Муниципальное бюджетное общеобразовательное учреждение города Владимира «Средняя общеобразовательная школа №31 имени Героя Советского Союза Сергея Дмитриевича Василисина»

Номинация: «Генетика»

**ОПРЕДЕЛЕНИЕ PAM-ПОСЛЕДОВАТЕЛЬНОСТЕЙ БЕЛКОВ CAS-9 РАЗЛИЧНЫХ ШТАММОВ БАКТЕРИЙ БИОИНФОРМАТИЧЕСКИМИ МЕТОДАМИ**

**Автор работы:**Юдин Владислав Ильич учащийся 11А класса МБОУ «СОШ №31»

**Научный руководитель:**Суслина Светлана Анатольевна учитель химии и биологии  
МБОУ «СОШ №31»

**Научный консультант:**  
Запруднова Елена Александровна  
Кандидат биологических наук  
Педагог дополнительного образования ДТ «Кванториум33»

Владимир – 2022

**СОДЕРЖАНИЕ**

ВВЕДЕНИЕ1

ЛИТЕРАТУРНЫЙ ОБЗОР2

1. CRISPR Cas2
2. Биоинформатические методы4
   1. Выравнивание4
   2. Филогенетический анализ5
   3. Получение последовательностей геномов6
      1. NGS-секвенирование6
      2. Сборка геномов8

МАТЕРИАЛЫ И МЕТОДЫ10

РЕЗУЛЬТАТЫ11

ВЫВОДЫ19

ЗАКЛЮЧЕНИЕ19

СПИСОК ЛИТЕРАТУРЫ20

**ВВЕДЕНИЕ**

Технологии геномного редактирования в настоящее время являются одними из самых стремительно развивающихся направлений современной науки. Системы редактирования генома активно применяются для модификации и улучшения ценных признаков сельскохозяйственных животных и растений, для синтеза необходимых биологически активных веществ. Активно разрабатываются методы генной терапии с помощью CRISPR Cas-9, позволяющие избавить человечество от многих наследственных заболеваний. Главным инструментом в арсенале генетической инженерии являются, так называемые, «молекулярные ножницы» - CRISPR Cas-9.

Как и любая технология, CRISPR Cas-9 имеет свои ограничения. Cas-эффектор способен разрезать целевую последовательность только при выполнении двух условий: комплементарном связывании sgРНК с таргетируемой последовательностью и обнаружении определенного сочетания нуклеотидов, 3’-фланкирующих участок связывания с sgРНК. Данный мотив называется PAM-последовательностью и является строго специфичным для каждого Cas-9. Соответственно, возможности редактирования будут ограничены теми местами генома, где встречается данное сочетание нуклеотидов.

В настоящее время активно ведется поиск новых белков-эффекторов, с помощью биоинформатических инструментов обнаружено более 3,5 тысяч Cas-9 белков разных видов бактерий, из них в лабораторной практике применяется менее 10. Чтобы использование белка в экспериментах стало возможным, необходимо подобрать оптимальные условия работы, установить структуру направляющей РНК и определить PAM-последовательность. Так как PAM-последовательность связывается с определенным доменом белка Cas-9, знание места дислокации и структуры этого

домена позволило бы предсказывать PAM-последовательность для любого белка Cas-9 биоинформатически.

**Цель исследования:** создать библиотеку PAM-последовательностей белков Cas-9 различных штаммов бактерий.

**Задачи исследования:**

1. Построить филогенетическое дерево по аминокислотным последовательностям белков Cas-9 различных родов бактерий и отследить закономерности их эволюции.
2. Собрать геномы бактерий с неизвестными PAM-последовательностями.
3. Определить PAM-последовательности для белков Cas-9 бактерий из различных родов.

**ЛИТЕРАТУРНЫЙ ОБЗОР**

**1. CRISPR Cas**

CRISPR Cas — это аналог противовирусного иммунитета у бактерий и архей. Когда вирусная ДНК или РНК попадает в клетку, бактериальные белки Cas-1 и Cas-2 образуют интегразный комплекс, который вырезает фрагмент вирусной ДНК длинною от 30 до 100 нуклеотидов и интегрирует его в специальный участок бактериального генома - CRISPR-кассету, фрагменты вирусной ДНК называются спейсерами (1-2 на рис.1). CRISPR-кассета состоит из множества спейсеров, разделенных повторяющимися последовательностями, комплементарными tracrРНК. В процессе экспрессии такой CRISPR-кассеты, получается длинная РНК, связанная с tracrРНК, которая называется pre-crРНК. В процессе созревания pre-crРНК разрезается РНКазойIII на множество малых sgРНК (3-4 на рис.1), состоящих из последовательности, комплементарной фрагменту вирусного генома (спейсеру) и двуцепочечного участка – скаффолда.

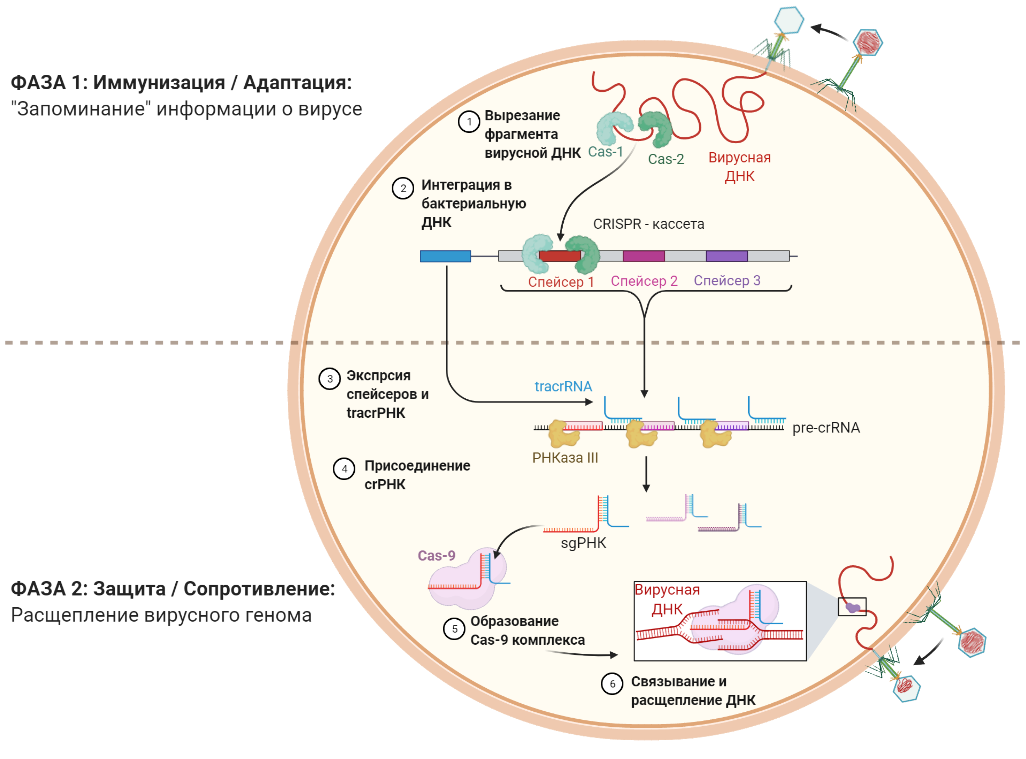


Рис.1. Схема работыCRISPR Cas системы в бактериальной клетке.

Затем sgРНК соединяется с белком Cas-9, образуя Cas-9 комплекс. Cas-9 комплекс связывается с находящейся в клетке ДНК и движется по ней, пока не наткнется область, комплементарную спейсеру (протоспейсер) и PAM-последовательность, которая расположена с 3’-конца протоспейсера в вирусном геноме. Для каждого Cas-9 белка PAM-последовательность строго специфична, обычно ее длинна не превышает 10 нуклеотидов, у самого известного белка Cas-9 из бактерий Streptococcus pyogenes PAM-последовательность - NGG, где N это любой нуклеотид. Cas-9 комплекс связывается с PAM-последовательностью и совершает двуцепочечный разрыв, приводя вирусную ДНК в негодность (рис.2).

На основе бактериальной CRISPR Cas системы была создана система внесения направленных изменений в ДНК любых организмов in vitro - CRISPR Cas-9. При использовании этой технологии конструируется sgРНК, комплементарная таргетируемому участку целевой ДНК, затем полученную sgРНК смешивают с белками Cas-9, находящимися в соответствующем буфере. После образования Cas-9 комплексов в реакционную смесь добавляют целевую ДНК, инкубируют, в результате чего происходит разрезание целевой ДНК в необходимом месте, затем в реакционную смесь вносится комбинативный шаблон, содержащий вставляемую последовательность, фланкированную фрагментами таргетированной ДНК. Системы репарации восстанавливают целостность ДНК, используя комбинативный шаблон в качестве матрицы, в результате получается целевая ДНК со вставкой (рис.3). Технология CRISPR Cas-9 позволяет вносить разрезы лишь в те места генома, где есть PAM-последовательность, поэтому чем большее количество Cas-9 белков будут исследованы, тем большее количество генно-инженерных манипуляций мы сможем совершать.

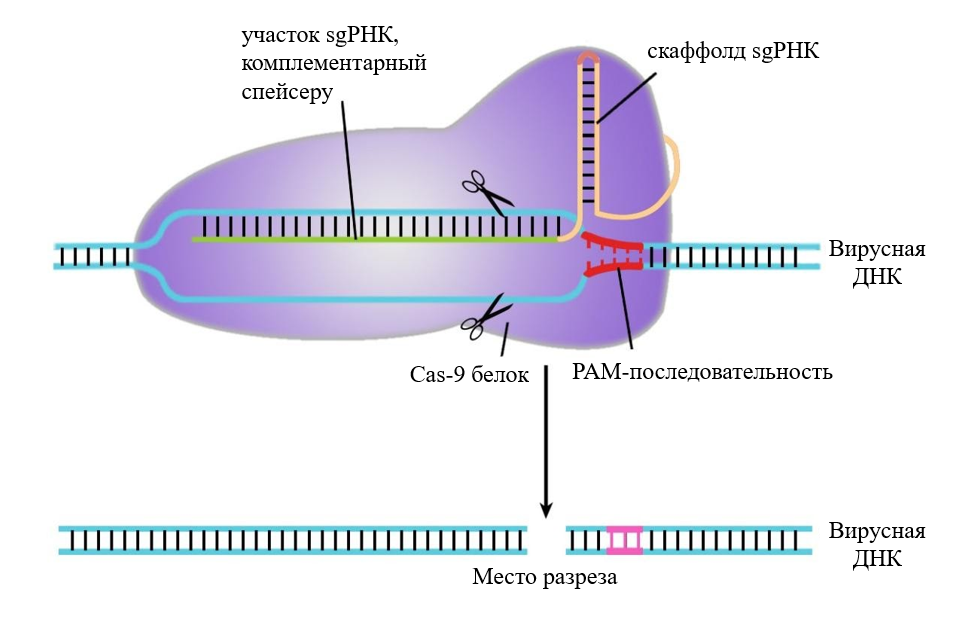


Рис.2. Схема работыCas-9 комплекса.

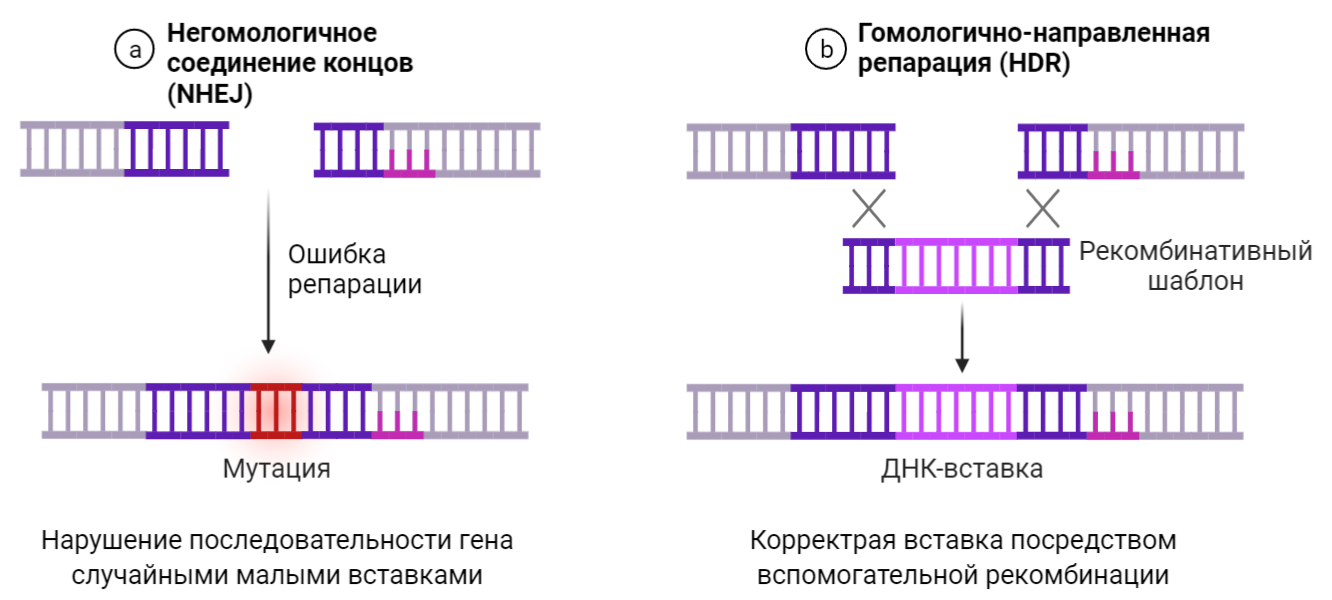


Рис.3. Схема комбинативной трансформации

**2. Биоинформатические методы**

**2.1. Выравнивание**

Выравнивание – это метод вычислительной биологии, позволяющий сопоставить несколько нуклеотидных или аминокислотных последовательностей между собой и найти участки сходства и различия (рис.4).

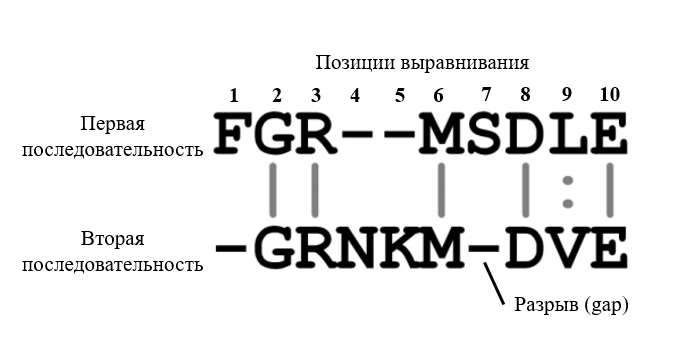


Рис.4. Выравнивание двух фрагментов аминокислотных последовательностей.

В основе алгоритмов выравнивания лежит точечная матрица (dot plot). «Точечная матрица представляет собой таблицу или матрицу, в которой строки соответствуют элементам одной последовательности, а колонки – элементам другой последовательности. В простейшем варианте ячейки точечной матрицы оставляют пустыми, если сравниваемые элементы различны, и заполняются, если они совпадают. Совпадающие фрагменты последовательностей отображаются в виде диагоналей, идущих из верхнего левого угла в нижний правый» (рис.5) [14].



Рис.5. Точечная матрица сходства 2 строк и итоговое выравнивание.

**2.2. Филогенетический анализ**

Филогенетический анализ – это метод вычислительной биологии, который позволяет выявить эволюционные взаимосвязи между нуклеотидными или аминокислотными последовательностями и восстановить эволюционную историю в виде разветвленной диаграммы (филогенетического дерева)(рис.6).

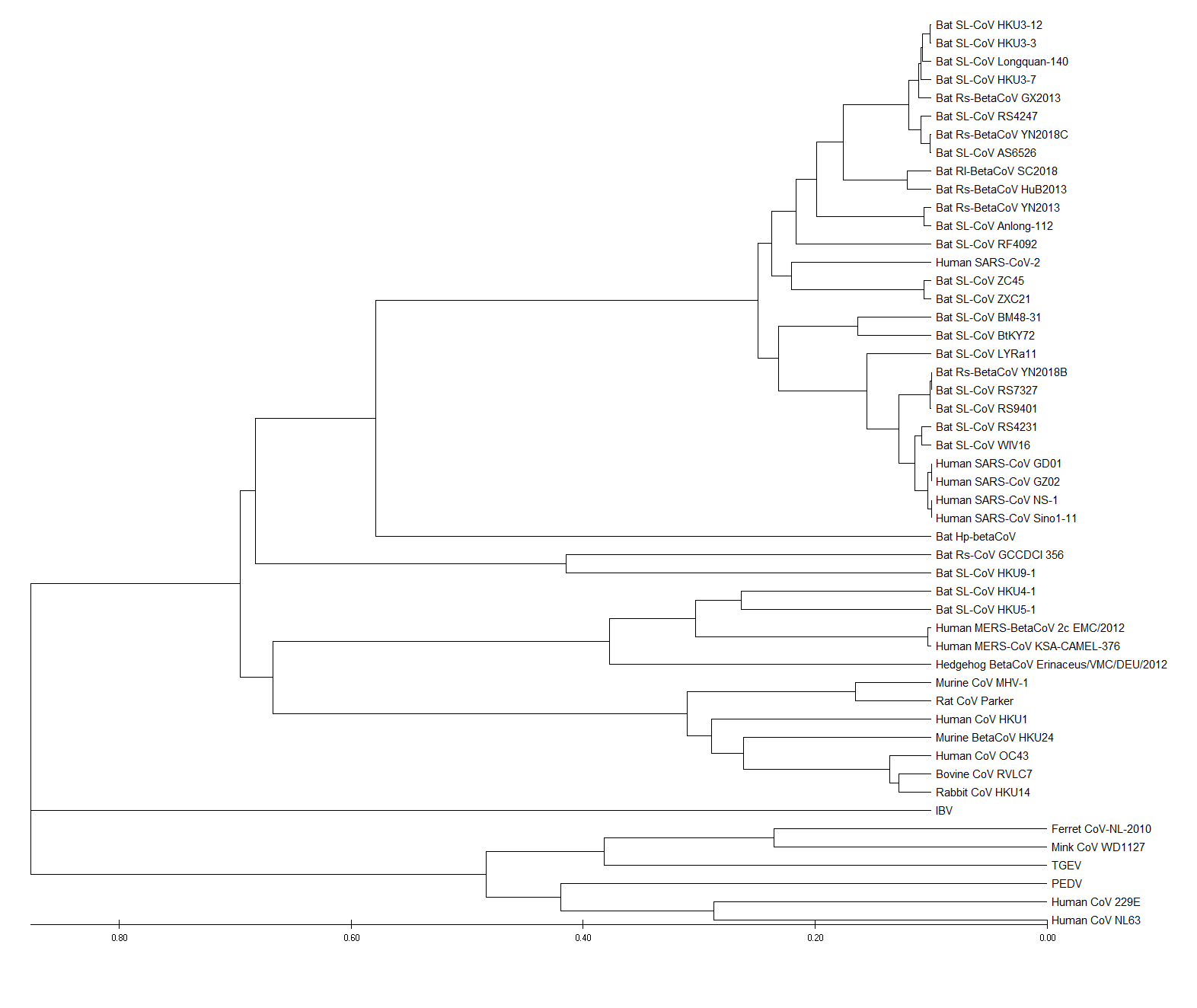


Рис.6. Филогенетическое дерево SARS-CoV-2.

Одним из наиболее популярных методов построения филогенетических деревьев является метод наибольшей парсимонии (метод минимальной эволюции). Он применяется для построения деревьев на основе минимального числа мутаций, необходимых для преобразования одной последовательности в другую (рис.7) [14].

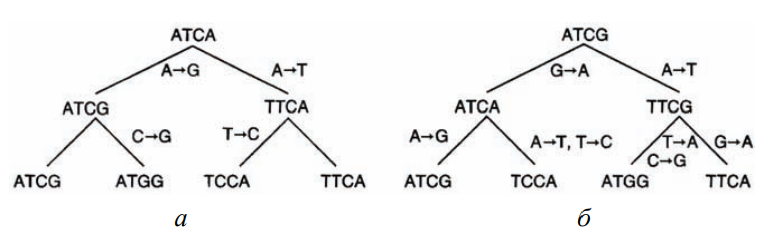


Рис.7. Два генеалогических дерева мутаций.

Филогенетика позволяет отследить не только видовую эволюцию, но и эволюцию отдельных генов и белков, которая далеко не всегда соответствует видовой из-за разной степени конформности последовательностей.

**2.3. Получение последовательностей геномов**

**2.3.1. NGS-секвенирование**

Секвенирование нового поколения (NGS) – это набор методов прочтения нуклеотидных последовательностей ДНК и РНК, основанных на параллельном прочтении сразу нескольких участков генома с помощью повторяющихся циклов комплиментарного удлинения цепи полимеразой или лигирования олигонуклеотидов, что позволяет получать последовательности целых геномов. Наиболее распространенным методом секвенирования в данный момент является illumina (рис.8).

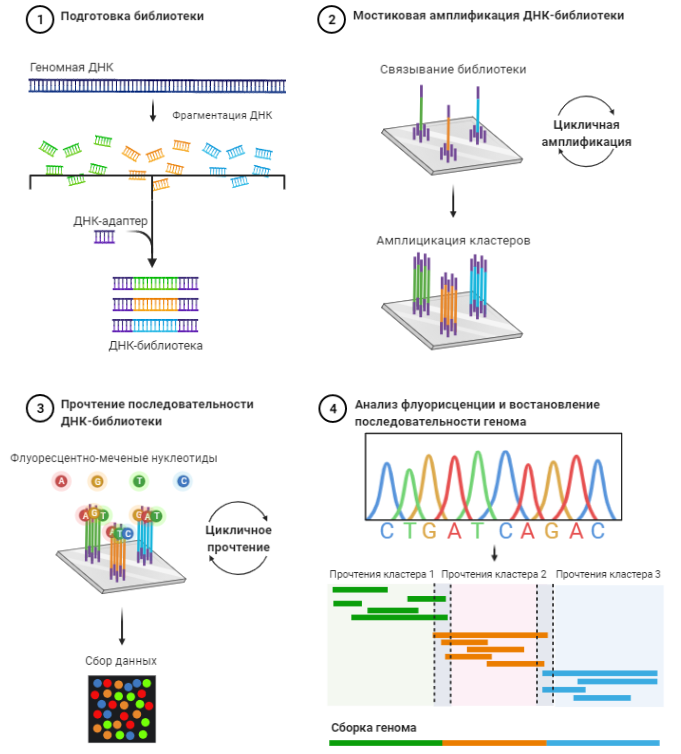


Рис.8.Схема метода секвенирования illumina.

На первом этапе производится фрагментация ДНК, последовательность которой необходимо установить. Наиболее популярным способом фрагментации является ультразвуковое воздействие, также фрагментацию осуществляют с помощью эндонуклеаз рестрикции. Далее к образовавшимся фрагментам целевой ДНК лигируют 3’ и 5’ ДНК-адаптеры, последовательности которых, в отличие от целевой ДНК, нам заранее известны, получается библиотека k-меров (1 на рис.8).

На втором этапе целевая ДНК с адаптерами попадает в ячейку, поверхность которой усеяна прикреплёнными олигонуклеотидами, комплиментарными 3’ и 5’ адаптерам. k-меры связываются с олигонуклеотидами ячейки и происходит амплификация, в процессе которой достраиваются вторые цепи ДНК. На каждом цикле происходит денатурация, новообразованные двуцепочечные молекулы разделяются и связываются с близлежащими олигонуклеотидами, за несколько циклов нарабатываются крупные кластеры, состоящие из одинаковых одноцепочечных фрагментов целевой ДНК с адаптерами (2 на рис.8).

На третьем этапе происходит прочтение последовательностей фрагментов ДНК. Вместо стандартных нуклеотидмонофосфатов в реакционную смесь поступают флуоресцентно-меченные обратимо терминированные нуклеотиды (dNTP), синтез на которых обрывается, а при облучении ультрафиолетом они издают свет определенной длинны волны, что позволяет отличить нуклеотиды одного типа от нуклеотидов других типов (рис.9).

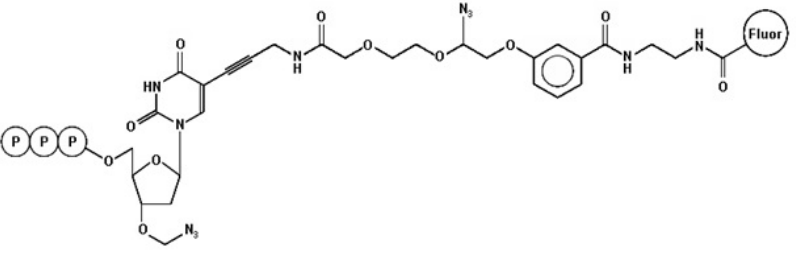
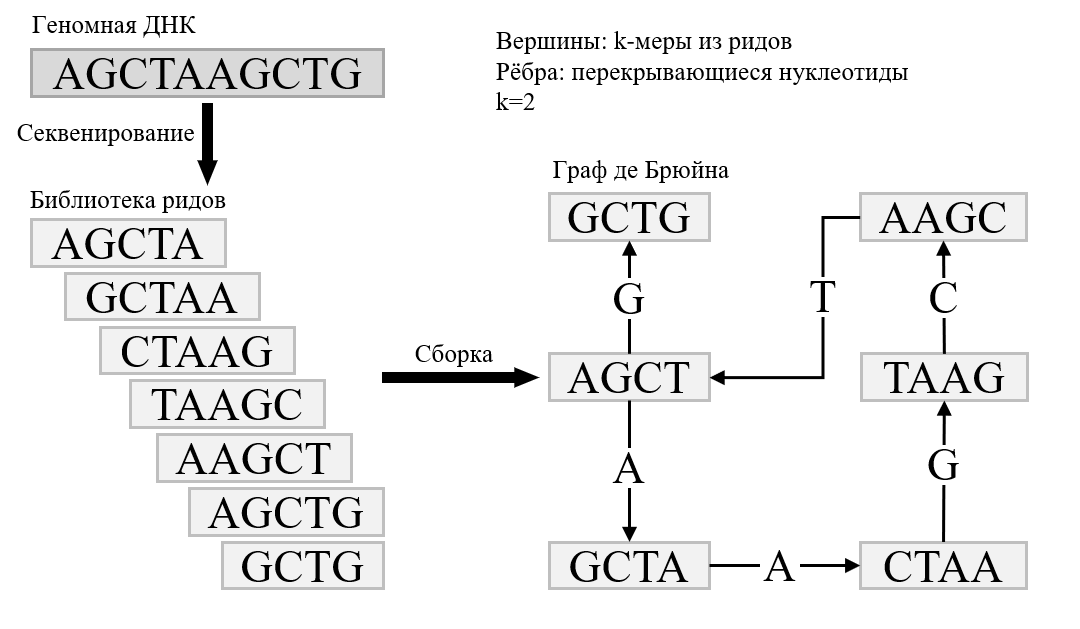


Рис.9 флуоресцентно-меченный обратимо терминированный тимин.

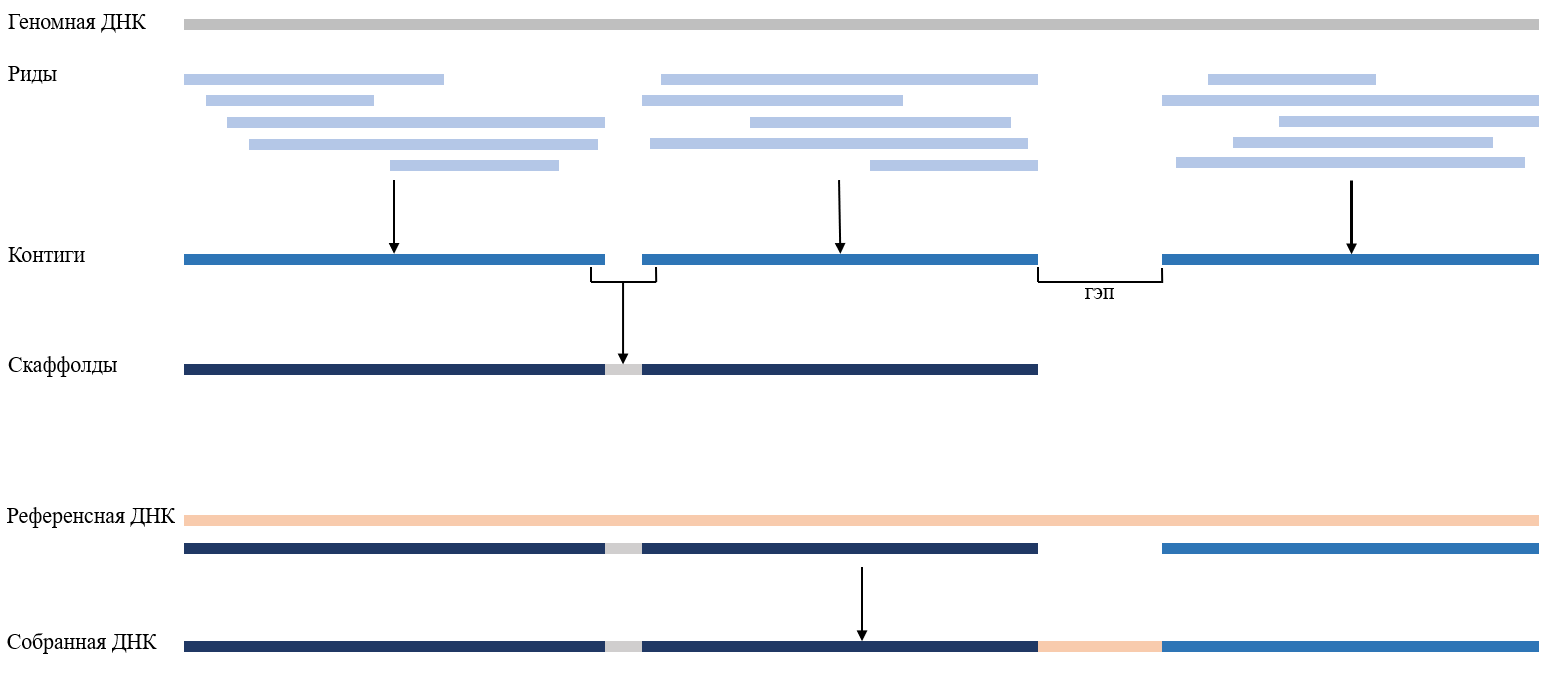
На каждом цикле происходит присоединение меченых нуклеотидов, регистрация свечения ПЗС-матрицей, затем через ячейку пропускают реагенты, отщепляющие флуорофор и терминатор, после чего снова становится возможным дальнейшее удлинение цепи полимеразой (3 рис.8). Такой цикл удлинения цепи и регистрации свечения повторяется ~100-200 раз, в результате чего получается установить последовательности всех кластеров в ячейке.

**2.3.2. Сборка геномов**

Выходными данным NGS-секвенирования служат библиотеки прочтений (ридов), каждое прочтение – это последовательность одного кластера с этапа секвенирования. Из библиотеки ридов необходимо собрать полную последовательность изучаемой ДНК. Для выполнения этой задачи существуют специальные программы – сборщики (ассемблеры). В основе алгоритмов сборщиков лежит разделение ридов на k-меры (последовательности из k символов) и построение графов де Брюйна (рис.10) [15].

Рис.10.Схема построения графа де Брюйна.

Далее ассемблер проводит последовательный анализ вероятностей разных сборок и определяет наиболее вероятный. В процессе сборки получаются крупные фрагменты целевой последовательности порядка нескольких сотен тысяч нуклеотидов (для бактериальных геномов), которые называют котингами, затем из контигов восстанавливают скаффолды - фрагменты порядка нескольких миллионов нуклеотидов (для бактериальных геномов). Картируя скаффолды и котинги на референтный геном, можно восстановить их относительное положение и заполнить пробелы (гэпы), получив полный геном (рис.11).

  
Рис.11.Схема сборки генома.

**МАТЕРИАЛЫ И МЕТОДЫ**

|  |  |
| --- | --- |
| Этап исследования | Сроки |
| Сбор библиотеки аминокислотных последовательностей белков Cas-9 и построение филогенетического дерева. | август – сентябрь 2021 |
| Сборка бактериальных геномов. | август – ноябрь 2021 |
| Создание библиотеки спейсеров. | август – декабрь 2021 |
| Разработка скриптов | декабрь 2021 |
| Создание библиотек фланкирующих последовательностей и определение PAM-последовательностей | декабрь 2021– январь 2022 |

В своем исследовании мы применяли методы вычислительной биологии (in silico), которые позволяют охватить огромное количество данных, полученных в различных лабораториях по всему миру.

Для построения филогенетического дерева были необходимы аминокислотные последовательности генов, кодирующих белки Cas-9, они были взяты из открытой базы данных «NCBI GeneBank». Выравнивание нуклеотидных последовательностей осуществлялось с помощью алгоритма «MUSCLE», далее проводился филогенетический анализ и построение дерева с помощью алгоритма «Maximum Likelihood» в приложении «MegaX».

Далее проводилась сборка геномов бактерий. «Сырые» fastq-файлы секвенирований были взяты из открытой базы данных «NCBI Assembly». Для сборки геномов было необходимо организовать сервер на операционной системе «Linux ubuntu». Для анализа качества прочтений был применен пакет «FastQC». Глобальные ошибки секвенирования были исправлены с помощью пакета «Trimmomatic». Сборка проводилась с помощью ассемблера «SPAdes». Качество сборки анализировалось с помощью пакета «Quast» и обратным картированием ридов с помощью пакета «Bowtie2».

Далее, с помощью сервиса «CRISPRs finder», приводился поиск CRISP-кассет в собранных геномах, из которых впоследствии были извлечены спейсеры.

Затем, с помощью сервиса «NCBI blast» по базе данных вирусных геномов, были найдены геномы вирусов, в которых, встречаются участки, совпадающие с бактериальными спейсерами.

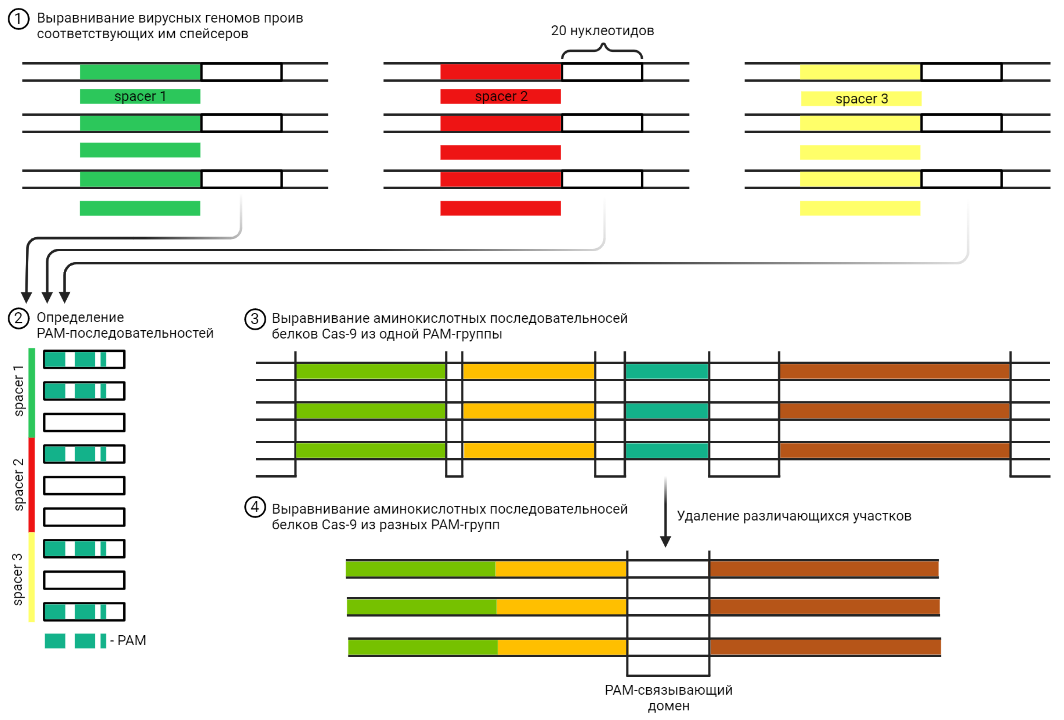


Рис.13. Схема поиска PAM-последовательностей и PAM-связывающих доменов.

Далее было проведено выравнивание полученных вирусных геномов с соответствующими им спейсерами с помощью алгоритма «ClustalW» в приложении «MegaX», отобраны 3’-фланкирующие участки длинной 20 нуклеотидов и объединены в отдельные для каждого вида бактерий библиотеки (1 на рис.13).

С помощью сервиса «WebLogo» были определены повторяющиеся мотивы – PAM-последовательности. По PAM-последовательностям бактерии были распределены по группам (далее PAM-группы) (2 на рис.13).

**РЕЗУЛЬТАТЫ**

Мы собрали библиотеку из 80 аминокислотных последовательностей белка Cas-9 бактерий из различных родов. Провели выравнивание (рис.14), затем филогенетический анализ и построили по нему филогенетическое дерево, отображающее родственность белков Cas-9 бактерий из различных родов (рис.15).

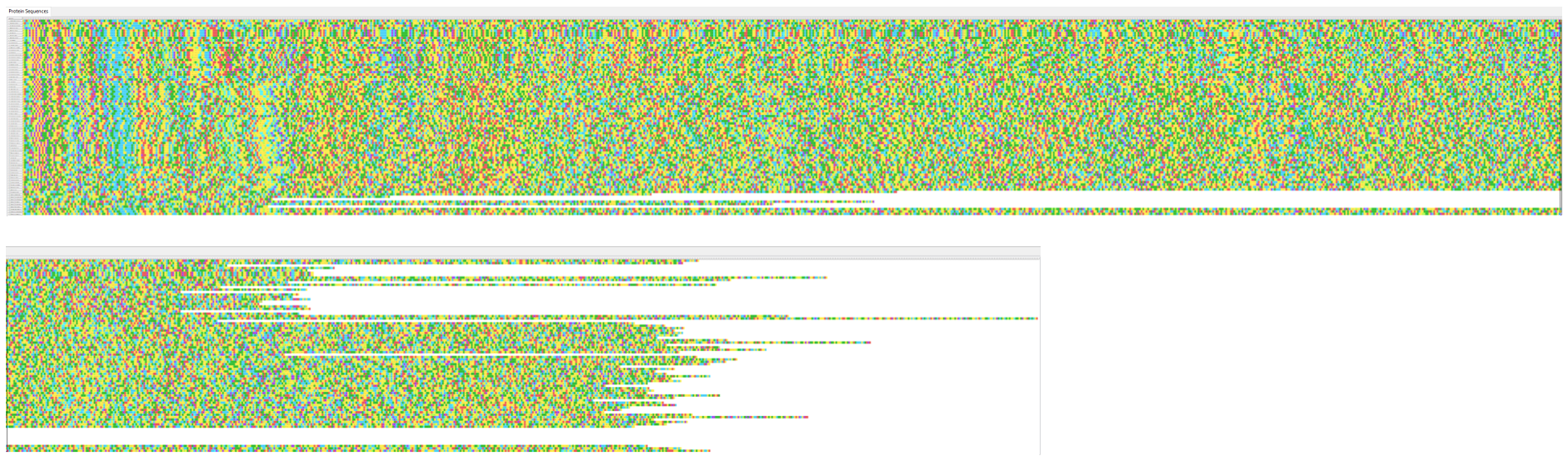
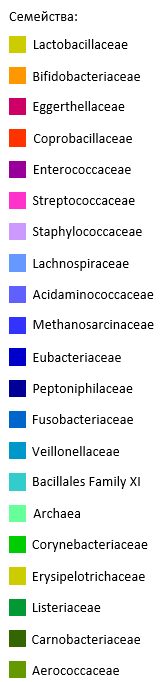
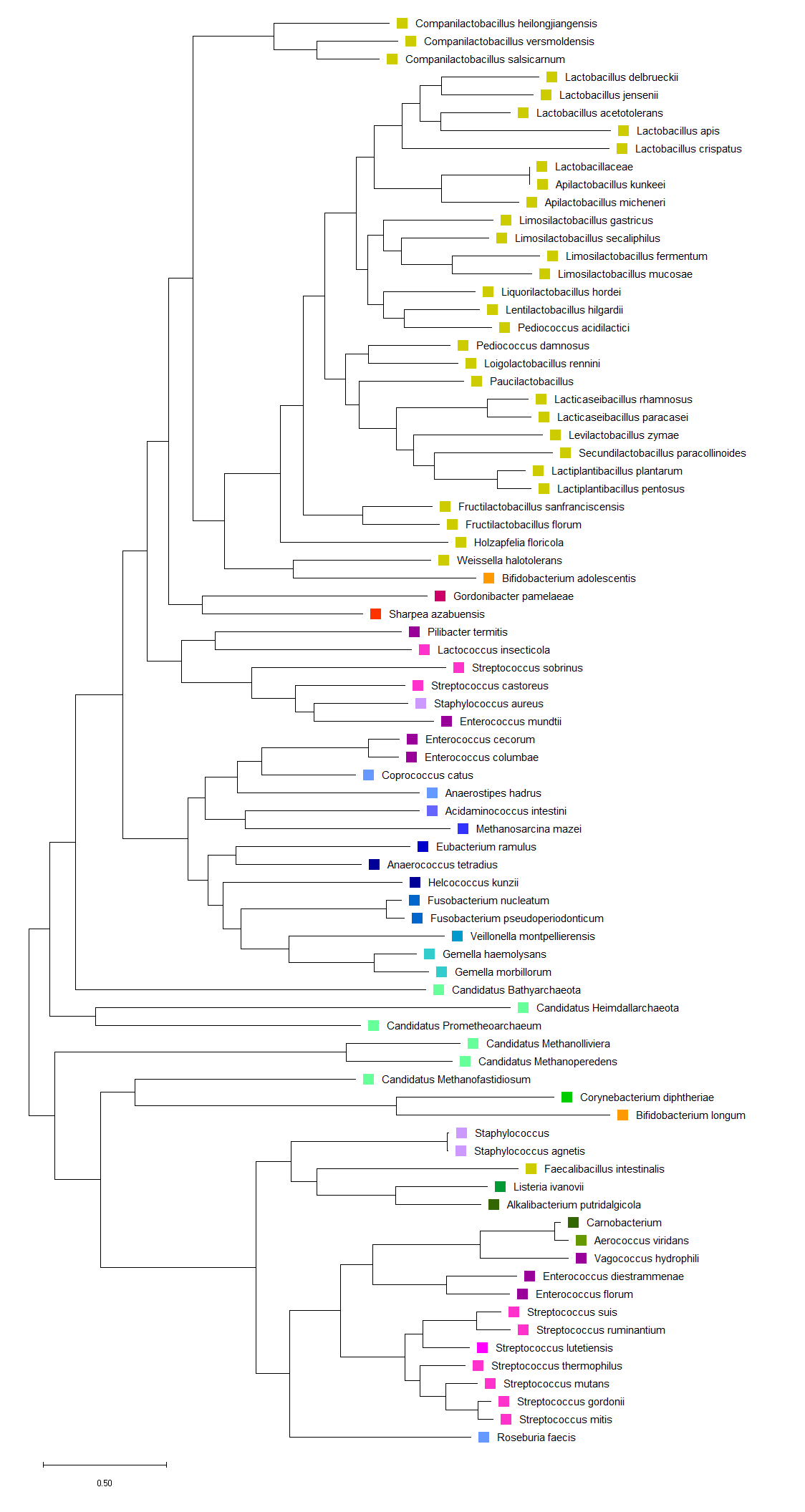


Рис.14. Выравнивание аминокислотных последовательностей Cas-9.

Из дерева видно, что прямой корреляцией между родом бактерий и их положением на дереве не существует, это, вероятно, связанно с тем, что скорость эволюции вида отличается от скорости эволюции Cas-генов.

  
Рис.15. Филогенетическое дерево последовательностей генов Cas-9 различных видов бактерий и архей.

Из дерева можно сделать ряд выводов:

1. Представители семейства Lachnospiraceae несколько раз встречаются в разных участках филогенетического дерева, вероятно Cas-9 белок этого семейства очень изменчив, из-за чего не удается объединить представителей этого семейства в одну таксономическую группу.
2. Представители семейства Bifidobacteriaceae несколько раз встречаются в разных участках филогенетического дерева, вероятно, Cas-9 белок этого семейства очень изменчив, из-за чего не удается объединить представителей этого семейства в одну таксономическую группу.
3. Представители царства архей образовали несколько параллельных ветвей, некоторые из которых совместны с другими видами бактерий из числа фирмикутов, это может свидетельствовать о взаимосвязи этих таксономических групп и о значительной вариативности Cas-9 среди царства архей.
4. Ветвь семейства Fusobacteriaceae, которое является ближайшим к грамм положительным фирмикутам по референсу, находится достаточно далеко, что свидетельствует о том, что Cas-9 меняется быстрее, чем само семейство.
5. Вероятно, грамм положительные фирмикуты на каком-то этапе эволюции разделились на две большие группы по строению Cas-9, в то время как по остальным генам они остались в одной группе.
6. Произошедшие от фирмикутов антибактерии соответственно также разделились на две группы.

Полученные результаты свидетельствуют о том, что распределение бактерий по группам с одинаковой PAM, вероятно, может отнести бактерии одинаковых родов в разные группы.

Мы собрали библиотеку геномов 110 бактерий из 22 различных родов, выделили CRISPR-кассеты и создали библиотеку спейсеров объёмом ~3000 штук, провели поиск вирусных геномов с помощью сервиса «NCBI blast». Написали скрипт на высокоуровневом языке «python3» с использованием библиотеки «bio-python3», который автоматически извлекает фланкирующие последовательности из вирусных геномов по данным hit-table.csv (рис.16).

 Рис.16. Скрипт, извлекающий 20-нкулеотидыне 3’-фланкирующие последовательности из вирусных геномов по данным hit-table.csv blast.

C его помощью установили PAM-последовательности для ряда бактерий:

*Bifidobacterium breve* – GAA (рис.17)

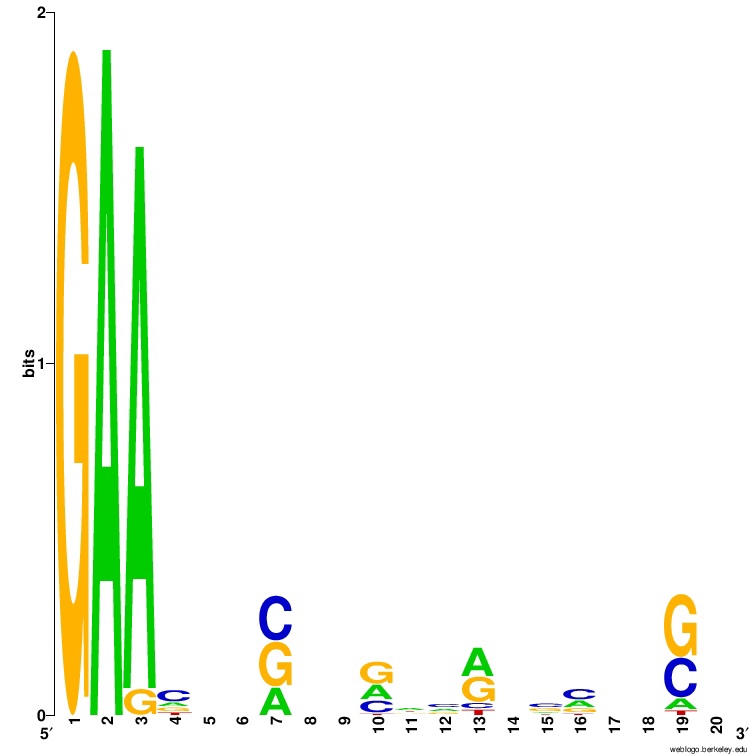


Рис.17. Logo фланкирующих последовательностей для *Bifidobacterium breve* сгенерированное с помощью «WebLogo».

*Bifidobacterium dentium* – GAA (рис.18)

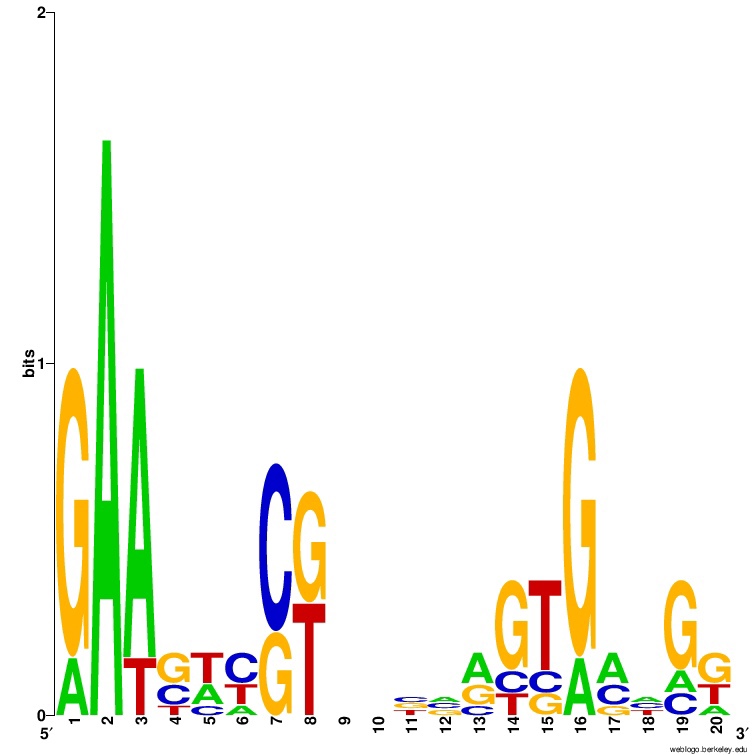


Рис.18. Logo фланкирующих последовательностей для *Bifidobacterium dentium* сгенерированное с помощью «WebLogo».

*Bifidobacterium pseudocatenulatum* – GAA (рис.19)

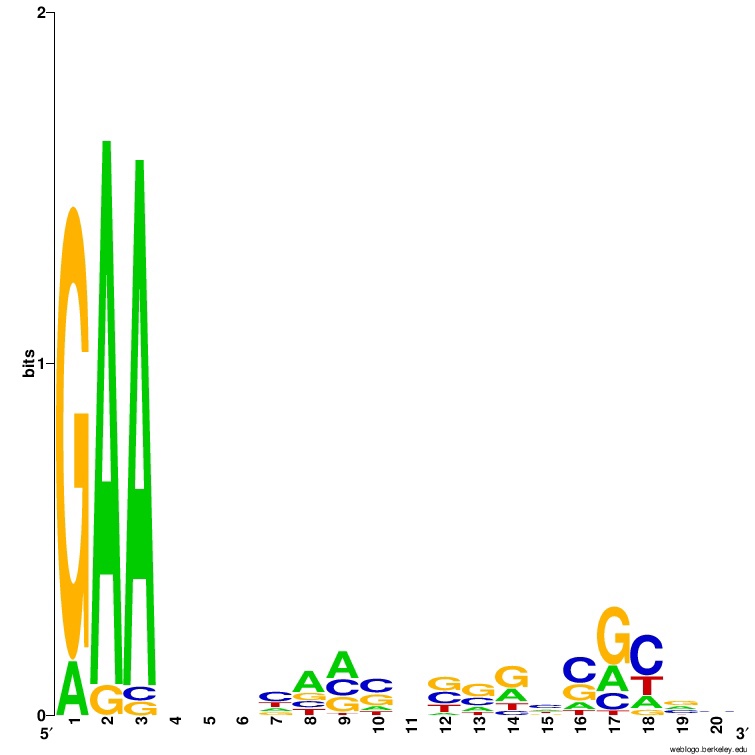


Рис.19. Logo фланкирующих последовательностей для *Bifidobacterium pseudocatenulatum* сгенерированное с помощью «WebLogo».

*Coprococcus catus* – AAA (рис.20)

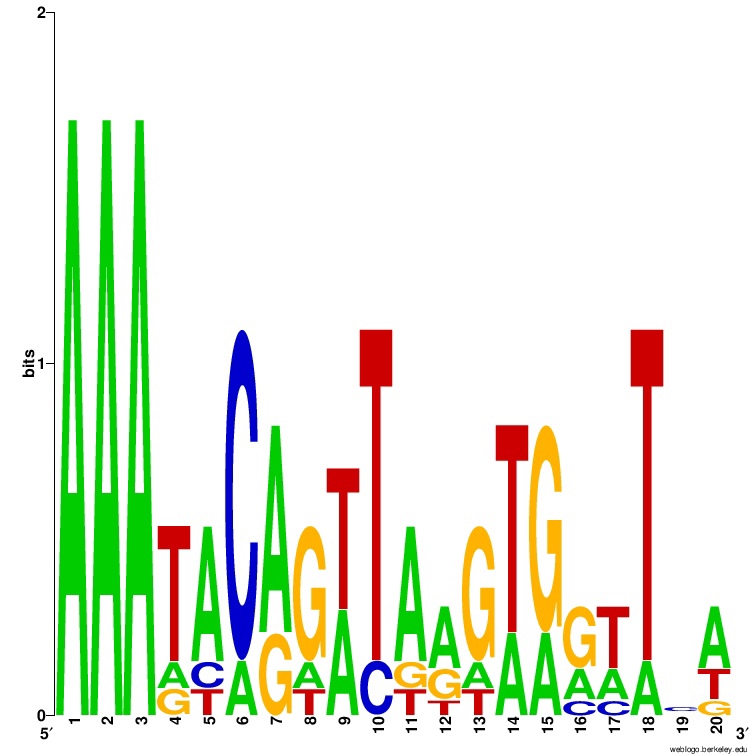


Рис.20. Logo фланкирующих последовательностей для *Coprococcus catus* сгенерированное с помощью «WebLogo».

*Gemella haemolysans* – TT (рис.21)

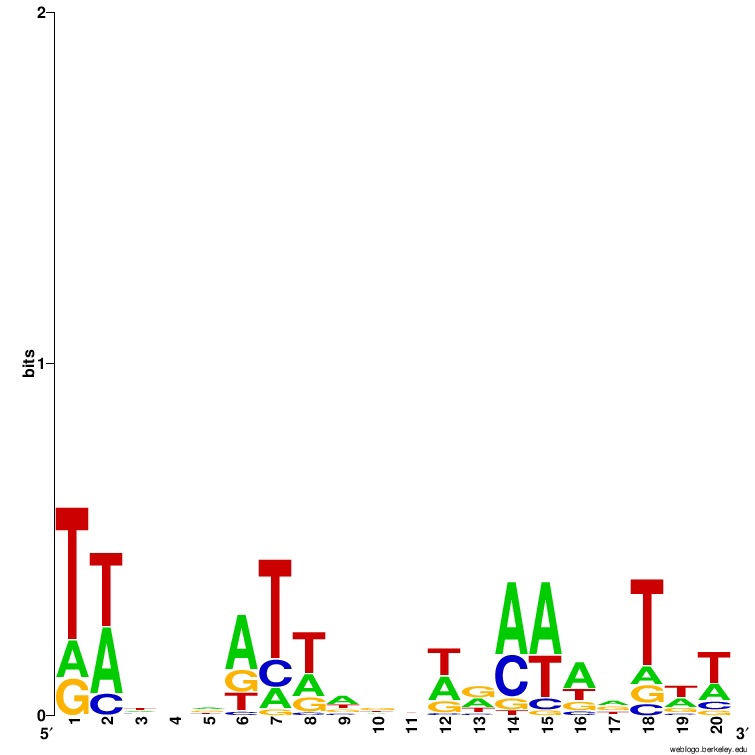


Рис.21. Logo фланкирующих последовательностей для *Gemella haemolysans* сгенерированное с помощью «WebLogo».

*Lentilactobacillus diolivorans* – NNGAAY (рис.22)

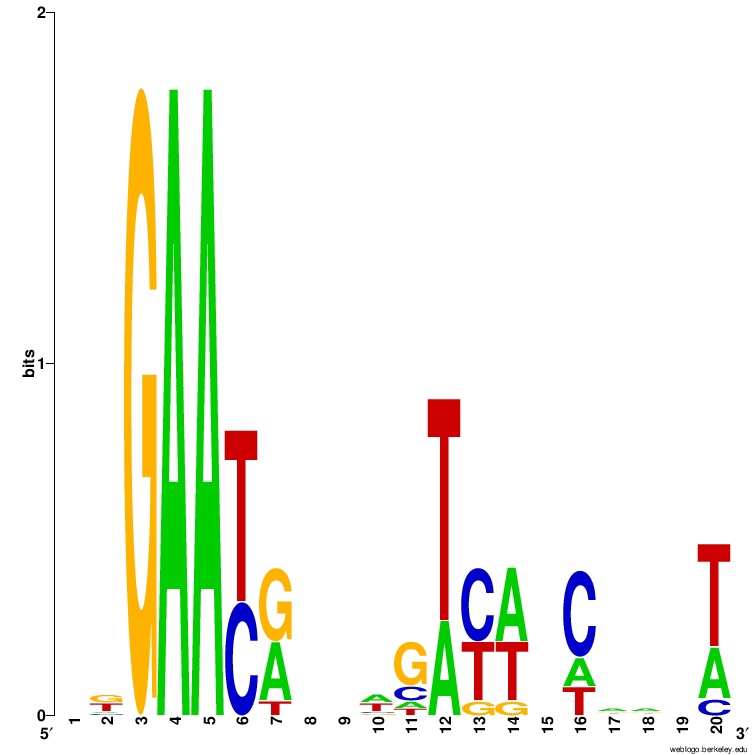


Рис.22. Logo фланкирующих последовательностей для *Lentilactobacillus diolivorans* сгенерированное с помощью «WebLogo».

*Limosilactobacillus oris* – TTNNNNK (рис.23)

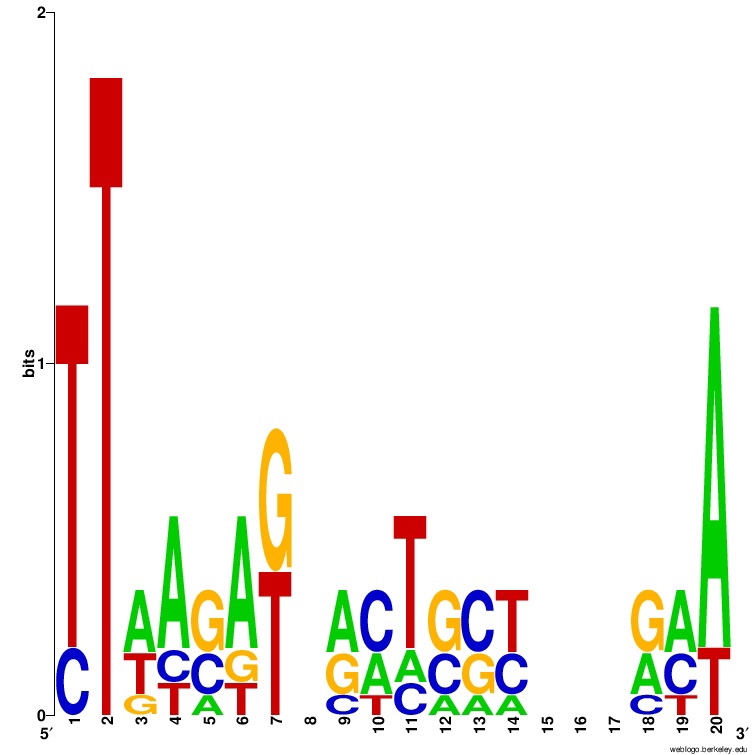


Рис.23. Logo фланкирующих последовательностей для *Limosilactobacillus oris* сгенерированное с помощью «WebLogo».

*Pediococcus parvulus* – SNAACAWW (рис.24)

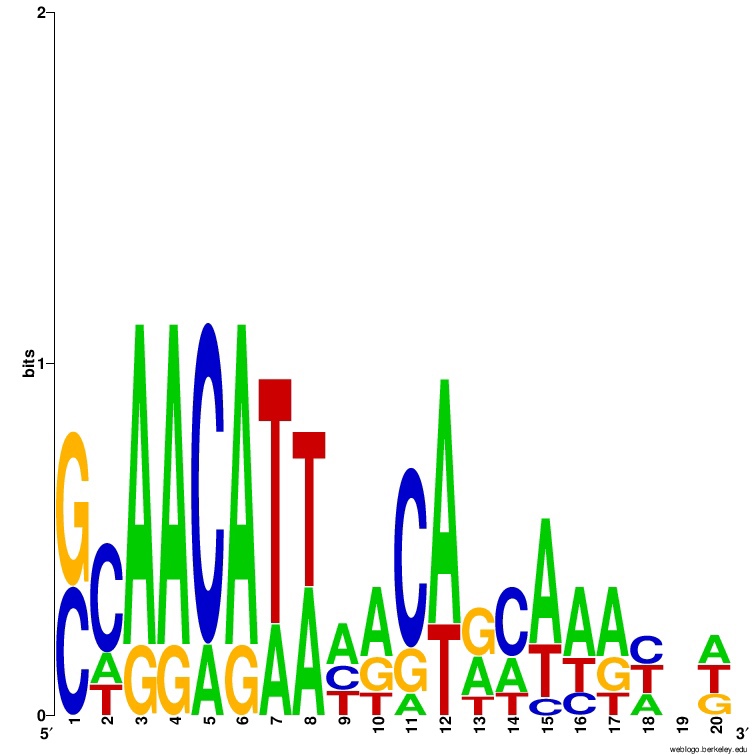


Рис.24. Logo фланкирующих последовательностей для *Pediococcus parvulus* сгенерированное с помощью «WebLogo».

*Roseburia hominis* – GAANMS (рис.25)

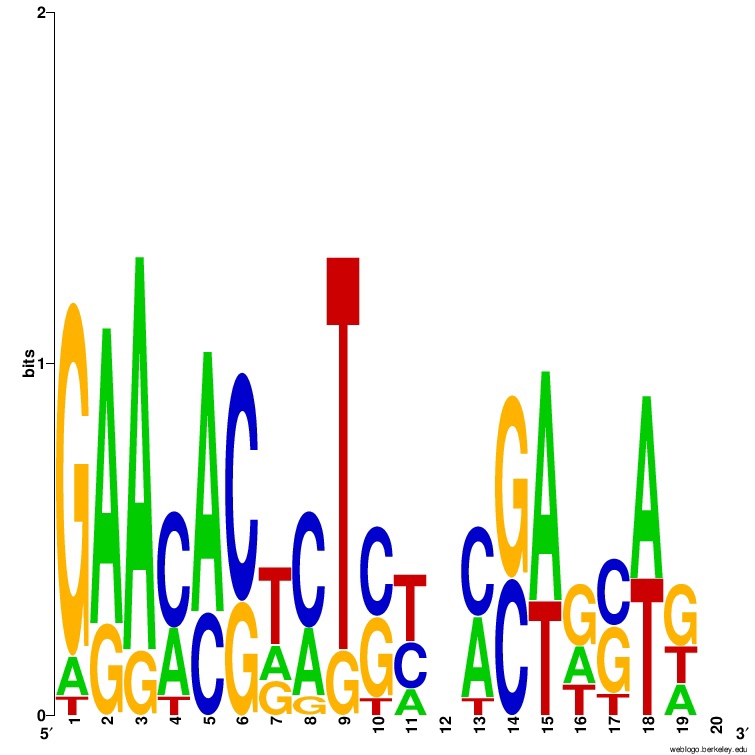


Рис.25. Logo фланкирующих последовательностей для *Roseburia hominis* сгенерированное с помощью «WebLogo».

*Streptococcus equi* – GAA (рис.26)

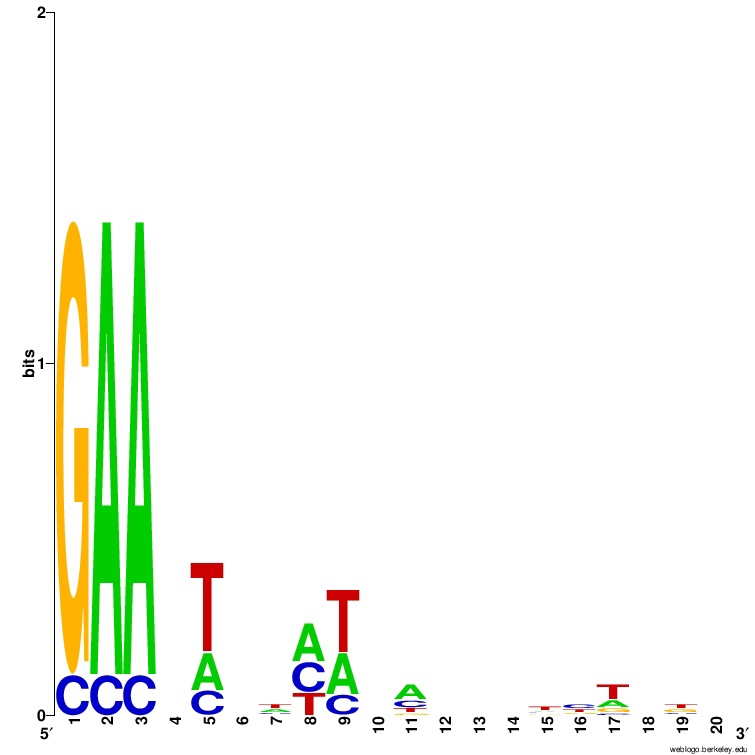


Рис.26. Logo фланкирующих последовательностей для *Streptococcus equi* сгенерированное с помощью «WebLogo».

Установленные PAM-последовательности свидетельствуют, что, как и предполагалось, PAM-последовательность никак не связанна с родовой принадлежностью.

Когда мы установим PAM-последовательности значительного количества видов бактерий, необходимо будет восстановить нативные структуры всех белков Cas-9 и локализовать PAM-связывающие домены. По их конформации можно будет судить о взаимосвязи PAM-последовательности и строения PAM-связывающих доменов.

**ВЫВОДЫ**

1. Был проведен филогенетический анализ и построено дерево по аминокислотным последовательностям белков Cas-9 различных родов бактерий. В результате анализа полученного дерева мы сделали вывод, что PAM-последовательность никак не связанна с родовой принадлежностью.
2. Была собрана библиотека бактериальных геномов объёмом 110 штук.
3. Определены PAM-последовательности некоторых видов бактерий из различных родов:

*Bifidobacterium breve* – GAA

*Bifidobacterium dentium* – GAA

*Bifidobacterium pseudocatenulatum* – GAA

*Coprococcus catus* – AAA

*Gemella haemolysans* – TT

*Lentilactobacillus diolivorans* – NNGAAY

*Limosilactobacillus oris* – TTNNNNK

*Pediococcus parvulus* – SNAACAWW

*Roseburia hominis* – GAANMS

*Streptococcus equi* – GAA

**ЗАКЛЮЧЕНИЕ**

Установленные в ходе нашего исследования PAM-последовательности позволят применять соответствующие Cas-9 в лабораторной практике, тем самым расширяя условия использования технологии CRISPR Cas-9, делая ее более гибкой и универсальной, а разработанные скрипты позволят облегчить определение PAM-последовательностей Cas-9 любых бактерий.

Далее необходимо провести выравнивание бактериальных геномов против эталонной последовательности гена Cas-9 с помощью алгоритма «ClustalW»». В результате выравнивания будут установлены последовательности генов Cas-9 каждого вида бактерий. Затем, с помощью сервиса «NCBI ORFfinder», в установленных последовательностях будут найдены открытые рамки считывания и установлены последовательности мРНК, по которым далее будут установлены аминокислотные последовательности белков Cas-9 каждого вида бактерий.

Затем необходимо провести выравнивание аминокислотных последовательностей белков Cas-9 внутри каждой PAM-группы с помощью алгоритма «MUSCLE». Сходные участки, вероятно, будут конформными доменами и PAM-связывающими доменами, Различающиеся участки будут вариабельными, их необходимо заменить буквами «NNN» и построить консенсусы белков Cas-9 для каждой PAM-группы.

Потом необходимо выровнять полученные консенсусы между собой с помощью алгоритма «MUSCLE». Сходные участки будут конформными доменами, а участки различия будут PAM-связывающими доменами.

Далее необходимо смоделировать третичные структуры белков Cas-9 по референсу с помощью сервиса «PyMOL», найти в этих структурах PAM-связывающие домены и отследить взаимосвязи PAM-последовательности и строения PAM-связывающих доменов.

Знание структур PAM-связывающих доменов, позволит выявлять закономерности и предсказывать PAM-последовательности методами биоинформатики и филогенетики. В перспективе возможно конструирование белков Cas-9 под условия конкретного эксперимента методами инженерной энзимологии.

**СПИСОК ЛИТЕРАТУРЫ**

1. FJM Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Alicante, 2009.
2. Luciano A. M, Ph.D. The CRISPR-Cas system of Streptococcus pyogenes: function and applications. New York, 2016.
3. Mendoza B. J., Trinh C. T. In Silico Processing of the Complete CRISPR-Cas Spacer Space for Identification of PAM Sequences. Knoxville, 2018.
4. Ryan T. L., Chase L. B. Deciphering, communicating, and engineering the CRISPR PAM. Raleigh, 2018.
5. Васильева Н.Ю. Биоинформатика: Множественное выравнивание. Филогенетические деревья: методическое пособие. Одесса, 2014.
6. Гоглева А.А. Исследование CRISPR-систем прокариотического иммунитета методами сравнительной геномики: диссертация на соискание учёной степени кандидата биологических наук. Москва, 2016.
7. Гребенкина Н.А., Андреюк Д.А. Генная инженерия
8. 8.Гребенкина Н.А., Глазова О.В., Митяева О.Н., Андреюк Д.С. Практикум по генному редактированию.
9. Зленко Д.В. Анализ последовательностей.
10. Немудрый А.А., Валетдинова К.Р., Медведев С.П., Закиян С.М. Системы редактирования геномов TALEN и CRISPR/Cas – инструменты открытий. Новосибирск, 2014.
11. Немудрый А.А. Исправление мутации в гене аргинин-вазопрессина крыс линии brattleboro in vitro: диссертация на соискание учёной степени кандидата биологических наук. Новосибирск, 2017.
12. Нурк С.Ю. Сборка генома и графы де Брюина.
13. Овчаренко Е.К. Роль гена ruvC в CRISPR-интерференции и адаптации у Escherichia coli. Москва, 2016.
14. Огурцов А.Н. Методы биоинформационного анализа: учебное пособие по курсу «Биоинформатика информационная биотехнология». Харьков, 2011.
15. Романенков К.В. Методы и алгоритмы автоматизированной сборки генома на вычислительном кластере.