

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра микробиологии и физиологии растений

**СОВЕРШЕНСТВОВАНИЕ СПОСОБОВ ДИФФЕРЕНЦИРОВАНИЯ ФИ-
ЛОГЕНЕТИЧЕСКИХ ГРУПП ТУЛЯРЕМИЙНОГО МИКРОБА С ИС-
ПОЛЬЗОВАНИЕМ БИОИНФОРМАТИЧЕСКИХ ПОДХОДОВ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 3 курса 331 группы
направления 06.04.01 Биология
биологического факультета
Катышева Семена Дмитриевича

Научный руководитель

д-р биол. наук, доц.

УМ 26.11.25.

Д.В. Уткин

Консультант

зав. отделом микробиологии

ФКУН Российский противочумный институт «Микроб»

Роспотребнадзора, канд. биол. наук

Осн 26.11.25

Н.А. Осина

Заведующий кафедрой

д-р биол. наук, доц.

УМ 26.11.25

Д.В. Уткин

Саратов 2025

Введение

Актуальность темы. Туляремия — зоонозная природно-очаговая инфекция, вызываемая грамотрицательной коккобациллой *Francisella tularensis*. Является особо опасной и относится к II группе патогенности. В последнее время на территории Российской Федерации ежегодно выявляется более 100 случаев заболевания туляремией. По данным референс-центра в период с 2000 по 2024 годы зафиксировано 4103 случая заражения в 67 субъектах РФ. По данной инфекции относительное благополучие определяется комплексом мероприятий по профилактике заболевания [13].

В связи с этим, задача идентификации штаммов туляремийного микроба имеет особую актуальность и заключается в определении их принадлежности к подвиду, биовару и субпопуляции в связи с существующими различиями в вирулентности между штаммами.[1].

На настоящий момент времени внутри вида *F. tularensis* особого внимания требуют два высокопатогенных подвида: *tularensis*, имеющий две субпопуляции AI и AII, *holarctica*, имеющая *japonica*, эритромицинустойчивый Ery^R, эритромицинчувствительный Ery^S биовары соответственно. Не используя методы *in silico*, крайне трудно определить данную внутривидовую принадлежность того или иного штамма [7].

Цель и задачи исследования. Целью данной работы стало совершенствование способов дифференцирования филогенетических групп туляремийного микроба с использованием биоинформатических подходов.

Для решения указанной цели были определены следующие задачи:

1. Сравнение эффективности способов внутривидовой дифференциации геномов штаммов *F. tularensis in silico*.
2. Разработка способа определения внутривидовой принадлежности *F. tularensis in silico* на основании выявления специфичных полиморфных нуклеотидов (SNP), делеций и вставок и других вариабельных участков генома.

3. Проведение апробации разработанного подхода на геномах штаммов из фонда «Государственной коллекции патогенных бактерий» ФКУН Российский противочумный институт «Микроб» Роспотребнадзора.

4. Определение внутривидовой принадлежности штаммов *F. tularensis*, полногеномные последовательности которых представлены в базе данных GenBankNCBI, с помощью предложенного метода.

5. Создание базы данных нуклеотидных последовательностей, с классификацией штаммов на основе геномных различий.

Материалы и методы исследования. Объектом исследования являлись генетические последовательности 106 полных геномов штаммов из фонда «Государственной коллекции патогенных бактерий» ФКУН Российский противочумный институт «Микроб» Роспотребнадзора и 914 полных геномов штаммов *F. tularensis* базы данных Genbank NCBI.

Полногеномное секвенирование

Полногеномное секвенирование осуществлялось на платформе DNBSEQ G50 (Mgi Tech, КНР).

Биоинформатический анализ

Для анализа данных и автоматизации использовался язык программирования Python (версия 3.11). Использовались библиотеки Biopython для работы с биологическими последовательностями, библиотека Scikit-learn использовался для воспроизведения методов кластеризации на геномных последовательностях. Также Pandas и SciPy для статистической обработки и манипуляций с данными. Библиотеки matplotlib и seaborn для визуализации данных. Язык программирования R (версия 4.3) в нем были задействованы пакеты ggplot2 для построения графиков и dplyr для обработки данных.

Анализа нуклеотидных последовательностей производился с помощью инструмента для локального выравнивания последовательностей BLAST (Basic Local Alignment Search Tool, версия 2.14), который применялся для поиска гомологий между целевыми последовательностями и базой данных NCBI. Это

позволило определить потенциальные филогенетические связи между изучаемыми генами и уже известными последовательностями. Так же он использовался для извлечения соответствующих генов-мишеней из всех штаммов, используемых в исследовании. Программа для множественного выравнивания MAFFT (версия 7.505) использовалась для точного и быстрого выравнивания нуклеотидных и аминокислотных последовательностей. MAFFT так же применялся для построения конкатенированных выравниваний, что обеспечивало надежность анализа филогенетических деревьев.

Попарное сравнение геномов штаммов относительно заданного референса проводилось программой Snippy (версия 4.6.).

Филогенетический анализ

Филогенетические деревья, были построены используя программу PhyML (версия 3.3.2023) методом максимального правдоподобия. UGENE (версия 52.0) — программное обеспечение с открытым исходным кодом так же применялся для комплексного анализа биоинформатических данных.

Пан-геномный анализ

Пан-геномный анализ проводился с помощью программы Roary (версия 3.11.2).

Мультилокусное сиквенс-типирование

Все файлы сборок полных геномов последовательностей подвергались MLST-типированию на открытом онлайн ресурсе PubMLST (<https://pubmlst.org/>) для сверки результатов.

Структура и объём работы. Работа изложена на 74 страницы, включает в себя введение, 3 главы, заключение, выводы, список использованных источников и приложение. Работа проиллюстрирована 8 таблицами и 16 рисунками. Список использованных источников включает 52 наименования.

Научная новизна.

Впервые было проведенное комплексное исследование геномного разнообразия, состоящее из семи подходов к внутривидовой дифференциации штаммов *F. tularensis*, что позволило получить качественную картину филогенетической структуры данного патогена.

Практическая значимость.

Был разработан алгоритм внутривидовой дифференциации на основе полиморфизмов гена 23S рРНК, который прошел апробацию, доказал свою эффективность и может быть использован не только для отдельного гена, но и для выбранных локусов, регионов дифференциации и даже для целой сети генетических элементов.

Разработана база данных, классифицированных по подвиговой принадлежности (с указанием субпопуляции и биоваров), указанием географической принадлежности, генетического варианта, носителя возбудителя, дате и места выделения туляремийных изолятов, позволит сильно облегчить дальнейшую работу исследователей данного патогена. Данная база данных является портативной, содержит встроенные инструменты анализа последовательностей, сортировки и пакетной выгрузки последовательностей полных геномов туляремийных изолятов.

Положения, выносимые на защиту.

1. Установлено, что наиболее точная реконструкция эволюционных взаимоотношений между близкородственными изолятами, среди всех используемых подходов, достигается при использовании анализа по коровым SNPs с построением филогенетических древ методом максимальной экономности.

2. Совершенствование процедуры анализа позволили идентифицировать ранее не описанные точечные мутации в консервативных участках 23S рРНК.

3. Разработанный алгоритм внутривидовой дифференциации на основе полиморфизмов гена 23S рРНК, который прошел апробацию, доказал свою эффективность и может быть использован не только для отдельного гена, но и

для выбранных локусов, регионов дифференциации и даже для целой сети генетических элементов.

4. Алгоритм внутривидовой дифференциации на основе полиморфизмов гена 23S рРНК, обобщен на все многообразие полных геномов штаммов *F. tularensis* содержащихся в базе данных GenBank NCBI.

5. Созданная база данных нуклеотидных последовательностей, с классификацией штаммов на основе геномных различий, позволит сильно облегчить дальнейшую работу исследователей данного патогена.

Основное содержание работы

В главе «Основная часть» представлен анализ литературных источников, своей тематикой касающихся выделения культуры, обнаружения туляремийного антигена, современного эпидемиологического состояния данной инфекции и типированию штаммов *F. tularensis*. В данной главе так же показана сравнительная характеристика методов внутривидовой дифференциации туляремийного микроба и их дискриминационная способность [5].

В главе «Результаты исследования» приводятся полученные экспериментальным путем данные разных подходов к внутривидовой дифференциации полных геномов туляремийных изолятов.

Кластерный анализ по нуклеотидному и GC-составу показал достаточно четкое деление на подвиды туляремийных изолятов, однако низкую дискриминационную способность касательно определения биоваров и субпопуляций.

Пан-геномный анализ, является крайне перспективным подходом и должен использоваться в исследованиях самого высшего уровня, так как способен предоставлять полную картину геномных различий даже между близкородственными штаммами патогена. Но вместе с этим пан-геномный анализ сопряжен со значительной трудоемкостью, требующей скоординированного междисциплинарного подхода в области классической микробиологии и математики, роль которого возрастает с повышением выборки исследуемых геномов.

Золотой стандарт типирования на основе канонических однонуклеотидных полиморфизмов (canSNPs) долгое время выступал как золотой стандарт для высокоточного типирования туляремийных изолятов, так как в отличие от методов, основанных на вариабельных повторах или других структурных различий, которые в свою очередь могут быть крайне неустойчивы, как например гены, полученные путем горизонтального переноса. Канонические SNPs представляют собой уникальные, стабильные и эволюционно обоснованные маркеры, которые возникают однократно и не подвержены обратным мутациям или независимому возникновению. Данный подход позволяет с высокой точностью разделить основные подвиды (*tularensis*, *holarctica*, *mediasiatica*) на четко определенные филогенетические клады и субклады и присвоить геноварианты конкретным исследуемым штаммам [7].

На настоящий момент времени дендрограммы на основе большого числа SNPs представляет более информативный подход по сравнению со схемой канонических, тем более что последний может вызывать существенные ошибки при дальнейшей интерпретации. Например, штаммы с разными дочерними canSNP-типами (B.199 и B.95), производными от B.65, если построить дерево по коровым SNPs (coreSNPs) формируют тесный кластер, что подтверждает их близкое родство, хотя исходя из указанных генотипов эта представлялось маловероятным. Также типирование по canSNPs может создать ложное представление об идентичности тех или иных штаммов, примером тому может стать вакцинный штамм 15НИИЭГ, который имеет такой же генотип что и три природных изолята, выделенных в Болгарии: FDC184, FDC186, FDC179. Этой ложной интерпретации можно избежать так же путем построения дерева по coreSNPs, где данные штаммы будут ложиться уже в отдельные субклады [15].

В отличие от подходов, использующих пан-геномное профилирование, вариабельные участки, канонические консервативные SNPs или кластеризация по тому или иному признаку, анализ по коровым SNPs (coreSNPs) делает фокус на всей полноте SNPs, присутствующие у выборки исследуемых полных геномов штаммов и разделяющих их друг от друга, обеспечивая высокую надежность и

воспроизводимость результатов. Данный анализ возможен, например при использовании метода «максимальной экономии», в основе данного метода лежит принцип экономности эволюционного любого живого организма. Или метода «максимального правдоподобия», в основе которого лежит крайне строгий математический метод переоценки правдоподобности дерева в процессе его построения. Это позволяет соблюсти все возможные признаки схожести или различия между полными геномами изолятов и обеспечить высокую дифференциацию между ними.

Анализ стабильных видоспецифичных SNPs обеспечивает воспроизводимые результаты и минимизирует риск ложного разделения. Такой метод особенно важен для быстрого отнесения новых клинических изолятов к известным генетическим группам. В качестве перспективной ДНК-мишени для внутривидовой идентификации туляремийного микроба была выбрана последовательность консервативного гена рибосомальной РНК 23S рРНК, в котором ранее были выявлены snp, специфичные для внутривидовых таксонов [2].

По результатам работы алгоритма из 106 штаммов из фонда «Государственной коллекции патогенных бактерий» ФКУН Российский противочумный институт «Микроб» Роспотребнадзора, 3 принадлежали FTT AI, 5 – FTT AII, 2 – FTM, 54 – FTH Ery^R, 38 – FTH Ery^S, 2 – FTH japonica и 2 – FTN. Аналогично по результатам работы алгоритма из 914 штаммов базы данных Genbank NCBI, 46 принадлежали FTT AI, 27 – FTT AII, 17 – FTM, 206 – FTH Ery^R, 573 – FTH Ery^S, 4 – FTH japonica и 41 – FTN. Так же в ходе проведения анализа выявлены дополнительные мутации (исключая аллели генов FTN) в последовательностях гена 23S рРНК – семь единичных полиморфизмов и три единичные вставки.

На момент начала работы в базе Genbank, процент идентифицированных на уровне подвидов (*tularensis*, *holarctica*, *mediasiatica* и *novicida*) штаммов составлял 53%, причем патогенные подвиды *tularensis* и *holarctica* не разграничивались на субпопуляции AI и AII и на Japonica, эритромицинустойчивый Ery^R, эритромицинчувствительный Ery^S биовары соответственно. В связи с этим со-

здание базы данных, включающих в себя все полседовательности полных геномов штаммов, представленных в GenBank NCBI с указанием данной дифференциации, является практически значимой задачей. Мы решили не останавливаться только на данном признаке, но и внести в базу метаданные по генетическому варианту, дате и месту выделения, номера ID (GenBank ID) и биологических проектов (BioProject GenBank) и носителю того или иного изолята возбудителя туляремии. Была создана портативная база данных нуклеотидных последовательностей полных геномов штаммов *F. tularensis* с их классификацией, возможностью сортировки, одиночной и пакетной выгрузки файлов (.fasta) указанных последовательностей.

Таким образом, была проведена комплексная характеристика геномного разнообразия штаммов *F. tularensis* с применением различных биоинформатических подходов.

Выводы

1. Представленные в исследовании способы внутривидовой дифференциации геномов штаммов *F. tularensis in silico* демонстрировали различную эффективность и трудоемкость их использования. Способ Кластеризации по нуклеотидному и GC-составу показал способность типирования полных геномов патогена только до уровня подвида. Пан-геномный анализ показал высокую дискриминационную способность и перспективность исследования данного метода, однако обладает достаточно трудоемкой реализацией. Анализ по каноническим SNPs является крайне эффективным способом идентификации, но может приводить в некоторых случаях к ошибочной интерпретации. Анализ по коровым SNP продемонстрировал самую высокую эффективность.

2. Был разработан способ определения внутривидовой принадлежности *F. tularensis in silico* на основании выявления специфичных полиморфных нуклеотидов (SNP), делеций и вставок и других вариабельных участков генома. В качестве перспективной ДНК-мишени была выбрана последовательность гена

23S рРНК, в котором были выявлены *snr*, специфичные для внутривидовых таксонов.

3. По способу указанному выше был разработан автоматизированный алгоритм, который был апробирован на штаммах из фонда «Государственной коллекции патогенных бактерий» ФКУН Российский противочумный институт «Микроб» Роспотребнадзора (ГКПБ). Результаты анализа полностью совпали с уже известными значениями таксономической геномов штаммов ГКПБ.

4. Данный подход был обобщен на все многообразие геномов штаммов представленных в базе данных GenBank NCBI на момент 15.09.2024 г. В связи с увеличением выборки исследуемых штаммов в ходе работы были выявлены дополнительные мутации в последовательности гена 23S *pРНК* - семь единичных полиморфизмов и три единичные вставки.

5. Была создана портативная базы данных нуклеотидных последовательностей, с классификацией штаммов на основе геномных различий, куда вошли 914 полных последовательностей туляремийных изолятов.

Список использованных источников

1 Внутривидовая дифференциация штаммов *Francisella tularensis* методом мультилокусной полимеразной цепной реакции с учетом результатов в режиме реального времени / Н.А. Осина [и др.] // Проблемы особо опасных инфекций. – 2023, № 1. – С. 132-141.

2 Осина, Н.А. Внутривидовая дифференциация *Francisella tularensis* с помощью полимеразной цепной реакции / Н.А. Осина, С.А. Портенко, Л.Н. Ивашенцева, А.Н. Куличенко // Генодиагност. инфекцион. болезни: Материалы Росс. науч.-практ. конф., 25-27 октября 2005 г. – Новосибирск, 2005. – С. 138-140.

3 Роль нуклеотидных замен в V домене 23S РНК *Francisella tularensis* в формировании резистентности туляремийного микроба к эритромицину / В.С. Тимофеев [и др.] // Актуальные проблемы эпидемиологии и профилактической

медицины: Материалы VI Всерос. науч.-практ. конф. молодых ученых и специалистов Роспотребнадзора. – 2014. – С. 112.

4 Выделение среднеазиатского подвида туляремийного микроба на территории Алтайского края / А.Н. Мокриевич [и др.] // Проблемы особо опасных инфекций. – 2013, № 1. – С. 66-69.

5 Мокриевич, А.Н. Молекулярно-генетические подходы к исследованию возбудителя туляремии для целей совершенствования диагностики и специфической профилактики: дис. ... канд. биол. наук / А.Н. Мокриевич. – М.: Московский науч.-исслед. ин-т эпидемиологии и микробиологии им. Г.Н. Габричевского, 2016. – 145 с.

6 Тимофеев, В.С. Генетическое разнообразие *Francisella tularensis* из природных очагов России: дис. ... канд. биол. наук / В.С. Тимофеев. – Оболенск: Гос. науч. центр прикладной микробиологии и биотехнологии, 2015. – 168 с.

7 Phylogeography of *Francisella tularensis*: global expansion of a highly fit clone / A.J. Vogler [et al.] // Journal of Bacteriology. – 2009. – V. 191, № 8. – P. 2474-2484.

8 Генотипирование методом мультилокусного VNTR-анализа штаммов *Francisella tularensis*, выделенных из природных очагов туляремии Ростовской области / А.С. Водопьянов [и др.] // Молекулярная генетика, микробиология и вирусология. – 2004, № 2. – С. 24-28.

9 Мультилокусное VNTR-типирование штаммов *Francisella tularensis* / А.С. Водопьянов [и др.] // Журнал микробиологии, эпидемиологии и иммунологии. – 2004, № 2. – С. 21-25.

10 ПЦР дифференциация подвидов *Francisella tularensis* с помощью одного праймера / Г.М. Вахрамеева [и др.] // Проблемы особо опасных инфекций. – 2011, № 1. – С. 46-48.

11 Differentiation of the *Francisella tularensis* subspecies by the INDEL typing method / V.M. Sorokin [et al.] // Journal of microbiology, epidemiology and immunobiology. – 2022. – V. 99, № 2. – P. 193-202.

12 Питательная среда для глубинного культивирования туляремийного

микроба / О.А. Волох [и др.] – Патент на изобретение RUS 2524687, 2014.

13 Анализ эпизоотологической и эпидемиологической ситуации по туляремии на территории Российской Федерации в 2023 г. и прогноз на 2024 г. / Т.Ю. Кудрявцева [и др.] // Проблемы особо опасных инфекций. – 2024, № 1. – С. 17-28.

14 A real-time PCR array for hierarchical identification of *Francisella* isolates / K. Svensson [et al.] // PloS one. – 2009. – V. 4, № 12. – P. E8360.

15 Сравнительный молекулярно-генетический анализ штаммов *Francisella tularensis*, изолированных в Ростовской области в 2020 г., и последовательностей геномов штаммов, выделенных в различных регионах мира / А.С. Водопьянов [и др.] // Проблемы особо опасных инфекций. – 2023, № 3. – С. 60-64.

