

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

Высшая инженерная школа агробιοтехнологий

Оценочные материалы по дисциплине

Практикум по биоинформатике

по специальности

36.05.01 Ветеринария

Специализация:

Ветеринария

Форма обучения

Очная

Квалификация

Ветеринарный врач

Год приема

2021

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

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 Способен использовать в профессиональной деятельности методы решения задач с использованием современного оборудования при разработке новых технологий и использовать современную профессиональную методологию для проведения экспериментальных исследований и интерпретации их результатов.

ОПК-7 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-4.3 Обосновывает применение специализированного оборудования для реализации поставленных задач при проведении исследований и разработке новых технологий.

ИОПК-7.1 Понимает особенности работы современных информационных технологий

ИОПК-7.2 Использует современные информационные технологии для решения задач профессиональной деятельности

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- практические задания
- устный опрос

ИОПК-4.3

1. Дайте определение биоинформатики.
2. Какую дату можно считать датой выделения биоинформатики в отдельную научную область, и почему?
3. Почему до настоящего времени нет общепринятого определения биоинформатики?
4. В чём состоит особенность биоинформационных данных по сравнению с другими (не биологическими) данными?
5. Что такое секвенирование и какую роль играет секвенирование в биоинформатике?
6. Где хранятся биоинформационные данные?
7. Какие три компонента включает в себя предмет биоинформатики?
8. Каковы цели биоинформатики?
9. Какие задачи стоят перед биоинформатикой?
10. В каких видах деятельности реализуется предмет биоинформатики?
11. Какие два различных аналитических направления существуют в расшифровке биологических последовательностей?
12. Какую роль играет анализ гомологических последовательностей в расшифровке биологической информации?
13. Каким образом биоинформатика способствует разработке лекарственных препаратов?
14. Перечислите медицинские применения биоинформатики.
15. Какие области биоинформатики нуждаются в защите интеллектуальной собственности?
16. Какие методы геномики и протеомики нуждаются в патентной защите, и почему?

ИОПК-7.1

17. В чём сходство и различие IP-адреса компьютера и его текстового имени в иерархической системе доменных имён?
18. Как записывается последовательность белка в формате FASTA?
19. Что такое GI-номер?
20. Что такое EMBnet и какой браузер используется в этой сети?
21. Что такое SRS (Sequence Retrieval System) и в какой сети она используется?
22. Что такое NCBI и какой браузер используется для поиска в сети баз данных NCBI?
23. Для чего используется Entrez?
24. Что такое контиг?
25. В чём состоит парадоксальное различие протекания процессов трансляции и фолдинга белка?
26. Анализом каких трёх информационно-управляемых процессов преимущественно занимается биоинформатика?

ИОПК-7.2

27. Что называется выравниванием биологических последовательностей?
28. Какие три типа изменений происходят в процессе эволюционного расхождения последовательностей от общего предка?
29. Что такое глобальное выравнивание последовательностей?
30. Что такое локальное выравнивание последовательностей?
31. В чём заключается поиск мотивов совпадения при выравнивании последовательностей?
32. Что такое множественное выравнивание?
33. Что такое точечная матрица сходства и для чего её используют?
34. Как проявляются инверсные и палиндромные последовательности в точечных матрицах сходства?
35. Как проявляются инсерции и делеции в точечных матрицах сходства?
36. Почему с точки зрения биолога поиск локального подобия может дать более значимые и точные результаты, чем оценка выравнивания по всей длине последовательностей?

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

100-80 баллов–Дан полный, развернутый ответ на поставленный вопрос, показана совокупность осознанных знаний об объекте, проявляющаяся в свободном оперировании понятиями.

80-60 - Дан полный, развернутый ответ на поставленный вопрос, доказательно раскрыты основные положения темы; в ответе прослеживается четкая структура, логическая последовательность

60-40 Дан неполный ответ, представляющий собой разрозненные знания по теме вопроса с существенными ошибками в определениях.

Зачет проставляется при наборе 60 баллов и выше.

ИОПК-7.1

1. Для чего предназначена программа BLAST?
2. Сформулируйте определение понятия "информация", которое наиболее адекватно для биологических применений.

3. Почему не существует единого для всех наук определения понятия "информация"?
4. Запишите формулу Шеннона и объясните смысл входящих в неё параметров.
5. Проиллюстрируйте отличие между понятиями "количество информации" и "ценность информации", используя какое-либо предложение.
6. Что такое фиксируемость информации?
7. Что такое инвариантность информации?
8. Что такое брэнность информации?
9. Что такое изменчивость информации?
10. Что такое транслируемость информации?
11. Что такое размножаемость информации?
12. Что такое мультипликативность информации?
13. Что такое действенность информации?
14. Что такое оператор, порождаемый информацией?
15. Что такое семантика информации?
16. Что такое полипотентность информации?
17. Что такое полезность информации?
18. Что такое истинность информации?

ИОПК-4.3

19. Что такое ценность информации? В чём она проявляется? Как определяется мера ценности информации?
20. Что такое рецепция информации? Какие выделяют два способа рецепции?
21. Каким образом происходит запоминание информации?
22. Чем различаются макроинформация и микроинформация?
23. Оцените количество информации, содержащееся в гиперцикле биосинтеза белка.
24. В каких случаях проводится секвенирование биологических последовательностей?

ИОПК-7.2

25. Чем занимается структурная геномика?
26. Чем занимается функциональная геномика?
27. Чем занимается сравнительная геномика?
28. Что такое генетическая карта?
29. Что такое генетический маркер?
30. Какие четыре типа ДНК-маркеров используются при составлении карты генома человека?

ИОПК-4.3 Обосновывает применение специализированного оборудования для реализации поставленных задач при проведении исследований и разработке новых технологий.

ИОПК-7.1 Понимает особенности работы современных информационных технологий

ИОПК-7.2 Использует современные информационные технологии для решения задач профессиональной деятельности

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

ИОПК-4.3

1. Какой алгоритм используется для локального попарного выравнивания последовательностей?

- A) Needleman-Wunsch
- B) Smith-Waterman
- C) ClustalW

D) UPGMA

Ответ: B

2. Какая база данных НЕ является первичной (primary database) нуклеотидных последовательностей?

- A) GenBank
- B) EMBL
- C) DDBJ
- D) PROSITE

Ответ: D (PROSITE — база белковых доменов и семейств, вторичная база)

3. Какое значение E-value указывает на наиболее статистически значимую находку при поиске в BLAST?

- A) E = 5
- B) E = 0,11
- C) E = 0,02
- D) E = 1e-5

Ответ: D

4. Какой формат файла используется для хранения данных секвенирования вместе с качеством каждого нуклеотида (Phred score)?

- A) FASTA
- B) FASTQ
- C) GenBank flatfile
- D) PDB

Ответ: B

5. Какая программа используется для предсказания белок-кодирующих генов в бактериальном геноме?

- A) BLAST
- B) Prodigal
- C) Barrnap
- D) ClustalW

Ответ: B

ИОПК-7.1

6. Две белковые последовательности являются гомологичными, если:

- A) они имеют не менее 50% идентичных аминокислотных остатков
- B) они выполняют одинаковую функцию в клетке
- C) они произошли от общего предка
- D) их третичные структуры похожи больше, чем первичные

Ответ: С

7. Какой инструмент используется для поиска рибосомальных РНК (например, 16S rRNA) в геноме?

- A) Prodigal
- B) BLAST
- C) Barrnap
- D) HMMER

Ответ: С

8. Что обозначает аббревиатура SNP?

- A) Single Nucleotide Polymorphism
- B) Short Nucleotide Pattern
- C) Sequence Number Parameter
- D) Standard Nucleotide Position

Ответ: А

9. Какая база данных содержит трехмерные структуры биологических макромолекул?

- A) PubMed
- B) OMIM
- C) UniProt
- D) PDB (Protein Data Bank)

Ответ: D

10. Какой параметр рассчитывается по формуле $Q = -10 \times \log_{10}(p)$, где p — вероятность ошибки при определении нуклеотида?

- A) E-value
- B) Bit-score
- C) Phred quality score
- D) Coverage

Ответ: С

ИОПК-7.2

11. Для чего используется программа BLAST?

- A) Для сборки генома de novo
- B) Для поиска гомологичных последовательностей в базе данных
- C) Для визуализации трехмерных структур белков
- D) Для предсказания вторичной структуры РНК

Ответ: В

12. Какой метод секвенирования относится к первому поколению (исторически первый)?

- A) Секвенирование по Сангеру
- B) Секвенирование на платформе Illumina
- C) Пиросеквенирование (Roche 454)
- D) Секвенирование нанопорами (Oxford Nanopore)

Ответ: А

13. Какое расширение имеют файлы, содержащие информацию о выравнивании прочтений на референсный геном в бинарном формате?

- A) .sam
- B) .bam
- C) .vcf
- D) .bed

Ответ: В

14. Что такое контиг (contig) в сборке генома?

- A) Одиночное прочтение (рид) с секвенатора
- B) Непрерывная последовательность, собранная из перекрывающихся ридов
- C) Показатель качества сборки генома
- D) Участок генома, покрытый ридами

Ответ: В

15. Какая разновидность BLAST используется для поиска белка, транслируемого с нуклеотидной последовательности, по базе данных нуклеотидных последовательностей (с динамической трансляцией базы)?

- A) BLASTP
- B) BLASTN
- C) TBLASTN
- D) BLASTX

Ответ: С

ИОПК-4.3

16. Соотнесите формат файла (1-4) с его основным назначением (А-Г).

1.FASTA	А. Хранение данных секвенирования с качеством нуклеотидов
2.FASTQ	Б. Хранение последовательности в простом текстовом формате (строка ">")
3.SAM	В. Хранение информации о вариантах генома (SNP, INDEL)
4.VCF	Г. Хранение выравнивания прочтений на референс (текстовый формат)

Ответ: 1 → Б 2 → А 3 → Г 4 → В

17. Установите соответствие между типом BLAST (1-4) и его описанием (А-Г).

1.BLASTN	А. Белок ищет по базе белков
2.BLASTP	Б. Нуклеотидная последовательность ищет по базе нуклеотидов
3.BLASTX	В. Белок ищет по базе нуклеотидов (с трансляцией базы)
4.TBLASTN	Г. Нуклеотидная последовательность транслируется в белок и ищет по базе белков

Ответ:

1 → Б 2 → А 3 → Г 4 → В

18. Соотнесите базу данных (1-4) с типом информации (А-Г).

1.PDB	А. Белковые последовательности (аннотированные)
2. UniProtKB/Swiss-Prot	Б. Трёхмерные структуры белков
3. Pfam	В. Медико-биологическая литература
4. PubMed	Г. Семейства белков (HMM-профили)

Ответ: 1 → Б 2 → А 3 → Г 4 → В

ИОПК-7.1

19. Установите соответствие между параметром качества (1-4) и его характеристикой (А-Г).

1. Phred score (Q20)	А. Вероятность ошибки 1:1000 (точность 99,9%)
2. Phred score (Q30)	Б. Вероятность ошибки 1:100 (точность 99%)
3. Coverage (глубина)	В. Доля генома, собранная в контиги
4. N50	Г. Среднее число прочтений, покрывающих нуклеотид

Ответ:

1 → Б 2 → А 3 → Г 4 → В

20. Соотнесите программу (1-4) с её функцией (А-Г).

1. FastQC	А. Сборка генома de novo
2. Trimmomatic	Б. Оценка качества ридов
3. SPAdes	В. Аннотация бактериального генома
4. Prokka	Г. Обрезка адаптеров и фильтрация ридов

Ответ:

1 → Б 2 → Г 3 → А 4 → В

21. Как называется база данных, в которой хранятся трёхмерные структуры биологических макромолекул (белков и нуклеиновых кислот)?

Ответ: PDB (Protein Data Bank)

22. Какая разновидность BLAST используется для поиска белковой последовательности по базе данных нуклеотидных последовательностей (с динамической трансляцией базы)?

Ответ: TBLASTN

ИОПК-7.2

23. Как называется метрика качества сборки генома, представляющая собой длину, при которой сумма длин всех контигов этой длины и больше достигает 50% от общего размера сборки?

Ответ: N50

24. Какой параметр в выдаче BLAST показывает ожидаемое количество случайных совпадений с таким же или лучшим скором?

Ответ: E-value

25. Какой тип РНК является компонентом рибосом и используется в качестве филогенетического маркера для идентификации бактерий?

Ответ: 16S или рРНК,

Информация о разработчиках

Бойко Екатерина Владимировна, к.б.н., доцент кафедры биотехнологии и биоинформатики ВИША ТГУ