Приложение № 4
к образовательной программе высшего образования – программе магистратуры
по направлению подготовки 06.04.01 Биология
(направленность (профиль) «Генетика и генетические технологии»)

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

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат: 7835bcdac1c89a3da835a18f6045494f3f404337
Владелец: Некрасова Людмила Андреевна
Действителен: с 08.07.2026 до 08.07.2027

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Алгоритмы в биоинформатике»

1. Трудоемкость дисциплины (модуля): 3 з.е.

2. Место дисциплины (модуля) в учебном плане:

Обязательная часть «Профессиональной подготовки» в рамках Блока 1. Дисциплины (модули), реализуется во 2 семестре.

3. Цель дисциплины (модуля): подготовка обучающихся к осмысленному применению программных инструментов вычислительной биологии, а также навыки разработки собственных инструментов или модификации существующих.

4. Задачи дисциплины (модуля):

- дать основные представления о подходах, методах и алгоритмах анализа биологических последовательностей;
- дать основные представления об анализе структур биополимеров;
- дать основные представления о системной биологии.

5. Перечень разделов (тем) дисциплины (модуля) и их краткое содержание:

Наименование раздела (темы) дисциплины (модуля)	**Краткое содержание**
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.

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. Алгоритм BLAST.
Раздел 5. Кластеризация последовательностей	Задача кластеризации биологических последовательностей. Двоичная кластеризация. Ультраметричность кластеров гипотеза молекулярных часов. Алгоритм UPGMA. Дендрограмма. Корень дерева. Алгоритм ближайшего соседа (NJ). Неукорененное дерево.
Раздел 6. Множественное выравнивание	Задача множественного выравнивания. Критерий оптимальности множественного выравнивания. Сведение задачи множественного выравнивания к задаче парного выравнивания кластеров. Подход прогрессивного выравнивания.
Раздел 7. Локальное выравнивание для NGS. Преобразование Барроуза-Уилера	Биологическая задача быстрого полуглобального выравнивания коротких последовательностей на одну длинную. Понятие суффиксного массива. Алгоритм поиска подстроки с использованием FM-индекса. Реализация алгоритма в инструменте BWA (bwa-bt). Поиск длинных подстрок. Алгоритм BWA-SW.
Раздел 8. Задача сборки геномов. Контиг и прочтение. Выравнивание перекрытий. Методы на графах	Задача сборки – восстановление последовательности по прочтениям секвенирования. Понятие контига и скаффолда. Технические источники сложности задачи сборки: распределение длин франментов, распределение ошибок секвенирования, обогащение библиотек. Алгоритмы выравнивания перекрытий – динамическое программирование и суффиксное дерево.
Раздел 9. Аннотация геномов. Марковские модели последовательностей. HMM	Задача аннотации функциональных участков геномных последовательностей. Статистическая модель последовательности как стохастического процесса. Современные марковские модели генов эукариот.

6. Образовательные результаты освоения дисциплины (модуля):

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
ПК-5. Способен использовать современные информационные технологии, в том числе специализированное программное обеспечение, для решения прикладных задач в научно-исследовательской (проектно-исследовательской) деятельности	ИПК-5.1. Знает принципы написания программ на различных языках программирования
	ИПК-5.3. Использует современные информационные технологии, в том числе специализированное программное обеспечение, для проведения биоинформатического анализа больших объемов геномных данных