

**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**

Создание базы данных бактериоцинов и их продуцентов

Обучающийся / Student Толстая Василиса Олеговна
Факультет/институт/кластер/ Faculty/Institute/Cluster факультет биотехнологий
Группа/Group Т34305
Направление подготовки/ Subject area 19.03.01 Биотехнология
Образовательная программа / Educational program Биотехнология 2021
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Бакалавр
Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного инжиниринга, доцент (квалификационная категория "ординарный доцент")

Обучающийся/Student

Документ подписан	
Толстая Василиса Олеговна	
08.05.2025	

(эл. подпись/ signature)

Толстая
Василиса
Олеговна

(Фамилия И.О./ name
and surname)

Руководитель ВКР/
Thesis supervisor

Документ подписан	
Райко Михаил Петрович	
04.05.2025	

(эл. подпись/ signature)

Райко Михаил
Петрович

(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 Толстая Василиса Олеговна
Факультет/институт/кластер/ Faculty/Institute/Cluster факультет биотехнологий
Группа/Group Т34305
Направление подготовки/ Subject area 19.03.01 Биотехнология
Образовательная программа / Educational program Биотехнология 2021
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Бакалавр
Тема ВКР/ Thesis topic Создание базы данных бактериоцинов и их продуцентов
Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного инжиниринга, доцент (квалификационная категория "ординарный доцент")

Характеристика темы ВКР / Description of thesis subject (topic)

Название организации-партнера / Name of partner organization: Передовая инженерная школа ИТМО

Тема в области фундаментальных исследований / Subject of fundamental research: нет / not

Тема в области прикладных исследований / Subject of applied research: да / yes

Основные вопросы, подлежащие разработке / Key issues to be analyzed

Разработать базу данных бактериоцинов и их продуцентов, а также авторские классификации бактериоцинов, обеспечивающие удобный поиск бактериоцина под задачу пользователя базой данных.

Техническое задание:

1. Оценить доступность и полноту существующих баз данных бактериоцинов;
2. Рассмотреть имеющиеся классификации бактериоцинов и выделить ключевые параметры классификации;
3. Составить перечень бактериоцинов на основе литературных источников и провести их классификацию с учетом совокупности выявленных параметров классификации;
4. Получить геномы и аннотации геномов бактерий, потенциально способных к продукции бактериоцинов, из базы данных National Center for Biotechnology Information (NCBI);
5. Идентифицировать бактериоцины в аннотациях бактериальных геномов при помощи биоинформатических инструментов;
6. Провести поиск бактериоцинов в геномах бактерий, исходя из гомологии их последовательностей;

7. Разработать базу данных бактериоцинов и бактерий-продуцентов на основе полученных данных.

Первая часть работы будет посвящена обзору литературы по бактериоцинам и существующим базам данных. В ней будут рассмотрены современные подходы к систематизации бактериоцинов, описаны основные классы этих соединений, а также проведён анализ существующих ресурсов, содержащих информацию о бактериоцинах.

Во второй части работы будет представлен процесс поиска и систематизации бактериоцинов на основе литературных источников и баз данных. Будут рассмотрены возможные способы улучшения структуры и наполнения новой базы данных, а также предложены новые варианты классификации бактериоцинов.

Третья часть работы будет включать поиск бактериальных геномов и их аннотаций в специализированных базах данных. В ней будет подробно описан процесс идентификации бактериоцинов в аннотациях геномов по ключевым словам и гомологичным последовательностям. Также будет проведено извлечение генов и белковых последовательностей бактериоцинов.

Заключительная часть работы будет посвящена формированию структуры и созданию базы данных бактериоцинов. В ней будут приведены основные этапы разработки базы данных бактериоцинов, описаны использованные методы и инструменты, а также обсуждены перспективы её дальнейшего развития.

Цель работы: разработать пилотную базу данных, содержащую классифицированные последовательности бактериоцинов и аннотированные геномы бактерий-продуцентов.

Задачи:

1. Составить перечень бактериоцинов на основе литературных источников и провести их классификацию;
2. Получить аннотации геномов бактерий, потенциально способных к продукции бактериоцинов;
3. Идентифицировать бактериоцины в геномах бактерий при помощи биоинформатических инструментов;
4. Разработать базу данных бактериоцинов и бактерий-продуцентов на основе полученных данных.

Дата выдачи задания / Assignment issued on: 01.12.2024

Срок представления готовой ВКР / Deadline for final edition of the thesis 08.05.2025

СОГЛАСОВАНО / AGREED:


Руководитель ВКР/
Thesis supervisor

Документ подписан	
Райко Михаил Петрович	
21.03.2025	

Райко Михаил
Петрович

(эл. подпись)

Задание принял к
исполнению/ Objectives
assumed BY

Документ подписан	
Толстая Василиса Олеговна	
01.04.2025	

Толстая
Василиса
Олеговна

(эл. подпись)

Руководитель ОП/ Head
of educational program

Документ подписан	
Волох Олеся Игоревна	
07.04.2025	

Волох Олеся
Игоревна

(эл. подпись)

**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 Толстая Василиса Олеговна
Факультет/институт/кластер/ Faculty/Institute/Cluster факультет биотехнологий
Группа/Group Т34305
Направление подготовки/ Subject area 19.03.01 Биотехнология
Образовательная программа / Educational program Биотехнология 2021
Язык реализации ОП / Language of the educational program Русский
Квалификация/ Degree level Бакалавр
Тема ВКР/ Thesis topic Создание базы данных бактериоцинов и их продуцентов
Руководитель ВКР/ Thesis supervisor Райко Михаил Петрович, кандидат биологических наук, Университет ИТМО, передовая инженерная школа ИТМО интердисциплинарного инжиниринга, доцент (квалификационная категория "ординарный доцент")

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

Цель исследования / Research goal

разработать пилотную базу данных, содержащую классифицированные последовательности бактериоцинов и аннотированные геномы бактерий-продуцентов.

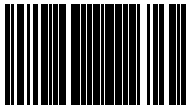
Задачи, решаемые в ВКР / Research tasks

1. Составить перечень бактериоцинов на основе литературных источников и провести их классификацию; 2. Получить аннотации геномов бактерий, потенциально способных к продукции бактериоцинов; 3. Идентифицировать бактериоцины в геномах бактерий при помощи биоинформатических инструментов; 4. Разработать базу данных бактериоцинов и бактерий-продуцентов на основе полученных данных.

Краткая характеристика полученных результатов / Short summary of results/findings

В ходе данной работы на основе литературных источников был составлен перечень известных на сегодняшний день бактериоцинов, а также получены их аннотации. Помимо этого по гомологии с уже известными бактериоцинами и наличию белковых мотивов, характерных для бактериоцинов, в геномах бактерий было найдены белки, которые потенциально могут быть неаннотированными бактериоцинами. На основе этой информации, а также информации об аннотированных геномах бактерий-продуцентов бактериоцинов была создана база данных бактериоцинов bacteriocin_database.

Обучающийся/Student

Документ подписан	
Толстая Василиса	

Толстая

Руководитель ВКР/
Thesis supervisor

Олеговна	
08.05.2025	

(эл. подпись/ signature)

Василиса
Олеговна

(Фамилия И.О./ name
and surname)

Документ подписан	
Райко Михаил Петрович	
04.05.2025	

(эл. подпись/ signature)

Райко Михаил
Петрович

(Фамилия И.О./ name
and surname)

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	9
1. ОБЗОР ЛИТЕРАТУРЫ	11
1.1. Антибиотики и проблема антибиотикорезистентности	11
1.1.2. Причины глобальной проблемы антибиотикорезистентности	12
1.1.3. Современные альтернативы антибиотиков	15
1.2. Бактериоцины	18
1.2.1. Применение бактериоцинов	20
1.2.2. Классификации бактериоцинов	24
1.3. Существующие базы данных бактериоцинов	28
2. МАТЕРИАЛЫ И МЕТОДЫ	32
2.1. Первичный анализ встречаемости бактериоцинов в геномах бактерий	32
2.2. Сбор данных об аннотированных бактериоцинах	34
2.3. Поиск новых потенциальных бактериоцинов	34
2.3.1. Сбор и обработка данных	34
2.3.2. Поиск новых потенциальных бактериоцинов по гомологии с уже известными бактериоцинами	35
2.3.3. Проверка новых потенциальных бактериоцинов на наличие белковых мотивов, характерных для бактериоцинов	36
2.4. Объединение данных по аннотированным и предсказанным бактериоцинам	37
2.5. Фильтрация данных по аннотированным и предсказанным бактериоцинам	39
2.6. Сбор данных об аннотированных геномах бактерий-продуцентов бактериоцинов	41
2.7. Создание базы данных бактериоцинов	42
3. РЕЗУЛЬТАТЫ	43
3.1. Первичная оценка встречаемости бактериоцинов в геномах бактерий	43
3.2. Характеристика данных, полученных о бактериоцинах при помощи биоинформатических инструментов	47
3.3. Характеристика объединенных и отфильтрованных данных по аннотированным и предсказанным бактериоцинами	47
3.4. Характеристика данных об аннотированных геномах бактерий-продуцентов бактериоцинов	49
3.5. Описание созданной базы данных бактериоцинов	50
ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ	52

ЗАКЛЮЧЕНИЕ	53
БЛАГОДАРНОСТИ	54
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	55

ВВЕДЕНИЕ

Бактериоцины – обширная группа пептидов и белков, секретируемых рибосомами бактерий, которые обладают антимикробной активностью в отношении как подобных или близкородственных бактериальных штаммов, так и других видов и родов бактерий посредством воздействия на специфические рецепторы бактериальной клетки [1, 2, 3]. Бактериоцины были открыты в 1925 г. французским ученым Андре Грация, и с того времени активно изучались не только механизмы их действия, но и потенциал их применения в промышленности [4].

На сегодняшний день бактериоцины уже нашли применение в пищевой промышленности (низин, микоцин и др.), медицине (грамицидин, колистин и др.) и ветеринарии (энтероцин, лактицин и др.), а также продолжает исследоваться потенциал применения бактериоцинов и в других отраслях (например, агрономии).

Однако наиболее актуальной областью исследования бактериоцинов на сегодняшний день является разработка новых антибактериальных препаратов. Бактериоцины, являющиеся пептидами с антимикробной активностью, могут сыграть ключевую роль в лечении бактериальных заболеваний в контексте глобальной проблемы антибиотикорезистентности.

Итак, в связи с высоким потенциалом применения бактериоцинов в различных областях промышленности была сформулирована цель данной выпускной квалификационной работы – разработать пилотную базу данных, содержащую классифицированные последовательности бактериоцинов и аннотированные геномы бактерий-продуцентов.

Для осуществления этой цели были поставлены следующие задачи:

1. Составить перечень бактериоцинов на основе литературных источников;
2. Получить аннотации геномов бактерий, потенциально способных к продукции бактериоцинов;
3. Идентифицировать бактериоцины в геномах бактерий при помощи биоинформатических инструментов;

4. Разработать базу данных бактериоцинов и бактерий-продуцентов на основе полученных данных.

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Антибиотики и проблема антибиотикорезистентности

В современном мире антибиотики играют огромную роль не только в медицине, но и животноводческой, аграрной, пищевой промышленности, предотвращая заражения патогенными бактериями или помогая справляться с их последствиями.

Пенициллин, первый в мире антибиотик, открытый сэром Флемингом в 1928 г., впервые начал массово применяться в 1940-х годах для лечения серьезных инфекций у солдат во время Второй Мировой войны. Однако уже в 1950-х годах человечество впервые столкнулось с проблемой антибиотикорезистентности – ситуацией, при которой бактерии, вызывающие инфекционные заболевания, становятся устойчивы к специфичным к ним антибиотикам [5]. Проблема резистентности к пенициллину побудила ученых создать новый класс антибиотиков – бета-лактамы, однако в то же десятилетие были выявлены случаи антибиотикорезистентности и к новому препарату. В дальнейшем та же картина наблюдалась и с другими открытыми антибиотиками. В конечном счете к 2015 году, почти через 70 десятилетий после открытия первого антибиотика, человечество вновь оказалось под серьезной угрозой неконтролируемых бактериальных инфекций [6]. Согласно исследованию Global Burden of Disease (2022) [7], в 2019 году в мире было зафиксировано 1,27 млн смертей непосредственно из-за устойчивости к противобактериальным препаратам, а также 4,95 млн смертей, которым способствовала антибиотикорезистентность.

Среди грамположительных бактерий в настоящее время наибольшую опасность представляют антибиотикорезистентные виды *Staphylococcus aureus* и *Enterococcus sp.* [6]: согласно исследованиям Golkar et al. (2014) и Gross (2013) [8, 9], метициллинрезистентный золотистый стафилококк

убивает больше американцев каждый год, чем ВИЧ/СПИД, болезнь Паркинсона, эмфизема и убийства вместе взятые.

Однако наибольшую опасность человечеству в контексте антибиотикорезистентности представляют грамотрицательные бактерии: они способны вырабатывать резистентность практически ко всем известным антибиотикам и создавать опасные эпидемиологические ситуации подобно тем, что были еще до открытия антибиотиков [6]. Наиболее серьезные вспышки инфекционных заболеваний, вызываемых антибиотикорезистентными грамотрицательными бактериями, регистрируются в больницах и чаще всего вызываются Enterobacteriaceae (в основном *Klebsiella pneumoniae*), *Pseudomonas aeruginosa* и *Acinetobacter* [10]. За пределами медицинских учреждений активно распространяются антибиотикорезистентные штаммы *Escherichia coli*, продуцирующие бета-лактамазу расширенного спектра, и *Neisseria gonorrhoeae* [11].

1.1.2. Причины глобальной проблемы антибиотикорезистентности

Гены бактерий могут наследоваться не только от их непосредственных бактерий-предшественников (родственников), но и путем горизонтального переноса генов при помощи плазмид – мобильных генетических элементов, передаваемых от одной бактерии к другой, даже если они принадлежат к разным видам. Таким образом бактерии, выжившие в среде с антибиотиком в результате естественного отбора, могут активно размножаться и делиться плазмидами с бактериями других штаммов [6]. Помимо механизма наследования или получения генов через плазмиды антибиотикорезистентность также может возникать благодаря спонтанным мутациям в геноме бактерий, что в конечном счете приводит к увеличению числа бактерий, устойчивых к антибиотикам, и росту проблемы антибиотикорезистентности.

Таким образом, формирование антибиотикорезистентности является естественным эволюционным процессом, который со временем протекает в геноме бактерий. Однако такая глобальная и стремительная проблема антибиотикорезистентности, которую мы наблюдаем сейчас во всем мире, спровоцирована чрезмерным потреблением антибиотиков.

Во многих странах продажа антибиотиков осуществляется без рецепта, что еще больше способствует увеличению и без того высокого уровня потребления препаратов этого класса. Однако и лечение антибиотиками под наблюдением врачей не всегда гарантирует 100% эффективность и рациональный подход к использованию антибиотиков [6]: согласно исследованиям Luyt et al. (2014) [12], подбор препарата или продолжительность лечения антибиотиками даже под наблюдением врача оказываются неверными в 30–50 % случаев, а 30-60% препаратов на основе антибиотиков, назначаемых в отделениях интенсивной терапии, оказываются избыточными, неподходящими или неоптимальными.

Неправильно подобранные антибиотики не только не лечат заболевание, но и могут вызывать осложнения: так, субингибиторные и субтерапевтические концентрации антибиотиков способствуют накоплению изменений в геноме бактерий посредством мутаций или горизонтального переноса генов, что в конечном счете может привести к повышению вирулентности и развитию антибиотикорезистентности [6]. Так, например, согласно Vishwanathan (2014) [13], низкие дозы антибиотиков способствуют диверсификации штаммов *Pseudomonas aeruginosa* и вызывают широкие протеомные изменения у *Bacteroides fragilis*.

Кроме того, антибиотики широко применяются в сельском хозяйстве для предотвращения инфицирования и стимуляции роста скота. Как показало исследование Klein и соавторов (2023) [14], по состоянию на 2020 год во всем мире около около 99 500 тонн антибиотиков использовалось именно на скоте, а к 2030 году научное сообщество прогнозирует рост этого показателя еще на 8%.

Таким образом, даже если человек не принимает антибактериальные препараты, какое-то количество антибиотиков попадает в его организм вместе с мясом и молочной продукцией, полученной от сельскохозяйственных животных. Более того, через пищу животного происхождения в организм человека попадают и резистентные бактерии скота, эволюционировавшие благодаря активному использованию антибиотиков.

Антибиотики, назначаемые скоту, влияют не только на микробиом отдельно взятых людей, но на микробиом окружающей среды, попадая в нее через продукты их метаболизма и способствуя росту числа антибиотикорезистентных бактерий.

Что касается фармпромышленности, то в последние десятилетия наблюдается существенный спад в разработке новых антибиотиков. Это обусловлено как нормативными, так и экономическими аспектами: в силу краткосрочности лечения, возникновения антибиотикорезистентности и относительно низкой стоимости за дозу разработка и производство новых антибиотиков не так экономически выгодны, как, например, разработка препаратов против хронических заболеваний и рака.

Антибиотикорезистентность является глобальной проблемой, затрагивающей не только отдельные группы населения, непосредственно столкнувшихся с этой проблемой при лечении заболевания, но и всю систему здравоохранения страны. Так, например, для пациентов с антибиотикорезистентными инфекциями время пребывания в больнице увеличивается на 6,4–12,7 дня, что в совокупности добавляет восемь миллионов дополнительных больничных дней [15]. В конечном счете по состоянию на 2013 год антибиотикорезистентные инфекции наносят ежегодный урон экономике США в 20 миллиардов долларов США в виде расходов на здравоохранение и 35 миллиардов долларов США в виде потери производительности [6]. По оценкам Всемирного банка, к 2030 году проблема антибиотикорезистентности может привести к потерям мирового внутреннего

валового продукта в размере от 1 триллиона до 3,4 триллиона долларов США [16].

1.1.3. Современные альтернативы антибиотиков

Для решения проблемы антибиотикорезистентности важно не только снижать уровень использования антибиотиков в медицине и сельском хозяйстве, разрабатывать новые антибиотики, совершенствовать диагностику бактериальных инфекций и осуществлять более строгий контроль использования препаратов на основе антибиотиков, но и разрабатывать альтернативные подходы для решения проблем, вызванных бактериями [17].

На сегодняшний день существует ряд соединений, которые можно использовать как альтернативы антибиотикам. Рассмотрим некоторые примеры таких альтернатив:

1. Антимикробные адъюванты – соединения, не обладающие непосредственным антимикробным действием, но способные ингибировать механизмы резистентности бактерий, тем самым увеличивая эффективность антибиотиков [18]. Механизмы действия адъювантов могут быть различны: так, некоторые адъюванты воздействуют на клеточную стенку бактерии с целью увеличения ее проницаемости для антибиотика, тогда как другие блокируют эффлюксные насосы, выводящие токсичные для бактериальной клетки вещества наружу [19].

В качестве примера антимикробных адъювантов можно привести производные пиррола, которые ингибируют эффлюксные транспортеры семейства RND (Resistance-Nodulation-Division), обеспечивающие выведение антибиотиков из бактериальных клеток и обуславливающие резистентность грамотрицательных бактерий к множеству антибиотиков [18]. Введение антимикробных адъювантов на основе

пиррола способствовало увеличению восприимчивости мультирезистентных штаммов *E. coli*, *P. aeruginosa* и *K. pneumoniae* к нескольким антибиотикам *in vitro* и снижению бактериальной нагрузки *in vivo* [18].

В качестве главных недостатков терапии антимикробными адъювантами [20] выделяют их токсичность и потенциальный чрезмерный иммунный ответ со стороны организма человека, что существенно ограничивает их применение адъювантов в лечении бактериальных инфекций.

2. Бактериофаги – вирусы, специфично поражающие определенные группы бактерий. Механизм воздействия бактериофага на бактериальную клетку основан на адсорбции бактериофага на поверхности клеточной стенки бактерии, введении генетического материала внутрь клетки с его последующей репликацией и лизисом бактериальной клетки [19]. Высокая специфичность бактериофагов обуславливает потенциал их использования в антибактериальной терапии. Еще одним преимуществом бактериофагов является их экологическая безопасность и эволюционная адаптивность, позволяющая стремительно приспосабливаться к новым защитным механизмам бактерий [18]. Однако у фаготерапии существует и ряд серьезных недостатков: низкая стабильность, высокие требования к условиям хранения и транспортировки, иммунный ответ на бактериофаги при их введении пациенту и, безусловно, сложности с нормативно-правовыми аспектами введения таких препаратов на фармацевтический рынок [18]. На данный момент фаготерапия используется преимущественно как один из методов персонализированной медицины, нежели как общедоступная практика.

3. Наноантибиотики – молекулы антибиотиков размером 1–100 нм или наночастицы с прикрепленными к ним антибиотиками. В отличие от классических антибиотиков, антибиотик размером с наночастицу или доставка антибиотиков при помощи наночастиц обеспечивает большее антибактериальное действие [18]. Механизм воздействия наноантибиотиков на бактериальные клетки может быть различным: наночастицы серебра, меди, цинка и ряда других металлов, потенциально входящих в состав наноантибиотиков, повреждают клеточную стенку бактерий, тогда как наночастицы оксидов металлов вызывают окислительный стресс клетки, а другие наночастицы способствуют разрушению защитной биопленки бактерий [21].

Так, например, наноантибиотики на основе наночастиц оксида цинка показали хорошие результаты в подавлении роста *S. aureus*, *S. pyogenes*, *E. coli*, *K. pneumoniae*, при этом с увеличением дозы росла и эффективность наночастиц [18].

Однако у наноантибиотиков существует и ряд недостатков, ограничивающих их широкое применение в медицине. К недостаткам наноантибиотиков относятся их токсичность, низкая стабильность в биологических средах, дороговизна и сложность масштабирования производства, а также регуляторные барьеры со стороны контролирующих органов [21].

4. Иммуноterapia – подход в лечении бактериальных и вирусных инфекций, основанный на моделировании ответов как врожденного, так и приобретенного иммунитета человека и включающий в себя терапию препаратами на основе антител или Т-клеток, вакцин, регулирования уровня цитокинов и др. подходами [18, 22].

Несмотря на перспективы использования иммунотерапии в качестве альтернативы антибиотикам, у каждого из вариантов иммунотерапии есть и свои недостатки. Серьезными недостатками лечения цитокинами являются различные системные воспаления

(вплоть до повреждения отдельных органов) и проблемы с таргетной доставкой, из-за чего такой подход к лечению инфекций на данный момент целесообразен лишь в виде комбинированной терапии с другими препаратами. Что касается лечения бактериальных инфекций препаратами на основе антител, Т-клеток или вакцин, то главными недостатками таких подходов являются постоянные мутации патогенов, и, соответственно, потеря специфичности таких препаратов [22].

5. Бактериоцины – антимикробные пептиды, вырабатываемые бактериями против узкого или, напротив, широкого круга других бактерий. Существует целый ряд механизмов цитотоксичности бактериоцинов в отношении бактериальных клеток, однако все они основаны на взаимодействии бактериоцина со специфическими рецепторами на поверхности клетки с последующим нарушением ее целостности через воздействие на клеточную стенку или мембрану [23].

В отличие от других альтернатив антибиотиков, к главным преимуществам бактериоцинов относят их нетоксичность и совместимость с нормальной микробиотой человека, однако по сей день все еще существует ряд сложностей, не позволяющих в полной мере вывести бактериоцины на рынок антимикробных препаратов. Об этом подробнее будет рассказано в следующем разделе.

1.2. Бактериоцины

Бактериоцины – обширная группа пептидов и белков, секретируемых рибосомами бактерий, которые обладают антимикробной активностью в отношении как подобных или близкородственных бактериальных штаммов, так и других видов и родов бактерий посредством воздействия на специфические рецепторы бактериальной клетки [1, 2, 3].

Гены, кодирующие бактериоцины, в зависимости от типа бактерии располагаются в нуклеотиде, плазмидах или других мобильных генетических элементах, и вместе с вспомогательными белками формируют опероны. Для их транскрипции необходима аутоиндукция или внешняя индукция посредством пептидов, накопленных в окружающей среде [1].

Механизм разрушения бактериальной клетки под действием бактериоцинов основан на способности бактериоцинов связываться с бактериальной клеткой через специфические рецепторы на ее поверхности, а затем биохимически разрушать клетку через образование мембранных каналов и пор на ее поверхности [1].

Наиболее широко изученным источником бактериоцинов в природе является почва, однако также исследуются бактериоцины в донных отложениях, микробиоте человека и животных [1]. Наибольшая концентрация бактериоцинов в теле человека зафиксирована в микробиоте влагалища, дыхательных путей и полости рта; наименьшая - в микробиоте кишечника [2]. Бактериоцины, входящие в состав микробиоты, согласно исследованию отечественных ученых (Иванова, Петров, 2021) [2], не только способствуют снижению числа патогенных и условно-патогенных бактерий, но также принимают участие в стимуляции лимфоидной ткани, образовании витаминов и др.

Что касается продуцентов бактериоцинов, то гены, кодирующие эти антимикробные соединения, были обнаружены в геномах как грамположительных, так и в грамотрицательных бактерий и архей, однако большая часть известных ныне бактериоцинов синтезирована именно грамположительными бактериями. Это обуславливает потенциал широкого применения бактериоцинов в биотехнологической промышленности как в фармацевтической, так и пищевой отраслях. Подробнее о применении бактериоцинов описано в разделе 1.2.1 «Применение бактериоцинов».

Основными преимуществами бактериоцинов являются их высокая специфичность действия, низкая токсичность, совместимость с резидентной

микробиотой, а также возможность внесения различных модификаций в их структуру при помощи генной инженерии [24]. К недостаткам бактериоцинов можно отнести их ограниченную активность *in vivo*, а также сложности с таргетной доставкой и высокую стоимость производства [24].

1.2.1. Применение бактериоцинов

1. Пищевая промышленность

Первым бактериоцином, который нашел широкое применение в промышленности, стал низин – наряду с лизоцимом, лактопероксидазой, хитозаном и другими альтернативами классических консервантов, негативно сказывающимися на здоровье человека, этот бактериоцин стал применяться в пищевой промышленности в качестве пищевой добавки E234 [25].

Выделенный из *Lactococcus lactis*, низин является натуральным консервантом микробного происхождения: он подавляет рост клеток широкого спектра грамположительных бактерий и их спор, тем самым и продлевая срок годности мясных продуктов, сыра Чеддер и пива [25]. Наряду с низином в качестве пищевых добавок FDA (Food and Drug Administration) также одобрила бактериоцины педиоцин PA-1 и микоцин [26], которые подавляют рост грамположительных бактерий (например, *Listeria monocytogenes*) и используются при производстве консерв, плавленного сыра и других молочных продуктов [26].

Бактериоцины нашли широкое применение в пищевой промышленности в роли консервантов не только ввиду своего природного происхождения, но также благодаря тому, что они не имеют ни цвета, ни запаха, ни вкуса и остаются стабильны при высоких температурах, низком pH и широком диапазоне концентраций солей [26]. Однако при использовании бактериоцинов в качестве консервантов важно помнить, что в силу своей природы они обладают ограниченным спектром действия и не могут быть

применимы против широкого диапазона бактерий, а также обладают гидрофобными свойствами, что важно учитывать во избежание фракционирования жиров в продукте [26].

2. Медицина

Помимо пищевой промышленности некоторые бактериоцины нашли широкое применение в медицине и ветеринарии в роли антимикробных препаратов. Так, например, бактериоцин грамицидин, синтезируемый бактериями *Bacillus brevis*, используется для производства препарата Грамицидин С (Gramisidin S) в форме сублингвальных таблеток, которые назначаются при инфекционно-воспалительных заболеваниях рта и горла, или раствора и пасты для наружного применения [27]. Бактериоцин грамицидин и лекарственные препараты на его основе обладают бактериостатическим действием против *Streptococcus pneumoniae*, менингококков и гонококков, бактерицидным действием против *Streptococcus spp.* и *Staphylococcus spp.*, а также способствует обезболиванию и увеличению слюноотделения. [27].

Еще одним бактериоцином, используемым в фармпромышленности, является колистин. Этот бактериоцин продуцируется бактериями *Bacillus subtilis* и *Bacillus polymyxa* и используется для производства одноименного препарата Колистин (Colistin) в форме порошка для приготовления ингаляционных растворов. Препарат активен против *Pseudomonas aeruginosa*, *Haemophilus influenzae* и *Acinetobacter spp.* и назначается для лечения инфекций дыхательных путей и муковисцидоза [27].

Колистин, выделенный из *Bacillus colistinus*, а также полимиксин из *Bacillus polymyxa* активны против широкого спектра грамотрицательных бактерий, что позволяет использовать препараты на основе этих бактериоцинов в качестве монотерапии или в комбинации с другими

антибактериальными средствами (например, антибиотиками тетрациклином и сульфонамидами) [27].

Банеоцин (Baneocin, «Сандоз»), один из наиболее широко применяемых препаратов для лечения инфекций кожных покровов, также имеет в своем составе бактериоцины. Так, например, для производства этого препарата из *Bacillus subtilis* выделяют бактериоцин бацитрацин, а из *Streptomyces fradiae* - неомицин. Эти бактериоцины имеют различные механизмы воздействия на бактериальные клетки: бацитрацин препятствует синтезу клеточной стенки бактерий путем нарушения транспорта пептидогликанов, а неомицин воздействует на рибосомы бактерий, тем самым влияя на синтез белков; однако оба этих бактериоцина оказывают антибактериальное действие против клостридий, коринебактерий, стафилококков и др. бактерий [27].

3. Ветеринария

В настоящее время также широко исследуется потенциал применения бактериоцинов для терапии бактериальных инфекций животных. Так, например, одним из ведущих направлений, в котором используются антибиотики в сельском хозяйстве, является лечение мастита высокопродуктивных коров [28]. Мастит – это заболевание молочных желез, чаще всего вызываемое бактериями *Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcus dysgalactiae* и *Escherichia coli*, которое приводит к снижению объемов и качества удоев. Наибольшую опасность представляет субклиническая форма мастита, которую крайне сложно диагностировать из-за отсутствия видимых признаков, и, соответственно, предотвратить бактериальную обсемененность молока, взятого от таких коров [28].

Ввиду растущей мировой проблемы антибиотикорезистентности разрабатываются все больше новых подходов к альтернативному лечению и профилактике мастита коров. К таким подходам относится вакцинирование, применение бактериофагов и бактериоцинов. В статье Сидоровой (2022) [28]

описывается применение композиции на основе бактериоцина низина для терапии мастита, вызванного различными ассоциациями бактерий, демонстрирующих различный уровень антибиотикорезистентности (от 6,3% до 62,5% в зависимости от вида бактерий и группы антибиотиков). Согласно этому исследованию, на 14 день лечения мастита композицией на основе бактериоцина низина в 88,6% проб было зафиксировано отсутствие роста патогенной микрофлоры вымени.

Помимо низина в ветеринарии широко исследуется применение и других бактериоцинов, выделенных из молочнокислых бактерий. В статье [29] описывается потенциал использования бактериоцинов энтероцина, лактицина и мацедоцина для антибактериальной терапии как скота, так и птиц и мелких животных. Так, например, бактериоцин энтероцин, подавляющий рост бактерий *Listeria monocytogenes*, показал хорошие результаты в исследованиях альтернативной терапии зубного пародонтита собак, а бактериоцины лактицин и мацедоцин – потенциал применения для лечения мастита крупного рогатого скота [29].

4. Агрономия

Помимо пищевой промышленности, медицины и ветеринарии в настоящее время также исследуется потенциал применения бактериоцинов для борьбы с заболеваниями растений. Бактериоцины являются многообещающими альтернативами агрохимикатов, широко используемых в сельском хозяйстве во всем мире.

Так, например, в исследовании [30] описано применение тайлоцинов SF4с против бактерий *Xanthomonas sp.* и *Pseudomonas sp.*, вызывающих бактериальную пятнистость тепличных томатов. Было установлено, что при внекорневой обработке томатов эти бактериоцины не только уменьшали симптомы бактериальной пятнистости у инфицированных томатов, но также препятствовали распространению этой инфекции на здоровые растения.

Механизм бактерицидного действия тайлоцинов SF4с основан на избирательном разрушении клеточной стенки восприимчивых к этим бактериоцинам бактерий, что делает тайлоцины SF4с перспективным объектом исследования в качестве альтернатив широко используемых бактерицидов на основе меди: повсеместное использование этих агрохимикатов не только способствует возникновению плазмидных генов устойчивости к меди, но и наносит вред здоровью человека и животных [30].

Помимо *Xanthomonas sp.* и *Pseudomonas sp.* существует ряд других фитопатогенных бактерий, относящихся к родам *Pectobacterium* и *Agrobacterium*. Эти бактерии вызывают бактериальные заболевания не только томатов, но и риса, банана, картофеля и ряда других важных для мировой экономики сельскохозяйственных растений [31]. По этой причине поиск альтернативного антибактериального препарата остается крайне актуальной задачей. Бактериоцины же, в свою очередь, являются перспективными антимикробными агентами, которые могут как наноситься непосредственно на растение, так и экспрессироваться в его геноме, являясь безопасной природной альтернативой агрессивных химикатов [31].

1.2.2. Классификации бактериоцинов

На данный момент известно 63 различных бактериоцинов: ацидозин (acidocin), актагардин (actagardine), агроцин (agrocin), альвеицин (alveicin), ауреоцин (aureocin), ауреоцин A53 (aureocin A53), ауреоцин A70 (aureocin A70), бисин (bisin), карнозин (carnocin), карноциклин (carnocyclin), казеицин (caseicin), цереин (cerein), циркулярин A (circularin A), колицин (colicin), курвацицин (curvaticin), диверцин (divercin), дурамицин (duramycin), энтероцин (enterocin), энтеролизин (enterolysin), эпидермин (epidermin), галлидермин (gallidermin), эрвиниоцин (erwinioicin), гардимицин (gardimycin), гассерицин A (gassericin A), глицинецин (glycinecin), галоцин (halocin), галоурацин (haloduracin), клебицин (klebicin), лактоцин S (lactocin S),

лактококцин (lactococcin), лактицин (lacticin), лейкоцин (leucococcin), лизостафин (lysostaphin), македонцин (macedocin), мерсацидин (mersacidin), мезентерицин (mesentericin), микробиспорицин (microbisporicin), микроцин (microcin), микроцин S (microcin S), мутацин (mutacin), низин (nisin), пенибациллин (paenibacillin), планоспорицин (planosporicin), педиоцин (pediocin), пентозин (pentocin), плантарицин (plantaricin), пневмоциклицин (pneumocyclacin), пиоцин (pyocin), реутерицин 6 (reuterin 6), реутерицилин (reutericyclin), реутерин (reuterin), сакацин (sakacin), саливарицин (salivaricin), субланцин (sublancin), субтилилин (subtilin), сульфолобицин (sulfolobacin), тасманцин (tasmanicin), турицин 17 (thuricin 17), трифолитоксин (trifolitoxin), вариацин (variacin), вибриоцин (vibriocin), варнерицин (warnericin) и варнерин (warnerin) [32].

Наиболее широко используемая классификация бактериоцинов разделяет все бактериоцины на 3 группы в зависимости от типа организма, из которого они были получены: грамотрицательные бактерии, грамположительные бактерии и археи. Внутри этих больших групп также выделяются классы: так, грамотрицательные бактерии делятся на 4 класса, грамположительные – на 3 класса с подклассами, археи – на 2 класса [33]. Такая классификация основана на химическом строении бактериоцинов.

Классификация уделяет внимание посттрансляционным модификациям и третичной структуре белка, аминокислотному составу, полярности, молекулярной массе и другим физико-химическим характеристикам бактериоцинов [34]. Такая классификация приведена в таблице 1 [33]:

Таблица 1 – Наиболее широко используемая классификация бактериоцинов

Тип бактерии	Класс	Размер (кДа)	Примеры
Грамотрицательные бактерии	Колицины	30–80	Колицины А, В, Е2, Е3
	Колицин-подобные бактериоцины	30–80	S-пиоцины, клебицины

Продолжение таблицы 1

Грамотрицательные бактерии	Бактериоцины, подобные хвосту фага	20–100	R- и F-пиоцины
	Микроцины	<10	Микроцин C7, микроцин B17, колицин V
Грамположительные бактерии	Класс I	<5	Низин, мерсадин, лакцин 3147
	Класс II	<10	Педиоцин RA1, карнобактериоцин B2
	Класс III	>10	Гельвецин, энтероцин AC-48
Археи	Галоцины	>5	Галоцин A4, C8, H1, H4
	Sulfolobaceae	~20	Sulfolobaceae

Существует и ряд альтернативных классификаций бактериоцинов. Самую первую классификацию бактериоцинов предложил Д. Бредли в 1967 г. Эта классификация была основана на химической природе и свойствах молекул и разделяла все известные на тот момент бактериоцины на 3 класса [35]:

1. Термостабильные и невидимые под электронным микроскопом бактериоцины с низкой молекулярной массой, чувствительные к трипсину и не осаждающиеся при центрифугировании;
2. Термолабильные и видимые под электронным микроскопом бактериоцины с высокой молекулярной массой, нечувствительные к трипсину и осаждающиеся при центрифугировании;
3. Бактериоцины с ярко выраженной ферментативной активностью.

В настоящее время классификация Д. Бредли уже не используется, однако ввиду отсутствия единой общепринятой классификации

бактериоцинов в научной литературе можно встретить авторские классификации бактериоцинов, похожие на классификацию Д. Бредли [36].

Кроме уже перечисленных классификаций существует классификация бактериоцинов, в основе которой лежит механизм воздействия бактериоцина на бактериальную клетку. Согласно этой классификации, все бактериоцины можно разделить на 4 класса [35]:

1. Бактериоцины, ингибирующие окислительное фосфорилирование;
2. Бактериоцины, разрушающие ДНК;
3. Бактериоцины, блокирующие синтез белков;
4. Бактериоцины, нарушающие проницаемость цитоплазматической мембраны.

Бактериоцины также можно подразделять на классы согласно типу их химической структуры. Так, выделяют циклические и линейные бактериоцины. С точки зрения практического применения наибольший интерес представляют бактериоцины циклической структуры, имеющие связь между N- и C-концами пептида и подразделяющиеся на 2 подкласса в зависимости от значения их изоэлектрической точки. Бактериоцины линейной структуры подразделяются на 3 подгруппы в зависимости от их трехмерной структуры [37].

Помимо различий в химической структуре в [34, 37] также описывается классификация бактериоцинов, основанная на взаимосвязи их последовательности и структуры: так, существуют α -спиральные, β -листовые, петлевые или расширенные структуры, а также альтернативная структурная классификация с использованием углов скручивания пептидного остова.

В обзорной статье журнала *Antibiotics* (Garcia-Gutierrez et al., 2020) [38] упоминается классификация бактериоцинов по консенсуальным

последовательностям, то есть усредненным нуклеотидным последовательностям, которые чаще всего встречаются в бактериоцинах.

В статье Brown и коллег (2024) [34] описывается классификация бактериоцинов как post-translationally modified peptides (RiPP), так как все бактериоцины являются RiPP. Однако не все RiPP являются бактериоцинами, поэтому для разделения бактериоцинов в рамках RiPP используется суженная версия классификации RiPP, описывающая только те RiPP, к которым относятся бактериоцины.

Таким образом, несмотря на распространенность отдельных классификаций бактериоцинов, разработка общепринятой классификации бактериоцинов всё ещё остаётся актуальной задачей [34, 39].

1.3. Существующие базы данных бактериоцинов

База данных – это структура, предназначенная для упорядоченного хранения и модификации каких-либо данных. Стоит отличать этот термин от системы управления базой данных (СУБД), которая является программным комплексом для управления базами данных [40].

На текущий момент существует множество видов баз данных: графовые, объектные, реляционные и др., и каждый из этих видов имеет свои особенности, подходящие под конкретные данные [40]. Данная работа фокусируется на реляционных базах данных. Главным элементом реляционной базы данных являются таблицы: таблицы состоят из строк (записей) и столбцов (полей), и все таблицы базы данных связаны между собой ключами [40]. Базы данных находят применение в медицине и науке, упорядочивая массивы накопленных данных.

Одной из наиболее известных баз данных бактериоцинов, которая часто упоминается в контексте их исследования, является BACTIBASE [41]. Согласно работе Smith и соавторов (2010) [41], эта база данных позволяет не

только получить информацию об аннотированных бактериоцинах, но также произвести поиск гомологичных последовательностей при помощи BlastP, применения скрытых марковских моделей и молекулярного моделирования.

BACTIBASE содержит не только аминокислотные последовательности бактериоцинов, выделенные из грамположительных и грамотрицательных бактерий, из базы данных UniProt и научных статей PubMed, но и рассчитанные и предсказанные физико-химические свойства ряда бактериоцинов [41].

Запись о каждом бактериоцине содержит информацию о его уникальном шестизначном идентификационном номере (ID), названии пептида, последовательности, классе, микробных данных, физико-химических данных (эмпирическая формула, масса, изоэлектрическая точка и др.) и веб-ссылку на UniProt, если пептид уже существует в этой базе данных [42].

Согласно информации Anderson и соавторов (2007) [42], BACTIBASE находится в свободном доступе по адресу <http://bactibase.pfba-lab.org>, однако этот сайт не доступен ни на территории РФ, ни на территории Италии, Голландии, Германии, Англии (актуально на весну 2025 года).

Другая существующая база данных BUR (Bacteriocins of URMITE) фокусируется на бактериоцинах из микробиоты кишечника, и содержит как известные в BAGEL, BACTIBASE, NCBI бактериоцины, так и предсказанные: новые бактериоцины искались помощи BLAST по гомологии с аминокислотными последовательностями известных бактериоцинов с совпадением аминокислот более 50% и покрытием более 70% [43].

Согласно заявлениям авторов [43], последовательности геномов организмов из желудочно-кишечного тракта брались не только из NCBI, но также секвенировались в их лаборатории.

База данных BUR содержит информацию о названии бактериоцина, его классе, организме-продуcente, а также физико-химические данные, такие как

длина, масса, аминокислотный состав, заряженные, гидрофобные и гидрофильные остатки [43].

Ссылка <https://drissifatima.wix.com/bacteriocins> на базу данных BUR также неактивна (актуально на весну 2025 года), и получить доступ к этой базе данных не представляется возможным из-за отсутствия подключения домена к сайту.

Следующая база данных бактериоцинов NucleBact находится в разработке. По заявлениям разработчиков [44], эта база данных будет содержать информацию о нуклеазных бактериоцинах, специфичных к близкородственным штаммам бактерий.

При помощи NucleBact можно будет проводить филогенетический анализ нуклеазных бактериоцинов и последовательностей иммунитета, а также прогнозировать восприимчивость к бактериоцину среди γ -протеобактерий, что откроет новые возможности для разработки антибактериальных и онкологических препаратов узкого спектра действия, а также пищевых консервантов [44].

Четвертая база данных бактериоцинов называется BAGEL. BAGEL – это инструмент, разработанный для идентификации предполагаемых открытых рамок считывания бактериоцинов в последовательности ДНК. Для успешной реализации этого инструмента разработчики также собрали базу данных мотивов и бактериоцинов, с данными из которых и производится сравнение загружаемых пользователем последовательностей пептидов при скрининге [45].

По заявлениям разработчиков [45], последовательности бактериоцинов в базе данных были получены из SRS ExPASy, NCBI, UniProt, а также различных научных публикаций. Помимо аннотаций последовательностей бактериоцинов в базе данных бактериоцинов BAGEL также содержится информация об их классификации и ссылка на соответствующий белок в UniProt [45].

К сожалению, база данных бактериоцинов BAGEL также недоступна по указанной в [45] ссылке <http://bioinformatics.biol.rug.nl/bacteriocin> ни на территории РФ, ни на территории Германии, Голландии, Италии, Англии (актуально на весну 2025 года).

Еще одна база данных LABiocin содержит информацию исключительно о бактериоцинах, выделенных из молочнокислых бактерий, и создана для содействия разработки новых лекарственных препаратов и пищевых консервантов [46].

Каждая запись о бактериоцине в базе данных LABiocin помимо его аминокислотной последовательности также сопровождается информацией о классе, физико-химических свойствах, бакетрии-продуcente и ссылками на данные об этом пептиде на NCBI, UnitProt, Bactibase. Помимо классического поиска интересующей информации по названию бактериоцина в LABiocin также доступен BLAST-поиск гомологий для загружаемого пользователем сиквенса [46].

База данных LABiocin доступна по ссылке <https://labiocin.univ-lille.fr/> по перварительной регистрации на сайте.

Таким образом, на территории РФ из всех описанных баз данных доступна только база данных LABiocin, содержащая информацию о бактериоцинах, полученных из молочнокислых бактерий.

Подводя итог, разработка доступной на территории РФ базы данных бактериоцинов, содержащей информацию о бактериоцинах как из грамположительных, так и грамотрицательных бактерий, является актуальной задачей.

2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Первичный анализ встречаемости бактериоцинов в геномах бактерий

Перед началом работы с бактериальными геномами на основании [32] был составлен удобный для работы в Python список названий всех известных на данный момент бактериоцинов, описанных в разделе 1.2.2. «Классификации бактериоцинов». Список включал названия 63 различных бактериоцинов на латинском языке.

Далее для поиска этих бактериоцинов в геномах бактерий на персональный компьютер через Windows Subsystem for Linux (WSL) были скачаны аннотации всех бактериальных геномов с полными геномными сборками из базы данных NCBI [47]. Всего было получено порядка 100 000 геномных сборок и аннотаций. Геном каждой бактерии был помещен в отдельную папку, название которой – уникальный идентификатор (ID) сборки генома, присваиваемый базой данных NCBI.

Далее был написан скрипт на Python [48] для прямого поиска бактериоцинов в аннотациях геномов бактерий по названиям бактериоцинов с использованием библиотеки Biopython [49]. Из каждой аннотации была получена следующая информация: название бактериоцина (например, microcin), полное имя бактерии (например, *Escherichia coli* K12) и ID сборки генома (например, GCA_000005825.2). Результаты работы скрипта изображены на рисунке 1.


```

Bacteriocin: microcin, Organism: 398511, Folder: GCA_000005825.2
Bacteriocin: colicin, Organism: 398511, Folder: GCA_000005825.2
Bacteriocin: microcin, Organism: 511145, Folder: GCA_000005845.2
Bacteriocin: colicin, Organism: 511145, Folder: GCA_000005845.2
Bacteriocin: colicin, Organism: 511145, Folder: GCA_000005845.2
Bacteriocin: colicin, Organism: 511145, Folder: GCA_000005845.2
Bacteriocin: aureocin, Organism: 306537, Folder: GCA_000006605.1
Bacteriocin: colicin, Organism: 187410, Folder: GCA_000006645.1
Bacteriocin: colicin, Organism: 187410, Folder: GCA_000006645.1

```

Рисунок 1 – Выходные данные авторского скрипта на Python для прямого поиска бактериоцинов в аннотациях геномов бактерий

Для удобства дальнейшего анализа все ID сборок генома были конвертированы в названия бактерий при помощи Biopython [49] и API NCBI Taxonomy [50]. Помимо конвертации было произведено удаление дубликатов строк: дубликатами считались строки, содержащие одно и то же ID сборки и повторяющиеся названия бактериоцинов. Принцип преобразования представлен на рисунке 2.

```

Bacteriocin: circularin, Organism: [Bacillus] caldolyticus, Folder: GCA_003595605.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus SCF1, Folder: GCA_000164865.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus SCF1, Folder: GCA_000164865.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus SCF1, Folder: GCA_000164865.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus SCF1, Folder: GCA_000164865.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus SCF1, Folder: GCA_000164865.1
Bacteriocin: colicin, Organism: [Acidovorax] ebreus TPSY, Folder: GCA_000022305.1
Bacteriocin: microcin, Organism: [Enterobacter] lignolyticus, Folder: GCA_001461805.1
Bacteriocin: colicin, Organism: [Enterobacter] lignolyticus, Folder: GCA_001461805.1

```

Рисунок 2 – Результаты работы скрипта на Python для конвертации ID сборки генома в названия бактерий и удаления дубликатов строк

Для дальнейшей работы важно понимать, что получившиеся данные можно использовать только для первичной оценки встречаемости бактериоцинов в геномах бактерий: помимо собственно бактериоцинов прямой поиск по названиям находит вспомогательные белки, ассоциированные с бактериоцинами. Так, например, слово «colicin» несколько раз встретилось в аннотации генома GCA_000164865, однако при более детальном ее изучении часть совпадений оказывается продуктами, которые, например, помогают поглощать колицин или связываться с ним.

Вспомогательные белки, не являющиеся бактериоцинами, были убраны на последующих этапах фильтрации данных.

2.2. Сбор данных об аннотированных бактериоцинах

На следующем этапе работы были получены аминокислотные последовательности уже известных бактериоцинов из базы данных UniProt [51] путем применения фильтра по запросу «Bacteriocin» в аннотациях белков. Все аминокислотные последовательности аннотированных бактериоцинов были собраны в один файл формата fasta. Этот файл впоследствии стал основой для фильтрации известных бактериоцинов.

2.3. Поиск новых потенциальных бактериоцинов

2.3.1. Сбор и обработка данных

Для поиска новых потенциальных бактериоцинов необходимо получить данные о всех доступных белках бактерий, исключив из них белки, уже аннотированные как бактериоцины. Для этого из базы данных RefSeq NCBI [52] были скачаны аминокислотные последовательности всех белков (в том числе бактериоцинов) всех бактерий, для которых доступны полные геномные сборки.

Далее при помощи скрипта на Python [48] с использованием библиотеки Biopython [49] из fasta-файла с аминокислотными последовательностями всех белков бактерий из NCBI были удалены аминокислотные последовательности известных бактериоцинов, содержащихся в fasta-файле, полученном из базы данных UniProt. Таким образом, на этом этапе был получен fasta-файл с аминокислотными последовательностями белков бактерий, которые потенциально могут быть гомологичны бактериоцинам.

2.3.2. Поиск новых потенциальных бактериоцинов по гомологии с уже известными бактериоцинами

Для поиска новых потенциальных бактериоцинов по гомологии с уже известными бактериоцинами использовался биоинформатический инструмент DIAMOND [53].

Для работы с DIAMOND файл с аминокислотными последовательностями потенциальных бактериоцинов, полученный на предыдущем этапе, был разделен на 13 примерно равных частей по записям формата fasta, и для каждой из этих частей была создана база данных DIAMOND [53]. Далее по каждой из этих баз данных был запущен Diamond BLASTp [54] – инструмент, предназначенный для поиска гомологичных белковых последовательностей в больших базах данных. Для оптимальной работы Diamond BLASTp были заданы параметры, подобранные на основании инструкции [55], которые описаны в таблице 2.

Таблица 2 – Параметры для поиска белков, гомологичных известным бактериоцинам, с помощью Diamond BLASTp

Параметр	Значение	Описание
Порог e-value	1e-5	Максимальное ожидаемое количество случайных совпадений. Чем меньше, тем строже фильтрация.
Минимальный процент идентичности	30	Совпадения должны иметь $\geq 30\%$ идентичности аминокислот
Минимальное покрытие запроса	50	Совпадение должно покрывать $\geq 50\%$ длины референсного белка

Затем для потенциальных гомологов бактериоцинов были получены аминокислотные последовательности с помощью NCBI datasets CLI [56].

2.3.3. Проверка новых потенциальных бактериоцинов на наличие белковых мотивов, характерных для бактериоцинов

Белки, отобранные на основе сходства аминокислотных последовательностей с известными бактериоцинами в разделе 2.3.2. «Поиск новых потенциальных бактериоцинов по гомологии с уже известными бактериоцинами», были проверены путем поиска функциональных мотивов, характерных для бактериоцинов.

Самый классический способ поиска мотивов у белков – это использование скрытых Марковских моделей, которые также называются НММ (hidden Markov model) [57]. Принцип этого метода следующий: белки-кандидаты в бактериоцины последовательно сопоставляются с базой данных уже известных доменов бактериоцинов. Если у белка-кандидата найдется бактериоцинный домен, то такой белок выносится в отдельный файл как потенциальный бактериоцин.

Итак, для реализации поиска потенциальных бактериоцинов с помощью НММ [57] был использован биоинформатический инструмент HMMER [58]. Перед использованием HMMER было необходимо создать базу данных известных бактериоцинов: для этого сначала была целиком скачана база данных Pfam [59]. Далее из этой базы были выбраны только мотивы, характерные для бактериоцинов из списка из раздела 1.2.2. «Классификации бактериоцинов».

С выбранными мотивами, характерным для бактериоцинов, проводилось сравнение потенциальных белков-гомологов бактериоцинов с целью поиска у них специфических функциональных мотивов.

Найденные совпадения фильтровались по критерию $e\text{-value} < 1e-5$, показывающему ожидаемое количество случайных совпадений.

Для получения аминокислотной последовательности и другой информации по белкам-гомологам, у которых были обнаружены мотивы бактериоцинов, далее использовался NCBI datasets CLI [56].

2.4. Объединение данных по аннотированным и предсказанным бактериоцинам

Для обработки полученных данных по аннотированным и предсказанным бактериоцинам использовался Python и библиотека для анализа данных pandas [60]. При помощи этой библиотеки, а также библиотеки Biopython из tsv- и fasta-файлов были собраны датафреймы — структуры для работы с табличными данными, и вся последующая работа велась именно с датафреймами.

Первым делом из полученных датафреймов были удалены столбцы с избыточной в контексте этой работы информацией, содержащейся в обрабатываемых файлах. Датафреймы с отобранными колонками, содержащими информацию об аминокислотных последовательностях и описанием предсказанных бактериоцинов были объединены друг с другом посредством объединения по пересечению идентичных ID белков. Далее была создана новая колонка, характеризующая все данные в получившемся датафрейме как предсказанные. В соседнюю с ней колонку была добавлена информация о точности предсказания того, что данный белок может потенциально быть неаннотированным бактериоцином — числовые значения для этой метрики были предсказаны с помощью инструмента HMMER [58].

Те же самые действия были проделаны и с данными по аннотированным бактериоцинам. В отдельную колонку была добавлена информация о том, что все бактериоцины в данном датафрейме являются существующими, а в соседнюю колонку с информацией о точности

предсказания того, что данный белок является бактериоцином, – значение 1, означающее 100% точность.

После этого был произведен поиск дубликатов в обоих датафреймах: дубликатами считались те строки, у которых полностью совпадал ID белка и его аминокислотная последовательность у нескольких строк. В случае обнаружения таких дубликатов в датафрейме с предсказанными бактериоцинами из всех дубликатов оставлялась только одна строка, имеющая наивысшую точность предсказания в соответствующей колонке, остальные дубликаты удалялись из датафрейма. В случае обнаружения таких дубликатов в датафрейме с аннотированными бактериоцинами сохранялась первая такая строка, а остальные дубликаты удалялись.

Далее в датафрейм с предсказанными бактериоцинами необходимо было добавить колонки, содержащие название бактерии-продуцента каждого бактериоцина и таксономический ID этой бактерии. Для этого был получен список ID всех предсказанных бактериоцинов в виде текстового файла, по которому при помощи NCBI [47] и был произведен поиск этой недостающей информации.

Затем как в датафрейм с аннотированными, так и в датафрейм с предсказанными бактериоцинами была добавлена колонка с информацией об исходном источнике каждого белка. Для этого была написана отдельная функция: если ID белка начинается с символов "WP_", "NP_", "XP_", "YP_", источником белка является RefSeq. Если ID белка начинается с символов "A0A", "D3V", "I6W", "O", "P", "Q", "Z9J", а также ID белка состоит из 6 символов, источником белка является UniProtKB/TrEMBL. Если ID белка начинается с символов "P", "Q" и не имеет символа "_" в ID, источником белка является UniProtKB/Swiss-Prot. Для остальных случаев присваивалось значение «неизвестный источник».

После этого датафреймы, содержащие информацию об аннотированных и предсказанных бактериоцинах, были соединены друг с другом в общий датафрейм.

2.5. Фильтрация данных по аннотированным и предсказанным бактериоцинам

Как уже было описано в разделе 2.1. «Первичный анализ встречаемости бактериоцинов в геномах бактерий», помимо собственно бактериоцинов в собранных данных содержатся и вспомогательные белки, ассоциированные с бактериоцинами. К таким вспомогательным белкам относятся иммунные, транспортные, регуляторные белки, протеазы, ферменты модификации и др. белки, участвующие в биосинтезе бактериоцинов, их посттрансляционных модификациях и защите бактерий от собственных бактериоцинов.

Для фильтрации данных по аннотированным и предсказанным бактериоцинам использовалась библиотека `re` [61]. Регулярное выражение, по которому производилась первичная фильтрация, содержало перечисление регистронезависимых названий всех известных бактериоцинов, а также само слово «бактериоцин» через логическое «ИЛИ» с экранированием спецсимволов в названиях. Код, при помощи которого производилась первичная фильтрация бактериоцинов, представлен ниже (листинг 1):

```
import re
```

```
# Список "настоящих" бактериоцинов
```

```
bacteriocins = [
```

```
    "acidocin", "actagardine", "agrocin", "alveicin", "aureocin", "aureocin  
A53", "aureocin A70", "bisin",
```

```
    "carnocin", "carnocyclin", "caseicin", "cerein", "circularin A", "colicin",  
    "curvaticin", "divercin",
```

```
    "duramycin", "enterocin", "enterolysin", "epidermin", "gallidermin",  
    "erwiniocin", "gardimycin",
```

```
    "gassericin A", "glycinecin", "halocin", "haloduracin", "klebicin",  
    "lactocin S", "lactococcin",
```

```

        "lacticin", "leucococcin", "lysostaphin", "macedocin", "mersacidin",
        "mesentericin", "microbisporicin",
        "microcin", "microcin S", "mutacin", "nisin", "paenibacillin",
        "planosporicin", "pediocin", "pentocin",
        "plantaricin", "pneumocyclicin", "pyocin", "reuterin 6", "reutericyclin",
        "reuterin", "sakacin",
        "salivaricin", "sublancin", "subtilin", "sulfolobacin", "tasmancin", "thuricin
        17", "trifolitoxin",
        "variacin", "vibriocin", "warnericin", "warnerin", 'bacteriocin'
    ]

```

Регулярное выражение для поиска бактериоцинов

```

pattern = re.compile(r"\b(" + "|".join([re.escape(b.lower()) for b in
bacteriocins]) + r")\b", re.IGNORECASE)

```

Фильтрация строк по регулярному выражению

```

df_bacteriocins_only =
df3_copy[df3_copy["Protein_Name"].str.contains(pattern, na=False)]

```

Листинг 1 – Код на Python для первичной фильтрации бактериоцинов

Далее производилась итоговая фильтрация: из всех данных удалялись строки, соответствующие регулярному выражению, содержащему ключевые слова из названий вспомогательных белков. Само регулярное выражение было составлено аналогично регулярному выражению для первичной фильтрации. Код, при помощи которого производилась итоговая фильтрация бактериоцинов, представлен ниже (листинг 2):

Обратный паттерн: ищем строки, содержащие хотя бы одно из "вспомогательных" слов

```

fake_keywords = [
    "immunity", "transporter", "secretion", "processing",
    "ATP-binding", "permease", "biosynthesis", "leader", 'protection',
    'transport'
]

```



```

fake_pattern = re.compile(r"(" + "|".join([re.escape(word.lower()) for word in
fake_keywords]) + r")", re.IGNORECASE)

# Оставим только строки, НЕ содержащие слова из списка
fake_keywords
df_without_fake =
df_bacteriocins_only[~df_bacteriocins_only["Protein_Name"] \
.str.contains(fake_pattern, na=False)]

```

Листинг 2 – Код на Python для итоговой фильтрации бактериоцинов

2.6. Сбор данных об аннотированных геномах бактерий-продуцентов бактериоцинов

Для поиска информации о бактериях-продуцентах бактериоцинов из отфильтрованного датафрейма, содержащего информацию об аннотированных и предсказанных бактериоцинах, были взяты уникальные таксономические ID всех бактерий.

Далее через командную строку WSL из RefSeq NCBI [52] и GenBank [62] были получены все сборки бактерий, имеющие полностью аннотированные геномы. Из этих геномов затем были взяты данные только о тех бактериях, таксономические ID которых были отобраны из датафрейма в отдельный файл на предыдущем шаге. Скрипт для получения данных был настроен так, что в случае наличия нескольких сборок для одного таксономического ID бактерии выбиралась одна наилучшая сборка. Наилучшая сборка определялась так: наивысший приоритет отдавался Reference Genome, то есть проверенной сборке, далее вторым по качеству считался Representative Genome. В случаях, когда отсутствовали такие «эталонные» сборки, брались сборки с самыми новыми датами.

2.7. Создание базы данных бактериоцинов

Для создания базы данных бактериоцинов использовалась бесплатная онлайн-платформа SQLiteOnline, доступная по ссылке <https://sqliteonline.com/>.

Создаваемая база данных бактериоцинов будет работать в браузере, то есть локально на персональном компьютере до тех пор, пока пользователь не покинет сайт. Инструкция, по которой была создана база данных, может быть использована любыми другими пользователями для создания этой базы данных локально на своем персональном компьютере.

Итак, для создания базы данных бактериоцинов для начала требуется зайти на онлайн-платформу SQLiteOnline по ссылке <https://sqliteonline.com/>. В открывшейся вкладке необходимо нажать на кнопку «Import», далее «Open» и выбрать скачанный csv-файл, содержащий информацию о бактериоцинах.

После этого необходимо настроить параметры импорта этого файла: для параметра «Column name», отвечающего за название колонок, необходимо выбрать «First line», что означает, что названиями колонок в таблице будут первые строки загружаемого датафрейма. Все остальные параметры нужно оставить без изменений.

Те же самые действия необходимо проделать с датафреймом, содержащим информацию об аннотированных геномах бактерий-продуцентов бактериоцинов.

В основном окне сайта, открывающимся после загрузки всех таблиц создаваемой базы данных из датафреймов, далее можно писать запросы к базе данных бактериоцинов на языке SQL (Structured Query Language) [63]. Примеры таких SQL-запросов описаны в разделе 3.5. «Описание созданной базы данных бактериоцинов»

3. РЕЗУЛЬТАТЫ

3.1. Первичная оценка встречаемости бактериоцинов в геномах бактерий

При помощи скрипта на Python была произведена первичная оценка встречаемости бактериоцинов в геномах бактерий. Получившиеся результаты можно увидеть на рисунках 3 и 4.

Bacteriocin	Amount_of_bacteria	Bacteriocin	Amount_of_bacteria
microcin	5305	klebicin	50
colicin	4948	sublancin	31
pyocin	2523	halocin	26
lysostaphin	1111	agrocin	25
lactococcin	816	pediocin	25
nisin	393	epidermin	24
cerein	391	enterolysin	17
carnocyclin	233	mutacin	12
gallidermin	193	paenibacillin	11
plantaricin	184	mesentericin	10
mersacidin	177	alveicin	5
subtilin	146	macedocin	3
trifolitoxin	141	vibriocin	2
sakacin	119	actagardine	2
enterocin	109	glycinecin	2
circularin A	81	gassericin A	1
lacticin	64	divercin	1
salivaricin	58	bisin	1
aureocin	58	acidocin	1

Рисунок 3 – Первичная оценка встречаемости бактериоцинов в геномах бактерий

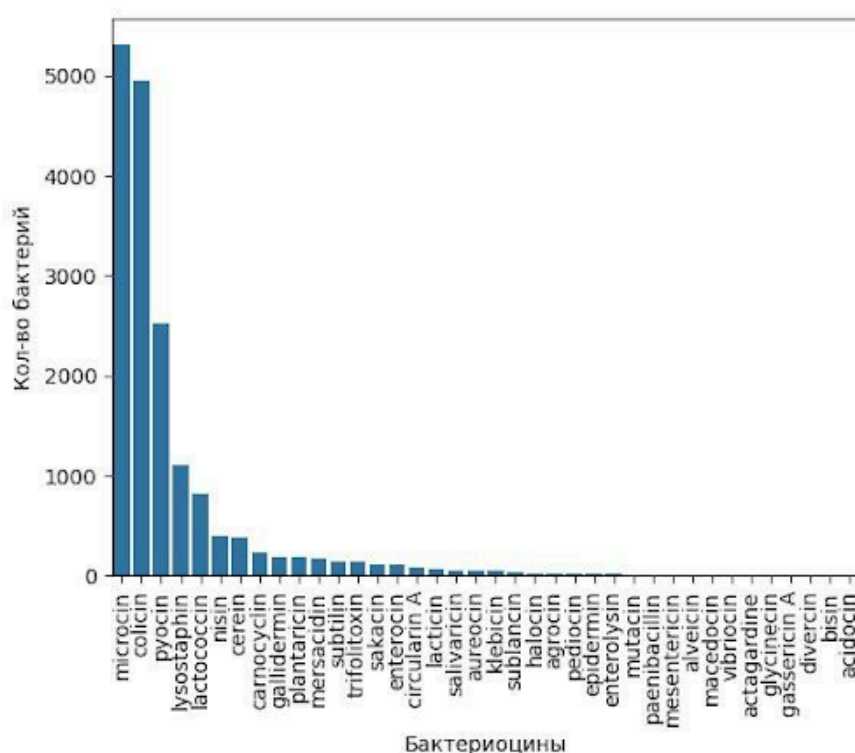


Рисунок 4 – Первичная оценка встречаемости бактериоцинов в геномах бактерий

Как видно из рисунка 4, некоторых бактериоцинов (например, микроцина, колицина) встречается в геномах бактерийкратно больше, чем остальных бактериоцинов. Для исследования причин наблюдаемых различий все названия бактерий в датафрейме из разупорядоченных названий, некоторые из которых содержали штамм, некоторые вид или род, были приведены к единому стилю: все названия стали содержать информацию только о роде исследуемых бактерий. Так, в датафрейме получилось 1392 различных родов бактерий, и самыми часто встречающимися родами в геномных сборках стали *Escherichia* и *Klebsiella*. Другие часто встречающиеся рода бактерий проиллюстрированы на графике, отражающем топ 1% встречаемости родов бактерий по числу сборок (рис. 5):

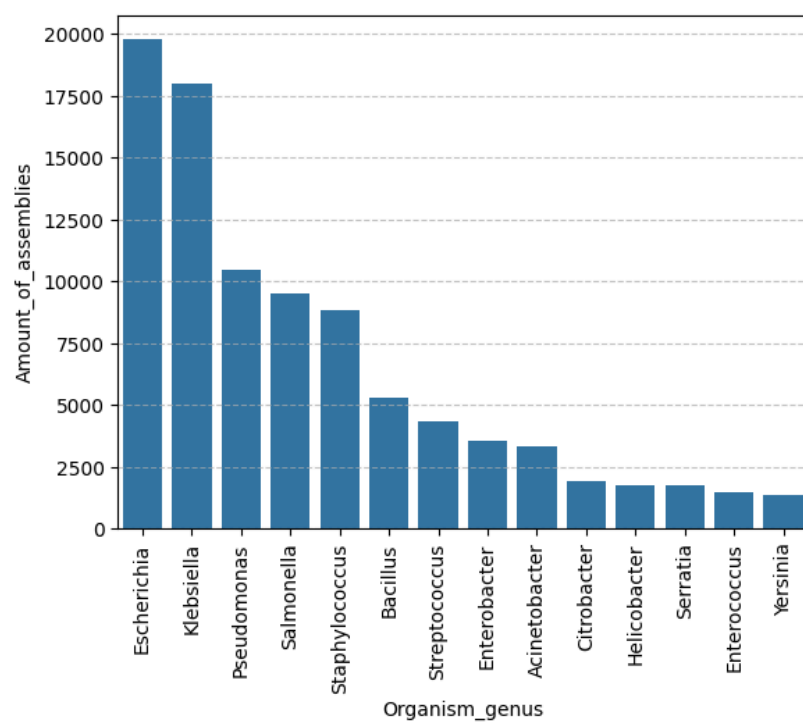


Рисунок 5 – Топ 1% родов бактерий по числу сборок

Далее для топ 1% родов бактерий по числу сборок был построен график встречаемости бактериоцинов в этих родах бактерий. Результаты продемонстрированы на рисунке 6.

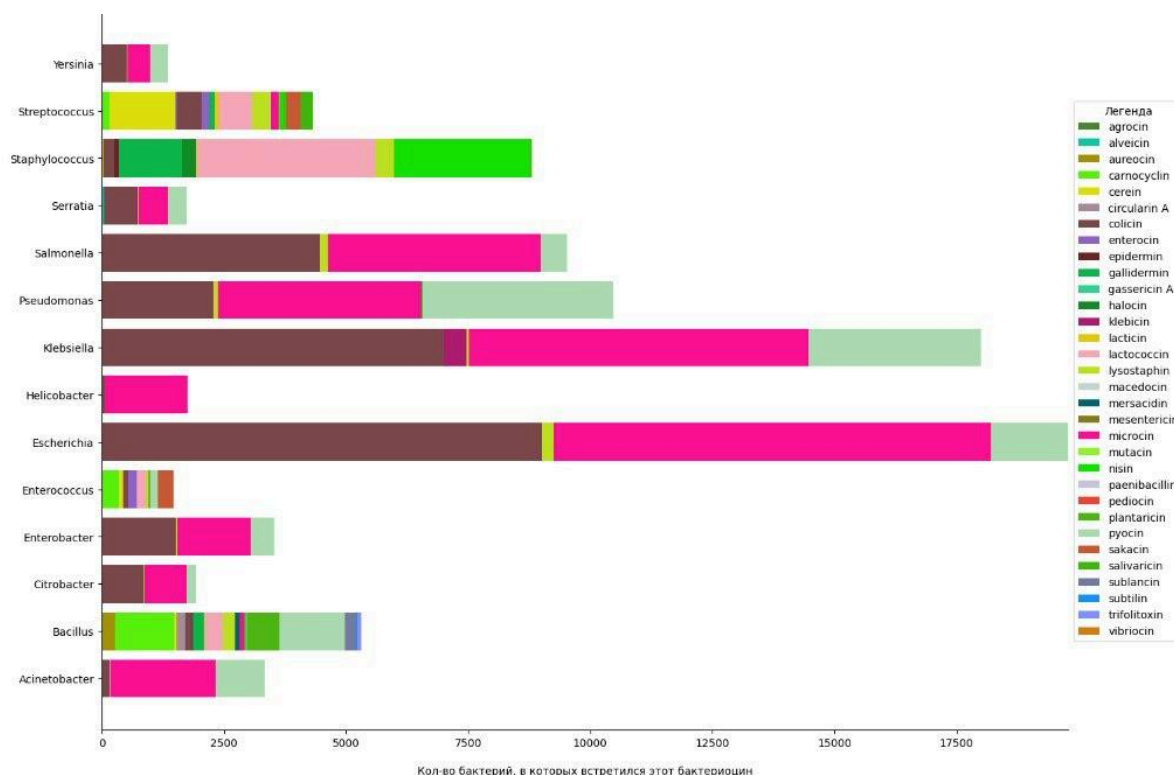


Рисунок 6 – Встречаемость бактериоцинов в топ 1% родов бактерий по количеству сборок

Из рисунка 6 видно, что доминирующими бактериоцинами в родах бактерий с наибольшим числом геномных сборок являются микроцин и колицин, что согласуется с данными рисунка 5.

Таким образом, обилие микроцина, колицина и других бактериоцинов, преобладающих на графике первичной оценки встречаемости бактериоцинов в геномах бактерий (рис. 4) обусловлено различной степенью изученности каждого рода бактерий: чем больше сборок какого-то конкретного рода бактерий, содержащего определенный бактериоцин, тем чаще этот бактериоцин будет встречаться в общей статистике по встречаемости всех бактериоцинов. Второй возможной причиной наблюдаемых различий может являться разная степень изученности бактериоцинов: некоторые из них могут аннотироваться достаточно достоверно, в то время как другие не аннотироваться вовсе. Ввиду этих причин наряду с прямым поиском бактериоцинов по названиям в геномных аннотациях был введен поиск

потенциальных бактериоцинов на основании сходства аминокислотных последовательностей этих белков.

3.2. Характеристика данных, полученных о бактериоцинах при помощи биоинформатических инструментов

Из базы данных UniProt [51] путем применения фильтра по запросу «Bacteriocin» в аннотациях белков было получено 127 629 записей об аминокислотных последовательностях уже известных бактериоцинов.

Что касается поиска новых потенциальных бактериоцинов, то после удаления аминокислотных последовательностей бактериоцинов из fasta-файла с аминокислотными последовательностями всех белков бактерий было получено 155 723 827 белков, которые потенциально могли бы быть бактериоцинами.

После проведения поиска гомологичных белков по отношению к известным аминокислотным последовательностям бактериоцинов при помощи Diamond BLASTp [54] было получено 15 667 525 белков, которые потенциально могут быть бактериоцинами. То есть порядка 10% всех исходных белков бактерий за исключением уже аннотированных бактериоцинов продемонстрировали сходство с бактериоцинами по аминокислотным последовательностям.

В результате применения НММ [56] и дальнейшей фильтрации результатов по e-value было получено 12707 уникальных белков, имеющих совпадения с НММ-профилями бактериоцинов. Это составляет 0.008% от исходного количества всех анализируемых белков и 0.08% от белков, отобранных на основании гомологии с аннотированными бактериоцинами.

3.3 Характеристика объединенных и отфильтрованных данных по аннотированным и предсказанным бактериоцинами

После удаления дубликатов и фильтрации в итоговом датафрейме осталось 8105 различных предсказанных бактериоцинов. На рисунке 7 представлена визуализация встречаемости различных предсказанных бактериоцинов в датафрейме.

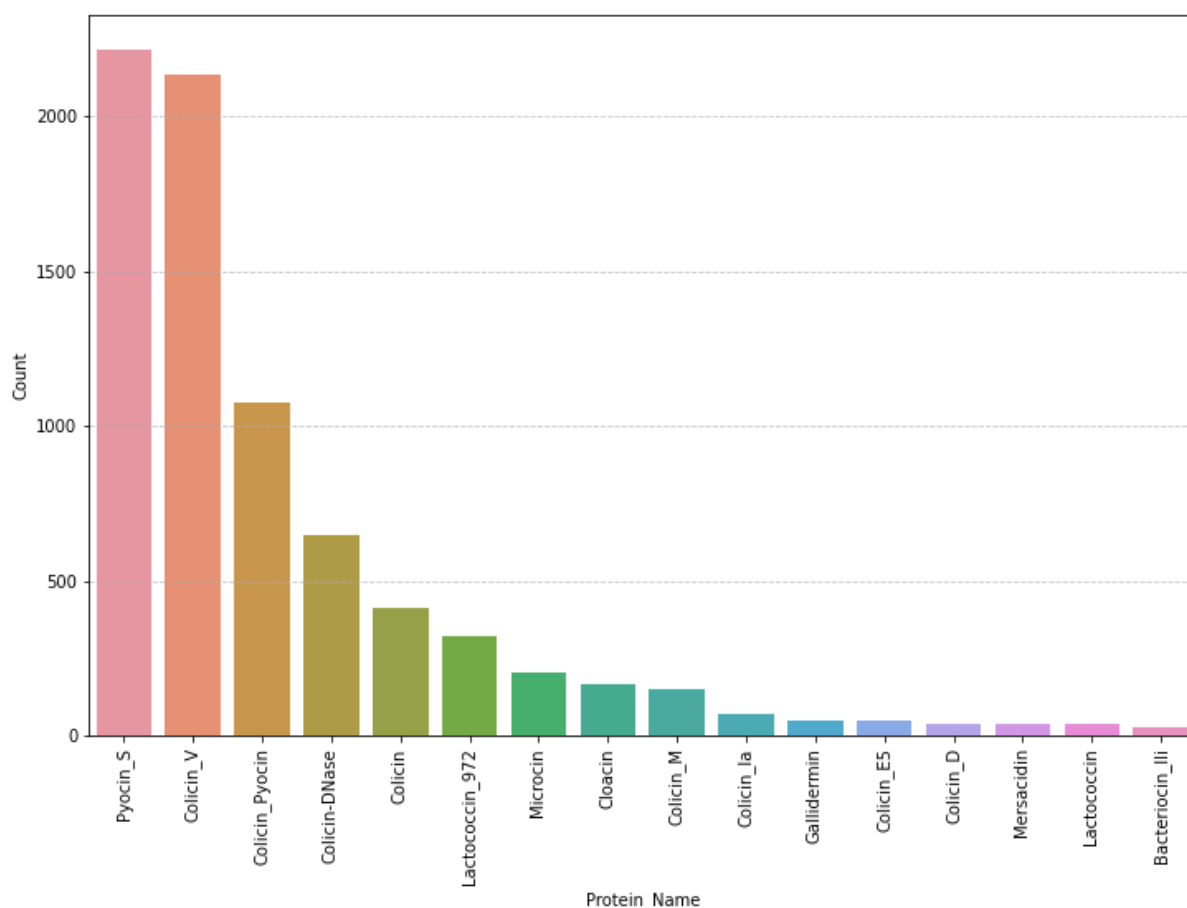


Рисунок 7 – Встречаемость предсказанных бактериоцинов в геномных данных бактерий из базы данных NCBI. Поиск проведен на основе сходства аминокислотных последовательностей потенциальных бактериоцинов с последовательностями известных бактериоцинов из базы данных UniProt.

Ось X - название бактериоцина. Ось Y - количество бактериоцина.

Как видно из рисунка 7, наиболее часто встречающимися предсказанными бактериоцинами в геномах бактерий являются различные виды колицина и пиоцин-S. Вместе они образуют порядка 89% всех предсказанных бактериоцинов.

Что касается аннотированных бактериоцинов, то в итоговом датафрейме после удаления дубликатов и фильтрации по ним осталось 33 195 уникальных записей, что составляет порядка 80% от всех собранных данных

о бактериоцинах. Распределение количества аннотированных и предсказанных бактериоцинов в итоговом датафрейме представлено на рисунке 8:

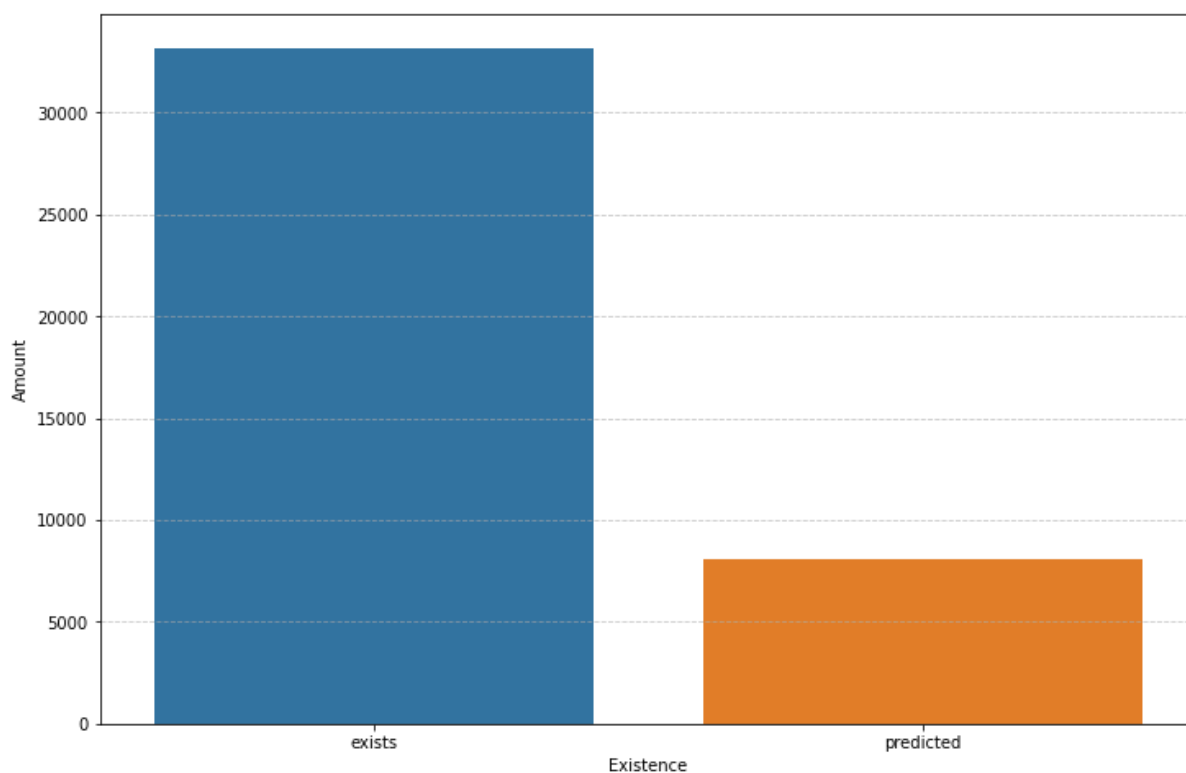


Рисунок 8 – Количество аннотированных (exists) и предсказанных (predicted) бактериоцинов, входящих в датафрейм после фильтрации дубликатов.

3.4. Характеристика данных об аннотированных геномах бактерий-продуцентов бактериоцинов

В результате сбора аннотаций геномов бактерий-продуцентов бактериоцинов была получена информация о 11 175 бактериях из 13 389. Примерно 16.5% пропусков в данных можно объяснить тем, что для этих белков не существует опубликованных геномных сборок.

Итоговый файл с информацией об аннотированных геномах бактерий-продуцентов содержит данные о таксономическом ID бактерии, ID сборки ее генома, категории этой сборки (reference/representative/common), названии организма, а также FTP-ссылку на директорию со всеми файлами

данной сборки, прямую ссылку на геном (FASTA) и прямую ссылку на аннотацию (GFF).

3.5. Описание созданной базы данных бактериоцинов

Из данных об аннотированных и предсказанных бактериоцинах, собранных в датафрейм `real_and_predicted_bacteriocins`, а также данных об аннотированных геномах их бактерий-продуцентов, собранных в датафрейм `bacteria_annotations`, с помощью СУБД SQLite была создана реляционная база данных `bacteriocin_database`.

Данная база данных состоит из двух таблиц `real_and_predicted_bacteriocins` и `bacteria_annotations`, которые имеют связь «один ко многим»: полю `TaxID`, содержащему уникальные значения таксономического ID бактерии-продуцента в таблице `bacteria_annotations`, может соответствовать несколько записей в поле `BacteriaTaxID` таблицы `real_and_predicted_bacteriocins`. Графическое изображение связи между таблицами, а также перечень других полей этих таблиц представлены на рисунке 9:

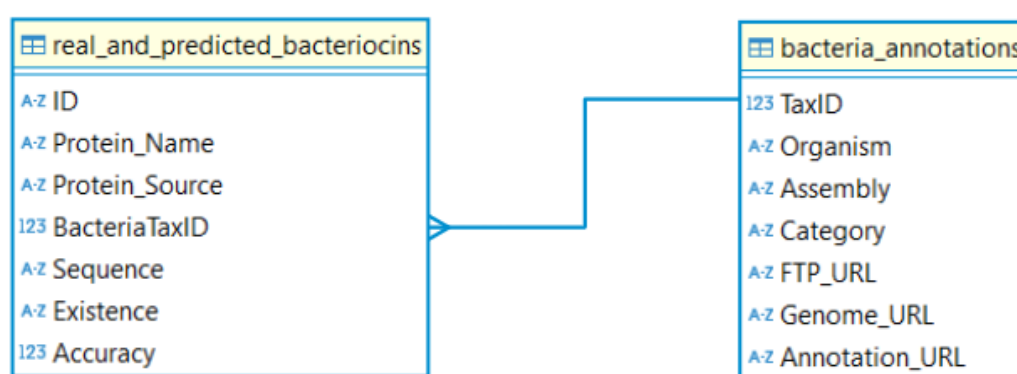


Рисунок 9 – Схема базы данных `bacteriocin_database`

База данных `bacteriocin_database`, а также исходные csv-файлы с датафреймами `real_and_predicted_bacteriocins` и `bacteria_annotations` доступны

по ссылке

https://github.com/Vas111sa/Bacteriocin_database_ITMO_Final_qualifying_work.

С данными датафреймами можно работать в Python при помощи библиотеки pandas, а базу данных развернуть локально на своем персональном компьютере при помощи бесплатной онлайн-платформы SQLiteOnline или ее аналога. Для получения информации из базы данных bacteriocin_database к ней необходимо писать SQL-запросы. Примеры запросов, которые могли бы быть полезны при работе с базой данных bacteriocin_database, описаны в таблице 3:

Таблица 3 – Примеры SQL-запросов к базе данных bacteriocin_database

Цель запроса	SQL-запрос
Получить всю информацию из таблицы real_and_predicted_bacteriocins	SELECT * FROM real_and_predicted_bacteriocins;
Получить всю информацию из таблицы real_and_predicted_bacteriocins, а также название бактерии-продуцента	SELECT r.*, b.Organism FROM real_and_predicted_bacteriocins as r JOIN bacteria_annotations AS b ON r.BacteriaTaxID = b.TaxID;
Получить всю информацию из таблицы real_and_predicted_bacteriocins о существующих бактериоцинах, у которых бактерией-продуцентом является бактерия рода <i>Escherichia</i>	SELECT r.*, b.Organism FROM real_and_predicted_bacteriocins as r JOIN bacteria_annotations AS b ON r.BacteriaTaxID = b.TaxID WHERE r.Existence = 'exists' AND LOWER(b.Organism) LIKE 'escherichia%';

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Созданная база данных *bacteriocin_database*, в отличие от BACTIBASE, BUR, NucleBact и BAGEL, доступна на территории РФ и других стран, а также, в отличие от LABiocin, содержит информацию о бактериоцинах не только из молочнокислых бактерий, а также из других грамположительных и грамотрицательных бактерий. Помимо этого в *bacteriocin_database* содержится информация о белках, потенциально способных проявлять свойства бактериоцинов, с указанием статистической точности такого предсказания, а также обширная информация об аннотированных геномах всех бактерий-продуцентов бактериоцинов, чего нет в других базах данных бактериоцинов.

Однако, в отличие от вышеперечисленных существующих баз данных бактериоцинов, *bacteriocin_database* создается локально на персональном компьютере пользователя, тогда как другие базы данных имеют веб-интерфейс. Помимо этого *bacteriocin_database* не содержит данных о классификации бактериоцинов и их физико-химических свойствах. Все это будет учтено при дальнейшей доработке этой базы данных.

Созданная база данных *bacteriocin_database* может быть использована для разработки новых антибактериальных и пробиотических препаратов, альтернатив химическим пестицидам и новых пищевых консервантов, а также для исследования механизмов взаимодействия между бактериями.

ЗАКЛЮЧЕНИЕ

В ходе данной работы на основе литературных источников был составлен перечень известных на сегодняшний день бактериоцинов, состоящий из 63 уникальных белков. На основании этого списка из всех полностью аннотированных геномов бактерий была получена информация о 33 195 уникальных аннотированных бактериоцинах. Помимо этого по гомологии с уже известными бактериоцинами и наличию белковых мотивов, характерных для бактериоцинов, в геномах бактерий было найдено 8105 уникальных белков, которые потенциально могут быть неаннотированными бактериоцинами.

На основе собранной информации об аннотированных и предсказанных бактериоцинах, а также информации об аннотированных геномах бактерий-продуцентов бактериоцинов была создана база данных бактериоцинов `bacteriocin_database`, а также датафреймы `real_and_predicted_bacteriocins` и `bacteria_annotations` для работы в Python при помощи библиотеки `pandas`. База данных `bacteriocin_database` и датафреймы `real_and_predicted_bacteriocins`, `bacteria_annotations` доступны по ссылке https://github.com/Vas111sa/Bacteriocin_database_ITMO_Final_qualifying_work.

БЛАГОДАРНОСТИ

Выражаю искреннюю признательность своему научному руководителю к.б.н. Райко Михаилу Петровичу, являющемуся доцентом ПИШ ИТМО и руководителем образовательной программы «Прикладная геномика», за участие и отзывчивость в ходе выполнения данной работы, а также его коллеге, преподавателю из ПИШ ИТМО – Кучур Полине Дмитриевне, за неоценимую помощь и наставничество.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Smith A.B., Jones C.D. Antimicrobial Resistance: Review of the Current Situation and Future Perspectives [Электронный ресурс] // PMC. – 2020. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7559574/> (дата обращения: 01.11.2024).
2. Иванова О.П., Петров С.К. Перспективы использования бактериоцинов нормальной микробиоты в антибактериальной терапии (обзор) [Электронный ресурс] // КиберЛенинка. – 2021. – URL: <https://cyberleninka.ru/article/n/perspektivy-ispolzovaniya-bakteriotsinov-normalnoy-mikrobioty-v-antibakterialnoy-terapii-obzor> (дата обращения: 01.11.2024).
3. Brown E.L., Williams D.M. Bacteriocins as Novel Antimicrobial Agents: Current Status and Future Potential [Электронный ресурс] // Antibiotics. – 2024. – Vol. 13, N 7. – P. 666. – URL: <https://www.mdpi.com/2079-6382/13/7/666> (дата обращения: 01.11.2024).
4. Johnson T.R., Miller S.F. Bacteriocins of Gram-Positive Bacteria: Mode of Action and Applications in Medicine [Электронный ресурс] // Microbiology and Molecular Biology Reviews. – 2006. – Vol. 70, N 3. – P. 564–582. – URL: <https://journals.asm.org/doi/10.1128/mmbr.00036-06> (дата обращения: 01.11.2024).
5. Smith J.R., Patel R.K. Antimicrobial Resistance in Emerging Pathogens: A Global Challenge [Электронный ресурс] // Emerging Infectious Diseases. – 2017. – Vol. 23, N 5. – P. 832-839. – URL: https://wwwnc.cdc.gov/eid/article/23/5/16-1556_article (дата обращения: 05.11.2024).
6. Anderson L.M., White C.D. Novel Approaches to Combat Antibiotic Resistance [Электронный ресурс] // PMC. – 2015. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/> (дата обращения: 05.11.2024).

7. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis // The Lancet. – 2022. – Vol. 399, N 10325. – P. 629-655. – DOI: 10.1016/S0140-6736(21)02724-0.
8. Golkar Z., Bagazra O., Pace D.G. Phage Therapy: A Potential Solution to the Antibiotic Resistance Crisis [Электронный ресурс] // Journal of Infection in Developing Countries. – 2014. – Vol. 8, N 2. – P. 129-136. – DOI: 10.3855/jidc.3573.
9. Gross M. Antibiotics in Crisis [Электронный ресурс] // Current Biology. – 2013. – Vol. 23, N 24. – P. R1063-R1065. – DOI: 10.1016/j.cub.2013.11.057.
10. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2013 [Электронный ресурс]. – 2013. – URL: <http://www.cdc.gov/drugresistance/threat-report-2013> (дата обращения: 12.01.2025).
11. Rossolini G.M., Arena F., Pecile P., Pollini S. Update on the antibiotic resistance crisis // Clinical Opinion in Pharmacology. – 2014. – Vol. 18. – P. 56-60. – DOI: 10.1016/j.coph.2014.09.006.
12. Luyt C.E., Brechot N., Trouillet J.L., Chastre J. Antibiotic stewardship in the intensive care unit // Critical Care. – 2014. – Vol. 18, N 5. – P. 480. – DOI: 10.1186/s13054-014-0480-6.
13. Vishwanathan V.K. Misuse of antibiotics by bacteria [Электронный ресурс] // Gut Microbes. – 2014. – Vol. 5, N 1. – P. 3-4. – DOI: 10.4161/gmic.28027.
14. Klein E.Y. et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2018 [Электронный ресурс] // PLOS Global Public Health. – 2023. – URL: <https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0001305> (дата обращения: 12.01.2025).
15. Golkar Z., Bagazra O., Pace D.G. Phage Therapy: A Potential Solution to the Antibiotic Resistance Crisis [Электронный ресурс] // Journal of Infection in Developing Countries. – 2014. – Vol. 8, N 2. – P. 129-136. – DOI: 10.3855/jidc.3573.

16. World Bank. Drug-Resistant Infections: A Threat to Our Economic Future [Электронный ресурс]. – 2017. – URL: <https://www.worldbank.org/en/topic/health/publication/drug-resistant-infections-a-threat-to-our-economic-future> (дата обращения: 21.01.2025).
17. Ventola C.L. The antibiotic resistance crisis [Электронный ресурс] // *Pharmacy and Therapeutics*. – 2015. – Vol. 40, N 4. – P. 277-283. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4422635/> (дата обращения: 21.01.2025).
18. Иванов А.А., Петрова В.М. Современные подходы к преодолению антибиотикорезистентности [Электронный ресурс] // *Вестник СПбГУ*. – 2024. – Т. 11, № 3. – С. 75-87. – URL: https://dspace.spbu.ru/bitstream/11701/48395/1/11-3-2024%2B%D0%B4%D0%B%D1%8F%2B%D1%81%D0%B0%D0%B9%D1%82%D0%B0_75-87.pdf (дата обращения: 21.01.2025).
19. Смирнов В.В. и др. Новые стратегии борьбы с антимикробной резистентностью [Электронный ресурс] // *Acta Naturae*. – 2023. – Т. 15, № 2. – С. 34-45. – URL: <https://actanaturae.ru/2075-8251/article/view/10955> (дата обращения: 21.01.2025).
20. Sharma A. et al. Antimicrobial adjuvants: An innovative strategy for handling antimicrobial resistance displayed by microbes [Электронный ресурс] // *Journal of Bacteriology & Mycology*. – 2020. – URL: <https://medcraveonline.com/JBMOA/antimicrobial-adjuvants--an-innovative-strategy-for-handling-antimicrobial-resistance-displayed-by-microbes.html> (дата обращения: 21.01.2025).
21. Zhang L. et al. Nanomaterials for combating antibiotic resistance [Электронный ресурс] // *Nanoscale Advances*. – 2023. – Т. 5. – С. 123-135. – URL: <https://pubs.rsc.org/en/content/articlelanding/2023/na/d3na00530e> (дата обращения: 21.01.2025).
22. Johnson M.K. Novel approaches to overcome multidrug resistance in bacteria [Электронный ресурс] // *Preprints*. – 2024. – URL:

- <https://www.preprints.org/manuscript/202411.0105/v1> (дата обращения: 21.01.2025).
23. Brown E.L. et al. Emerging strategies to combat antimicrobial resistance [Электронный ресурс] // Nature Reviews Microbiology. - 2024. - URL: <https://www.nature.com/articles/s41579-024-01045-x> (дата обращения: 21.01.2025).
24. Davies J., Davies D. Origins and Evolution of Antibiotic Resistance // Microbiology and Molecular Biology Reviews. – 2010. – Vol. 74, N 3. – P. 417-433. – DOI: [10.1128/MMBR.00016-10](<https://doi.org/10.1128/MMBR.00016-10>).
25. Соколова М.И., Петров К.Л. Современные методы преодоления антибиотикорезистентности [Электронный ресурс] // Вестник биохимии. – 2023. – URL: <https://vuzbiochemi.elpub.ru/jour/article/view/810> (дата обращения: 21.01.2025).
26. Oliveira A.S. et al. Antimicrobial Peptides as New Therapeutic Alternatives // Brazilian Journal of Microbiology. – 2021. – Vol. 52, N 3. – P. 1011-1024. – DOI: [10.1007/s42770-021-00464-9](<https://doi.org/10.1007/s42770-021-00464-9>).
27. Кузнецов А.А. и др. Биосинтез антибиотиков и механизмы резистентности [Электронный ресурс] // Вестник ВГУ. Серия: Химия. Биология. – 2020. – № 4. – С. 17-25. – URL: <http://www.vestnik.vsu.ru/pdf/chembio/2020/04/2020-04-17.pdf> (дата обращения: 21.01.2025).
28. Сидорова Е.В. Применение бактериофагов в ветеринарии [Электронный ресурс] // Российский ветеринарный журнал. – 2022. – № 3. – С. 45-52. – URL: <https://veterinary.arriah.ru/jour/article/view/847> (дата обращения: 15.02.2025).

29. Johnson R.B. et al. Alternatives to Antibiotics in Animal Husbandry // Animals. – 2021. – Vol. 11, N 4. – P. 979. – DOI: [10.3390/ani11040979](https://doi.org/10.3390/ani11040979).
30. Müller E. et al. Novel Approaches for Rapid Detection of Antibiotic Resistance // Journal of Microbiological Methods. – 2018. – Vol. 144. – P. 120-128. – DOI: [10.1016/j.mimet.2017.11.015](https://doi.org/10.1016/j.mimet.2017.11.015).
31. Chen X. et al. Phage Therapy: A Renewed Approach Against Antibiotic-Resistant Bacteria // Frontiers in Microbiology. – 2020. – Vol. 11. – P. 575981. – DOI: [10.3389/fmicb.2020.575981](https://doi.org/10.3389/fmicb.2020.575981).
32. [Электронный ресурс] // Википедия. – URL: <https://ru.wikipedia.org/wiki/Бактериоцины> (дата обращения: 18.02.2025).
33. Smith A.B. et al. Antimicrobial Peptides: Mechanisms and Clinical Applications [Электронный ресурс] // PMC. – 2020. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7559574/> (дата обращения: 18.02.2025).
34. Brown E.L. et al. Bacteriocins as Novel Antimicrobial Agents [Электронный ресурс] // Antibiotics. – 2024. – Vol. 13, № 7. – P. 666. – DOI: 10.3390/antibiotics13070666.
36. Суздальцев А.Е. Микробиология и антимикробные препараты [Электронный ресурс]. – 2021. – URL: https://pharmprobeg.ru/download/Suzdal_2021-05/Mikrobiologiya..pdf (дата обращения: 18.02.2025).
37. Абрамова Е.С. Бактериоцины: перспективы использования в России и за рубежом // Вестник МГУ. Серия Биология. – 2022. – № 3. – С. 45-52.
38. Garcia-Gutierrez E. et al. Bacteriocins: Current Trends and Potential Applications [Электронный ресурс] // Antibiotics. – 2020. – Vol. 9, № 9. – P. 553. – DOI: 10.3390/antibiotics9090553.

39. Петрова И.К. Потенциальное значение бактериоцинов при этиологическом лечении инфекционных заболеваний респираторного тракта [Электронный ресурс] // Медицинская наука и образование. – 2021. – URL: <https://cyberleninka.ru/article/n/potentsialnoe-znachenie-bakteriotsinov-pri-etiologicalheskom-lechenii-infektsionnyh-zabolevaniy-respiratornogo-trakta> (дата обращения: 02.03.2025).
40. Дейт К. Введение в системы баз данных. – М.: Вильямс, 2020. – 1328 с. – ISBN 978-5-8459-2088-3.
41. Smith J.R. et al. Molecular characterization of bacterial strains // BMC Microbiology. – 2010. – Vol. 10. – Art. 22. – DOI: [10.1186/1471-2180-10-22](<https://doi.org/10.1186/1471-2180-10-22>).
42. Anderson L.M. et al. Comparative genomics of pathogenic bacteria // BMC Microbiology. – 2007. – Vol. 7. – Art. 89. – DOI: [10.1186/1471-2180-7-89](<https://doi.org/10.1186/1471-2180-7-89>).
43. Brown E.L. et al. Advances in microbial genomics // Frontiers in Microbiology. – 2015. – Vol. 6. – P. 441. – DOI: [10.3389/fmicb.2015.00441](<https://doi.org/10.3389/fmicb.2015.00441>).
44. Jolley K.A. et al. Nucleotide sequence databases for bacterial typing [Электронный ресурс]. – 2022. – URL: <https://pubmlst.org/projects/nucleobact> (дата обращения: 15.04.2025).
45. Taylor D.E. et al. Bioinformatics tools for DNA analysis // Nucleic Acids Research. – 2006. – Vol. 34 (Suppl. 2). – P. W273-W279. – DOI: [10.1093/nar/gkl244](<https://doi.org/10.1093/nar/gkl244>).
46. Martinez J.L. et al. Computational approaches to antimicrobial resistance // International Journal of Antimicrobial Agents. – 2019. – Vol. 54, № 4. – P. 401-412. – DOI: [10.1016/j.ijantimicag.2019.06.013](<https://doi.org/10.1016/j.ijantimicag.2019.06.013>).

47. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information // Nucleic Acids Research. – 2018. – Vol. 46, N 1. – С. D8-D13. – URL: <https://www.ncbi.nlm.nih.gov/> (дата обращения: 15.04.2025).
48. Python Software Foundation. Python Language Reference, version 3.11. – 2023. – URL: <https://docs.python.org/3/> (дата обращения: 21.04.2025).
49. Cock P.J.A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics // Bioinformatics. – 2009. – Vol. 25, N 11. – С. 1422-1423. – URL: <https://biopython.org> (дата обращения: 21.04.2025).
50. Schoch C.L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools // Database. – 2020. – Vol. 2020. – URL: <https://www.ncbi.nlm.nih.gov/taxonomy> (дата обращения: 28.04.2025).
51. UniProt Consortium. UniProt: the universal protein knowledgebase in 2023 // Nucleic Acids Research. – 2023. – Vol. 51, N 1. – С. D523-D531. – URL: <https://www.uniprot.org/> (дата обращения: 28.04.2025).
52. O'Leary N.A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation // Nucleic Acids Research. – 2016. – Vol. 44, N 1. – С. D733-D745. – URL: <https://www.ncbi.nlm.nih.gov/refseq/> (дата обращения: 28.04.2025).
53. Buchfink B. et al. Fast and sensitive protein alignment with DIAMOND // Nature Methods. – 2015. – Vol. 12, N 1. – С. 59-60. – URL: <https://github.com/bbuchfink/diamond> (дата обращения: 28.04.2025).
54. DIAMOND blastp command documentation [Электронный ресурс]. – 2023. – URL: <https://github.com/bbuchfink/diamond/wiki> (дата обращения: 29.04.2025).
55. McArthur A.G. et al. The comprehensive antibiotic resistance database // Antimicrobial Agents and Chemotherapy. – 2013. – Vol. 57, N 7. – С. 3348-3357. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3820096/> (дата обращения: 29.04.2025).
56. NCBI Datasets Command Line Tools [Электронный ресурс]. – URL: <https://www.ncbi.nlm.nih.gov/datasets/docs/v2/reference-docs/command-line/datasets/> (дата обращения: 29.04.2025).

57. Eddy S.R. Hidden Markov models // Current Opinion in Structural Biology. – 1996. – Vol. 6, N 3. – С. 361-365. – URL: <https://www.sciencedirect.com/science/article/abs/pii/S09594440X96800309> (дата обращения: 29.04.2025).
58. HMMER: biosequence analysis using profile hidden Markov models [Электронный ресурс]. – URL: <http://hmmer.org/> (дата обращения: 30.04.2025).
59. Pfam: the protein families database [Электронный ресурс]. – URL: <https://pfam.xfam.org/> (дата обращения: 30.04.2025).
60. pandas: Python Data Analysis Library [Электронный ресурс]. – URL: <https://pandas.pydata.org/docs/> (дата обращения: 30.04.2025).
61. Python re module documentation [Электронный ресурс]. – URL: <https://docs.python.org/3/library/re.html> (дата обращения: 30.04.2025).
62. Benson D.A. et al. GenBank // Nucleic Acids Research. – 2023. – Vol. 51, N 1. – С. D36-D42. – URL: <https://www.ncbi.nlm.nih.gov/genbank/> (дата обращения: 30.04.2025).
63. SQL documentation [Электронный ресурс]. – URL: <https://www.iso.org/standard/76583.html> (дата обращения: 30.04.2025).