

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

Высшая инженерная школа агробιοтехнологий

Оценочные материалы по дисциплине

Введение в биоинформатику

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.

УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий.

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

ИУК-1.1 Применяет алгоритмы анализа задач, выделяя их базовые составляющие

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

Темы практических работ

ИУК-1.1

1. Средства работы с базами данных I (Entrez).
2. Средства работы с базами данных II (SRS).
3. Сервис GeneVee. Основные поля записи SwissProt.
4. Поиск гомологов (интерпретация результатов, сравнение алгоритмов, зависимость от параметров).
5. Построение выравниваний, реконструкция филогенетических деревьев (сравнение локальных и глобальных выравниваний, зависимость выравнивания от параметров, оценка статистической значимости).
6. Работа с базой пространственных структур PDB.

ИОПК-4.3

7. Структуры белков (RASMOL, SwissPDBViewer). Работа с программой визуализации макромолекул RasMol I.

8. Аннотирование последовательности (поиск белок-кодирующих областей, поиск функциональных сайтов).

ИОПК-7.1

9. Работа с программой визуализации макромолекул RasMol II.
10. Поиск слабых сигналов в биологических последовательностях. Интернет-ресурсы работы с полными геномами.
11. Вторичные структуры РНК.
12. Предсказание структурных особенностей белков

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

Работы оцениваются максимум в 5 баллов, зачет проставляется в случае набора 2 и более баллов.

- 4 – 5 баллов ставится в том случае, если:
работа выполнена в полном объеме с соблюдением необходимой последовательности;
- отчет правильно оформлен, заполнены все таблицы протоколов, правильно проведены вычисления;
 - качественно и грамотно сделаны выводы.
- 2 – 3 балла ставится в том случае, если выполнены предыдущие требования, но:
- было допущено два-три недочета при выполнении расчетов;
 - формулировка выводов не полностью соответствует поставленным задачам;
- 1 балл ставится, если:
- работа выполнена не полностью, и объем выполненной части позволяет получить правильные выводы;
 - в отчете были допущены ошибки (в записях единиц, измерениях, в вычислениях, графиках, таблицах и т. д.), повлиявшие на результат выполнения работы.
 - возникли затруднения с формулировкой выводов;
- 0 баллов ставится в том случае, если:
- работа выполнена не полностью, и объем выполненной части работы не позволяет сделать правильных выводов;
 - опыты, измерения, вычисления, наблюдения производились неправильно;
- работа не выполнена

Примеры вопросов и заданий:

- 4.3
- Биоинформатика и биоинженерия: возникновение, цели, задачи, методы. ИОПК-
 - Базы данных: классификация, основы структур. ИУК-1.1
 - Базы данных белковых последовательностей ИОПК-7.1
 - Банки данных метаболических путей. ИОПК-7.1

В расширенном режиме выполните поиск записей, соответствующих гемоглобину человека. В режиме обычного поиска найдите 4 белка – альфа- и бета-глобины человека и быка. ИОПК-7.1

С помощью серии парных выравниваний выясните, какие белки более сходны – одинаковые субъединицы разных организмов или разные субъединицы одного организма. ИОПК-4.3

Используя систему поиска Entrez оцените количество белков человека, которые имеют массу более 300 000 дальтон. ИУК-1.1

АТР-binding cassette (ABC) – пример белкового домена в широко-представленном семействе ABC transporters. Используя систему поиска Entrez оцените количество белков человека, которые имеют такой домен? Сколько белков бактерий имеют такой домен? ИОПК-7.1

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

ИОПК-4.3

1. Основной задачей биоинформатики как научной дисциплины является:
 - а) Проведение полевых исследований биологического разнообразия
 - б) Разработка математических методов и компьютерных технологий для анализа биологических данных
 - в) Создание нового лабораторного оборудования для секвенирования
 - г) Изучение анатомии и морфологии многоклеточных организмов

2. Согласно центральной догме молекулярной биологии, передача генетической информации осуществляется по схеме:
 - а) Белок → РНК → ДНК
 - б) ДНК → белок → РНК
 - в) ДНК → РНК → белок
 - г) РНК → ДНК → белок

3. Для эффективного поиска специализированной научной информации в базах данных наименее эффективно использование:
 - а) Поисковых операторов (логических AND, OR, NOT)
 - б) Поиска по ключевым словам в определенных полях (название, автор, аннотация)
 - в) Полнотекстового поиска по всем статьям без ограничений
 - г) Ограничения поиска по дате публикации или типу публикации

ИОПК-7.1

4. К специализированным базам данных, содержащим информацию о нуклеотидных последовательностях, относится:
 - а) Scopus
 - б) GenBank
 - в) PubMed
 - г) Mendeley

5. Математический алгоритм, лежащий в основе попарного выравнивания биологических последовательностей (например, алгоритм Нидлмана-Вунша), — это:
 - а) Машинное обучение
 - б) Динамическое программирование
 - в) Байесовский вывод
 - г) Кластерный анализ

6. Процесс предсказания пространственной трехмерной структуры белка на основе его аминокислотной последовательности наиболее точно характеризует термин:
 - а) Секвенирование
 - б) Фолдинг
 - в) Моделирование
 - г) Денатурация

7. Программное обеспечение, предназначенное для хранения, организации и оформления библиографических ссылок, называется:
 - а) Операционная система
 - б) Текстовый процессор
 - в) Система управления библиографической информацией (менеджер цитирования)
 - г) База данных метаболических путей

Ключи к тесту: 1 б), 2 в), 3 в), 4 б), 5 б), 6 в), 7 в)

ИОПК-4.3

8. Основной задачей биоинформатики является:
- а) Проведение полевых исследований биологического разнообразия.
 - б) Разработка математических методов и компьютерных технологий для анализа биологических данных.
 - в) Создание нового лабораторного оборудования для секвенирования.
 - г) Изучение анатомии и морфологии многоклеточных организмов.
9. Для поиска информации о нуклеотидных последовательностях используется:
- а) Баз данных научных статей Scopus и Web of Science.
 - б) Текстовый редактор.
 - в) Специализированные геномные базы данных, таких как GenBank.
 - г) Социальные сети и мессенджеры.
10. Для попарного выравнивания биологических последовательностей алгоритма используется: (ИОПК-8.2)
- а) Машинное обучение.
 - б) Динамическое программирование.
 - в) Байесовский вывод.
 - г) Кластерный анализ.

ИУК-1.1

11. Для установления эволюционных взаимоотношений между генами используется: (ИПК-2.3)
- а) Проведение химического синтеза новых соединений.
 - б) Построение и анализ филогенетических деревьев.
 - в) Измерение оптической плотности растворов.
 - г) Взвешивание биологических образцов.
12. Для организации библиографических ссылок используется: (ИПК-2.4)
- а) Графический редактор.
 - б) Система управления базами данных (СУБД).
 - в) Система управления библиографической информацией (менеджеров цитирования).
 - г) Электронные таблицы.
13. Для поиска информации о трехмерной структуре белков предназначена база данных: (ИОПК-6.2)
- а) GenBank
 - б) UniProt
 - в) PDB (Protein Data Bank)
 - г) PubMed
- Ключи: 8 б), 9 в), 10 б), 11 б), 12 в), 13 в)

ИУК-1.1

14. Раскройте суть центральной догмы молекулярной биологии. Объясните, как вычислительные методы биоинформатики используются для изучения каждого из этапов реализации генетической информации.

Ответ должен содержать описание потоков информации: репликация, транскрипция, трансляция,

15. Соотнесите инструмент (1-4) с его ключевой характеристикой или алгоритмом (А-Д).
ИОПК-4.3

1.BLAST	А. Использует позиционно-специфические матрицы (PSSM) для поиска удаленных гомологов
2.FASTA	Б. Эвристический метод, использующий разбиение запроса на короткие слова (k-tuple) для поиска совпадений
3.PSI-BLAST	В. Метод, основанный на поиске коротких совпадений (seed), которые затем расширяются
4.HMMER	Г. Использует скрытые марковские модели (HMM) для поиска гомологии

Ответ:

1 → **В**, 2 → **Б**, 3 → **А**, 4 → **Г**

16. Установите соответствие между термином, описывающим тип гомологии генов (1-4), и его определением (А-Д). ИОПК-7.1

1.Ортологи	А. Гены, возникшие в результате дупликации в пределах одного генома
2.Паралоги	Б. Гены у разных видов, произошедшие от одного общего предкового гена в результате видообразования
3.Ксенологи	В. Гомологичные гены, приобретенные горизонтальным переносом
4.Гомологи	Г. Общий термин для генов, имеющих общее эволюционное происхождение

Ответ:

1 → **Б** 2 → **А**, 3 → **В** 4 → **Г**

Вопрос 3.

17. Соотнесите название базы данных (1-4) с её основным содержимым (А-Д). ИУК-1.1

1.GenBank / EMBL / DDBJ	А. База данных трехмерных структур белков
2.UniProtKB/Swiss-Prot	Б. Аннотированная база данных белковых последовательностей с высокой степенью достоверности
3.PDB	В. Интегрированная база данных белковых семейств, доменов и сайтов
4.InterPro	Г. Первичный архив нуклеотидных последовательностей

Ответ:

1 → **Г** 2 → **Б** 3 → **А** 4 → **В**

Вопрос 4.

18. Установите соответствие между форматом файла (1-4) и его типом или применением (А-Д). ИОПК-4.3

1.FASTA	А. Стандартный формат для хранения и аннотации последовательностей, использующий двухбуквенные коды (ID, AC, FT и др.)
---------	--

2.EMBL / GenBank flatfile	Б. Простой текстовый формат, где первая строка начинается с символа ">" и содержит описание, а последующие — саму последовательность
3.PDB	В. Формат файла для выравнивания последовательностей, показывающий совпадения, замены и делеции
4.ALN/FASTA alignment	Г. Специализированный формат для координат атомов трехмерных структур

Ответ:

1 → **Б** 2 → **А** 3 → **Г** 4 → **В**

Вопрос 5.

19. Соотнесите тип матрицы замен (1-4) с её характеристикой (А-Д). ИОПК-7.1

1.PAM	А. Семейство матриц, полученное путем наблюдения за заменами в блоках консервативных участков без глобального выравнивания
2.BLOSUM	Б. Единичная матрица, присваивающая 1 за совпадение и 0 за несовпадение
3.Unitary matrix	В. Матрица, полученная путем моделирования эволюционного процесса на основе точечных мутаций
4.BLOSUM 62	Г. Матрица по умолчанию для поиска BLAST, созданная на основе кластеризации последовательностей с идентичностью $\geq 62\%$

Ответ:

1 → **В** 2 → **А** 3 → **Б** 4 → **Г**

Вопрос 6.

20. Установите соответствие между задачей в биоинформатике (1-4) и наиболее подходящим методом или технологией (А-Д). ИУК-1.1

1.Предсказание трехмерной укладки белка при отсутствии гомологичной структуры	А. Сборка генома de novo
2.Поиск удаленно родственных белков (remote homology)	Б. Трединг (Threading) или фолдинг (розетка)
3.Реконструкция последовательности генома без референса	В. Анализ SNP-чипов или секвенирование экзона
4.Идентификация однонуклеотидных полиморфизмов (SNPs), связанных с заболеванием	Г. PSI-BLAST или HMMER

Ответ:

1 → **Б**, 2 → **Г**, 3 → **А**, 4 → **В**

ИУК-1.1

21. Какой тип РНК является конечным продуктом экспрессии белок-кодирующих генов и служит матрицей для синтеза белка?

Ответ: мРНК

22. Как называется свойство последовательностей, имеющих общего эволюционного предка?

Ответ: Гомология

23. Какой аббревиатурой обозначается самый популярный инструмент для поиска локального выравнивания последовательностей?

Ответ: BLAST

24. Как называется статистический показатель в выдаче BLAST, который отражает ожидаемое число случайных совпадений?

Ответ: E-value

ИОПК-7.1

25. Как называется основной формат файла базы данных GenBank для хранения аннотированных последовательностей?

Ответ: Flatfile

26. Какая матрица замен (PAM или BLOSUM) считается более современной и лучше подходит для поиска гомологов в BLAST?

Ответ: BLOSUM

27. Как называется сборка генома без использования референсной последовательности?

Ответ: Denovo

ИОПК-4.3

28. Какой старт-кодон (триплет) чаще всего инициирует трансляцию у прокариот и эукариот?

Ответ: ATG

29. Как называется база данных, в которой хранятся трехмерные структуры биологических молекул?

Ответ: PDB

30. Как называются гены у разных видов, произошедшие от общего предкового гена в результате видообразования?

Ответ: Ортологи

Информация о разработчиках

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