ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

Биоинформатика

Код модуля	Модуль
1157971(0)	Информационно-аналитические методы в науке, медицине, фармацевтике и образовании

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Безматерных Максим Алексеевич	кандидат химических наук, доцент	Доцент	технологии органического синтеза

Согласовано:

Управление образовательных программ С.А. Иванченко

Авторы:

- **Безматерных Максим Алексеевич**, **Доцент**, **технологии органического синтеза**

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Биоинформатика

1.	**Объем дисциплины в зачетных единицах**	4	
2.	**Виды аудиторных занятий**	Лекции Практические/семинарские занятия Лабораторные занятия	
3.	**Промежуточная аттестация**	Зачет	
4.	**Текущая аттестация**	Контрольная работа	1
		Домашняя работа	2

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Биоинформатика

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	**Планируемые результаты обучения (индикаторы)**	**Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине**
1	**2**	**3**
ОПК-2 -Способен самостоятельно ставить, формализовывать и решать задачи, относящиеся к профессиональной деятельности, используя методы моделирования и математического анализа (Живые системы. Перспективные химико-фармацевтические и биотехнологии: исследования и разработки)	Д-1 - Проявлять ответственность и настойчивость в достижении цели З-1 - Сделать обзор основных методов моделирования и математического анализа, применимых для формализации и решения задач профессиональной деятельности З-2 - Характеризовать сферы применения и возможности пакетов прикладных программ для решения задач профессиональной деятельности П-1 - Решать самостоятельно сформулированные	Домашняя работа № 1 Домашняя работа № 2 Зачет Контрольная работа Лабораторные занятия Лекции Практические/семинарские занятия

	практические задачи, относящиеся к профессиональной деятельности методами моделирования и математического анализа, в том числе с использованием пакетов прикладных программ У-1 - Самостоятельно сформулировать задачу области профессиональной деятельности, решение которой требует использования методов моделирования и математического анализа У-2 - Использовать методы моделирования и математического анализа, в том числе с использованием пакетов прикладных программ для решения задач профессиональной деятельности	
ОПК-4 -Способен выбирать и использовать существующие информационно-коммуникационные технологии и вычислительные методы для решения задач в области профессиональной деятельности (Живые системы. Перспективные химико-фармацевтические и биотехнологии: исследования и разработки)	Д-1 - Демонстрировать аналитические и системные умения, способность к поиску информации З-1 - Представлять возможности современных информационно-коммуникационных средств и технологий сбора, передачи, обработки и накопления информации, создания баз данных, используемых в области профессиональной деятельности П-1 - Иметь опыт сбора, анализа и обработки информации при решении задач профессиональной деятельности с использованием современных информационно-коммуникационных технологий и баз данных У-1 - Выбирать и использовать современные IT-технологии и базы данных при сборе, анализе, обработке и представлении информации для решения задач	Домашняя работа № 1 Домашняя работа № 2 Зачет Лабораторные занятия Лекции Практические/семинарские занятия

	профессиональной деятельности	
ОПК-5 -Способен готовить публикации, участвовать в профессиональных дискуссиях, представлять результаты профессиональной деятельности в виде докладов на российских и международных конференциях (Живые системы. Перспективные химико-фармацевтические и биотехнологии: исследования и разработки)	Д-1 - Демонстрировать аналитические умения и креативное мышление Д-2 - Проявлять внимательность и ответственность в подготовке материалов научных исследований к публичному доступу З-1 - Демонстрировать понимание правил оформления различных видов и способов представления результатов: научных и научно-технических отчетов, презентаций, публикаций (докладов, статей, тезисов к конференциям, обзоров), стилей и норм научного письма на русском и английском языках З-2 - Соотносить правила проведения профессиональных дискуссий с их характером, и демонстрировать понимание особенностей научных дискуссий П-2 - Иметь опыт подготовки выступлений и ведения профессиональных дискуссий, выступлений на семинарах и/или конференциях У-1 - Оценивать выполненные отчеты, презентации, научные публикации (доклады, статьи, тезисы к конференциям, обзоры) на соответствие нормам научного письма на русском и английском языках	Домашняя работа № 1 Домашняя работа № 2 Зачет Лабораторные занятия Лекции Практические/семинарские занятия
УК-1 -Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий, в том числе в цифровой среде (Живые	З-2 - Определять этапы разработки стратегии действий, в том числе в цифровой среде, и методы решения проблемных ситуаций П-1 - Использовать эффективные стратегии действий для решения проблемной ситуации, в том числе в цифровой среде, с	Зачет Контрольная работа Лабораторные занятия Лекции Практические/семинарские занятия

системы. Перспективные химико-фармацевтические и биотехнологии: исследования и разработки)	учетом оценки ограничений, рисков и моделируемых результатов У-2 - Обосновывать выбор стратегии для достижения поставленной цели, в том числе в цифровой среде, с учетом ограничений, рисков и моделируемых результатов	
УК-7 -Способен обрабатывать, анализировать, передавать данные и информацию с использованием цифровых средств для эффективного решения поставленных задач с учетом требований информационной безопасности (Живые системы. Перспективные химико-фармацевтические и биотехнологии: исследования и разработки)	З-3 - Сделать обзор современных цифровых средств и технологий, используемых для обработки, анализа и передачи данных при решении поставленных задач П-1 - Обосновать выбор технических и программных средств защиты персональных данных и данных организации при работе с информационными системами на основе анализа потенциальных и реальных угроз безопасности информации П-2 - Решать поставленные задачи, используя эффективные цифровые средства и средства информационной безопасности У-2 - Выбирать современные цифровые средства и технологии для обработки, анализа и передачи данных с учетом поставленных задач	Домашняя работа № 1 Домашняя работа № 2 Зачет Лабораторные занятия Лекции Практические/семинарские занятия

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

1. **Лекции**: **коэффициент значимости совокупных результатов лекционных занятий – 0.6**		
Текущая аттестация на лекциях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
контрольная работа	*4*	*40*

Ведение конспекта	8	*20*
домашняя работа	8	*40*
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.4		
Промежуточная аттестация по лекциям – зачет **Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.6**		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.2		
Текущая аттестация на практических/семинарских занятиях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
домашняя работа	12	40
Работа на занятиях	16	60
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям–нет **Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– 1**		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –0.2		
Текущая аттестация на лабораторных занятиях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
работа на занятиях	17	60
защита отчетов	17	40
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям -1		
Промежуточная аттестация по лабораторным занятиям –нет **Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено**		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено		
Текущая аттестация на онлайн-занятиях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет **Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено**		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	**Сроки − семестр, учебная неделя**	**Максимальная оценка в баллах**
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		

Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	**Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам**
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)			
№ п/п	**Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)**	**Шкала оценивания**	
		Традиционная характеристика уровня	**Качественная характеристика уровня**

1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Практические/семинарские занятия

Примерный перечень тем

1. Секвенирование. Контроль качества данны
2. Anaconda, среды, предобработка данных: QC и тримминг
3. Анализ метагеномных данных. Сборка метагенома
4. Сценарии обработки и анализа метагеномных данных
5. Классификация генетических последовательностей Подходы на основе выравниваний. BLAST
6. Классификация генетических последовательностей Подходы без использования выравнивания. Анализ спектра k-меров.
7. Вирусная метагеномика

Примерные задания

Тестовые задания

Метагеномика занимается изучением:

а) Генетического материала любых сообществ

б) Генетического материала только сообществ животных и человека

в) Бактерий, вирусов и других микроорганизмов, населяющих сообщество

г) Только бактерий, населяющих сообщество

д) Функциональных связей между микроорганизмами в сообществе

у)Только таксономическим анализом микроорганизмов, населяющих сообщества

Выберите верные суждения о геноме :

а) Геном - совокупность всего генетического материала организма/отдельной клетки

б) Геном может быть представлен только ДНК

в) У некоторых вирусов геном представлен РНК

г) Размер генома не зависит от эволюционного возраста вида

д) Геном человека впервые был расшифрован в 1953 году

у) Геном человека самый большой среди живых организмов

ж) Самый маленький геном принадлежит представителю вирус

Выберете верные суждения о репликации ДНК *

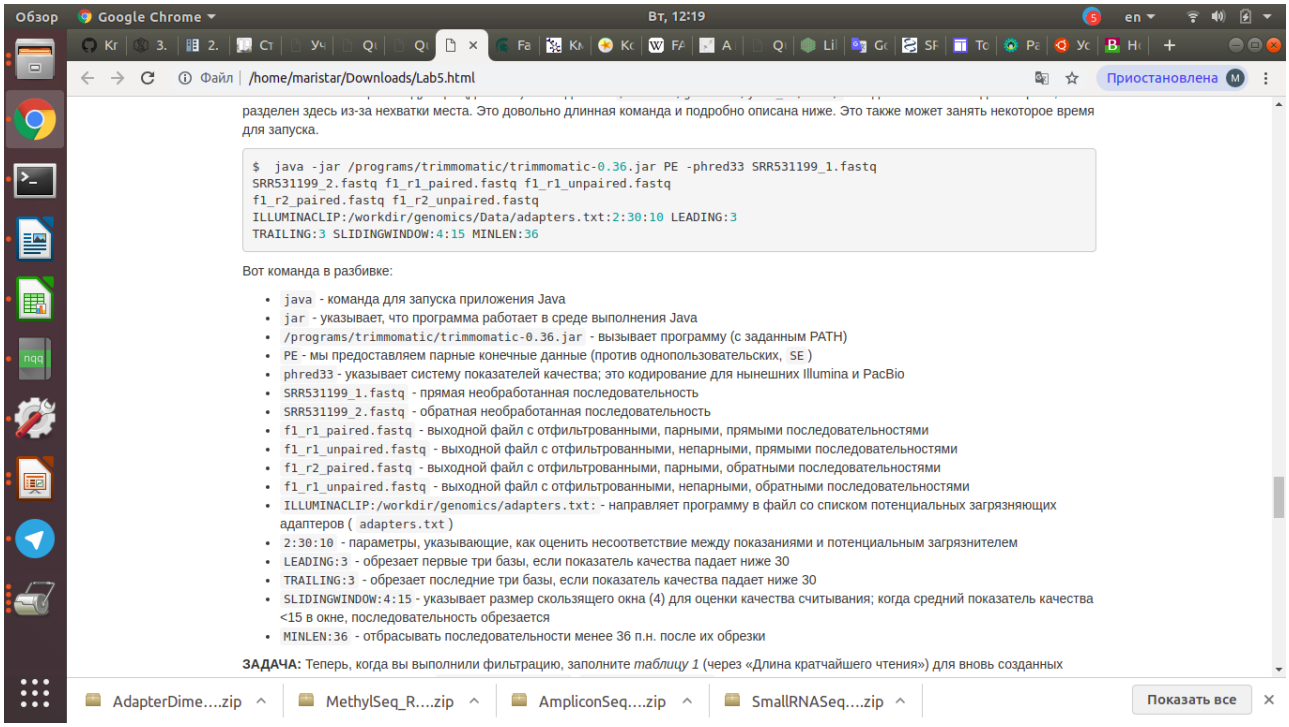
а) Репликация - процесс синтеза РНК по матрице ДНК

б) Для репликации необходима ДНК-полимераза, которая способна пришивать нуклеотиды только к 5'-концу

в) Для репликации необходима ДНК-полимераза, которая способна пришивать нуклеотиды только к 3'-концу

г) Для репликации необходима ДНК-полимераза, которая способна пришивать нуклеотиды и к 3'-концу, и к 5'-концу

в) Репликация происходит всякий раз, когда размножаются клетки организма



Работа с файлами.

1. Скачайте файлы list.viruses.txt, Data.csv, viral.3.1.genomic.fna и miseq_reads_metagenome_fastp_1/2.fastq и поместите в папку Documents/Viruses

2. Переименуйте файлы viral.3.1.genomic.fna и miseq_reads_metagenome_fastp_1/2.fastq на viruses.fasta и viruses_1/2.fastq соответственно.

mv viral.3.1.genomic.fna viruses.fasta

mv miseq_reads_metagenome_fastp_1.fastq viruses_1.fastq

mv miseq_reads_metagenome_fastp_2.fastq viruses_2.fastq

3. Подсчитайте число строк в каждом файле, просмотрите содержимое с помощью mc
wc -l viruses.fasta
wc -l viruses_1.fastq
wc -l viruses_2.fastq

4. Отсортируйте строки в файле list.viruses.txt и запишите в новый файл list.viruses.sort.txt
sort list.viruses.txt > list.viruses.sort.txt

5. Выведите первые 5 строк из файла list.viruses.txt и дозапишите в файл list_F15.txt, переименуйте файл на list_F20.txt
head -n 5 list.viruses.txt > list_F15.
mv list_F15.txt list_F20.txt

6. Удалите дубли строк из файла list_F20.txt и запишите уникальные строки в файл uniq.txt
cat list_F20.txt | sort | uniq > uniq.txt

Перед следующим заданием посмотрим справку о команде cut
man cut или cut --help

7. Выведите 1, 2 и 4 столбец из файла Data.csv
cut -f 1,2,4 Data.csv

8. Удалить идентификаторы из файла list.viruses.txt, оставив только названия
cat list.viruses.txt | cut -c 12-
cat list.viruses.txt | cut -d' ' -f 2-

Скопируйте файл viral.3.1.genomic.fna в рабочую директорию
Ознакомьтесь с командой grep

9. Выведете все последовательности, содержащие слово 'arenavirus' из файла viral.3.1.genomic.fna
grep 'arenavirus' -A 1 viral.3.1.genomic.fna

10. Выведете все последовательности, НЕсодержащие слово 'arenavirus' из файла viral.3.1.genomic.fna
grep -v 'arenavirus' -A 1 viral.3.1.genomic.fna

11. Посчитайте количество последовательностей в файле viral.3.1.genomic.fna
grep -c '>' viral.3.1.genomic.fna
Работа с файлами.

1. Скачайте файлы list.viruses.txt, Data.csv, viral.3.1.genomic.fna и miseq_reads_metagenome_fastp_1/2.fastq и поместите в папку Documents/Viruses

2. Переименуйте файлы viral.3.1.genomic.fna и miseq_reads_metagenome_fastp_1/2.fastq на viruses.fasta и viruses_1/2.fastq соответственно.

mv viral.3.1.genomic.fna viruses.fasta
mv miseq_reads_metagenome_fastp_1.fastq viruses_1.fastq
mv miseq_reads_metagenome_fastp_2.fastq viruses_2.fastq

3. Подсчитайте число строк в каждом файле, просмотрите содержимое с помощью mc

wc -l viruses.fasta
wc -l viruses_1.fastq
wc -l viruses_2.fastq

4. Отсортируйте строки в файле list.viruses.txt и запишите в новый файл list.viruses.sort.txt

sort list.viruses.txt > list.viruses.sort.txt

5. Выведите первые 5 строк из файла list.viruses.txt и дозапишите в файл list_F15.txt, переименуйте файл на list_F20.txt

head -n 5 list.viruses.txt > list_F15.
mv list_F15.txt list_F20.txt

6. Удалите дубли строк из файла list_F20.txt и запишите уникальные строки в файл uniq.txt

cat list_F20.txt | sort | uniq > uniq.txt

Перед следующим заданием посмотрим справку о команде cut

man cut или cut --help

7. Выведите 1, 2 и 4 столбец из файла Data.csv

cut -f 1,2,4 Data.csv

8. Удалить идентификаторы из файла list.viruses.txt, оставив только названия

cat list.viruses.txt | cut -c 12-
cat list.viruses.txt | cut -d' ' -f 2-

Скопируйте файл viral.3.1.genomic.fna в рабочую директорию

Ознакомьтесь с командой grep

9. Выведете все последовательности, содержащие слово 'arenavirus' из файла viral.3.1.genomic.fna

grep 'arenavirus' -A 1 viral.3.1.genomic.fna

10. Выведете все последовательности, НЕсодержащие слово 'arenavirus' из файла viral.3.1.genomic.fna

grep -v 'arenavirus' -A 1 viral.3.1.genomic.fna

11. Посчитайте количество последовательностей в файле viral.3.1.genomic.fna

grep -c '>' viral.3.1.genomic.fna

12. Выведите все нуклеотидные последовательности по идентификаторам из файла uniq.txt
grep -f uniq.txt -A 1 viral.3.1.genomic.fna > Uniq.fasta

sed и awk
https://www.opennet.ru/docs/RUS/bash_scripting_guide/a14586.html
https://bioinformatics.cvr.ac.uk/essential-awk-commands-for-next-generation-sequence-analysis/

13. Добавьте в начало всех строк list.viruses.txt символ '>' и запишите в файл id.txt
sed 's/^/>/' list.viruses.txt > id.txt

14. Выведите каждую вторую строку из файла viruses.fasta
awk 'FNR%2' viruses.fasta

15. Посмотреть статистику файлов miseq_reads_metagenome_fastp_1/2.fastq
seqkit stat -a miseq_reads_metagenome_fastp_1.fastq
seqkit stat -a miseq_reads_metagenome_fastp_2.fastq

16. Извлечь из miseq_reads_metagenome_fastp_1/2.fastq все последовательности, содержащие подпоследовательность ATCGAAG
seqkit grep --by-seq --max-mismatch 1 --pattern "ATCGAAG" miseq_reads_metagenome_fastp_1.fastq > ATCGAAG_1.fastq
seqkit grep --by-seq --max-mismatch 1 --pattern "ATCGAAG" miseq_reads_metagenome_fastp_2.fastq > ATCGAAG_2.fastq

12. Выведите все нуклеотидные последовательности по идентификаторам из файла uniq.txt
grep -f uniq.txt -A 1 viral.3.1.genomic.fna > Uniq.fasta

sed и awk
https://www.opennet.ru/docs/RUS/bash_scripting_guide/a14586.html
https://bioinformatics.cvr.ac.uk/essential-awk-commands-for-next-generation-sequence-analysis/

13. Добавьте в начало всех строк list.viruses.txt символ '>' и запишите в файл id.txt
sed 's/^/>/' list.viruses.txt > id.txt

14. Выведите каждую вторую строку из файла viruses.fasta
awk 'FNR%2' viruses.fasta

15. Посмотреть статистику файлов miseq_reads_metagenome_fastp_1/2.fastq
seqkit stat -a miseq_reads_metagenome_fastp_1.fastq
seqkit stat -a miseq_reads_metagenome_fastp_2.fastq

16. Извлечь из miseq_reads_metagenome_fastp_1/2.fastq все последовательности, содержащие подпоследовательность ATCGAAG
seqkit grep --by-seq --max-mismatch 1 --pattern "ATCGAAG" miseq_reads_metagenome_fastp_1.fastq > ATCGAAG_1.fastq

seqkit grep --by-seq --max-mismatch 1 --pattern "ATCGAAG" miseq_reads_metagenome_fastp_2.fastq > ATCGAAG_2.fastq

LMS-платформа − не предусмотрена

5.1.3. Лабораторные занятия

Примерный перечень тем

1. Секвенирование по Сэнгеру
2. Инструменты для обработки данных
3. Алгоритм OLC - Overlap–Layout-Consensus
4. Поиск оптимального выравнивания S1 [1, i] с S2
5. Базы данных аминокислотных последовательностей
6. Анализ спектра k-меров. Сравнение метагеномных образцов
7. Таргетное секвенирование

LMS-платформа − не предусмотрена

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа

Примерный перечень тем

1. Методы оптимального выравнивания (Megahit и Spades)
2. Построение индентификаторов

Примерные задания

1. Соберите прочтения (согласно вашему варианту) с помощью SPAdes и MEGAHIT (проделали на занятии, контиги лежат на гугл-диске в папке data). В Отчете укажите количество контигов для каждого инструмента, среднюю длину контигов и длину самого длинного контига (можно воспользоваться seqkit stat).

	Megahit	Spades
количество контигов	3064	1263
средняя длина контигов	929,7	1943
длина самого длинного контига	284058	284110

2. Оцените полноту сборки, полученной с помощью SPAdes и MEGAHIT (обратным выравниванием прочтений на контиги). В отчете укажите процент выравненных прочтений (процент полноты сборки).

Процент сборки ридов в контиги Megahit
145824/150900*100=96,63%

Процент сборки ридов в контиги Spades
145530/150900*100=96,44%

3. На основании п. 1 о количестве и длине контигов, а также п.2 на основании полноты сборки, сделайте выводы о том, какая сборка является лучше.

Процент сборки ридов в кинтиги с помощью Megahit и Spades почти одинаковый. количество кинтиов больше со сборки Megahit, а средняя длина контигов меньше. Максимальная длина длина также больше у Spades. исходя из этого делаем вывод, что сборка с помощью Spades лучше.

Задание 1. Соберите прочтения (согласно вашему варианту) с помощью SPAdes и MEGAHIT (проделали на занятии, контиги лежат на гугл-диске в папке data). В Отчете укажите количество контигов для каждого инструмента, среднюю длину контигов и длину самого длинного контига (можно воспользоваться seqkit stat).

Таблица 1. Параметры сборок, полученных с помощью Megahit и SPAdes:

№	Параметр	Megahit	SPAdes
1	Количество контигов	3064	1263
2	Средняя длина контига	929,7	1943
3	Максимальная длина	284058	284110

Задание 2. Оцените полноту сборки, полученной с помощью SPAdes и MEGAHIT (обратным выравниванием прочтений на контиги). В отчете укажите процент выравненных прочтений (процент полноты сборки).

Таблица 2. Полнота сбороки, полученных с помощью Megahit и SPAdes:

№	Megahit	SPAdes
1	96,63 %	96,44 %

Задание 3. На основании п. 1 о количестве и длине контигов, а также п.2 на основании полноты сборки, сделайте выводы о том, какая сборка является лучше.

Согласно таблице 2 процент выравненных прочтений при сборке с помощью Megahit и SPAdes (96,63 % и 96,44 % соответственно) практически не отличается. Однако, сборка, полученная с помощью SPAdes считается лучше, так как при осуществлении сборки с помощью SPAdes средняя длина контига больше, а количество контигов меньше, чем при сборке с помощью Megahit (Таблица 1).

Задание 1. Соберите прочтения (согласно вашему варианту) с помощью SPAdes и MEGAHIT (проделали на занятии, контиги лежат на гугл-диске в папке data). В Отчете укажите количество контигов для каждого инструмента, среднюю длину контигов и длину самого длинного контига (можно воспользоваться seqkit stat).

Таблица 1. Параметры сборок, полученных с помощью Megahit и SPAdes:

№	Параметр	Megahit	SPAdes
1	Количество контигов	3064	1263
2	Средняя длина контига	929,7	1943
3	Максимальная длина	284058	284110

Задание 2. Оцените полноту сборки, полученной с помощью SPAdes и MEGAHIT (обратным выравниванием прочтений на контиги). В отчете укажите процент выравненных прочтений (процент полноты сборки).

Таблица 2. Полнота сбороки, полученных с помощью Megahit и SPAdes:

№	Megahit	SPAdes
1	96,63 %	96,44 %

Задание 3. На основании п. 1 о количестве и длине контигов, а также п.2 на основании полноты сборки, сделайте выводы о том, какая сборка является лучше.

Согласно таблице 2 процент выравненных прочтений при сборке с помощью Megahit и SPAdes (96,63 % и 96,44 % соответственно) практически не отличается. Однако, сборка, полученная с помощью SPAdes считается лучше, так как при осуществлении сборки с помощью SPAdes средняя длина контига больше, а количество контигов меньше, чем при сборке с помощью Megahit (Таблица 1).

LMS-платформа − не предусмотрена

5.2.2. Домашняя работа № 1

Примерный перечень тем

1. Основыные работы в операционной системе Linux

Примерные задания

Терминал:

```
evgeny@evgeny-VirtualBox:~/metagenomics$ grep 'adenovirus' vir.fasta >> new.txt
evgeny@evgeny-VirtualBox:~/metagenomics$ grep 'adenovirus' vir.fasta | wc -l >> new.txt
```

Содержимое файла “new.txt”:

```
>NC_006144.1 Simian adenovirus 3, complete genome
>AC_000010.1 Simian adenovirus 21, complete genome
>NC_001720.1 Fowl adenovirus A, complete genome
>NC_014899.1 Murine adenovirus 2, complete genome
>NC_016437.1 South polar skua adenovirus-1, complete genome
>NC_001813.1 Duck adenovirus A, complete genome
>NC_012959.1 Human adenovirus 54, complete genome
>NC_020487.1 Titi monkey adenovirus ECC-2011, complete genome
>AC_000020.1 Canine adenovirus type 2, complete genome
>AC_000017.1 Human adenovirus type 1, complete genome
```

>AC_000018.1 Human adenovirus type 7, complete genome
>AC_000019.1 Human adenovirus type 35, complete genome
>AC_000004.1 Duck adenovirus A, complete genome
>AC_000006.1 Human adenovirus D, complete genome
>AC_000009.1 Porcine adenovirus C, complete genome
>AC_000016.1 Turkey adenovirus A, complete genome
>AC_000011.1 Simian adenovirus 25, complete genome
>AC_000013.1 Fowl adenovirus D, complete genome
>AC_000014.1 Fowl adenovirus A, complete genome
>NC_011202.1 Human adenovirus B2, complete genome
>NC_015323.1 Fowl adenovirus C, complete genome
>AC_000001.1 Ovine adenovirus A, complete genome
>AC_000189.1 Porcine adenovirus 3, complete genome
>AC_000190.1 Tree shrew adenovirus 1, complete genome
>AC_000191.1 Bovine adenovirus A, complete genome
>NC_002501.1 Frog adenovirus 1, complete genome
>NC_015225.1 Simian adenovirus 49, complete genome
>AC_000003.1 Canine adenovirus 1, complete genome
>NC_002685.2 Bovine adenovirus D, complete genome
>NC_038332.1 Fowl adenovirus 6 strain CR119, complete genome
>NC_038333.1 Bottlenose dolphin adenovirus 1 strain Tt11018, partial genome
>NC_038334.1 Porcine adenovirus 4 putative fiber protein gene, complete cds
>NC_039032.1 Psittacine aviadenovirus B isolate CS15-4016, complete genome

Приведите команды, которые выполнят следующие задачи:

1. Из файла viruses.fasta извлеките все строки, которые содержат слово "adenovirus" И подсчитайте их количество. Запишите результат (строки и их количество) в отдельный файл.
2. Выведите первые 10 и последние 5 строк из файла id.txt и найдите под какими номерами в файле viruses.fasta они находятся.

!grep 'adenovirus' vir.fna
>NC_006144.1 Simian adenovirus 3, complete genome
>AC_000010.1 Simian adenovirus 21, complete genome
>NC_001720.1 Fowl adenovirus A, complete genome
>NC_014899.1 Murine adenovirus 2, complete genome
>NC_016437.1 South polar skua adenovirus-1, complete genome
>NC_001813.1 Duck adenovirus A, complete genome
>NC_012959.1 Human adenovirus 54, complete genome
>NC_020487.1 Titi monkey adenovirus ECC-2011, complete genome
>AC_000020.1 Canine adenovirus type 2, complete genome
>AC_000017.1 Human adenovirus type 1, complete genome
>AC_000018.1 Human adenovirus type 7, complete genome
>AC_000019.1 Human adenovirus type 35, complete genome

>AC_000004.1 Duck adenovirus A, complete genome
>AC_000006.1 Human adenovirus D, complete genome
>AC_000009.1 Porcine adenovirus C, complete genome
>AC_000016.1 Turkey adenovirus A, complete genome
>AC_000011.1 Simian adenovirus 25, complete genome
>AC_000013.1 Fowl adenovirus D, complete genome
>AC_000014.1 Fowl adenovirus A, complete genome
>NC_011202.1 Human adenovirus B2, complete genome
>NC_015323.1 Fowl adenovirus C, complete genome
>AC_000001.1 Ovine adenovirus A, complete genome
>AC_000189.1 Porcine adenovirus 3, complete genome
>AC_000190.1 Tree shrew adenovirus 1, complete genome
>AC_000191.1 Bovine adenovirus A, complete genome
>NC_002501.1 Frog adenovirus 1, complete genome
>NC_015225.1 Simian adenovirus 49, complete genome
>AC_000003.1 Canine adenovirus 1, complete genome
>NC_002685.2 Bovine adenovirus D, complete genome
>NC_038332.1 Fowl adenovirus 6 strain CR119, complete genome
>NC_038333.1 Bottlenose dolphin adenovirus 1 strain Tt11018, partial genome
>NC_038334.1 Porcine adenovirus 4 putative fiber protein gene, complete cds
>NC_039032.1 Psittacine aviadenovirus B isolate CS15-4016, complete genome
>NC_034382.1 Cynomolgus adenovirus 1 isolate UK/UK-1/2004, complete genome

Приведите команды, которые выполнят следующие задачи:

1. Из файла viruses.fasta извлеките все строки, которые содержат слово "adenovirus" И подсчитайте их количество. Запишите результат (строки и их количество) в отдельный файл.
2. Выведите первые 10 и последние 5 строк из файла id.txt и найдите под какими номерами в файле viruses.fasta они находятся.

LMS-платформа − не предусмотрена

5.2.3. Домашняя работа № 2

Примерный перечень тем

1. Основы работы в ОС Linux. Создание директорий, пути абсолютные и относительные. Базовые команды

Примерные задания

Создайте ветку директорий в папке Documents следующего вида:

-- Bacteria
|
Documents -- Metagenomic -- Viruses -- viruses_hg
|
-- Contamination -- human_genome
|

-- adapters

Сортируйте файл list.viruses.txt по названиям вирусов (т.е.сортировка НЕ по иднетификаторам "NC_").

Из Файла Data.csv вывести только значения "Area" для которых ec_count равно 0.

Вывести 8, 37 и 101 последовательности из viral.3.1.genomic.fna (т.е. идентификатор последовательности + последовательность)

Добавить к каждому обязательному идентификатору (т.е. 1,5,9 и т.д. строки) файла miseq_reads_metagenome_fastp_1.fastq слово "ILLUMINA"

Определить количество нуклеотидов G в файле miseq_reads_metagenome_fastp_2.fastq

LMS-платформа − не предусмотрена

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Зачет

Список примерных вопросов

1. Методы и платформы (Illumina, ONT) секвенирования нуклеиновых кислот. 2. Структура адаптера для платформы Illumina. 3. Программы для контроля качества данных и тримминга. 4. Формат fastq 5. Выравнивание нуклеотидных последовательностей. Алгоритмы глобального и локального выравнивания (в общих чертах), выравнивание прочтений на геном. 6. Форматы sam/bam 7. Алгоритм и инструменты BLAST 8. Сборка генома de novo. Оценка качества сборки. 9. Классифицирование нуклеотидных последовательностей. Подходы на основе выравниваний и k-мерный анализ. 10. Особенности вирусной метагеномики.

LMS-платформа − не предусмотрена

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.