

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

Введение в биоинформатику

по направлению подготовки

06.04.01 Биология

Направленность (профиль) подготовки:

**Физиология, биохимия, биотехнология и биоинформатика растений и
микроорганизмов**

Форма обучения
Очная

Квалификация
Магистр

Год приема
2024

СОГЛАСОВАНО:
Руководитель ОП
О.В. Карначук

Председатель УМК
А.Л. Борисенко

Томск – 2024

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

Целью освоения дисциплины является формирование следующих компетенций:

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

ОПК-8 Способен использовать современную исследовательскую аппаратуру и вычислительную технику для решения инновационных задач в профессиональной деятельности.

ПК-2 Способен проводить основные этапы полевых и лабораторных исследований в соответствии с профилем (направленностью) магистерской программы.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

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

ИОПК-8.2 Применяет современную исследовательскую аппаратуру и вычислительную технику при решении стандартных и инновационных задач в профессиональной деятельности

ИПК-2.3 Получает научно значимые результаты при использовании полевых и лабораторных методов исследования биологических объектов, в том числе применяя современную аппаратуру и оборудование

ИПК-2.4 Описывает, обобщает и делает выводы на основе результатов исследования, в том числе с помощью современных компьютерных технологий

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- тесты;
- контрольная работа;

Тест (ИОПК-6.1; ИОПК-6.2; ИОПК-8.2; ИПК-2.3; ИПК-2.4).

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

4. К специализированным базам данных, содержащим информацию о нуклеотидных последовательностях, относится:
- a) Scopus
 - б) GenBank
 - в) PubMed
 - г) Mendeley
5. Математический алгоритм, лежащий в основе попарного выравнивания биологических последовательностей (например, алгоритм Нидлмана-Вунша), — это:
- a) 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:
- a) Секвенирование
 - б) Фолдинг
 - в) Моделирование
 - г) Денатурация
7. Программное обеспечение, предназначенное для хранения, организации и оформления библиографических ссылок, называется:
- a) 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)

Критерии оценивания: тест считается пройденным, если обучающий ответил правильно как минимум на половину вопросов.

Контрольная работа (ИОПК-6.1, ИОПК-6.2, ИОПК-8.2, ИПК-2.3, ИПК-2.4)
Контрольная работа состоит из 2 теоретических вопросов и 1 задачи.

Перечень теоретических вопросов:

1. История, основные задачи и области применения биоинформатики в современной биологии. Приведите примеры профессиональных баз данных.
2. Суть центральной догмы молекулярной биологии. Объясните, что изучают геномика и протеомика, и какое отношение они имеют к биоинформатике.
3. Принципы организации и индексирования научной информации. Опишите назначение и различия между базами данных научной литературы (например, Scopus, Web of Science) и специализированными биологическими базами данных (например, GenBank, UniProt).
4. Алгоритмы выравнивания биологических последовательностей. Что такое динамическое программирование и как оно применяется для поиска гомологичных последовательностей?
5. Основные этапы и методы молекулярно-филогенетического анализа. Какова роль биоинформатических программ в построении и интерпретации филогенетических деревьев?
6. Проблема фолдинга белка. Какие вычислительные методы и базы данных используются для предсказания и анализа трехмерной структуры белков?

7. Назначение и функциональные возможности систем управления библиографической информацией (менеджеров цитирования). Как они интегрируются в процесс научного исследования?

Задачи:

Задача 1. Исследователю необходимо найти все научные публикации, посвященные филогенетическому анализу вируса гриппа H1N1, опубликованные за последние 5 лет. Опишите пошаговый алгоритм его действий с использованием баз данных научной литературы (например, PubMed) и менеджера цитирования. Какие поисковые операторы и фильтры следует применить для максимально релевантного результата?

Задача 2. Вам предоставлены две нуклеотидные последовательности:

Последовательность 1: ---ATGCTAGCT

Последовательность 2: ATGCGAGCT---

Выровняйте последовательности вручную, определив оптимальное выравнивание с учетом одного возможного пропуска (делеции/инсерции). Определите тип и количество найденных нуклеотидных замен. Укажите, в какую базу данных Вы бы депонировали последовательности для дальнейшего сравнительного анализа.

Задача 3. При анализе нуклеотидной последовательности гена, кодирующего важный фермент у бактерии *E. coli*, в базе данных UniProt было найдено несколько десятков его гомологов у других видов. Сформулируйте план действий для установления эволюционных взаимоотношений между этими последовательностями. Перечислите основные этапы работы: от множественного выравнивания последовательностей до построения и интерпретации филогенетического дерева, указав, какое программное обеспечение можно использовать на каждом этапе.

Задача 4. Для белка с известной аминокислотной последовательностью, но неизвестной структурой необходимо предсказать его третичную структуру. Перечислите доступные вычислительные подходы (от гомологичного моделирования до методов *ab initio*), ранжируя их по ожидаемой точности и требовательности к вычислительным ресурсам. В каких базах данных структур белков следует искать шаблоны для моделирования?

Ответы к задачам:

Задача 1. Алгоритм: 1) Выбор БД (PubMed); 2) Поисковый запрос: "Influenza A Virus, H1N1 Subtype"[Mesh] AND phylogenetic analysis; 3) Применение фильтров: дата публикации (5 лет), тип статьи (исследование, обзор); 4) Экспорт результатов в менеджер цитирования (например, Zotero, Mendeley) для организации списка литературы.

Задача 2. Оптимальное выравнивание:

Последовательность 1: ATGCTAGCT

Последовательность 2: ATGCGAGCT

Обнаружена одна транзиция (замена пиримидина на пиримидин): T → G в пятой позиции. Для депонирования и сравнительного анализа следует использовать базы данных нуклеотидных последовательностей, такие как GenBank или ENA (European Nucleotide Archive).

Задача 3. План: 1) Множественное выравнивание (MAFFT, ClustalOmega); 2) Выбор модели эволюции (ModelTest, ProtTest); 3) Построение дерева (Maximum Likelihood в IQ-TREE, Neighbor-Joining); 4) Визуализация и оценка поддержки узлов (FigTree, iTOL).

Задача 4. Подходы по возрастанию точности/сложности: 1) *Ab initio* (низкая точность); 2) Моделирование на основе threading; 3) Гомологичное моделирование (высокая точность при >25% идентичности). Поиск шаблонов осуществляется в БД PDB с помощью инструментов поиска гомологий (BLAST, HHblits).

Критерии оценивания контрольной работы

Результаты контрольной работы определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» выставляется, если:

- Даны полные, развернутые и правильные ответы на 2 теоретических вопроса.
- Верно решена задача, приведены необходимые расчеты и аргументированные пояснения.
- Продемонстрировано глубокое понимание принципов работы методов и умение их применять для анализа.

Оценка «хорошо» выставляется, если:

- Даны в основном правильные ответы на 2 теоретических вопроса, возможны незначительные неточности или неполнота.
- Верно решена задача, в решениях допущены незначительные ошибки в расчетах или формулировках, но общий принцип решения верен.

Оценка «удовлетворительно» выставляется, если:

- Теоретические вопросы раскрыты поверхностно, с существенными неточностями, дан правильный ответ только на 1 вопрос.
- В решении задач допущены существенные ошибки, но продемонстрировано частичное понимание подхода.

Оценка «неудовлетворительно» выставляется, если:

- Ответы на теоретические вопросы неверны или отсутствуют.
- Задача не решена или решение фундаментально неверно.
- Продемонстрировано полное непонимание принципов работы методов.

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

Зачетный билет содержит 2 вопроса.

Перечень вопросов:

1. Дайте определение биоинформатики как научной дисциплины. Каковы были исторические предпосылки её возникновения?
2. Назовите и охарактеризуйте основные области интереса и фундаментальные задачи, решаемые методами биоинформатики.
3. Какие современные компьютерные технологии являются ключевыми для развития биоинформатики на современном этапе?
4. Опишите основные принципы биологической классификации и номенклатуры. Как биоинформатика способствует решению задач таксономии?
5. В чём заключается суть центральной догмы молекулярной биологии? Какова роль биоинформатики в изучении процессов, описываемых этой догмой?
6. Что такое генетический код? Какие его свойства имеют ключевое значение для разработки вычислительных методов анализа последовательностей?
7. Дайте определения понятиям «геномика» и «секвенирование». Как связаны развитие технологий секвенирования и прогресс в биоинформатике?
8. Опишите основные типы изменчивости на молекулярном уровне (мутации, нуклеотидные замены). Как они классифицируются?
9. В чём заключаются основные принципы молекулярной эволюции? Дайте определения понятиям «гомология» и «сходство».
10. Объясните, чем отличаются процессы дивергенции и конвергенции на молекулярном уровне. Как эти процессы учитываются при построении филогенетических деревьев?
11. Опишите основные принципы теории поискового индексирования информации.

12. Как применение поисковых операторов и технологий расширенного поиска повышает эффективность работы с профессиональными базами данных?
13. Назовите и охарактеризуйте основные типы поисковых систем, используемые в научной деятельности.
14. Как организованы системы доступа к профессиональным базам данных? Каковы общие принципы их организации?
15. Какие существуют основные архивы и базы данных научной информации? Дайте их сравнительную характеристику.
16. Что такое базы данных по цитированию академической литературы (например, Scopus, Web of Science)? В чём их основное назначение?
17. Опишите назначение, структуру и принципы использования ключевых геномных баз данных (например, GenBank, ENSEMBL).
18. Опишите назначение, структуру и принципы использования ключевых протеомных баз данных (например, UniProt, PDB).
19. Для решения каких задач предназначены банки данных метаболических путей (например, KEGG)? Приведите примеры использования.
20. Что такое программы-менеджеры цитирования (системы управления библиографической информацией)? Опишите их основные функции и значение для исследовательской деятельности.
21. В чём заключается задача выравнивания первичных последовательностей биомолекул? Почему она является фундаментальной для биоинформатики?
22. Какой математический аппарат (алгоритм) лежит в основе оптимального попарного выравнивания последовательностей? Опишите его базовые принципы.
23. Что такое точечные матрицы сходства (например, PAM, BLOSUM)? Как они построены и для чего применяются при выравнивании?
24. Какое программное обеспечение используется для выравнивания аминокислотных и нуклеотидных последовательностей? Дайте примеры программных пакетов.
25. Дайте определение молекулярной филогенетики. Какие этапы включает в себя филогенетический анализ?
26. Что такое генетическая дистанция? Какие методы и модели используются для её оценки?
27. Опишите основные эволюционные модели замен нуклеотидов и аминокислот. Как выбирается адекватная модель для конкретного анализа?
28. Какие методы используются для построения филогенетических деревьев? Опишите достоинства и недостатки дистанционных и вероятностных методов.
29. Как результаты филогенетического анализа применяются в таксономии и кладистике?
30. Что изучает протеомика? Какова роль биоинформатики в протеомных исследованиях?
31. Опишите процесс фолдинга белка – формирования трёхмерной структуры. Какие факторы влияют на стабильность этой структуры?
32. Что такое денатурация белка и каковы её причины? Как это явление изучается вычислительными методами?
33. Что такое изоформы белков? Как биоинформатика помогает в анализе и предсказании альтернативного сплайсинга?
34. Чем структурное выравнивание белков принципиально отличается от выравнивания первичных последовательностей?
35. Опишите основные вычислительные методы и подходы, используемые для предсказания и моделирования трёхмерной структуры белков (гомологичное моделирование, предсказание *de novo*).

36. Какие базы данных используются для хранения и анализа информации о трёхмерной структуре биомолекул? Опишите их содержание и возможности.

37. Как осуществляется совместная интеграция различных систем по научной литературе? Каковы форматы обмена данными?

38. Опишите процесс экспорта и конвертации библиографических данных между различными платформами и системами.

39. Как современные компьютерные технологии и базы данных используются для профессионального оформления и представления результатов научных исследований?

40. Приведите пример использования современной исследовательской аппаратуры и вычислительной техники для решения конкретной инновационной задачи в области биоинформатики.

Критерии оценивания:

Результаты зачета определяются оценками «зачтено» или «не зачтено».

Оценка «зачтено» если дан как минимум один правильные ответы на теоретические вопросы с незначительными неточности или один из ответов неполный, в ином случае выставляется оценка «не зачтено».

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

Тест

1. Основной задачей биоинформатики является: (ИОПК-6.1)
 - 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. Для поиска информации о нуклеотидных последовательностях используется: (ИОПК-6.2)
 - 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.
3. Для попарного выравнивания биологических последовательностей алгоритма используется: (ИОПК-8.2)
 - а) Машинное обучение.
 - б) Динамическое программирование.
 - в) Байесовский вывод.
 - г) Кластерный анализ.
4. 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.
5. 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.

6. 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: 1 б), 2 в), 3 б), 4 б), 5 в), 6 в)

Теоретические вопросы:

1. Опишите предмет, цели и основные задачи биоинформатики. Приведите примеры профессиональных баз данных, используемых в каждой из областей ее применения (геномика, протеомика). (ИОПК-6.1)

Ответ должен содержать определение биоинформатики как дисциплины на стыке наук, перечень основных задач (анализ последовательностей, предсказание структуры, филогенетика), а также примеры баз данных (напр., GenBank для геномики, UniProt и PDB для протеомики).

2. Раскройте суть центральной догмы молекулярной биологии. Объясните, как вычислительные методы биоинформатики используются для изучения каждого из этапов реализации генетической информации. (ИОПК-6.1, ИОПК-6.2)

Ответ должен содержать описание потоков информации: репликация (ДНК → ДНК), транскрипция (ДНК → РНК), трансляция (РНК → белок), а также примеры методов (алгоритмы выравнивания для анализа последовательностей ДНК/РНК, предсказание структуры белка).

3. Опишите принципы организации и поиска информации в основных биологических базах данных. Какой алгоритм действий следует применять исследователю для эффективного поиска литературных источников по конкретной научной проблеме? (ИОПК-6.2)

Ответ должен содержать описание устройства баз данных (первичные, вторичные), принципов индексирования, а также пошаговый алгоритм: выбор БД (PubMed/Scopus), формирование запроса с использованием операторов, применение фильтров, использование менеджера цитирования.

4. В чем заключается задача выравнивания биологических последовательностей? Опишите математическую основу (алгоритм) и критерии качества, используемые для построения оптимального выравнивания. (ИОПК-8.2)

Ответ должен содержать определение выравнивания, его цели (поиск гомологии, установление консервативных участков), описание алгоритма динамического программирования и понятия о *scoring*-матрицах (PAM, BLOSUM) и штрафах за гэпы.

5. Что такое молекулярная филогенетика? Опишите основные этапы построения филогенетического дерева: от получения последовательностей до интерпретации результатов. (ИПК-2.3, ИПК-2.4)

Ответ должен содержать определение филогенетики, этапы: получение и множественное выравнивание последовательностей, выбор эволюционной модели, собственно построение дерева (максимальное правдоподобие, байесовские методы), оценка достоверности узлов (*bootstrap*), визуализация и интерпретация.

6. Какие вычислительные методы и базы данных используются для предсказания трехмерной структуры белка? Опишите логику выбора метода в зависимости от наличия гомологичных последовательностей с известной структурой. (ИОПК-8.2, ИПК-2.4)

Ответ должен содержать сравнение методов: гомологичное моделирование (при наличии гомолога), предсказание *de novo* (*ab initio*), а также названия ключевых баз данных (PDB) и программного обеспечения (напр., SWISS-MODEL, Rosetta).

Информация о разработчиках

Слепцов Алексей Анатольевич, кандидат медицинских наук, кафедра физиологии растений, биотехнологии и биоинформатики Биологического института Национального исследовательского Томского государственного университета, доцент.