

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
(АНОО ВО «УНИВЕРСИТЕТ «СИРИУС»)**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Алгоритмы в биоинформатике

Уровень образования: высшее образование – программа магистратуры
Направление подготовки: 06.04.01 Биология
09.04.03 Прикладная информатика
Направленность (профиль): Биоинформатика

1. Общая характеристика дисциплины (модуля)

1.1. Цель: формирование у студентов теоретических знаний и практических навыков по алгоритмам, необходимым для успешного решения биоинформатических задач.

1.2. Задачи: знание основных алгоритмов, применяемых в биоинформатике, понимание принципов поиска последовательностей в банке последовательностей, получение навыков по анализу вторичной структуры РНК.

1.3. Общая трудоемкость: 3 з.е.

1.4. Планируемые результаты обучения:

Формируемые компетенции (код компетенции, формулировка)	Планируемые результаты обучения по дисциплине (модулю) (индикаторы достижения компетенций)
ПК-2. Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных информационных технологий, для решения профессиональных задач в области биоинформатики, биоинженерии, биотехнологии и фарминдустрии	ИПК-2.1. Знает современные алгоритмы, средства разработки и программные средства, а также принципы написания программ на различных языках программирования
	ИПК-2.2. Осуществляет анализ и выбор методов решения профессиональных задач на основе теоретических знаний в области информационных технологий
	ИПК-2.3. Разрабатывает оригинальные алгоритмы и программные средства для решения профессиональных задач в области биоинформатики, биоинженерии, биотехнологии и фарминдустрии
ПК-5. Способен определять необходимые системные и программные средства для разработки и отладки прикладного программного обеспечения в современных специализированных программных комплексах, а также реализовывать в них новые алгоритмы	ИПК-5.1. Знает математические алгоритмы и принципы определения необходимых системных и программных средств для решения профессиональных задач
	ИПК-5.2. Определяет необходимые системные и программные средства для разработки и отладки прикладного программного обеспечения в современных специализированных программных комплексах
	ИПК-5.3. Реализует новые алгоритмы в современных специализированных программных комплексах

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 3 Листов 10
-------------------------------	---	---------------------

2. Структура и содержание дисциплины (модуля)

2.1. Объем дисциплины и виды учебной деятельности:

Виды учебной деятельности	2 семестр	Всего
Контактная работа обучающихся с преподавателем, всего ч.	60	60
Лекционные занятия, ч.	22	22
Практические (семинарские) занятия, ч.	34	34
Лабораторные занятия, ч.	х	х
Промежуточная аттестация – экзамен, ч	4	4
Промежуточная аттестация – зачет с оценкой, ч	х	х
Промежуточная аттестация – зачет, ч	х	х
Самостоятельная работа обучающихся, всего ч.	48	48
Общая трудоемкость, ч.	108	108
Общая трудоемкость, з.е.	3	3

2.2. Структура дисциплины (модуля) по разделам (темам) и видам учебной деятельности:

Наименования разделов (тем) дисциплины (модуля)	Лекционные занятия, ч	Практические (семинарские) занятия, ч	Лабораторные занятия, ч	Промежуточная аттестация, ч	Самостоятельная работа, ч	Всего, ч	Форма текущего контроля / промежуточной аттестации
Раздел 1. Введение. Сходство последовательностей	2	3	0		4	9	письменное домашнее задание
Раздел 2. Задача парного выравнивания. Локальное и глобальное выравнивание.	2	3	0		4	9	письменное домашнее задание
Раздел 3. Матрицы замен. PAM. BLOSUM.	2	3	0		5	10	письменное домашнее задание
Раздел 4. Поиск в базах. BLAST.	3	4	0		6	13	письменное домашнее задание
Раздел 5. Кластеризация последовательностей	3	4	0		6	13	письменное домашнее задание
Раздел 6. Множественное выравнивание	2	4	0		5	11	письменное домашнее задание

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 4 Листов 10
-------------------------------	---	---------------------

Раздел 7. Локальное выравнивание для NGS. Преобразование Барроуза-Уилера	3	4	0		6	13	письменное домашнее задание
Раздел 8. Задача сборки геномов. Контиг и прочтение. Выравнивание перекрытий. Методы на графах	3	5	0		6	14	письменное домашнее задание
Раздел 9. Аннотация геномов. НММ	2	4	0		6	12	письменное домашнее задание
Промежуточная аттестация				4		4	Экзамен
Итого	22	34	0	4	48	108	

2.3. Содержание разделов (тем) дисциплины (модуля):

Наименования разделов (тем) дисциплины (модуля)	Содержание разделов (тем) дисциплины (модуля)
Раздел 1. Введение. Сходство последовательностей	Понятие гомологии последовательностей биополимеров. Задача и мера сходства. Понятие алгоритма. Расстояние Левенштейна между последовательностями. Переход от рекурсии к рекуррентности. Динамическое программирование. Алгоритм Вагнера-Фишера. Штрафы за несовпадение как модель биологических замен. Расстояние между последовательностями и эволюционное расстояние. Модель нейтральной эволюции.
Раздел 2. Задача парного выравнивания. Локальное и глобальное выравнивание.	Задачи глобального, локального и полуглобального выравнивания, их биологический смысл и формальная постановка. Алгоритмы Нидлмана-Вунша и Смита-Уотермана. Проблема оптимальной модели инсерций и делеций. Линейная и аффинная модели пропусков.
Раздел 3. Матрицы аминокислотных замен. PAM. BLOSUM.	Понятие консервативности последовательностей. Неравноценность замен в первичной структуре белка. Понятие мутабельности аминокислот, оценка доли консервативных участков белка. Вероятность замены. Количество допустимых точечных замен. Единица PAM. Построение матрицы PAM. Задача учета вариабельности аминокислот при поиске в базах данных. Вероятность мутации и вероятность наблюдения пары аминокислот. Матрица BLOSUM и ее построение. Различные смысл и предназначение матриц PAM и BLOSUM.
Раздел 4. Поиск в базах. BLAST.	Биологическая задача аннотации последовательностей через поиск сходства с известными предположительными гомологами. Алгоритмическая задача быстрого поиска запроса в базе данных локальным выравниванием. «Google для биологов». Алгоритм BLAST. Статистическая и алгоритмическая сложность поиска в биологических последовательностях. Повторы, участки с низкой сложностью. Мера сложности последовательности и фильтрация

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 5 Листов 10
-------------------------------	---	---------------------

	<p>низкой сложности. Поиск пар-кандидатов (HSP). Статистическая значимость пары в последовательности по Гумбелю и значимость совпадения среди базы по Пуассону. E-значение как мат. ожидание. Меры идентичности и процента покрытия.</p>
<p>Раздел 5. Кластеризация последовательностей</p>	<p>Задача кластеризации биологических последовательностей. Двоичная кластеризация. Проблема пересчета эволюционных расстояний и подход пропорционального взвешивания ветвей. Ультраметричность кластеров гипотеза молекулярных часов. Алгоритм UPGMA. Дендрограмма. Корень дерева. Алгоритм ближайшего соседа (NJ). Неукорененное дерево.</p>
<p>Раздел 6. Множественное выравнивание</p>	<p>Задача множественного выравнивания. Критерий оптимальности множественного выравнивания. Сведение задачи множественного выравнивания к задаче парного выравнивания кластеров. Подход прогрессивного выравнивания. Модификация алгоритма Нидлмана-Вунша для парного выравнивания кластеров мощностью более 1. Порядок прогрессивного выравнивания кластеров последовательностей по известному дереву кластеризации.</p>
<p>Раздел 7. Локальное выравнивание для NGS. Преобразование Барроуза-Уилера</p>	<p>Биологическая задача быстрого полуглобального выравнивания коротких последовательностей на одну длинную. Понятие суффиксного массива. Использование суффиксного массива для бинарного поиска короткой подстроки. Понятие вращения строки. Построение преобразования Барроуза-Уилера (Burrows & Wheeler). FM-индекс (Ferragina & Manzini). Алгоритм поиска подстроки с использованием FM-индекса. Реализация алгоритма в инструменте BWA (<i>bwa-bt</i>). Поиск длинных подстрок. Алгоритм BWA-SW.</p>
<p>Раздел 8. Задача сборки геномов. Контиг и прочтение. Выравнивание перекрытий. Методы на графах</p>	<p>Задача сборки – восстановление последовательности по прочтениям секвенирования. Понятие контига и скаффолда. Технические источники сложности задачи сборки: распределение длин фрагментов, распределение ошибок секвенирования, обогащение библиотек. Алгоритмы выравнивания перекрытий – динамическое программирование и суффиксное дерево. Граф перекрытий и его раскладка. Подход «перекрытие-раскладка-консенсус» (<i>overlap-layout-consensus</i>). Ограничения подхода. Проблема восстановления повторов при сборке. Проблема вычислительных ресурсов. Граф подслов де Брёйна и инструменты на его основе. Граф де Брёйна как эйлеров граф. Алгоритмы построения графа де Брёйна и сборки с его помощью. Ограничения подхода на графах де Брёйна. Сборка сверхдлинных прочтений одномолекулярного секвенирования. Инструмент <i>Flye</i> и граф повторов. Использование графа повторов для разрешения повторенных участков в восстанавливаемой последовательности.</p>

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 6 Листов 10
-------------------------------	---	---------------------

Раздел 9. Аннотация геномов. Марковские модели последовательностей. НММ	Задача аннотации функциональных участков геномных последовательностей. Статистическая модель последовательности как стохастического процесса. Условие марковости. Гипотеза марковости и марковская модель последовательности. Явная марковская цепь. Аннотация участков с помощью явных марковских моделей. Калибровка и ограничения подхода явных марковских моделей. Скрытые состояния, их расшифровка как подход к аннотации с точностью до нуклеотида. Скрытые марковские модели последовательностей (НММ). Алгоритм Витерби. Современные марковские модели генов эукариот.
---	---

2.4. Самостоятельная работа

Самостоятельная работа предусматривает: самостоятельное изучение теоретического материала, подготовку к ответам на семинарских заданиях, подготовку к текущему контролю и промежуточной аттестации, выполнение тестовых заданий по пройденным темам курса.

3. Текущий контроль и промежуточная аттестация по дисциплине (модулю). Оценочные материалы

3.1. Текущий контроль успеваемости проводится в течение семестра в следующих формах:

Наименования разделов (тем) дисциплины (модуля)	Форма текущего контроля	Оценочные материалы
Раздел 1. Введение. Сходство последовательностей	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 2. Задача парного выравнивания. Локальное и глобальное выравнивание.	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 3. Матрицы замен. PAM. BLOSUM	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 7 Листов 10
-------------------------------	---	---------------------

Раздел 4. Поиск в базах. BLAST.	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 5. Кластеризация последовательностей	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 6. Множественное выравнивание	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 7. Локальное выравнивание для NGS. Преобразование Барроуза-Уилера	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 8. Задача сборки геномов. Контиг и прочтение. Выравнивание перекрытий. Методы на графах	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования
Раздел 9. Аннотация геномов. Марковские модели последовательностей. HMM	Устный опрос, проектные работы, контрольные работы, тестирование	Перечень вопросов для опроса и тем для групповых и индивидуальных проектов, комплект заданий для контрольных работ и тестирования

3.2. Оценочные материалы для текущего контроля

Примерный перечень заданий для контрольных работ:

1. Расстояние В.И. Левенштейна.
2. Понятие глобального выравнивания.
3. Понятие локального выравнивания.

4. Построение матрицы замен PAM1.
5. Построение матрицы замен BLOSUM. BLOSUM62
6. Алгоритм локального поиска в базах последовательностей BLAST
7. Алгоритм преобразования Барроуза-Уилера, FM-индекс Феррагины и Манцини (Ferragina & Manzini)
8. Алгоритм поиска подстрок в строке с использованием FM-индекса.
9. Задача сборки последовательности из перекрывающихся прочтений.
10. Алгоритмы выравнивания перекрытий – динамическое программирование и суффиксное дерево.
11. Граф перекрытий. Подход «перекрытие-раскладка-консенсус» (*overlap-layout-consensus, OLC*).
12. Граф подслов де Брёйна. Алгоритм построения. Алгоритм восстановления исходной последовательности обходом по эйлерову пути.
13. Сборка длинных прочтений.
14. Задача классификации функциональных участков геномных последовательностей.
15. Задача аннотации функциональных участков.

3.3. Формой промежуточной аттестации является экзамен.

Результаты промежуточной аттестации оцениваются как «отлично», «хорошо», «удовлетворительно» и «не удовлетворительно».

Оценка «отлично», «хорошо», «удовлетворительно» означает успешное прохождение промежуточной аттестации.

3.4. Оценочные материалы для промежуточной аттестации

Перечень вопросов для подготовки к экзамену:

1. Расстояние В.И. Левенштейна. Алгоритм Вагнера-Фишера (Wagner & Fisher)
2. Понятие глобального выравнивания. Алгоритм Нидлмана-Вунша (Needleman & Wunsch)
3. Понятие локального выравнивания. Алгоритм Смита-Уотермана (Smith & Waterman).
4. Построение матрицы замен PAM1. Произвольное количество PAM
5. Построение матрицы замен BLOSUM. BLOSUM62
6. Алгоритм локального поиска в базах последовательностей BLAST
7. Алгоритм преобразования Барроуза-Уилера, FM-индекс Феррагины и Манцини (Ferragina & Manzini)
8. Алгоритм поиска подстрок в строке с использованием FM-индекса.

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Алгоритмы в биоинформатике»	Лист 9 Листов 10
-------------------------------	---	---------------------

9. Задача сборки последовательности из перекрывающихся прочтений. Выравнивание перекрытия. Контиг, синглтон, скаффолд, хромосома.

10. Алгоритмы выравнивания перекрытий – динамическое программирование и суффиксное дерево.

11. Граф перекрытий. Подход «перекрытие-раскладка-консенсус» (*overlap-layout-consensus, OLC*). Проблема повторов. Построение консенсуса

12. Граф подслов де Брёйна. Алгоритм построения. Алгоритм восстановления исходной последовательности обходом по эйлерову пути. Применение для сборки. Проблема неравномерного секвенирования.

13. Сборка длинных прочтений. Граф повторов и его применение для разрешения повторов. Схема работы инструмента Flye.

14. Задача классификации функциональных участков геномных последовательностей. Явная марковская модель последовательности как стохастического процесса. Алгоритм классификации последовательностей (на примере CpG-островков).

15. Задача аннотации функциональных участков. Скрытые марковские модели последовательностей (НММ). Алгоритм Витерби.

4. Учебно-методическое и информационное обеспечение дисциплины (модуля)

4.1. Перечень основной литературы:

Стефанов, В. Е. Биоинформатика / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. — Москва : Издательство Юрайт, 2025. — 249 с.

4.2. Перечень дополнительной литературы:

1. Compeau P, Pevzner P. Bioinformatics Algorithms / Active Learning Publishers, 2018.

2. Mount D. Bioinformatics: Sequence and Genome Analysis / Cold Spring Harbor Laboratory Press, 2004.

3. Dasgupta S, Papadimitriou CH & Vazirani M, Algorithms / McGraw-Hill Education, 2016.

4. D. MacLean, R Bioinformatics Cookbook: Use R and Bioconductor to perform RNAseq, genomics, data visualization, and bioinformatic analysis.

4.3. Перечень современных профессиональных баз данных и ресурсов информационно-телекоммуникационной сети «Интернет»:

1. SEQanswers [Электронный ресурс] <http://seqanswers.com/> - информационный ресурс (дискуссионный форум) о методах высокопроизводительного секвенирования.

5. Материально-техническое и программное обеспечение дисциплины (модуля)

5.1. Материально-техническое обеспечение:

Вид аудитории	Технические средства и оборудование
<i>Учебная аудитория для проведения лекционных занятий</i>	Альфа 5.1 - учебная аудитория для проведения учебных занятий, предусмотренных программой магистратуры. Доска магнитно-маркерная поворотная BoardSYS Twist 100x160 ПО-15Ф 1 шт. Флипчарт 70*100 на роликах 1 шт. Стол-кафедра 1 шт. Стол аудиторный 1 шт. Столы-трансформеры Summa GA ученические 25 шт. Стулья на колесах ученические 25 шт. Ноутбук HP 1 шт. Интерактивная панель NexTouch Nextpanel 86” 1 шт. Радиосистема Arthur Forty U-9700C PSC (UHF) в комплекте. Акустическая система Behringer B215D 2 шт. Веб-камера 4К с технологией искусственного интеллекта JazzTel JT-Vintage-4K 1 шт. Комплект электронных презентаций.
<i>Учебная аудитория для проведения практических занятий</i>	Дельта 2.3 – учебная аудитория для проведения практических занятий (компьютерный класс). Доска магнитно-маркерная поворотная BoardSYS Twist 100x160 ПО-15Ф 1 шт. Флипчарт 70*100 на роликах 1 шт. Стол преподавателя аудиторный 1 шт. Столы и стулья ученические 45 шт. Компьютеры Lenovo ThinkCentre M920s SFF в комплекте с мониторами IIYAMA 27” и периферией – 45 шт. Интерактивная панель NexTouch Nextpanel 86” 1 шт. Радиосистема Arthur Forty U-9700C PSC (UHF) в комплекте. Акустическая система Behringer B215D 2 шт. Веб-камера 4К с технологией искусственного интеллекта JazzTel JT-Vintage-4K 1 шт. Комплект электронных презентаций.

5.2. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе российского производства:

- компьютерные системы BioUML;
- COPASI;
- пакет программ CobraPy;
- MassPy.