

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт биологии, экологии, почвоведения, сельского и лесного хозяйства (Биологический институт)

УТВЕРЖДЕНО:
Директор
Д. С. Воробьев

Оценочные материалы по дисциплине

Практикум по геномике

по направлению подготовки

06.04.01 Биология

Направленность (профиль) подготовки:
Генетика, геномика и синтетическая биология

Форма обучения
Очная

Квалификация
Магистр

Год приема
2024

СОГЛАСОВАНО:
Руководитель ОП
Г. Н. Артемов

Председатель УМК
А. Л. Борисенко

Томск – 2025

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен творчески использовать в профессиональной деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность программы магистратуры.

ОПК-5 Способен участвовать в создании и реализации новых технологий в сфере профессиональной деятельности и контроле их экологической безопасности с использованием живых объектов.

ОПК-6 Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.2 Демонстрирует понимание методологических основ дисциплин, определяющих направленность программы магистратуры

ИОПК-2.3 Использует фундаментальные знания, практические наработки и методический базис специальных дисциплин, определяющих направленность программы магистратуры, при планировании и реализации профессиональной деятельности

ИОПК-5.1 Понимает теоретические принципы и современный практический опыт использования биологических объектов в сфере профессиональной деятельности

ИОПК-6.1 Описывает разнообразие, пути и перспективы применения компьютерных технологий в современной биологии

ИОПК-6.2 Использует компьютерные технологии и профессиональные базы данных при планировании профессиональной деятельности, обосновывает их выбор

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- тест;
- реферат;
- практическое задание

Текущий контроль считается пройден, если студент набрал 80% от максимально возможной суммы баллов.

ИОПК-2.2, ИОПК-2.3

Тест с развернутым ответом:

1. Опишите формат хранения прочтений FASTQ, как формируются поля данного формата

Пример полного ответа:

Формат FASTQ используется для хранения последовательностей нуклеотидов и их качественных оценок (прочтений). Каждое прочтение в файле FASTQ представлено четырьмя строками:

1. Строка 1: начинается с символа "@" и содержит уникальный идентификатор прочтения (например, название секвенатора и координаты прочтения).
2. Строка 2: содержит последовательность нуклеотидов (A, T, C, G, N).
3. Строка 3: начинается с символа "+" и может повторять идентификатор или оставаться пустой.
4. Строка 4: содержит качественные оценки (ASCII-символы), соответствующие каждому нуклеотиду из строки 2, которые показывают вероятность ошибок в прочтении.

Формат сохраняет как последовательности, так и их качество в одном файле.

Критерии оценивания: полностью развернутый ответ оценивается в 2 балла. Частично правильный ответ оценивается в 1 балл. Неверный ответ оценивается в 0 баллов.

Пример частично правильного ответа:

Формат FASTQ используется для хранения данных секвенирования, в нем записаны как нуклеотиды каждого прочтения, так и их качество.

2. Как оценивается качество прочтений после секвенирования

Пример полного ответа:

Качество прочтений после секвенирования оценивается с использованием Phred-score, которые указывают вероятность ошибки в каждом нуклеотиде. Оценка преобразуется по формуле: $Q = -\log_{10}P$, где Q — Phred-score, а P — вероятность ошибки.

Для разных платформ оценка качества варьируется:

Illumina:

1. Оценка качества основана на интенсивности флуоресцентных сигналов. В процессе секвенирования к каждому нуклеотиду присоединяется флуоресцентный маркер, который испускает свет при прочтении. Камера фиксирует интенсивность сигнала на каждом цикле. Чем выше интенсивность, тем выше качество прочтения. Ошибки могут возникать из-за наложения сигналов или ослабления интенсивности флуоресценции.

PacBio (Pacific Biosciences):

2. Здесь используется технология SMRT (Single Molecule Real-Time), где качество прочтений зависит от точности измерения времени удержания (время инкорпорации) каждого нуклеотида ферментом ДНК-полимеразой. Флуоресцентные метки испускают свет при инкорпорации нуклеотида, и длительность этого сигнала помогает оценивать корректность прочтения. PacBio также использует круговые консенсусные прочтения (HiFi reads), которые многократно прочитывают один и тот же участок, что повышает точность.

3. Oxford Nanopore Technologies (ONT):

В ONT используется технология измерения изменений электрического тока, когда нуклеотиды проходят через нанопоры. Качество прочтений основано на моделировании изменений тока, связанных с каждым нуклеотидом. Ошибки могут возникать из-за шумов сигнала или ошибок в моделировании тока, поэтому качество может сильно варьироваться. Более высокая частота ошибок в длинных прочтениях компенсируется улучшением алгоритмов для анализа сигналов.

Критерии оценивания: полностью развернутый ответ оценивается в 4 балла.

Частично правильный ответ оценивается от 1 до 3 баллов в зависимости от степени освещения вопроса. Неверный ответ оценивается в 0 баллов.

ИОПК-5.1

Реферат по теме:

Использование биоинформатических инструментов в анализе геномных данных: теоретические основы и практические применения в современной науке.

Критерии оценивания: выполнение задания оценивается от 1 до 5 баллов.

ИОПК-6.1, ИОПК-6.2

Практическое задание:

На основе данных секвенирования пациента, болеющего гриппом, определите по какой причине пациент болен, если он прививался от гриппа в текущем сезоне
Последовательность действий, по которым будет производиться оценивание:

1. Скачать последовательности.
2. Провести оценку качества прочтений.
3. Вырезать низкокачественные последовательности и удалить адаптеры с прочтений.
4. Провести картирование прочтений на референсный геном.
5. Определить, какие замены наблюдаются у просеквенированного штамма.
6. Проанализировать, к чему могут привести такие замены. Сделать заключения по исследуемой работе.

Полностью выполненное задание по всем пунктам оценивается в 10 баллов. Частично правильное задание оценивается от 1 до 9 баллов. Полностью невыполненное задание оценивается в 0 баллов.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Промежуточная аттестация в форме зачета проводится во втором семестре на основе суммы баллов, которые студент получил за выполнение всех заданий, тестов, рефератов. Если студент сдал тесты и выполнил задания на общую сумму баллов, равную 80 % от максимально возможной суммы баллов, то он получает зачет:

Компетенция	Индикатор компетенции	Не зачтено	Зачтено
ОПК-2	ИОПК-2.2.	Менее 35 баллов	35 баллов и выше
	ИОПК-2.3.	Менее 35 баллов	35 баллов и выше
ОПК-5	ИОПК-5.1.	Менее 25 баллов	25 баллов и выше
ОПК-6	ИОПК-6.1.	Менее 50 баллов	50 баллов и выше
	ИОПК-6.2.	Менее 50 баллов	50 баллов и выше
Итого		Менее 195 баллов	195 баллов и выше

Если набрано меньше 80% баллов, студент сдает устный зачет по билетам. Каждый билет содержит 2 вопроса из перечня:

ИОПК-2.2 Демонстрирует понимание методологических основ дисциплин, определяющих направленность программы магистратуры;

1. Направления исследований в прикладной геномике.
2. Основные технологии секвенирования
3. Место геномики в современной науке
4. Форматы fasta и fastq
5. Форматы SAM, BAM.

6. Скриптовые языки программирования. Как применяются в геномике, какие наиболее часто используются и при решении каких задач.
7. Основные характеристики языка программирования Python.
8. Основные характеристики языка программирования R.
9. Язык Bash и основы работы в командной строке.

ИОПК-2.3 Использует фундаментальные знания, практические наработки и методический базис специальных дисциплин, определяющих направленность программы магистратуры, при планировании и реализации профессиональной деятельности;

10. Оценка качества прочтений. Программы для оценки качества прочтений.
11. Выравнивание последовательностей. Парное и множественное выравнивание.
12. Выравнивание последовательностей. Глобальное и локальное выравнивание.
13. Отличие картирования прочтений от выравнивания прочтений.
14. Картирование РНК-секвенирования на референсный геном, особенности картирования.
15. Программы для работы с SAM, BAM.
16. Поиск геномных вариантов. Основные направления изучения.
17. Классификация геномных вариантов.

ИОПК-5.1 Понимает теоретические принципы и современный практический опыт использования биологических объектов в сфере профессиональной деятельности

18. Сборка генома из коротких прочтений.
19. Контиги, скэффолды. Преимущество парноконцевых прочтений при сборке генома.
20. Сборка de novo.
21. Сборка уже собранного генома. В каких задачах имеется такая необходимость.
22. Графы де Брюйна, особенности применения.
23. Преимущества и недостатки сборки генома длинными прочтениями.
24. Гибридная сборка генома. Метод Hi-C в приложении сборки генома.

ИОПК-6.1 Описывает разнообразие, пути и перспективы применения компьютерных технологий в современной биологии;

ИОПК-6.2 Использует компьютерные технологии и профессиональные базы данных при планировании профессиональной деятельности, обосновывает их выбор;

25. Формат VCF, программы для работы с VCF.
26. Метагеномика. Приложение и методы.
27. Метагеномный анализ на основе 16S рРНК.
28. Статистический анализ метагеномных данных.
29. Анализ дифференциальной экспрессии генов. Основы.
30. Секвенирование одиночных клеток. Приложение, преимущества и недостатки.
31. Машинное обучение в биологии. Общий обзор.

Критерии оценивания: результаты определяются оценками «зачтено», «не зачтено». Оценка «зачтено» выставляется, если даны исчерпывающие ответы на оба вопроса в билете. Оценка «не зачтено» выставляется, если студент не может дать ответ на оба вопроса в билете.

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

ИОПК-2.2; ИОПК-2.3

- 1) Чем обусловлен бурный рост геномики в XX веке.
- 2) Опишите основные шаги в процессе секвенирования ДНК.
- 3) Перечислите задачи, которые решает геномика.
- 4) Сравните преимущества и недостатки различных методов секвенирования (Sanger, Illumina, PacBio, ONT).
- 5) Назовите основные форматы данных в биоинформатике.
- 6) Назовите основные направления геномных исследований.

ИОПК-5.1

- 1) Отличие формата fasta от fastq:
 - а) fasta содержит информацию о качестве
 - б) в fasta 4 строки для каждого прочтения
 - в) fastq начинается с ">"
 - г) fasta не содержит информацию о качестве прочтений
- 2) Основное преимущество парноконцевых прочтений (paired-end) при сборке генома:
 - а) Увеличение длины отдельных прочтений
 - б) Повышение точности определения последовательности
 - в) Возможность выравнивания коротких фрагментов
 - г) Помогают лучше определять структурные варианты
- 3) Картирование РНК-секвенирования на референсный геном отличается от обычного картирования ДНК тем, что:
 - а) Необходимо учитывать экзоны и интроны
 - б) Требуется больше данных для картирования
 - в) Используются специальные секвенаторы
 - г) Оно проводится исключительно для прокариот
- 4) Основная задача форматов SAM/BAM:
 - а) Хранение последовательностей нуклеотидов
 - б) Хранение выровненных на референсный геном прочтений
 - в) Хранение последовательностей белков
 - г) Анализ филогенетических данных
- 5) Какое из перечисленных является примером метода для оценки качества прочтений?
 - а) FastQC
 - б) BWA
 - в) IGV
 - г) BLAST

- 6) Что делает программа BWA?
- а) Визуализирует результаты секвенирования
 - б) Проводит выравнивание последовательностей на референсный геном
 - в) Проводит сборку генома de novo
 - г) Анализирует статистику геномных данных
- 7) Чем отличаются графы де Брюйна от других методов сборки генома?
- а) Используются для анализа длинных прочтений
 - б) Составляют цепочки на основе частично перекрывающихся фрагментов
 - в) Сравнивают прочтения с готовым референсным геномом
 - г) Изначально применялись только для белков
- 8) В чем заключается основная идея метагеномного анализа?
- а) Сборка генома из отдельных прочтений
 - б) Анализ геномных данных одного организма
 - в) Исследование и анализ генетического материала из смешанных сообществ
 - г) Выявление мутаций в геноме конкретного организма

Ключи: 1 г); 2 г); 3 а); 4 б); 5 а); 6 б); 7 б); 8 в).

ИОПК-6.1, ИОПК-6.2

- 1) Выровняйте прочтения Illumina на референсный геном и визуализируйте файл выравнивания в IGV.

Используйте программу BWA для выравнивания, а затем откройте файл выравнивания (SAM/BAM) в программе IGV для визуализации.

- 2) Оцените качество прочтений в файле FASTQ с помощью FastQC.

Проанализируйте результат и определите, есть ли необходимость в фильтрации низкокачественных прочтений.

- 3) Соберите геном de novo из коротких прочтений с помощью программы SPAdes.

Проверьте наличие контигов и скэффолдов в собранном геноме.

- 4) Выполните картирование прочтений РНК-секвенирования на референсный геном с использованием программы STAR.

Проверьте корректность выравнивания и получите информацию о выраженности генов.

Информация о разработчиках

Кириленко Кирилл Михайлович, ассистент каф. генетики и клеточной биологии