

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
GRADUATION THESIS

Сравнительный анализ геномных и протеомных данных для оценки активности
систем антифаговой защиты в различных штаммах O1/O139 серогрупп *Vibrio*
cholerae

Обучающийся / Student Козырева Анфиса Ильинична

Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа

ИТМО интердисциплинарного инжиниринга

Группа/Group A4251

Направление подготовки/ Subject area 06.04.01 Биология

Образовательная программа / Educational program Прикладная геномика 2023

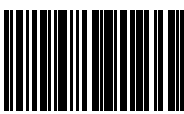
Язык реализации ОП / Language of the educational program Русский

Квалификация/ Degree level Магистр

Руководитель ВКР/ Thesis supervisor Мелешко Дмитрий Алексеевич, PhD, науки,
Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного
инжиниринга, доцент (квалификационная категория "ординарный доцент")

Консультант не из ИТМО / Third-party consultant Говорун Вадим Маркович, Research
Institute of Systems Biology and Medicine of Rospotrebnadzor, Director, Head of the Laboratory
of Simple Systems and Head of the Multi-mix Research Laboratory, доктор биологических
наук, профессор, академик РАН

Обучающийся/Student

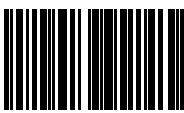
Документ подписан	
Козырева Анфиса Ильинична	
21.05.2025	

(эл. подпись/ signature)

Козырева
Анфиса
Ильинична

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Мелешко Дмитрий Алексеевич	
21.05.2025	

(эл. подпись/ signature)

Мелешко
Дмитрий
Алексеевич

(Фамилия И.О./ name
and surname)

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**ЗАДАНИЕ НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ /
OBJECTIVES FOR A GRADUATION THESIS**

Обучающийся / Student Козырева Анфиса Ильинична
Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа
ИТМО интердисциплинарного инжиниринга
Группа/Group A4251
Направление подготовки/ Subject area 06.04.01 Биология
Образовательная программа / Educational program Прикладная геномика 2023
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Магистр
Тема ВКР/ Thesis topic Сравнительный анализ геномных и протеомных данных для
оценки активности систем антифаговой защиты в различных штаммах O1/O139 серогрупп
Vibrio cholerae
Руководитель ВКР/ Thesis supervisor Мелешко Дмитрий Алексеевич, PhD, науки,
Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного
инжиниринга, доцент (квалификационная категория "ординарный доцент")
Консультант не из ИТМО / Third-party consultant Говорун Вадим Маркович, Research
Institute of Systems Biology and Medicine of Rospotrebnadzor, Director, Head of the Laboratory
of Simple Systems and Head of the Multi-mix Research Laboratory, доктор биологических
наук, профессор, академик РАН

Характеристика темы ВКР / Description of thesis subject (topic)

Тема в области фундаментальных исследований / Subject of fundamental research: да /
yes

Тема в области прикладных исследований / Subject of applied research: нет / not

Основные вопросы, подлежащие разработке / Key issues to be analyzed

В ВКР проводится анализ геномной структуры штаммов патогенного микроорганизма *Vibrio cholerae* и проводится соотнесение полученных результатов с протеомными данными для оценки реально экспрессирующихся систем бактериального организма. Основное внимание посвящено разнице между потенциальным дефенсомом бактериального организма и постоянно-экспрессирующимися частями этой системы.

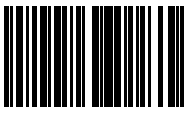
В рамках работы были выполнены: черновая сборка замкнутых геномов, структурная и функциональная аннотация драфтовых геномов нескольких штаммов *Vibrio cholerae*, масс-спектрометрический анализ для получения протеомных данных, анализ протеомных данных в контексте геномных данных, соотнесение с данными литературы.

Дата выдачи задания / Assignment issued on: 01.04.2025

Срок представления готовой ВКР / Deadline for final edition of the thesis 31.05.2025

СОГЛАСОВАНО / AGREED:

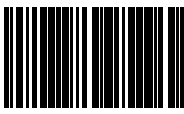
Руководитель ВКР/
Thesis supervisor

Документ подписан	
Мелешко Дмитрий Алексеевич	
22.04.2025	

(эл. подпись)

Мелешко
Дмитрий
Алексеевич

Задание принял к
исполнению/ Objectives
assumed BY

Документ подписан	
Козырева Анфиса Ильинична	
25.04.2025	

(эл. подпись)

Козырева
Анфиса
Ильинична

Руководитель ОП/ Head
of educational program

Документ подписан	
Райко Михаил Петрович	
22.05.2025	

(эл. подпись)

Райко Михаил
Петрович

**Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО
ITMO University**

**АННОТАЦИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
SUMMARY OF A GRADUATION THESIS**

Обучающийся / Student Козырева Анфиса Ильинична
Факультет/институт/кластер/ Faculty/Institute/Cluster передовая инженерная школа
ИТМО интердисциплинарного инжиниринга
Группа/Group A4251
Направление подготовки/ Subject area 06.04.01 Биология
Образовательная программа / Educational program Прикладная геномика 2023
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Магистр
Тема ВКР/ Thesis topic Сравнительный анализ геномных и протеомных данных для
оценки активности систем антифаговой защиты в различных штаммах O1/O139 серогрупп
Vibrio cholerae
Руководитель ВКР/ Thesis supervisor Мелешко Дмитрий Алексеевич, PhD, науки,
Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного
инжиниринга, доцент (квалификационная категория "ординарный доцент")
Консультант не из ИТМО / Third-party consultant Говорун Вадим Маркович, Research
Institute of Systems Biology and Medicine of Rospotrebnadzor, Director, Head of the Laboratory
of Simple Systems and Head of the Multi-mix Research Laboratory, доктор биологических
наук, профессор, академик РАН

**ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
DESCRIPTION OF THE GRADUATION THESIS**

Цель исследования / Research goal

Основной целью работы является выявление бактериальных белков Vibrio cholerae, экспрессируемых в стандартных условиях (отсутствие контакта с бактериофагами) и относящихся к компонентам противофаговых систем защиты (АФЗ).


Задачи, решаемые в ВКР / Research tasks

1. Характеристика геномной структуры изучаемых штаммов Vibrio cholerae серогруппы O1 El Tor токсигенной и нетоксигенной природы. 2. Типирование изучаемых штаммов по геномной последовательности основных факторов вирулентности: tcpA и ctxB генам. 3. Описание разнообразия компонентов противофаговых систем защит, обнаруженных in silico в изучаемых штаммах. 4. Описание разнообразия постоянно экспрессируемых компонентов систем АФЗ в условиях отсутствия контакта исследуемой бактериальной культуры с бактериофагом. 5. Предположение об альтернативных функциях, реализуемых наблюдаемыми компонентами АФЗ в бактериальной клетке, и скорости их вовлечения в противофаговый иммунный ответ.

Краткая характеристика полученных результатов / Short summary of results/findings

1. Штаммы *Vibrio cholerae* обладают высокой степенью вариабельности генома, ассоциированной с множественными включениями мобильных геномных элементов (МГЭ). 2. Геномы выборки кодируют 11 условных групп АФЗ различных по механизму действия, включая системы типов: Restriction-Modification (RM), CRISPR-cas, Toxin-antitoxin (TA), Retron, CBASS, Brex, Menshen, Lamassu-Family, AVAST, dXTPase. 3. Для всех изученных штаммов характерна экспрессия более половины компонентов всех предсказанных с использованием геномных данных систем АФЗ. 4. Согласно геномным данным, общей тенденцией для всех штаммов *V.cholerae* является наличие в геноме системы: рестрикции-модификации первого типа (RM-I), dXTPase, DMS_other, и экспрессия их компонентов на фоновом уровне в условиях отсутствия контакта с фаговым агентом. 5. Среди изученных представителей токсигенных штаммов встречаются случаи полной экспрессии крупной системы защиты типа brex, а также относительно менее распространённых систем типов Menshen, Lamassu-Family, retron даже в условии отсутствия бактериофагов в культуре. 6. Среди изученных представителей нетоксигенных штаммов из внешней среды предсказана система septu, однако наблюдается экспрессия только одного функционального компонента. 7. Только один из штаммов - нетоксигенный штамм M1526, содержащий плазмиду, содержит две АФЗ системы типа CRISPR-cas, из которых полностью подтверждается экспрессия всех компонентов системы cas-I-F1, расположенной на chr1, тогда как система cas_type_other, расположенная на плазмиде, экспрессируется частично. 8. Расчёт индекса адаптации кодонов CAI демонстрирует высокую степень сродства систем антифаговой защиты и остального генома исследуемых штаммов, однако ожидаемый по результатам расчёта CAI уровень экспрессии АФЗ-систем не соответствует реально наблюдаемому по анализу протеомных данных уровню экспрессии АФЗ.

Обучающийся/Student


Документ подписан	
Козырева Анфиса Ильинична	
21.05.2025	

(эл. подпись/ signature)

Козырева
Анфиса
Ильинична

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Мелешко Дмитрий Алексеевич	
21.05.2025	

(эл. подпись/ signature)

Мелешко
Дмитрий
Алексеевич

(Фамилия И.О./ name
and surname)

О РАБОТЕ

В работе представлен XXXXXXXXXXXXXXXXXXXXXXXX.

Выпускная квалификационная работа магистра изложена на ?? листах, включает X таблиц, X рисунков, X приложения, X литературных источников.

Ключевые слова: микробиология, медицинская микробиология, биобезопасность, секвенирование третьего поколения, ONT.

СОДЕРЖАНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

- АФЗ — Antiphage defense systems, Антифаговые системы защиты
- BREX — Bacteriophage Exclusion system, антифаговая система с эпигенетическим механизмом защиты
- CAI — Codon Adaptation Index, индекс адаптации кодонов
- CBASS — Cyclic-oligonucleotide-based antiphage signaling system, антифаговая система с использованием циклических сигнальных молекул
- CDS — Coding DNA Sequence, Кодированная последовательность
- dGTP — дезоксигуанозинтрифосфат, нуклеозидтрифосфат и предшественник нуклеотидов, используемый в клетках для синтеза ДНК
- dGTPase — фермент, гидролизующий dGTP, регулирующий пул нуклеотидов
- ICE — Integrative and Conjugative Element, интегрируемый конъюгативный элемент
- IS — Insertion Sequence, вставочный элемент
- Menshen — антифаговая система, связанная с регуляцией клеточного стресса
- PLE — Phage-inducible Chromosomal Island-like Elements, фагоиндуцируемые острова
- PDC-системы — предсказанные антифаговые комплексы с неподтверждённой экспериментально функцией
- RM-система — Restriction-Modification system, система рестрикции-модификации
- Retron-I — ретронная антифаговая система, синтезирующая msDNA
- Septu — антифаговая система, выявленная в нетоксигенных штаммах
- tcpA* — ген, кодирующий субъединицу токсин-корегулируемых пилей
- vps* — кластер генов, ответственных за биосинтез виброполисахарида
- hapR* — регулятор quorum sensing, влияющий на синтез ЭПС и формирование ружного фенотипа
- SXT/R391 — семейство интегрированных конъюгативных элементов, часто несущих гены резистентности
- ctxB* — ген субъединицы В холерного токсина
- Жгутиковые кластеры Flg и Fli — генные кластеры, кодирующие белки жгутикового аппарата
- ЛПС — липополисахарид, компонент внешней мембраны грамотрицательных бактерий
- О-антиген — полисахаридный антиген, компонент липополисахарида, определяющий серогруппу
- Дефенсом — Defensome, совокупность систем антифаговой защиты организма
- ДНК — дезоксирибонуклеиновая кислота
- РНК — рибонуклеиновая кислота

Транзиция — точечная мутация, обусловленная заменой одного пуринового основания на другое (аденина на гуанин или наоборот) или одного пиримидинового основания на другое (тимина на цитозин или наоборот)

Трансверсия — точечная мутация, обусловленная заменой пуринового основания (аденин, гуанин) на пиримидиновое (тимин, цитозин) и наоборот

рРНК, тРНК — рибосомная РНК, транспортная РНК

ЭПС — экзополисахарид, виброполисахарид

R-форма — форма липополисахарида с отсутствующим или сокращённым O-антигеном

Эффлюксный насос — белок, обеспечивающий активный вынос токсинов и антибиотиков из клетки

Фаг — бактериофаг, вирус, инфицирующий бактерии

Протеомика — набор методов и технологий определения и анализа белков

Геномика — совокупность методов исследования геномной структуры и функций организма

Норма, нормальные условия, стандартные условия — в данной работе - условия культивирования бактерий: немодифицированная среда LB, 37°C, в течение 18 часов, в условиях отсутствия специфических стрессовых факторов, в частности контакта с бактериофагами

ВВЕДЕНИЕ

[illegible]

Цели и задачи настоящей работы

XX

В задачи настоящей работы входит:

- a) XX
b) YY
B) VVV

Положения выносимые на защиту

1. AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA.
2. BBBBBBBBBBBBBBBBBBBBBBBBBBBBBB.
3. CCCCCCCCCCCCCCCCCCCCCCCCCCCC.

Научная новизна работы

1. AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA.
2. BBBBBBBBBBBBBBBBBBBBBBBBBBBBBB.
3. CCCCCCCCCCCCCCCCCCCCCCCCCCCC.

1 ЛИТЕРАТУРНЫЙ ОБЗОР

1.1 Подраздел 1

Генетический материал холерного вибриона состоит из двух хромосом, может содержать плазмиды. Суммарный размер хромосом оценивается в 4 миллиона пар оснований. Обе хромосомы кодируют элементы, необходимые для жизнедеятельности клетки, например рРНК опероны, имеют постоянный размер и структуру, встречаются во всех представителях таксона.

1.2 Подраздел 2

Всемирная организация здравоохранения рекомендовала использовать антибиотические средства для лечения только тяжелых случаев холеры, и в современной стратегии борьбы с заболеванием массовое использование антибиотиков не рекомендуется [1]. Тем не менее, в ряде случаев антибиотики продемонстрировали свою эффективность, а потому стали применяться, наряду с регидратационными растворами, как основное средство в борьбе этим заболеванием [2]. Неизбежным итогом широкого применения антибиотиков против бактериальных инфекционных агентов становится эволюция и приспособление последних к применяемым против них препаратам [?].

1.3 Подраздел 3

Применение бактериофагов в медицине известно с начала XX века. С момента их открытия фаги успешно использовались для терапии инфекций, вызванных различными патогенными бактериями, включая *Staphylococcus spp.*, *Streptococcus spp.* и *Pseudomonas aeruginosa*. Эффективность бактериофаги продемонстрировали против возбудителей дизентерии (*Shigella spp.*), холеры (*Vibrio cholerae*), брюшного тифа (*Salmonella typhi*), а также для лечения раневых инфекций в военно-полевых условиях [3].

1.3.1 Секция 1 внутри подраздела 3

Однако, широкое применение такого подхода ограничивается недостаточной изученностью молекулярных механизмов работы ряда антифаговых систем.

1.4 Подраздел 4

1.4.1 Секция 1 внутри подраздела 4

Описанный подход подразумевает цикл разнообразных экспериментальных и вычислительных работ.

1.4.2 Секция 2 внутри подраздела 4

Вычисление CAI включает формирование эталонного набора генов.

2 МАТЕРИАЛЫ И МЕТОДЫ

В рамках экспериментальной части работы было проведено выращивание чистой культуры, тотальное выделение ДНК.

2.1 Микробиология

Для культивирования и проведения дальнейшей работы были выбраны 5 штаммов.

Таблица 2.1 – Штаммы

Наименование	Год	Место
АФФ	2011	Калмыкия
МАО	2016	Москва
ИМА	1970	Саратов

Выращивание чистой культуры проводили с использованием твёрдой агаризованной не модифицированной среды LB при температуре 37°C, pH=7.4 в течение 18 часов.

2.2 Геномика

Выделение ДНК проводили стандартным методом [?].

Подготовка библиотек геномной ДНК бактерий для секвенирования осуществлялась по протоколу Rapid Barcoding Kit 96 V14 (SQK-RBK114.96) с использованием соответствующих реагентов, рекомендованных в протоколе. Секвенирование геномной ДНК проводилось с использованием прибора PromethION компании Oxford Nanopore Technologies.

2.3 Анализ данных

Оценка качества полученных в результате секвенирования прочтений проводилась с использованием fastQC (v0.12.1) [4], Nanoplot (v1.42.0) [5], pycoQC (v2.5.0.3) [6].

3 РЕЗУЛЬТАТЫ

3.1 Сборка генома

В работу были переданы 5 штаммов, культивированные в течение 18 часов на немодифицированной среде LB при 37°C. В результате по данным секвенирования с использованием ONT-прочтений были получены сборки последовательностей геномов, что согласуется с данными литературы (Таблица 3.1).

Таблица 3.1 – Результаты сборки геномов

Наименование	Суммарная длина, bp	Хромосома 1, bp	Хромосома 2, bp	Плазмида, bp
S1	4 109 167	2 980 553	1 129 214	-
S2	4 938 774	2 912 449	1 191 583	834 742
S3	4 064 585	2 993 592	1 070 993	-
S4	4 113 653	3 053 066	1 060 587	-
S5	4 091 768	3 042 768	1 049 000	-

Каждая сборка состоит из двух или трёх замкнутых кольцевых последовательностей (Рисунок 3.1) и не содержит крупных недосеквенированных участков генома, что подтверждается данными проверки целостности сборки с использованием BUSCO - количество обнаруженных в полном объеме групп ортологов составляет от 99.5 %.

3.1.1 Секция 1 Подраздела Сборка генома

Для более полной картины сравнения следует отметить, ген *flgN*, в референсе не представленный.

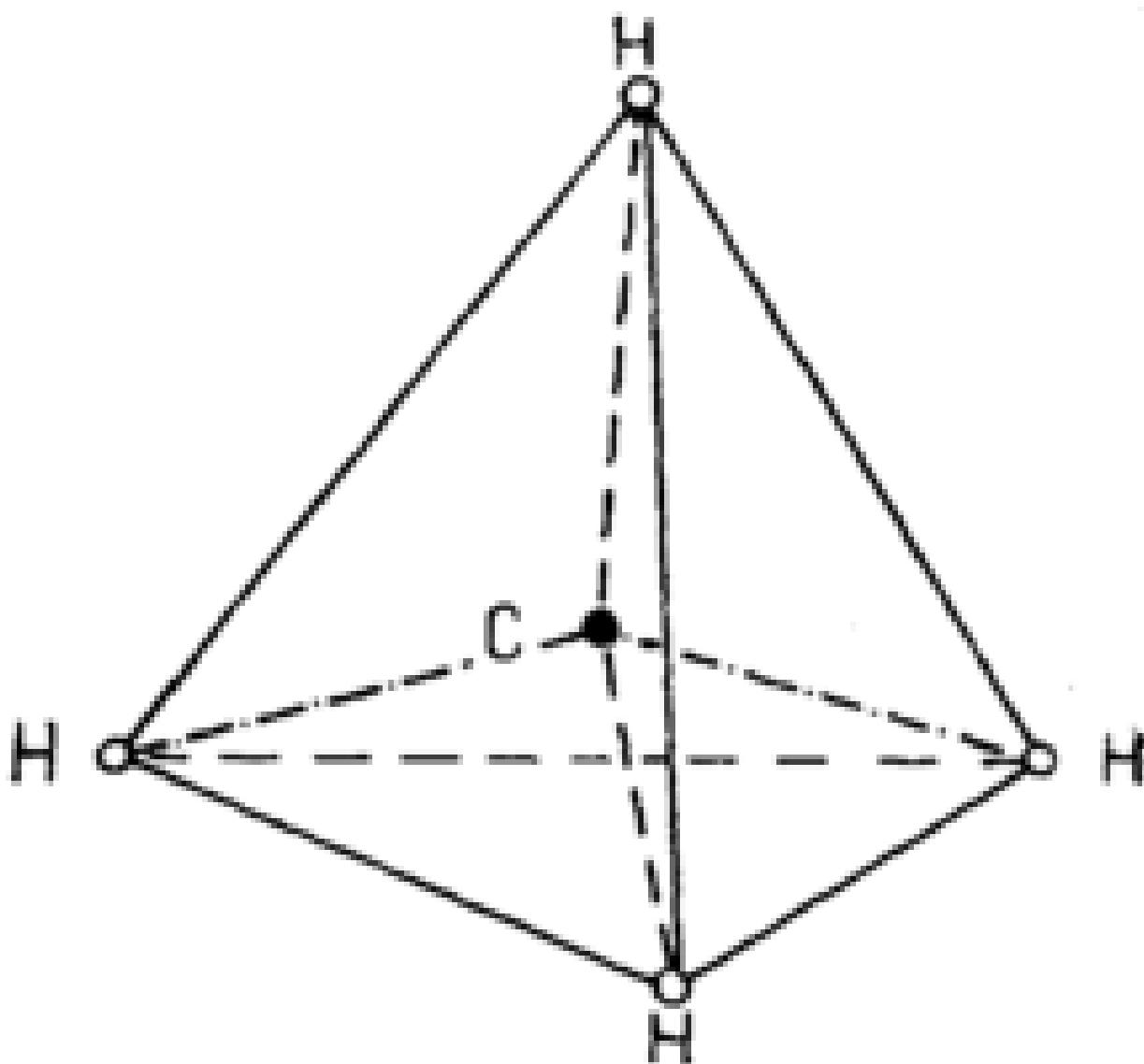


Рисунок 3.1 – Состав собранных геномов

4 ВЫВОДЫ

4.1 Геномная организация

Геномы большинства исследованных штаммов имеют характерную для вида организацию (2 кольцевые хромосомы) и суммарную длину около 4 млн пар оснований, что полностью соответствует современным представлениям о структуре генома этой бактерии.

4.2 Генетическое типирование

Типирование по ключевым генам полностью соответствует данным литературы и указывает на присутствие в выборке разнообразных по геномному составу вариантов патогена, характерных для разных волн пандемии.

4.3 Общие выводы и перспективы

Полученные результаты демонстрируют высокую степень геномного и функционального разнообразия штаммов как из клинических, так и из природных источников. Это подтверждает важность комплексного подхода к изучению эпидемиологии, эволюции и механизмов защиты возбудителя. Полученные данные могут быть использованы для дальнейших исследований по микробиологии.

ЗАКЛЮЧЕНИЕ

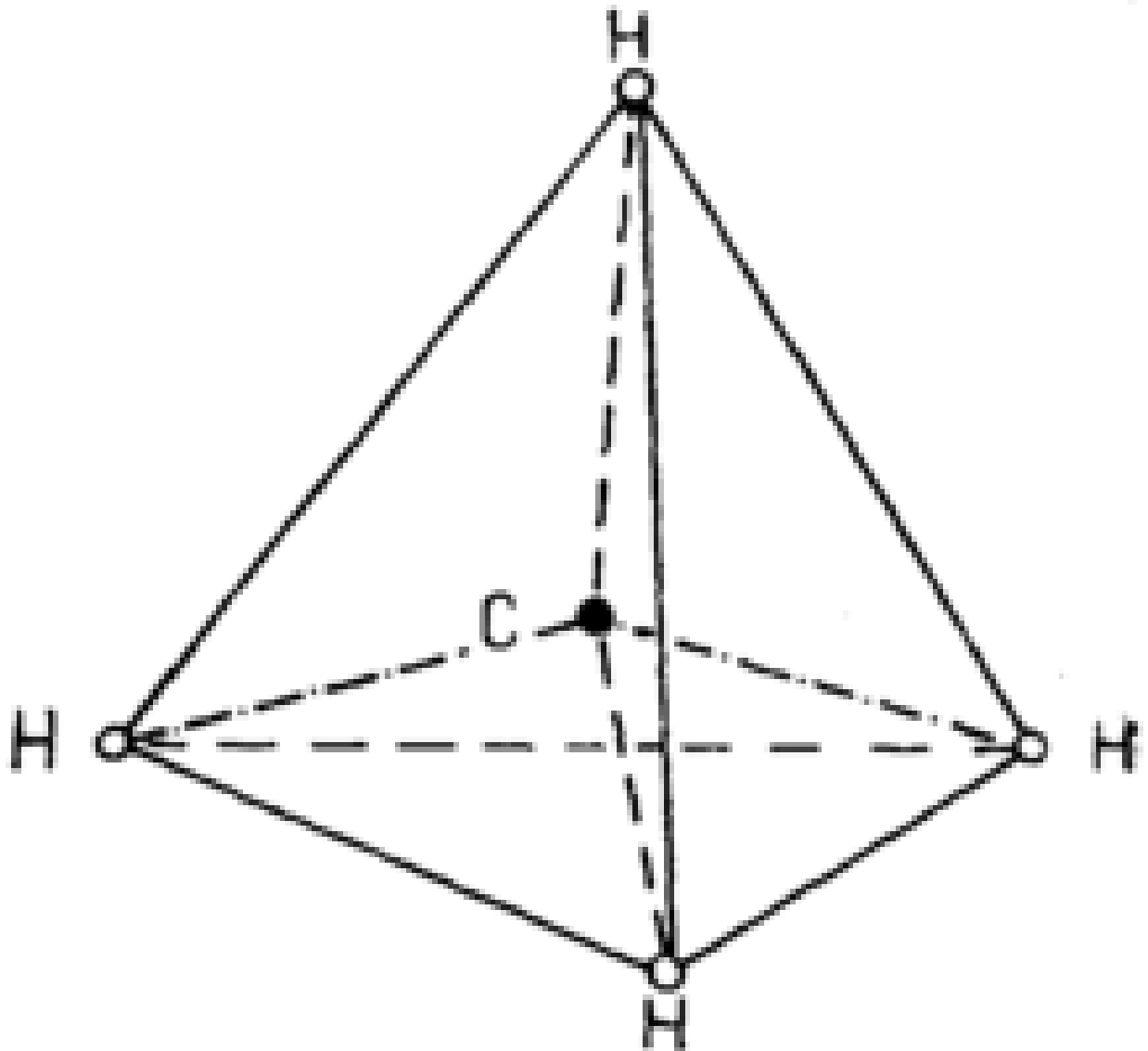
Исследование показало, что современные штаммы обладают сложной и динамичной системой геномной организации, включающей разнообразные механизмы защиты и адаптации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. П.Н Бургасов, В.И Покровский, А.П Авцын, В.А Шахламов, and Е.А Ведьмина. Большая медицинская энциклопедия. — Москва, 1986.
2. Всемирная организация здравоохранения. Отчет ВОЗ о внешней обстановке №23 : ВОЗ. — 2025. — Access mode: <https://www.who.int/publications/m/item/multi-country-cholera-outbreak--external-situation-report--23--20-feb> (online; accessed: 2025-03-03).
3. Trucksis Michele, Michalski Jane, Deng Ying Kang, and Kaper James B. The *Vibrio cholerae* genome contains two unique circular chromosomes // *Proceedings of the National Academy of Sciences*. — 1998. — Vol. 95, no. 24. — P. 14464–14469.
4. Andrews Simon. FastQC: A Quality Control Tool for High Throughput Sequence Data. — Software available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. — 2010.
5. De Coster Wouter and Rademakers Rosa. NanoPack2: population-scale evaluation of long-read sequencing data // *Bioinformatics*. — 2023. — 05. — Vol. 39, no. 5. — P. btad311. — <https://academic.oup.com/bioinformatics/article-pdf/39/5/btad311/50394865/btad311.pdf>.
6. Leger Adrien and Leonardi Tommaso. pycoQC, interactive quality control for Oxford Nanopore Sequencing // *Journal of Open Source Software*. — 2019. — Vol. 4, no. 34. — P. 1236. — Access mode: <https://doi.org/10.21105/joss.01236>.

ПРИЛОЖЕНИЕ 1

Полногеномные выравнивания против референсного генома



Приложение 1 - Схемы выравниваний: а - штамм S1; b - штамм S2; с - штамм S3; d - штамм S4; e
- штамм S5