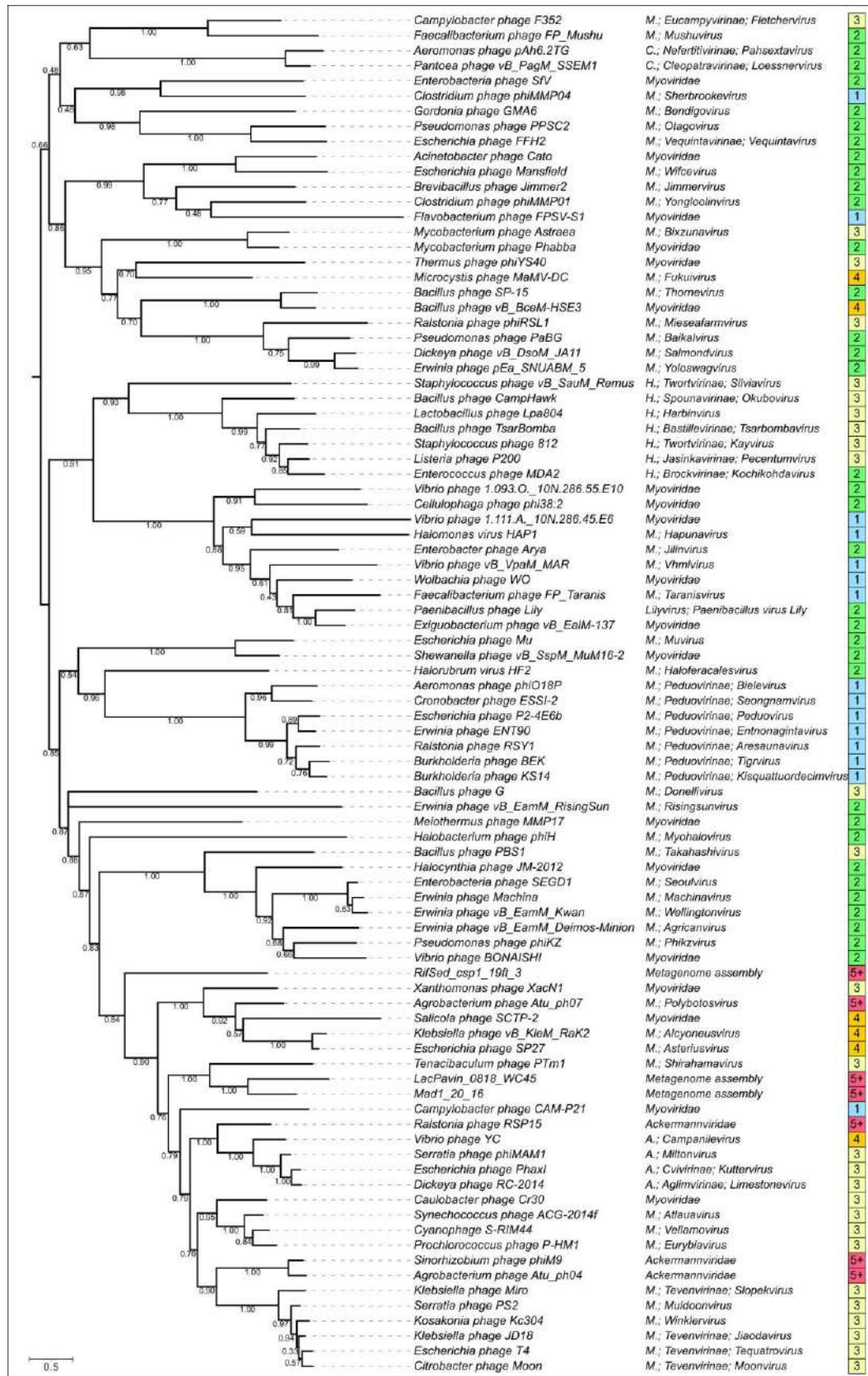


Рисунок Д.4. Филогенетическое дерево, полученное с использованием 90 аминокислотных последовательностей большой субъединицы терминазы. Таксономия указана справа от названия, число доменов ЧБ обозначено цифрой в квадрате.

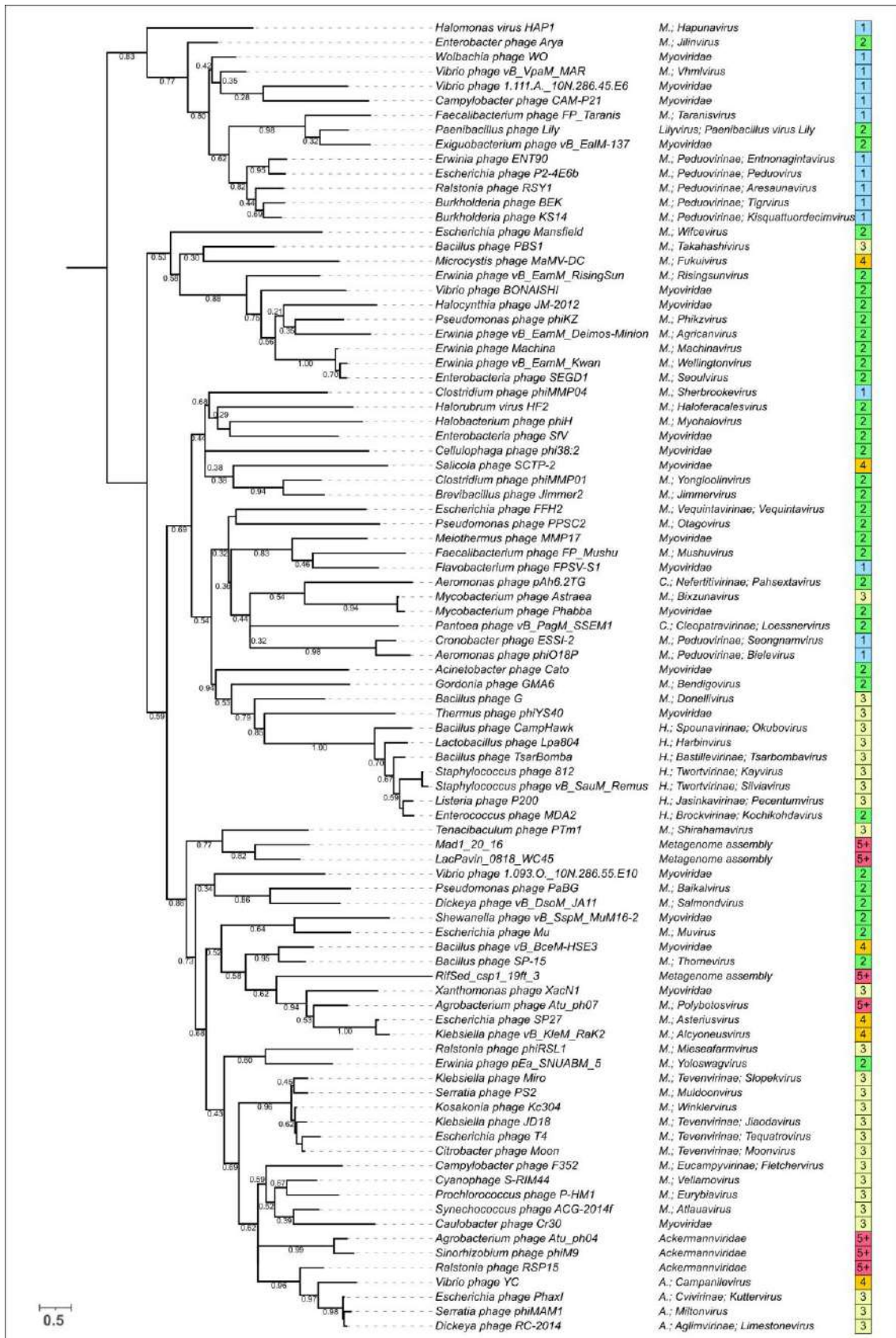


Рисунок Д.5. Филогенетическое дерево, полученное с использованием 90 аминокислотных последовательностей белка хвостовой трубки. Таксономия указана справа от названия, число доменов ЧБ обозначено цифрой в квадрате.

ПРИЛОЖЕНИЕ Е

Использование AlphaFold для классификации бактериофагов

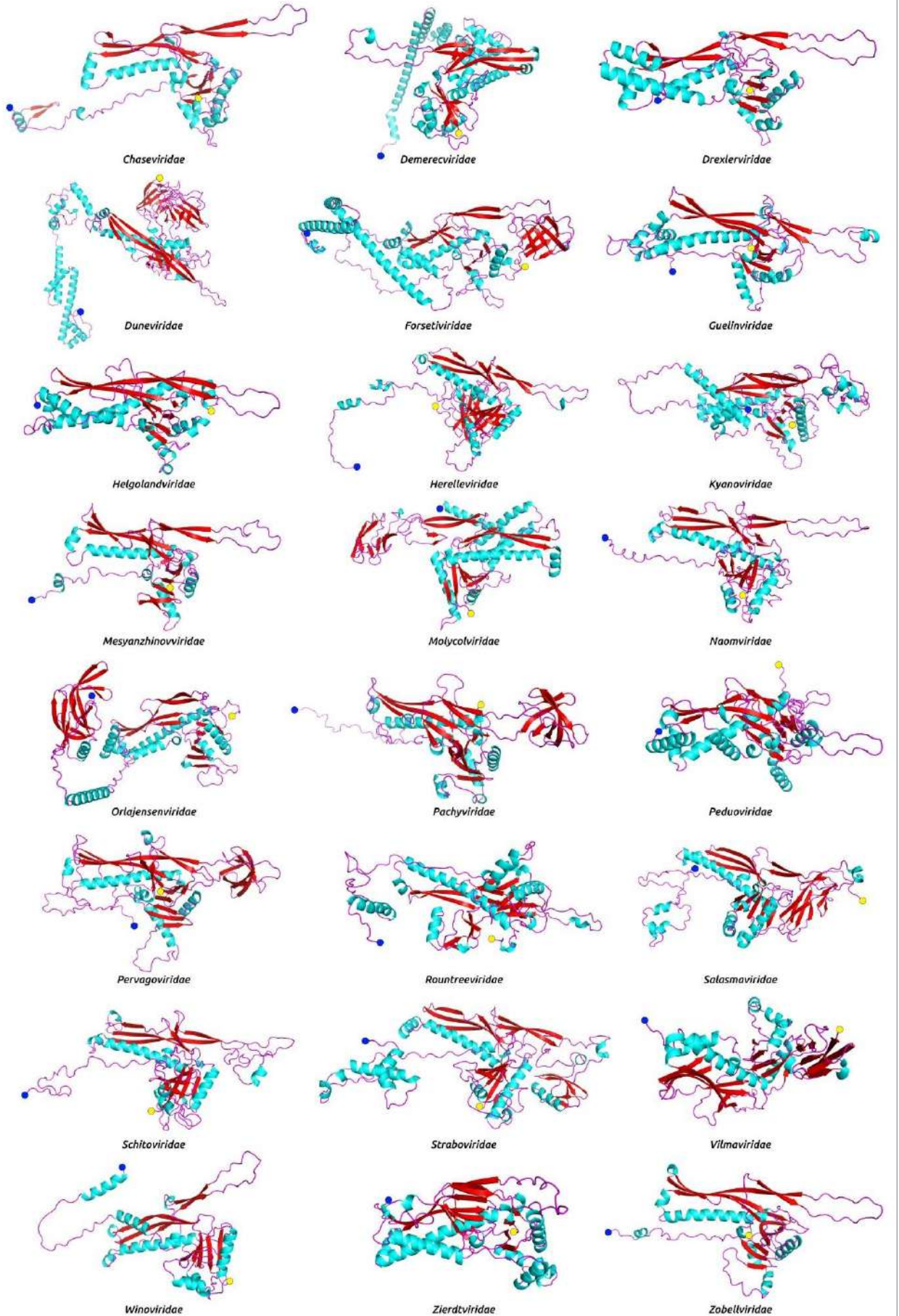
Таблица Е.1. Список вирусов, отобранных для сравнительного структурного анализа ГКП и терминазы. Для эукариотических вирусов использован синий фон, бактериальных – на зелёный, архейных – на жёлтый.

Вирус, название по номенклатуре 2022 г.	Вирус, оригинальное название	Таксономия по ICTV	Размер генома, н.п.	ГЦ-состав, %	Код NCBI
<i>Ranid herpesvirus 1</i>	<i>Lucke tumor herpesvirus-ranid herpesvirus 1</i>	<i>Herpesvirales; Alloherpesviridae</i>	220859	54,6	DQ665917.1
<i>Human alphaherpesvirus 1</i>	<i>Human herpesvirus 1 strain 17</i>	<i>Herpesvirales; Herpesviridae</i>	152222	68,3	JN555585.1
<i>Haliotid herpesvirus 1</i>	<i>Abalone herpesvirus Victoria/AUS/2009</i>	<i>Herpesvirales; Malacoherpesviridae</i>	211518	46,8	JX453331.1
<i>Curtobacterium phage Ayka</i>	<i>Curtobacterium phage Ayka</i>	<i>not classified</i>	18400	52,6	ON381767.1
<i>LacPavin</i>	<i>LacPavin_0818_WC45</i>	<i>not classified</i>	735411	32,2	LR756501.1
<i>Pseudomonas phage MD8</i>	<i>Pseudomonas phage MD8</i>	<i>not classified</i>	43277	61,1	KX198612.1
<i>Limestonevirus limestone</i>	<i>Dickeya phage vB-DsoM-LIMEstone1</i>	<i>Ackermannviridae</i>	152427	49,3	HE600015.1
<i>Harrekavirus harreka</i>	<i>Olleya phage Harreka_1</i>	<i>Aggregaviridae</i>	43175	32	MT732457.1
<i>Cebadecemvirus phi10una</i>	<i>Cellulophaga phage phi10:1</i>	<i>Assiduviridae</i>	53664	31,5	KC821618.1
<i>Teseptimavirus T7</i>	<i>Escherichia phage T7</i>	<i>Autographiviridae</i>	39937	48,4	V01146.1
<i>Chivirus chi</i>	<i>Salmonella phage χ</i>	<i>Casjensviridae</i>	59407	56,5	JX094499.1
<i>Lambdavirus lambda</i>	<i>Escherichia phage λ</i>	<i>Lambdavirus</i>	48502	49,9	J02459.1
<i>Suwonvirus PP101</i>	<i>Pectobacterium phage PP101</i>	<i>Chaseviridae</i>	53333	44,9	KY087898.2
<i>Junduvirus communis</i>	uncultured phage cr2_1	<i>Crassvirales; Crevaviridae</i>	95815	32,7	MZ130489.1

<i>Jahgtovirus gastrointestinalis</i>	uncultured phage cr36_1	<i>Crassvirales; Intestiviridae</i>	96466	32	MZ130479.1
<i>Kahnovirus copri</i>	uncultured phage cr44_1	<i>Crassvirales; Steigviridae</i>	93564	35,8	MZ130483.1
<i>Afonbuvirus coli</i>	uncultured phage cr35_1	<i>Crassvirales; Suoliviridae</i>	97706	31,4	MZ130499.1
<i>Cetovirus ceto</i>	<i>Vibrio</i> phage Ceto	<i>Demereviridae</i>	128241	39,9	MG649966.1
<i>Donellivirus gee</i>	<i>Bacillus</i> phage G	<i>Donellivirus</i>	497513	29,9	JN638751.1
<i>Tunavirus T1</i>	<i>Escherichia</i> phage T1	<i>Drexlerviridae</i>	48836	45,6	AY216660.1
<i>Unahavirus uv1H</i>	<i>Flavobacterium</i> phage 1H	<i>Duneviridae</i>	39290	31,4	KU599889.1
<i>Freyavirus freya</i>	<i>Polaribacter</i> phage Freya_1	<i>Forsetiviridae</i>	43978	28,9	MT732463.1
<i>Gregsiragusavirus CPS1</i>	<i>Clostridium</i> phage CPS1	<i>Guelinviridae</i>	19089	28,3	KY996523.1
<i>Leefvirus Leef</i>	<i>Polaribacter</i> phage Leef_1	<i>Helgolandviridae</i>	37547	29,7	MT732473.1
<i>Byrnievirus HK97</i>	<i>Escherichia</i> phage HK97	<i>Hendrixvirinae</i>	39732	49,8	AF069529.1
<i>Pecentumvirus P100</i>	<i>Listeria</i> phage P100	<i>Herelleviridae</i>	131384	36	DQ004855.1
<i>Beejeyvirus BJ1</i>	<i>Halorubrum</i> virus BJ1	<i>Kirjokansivirales; Graaviviridae</i>	42271	64,9	AM419438.1
<i>Retbasiphovirus HFTV1</i>	<i>Haloferax</i> tailed virus 1	<i>Kirjokansivirales; Haloferuviridae</i>	38059	54,1	MG550112.1
<i>Hatrivirus HATV3</i>	<i>Haloarcula</i> tailed virus 3	<i>Kirjokansivirales; Pyrstoviridae</i>	42293	51,1	MZ334527.1
<i>Lonfivirus HSTV1</i>	<i>Haloarcula sinaiensis</i> tailed virus 1	<i>Kirjokansivirales; Shortaselviridae</i>	32189	60,3	KC117378.1
<i>Bellamyvirus bellamy</i>	<i>Synechococcus</i> phage Bellamy	<i>Kyanoviridae</i>	204930	41,1	MF351863.1
<i>Clampvirus HHTV1</i>	<i>Haloarcula hispanica</i> tailed virus 1	<i>Madisaviridae</i>	49107	56,5	KC292025.1
<i>Pseudomonas virus Yua</i>	<i>Pseudomonas</i> phage YuA	<i>Mesyanzhinovviridae</i>	58663	64,3	AM749441.1
<i>Metforvirus Drs3</i>	<i>Methanobacterium</i> virus Drs3	<i>Methanobavirales; Anaerodiviridae</i>	37129	41,2	MH674343.1
<i>Psimunavirus psiM2</i>	<i>Methanobacterium</i> phage psiM2	<i>Methanobavirales; Leisingerviridae</i>	26111	46,3	AF065411.1
<i>Mollyvirus colly</i>	<i>Maribacter</i> phage Colly_1	<i>Molycolviridae</i>	124169	36,3	MT732450.1

<i>Noahvirus arc</i>	Bacteriophage DSS3_VP1	<i>Naomviridae</i>	75087	47,5	MN602266.1
<i>Bonaevitae bonaevitae</i>	<i>Microbacterium</i> phage BonaeVitae	<i>Orlajensenviridae</i>	17451	68,2	MH045556.1
<i>Bacelvirus phi46tres</i>	<i>Cellulophaga</i> phage phi46:3	<i>Pachyviridae</i>	72961	32,7	KC821622.1
<i>Peduovirus P2</i>	Bacteriophage P2	<i>Peduoviridae</i> ; <i>Peduovirus</i>	33593	50,2	AF063097.1
<i>Callevirus Calle</i>	<i>Cellulophaga</i> phage Calle_1	<i>Pervagoviridae</i>	72979	38,1	MT732432.1
<i>Phikzvirus phiKZ</i>	<i>Pseudomonas</i> phage phiKZ	<i>Phikzvirus</i>	280334	36,8	AF399011.1
<i>Rosenblumvirus rv66</i>	Bacteriophage 66	<i>Rountreeviridae</i>	18199	29,3	AY954949.1
<i>Salasvirus phi29</i>	<i>Bacillus</i> phage phi29	<i>Salasmaviridae</i>	19282	40	EU771092.1
<i>Halohivirus HHTV2</i>	<i>Haloarcula hispanica</i> tailed virus 2	<i>Saparoviridae</i>	52643	66,6	KC292024.1
<i>Enquatrovirus N4</i>	<i>Escherichia</i> phage N4	<i>Schitoviridae</i>	70153	41,3	EF056009.1
<i>Tequatrovirus T4</i>	<i>Escherichia</i> phage T4	<i>Straboviridae</i>	168903	35,3	AF158101.6
<i>Pormufvirus HRTV28</i>	<i>Halorubrum</i> tailed virus 28 isolate HRTV-28/28	<i>Suolaviridae</i>	35270	64,3	MZ334528.1
<i>Hacavirus HCTV1</i>	<i>Haloarcula californiae</i> tailed virus 1	<i>Thumleimavirales</i> ; <i>Druskaviridae</i>	103257	57	KC292029.1
<i>Haloferacalesvirus HF1</i>	<i>Halophage</i> HF1	<i>Thumleimavirales</i> ; <i>Hafunaviridae</i>	75898	55,8	AY190604.2
<i>Hagravirus HGTV1</i>	<i>Halogramum</i> tailed virus 1	<i>Thumleimavirales</i> ; <i>Halomagnusviridae</i>	143855	50,4	KC292026.1
<i>Eilatmyovirus HATV2</i>	<i>Haloarcula</i> tailed virus 2	<i>Thumleimavirales</i> ; <i>Soleiviridae</i>	63301	49,7	MZ334525.1
<i>Myohalovirus phiH</i>	<i>Halobacterium</i> phage phiH	<i>Vertoviridae</i>	58072	63,7	MK002701.1
<i>Bromdenvirus bromden</i>	<i>Mycobacterium</i> phage Bromden	<i>Vilmaviridae</i>	70183	58,2	MH576973.1
<i>Peternellavirus peternella</i>	<i>Winogradskyella</i> phage Peternella_1	<i>Winoviridae</i>	39649	35,4	MT732475.1
<i>Foxborovirus foxboro</i>	<i>Gordonia</i> phage Foxboro	<i>Zierdtviridae</i>	67773	65,8	MH727547.1
<i>Siovirus americense</i>	<i>Roseobacter</i> phage SIO1	<i>Zobellviridae</i>	39898	46,2	AF189021.1

bacterial *Uroviricota; Caudoviricetes* Viral families not assigned to orders



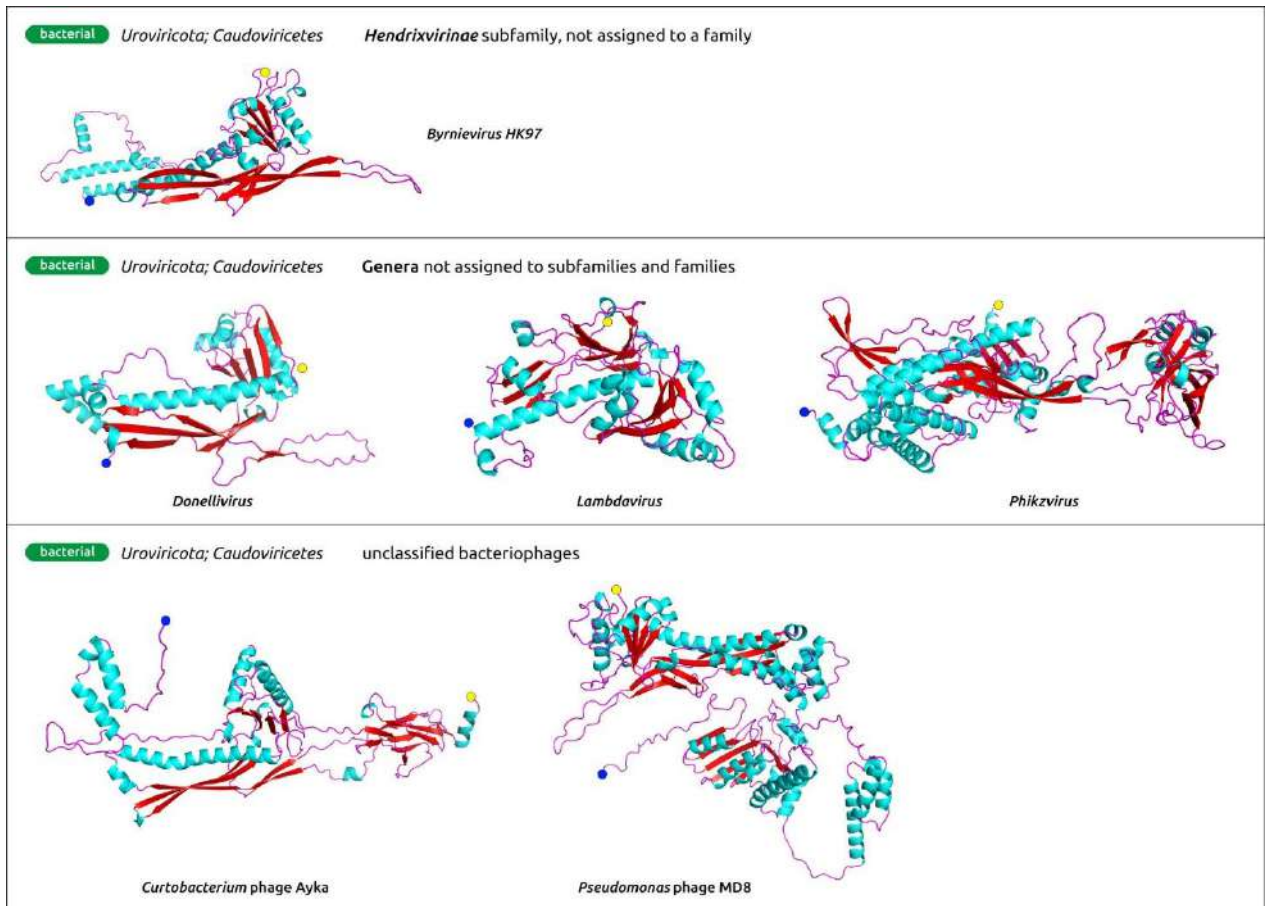
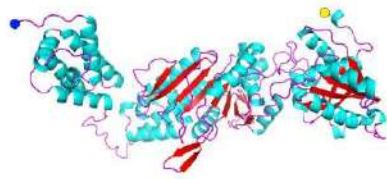
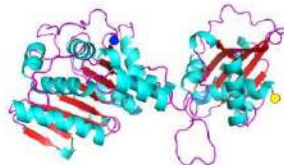


Рисунок Е.1. Структурные модели белков, кодируемых генами главного капсидного белка, полученные с помощью AlphaFold 2. N-конец полипептидной цепи отмечен синей, а С-конец – жёлтой меткой.

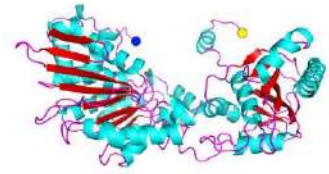
bacterial Uroviricota; Caudoviricetes Viral families not assigned to orders



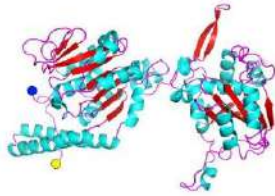
Chaseviridae



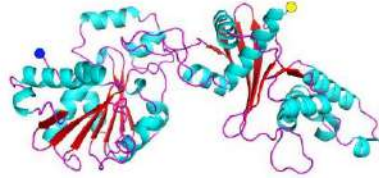
Demereviridae



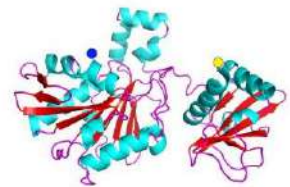
Drexelviriidae



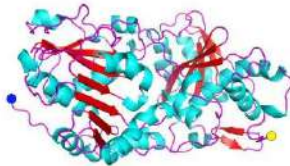
Duneviridae



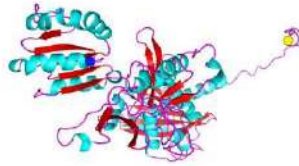
Forsetiviridae



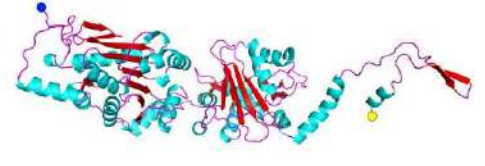
Guelniviriidae



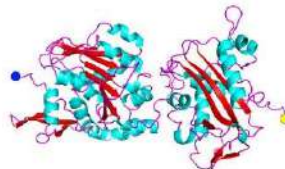
Helgolandviridae



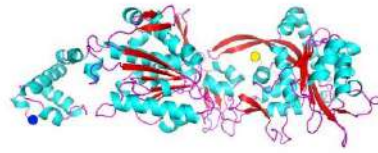
Herelleviridae



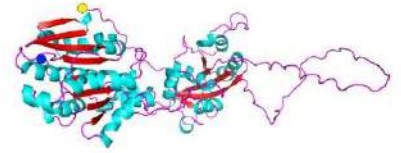
Kyanoviridae



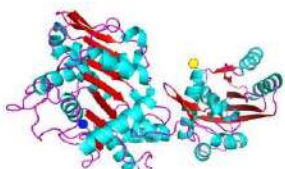
Mesyanzhinovviridae



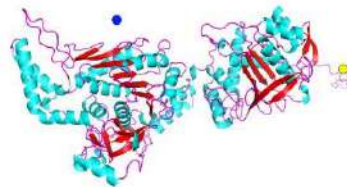
Molycolviridae



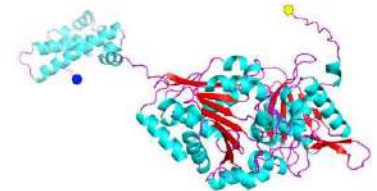
Naomviridae



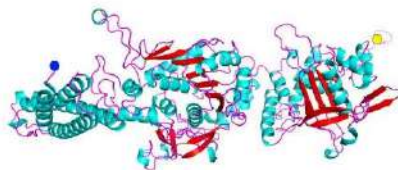
Orlajensenviridae



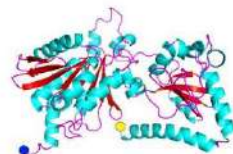
Pachyviridae



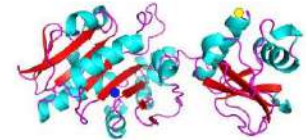
Peduoviridae



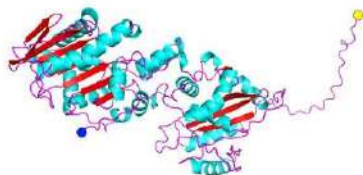
Pervagoviridae



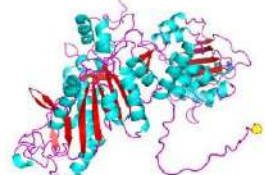
Rountreeviridae



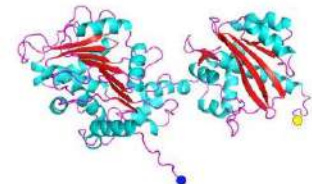
Salasmaviridae



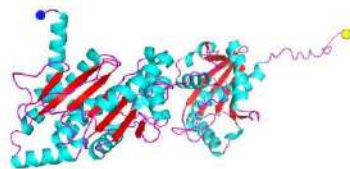
Schitoviridae



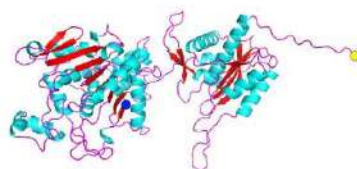
Straboviridae



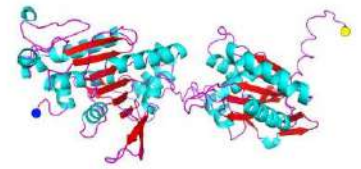
Vilmaviridae



Winoviridae



Zierdtviridae



Zobellviridae

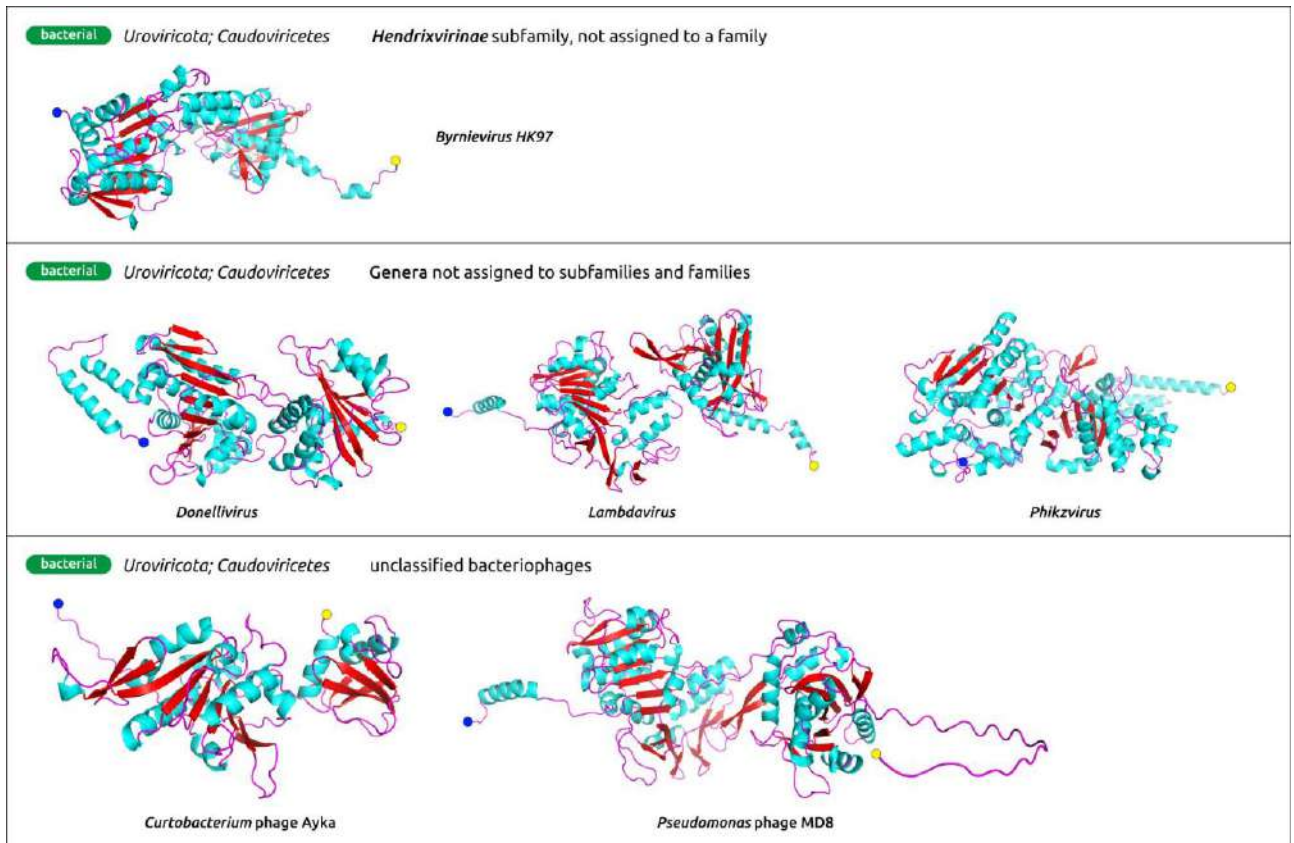


Рисунок Е.2. Структурные модели белков, кодируемых генами терминазы, полученные с помощью AlphaFold 2. N-конец полипептидной цепи отмечен синей, а С-конец – жёлтой меткой.

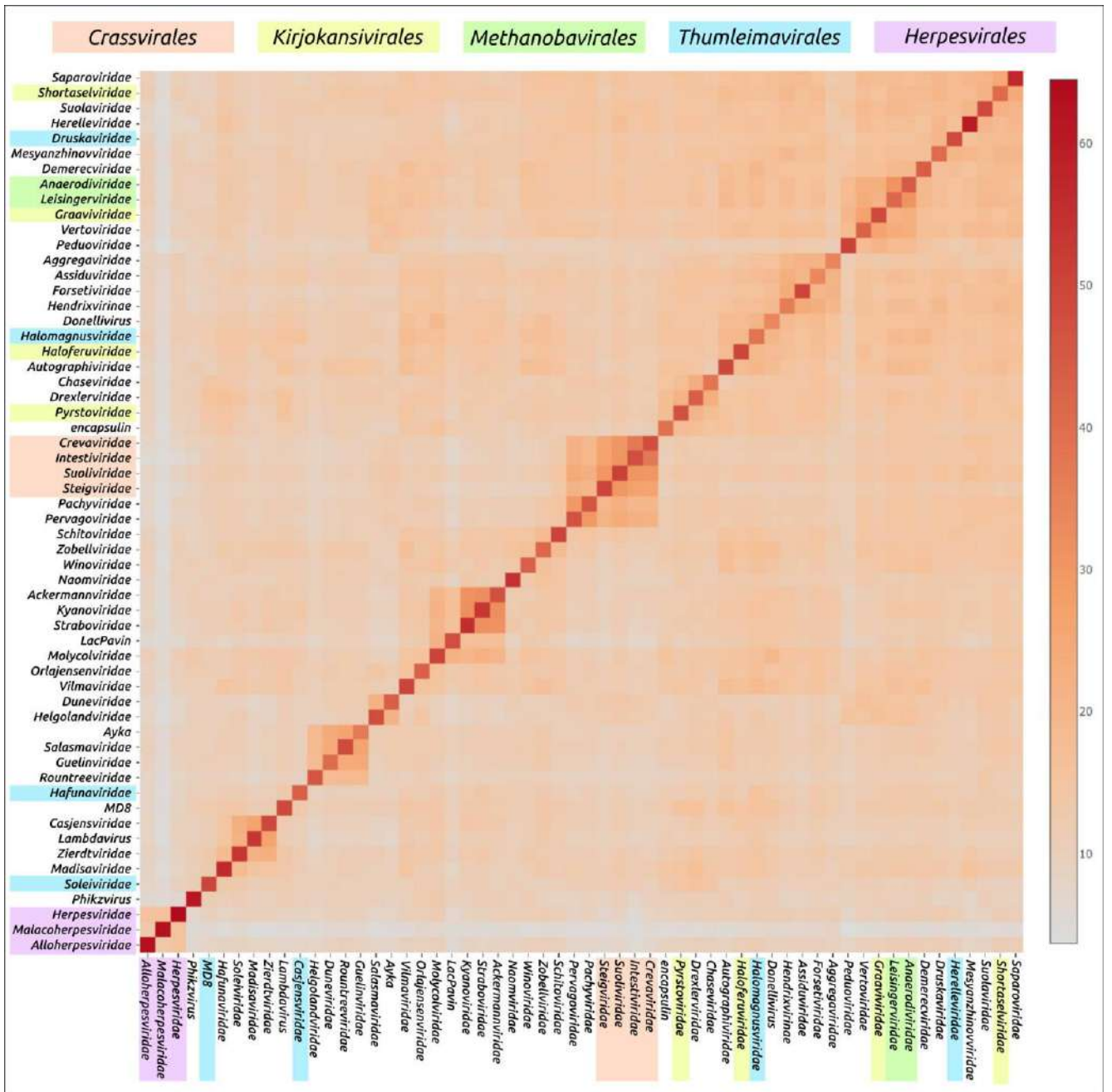


Рисунок Е.3. Кластерная тепловая карта, основанная на попарном сравнении с помощью DALI Z-score 57 моделей главных капсидных белков и моделей энкапсулина (с удалёнными доменами протеазы и белка скаффолда).

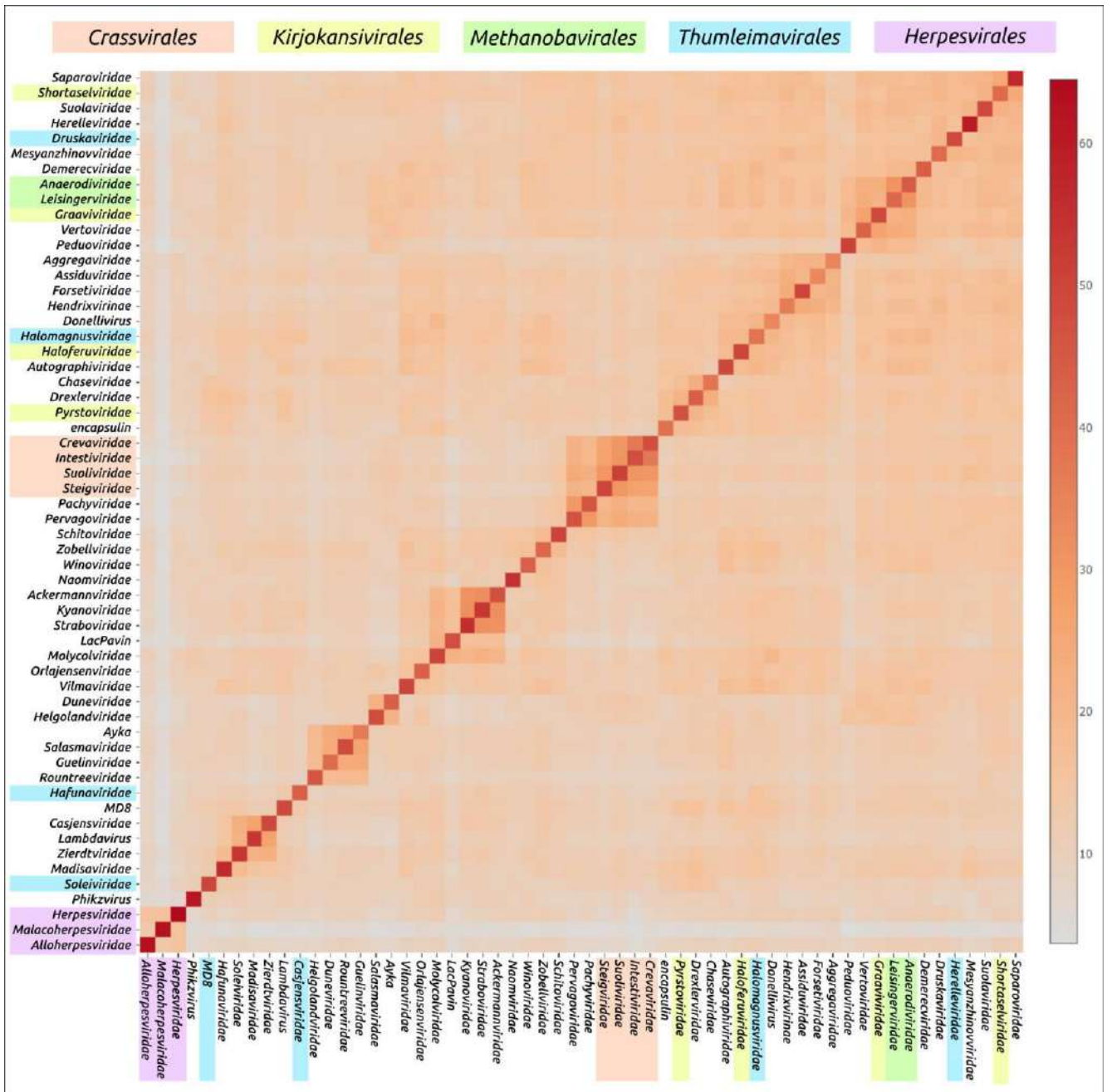


Рисунок Е.4. Кластерная тепловая карта, основанная на попарном сравнении с помощью DALI Z-score 58 моделей АТФазного домена терминазы.

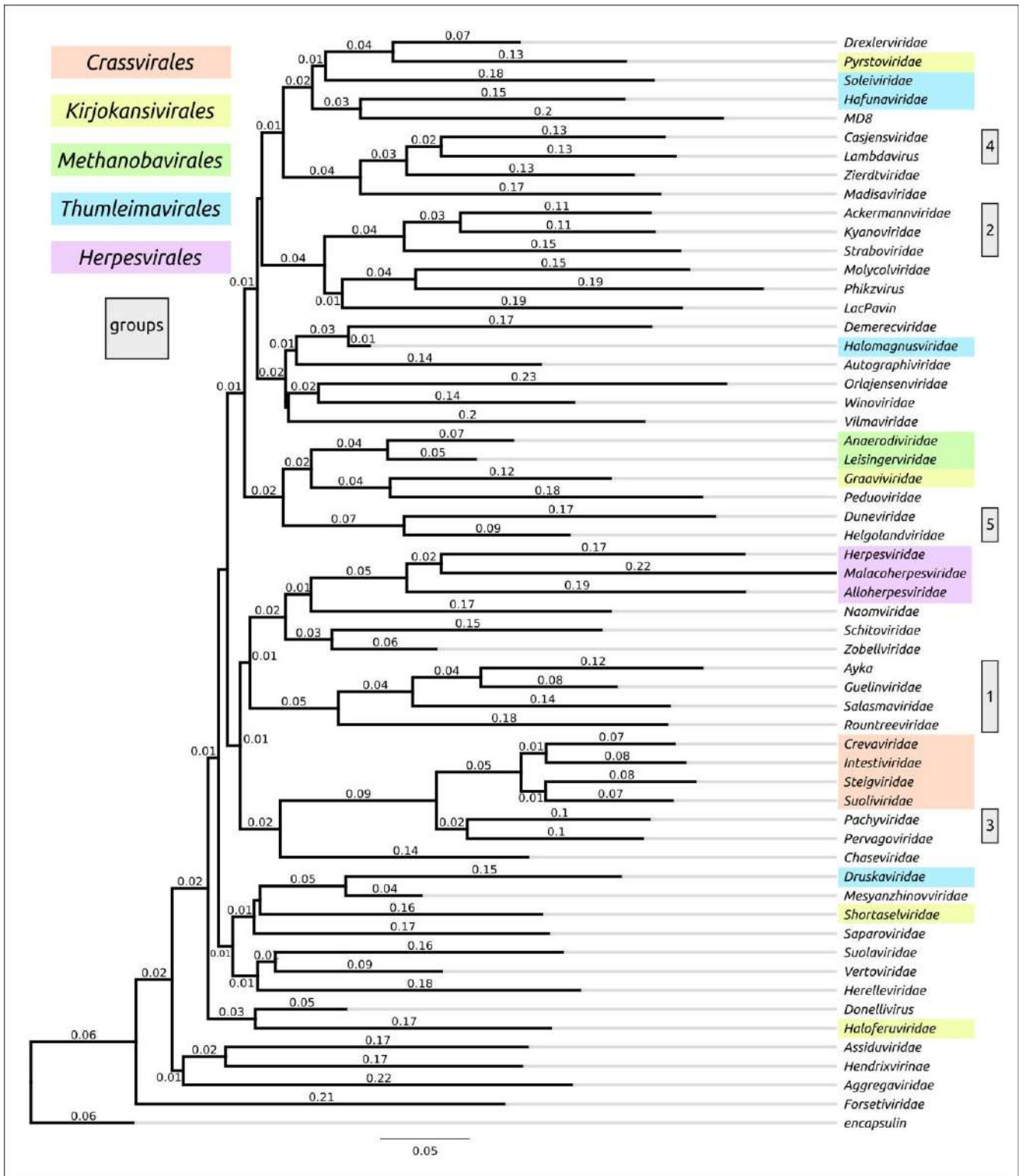


Рисунок Е.5. Дендрограмма, основанная на сравнении с помощью mTM-align 57 моделей ГКБ и энкапсулина.

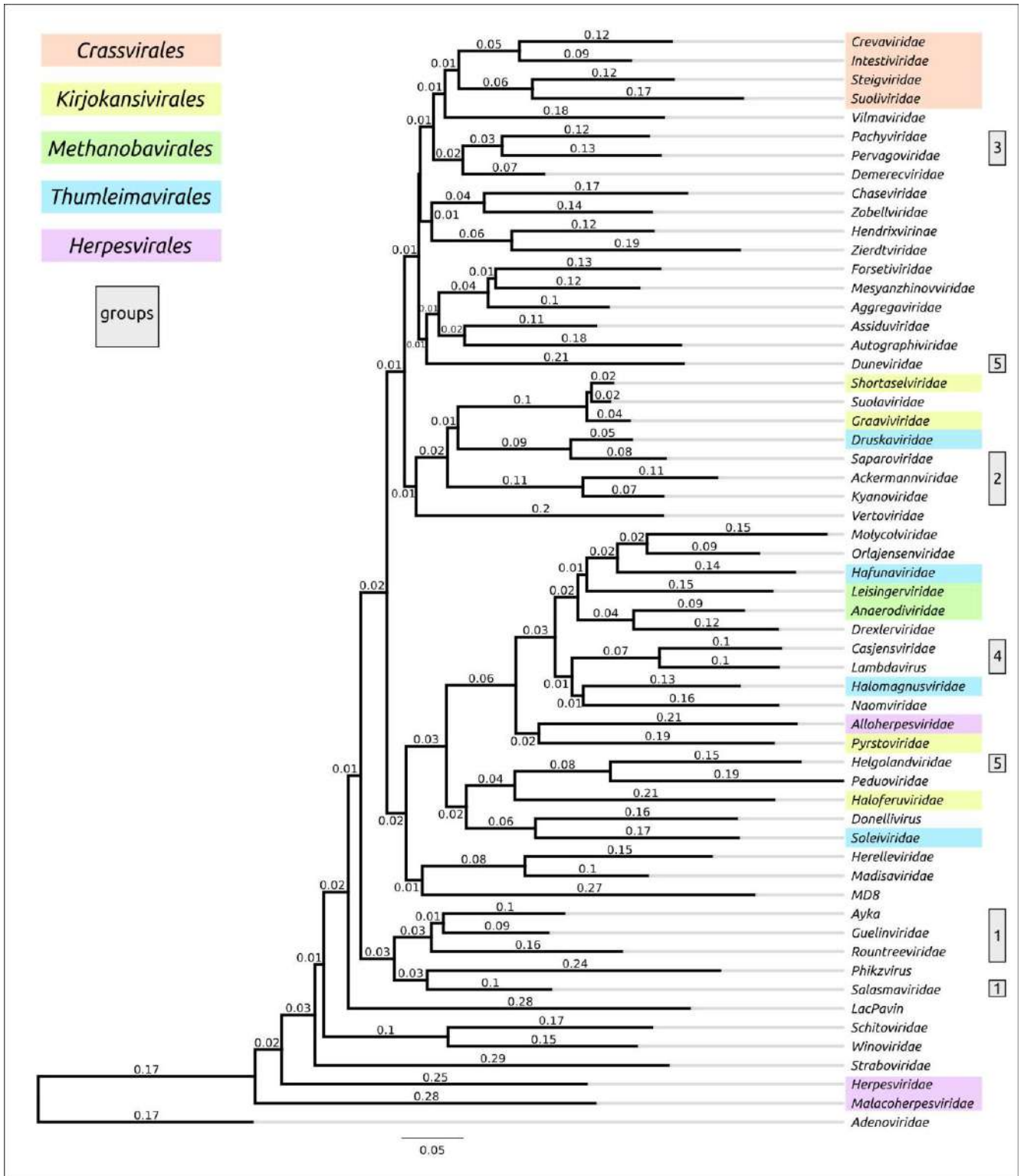


Рисунок Е.6. Дендрограмма, основанная на сравнении с помощью mTM-align 58 моделей терминазы.

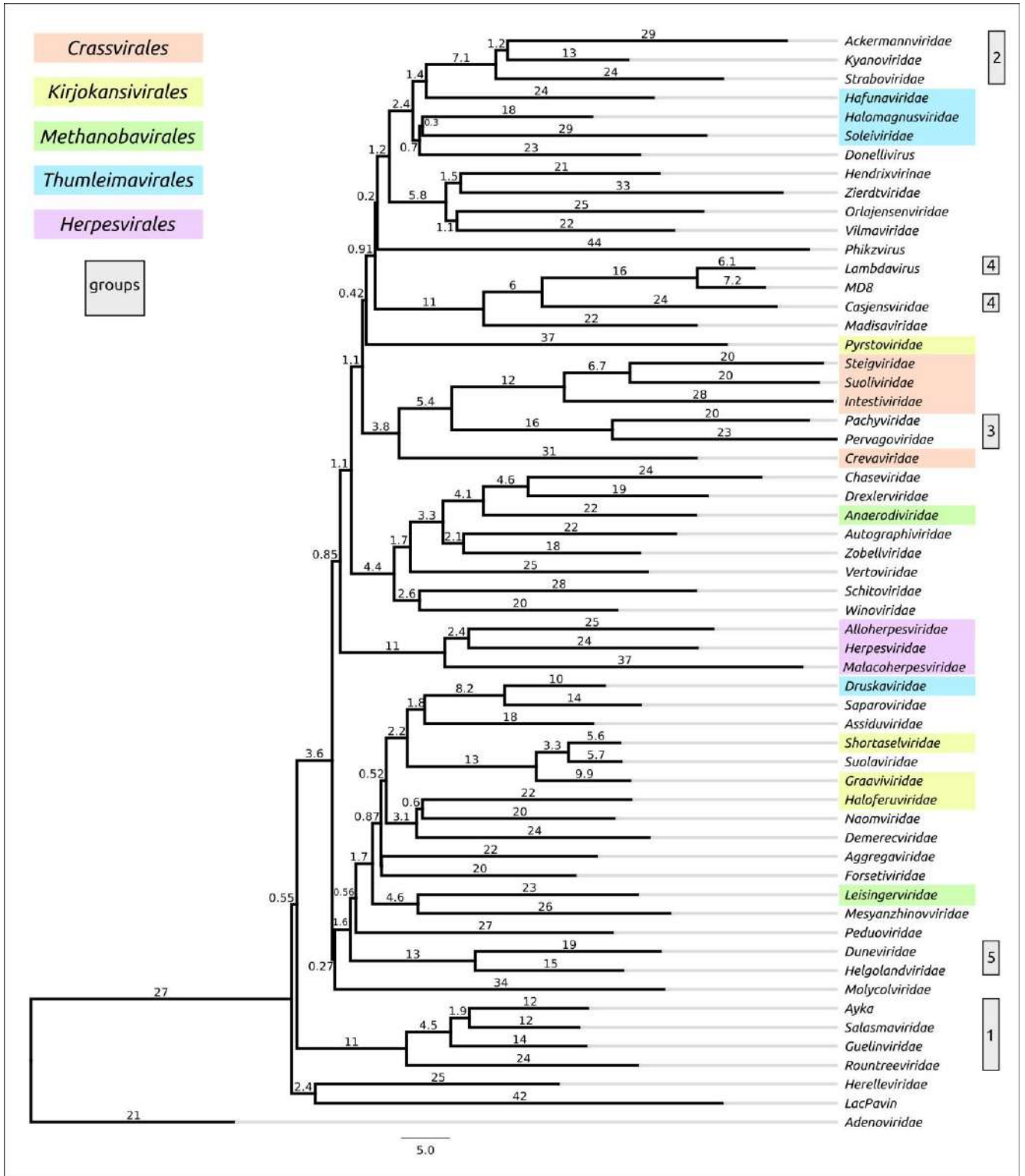


Рисунок Е.7. Дендрограмма, основанная на сравнении с помощью mTM-align 58 моделей АТФазной субъединицы терминазы.

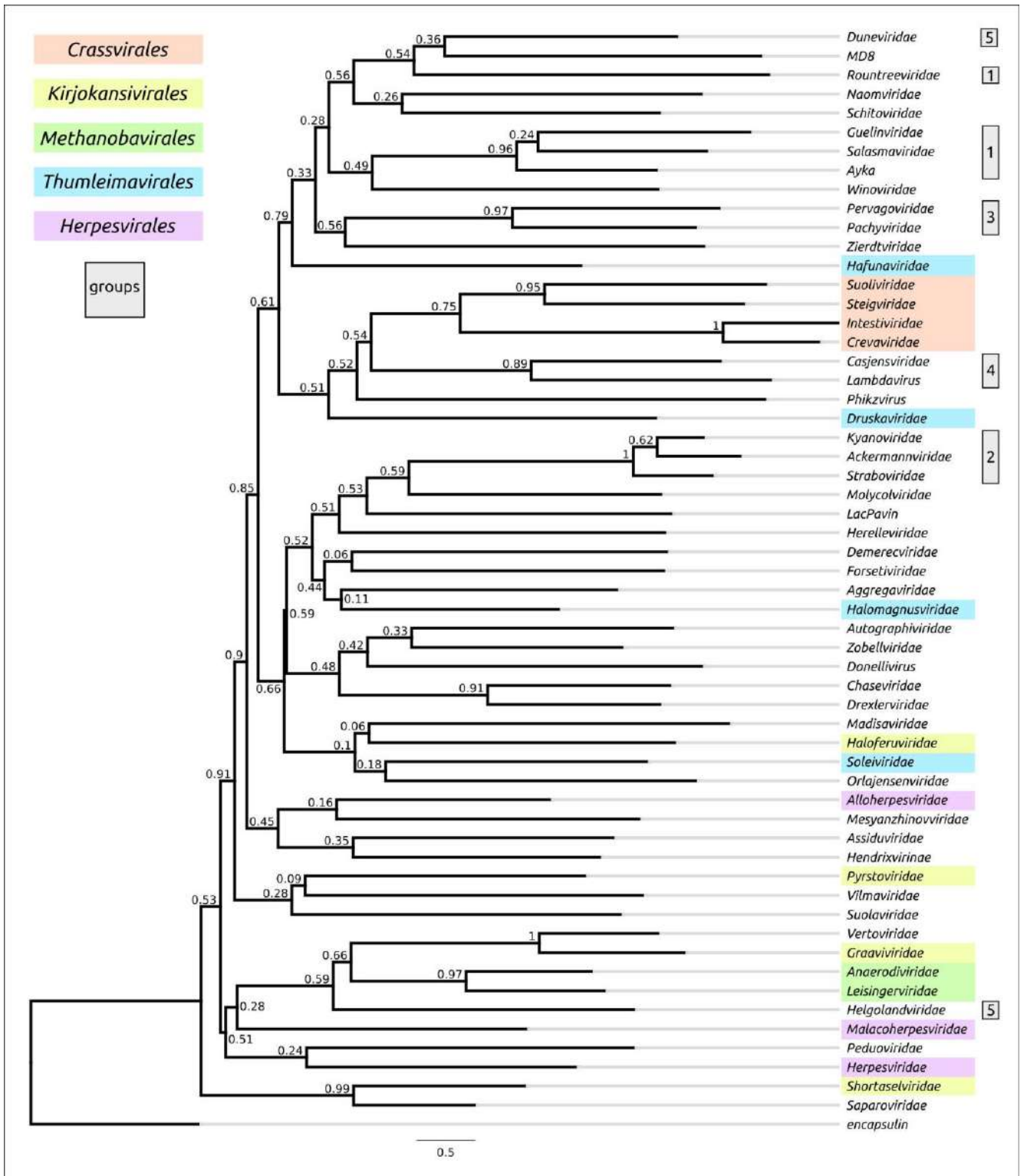


Рисунок Е.8. Филогенетическое дерево, построенное с использованием выравнивания 57 аминокислотных последовательностей главного капсидного белка и энкапсулина с помощью MAFFT. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

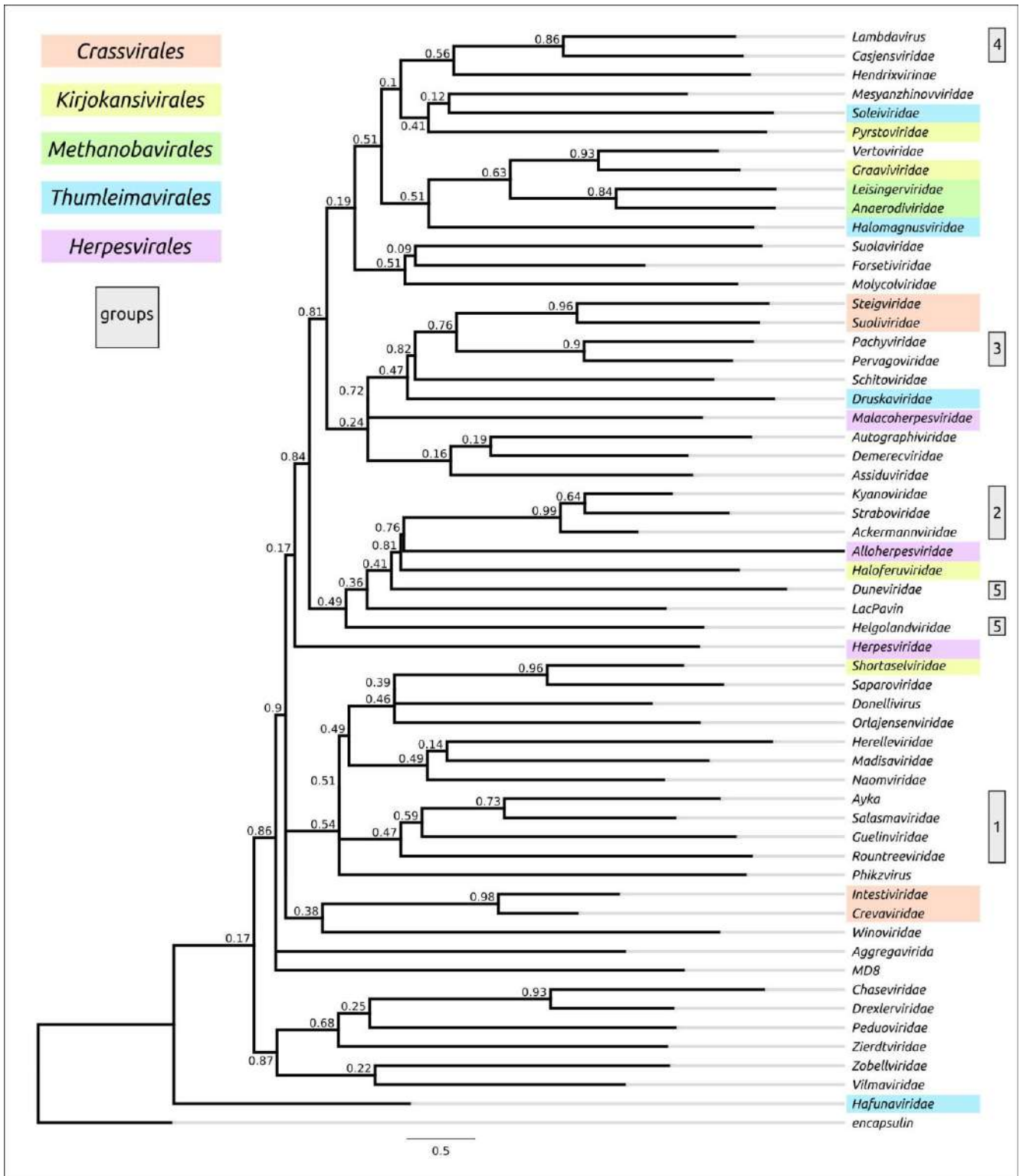


Рисунок Е.9. Филогенетическое дерево, построенное с использованием выравнивания 57 аминокислотных последовательностей главного капсидного белка и энкапсулина с помощью MUSCLE. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

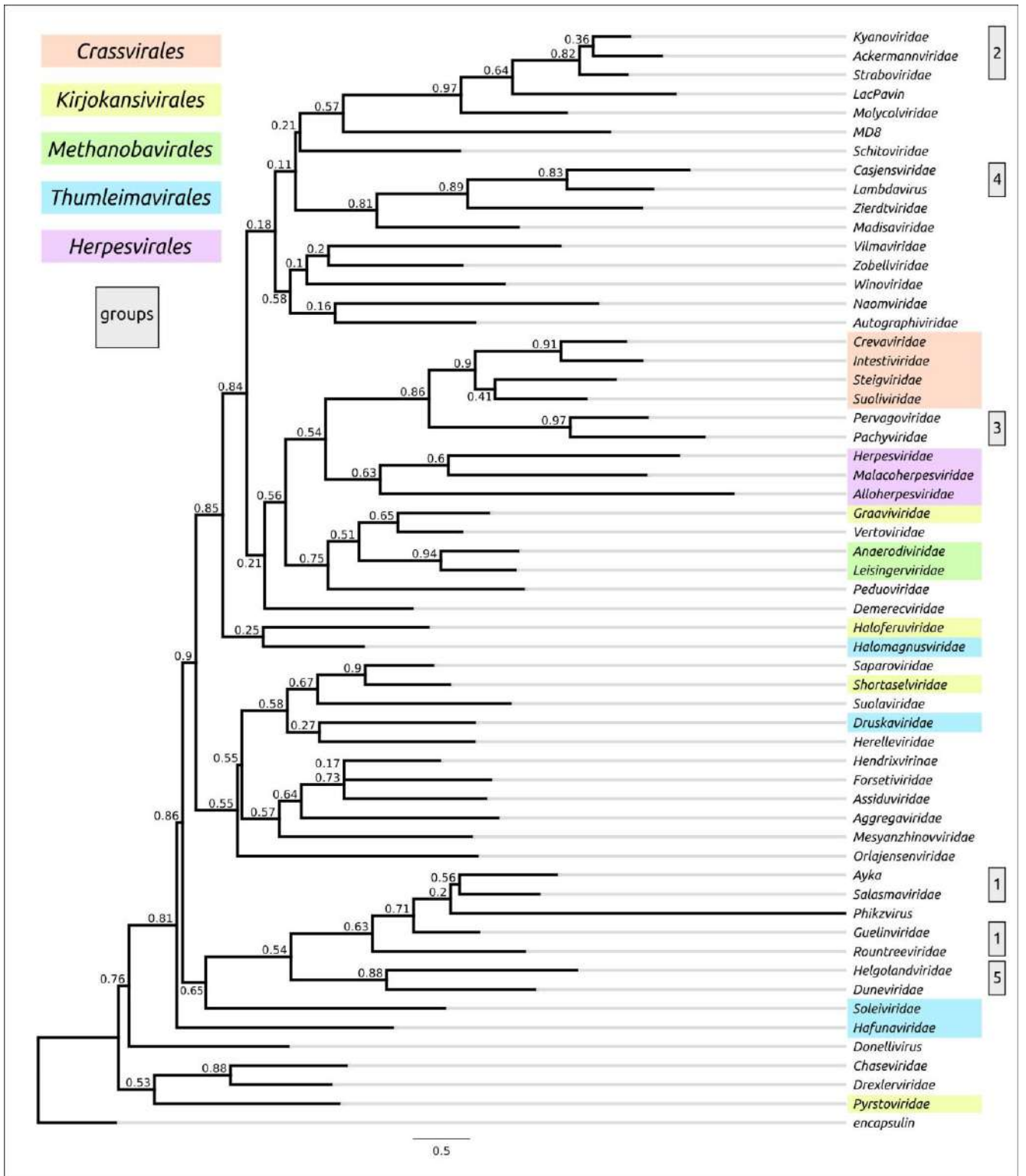


Рисунок Е.10. Филогенетическое дерево, построенное с использованием выравнивания 57 аминокислотных последовательностей главного капсидного белка и энкапсулина с помощью mTM-align. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

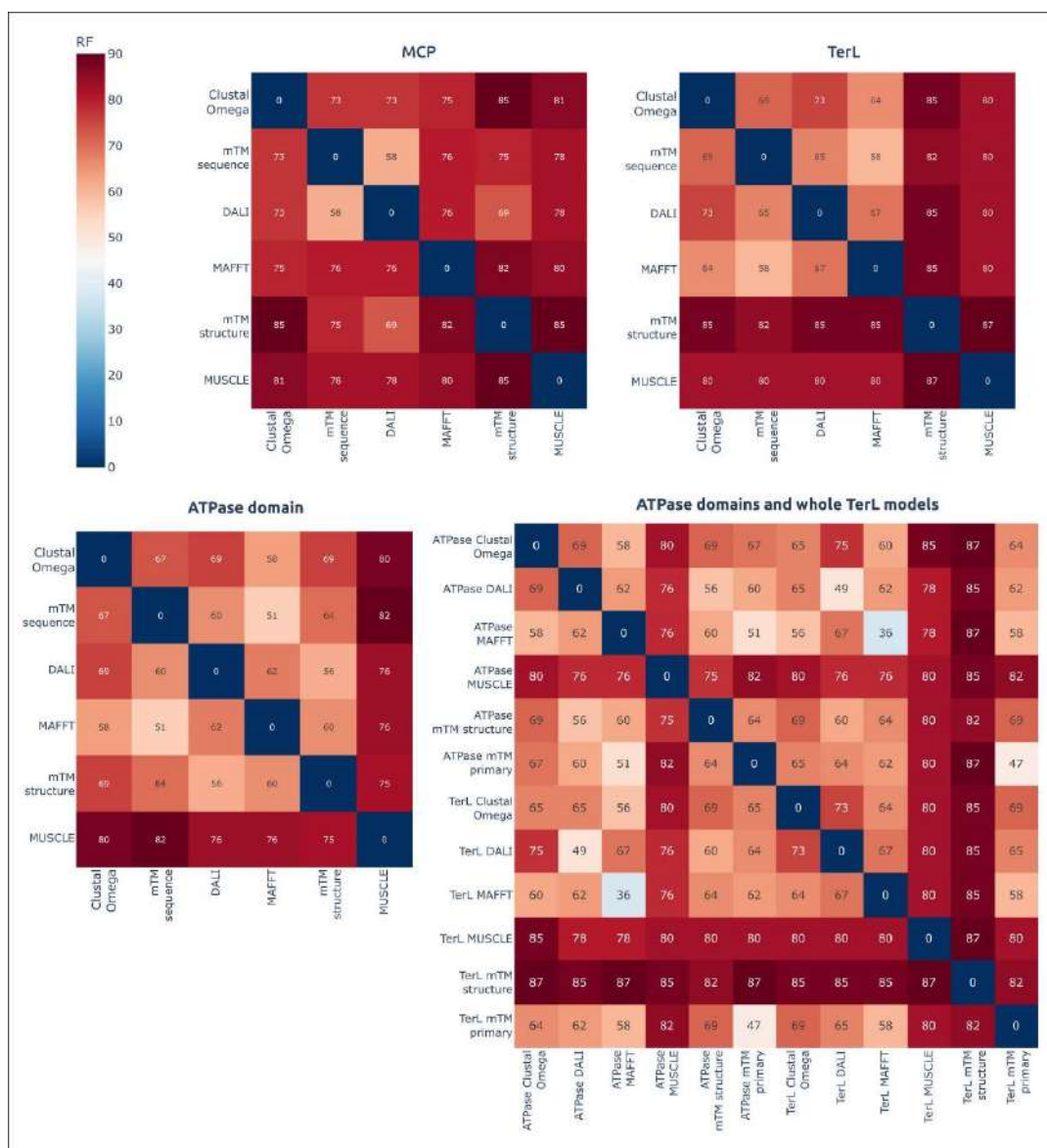


Рисунок E.11. Сравнение топологии деревьев, полученных с использованием аминокислотных последовательностей при помощи структурного выравнивания и других алгоритмов выравнивания, с помощью нормализованных дистанций Робинсона-Фаулдса, показанных в матрицах. В подписях указан белок и алгоритм выравнивания. «MCP» – главный капсидный белок, «TerL» – терминаза, «ATPase» – АТФазный домен терминазы. Обозначение «mTM primary» означает, что дерево было построено с использованием выравнивания аминокислотных последовательностей с помощью mTM-align; обозначение «mTM structure» означает, что дерево было построено с использованием структурного сходства, измеренного с помощью TM-score.

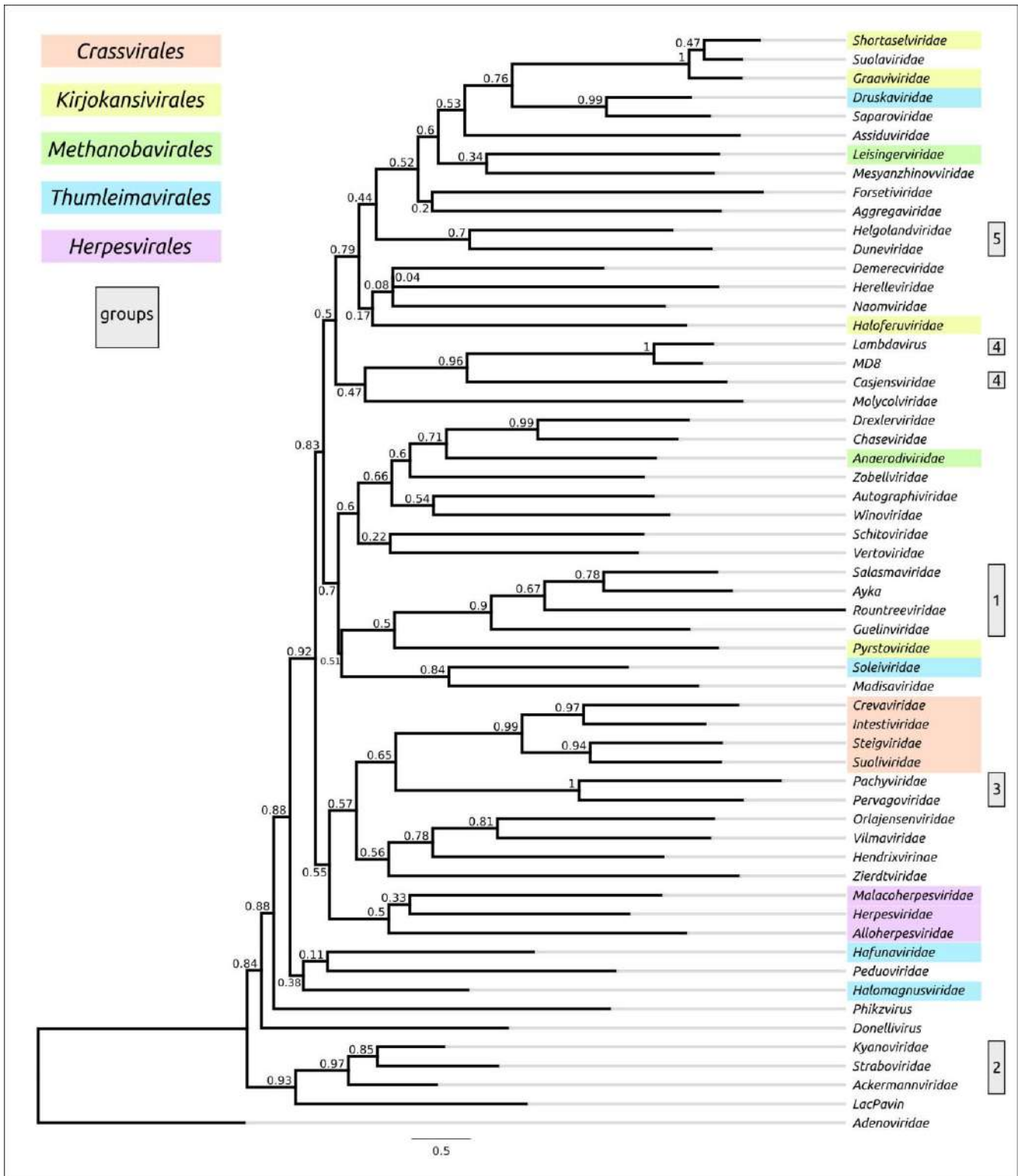


Рисунок Е.12. Филогенетическое дерево, построенное с использованием выравнивания 58 аминокислотных последовательностей терминазы с помощью MAFFT. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

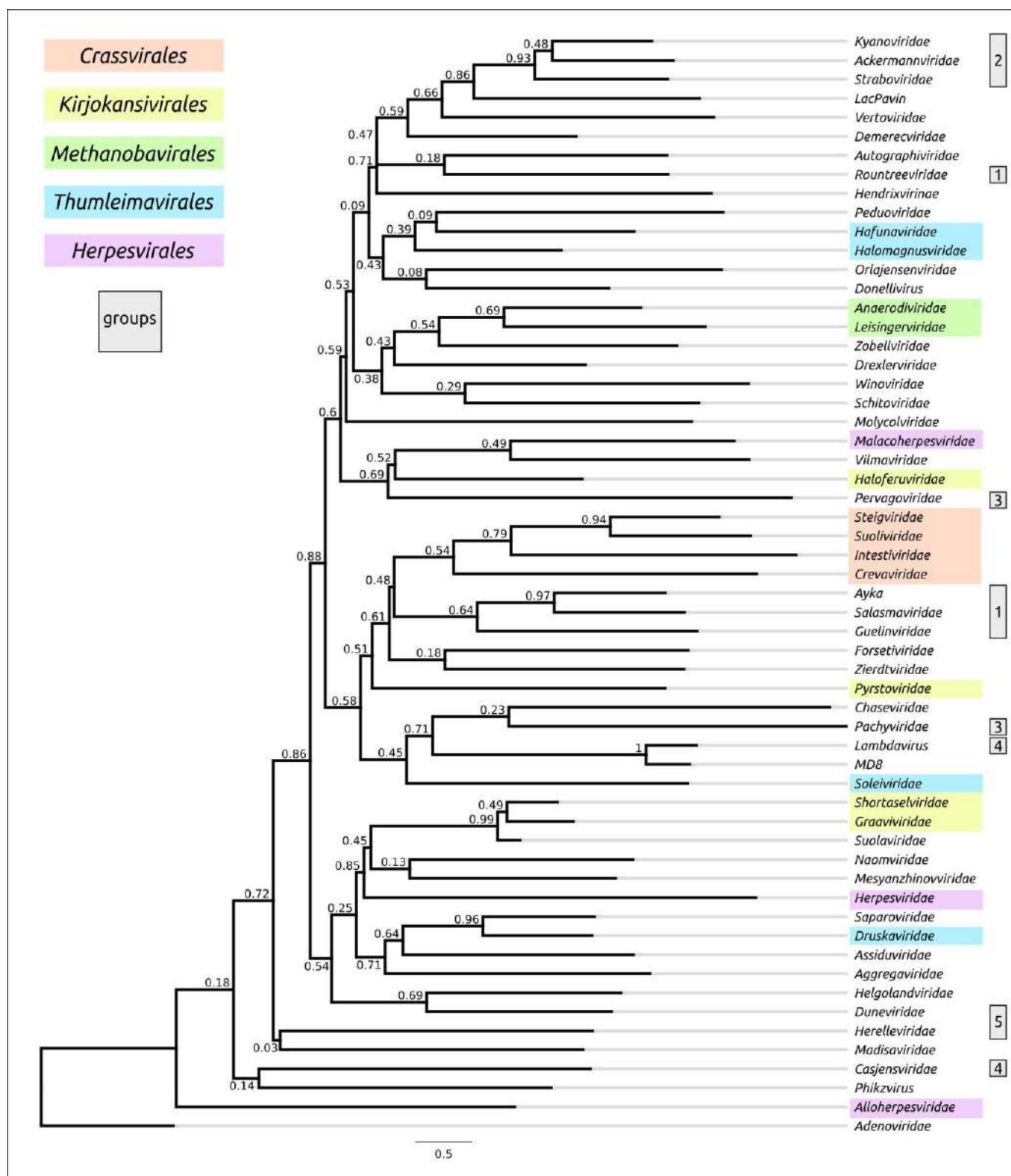


Рисунок Е.13. Филогенетическое дерево, построенное с использованием выравнивания 58 аминокислотных последовательностей терминазы с помощью MUSCLE. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.

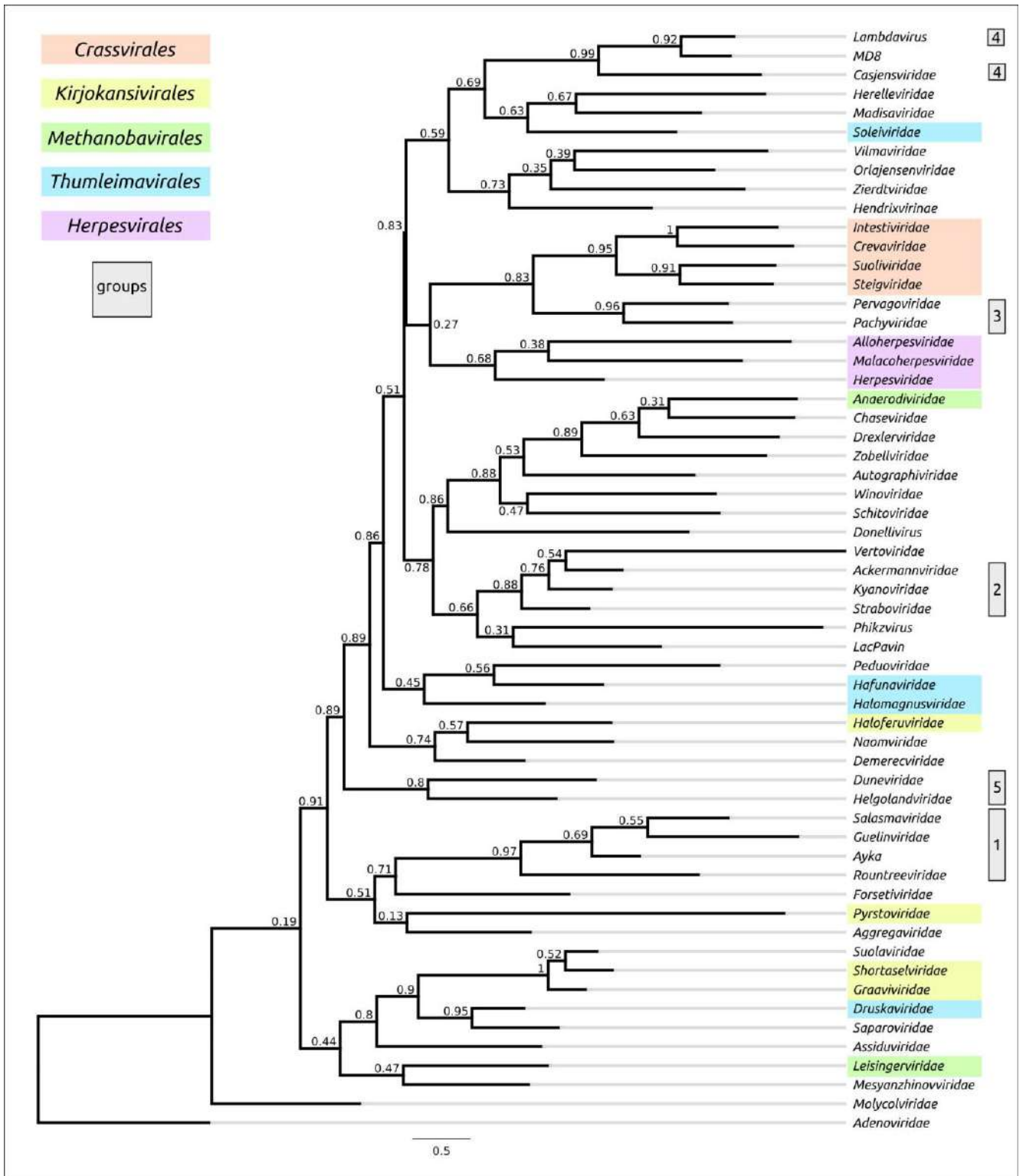
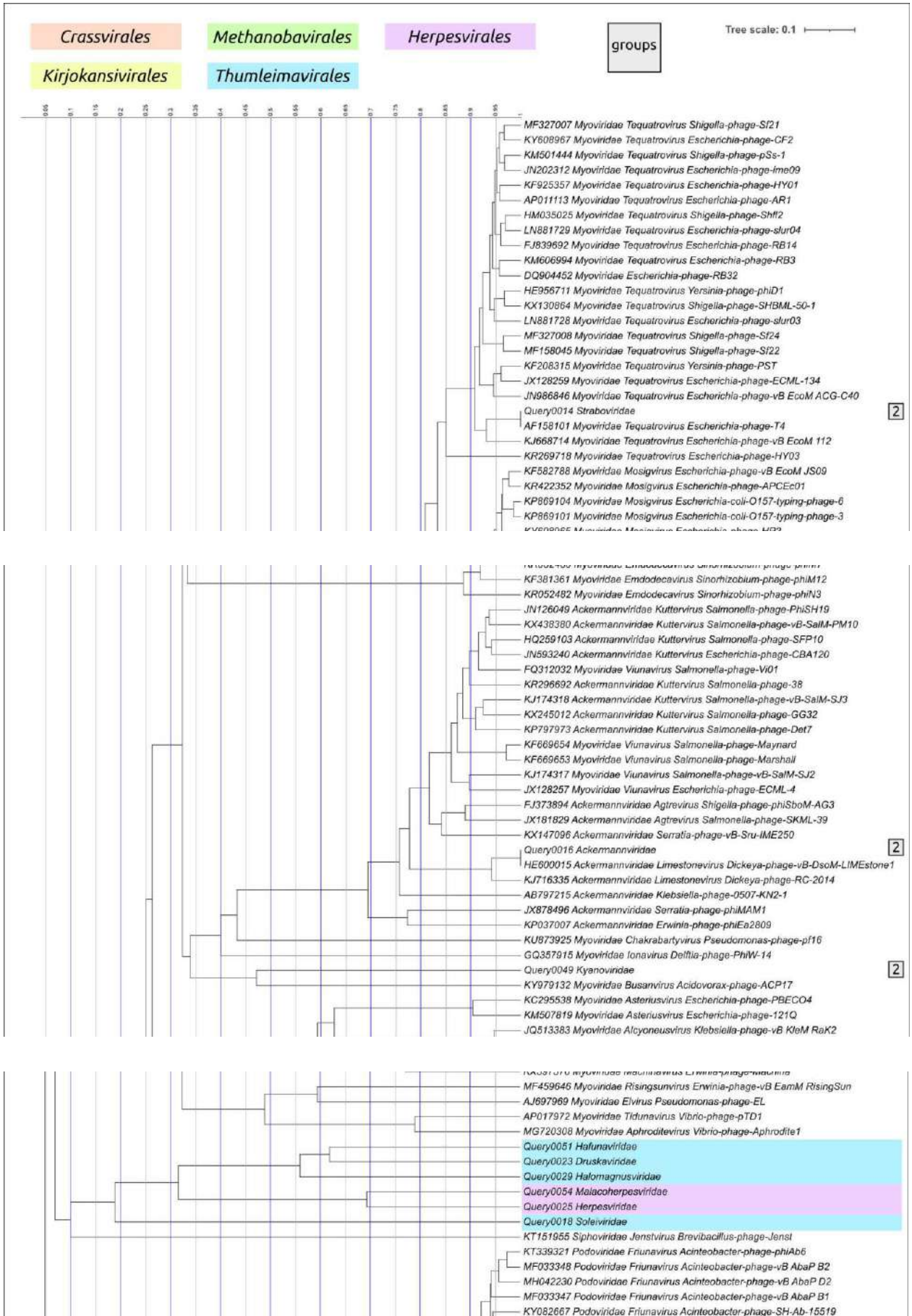
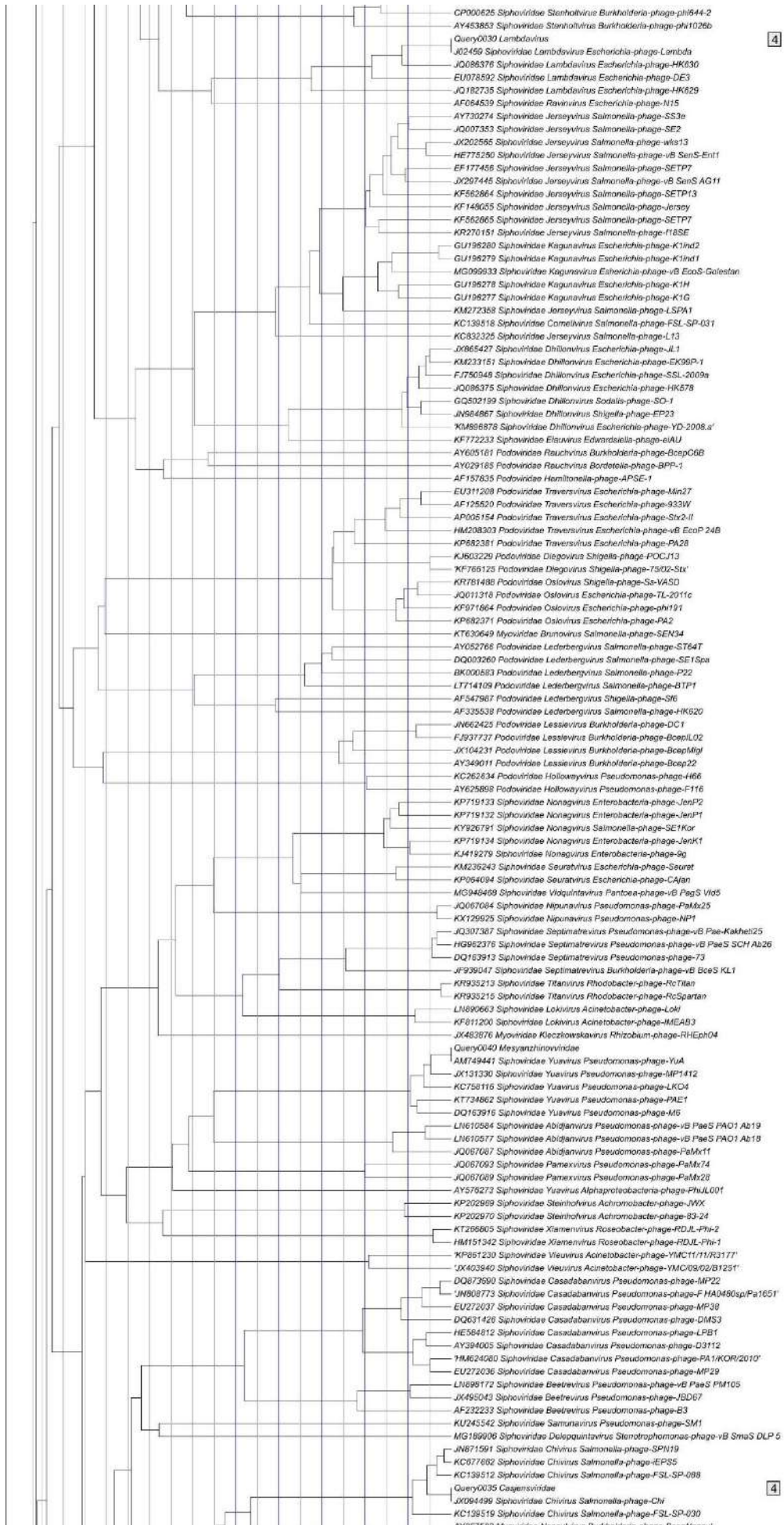
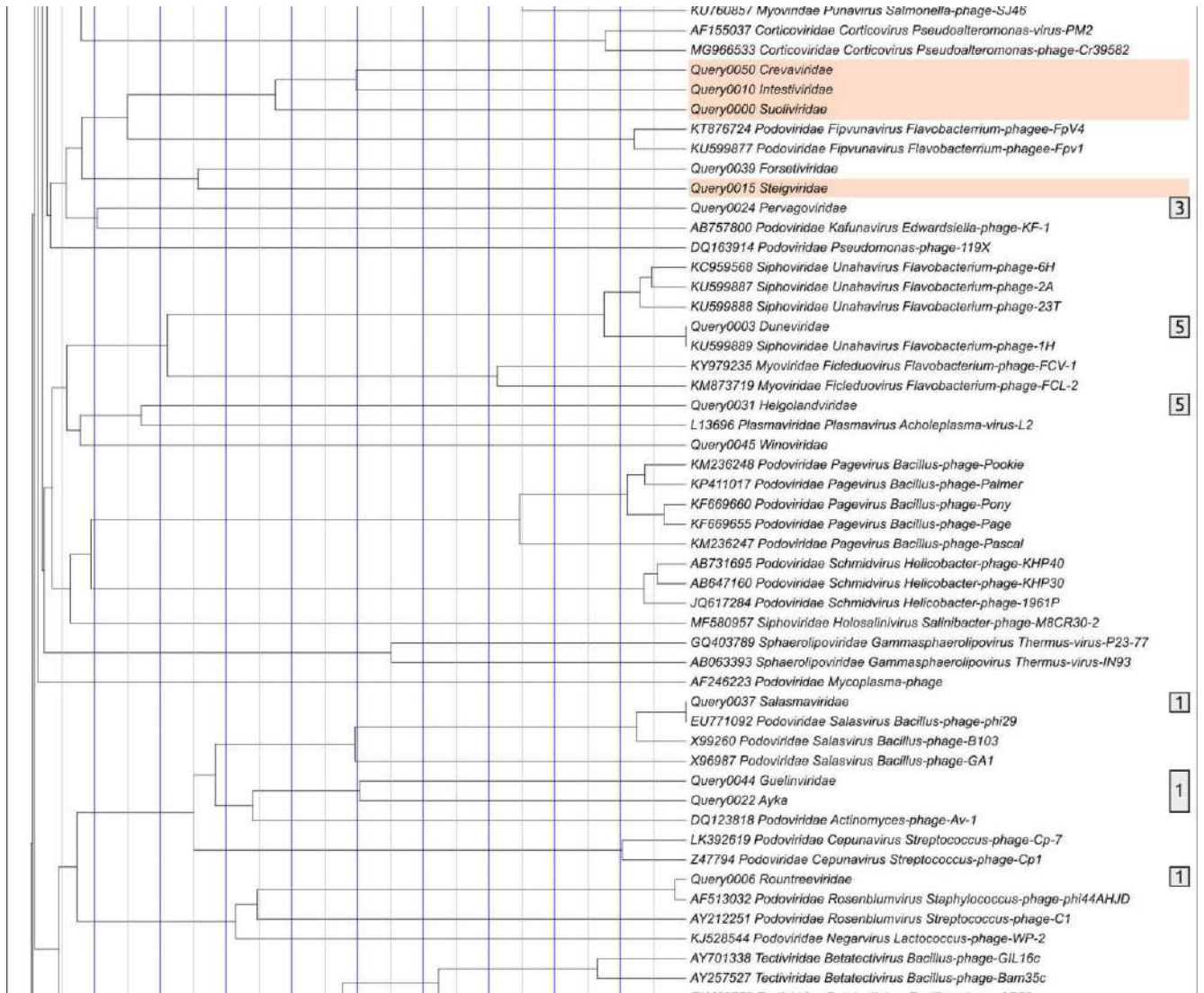
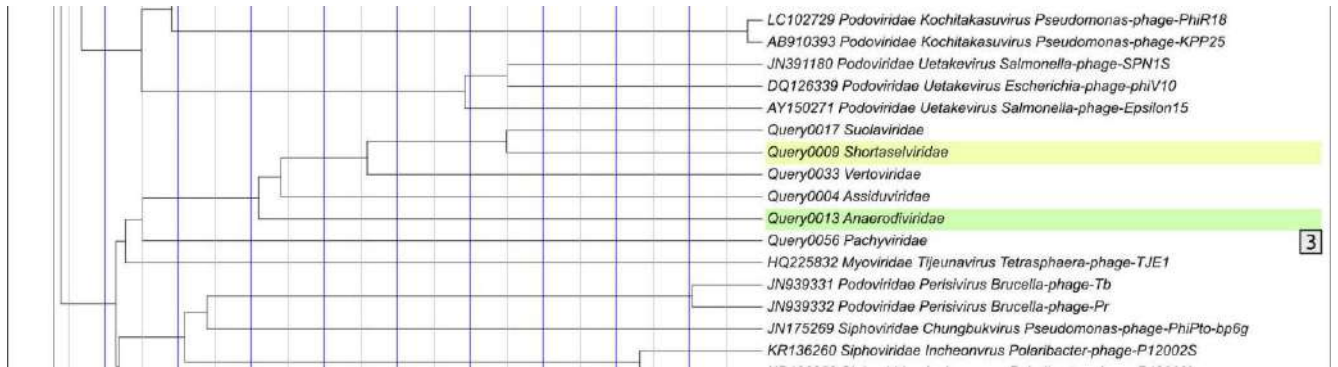


Рисунок Е.14. Филогенетическое дерево, построенное с использованием выравнивания 58 аминокислотных последовательностей терминазы с помощью mTM-align. Значения ожидаемых результатов бутстрэп-анализа (1000 реплик) указаны около соответствующих веток.







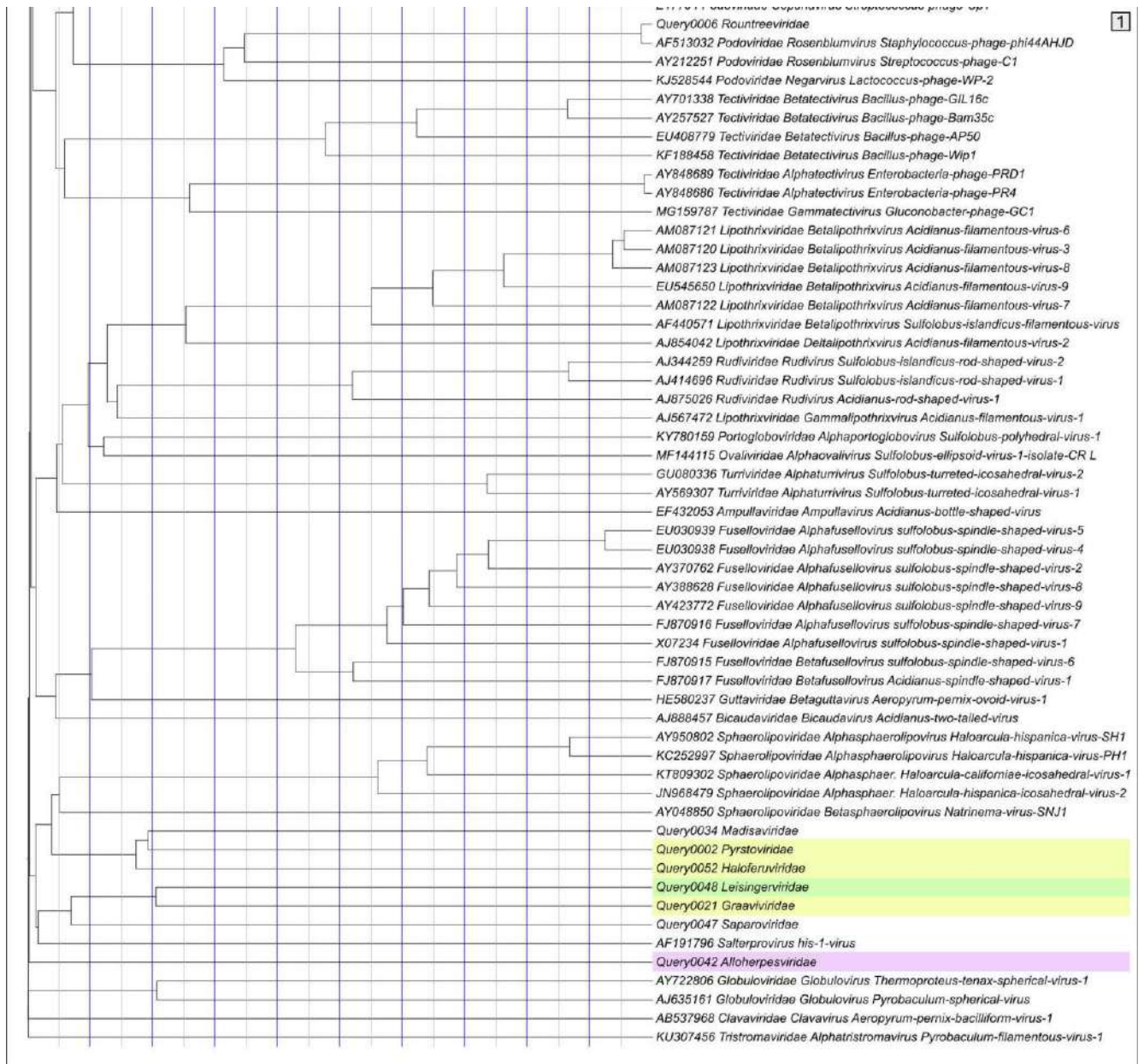


Рисунок Е.15. Фрагменты дерева GRAViTy (доступно в сети Интернет по адресу <https://www.mdpi.com/2218-273X/13/1/110/s1?version=1672913715>), полученного с использованием геномных последовательностей 57 представителей различных семейств и групп вирусов реалма *Duplodnaviria*.