



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»

Кафедра физико-химической биологии, биоинженерии и биоинформатики



ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ
для проведения текущего контроля и промежуточной аттестации по дисциплине:

Б1.О.41 БИОИНФОРМАТИКА

Специальность: 06.05.01 «Биоинженерия и биоинформатика»

Специализация: «Биоинженерия и биоинформатика»

Квалификация выпускника: биоинженер и биоинформатик

Форма обучения: очная с элементами электронного обучения и дистанционных образовательных технологий

Согласовано с УМК биологического факультета

Протокол № 5 от 21 марта 2025 г.

Председатель Матвеев А.Н. Матвеев

Рекомендовано кафедрой физико-химической биологии, биоинженерии и биоинформатики

Протокол № 12 от 19 марта 2025 г.

Зав. кафедрой Соловьева В.П. Соловьева

Иркутск 2025 г.

ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

Разработан для учебной дисциплины Б1.О.41 «БИОИНФОРМАТИКА», Специализация: «Биоинженерия и биоинформатика». Фонд оценочных материалов (ФОМ) включает оценочные материалы для проведения текущего контроля, промежуточной аттестации в форме экзамена.

Оценочные материалы соотнесены с требуемыми результатами освоения образовательной программы 06.05.01 «Биоинженерия и биоинформатика», в соответствии с содержанием рабочей программы учебной дисциплины Б1.О.41 «БИОИНФОРМАТИКА» с учетом ОПОП.

Нормативные документы, регламентирующие разработку ФОМ:

- статья 2, часть 9 Федерального закона «Об образовании в Российской Федерации», ФЗ-273, от 29.12.2012 г.;

- ФГОС ВО по специальности 06.05.01 «Биоинженерия и биоинформатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации 12 августа 2020 г. № 973.

1. Компетенции, формируемые в процессе изучения дисциплины (3 курс, 6 семестр)

ОПК-4: Способен применять методы биоинженерии и биоинформатики для получения новых знаний и для получения биологических объектов с целенаправленно измененными свойствами, проводить анализ результатов и методического опыта исследования, определять практическую значимость исследования.

ОПК-5: Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформационными средствами анализа..

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

| Компетенции | Индикаторы компетенций | Планируемые результаты обучения | Формы и методы контроля и оценки |
|---|--|---|--|
| ОПК-4 Способен применять методы биоинженерии и биоинформатики для получения новых знаний и для получения биологических объектов с целенаправленно измененными свойствами, проводить анализ результатов и методического опыта исследования, определять практическую значимость исследования.; | <i>ИДК</i> опк 4.1 Демонстрирует навыки использования методов биоинженерии и биоинформатики для получения новых фундаментальных знаний | Знать: основы биоинженерии Уметь: использовать методы биоинженерии и биоинформатики Владеть: методами для получения новых фундаментальных знаний | Текущий контроль: - контрольные вопросы, - тестирование Промежуточная аттестация: экзамен |
| | <i>ИДК</i> опк 4.2 Применяет методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами | Знать: методы биоинженерии и биоинформатики Уметь: применять методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами. Владеть: методами биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами. | Текущий контроль: - контрольные вопросы, - тестирование Промежуточная аттестация: экзамен |

| | | | |
|--|--|--|---|
| | <p>ИДК ОПК 4.3 Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования</p> | <p>Знать: методы анализа и интерпретации результатов исследования с целью определения практической значимости исследования Уметь: применять методы анализа и интерпретации результатов исследования с целью определения практической значимости исследования Владеть: методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования</p> | <p>Текущий контроль: - контрольные вопросы, - тестирование</p> <p>Промежуточная аттестация: экзамен</p> |
| <p>ОПК-5 Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформационными средствами анализа;</p> | <p><i>ИДК ОПК 5.1</i> Использует информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию</p> | <p>Знать: информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию Уметь: применять информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию Владеть: методами анализа информации, накопленной в базах данных по структуре геномов, белков и другую биологическую информацию</p> | <p>Текущий контроль: - контрольные вопросы, - тестирование</p> <p>Промежуточная аттестация: экзамен</p> |
| | <p><i>ИДК ОПК 5.2</i> Умеет применять биоинформационные методы и полученные знания для анализа геномной, структурной и иной информации</p> | <p>Знать: биоинформационные методы и полученные знания для анализа геномной, структурной и иной информации Уметь: применять биоинформационные методы и полученные знания для анализа геномной, структурной и иной информации Владеть: биоинформационными методами для анализа геномной, структурной и иной информации</p> | <p>Текущий контроль: - контрольные вопросы, - тестирование</p> <p>Промежуточная аттестация: экзамен</p> |
| | <p><i>ИДК ОПК 5.3</i> Демонстрирует навыки владения основными биоинформационными средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p> | <p>Знать: основные биоинформационные средства анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей. Уметь: применять основные биоинформационные средства анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей. Владеть: основными биоинформационными средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p> | <p>Текущий контроль: - контрольные вопросы, - тестирование</p> <p>Промежуточная аттестация: экзамен</p> |
| <p>ОПК-7: Способен понимать принципы работы современных информационных технологий и</p> | <p><i>ИДК ОПК 7.1</i> Демонстрирует теоретические и практические навыки использования современных информационных</p> | <p>Знать: теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности Уметь: применять теоретические и практические навыки использования современных информационных</p> | <p>Текущий контроль: - контрольные вопросы, - тестирование</p> |

| | | | |
|--|--|---|--|
| использовать их для решения задач профессиональной деятельности; | технологий в области профессиональной деятельности | технологий в области профессиональной деятельности Владеть: теоретическими и практическими навыками использования современных информационных технологий в области профессиональной деятельности | Промежуточная аттестация: экзамен |
| | <i>ИДК ОПК 7.2</i> Использует современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности. | Знать: современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности. Уметь: применять современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности. Владеть: методами анализа современных информационных технологий в рамках освоения материала и реализации задач в области профессиональной деятельности. | Текущий контроль: - контрольные вопросы, - тестирование Промежуточная аттестация: экзамен |

2. Оценочные материалы для проведения текущего контроля

2.1 Контрольные вопросы

1. Что такое Биоинформатика?
2. Какова заслуга Л. Полинга в развитии биоинформатики?
3. Каких еще пионеров биоинформатики вы знаете?
4. Кем и когда был получен первый организм с рекомбинантной ДНК?
5. Какие компьютерные программы, используемые в биоинформатике, появились первыми?
6. В чем заключалась суть программы «Геном человека», и каковы ее результаты?
7. Опишите вкратце историю развития технологии секвенирования.
8. В чем состоят цели и задачи биоинформатики?
9. В каких областях Биоинформатика находит применение?
10. Каким образом защищается интеллектуальная собственность в биоинформатике?
11. Что такое компьютер?
12. Что такое программное обеспечение? Какие виды ПО выделяют?
13. Приведите примеры языков программирования и назовите их отличительные особенности.
14. Что такое HTML и BIOML?
15. Что такое Internet и WWW? В чем их различия?
16. Что такое сетевые протоколы? Назовите самые популярные.
17. Опишите принцип работы Internet.
18. В чем оказалось преимущество WWW по сравнению с другими конкурирующими системами?
19. Что такое браузер? Назовите наиболее популярные обозреватели сети.
20. Какую пользу приносит система выборки последовательностей в биоинформатике?
21. Из каких операций состоит основная реакция секвенирования ДНК?
22. Опишите полный процесс секвенирования ДНК.
23. Что такое открытая рамка считывания и какова ее роль?
24. Опишите метод определения последовательности клона.
25. Что такое ярлыки экспрессионных последовательностей? Каким образом секвенируют EST?
26. Какие методы используют для секвенирования белков?
27. Опишите суть процесса гибридизации микроматриц ДНК.
28. В чем состоит анализ экспрессии белков?
29. Какие подходы к обнаружению генов вам известны?

30. Приведите пример организмов, гены которых были успешно расшифрованы.

31. Что такое база данных?

32. Какие типы баз данных вы знаете? Какие их функции?

33. Что такое СУБД? Какие типы СУБД вы знаете?

34. Что такое SQL и какую функцию он выполняет?

35. Исторические аспекты возникновения и развития биоинформационных БД.

36. Что такое EMBL и NCBI? Какие базы данных и программные продукты они поддерживают?

37. Что такое метод-ориентированные базы данных?

38. Приведите пример организм-ориентированных БД.

39. Приведите примеры баз данных последовательностей нуклеиновых кислот. Для каких целей они созданы?

40. Каковы функции баз данных белковых последовательностей? Назовите несколько ресурсов.

41. Какая база данных белковых последовательностей является наиболее полной? Какие функциональные разделы в ней выделяют?

42. Что такое базы данных структур? Приведите примеры.

43. Что общего и в чем различия между структурными классификациями белков в БД SCOP и САТН?

44. Что такое библиографическая база данных. Приведите несколько примеров.

45. Каковы особенности вторичных баз данных?

46. Что такое выравнивание последовательностей?

47. Каковы цели выравнивания последовательностей?

48. Какие типы выравнивания последовательностей вы знаете?

49. Опишите этапы анализа точечной диаграммы.

50. Опишите принцип назначения счета мутациям, выпадениям и заменам.

51. Какие программы применяются для попарного выравнивания последовательностей в базах данных?

52. Что такое матрица процентов точечных мутаций?

53. Что такое динамическое программирование?

54. Какие алгоритмы используются в динамическом программировании? Чем они отличаются?

55. Почему для оценки выравненности последовательностей нельзя использовать традиционные статистические критерии?

56. На каких критериях основано сравнение последовательностей?

57. Что такое расстояния Хемминга и Ливенштейна? Чем они отличаются?

58. Что такое множественное выравнивание выравнивание последовательностей?

59. Перечислите основные этапы в построении множественного выравнивания.

60. Какие программы применяют для множественного выравнивания?

61. Как вы понимаете разницу между понятиями «гомология» и «подобие»?

62. Чем отличаются ортологи, паралоги и ксенологи?

63. Что такое филогения?

64. В чем сущность фенетического подхода?

65. Назовите характерные особенности кладистического подхода.

66. Дайте определение следующим терминам: орграф, узел (вершина), дуга, путь, контур.

67. Что такое филогенетическое дерево? Каковы его отличительные свойства?

68. Какие виды филогенетических деревьев вы знаете?

69. Какие допущения приняты в построении филогенетических деревьев?

70. В чем молекулярная филогенетика превосходит традиционную?

71. Какие базы данных используют в филогенетическом анализе?

72. Что такое Newick format?

73. Назовите группы методов построения филогенетических деревьев?

74. Чем различаются кладистические и переборные методы филогенетики? На чем основан метод наибольшего правдоподобия?

75. Каким образом можно оценить достоверность построенного филогенетического дерева?

Критерии оценивания ответов на контрольные вопросы:

- 1) полнота и правильность ответа;
- 2) степень осознанности, понимания изученного;
- 3) языковое оформление ответа.

Критерии оценивания ответов на контрольные вопросы

| Критерий | Оцениваемые компетенции | Оценка |
|---|---------------------------|---------------------|
| Ответ оценивается на «отлично», если студент: полно излагает изученный материал, дает правильное определенное понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только по учебнику, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка. | ОПК-4, ОПК-5, ОПК-7 | отлично |
| Ответ оценивается на «хорошо», если студент даёт ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1-2 ошибки, которые сам же исправляет, и 1-2 недочёта в последовательности и языковом оформлении излагаемого. | | хорошо |
| «Удовлетворительно» ставится, если студент обнаруживает знание и понимание основных положений темы, но при этом: излагает материал неполно и допускает неточности в определении понятий или формулировке теорий; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого. | | удовлетворительно |
| Оценка «неудовлетворительно» ставится, если ответ не удовлетворяет требованиям положительной оценки или студент отказывается отвечать на контрольные вопросы | | неудовлетворительно |

2.2. Перечень докладов (по теме № 3)

1. Классические и новейшие методы секвенирования нуклеиновых кислот
2. Метод «Expressed sequence tags»: принципы и область применения
3. Методы определения первичной структуры белка
4. Методы изучения экспрессии генов: Нозерн- и Вестерн-блоттинги
5. Методы изучения экспрессии генов: Serial Analysis of Gene Expression и двумерный электрофорез в полиакриламидном геле
6. Методы изучения экспрессии генов: ДНК-чипы

Критерии оценки доклада:

- Новизна текста: а) актуальность темы; б) новизна и самостоятельность в постановке проблемы, формулирование нового аспекта известной проблемы в установлении новых связей (межпредметных, внутрипредметных, интеграционных); в) умение работать с исследованиями, критической литературой, систематизировать и структурировать материал; г) явленность авторской позиции, самостоятельность оценок и суждений; д) стилевое единство текста, единство жанровых черт.

- Степень раскрытия сущности вопроса: а) соответствие содержания теме и плану доклада; б) полнота и глубина знаний по теме; в) обоснованность способов и методов работы с материалом; г) умение обобщать, делать выводы, сопоставлять различные точки зрения по одному вопросу (проблеме).
- Обоснованность выбора источников: а) оценка использованной литературы: привлечены ли наиболее известные работы по теме исследования (в т.ч. журнальные публикации последних лет, последние статистические данные, сводки, справки и т.д.).
- Соблюдение требований к оформлению: а) насколько верно оформлены ссылки на используемую литературу, список литературы; б) оценка грамотности и культуры изложения (в т.ч. орфографической, пунктуационной, стилистической культуры), владение терминологией; в) соответствие презентации содержанию доклада и рекомендациям по ее подготовке (см. п. 4.4).

Критерии оценивания доклада

| Критерий | Оцениваемые компетенции | Оценка |
|--|---------------------------|---------------------|
| В докладе полностью раскрыта тема, проанализировано современное состояние вопроса; студент свободно владеет материалом, излагает его логично, последовательно, лаконично, хорошим научным языком. Доклад сопровождается презентацией, которая составлена с соблюдением общих требований оформления, содержит ссылки на приведенные фото, рисунки, схемы и т.д. При обсуждении студент демонстрирует понимание изучаемой проблемы и методологии научного исследования, владение профессиональной терминологией и умение грамотно отвечать на вопросы аудитории. | ОПК-4, ОПК-5, ОПК-7 | отлично |
| Тема раскрыта, приведено достаточное количество материала, но при этом материал в недостаточной степени проанализирован автором. Имеются недочеты в оформлении презентации или презентация не в полной степени соответствует общим требованиям. Ответы студента на вопросы не являются исчерпывающими и аргументированными. | | хорошо |
| Тема раскрыта не полностью, материал не проанализирован, студент показывает поверхностные знания. Презентация частично соответствует установленным требованиям. При обсуждении доклада студент дает неправильные или исчерпывающие ответы. | | удовлетворительно |
| Тема не раскрыта, приведен скудный объем материала; презентация отсутствует или не соответствует требованиям. При обсуждении доклада студент не дает ответы или они не соответствуют вопросам. | | неудовлетворительно |

2. 3 Демонстрационные практические задания для текущего контроля

Задание 1.

На конце гена β -гемоглобина человека имеется следующая последовательность:

...ctggccsacaagtatcaaa

а) Какова аминокислотная последовательность, соответствующая представленной?

б) Напишите нуклеотидную последовательность, в которой единичная замена нуклеотида приводит к «молчащей» мутации в этом фрагменте.

в) Напишите нуклеотидную последовательность, в которой единичная замена основания может привести к миссенс-мутации в этом фрагменте.

г) Напишите нуклеотидную последовательность, в которой единичная замена основания приводит к преждевременной остановке синтеза белка.

д) Напишите нуклеотидную последовательность, в которой единичная замена нуклеотида приводит к ошибке терминации и продолжению синтеза цепи.

Задание 2.

Найдите в БД UniProtKB документ, содержащий информацию о белке, указанно в файле «Индивидуальное распределение белков». Изучите полученный документ и заполните таблицу.

| Поле | Метка поля | Содержание |
|---|------------|------------|
| Код доступа | | |
| Идентификатор последовательности в БД | | |
| Название и краткое описание белка | | |
| Дата создания документа | | |
| Дата последнего исправления аннотации | | |
| Название организма | | |
| Полное название таксона | | |
| Длина последовательности | | |
| Молекулярная масса белка | | |
| Название гена, кодирующего этот белок | | |
| Число публикаций, использованных при создании документа | | |
| Выходные данные самой свежей публикации | | |
| Ключевые слова | | |
| Содержание поля комментариев | | |
| Особенности белка | | |
| Перекрестные ссылки | | |

Задание 3.

В файле «Альбом белков» найдите «свой» белок и на его условном изображении обозначьте цветом α -спирали (красным) и тяжи β -листа (синим). Если в молекуле имеется несколько доменов, выделите их. Объедините белковые структуры, представленные на рисунках, в категории: α -спираль, β -структура, α/β и $\alpha+\beta$.

Задание 4.

Путем горизонтального переноса в геном *E. coli* в течение 14,4 млн лет были внесены 755 ORF, что является причиной дивергенции *E. coli* от *Salmonella*. Оцените среднюю скорость горизонтального переноса в (парах нуклеотидов)/год. Каков процент генов, внесенных в геном *E. coli* посредством горизонтального переноса?

Задание 5.

Мутацией, приводящей к серповидно-клеточной анемии, является единственная замена основания А на Т, приводящая к замене Glu \rightarrow Val в β -цепи гемоглобина. Замена оснований происходит в последовательности 5'-GTGAG-3' (нормальная) - GTGTG (мутантная). Какая рестриктаза может быть использована для обнаружения этого различия? Какова ее специфичность?

Задание 6.

Получите и выровняйте последовательность двух белков из лошади, кита и кенгуру: митохондриального цитохрома- b и панкреатической рибонуклеазы.

Сравните степень попарного сходства каждого белка для каждой пары видов. Совместимы ли выводы, полученные в результате анализа последовательностей митохондриального цитохрома- b , с

выводами, полученными в результате анализа последовательностей панкреатических рибонуклеаз этих видов?

Задание 7.

В библиографической базе данных PubMed найдите ссылки на статьи, в которых упоминается название вашего белка. Укажите их общее количество. Из списка найденных статей выберите одну и сохраните ее аннотацию в формате txt. Найдите ссылки на все публикации последнего автора этой статьи за последние 3 года. Повторите алгоритм поиска в отечественной библиографической системе e-library.

Задание 8.

С помощью программы BLAST найдите последовательности, гомологичные вашему белку (файл «Индивидуальное распределение белков»). Сформируйте файл с набором последовательностей, для которых будет проводится множественное выравнивание. Проведите множественное выравнивание этих последовательностей с помощью программ семейства CLUSTAL. Повторите выравнивание при разных настройках параметров. Сделайте выводы.

Задание 9.

Гормоны роста большинства млекопитающих имеют очень сходные аминокислотные последовательности. Человеческие гормоны роста отличаются гораздо сильнее. Эволюция гормонов роста ускорялась резко в той линии, которая вела к человеку. Путем выравнивания последовательностей гормонов роста из организмов близкородственных человеку видов и более отдаленных таксонов определите, где в эволюционном древе человека имело место ускорение эволюции гормонов роста.

Задание 10.

Хордовых относят к группе вторичноротовых, которая включает хордовых, оболочечников и иглокожих. Есть систематические различия между этими тремя типами в их митохондриальном генетическом коде. Определите примеры организмов из каждого типа, чьи аминокислоты соответствуют кодонам ATA и AGA. Исходя из этого получите филогенетическое дерево для этих типов вторичноротовых.

Критерии оценивания заданий

| Критерий | Оцениваемые компетенции | Оценка |
|--|---------------------------|---------------------|
| Оценка «отлично» выставляется, если задание выполнено правильно, в ходе решения продемонстрированы понимание метода решения, правильность выбора и использования программного обеспечения, способность интерпретировать результаты, приведено детальное и полное описание решения. | ОПК-4, ОПК-5, ОПК-7 | отлично |
| Оценка «хорошо» выставляется, если задание выполнено правильно, но студент затрудняется изложить и обосновать алгоритм решения и / или интерпретировать полученные результаты | ОПК-4, ОПК-5, ОПК-7 | хорошо |
| Оценка «удовлетворительно» выставляется, если задание выполнено неправильно, но студент демонстрирует верный подход к проблеме, поставленной в задаче | ОПК-4, ОПК-5, ОПК-7 | удовлетворительно |
| Оценка «неудовлетворительно» выставляется, если задание выполнено неправильно или не выполнено совсем. | ОПК-4, ОПК-5, ОПК-7 | неудовлетворительно |

3. Оценочные материалы для проведения промежуточной аттестации (экзамена)

Промежуточная аттестация проходит в форме экзамена (6 семестр), к которому допускаются студенты, выполнившие в полном объеме аудиторную нагрузку, самостоятельную работу. Студенты, имеющие задолженность, должны выполнить все обязательные виды деятельности. Экзамен проводится в форме тестирования

3.1 Тестирование

Вариант 1

| Индекс и содержание формируемой компетенции | Индикаторы компетенций | Тестовые задания для промежуточной аттестации |
|---|---|--|
| ОПК-4 Способен применять методы биоинженерии и биоинформатики для получения новых знаний и для получения биологических объектов с целенаправленно измененными свойствами, проводить анализ результатов и методического опыта исследования, определять практическую значимость исследования | <i>ИДК</i> ОПК 4.1 Демонстрирует навыки использования методов биоинженерии и биоинформатики для получения новых фундаментальных знаний <i>ИДК</i> ОПК 4.2 Применяет методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами <i>ИДК</i> ОПК 4.3 Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования | Задание комбинированного типа с выбором одного или нескольких верных ответов из четырех предложенных с аргументацией выбора Вопрос 1. <i>Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i> Что является предметом изучения биоинформатики? a) Хранение и анализ биологических данных b) Разработка методов секвенирования ДНК <i>in vitro</i> c) Прогноз структуры и функций белков d) Проведение генетических экспериментов на клеточном уровне Ответ _____ Правильный ответ: a, c Аргументация: Биоинформатика связана с анализом данных и моделированием, а не с экспериментальными манипуляциями. Вопрос 2. <i>Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i> Какое открытие оказало наибольшее влияние на развитие биоинформатики? a) Изобретение микроскопа b) Расшифровка структуры ДНК c) Создание метода ПЦР d) Разработка метода КРИСПР-Cas9 Ответ _____ Правильный ответ: b, c Аргументация: Открытие двойной спирали ДНК (1953) и ПЦР (1983) стали основой для секвенирования и хранения данных. |

| | | |
|--|--|--|
| <p>ОПК-5 Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформатическими средствами анализа</p> | <p><i>ИДК</i> опк 5.1 Использует информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию</p> <p><i>ИДК</i> опк 5.2 Умеет применять биоинформатические методы и полученные знания для анализа геномной, структурной и иной информации</p> <p><i>ИДК</i> опк 5.3 Демонстрирует навыки владения основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p> | <p>Вопрос 3. Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> <p>Какие языки программирования традиционно применяются в биоинформатике?</p> <p>a) PERL b) Python c) Java d) COBOL</p> <p>Ответ _____</p> <p>Правильный ответ: a, b, c</p> <p>Аргументация: PERL, Python и Java — широко применяются для анализа последовательностей и биоинформатического софта. COBOL в этой области не используется.</p> <p>Вопрос 4. Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> <p>Какой протокол используется для передачи гипертекста в Интернете?</p> <p>a) FTP b) POP c) HTTP d) TCP/IP</p> <p>Ответ _____</p> <p>Правильный ответ: c</p> <p>Аргументация: HTTP (HyperText Transfer Protocol) — стандарт передачи гипертекстовых документов.</p> <p>Вопрос 5. Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> <p>Метод Сэнгера для секвенирования ДНК основан на:</p> <p>a) Использовании дидезоксинауклеотидов b) Флуоресцентной детекции c) Рестриктазном расщеплении ДНК d) Автоматическом чтении последовательности</p> <p>Ответ _____</p> <p>Правильный ответ: a, b, d</p> <p>Аргументация: Метод Сэнгера использует ddNTP, автоматическое определение и метки. Рестриктазы применяются в других методах.</p> <p>Вопрос 6. Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> |
| <p>ОПК-7: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности</p> | <p><i>ИДК</i> опк 7.1 Демонстрирует теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности</p> | |

| | |
|---|---|
| <p><i>ИДК опк 7.2</i></p> <p>Использует современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности.</p> | <p>Какая техника применяется для анализа экспрессии белков?</p> <p>a) Вестерн-блоттинг b) Нозерн-блоттинг c) EST d) ДНК-чипы</p> <p>Ответ _____</p> <p>Правильный ответ: a</p> <p>Аргументация: Вестерн-блоттинг применяется для выявления белков, остальные методы — для анализа РНК или ДНК.</p> <p>Вопрос 7.</p> <p><i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>Какая база данных относится к первичным?</p> <p>a) GenBank b) UniProtKB c) SCOP d) Pfam</p> <p>Ответ _____</p> <p>Правильный ответ: a</p> <p>Аргументация: GenBank хранит первичные нуклеотидные последовательности.</p> <p>Вопрос 8.</p> <p><i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>Какая СУБД наиболее распространена в биоинформатике?</p> <p>a) Иерархическая b) Реляционная c) Сетевая d) Древовидная</p> <p>Ответ _____</p> <p>Правильный ответ: b</p> <p>Аргументация: В большинстве биобаз данных используется реляционная модель.</p> <p>Вопрос 9.</p> <p><i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>Что является целью выравнивания последовательностей?</p> <p>a) Сравнение структур белков b) Поиск гомологичных последовательностей c) Определение экспрессии генов d) Определение степени сходства нуклеотидов</p> |
|---|---|

Ответ _____

Правильный ответ: b, d

Аргументация: Выравнивание применяется для поиска сходства и гомологии в последовательностях.

Вопрос 10.

Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Какой алгоритм используется для глобального выравнивания?

- a) Смит-Уотерман
- b) Нидлман-Вунш
- c) BLAST
- d) ClustalW

Ответ _____

Правильный ответ: b

Аргументация: Нидлман-Вунш — алгоритм глобального выравнивания, Смит-Уотерман — локального.

Вопрос 11.

Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Ортологи — это:

- a) Гены, возникшие в результате дупликации в одном геноме
- b) Гены, разделившиеся после видеообразования
- c) Гены, перенесённые горизонтальным переносом
- d) Гены, кодирующие идентичные белки

Ответ _____

Правильный ответ: b

Аргументация: Ортологи — гены у разных видов, происходящие от общего предка.

Вопрос 12.

Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Какой метод основан на минимизации числа мутаций при построении филогенетического дерева?

- a) Метод максимальной парсимонии
- b) Метод наибольшего правдоподобия
- c) Метод Байеса
- d) Метод объединения соседей

Ответ _____

Правильный ответ: a

Аргументация: Максимальная парсимония — минимизация эволюционных изменений.

Вопрос 13.

Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор

ответа.

Какая молекула часто используется в молекулярной филогении?

- a) мРНК
- b) 16S рРНК
- c) тРНК
- d) Полипептиды

Ответ _____

Правильный ответ: b

Аргументация: 16S рРНК — стандарт для построения филогенетических деревьев.

Задание закрытого типа на установление соответствия

Вопрос 14.

Прочтите вопрос и установите соответствие между учёным и его методом в биоинформатике:

Ученый:

- a) Ф. Сэнгер
- b) Э. Эдман
- c) Кэри Муллис
- d) GenBank

Методы в биоинформатике

- 1 — Метод ПЦР
- 2 — Секвенирование ДНК методом обрыва цепи
- 3 — Секвенирование белков
- 4 — Первая база данных нуклеотидных последовательностей

Ответ _____

Правильный ответ:

- a — 2
- b — 3
- c — 1
- d — 4

Аргументация: Сэнгер предложил метод ddNTP, Эдман — метод деградации белков, Муллис — изобрёл ПЦР, GenBank — база нуклеотидных последовательностей.

Вопрос 15.

Прочтите вопрос и установите соответствие между базой данных и её назначением:

База данных:

- a) PDB
- b) UniProtKB
- c) GenBank
- d) SCOP

Назначение:

- 1 — Структуры белков
- 2 — Белковые последовательности
- 3 — Нуклеотидные последовательности

4 — Классификация белков по структуре

Ответ _____

Правильный ответ:

a — 1

b — 2

c — 3

d — 4

Аргументация: PDB хранит 3D-структуры, UniProtKB — белковые последовательности, GenBank — ДНК/РНК, SCOP — иерархическая классификация белков.

Задание закрытого типа на установление последовательности

Вопрос 16.

Прочтите вопрос и установите последовательность шагов для этапов секвенирования по Сэнгеру в правильной последовательности:

- a) Добавление праймера
- b) Включение дидезоксинуклеотидов (ddNTP)
- c) Репликация ДНК
- d) Электрофоретическое разделение фрагментов

Ответ _____

Правильный ответ:

Правильная последовательность: a → c → b → d

Аргументация: сначала праймер, затем синтез цепи, встраивание ddNTP, а в конце электрофорез.

Вопрос 17.

Прочтайте вопрос и установите порядок действий при построении филогенетического дерева:

- a) Выравнивание последовательностей
- b) Выбор модели эволюции
- c) Расчёт матрицы расстояний
- d) Построение дерева

Ответ _____

Правильный ответ:

Правильная последовательность: a → c → b → d

Аргументация: сначала выравниваются последовательности, затем создаётся матрица различий, выбирается модель, и только потом строится дерево.

Задание открытого типа с развернутым ответом

Вопрос 18.

Прочтайте вопрос и запишите развернутый обоснованный ответ

Перечислите известные базы данных геномов и нуклеотидных последовательностей?

Ответ:

- **NCBI GenBank (National Center for Biotechnology Information GenBank):** Одна из самых крупных и всеобъемлющих баз данных нуклеотидных последовательностей, поддерживаемая NCBI (США). Включает

- GenBank, RefSeq и Trace Archive.
- **EMBL-EBI ENA (European Nucleotide Archive):** Европейский аналог GenBank, поддерживаемый EMBL-EBI (Европейский институт биоинформатики). Включает базы данных EMBL, Trace Archive и Sequence Read Archive (SRA).
- **DDBJ (DNA Data Bank of Japan):** Японская база данных нуклеотидных последовательностей, сотрудничающая с GenBank и ENA в рамках Международного сотрудничества баз данных нуклеотидных последовательностей (INSDC).

Вопрос 19.

Прочитайте вопрос и запишите развернутый обоснованный ответ

Сформулирует принцип работы алгоритмов локального выравнивания нуклеотидных и аминокислотных последовательностей?

Ответ:

Принцип работы алгоритмов локального выравнивания, таких как алгоритм Смита-Вотермана, заключается в поиске участков наибольшего сходства между двумя последовательностями, не требуя совпадения по всей длине последовательностей. Вот ключевые шаги:

1. **Матрица подсчета баллов (scoring matrix):**

- Определяется матрица подсчёта баллов, которая задаёт баллы за совпадения (match), несовпадения (mismatch) и гэпы (пропуски).
- Для нуклеотидных: Обычно используются простые схемы: +1 за совпадение, -1 за несовпадение, гап - штраф.
- Для аминокислотных: Используются более сложные матрицы, такие как BLOSUM или PAM, отражающие эволюционные взаимосвязи между аминокислотами (замена консервативной аминокислоты оценивается выше, чем замена на радикально отличающуюся).

1. **Создание матрицы выравнивания:**

- Строится матрица выравнивания, размерностью (длина sequenceA + 1) * (длина sequenceB + 1).
- Первая строка и первый столбец заполняются нулями (это ключевое отличие от глобального выравнивания).

1. **Заполнение матрицы выравнивания:**

- Каждая ячейка матрицы (i, j) рассчитывается по следующей формуле:
 - $Score(i, j) = \max \{ 0, // \text{Важно! Если все варианты дают отрицательный результат, начинаем с нуля - это позволяет "начать" выравнивание заново.} Score(i-1, j-1) + score(sequenceA[i], sequenceB[j]), // \text{Совпадение или несовпадение} Score(i-1, j) + gap_penalty, // \text{Гэп в sequenceA} Score(i, j-1) + gap_penalty // \text{Гэп в sequenceB} \}$
 - $Score(i-1, j-1) + score(sequenceA[i], sequenceB[j]) - \text{балл из матрицы подсчёта баллов для выравнивания символов sequenceA[i] и sequenceB[j].}$
 - $gap_penalty - \text{штраф за введение гэпа.}$

1. **Поиск максимального значения:**

- После заполнения матрицы находится ячейка с максимальным значением. Это значение представляет собой наилучший счет локального выравнивания.

- Координаты этой ячейки (i, j) являются конечной точкой оптимального локального выравнивания.

1. **Трассировка (backtracking):**
- Начиная с ячейки с максимальным значением, выполняется трассировка в обратном направлении, чтобы определить путь, который привёл к этому значению.
- На каждом шаге трассировки определяем, из какой ячейки $(i-1, j-1)$, $(i-1, j)$ или $(i, j-1)$ мы пришли в текущую ячейку, основываясь на том, какой вариант максимизировал формулу подсчета баллов.
- Трассировка продолжается до тех пор, пока не будет достигнута ячейка со значением 0.
- Путь трассировки определяет оптимальное локальное выравнивание.

Ключевые особенности локального выравнивания:

- **Начало с нуля:** Если при вычислении балла для ячейки все варианты дают отрицательный результат, в ячейку записывается 0. Это позволяет алгоритму игнорировать области с низким сходством и "начинать" выравнивание заново.
- **Максимальный балл:** Ищется максимальный балл *во всей* матрице, а не только в последней строке или столбце, как в глобальном выравнивании.
- **Трассировка до нуля:** Трассировка начинается с ячейки с максимальным значением и продолжается до ячейки со значением 0.

Преимущества локального выравнивания:

- Поиск консервативных доменов: Позволяет находить участки высокого сходства даже в дивергентных последовательностях.
- Работа с фрагментами: Подходит для выравнивания фрагментов последовательностей.
- Более гибкий подход: Не требует, чтобы выравнивание охватывало всю длину последовательностей.

Алгоритм Смита-Вотермана является мощным инструментом для выявления локальных сходств в биомолекулярных последовательностях, что делает его широко используемым в биоинформатике.

Вопрос 20.

Прочтите вопрос и запишите развернутый обоснованный ответ

Принцип работы дистанционных методов молекулярной филогенетии?

Ответ:

Дистанционные методы (Distance-based methods) молекулярной филогенетии - это группа методов, которые строят филогенетические деревья на основе матрицы попарных генетических дистанций между последовательностями. Эти дистанции отражают меру генетических различий между последовательностями, и, как правило, вычисляются на основе выравниваний последовательностей. Суть этих методов заключается в том, чтобы кластеризовать близкие последовательности вместе и строить дерево, отражающее эти отношения.

Основные этапы дистанционных методов:

1. **Выравнивание последовательностей:**
 - Первый шаг - это множественное выравнивание (multiple sequence alignment, MSA) последовательностей, которые будут использоваться для построения филогенетического дерева. Выравнивание позволяет определить соответствия между позициями в разных последовательностях.
2. **Расчет генетических дистанций:**
 - На основе MSA вычисляется матрица попарных генетических дистанций между всеми парами последовательностей.
 - Генетическая дистанция - это мера генетического различия между двумя последовательностями.

Существует множество способов оценки генетических дистанций, наиболее распространённые:

- **Пропорция несовпадающих сайтов (p-distance):** Просто доля сайтов, в которых последовательности различаются. Самый простой метод, но недостаточно точный при больших дистанциях, так как не учитывает множественные замены в одном и том же сайте.
- **Модель Джукса-Кантора (Jukes-Cantor model):** Учитывает множественные замены, предполагая, что все нуклеотидные замены равновероятны.
- **Модель Кимуры 2-параметра (Kimura 2-parameter model):** Различает переходы (пурин-пурин или пиримидин-пиримидин) и трансверсии (пурин-пиримидин или наоборот), так как переходы происходят чаще.
- **Более сложные модели:** Существуют и другие, более сложные модели (например, GTR), которые учитывают разные скорости замен для разных нуклеотидов и другие факторы. Выбор модели зависит от данных и цели анализа.

3. Построение филогенетического дерева:

- На основе матрицы дистанций строится филогенетическое дерево. Существует несколько алгоритмов:
 - **UPGMA (Unweighted Pair Group Method with Arithmetic Mean):**
 - Это иерархический аггломеративный (снизу-вверх) метод.
 - Начинает с того, что каждая последовательность рассматривается как отдельный кластер.
 - На каждом шаге объединяются два ближайших кластера (с наименьшей дистанцией).
 - Дистанция между новым кластером и другими кластерами вычисляется как среднее арифметическое дистанций между последовательностями в новом кластере и последовательностями в других кластерах.
 - Предполагает постоянную скорость эволюции по всем линиям (молекулярные часы), что часто не соответствует действительности. Поэтому UPGMA используется редко, кроме случаев, когда известно, что молекулярные часы работают.
 - **Метод соседнего соединения (Neighbor-Joining, NJ):**
 - Более продвинутый метод, который также является аггломеративным.
 - Не предполагает постоянную скорость эволюции.
 - На каждом шаге выбирает два кластера, которые минимизируют общую длину ветвей дерева.
 - NJ является одним из наиболее популярных дистанционных методов, так как он относительно быстрый и дает хорошие результаты, особенно когда последовательности не слишком сильно дивергировали.
 - **Метод наименьших квадратов (Least Squares):**
 - Сравнивает матрицу дистанций дерева с исходной матрицей дистанций, и стремится построить дерево, которое минимизирует разницу между ними (сумма квадратов разностей).
 - Более вычислительно сложный, чем NJ, но может давать более точные результаты, особенно если используется правильная модель эволюции.

4. Оценка достоверности дерева:

- После построения дерева важно оценить его достоверность. Это можно сделать с помощью:
 - **Бутстреп-анализа (Bootstrap analysis):** Создается множество (например, 1000) новых MSA путем случайного выбора колонок из исходного MSA с возвращением. Для каждого нового MSA строится дерево. Бутстреп-значение для каждой ветви - это доля деревьев, в которых эта ветвь встречается. Высокие бутстреп-значения (например, >70%) указывают на то, что ветвь хорошо поддержана данными.
 - **Другие методы:** Существуют и другие методы оценки достоверности дерева, такие как джекнайф (jackknife).

| | | |
|--|--|---|
| | | <p>Преимущества дистанционных методов:</p> <ul style="list-style-type: none"> • Скорость: Относительно быстры, что делает их пригодными для анализа больших наборов данных. NJ особенно быстр. • Простота: Концептуально простые. <p>Недостатки дистанционных методов:</p> <ul style="list-style-type: none"> • Потеря информации: Сводят информацию о последовательностях к попарным дистанциям, что приводит к потере информации, содержащейся в исходных данных. • Зависимость от выравнивания: Качество дерева сильно зависит от качества выравнивания последовательностей. Ошибки в выравнивании могут привести к неправильным дистанциям и, как следствие, к неверному дереву. • Чувствительность к выбору модели: Выбор модели эволюции для расчета дистанций может сильно влиять на результат. • Не позволяют учесть все возможные деревья: Методы NJ и UPGMA строят только одно дерево, в то время как другие методы, такие как максимальная правдоподобность и байесовские методы, позволяют оценить вероятность разных деревьев. |
|--|--|---|

Тестирование (Вариант 2).

| Индекс и содержание формируемой компетенции | Индикаторы компетенций | Тестовые задания для промежуточной аттестации |
|---|---|--|
| ОПК-4 Способен применять методы биоинженерии и биоинформатики для получения новых знаний и для получения биологических объектов с целенаправленно измененными свойствами, проводить анализ результатов и методического опыта исследования, определять практическую значимость исследования | <p><i>ИДК опк 4.1</i> Демонстрирует навыки использования методов биоинженерии и биоинформатики для получения новых фундаментальных знаний</p> <p><i>ИДК опк 4.2</i> Применяет методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными</p> | <p>Задание комбинированного типа с выбором одного или нескольких верных ответов из четырех предложенных с аргументацией выбора</p> <p>Вопрос 1. Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> <p>Какие задачи решает биоинформатика?</p> <p>a) Анализ геномных данных b) Сравнительный анализ последовательностей c) Прямое редактирование ДНК d) Прогноз структуры белков</p> <p>Ответ _____</p> <p>Правильный ответ: a, b, d</p> <p>Аргументация: Биоинформатика работает с данными, а не с редактированием ДНК.</p> <p>Вопрос 2. Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</p> |

| | | |
|--|---|--|
| | <p>свойствами</p> <p>ИДК ОПК 4.3 Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования</p> | <p>Ключевым историческим этапом в развитии биоинформатики стало:</p> <p>a) Создание GenBank b) Создание CRISPR-Cas9 c) Изобретение микроскопа d) Разработка метода Эдмана</p> <p>Ответ _____</p> <p>Правильный ответ: a, d</p> <p>Аргументация: GenBank — база последовательностей, а метод Эдмана позволил секвенировать белки.</p> <p>Вопрос 3. <i>Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> |
| <p>ОПК-5 Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформатическими средствами анализа</p> | <p>ИДК ОПК 5.1 Использует информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию</p> | <p>К операционным системам, применяемым в биоинформатике, относятся:</p> <p>a) Windows b) Unix c) BIOS d) Android</p> <p>Ответ _____</p> <p>Правильный ответ: a, b</p> <p>Аргументация: Windows и Unix активно применяются, BIOS не является ОС, Android — не используется в биоинформатике.</p> <p>Вопрос 4. <i>Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> |
| | <p>ИДК ОПК 5.2 Умеет применять биоинформатические методы и полученные знания для анализа геномной, структурной и иной информации</p> | <p>Для получения почты в Интернете используется:</p> <p>a) POP b) FTP c) TELNET d) HTTP</p> <p>Ответ _____</p> <p>Правильный ответ: a</p> <p>Аргументация: POP (Post Office Protocol) предназначен для загрузки писем.</p> <p>Вопрос 5. <i>Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> |
| | <p>ИДК ОПК 5.3 Демонстрирует навыки владения основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p> | <p>Какие методы относятся к анализу экспрессии генов?</p> <p>a) Нозерн-блоттинг b) Вестерн-блоттинг</p> |

| | | |
|--|--|---|
| <p>ОПК-7: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности</p> | <p><i>ИДК</i> опк 7.1 Демонстрирует теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности</p> <p><i>ИДК</i> опк 7.2 Использует современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности.</p> | <p>c) SAGE d) ДНК-чибы Ответ _____ Правильный ответ: a, c, d Аргументация: Эти методы анализируют экспрессию генов. Вестерн — для белков.</p> <p>Вопрос 6. <i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>Какие методы применяются для изучения 3D-структуры белков?</p> <p>a) Рентгеноструктурный анализ b) ЯМР-спектроскопия c) Вестерн-блоттинг d) Электронная микроскопия Ответ _____ Правильный ответ: a, b Аргументация: Эти методы позволяют получить пространственные структуры белков.</p> <p>Вопрос 7. <i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>К каким базам данных относится UniProtKB?</p> <p>a) Первичные b) Вторичные c) Смешанные d) Библиографические Ответ _____ Правильный ответ: b Аргументация: UniProtKB — вторичная база данных белков.</p> <p>Вопрос 8. <i>Прочтайте вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.</i></p> <p>Какой язык применяется для работы с реляционными базами данных?</p> <p>a) HTML b) SQL c) XML d) PERL Ответ _____ Правильный ответ: b Аргументация: SQL используется для запросов к реляционным БД.</p> |
|--|--|---|

Вопрос 9.

Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Локальное выравнивание применяется для:

- a) Поиска наиболее сходных регионов
- b) Сравнения геномов целиком
- c) Поиска уникальных белков
- d) Определения консервативных доменов

Ответ _____

Правильный ответ: a, d

Аргументация: Локальное выравнивание выявляет сходные фрагменты и консервативные области.

Вопрос 10.

Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Какой инструмент основан на эвристических методах поиска сходства?

- a) BLAST
- b) Нидлман-Вунш
- c) Смит-Уотерман
- d) Muscle

Ответ _____

Правильный ответ: a

Аргументация: BLAST — быстрый эвристический инструмент для поиска сходства.

Вопрос 11.

Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Паралоги — это:

- a) Гены, образовавшиеся в результате дупликации в одном геноме
- b) Гены, возникшие у разных видов от общего предка
- c) Гены, перенесённые горизонтально
- d) Идентичные последовательности ДНК

Ответ _____

Правильный ответ: a

Аргументация: Паралоги возникают в результате дупликации.

Вопрос 12.

Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Какой метод построения филогенетических деревьев является кластерным?

- a) Neighbor-Joining (объединение соседей)

- b) Максимальная parsimonия
- c) Наибольшее правдоподобие
- d) Байесовский метод

Ответ _____

Правильный ответ: а

Аргументация: Neighbor-Joining — один из кластерных методов.

Вопрос 13.

Прочтите вопрос, выберите правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа.

Какая база данных чаще всего используется для анализа белковых структур?

- a) PDB
- b) GenBank
- c) Ensemble
- d) DDBJ

Ответ _____

Правильный ответ: а

Аргументация: PDB (Protein Data Bank) хранит данные о пространственной структуре белков.

Задание закрытого типа на установление соответствия

Вопрос 14.

Прочтите вопрос и установите соответствие между алгоритмом и его применением:

Алгоритм:

- a) Нидлман–Вунш
- b) Смит–Уотерман
- c) BLAST
- d) ClustalW

Применение:

- 1 — Глобальное выравнивание
- 2 — Локальное выравнивание
- 3 — Быстрый поиск сходных последовательностей
- 4 — Множественное выравнивание

Ответ _____

Правильный ответ:

- a — 1
- b — 2
- c — 3
- d — 4

Аргументация: Нидлман–Вунш — глобальное выравнивание, Смит–Уотерман — локальное, BLAST — быстрый поиск, ClustalW — множественное выравнивание.

Вопрос 15.

Прочтите вопрос и установите соответствие между термином и его определением:

Термин:
a) Ортологи
b) Паралоги
c) Ксенологи
d) Фенетический метод

Определение:

- 1 — Гены, возникшие после дупликации в геноме
- 2 — Гены, разделившиеся после видеообразования
- 3 — Гены, перенесённые горизонтальным переносом
- 4 — Метод построения деревьев на основе сходства признаков

Ответ _____

Правильный ответ:

- a — 2
- b — 1
- c — 3
- d — 4

Аргументация: ортологи — у разных видов, паралоги — дупликация, ксенологи — горизонтальный перенос, фенетика — анализ сходства.

Задание закрытого типа на установление последовательности

Вопрос 16.

Прочтите вопрос и установите последовательность шагов ПЦР в правильном порядке:

- a) Денатурация ДНК
- b) Отжиг праймеров
- c) Элонгация цепи
- d) Анализ продуктов

Ответ _____

Правильный ответ:

Правильная последовательность: a → b → c → d

Аргументация: стандартный цикл ПЦР: денатурация, отжиг, элонгация, после чего анализируется результат.

Вопрос 17.

Прочтите вопрос и установите последовательность шагов при при создании базы данных последовательностей:

- a) Секвенирование образца
- b) Аннотация полученных данных
- c) Загрузка данных в GenBank
- d) Использование информации исследователями

Ответ _____

Правильный ответ:

Правильная последовательность: a → b → c → d

Аргументация: сначала секвенируют, затем аннотируют, далее загружают в базу и предоставляют доступ.

Задание открытого типа с развернутым ответом

Вопрос 18.

Прочтите вопрос и запишите развернутый обоснованный ответ

Перечислите известные белковые базы данных?

Ответ:

- **Protein Data Bank (PDB):** Хотя PDB в основном содержит информацию о структурах белков, она также может включать информацию о нуклеотидных последовательностях, связанных с этими белками.
- **UniProt (Universal Protein Resource):** База данных белковых последовательностей и аннотаций.
- **Pfam:** база данных семейств белков, представленных в виде множественных выравниваний последовательностей и скрытых марковских моделей.
- **InterPro:** база данных, объединяющая информацию о семействах, доменах и функциональных сайтах белков из нескольких других баз данных.

Вопрос 19.

Прочтите вопрос и запишите развернутый обоснованный ответ

Сформулируйте принцип работы алгоритмов глобального выравнивания нуклеотидных и аминокислотных последовательностей?

Ответ:

Принцип работы алгоритмов глобального выравнивания, таких как алгоритм Нидлмана-Вунша, заключается в построении оптимального выравнивания двух входных последовательностей по *всей* их длине. Это означает, что алгоритм стремится выровнять каждую единицу каждой последовательности, даже если для этого необходимо добавление гэпов. Вот основные шаги:

1. **Матрица подсчета баллов (scoring matrix):**
 - Определяется матрица подсчета баллов, которая задает баллы за совпадения (match), несовпадения (mismatch) и гэпы (пропуски).
 - Как и в локальном выравнивании:
 - Для нуклеотидных: Обычно используются простые схемы: +1 за совпадение, -1 за несовпадение, гап - штраф.
 - Для аминокислотных: Используются более сложные матрицы, такие как BLOSUM или PAM, отражающие эволюционные взаимосвязи между аминокислотами (замена консервативной аминокислоты оценивается выше, чем замена на радикально отличающуюся).
 - Выбор матрицы сильно влияет на результат, и, как правило, требует подбора под конкретную задачу.
2. **Создание матрицы выравнивания (alignment matrix):**
 - Строится матрица выравнивания (также называемая матрицей динамического программирования) размерностью (длина sequenceA + 1) * (длина sequenceB + 1).
 - Первая строка и первый столбец инициализируются таким образом, чтобы отражать штрафы за введение гэпов с самого начала выравнивания. Например, значение ячейки (i, 0) равно i * gap_penalty, а значение ячейки (0, j) равно j * gap_penalty. Это основное отличие от локального выравнивания.
3. **Заполнение матрицы выравнивания:**

- Каждая ячейка матрицы (i, j) рассчитывается по следующей формуле:
- $\text{Score}(i, j) = \max \{$
- $\text{Score}(i-1, j-1) + \text{score}(\text{sequenceA}[i], \text{sequenceB}[j]),$ // Совпадение или несовпадение
- $\text{Score}(i-1, j) + \text{gap_penalty},$ // Гэп в sequenceA
- $\text{Score}(i, j-1) + \text{gap_penalty}$ // Гэп в sequenceB
- $\}$
 - $\text{score}(\text{sequenceA}[i], \text{sequenceB}[j])$ - балл из матрицы подсчёта баллов для выравнивания символов $\text{sequenceA}[i]$ и $\text{sequenceB}[j]$. Другими словами - "награда" за совпадение, или штраф за несовпадение.
 - gap_penalty - штраф за введение гэпа. Этот штраф обычно отрицательный.
- При расчете каждой ячейки выбирается тот вариант, который дает максимальный балл.

4. **Трассировка (backtracking):**

- После заполнения матрицы выбирается ячейка в *правом нижнем углу* матрицы (последняя строка и последний столбец). Значение в этой ячейке представляет собой наилучший счет глобального выравнивания.
- Начиная с этой ячейки, выполняется трассировка в обратном направлении, чтобы определить путь, который привёл к этому значению.
- На каждом шаге трассировки определяем, из какой ячейки $(i-1, j-1)$, $(i-1, j)$ или $(i, j-1)$ мы пришли в текущую ячейку, основываясь на том, какой вариант максимизировал формулу подсчета баллов.
- Трассировка продолжается до достижения ячейки $(0, 0)$.
- Путь трассировки определяет оптимальное глобальное выравнивание.

Ключевые особенности глобального выравнивания:

- **Инициализация первой строки и столбца:** Первая строка и столбец инициализируются штрафами за гэпы, чтобы гарантировать, что выравнивание начинается с начала обеих последовательностей.
- **Оптимальный путь:** Алгоритм ищет оптимальный путь через матрицу, приводящий к наилучшему глобальному выравниванию.
- **Трассировка до начала:** Трассировка начинается с правого нижнего угла матрицы и продолжается до верхнего левого угла.

Преимущества и недостатки глобального выравнивания:

- **Преимущества:**
 - Оптимальное выравнивание по всей длине: Гарантирует наилучшее возможное выравнивание, учитывая всею длину последовательностей.
 - Подходит для гомологичных последовательностей: Особенно полезен для последовательностей, которые, как ожидается, имеют сходное происхождение и структуру.
- **Недостатки:**
 - Чувствительность к различиям в длине: Плохо работает с последовательностями сильно отличающимися по длине, так как вынужден вводить большое количество гэпов.
 - Не подходит для поиска локальных сходств: Не предназначен для поиска консервативных доменов внутри более дивергентных последовательностей.

Вопрос 20.

Прочтите вопрос и запишите развернутый обоснованный ответ

Принцип работы метода максимального правдоподобия для реконструкции эволюционных деревьев в молекулярной филогении?

Ответ:

Метод максимального правдоподобия (Maximum Likelihood, ML) - это статистический метод для оценки параметров модели, который в молекулярной филогении используется для поиска наиболее вероятного филогенетического дерева, учитывая наблюдаемые данные (выравнивание последовательностей) и модель эволюции. В отличие от дистанционных методов, которые преобразуют данные в матрицу дистанций, ML метод работает непосредственно с выравниванием последовательностей.

Основные этапы метода максимального правдоподобия:

1. **Выравнивание последовательностей (Multiple Sequence Alignment - MSA):**

- Как и в других филогенетических методах, первым шагом является получение множественного выравнивания последовательностей (MSA). Качество выравнивания напрямую влияет на результат филогенетического анализа.

1. **Выбор модели эволюции:**

- **Ключевой аспект ML:** Необходимость выбора модели эволюции, описывающей процесс изменения последовательностей со временем.
- Модель определяет вероятности различных типов замен нуклеотидов (или аминокислот).
- Существуют различные модели, различающиеся по сложности и реалистичности:

○ **Модели нуклеотидной эволюции:**

- **Jukes-Cantor (JC69):** Самая простая модель, предполагающая, что все нуклеотидные замены равновероятны.
- **Kimura 2-parameter (K80):** Различает переходы (пурин-пурин или пиримидин-пиримидин) и трансверсии (пурин-пиримидин или наоборот), так как переходы происходят чаще.
- **Hasegawa-Kishino-Yano (HKY85):** Учитывает разные частоты нуклеотидов.
- **General Time Reversible (GTR):** Самая общая обратимая модель, позволяющая задавать разные скорости для всех типов замен.

○ **Модели аминокислотной эволюции:**

- Существуют специализированные модели для аминокислотных последовательностей (например, PAM, BLOSUM, WAG, LG, JTT). Они разрабатываются на основе анализа больших наборов данных и учитывают частоту различных аминокислот и их склонность к замене.

- **Оценка параметров модели:** Модель эволюции может содержать параметры (например, частоты нуклеотидов, скорости замен), которые необходимо оценить на основе данных.

- **Выбор подходящей модели:** Выбор модели - сложный процесс. Существуют статистические тесты (например, likelihood ratio test, AIC, BIC), которые позволяют сравнить различные модели и выбрать ту, которая лучше всего соответствует данным.

1. **Вычисление правдоподобия для каждого дерева:**

- **Определение правдоподобия:** Правдоподобие (Likelihood, L) - это вероятность наблюдаемых данных (т.е. MSA) при заданном дереве, модели эволюции и параметрах этой модели.

- **Расчет для каждого сайта:** Правдоподобие вычисляется для каждого сайта (колонки) в MSA, а затем перемножается по всем сайтам. Это связано с тем, что предполагается, что сайты изменяются независимо друг от друга.

- **Суммирование по всем возможным состояниям внутренних узлов:** Поскольку мы не знаем состояния последовательностей в предковых узлах дерева, необходимо просуммировать вероятности всех возможных состояний для каждого внутреннего узла. Это делается с использованием алгоритма Felsenstein's pruning algorithm (алгоритм обрезки Фельсенштейна), который эффективно вычисляет правдоподобие для заданного

дерева.

- **Необходимость перебора деревьев:** Поскольку число возможных деревьев растет экспоненциально с увеличением числа последовательностей, полный перебор всех деревьев невозможен для больших наборов данных. Поэтому используются эвристические алгоритмы для поиска дерева с максимальным правдоподобием.
- 1. **Поиск дерева с максимальным правдоподобием:**
- **Эвристические алгоритмы:** Используются для поиска дерева, максимизирующего функцию правдоподобия.
 - **Nearest Neighbor Interchange (NNI):** Начинается с произвольного дерева, затем ветви меняются местами, чтобы найти дерево с более высоким правдоподобием.
 - **Subtree Pruning and Refactoring (SPR):** Поддеревья обрезаются и переносятся в другое место дерева, чтобы улучшить правдоподобие.
 - **Tree Bisection and Reconnection (TBR):** Дерево делится на две части, а затем соединяется снова разными способами в поисках лучшего дерева.
- **Множественные запуски (Multiple runs):** Часто алгоритм запускается несколько раз с разными начальными деревьями, чтобы увеличить вероятность нахождения глобального максимума правдоподобия.
- **Выбор лучшего дерева:** Из всех найденных деревьев выбирается дерево с наибольшим значением правдоподобия.
- 1. **Оценка достоверности дерева:**
- **Бутстреп-анализ (Bootstrap analysis):** Как и в дистанционных методах, бутстреп-анализ является распространенным способом оценки достоверности дерева. Создается множество новых MSA путем случайного выбора колонок из исходного MSA с возвращением. Для каждого нового MSA строится дерево ML. Бутстреп-значение для каждой ветви - это доля деревьев, в которых эта ветвь встречается. Высокие бутстреп-значения (>70%) указывают на то, что ветвь хорошо поддержана данными.
- **Байесовский анализ:** Иногда метод максимального правдоподобия комбинируют с байесовским анализом для получения апостериорных вероятностей ветвей.

Преимущества метода максимального правдоподобия:

- **Статистически обоснованный:** ML основан на строгих статистических принципах.
- **Использование всей информации:** Использует всю информацию, содержащуюся в выравнивании последовательностей, а не только попарные дистанции.
- **Гибкость:** Позволяет использовать сложные модели эволюции, которые лучше отражают реальные процессы изменения последовательностей.

Недостатки метода максимального правдоподобия:

- **Вычислительная сложность:** ML требует больших вычислительных ресурсов, особенно для больших наборов данных.
- **Зависимость от модели:** Результат сильно зависит от выбора модели эволюции. Неправильный выбор модели может привести к неверным результатам.
- **Эвристический поиск:** Эвристические алгоритмы не гарантируют нахождения глобального максимума правдоподобия. Существует риск "застрять" в локальном максимуме.
- **Сложность в понимании и интерпретации:** Требует хорошего понимания статистических принципов и моделей эволюции.

Критерии оценивания тестового задания

| № | Тип задания | Критерии оценки | Результат оценивания |
|---|---|--|---|
| 1 | Задание закрытого типа на установление соответствия | Считается верным, если правильно установлены все соответствия (позиции одного столбца верно соотнесены с позициями другого столбца) | Полное совпадение с верным ответом – 1 балл Все остальные случаи – 0 баллов |
| 2 | Задание закрытого типа на установление последовательности | Считается верным, если правильно указана вся последовательность цифр | Полное совпадение с верным ответом – 1 балл Все остальные случаи – 0 баллов |
| 3 | Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора | Считается верным, если правильно указана цифра (буква) правильного ответа и приведены корректные аргументы, используемые при выборе ответа | Полное совпадение с верным ответом – 1 балл Все остальные случаи – 0 баллов |
| 4 | Задание комбинированного типа с выбором нескольких верных ответов из четырех предложенных и обоснованием выбора | Считается верным, если правильно указаны цифры (буквы) правильного ответа и приведены корректные аргументы, используемые при выборе ответа | Полное совпадение с верным ответом – 1 балл Все остальные случаи – 0 баллов |
| 5 | Задание открытого типа с развернутым ответом | Считается верным, если ответ совпадает с эталонным ответом по содержанию и полноте | Полное соответствие эталонному ответу – 1 балл Все остальные случаи – 0 баллов |

| Процент результативности | Оцениваемые компетенции | Оценка | |
|--------------------------|-------------------------|----------------|---------------------|
| | | Балл (отметка) | Вербальный аналог |
| 91 % - 100 % | ПК-1, ПК-2 | 5 | отлично |
| 71 % - 90 % | | 4 | хорошо |
| 51 % - 70 % | | 3 | удовлетворительно |
| 0 % - 50 % | | 2 | неудовлетворительно |

Разработчик:

Букин Ю.С. доцент Букин Ю.С.
(подпись)