



## МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

ФГБОУ ВО «ИГУ»

**Кафедра физико-химической биологии, биоинженерии и биоинформатики**



### Рабочая программа дисциплины

Наименование дисциплины: 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 биолого-почвенного  
факультета  
Протокол № 4 от 20.04.2024  
Председатель \_\_\_\_\_ А. Н. Матвеев

Рекомендовано кафедрой физико-химической  
биологии, биоинженерии и биоинформатики  
Протокол № 15 от 17.04.2024  
Зав. кафедрой \_\_\_\_\_ В.П. Саловарова

Иркутск 2024 г.

## Содержание

	стр
I. Цель и задачи дисциплины .....	3
II. Место дисциплины в структуре ОПОП .....	3
III. Требования к результатам освоения дисциплины .....	3
IV. Содержание и структура дисциплины .....	5
4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов .....	5
4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине .....	6
4.3 Содержание учебного материала .....	7
4.3.1 Перечень семинарских, практических занятий и лабораторных работ .....	10
4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов .....	10
4.4. Методические указания по организации самостоятельной работы студентов .....	11
4.5. Примерная тематика курсовых работ (проектов) .....	11
V. Учебно-методическое и информационное обеспечение дисциплины .....	12
а) перечень литературы .....	12
б) периодические издания .....	12
в) список авторских методических разработок .....	15
г) базы данных, поисково-справочные и информационные системы.....	15
VI. Материально-техническое обеспечение дисциплины .....	14
6.1. Учебно-лабораторное оборудование .....	14
6.2. Программное обеспечение .....	14
6.3. Технические и электронные средства обучения .....	15
VII. Образовательные технологии .....	15
VIII. Оценочные материалы для текущего контроля и промежуточной аттестации .....	15

### I. Цель и задачи дисциплины:

**Цель:** Ознакомить студентов с современными представлениями о предмете, объектах, основных концепциях информационной биологии, методах и алгоритмах получения, представления и анализа данных в биоинформатике.

#### **Задачи:**

- рассмотреть основополагающие сведения о содержании и возможностях информационной биологии (биоинформатики);
- изучить понятийный аппарат и методологическую базу информационной биологии;
- освоить на практике базовые методы биоинформатики, включая работу с молекулярными базами данных, выравнивание последовательностей и молекулярную визуализацию;
- изучить возможности приложения методов информационной биологии, в том числе, теоретического анализа и компьютерного моделирования, к решению фундаментальных и прикладных проблем современной биологии, медицины, фармакологии и экологии;

- сформировать навыков использования сетевых технологий для эффективного поиска, передачи и обработки научной информации.

## **II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО**

2.1. Учебная дисциплина Б1.О.42 «Биоинформатика» относится к обязательной части учебного плана,. Изучается на 3 курсе в шестом семестре.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами специалитета («Математика», «Химия», «Общая биология», «Биохимия»).

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: «Биотехнология», «Большой практикум», «Геномные и постгеномные технологии», «Современные биомедицинские технологии», выполнение ВКР.

## **III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

Процесс освоения дисциплины направлен на формирование компетенции в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки 06.05.01 «Биоинженерия и биоинформатика»:

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 биоинформатическими средствами анализа;.

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

**Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций**

<b>Компетенция</b>	<b>Индикаторы компетенций</b>	<b>Результаты обучения</b>
<b>ОПК-4</b> 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 ОПК 4.1</i> 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
	<i>ИДК ОПК 4.2</i> 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.
	<i>ИДК ОПК 4.3</i> Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования	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
<b>ОПК-5</b> Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформатическими средствами анализа;	<i>ИДК ОПК 5.1</i> Использует информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию	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
	<i>ИДК ОПК 5.2</i> Умеет применять биоинформатические методы и полученные знания для анализа геномной, структурной и	Знать: биоинформатические методы и полученные знания для анализа геномной, структурной и иной информации Уметь: применять биоинформатические методы и полученные знания для анализа геномной, структурной и иной

	иной информации	информации Владеть: биоинформатическими методами для анализа геномной, структурной и иной информации
	<i>ИДК ОПК 5.3</i> Демонстрирует навыки владения основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.	Знать: основные биоинформатические средства анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей. Уметь: применять основные биоинформатические средства анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей. Владеть: основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.
<i>ОПК-7:</i> Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности;	<i>ИДК ОПК 7.3</i> Демонстрирует теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности	Знать: теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности Уметь: применять теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности Владеть: теоретическими и практическими навыками использования современных информационных технологий в области профессиональной деятельности
	<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.

#### IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 4 зачетных единицы, 144 часа, в том числе 17 часов на экзамен.

Из них реализуется с использованием электронного обучения и дистанционных образовательных технологий не менее 20% часов от аудиторной работы (14 часов)

Форма промежуточной аттестации: экзамен.

4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/тема	Семестр	Всего часов	Из них практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся , практическую подготовку и трудоемкость (в часах)				Форма текущего контроля успеваемости/ Форма промежуточной аттестации (по семестрам)
					Контактная работа преподавателя с обучающимися			Самостоятельн ая работа	
					Лекция	Семинар/ Практическое, лабораторное занятие/	Консультация		
1	2	3	4	5	6	7	8	9	10
1	Тема 1. Введение. История, предмет и значение биоинформатики	6	12		2	8		2	Контрольные вопросы
2	Тема 2. Основные инструменты биоинформатики.	6	14		2	8		4	Контрольные вопросы Практические задания
3	Тема 3. Лабораторные методы структурно-функциональных исследований нуклеиновых кислот и белков	6	14		2	8		4	Контрольные вопросы Доклады
4	Тема 4. Базы данных в биоинформатике	6	20		2	8		10	Контрольные вопросы Практические задания

<b>5</b>	Тема 5. Выравнивание последовательностей	6	27,5		4	8	0,5	15	Контрольные вопросы Практические задания
<b>6</b>	Тема 6. Гомология, филогения и эволюционные деревья	6	27,5		4	8	0,5	15	Контрольные вопросы Практические задания

#### 4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
6	Тема 1. Введение. История, предмет и значение биоинформатики	Подготовка к опросу	1	2	Контрольные вопросы	Раздел 5 а-г
6	Тема 2. Основные инструменты биоинформатики.	Подготовка к опросу Выполнение отчетов по практическим заданиям	2	4	Контрольные вопросы Отчет по ПР	- « -
6	Тема 3. Лабораторные методы структурно-функциональных исследований нуклеиновых кислот и белков	Подготовка к опросу Подготовка докладов	3-4	4	Контрольные вопросы Доклад	- « -
6	Тема 4. Базы данных в биоинформатике	Подготовка к опросу Выполнение отчетов по практическим заданиям	5-8	10	Контрольные вопросы Отчет по ПР	- « -
6	Тема 5. Выравнивание последовательностей	Подготовка к опросу Выполнение отчетов по практическим заданиям	9-13	15	Контрольные вопросы Отчет по ПР	- « -
6	Тема 6. Гомология, филогения и эволюционные деревья	Подготовка к опросу Выполнение отчетов по практическим заданиям Подготовка к промежуточному тестированию	14-16	15	Контрольные вопросы Отчет по ПР Тест	- « -
Общий объем самостоятельной работы по дисциплине (час) – 50						
Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий (час) - 20						

## 4.3 Содержание учебного материала

### **Тема 1. Введение. История, предмет и значение биоинформатики**

Биоинформатика как информационные технологии в приложении к управлению биологическими данными и их анализу. Геномика и протеомика. Предпосылки возникновения и развития биоинформатики. Развитие методов расшифровки последовательностей биополимеров – исторический аспект. Работы Ф. Сэнгера и Эдмана. Реакции обрыва цепи и химического расщепления. Полимеразная цепная реакция. Технологии автоматизированной регистрации результатов секвенирования. Закон Мура и эффективность секвенирования. Динамика накопления информации в базах данных последовательностей. Проект «Геном человека».

Цели и задачи биоинформатики. Предмет биоинформатики. Прикладное значение биоинформатики: анализ гомологичности последовательностей; анализ экспрессии генов; разработка лекарственных препаратов; функции прогноза.

### **Тема 2. Основные инструменты биоинформатики.**

Компьютер и компьютерная программа. Программное обеспечение. Языки программирования - HyperText Markup Language, Java Script, Java, PERL (Practical Extraction and Reporting Language), BSML (Bioinformatic Sequence Markup Language), BIOML (Biopolymer Markup Language). Операционные системы - BIOS, DOS, Windows, Unix.

Интернет. Сетевые протоколы - UUCP (Unix to Unix Copy Protocol), POP (Post Office Protocol), FTP (File Transfer Protocol), TELNET (TELEcommunication NETwork), Transmission Control Protocol / Internet Protocol (TCP/IP), HTTP (Hyper Text Transfer Protocol). История современного Интернета. ARPAnet. BITnet. IP-адреса и доменные номера. Виды подключения к Интернет. Всемирная паутина (World Wide Web) как информационная система, построенная на принципе гиперсреды. Веб-страницы и веб-узлы. Объектная сеть. Object Request Broker (ORB). Зеркала и Интранет.

Программы-обозреватели (браузеры): Lynx, Mosaic, Netscape Navigator, Internet Explorer, Opera, Mozilla. Гиперссылки. URL — Uniform Resource Locator. HyperText Markup Language (HTML). Extensible Markup Language (XML). Database Management System (DBMS).

Наиболее важные биоинформационные WWW-ресурсы. EMBnet (European Molecular Biology net). Центры и узлы. HNational Center for Biotechnology Information (NCBI) и сервисная программа Entrez. Выборка и применение информации.

### **Тема 3. Лабораторные методы структурно-функциональных исследований нуклеиновых кислот и белков**

Цели анализа последовательностей. «Омик» - революция: новая парадигма в биологии. Геномика – структурная, функциональная и сравнительная. Протеом и протеомика. Значение геномики и протеомики для анализа последовательностей и структур.

Методы картографирования генома: генетические карты, физические карты хромосом и секвенсовые карты. Основные технические подходы к секвенированию ДНК. Ферменты рестрикции, создание и клонирование гибридных молекул ДНК, амплификацией генов *in vivo* и *in vitro*, кДНК-синтез. Секвенирование с обрывом цепи (метод Сенгера). Секвенирование последовательности клона. Ярлыки экспрессируемых последовательностей (EST) – назначение и принцип. Методы секвенирования белков – прямой и косвенный метод. Определение пространственной структуры белка. Практические методы - рентгеноструктурный анализ и ЯМР-спектроскопия.

Анализ экспрессии генов. Нозерн- и Вестерн-блоттинг. Серийный анализ экспрессии генов (SAGE - Serial Analysis of Gene Expression). ДНК – чипы. Анализ экспрессии белков. Двумерный электрофорез в полиакриламидном геле.

### **Тема 4. Базы данных в биоинформатике**



База данных (БД) - функции и классификация. Реляционные, иерархические и сетевые базы данных. Первичные, вторичные и смешанные базы данных. Избыточные и безизбыточные базы данных. Раритетные базы данных. Записи базы данных.

Система управления базами данных (СУБД). Компоненты и функции СУБД. Типы СУБД – иерархические и реляционные. Язык структурированных запросов (SQL).

Современные тенденции в структурировании БД. Обзор основных биоинформационных БД. Многоцелевые базы данных последовательностей нуклеиновых кислот. EMBL (European Molecular Biology Laboratory). DDBJ (DNA DataBank of Japan). GenBank. GSDB (Genome Sequence DataBase). Ensemble. Специализированные БД последовательностей НК: SGD (Saccharomyces Genome Database), TDB (TIGR DataBase), EST. Базы данных белковых последовательностей. PIR (International Protein Sequence Database). БД UniProt, ее структура. Базы данных структур. PDB (Protein data base). MSD (Macromolecular Structure Database). SCOP (Structure classification of Protein). CATH (Class / Architecture / Topology / Homology). NDB (Nucleic Acid Database), CSD (Cambridge Structural Database). BMRB (BioMagResBank). Вторичные базы данных. Библиографические базы данных.. Виртуальная библиотека.

### **Тема 5. Выравнивание последовательностей**

Основные понятия и определения. Выравнивание, его цели. Последовательность запроса и предметная последовательность. Счет подобия (выравнивания). Близость последовательностей. Типы выравнивания - глобальное и локальное. Отличительные особенности и область применения. Оптимальное и субоптимальное выравнивание. Общие принципы выравнивания. Критерии определения меры сходства. Понятие расстояния в информатике. Хеммингово расстояние и расстояние Левенштейна (редактирующее расстояние). Счет выравнивания и факторы, влияющие на него. Матрица процентов точечных мутаций (ПТМ). Процент точечных мутаций как единица эволюционного расхождения. Матрицы блочных замен аминокислот.

Методы попарного выравнивания последовательностей. Точечная матрица – принцип метода, область применения. Динамическое программирование. Алгоритмы Нидлмана-Вунша и Сми-та-Уотермена. Матрица переходов. Метод k-кортежей. Алгоритмы программы BLAST (Basic Local Alignment Search Tool). PSI-BLAST.

Множественное выравнивание последовательностей. Цели и задачи. «Золотой стандарт». Методы множественного выравнивания: прогрессивное множественное выравнивание; локальные множественные выравнивания; итерационные методы; вероятностно-статистические методы. ПО для множественного выравнивания. Программы семейства CLUSTAL. Muscle. T-COFFEE.

### **Тема 6. Гомология, филогения и эволюционные деревья**

Гомология и подобие. Ортологи, Паралоги и Ксенологи. Филогения и родство. Критерии филогенетического анализа: морфологические характеристики, биохимические свойства и последовательности нуклеиновых кислот и белков. Молекулярные подходы к определению филогении. Типы макромолекул, используемых для филогенетического анализа. Митохондриальная и рибосомная ДНК. Эволюционная значимость и область применения в филогении.

Филогенетические деревья. Теория графов, ее значение для изучения филогенетических отношений. Граф, узлы, дуги, путь. Ориентированные и взвешенные графы. Характерные свойства филогенетических деревьев. Топология деревьев. Монофилия, парафилия и полифелия. Основные подходы к филогенетическому анализу: фенетический, кладистический и эволюционный. Клада, таксон и узел. Проблема скорости эволюции.

Методы построения деревьев. Кластерные методы. Групповой метод невзвешенных пар с вычислением среднего арифметического. Алгоритм объединения соседей. Переборные методы. Метод максимальной парсимонии. Метод наибольшего правдоподобия. Метод Баесовой вероятности. Проверка правильности построения

деревьев: методы jackknife и bootstrap. ПО, реализующее разные методы анализа молекулярной филогении

#### 4.3.1. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции (индикаторы)
			Всего часов	Из них практическая подготовка		
1	2	3	4	5	6	7
1	Тема 1	Информационные технологии в биологии	4	4	Контрольные вопросы	ОПК-4 ОПК-5 ОПК-7
2	Тема 2	Структура и свойства белков и нуклеиновых кислот	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
3	Тема 3	Физико-химические методы исследования структуры биомолекул	4	4	Контрольные вопросы Доклад	ОПК-4 ОПК-5 ОПК-7
4	Тема 3	Физико-химические методы анализа экспрессии генов	4	4	Контрольные вопросы Доклад	ОПК-4 ОПК-5 ОПК-7
5	Тема 4	Первичная структура белка. БД UniProtKB	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
6	Тема 4	Пространственная структура белка. БД SCOP и CATH.	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
7	Тема 5	Выравнивание последовательностей с помощью программы BLAST	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
8	Тема 5	Поиск белка в БД NCBI по его гомологу	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
9	Тема 5	Библиографические базы данных.	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
10	Тема 5	Множественное выравнивание последовательностей.	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
11	Тема 6	Построение филогенетических деревьев по последовательностям отдельных белков	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7
12	Тема 6	Установление филогенетических связей по наборам последовательностей ДНК	4	4	Контрольные вопросы Отчет по ПР	ОПК-4 ОПК-5 ОПК-7

**4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы (СРС)**

<b>№ п/п</b>	<b>Тема</b>	<b>Задание</b>	<b>Формируемая компетенция</b>	<b>ИДК</b>
1	Тема 1. Введение. История, предмет и значение биоинформатики	1.3. Контрольные вопросы по теме (1-10)	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3 <b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3
2	Тема 2. Основные инструменты биоинформатики.	1. Подготовка отчетов по ПР (задание 1) 2. Контрольные вопросы по теме (11-20)	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3 <b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3
3	Тема 3. Лабораторные методы структурно-функциональных исследований нуклеиновых кислот и белков	1. Подготовка устных докладов (задание 2) 2. Контрольные вопросы по теме (21-30)	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3 <b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3
4	Тема 4. Базы данных в биоинформатике	1. Подготовка отчетов по ПР (задания 3 и 4). 2. Контрольные вопросы по теме (31-45).	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3 <b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3
5	Тема 5. Выравнивание последовательностей	1. Подготовка отчетов по ПР (задания 5, 6, 7, 8). 2. Контрольные вопросы по теме (46-60)	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3

				<b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3
6	Тема 6. Гомология, филогения и эволюционные деревья	1. Подготовка отчетов по ПР (задание 9, 10, 11). 2. Контрольные вопросы по теме (61-75)	ОПК-4 ОПК-5 ОПК-7	<b>ОПК-4</b> ИДК ОПК 4.1 ИДК ОПК 4.2 ИДК ОПК 4.3 <b>ОПК-5</b> ИДК ОПК 5.1 ИДК ОПК 5.2 ИДК ОПК 5.3 <b>ОПК-7</b> ИДК ОПК 7.1 ИДК ОПК 7.2 ИДК ОПК 7.3

#### 4.4. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студента преследует следующие цели:

- совершенствование навыков самообразовательной работы как основного пути повышения уровня образования;
- углубление и расширение знаний по предмету.

По дисциплине «Биоинформатика» предлагаются следующие формы самостоятельной работы:

- Работа над конспектом лекции;
- Углубленный анализ научно-методической литературы и изучение учебного материала, предусмотренного рабочей программой;
- Самостоятельное изучение отдельных тем, параграфов, не изложенных в лекции;
- подготовка к контрольному опросу на практических занятиях;
- подготовка отчетов по практическим заданиям;
- подготовка к экзамену.

*Письменные работы.* Для самостоятельного изучения тем рекомендуется использовать основную и дополнительную литературу, а также источники, найденные при помощи информационно-справочных и поисковых. Для закрепления материала рекомендуется делать краткие конспекты по теме.

*Устный доклад* – это сообщение в течение 10-15 мин, в котором студент в лаконичной форме должен изложить материал по соответствующей теме, придерживаясь следующего плана: введение, основная часть, заключение. Доклад сопровождается презентацией, отражающей основные положения по соответствующей теме, включающей наглядные материалы (схемы, таблицы, фото и т.д.). По окончании доклада студенту задают вопросы, как преподаватель, так и студенты, на которые докладчик должен дать исчерпывающие ответы.

#### *Рекомендации по подготовке презентации.*

Презентации — способ представления информации, сочетающий в себе текст, гипертекстовые ссылки, компьютерную анимацию, графики, видео, музыку и звуковой ряд, которые организованы в единую среду. Презентация имеет сюжет, сценарий и структуру, организованную для удобного восприятия информации. Отличительной особенностью презентации является её интерактивность, то есть создаваемая для пользователя возможность взаимодействия через элементы управления.

Презентация всегда состоит из двух основных компонентов: информации, которую выступающий хочет донести до аудитории, и манеры изложения. Написанный на бумаге текст помогает более четко и последовательно изложить материал. Презентации обычно

делают в PowerPoint, в Impress, либо в Acrobat. Желательно придерживаться принципа: один слайд - одна мысль. Титульный слайд должен содержать название презентации, её автора, контактную информацию автора. На втором слайде обычно представлен план презентации, основные разделы или вопросы, которые будут рассмотрены. Остальные слайды нужно строить по модели: тезис - аргументы – вывод. Выводы всегда должны быть даны ясно и лаконично на отдельном слайде. Предпоследний слайд должен содержать информацию об использованных источниках литературы, интернет-ресурсах. Последний слайд может повторять титульный с добавлением фразы «Спасибо за внимание!»

На слайды должны попасть только самые важные тезисы и данные, а также графический материал: диаграммы, рисунки, фотографии. Старайтесь делать слайды на однородном светлом фоне с более контрастным текстом. Ключевые слова в предложении лучше выделять жирным шрифтом или цветом. Текст пишите крупно, плотно набранный текст сложнее воспринимается.

#### *Содержание и форма отчета по практической работе*

Отчет по лабораторной работе должен включать следующие разделы:

1. НАЗВАНИЕ РАБОТЫ
2. ЦЕЛЬ И ЗАДАЧИ РАБОТЫ
3. МАТЕРИАЛЫ И МЕТОДЫ

В данном разделе приводятся перечень использованных в работе компьютерных программ, иных электронных ресурсов и баз данных; описание методик. Не следует включать материалы, не использованные в работе.

4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В данном разделе приводятся результаты работы в виде таблиц, рисунков и схем. Дается обсуждение результатов работы: адекватность результатов поставленным задачам, интерпретация результатов с позиции основных биологических теорий и т.д.

5. ВЫВОДЫ

ПОДПИСЬ, ДАТА

**4.5. Примерная тематика курсовых работ (проектов):** не предусмотрены учебным планом.

## **V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **а) перечень литературы**

1. Леск А. Введение в биоинформатику : пер. с англ. / А. М. Леск ; ред.: А. А. Миронов, В. К. Швядаса. - М. : Бином. Лаборатория знаний, 2009. - 318 с. - ISBN 978-5-94774-501-6 (8 экз.)+
2. Приставка А.А. Большой практикум по биоинженерии и биоинформатике. В 3 ч. Ч. 1. Белки : учеб.-метод. пособие / А.А. Приставка, В.П. Саловарова. – Иркутск : Изд-во ИГУ, 2013. – 121 с. - ISBN 978-5-9624-0962-7 (69 экз.)+
3. Белькова, Наталья Леонидовна Большой практикум по биоинженерии и биоинформатике [Текст] : учеб.-метод. пособие : в 3 ч. / Н. Л. Белькова ; рец.: С. М. Попкова, Ю. М. Константинов ; Иркут. гос. ун-т, Биол.-почв. фак. - Иркутск : Изд-во ИГУ, 2013 - . - 20 см. - ISBN 978-5-9624-0956-6. **Ч. 2** : Нуклеиновые кислоты. - 2014. - 155 с. : ил. - ISBN 978-5-9624-1184-2 – 39 экз.
4. Стефанов В.Е. Биоинформатика [Электронный ресурс] : учебник для вузов / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. - Электрон. текстовые дан. - Москва : Юрайт, 2022. - 252 с. - ЭБС "Юрайт". - Неогранич. доступ. - ISBN 978-5-534-00860-9+

5. Биология клетки. Физико-химические, структурно-функциональные и информационные основы [Текст] : учеб. пособие / Г. Ф. Жегунов [и др.] ; ред. Г. Ф. Жегунов. - 5-е изд., стер. - М. : Ленанд, 2018. - 542 с. - ISBN 978-5-9710-4976-0 +

**б) периодические издания**

«Математическая биология и Биоинформатика», «Биохимия», «Биофизика», «Вопросы биологической, медицинской и фармацевтической химии», «Известия РАН. Серия биологическая», «Молекулярная биология», «Прикладная биохимия и микробиология», «Вопросы вирусологии».

**в) список авторских методических разработок:**

1. Приставка А.А. Большой практикум по биоинженерии и биоинформатике. В 3 ч. Ч. 1. Белки : учеб.-метод. пособие / А.А. Приставка, В.П. Саловарова. – Иркутск : Изд-во ИГУ, 2013. – 121 с. - ISBN 978-5-9624-0962-7 (69 экз.)

**г) базы данных, информационно-справочные и поисковые системы**

1. <http://www.protein.bio.msu.ru/biokhimiya/index.htm> - Интернет-версия международного журнала по биохимии и биохимическим аспектам молекулярной биологии, биоорганической химии, микробиологии, иммунологии, физиологии и биомедицинских исследований. Статьи в pdf-формате.
2. <http://www.6years.net/index.php> - портал бесплатной медицинской информации, содержит большое количество книг, учебных пособий биохимической направленности.
3. <http://www.chemexper.com/> - поиск химических соединений в различных базах данных
4. <http://www.dmb.biophys.msu.ru> - Информационная система «Динамические модели в биологии», рассчитанная на широкий круг пользователей, включает в себя гипертекстовые документы и реляционные базы данных и обеспечивает унифицированный доступ к разнообразной информации по данной предметной области.
5. <http://www.elibrary.ru/defaultx.asp> - Научная электронная библиотека, крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты более 12 млн научных статей и публикаций.
6. <http://www.tusearch.blogspot.com> - Поиск электронных книг, публикаций, законов, ГОСТов на сайтах научных электронных библиотек. В поисковике отобраны лучшие библиотеки, в большинстве которых можно скачать материалы в полном объеме без регистрации. В список включены библиотеки иностранных университетов и научных организаций.
7. <http://www.biengi.ac.ru/analyz.htm> - Биоинформатика в Центре «Биоинженерия» РАН
8. <http://www.bioinformatix.ru/> - Биоинформатика, геномика, протеомика, биософт, имейджинг — портал по биоинформатике, имейджингу и биософту.
9. <http://www.ebi.ac.uk/> - база данных EMBL EBI (European Bioinformatics Institute).
10. <http://www.expasy.ch/> - система анализа белка Expasy (Expert Protein Analysis System, SwissProt, TrEMBL)
11. <http://www.iscb.org/> - Международное сообщество вычислительной биологии.
12. <http://www.matbio.org/> - электронный журнал «Математическая биология и биоинформатика».
13. <http://www.ncbi.nlm.nih.gov/> - сайт NCBI (National Center Biotech Information)
14. <http://www.ncbi.nlm.nih.gov/blast> - программа выравнивания последовательностей BLAST (Basic Local Alignment Sequence Tool)

15. <http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html> - база данных GenBank
16. <http://www.ncbi.nlm.nih.gov/pubmed> - библиографическая база данных PUBMED
17. <http://www.protein.bio.msu.ru/biokhimiya/index.htm> - Интернет версия международного журнала по биохимии и биохимическим аспектам молекулярной биологии, биоорганической химии, микробиологии, иммунологии, физиологии и биоинформатике. Статьи в pdf-формате.
18. <http://www.rcsb.org/pdb/> - база данных по белкам PDB (Protein 3D Structure database)
19. <http://www.rusbiotech.ru/> - Российские биотехнологии и Биоинформатика
20. molbiol.ru - российский сервер с большим количеством справочной информации по биоинформатике на русском языке.
21. ЭБС «Издательство Лань». Адрес доступа <http://e.lanbook.com/>
22. ЭБС «Руконт». Адрес доступа <http://rucont.ru/>
23. ЭБС «Айбукс». Адрес доступа <http://ibooks.ru>
24. ЭБС «Юрайт». Адрес доступа: <http://biblio-online.ru/>

## **VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **6.1. Учебно-лабораторное оборудование:**

Аудитория для проведения занятий практического типа. Аудитория оборудована: *специализированной (учебной) мебелью* на 12 посадочных мест; оборудована *техническими средствами обучения*: Проектор Epson EB-X03, Экран ScreenMedia, Доска аудиторная меловая, магнитная, Лаборатория орган химии - Шкаф вытяжной АФ-221" - 2 шт., Химический шкаф (стеллаж) - 1 шт., Лабораторный стол с выкатными тумбами – 5 шт., Холодильник «Минск» - 2шт., Аппарат для вертикального электрофореза – 1 шт., Вакуумный испаритель РВО-64 – 1 шт., Вольметр ВУ-15 – 1 шт., Дезинтегратор УД-20 – 1 шт., Измеритель ионных сопротивлений (импеданса) - 1 шт., Источник питания для электрофореза "Эльф" – 1 шт., Осциллограф универсальный двухлучевой С-55 – 1 шт., Термостат ТС-80 – 1 шт., Центрифуга К-24 – 1 шт., Центрифуга МПВ-310 – 1 шт. Ноутбук Lenovo G580 – 1 шт., весы аналитические HR-200 – 1 шт., весы лабораторные ОНАУС – 2 шт., рефрактометр ИРФ 454Б2М – 1 шт., рефрактометр УРП – 1 шт., фотоэлектрокалориметр КФ 77 – 1шт., центрифуга лабораторная ОПК-8 – 1 шт., центрифуга лабор-я, медицин-я, настольная ЦЛн 16 с микропроцес-ной системой управл – 1 шт., спектрофотометр СФ-2000, ферментер Minifors Spesco бактериальный – 1шт., термостат WB4MS водный /с перемешиванием/ - 1 шт., термостат ТС-1/80 СПУ – 1 шт., служащими для представления учебной информации по дисциплине «Биоинформатика» *учебно-наглядными пособиями*, обеспечивающими тематические иллюстрации по дисциплине в виде презентации.

Компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, организации самостоятельной работы. Аудитория оборудована: *специализированной (учебной) мебелью* на 20 посадочных мест, доской меловой; оборудована *техническими средствами обучения*: Системный блок PentiumG850, Монитор BenQ G252HDA-1 шт.; Системный блокAthlon 2 X2 250, Монитор BenQ G252HDA – 8 шт.; Системный блок PentiumD 3.0GHz, Монитор Samsung 740N – 3 шт.; Моноблок IRU T2105P – 2 шт.; Системный блок Pentium G3250, Монитор BenQG955 – 1 шт.; Системный блок Pentium G3250, Монитор BenQ GL2250 – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T200 HD – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T190N – 1 шт.; Системный блок tium G3250, Монитор Samsung 740N – 1 шт.; с неограниченным доступом к сети Интернет; Проектор BenQ MX503; экран ScreenVtdiaEcot. Ноутбук Lenovo G580 – 1 шт. С неограниченным доступом к сети Интернет и обеспечением доступа в электронную информационно-образовательную среду организации.

Помещения для хранения и профилактического обслуживания учебного оборудования. Аудитория оборудована: специализированной мебелью на 8 посадочных мест; Вытяжной шкаф – 1шт., Ламинарный шкаф – 2 шт., Термостат ТС-80 – 2 шт., Лабораторный стол металлический – 3 шт., Лабораторный стол с резиновой поверхностью – 2 шт., Холодильник «Атлант» – 1шт. Микроскоп монокулярный – 8 шт, Микроскоп "Биолам"-1 шт., Стерилизатор паровой ВК-75 ПТ "ТЗМОИ" – 1шт., Пипетка автоматическая Ленпипет 0,5-10 м"-1 шт., Пипетка-дозатор"-1 шт., Микроскоп Levenhuk D870T тринокуляр"-1 шт., Проектор Оверхед"-1 шт., Проектор View Sonic"-1 шт., Проектор View Sonic"-1 шт., Ноутбук Lenovo"-2 шт., Принтер Brother -1 шт., Принтер Canon -1 шт.

#### **6.2. Программное обеспечение:**

- DreamSpark Premium Electronic Software Delivery (3 years) Renewal (Windows 10 Education 32/64-bit (Russian) - Microsoft Imagine, Windows 7 Professional with Service Pack 1 32/64-bit (English) - Microsoft Imagine, Windows Server 2008 Enterprise and Standard without Hyper-V with SP2 32/64-bit (English) - Microsoft Imagine, Access 2016 32/64-bit (Russian) - Microsoft Imagine, Access 2010 32/64-bit (Russian) - Microsoft Imagine). Договор №03-016-14 от 30.10.2014г.
- Kaspersky Endpoint Security для бизнеса - Стандартный Russian Edition. 250-499. Форус Контракт №04-114-16 от 14ноября 2016г KES. Счет №РСЦЗ-000147 и АКТ от 23ноября 2016г Лиц.№1В08161103014721370444.
- Microsoft Office Enterprise 2007 Russian Academic OPEN No Level. Номер Лицензии Microsoft 43364238.
- Microsoft Windows XP Professional Russian Upgrade Academic OPEN No Level. Номер Лицензии Microsoft 41059241.
- Office 365 профессиональный плюс для учащихся. Номер заказа: 36dde53d-7cdb-4cad-a87f-29b2a19c463e.

#### **6.3. Технические и электронные средства:**

- Презентации по всем темам курса;
- Система электронного тестирования на базе образовательного портала Educa;
- Онлайн версии программ для выравнивания последовательностей и филогенетического анализа (BLAST, CLUSTAL, PhyML, T-Coffee, MUSCLE, COBALT)

### **VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ**

Для освоения дисциплины «Биоинформатика» применяются следующие образовательные технологии:

- *Информационная лекция* - это сжатое изложение основных научных фактов, что является базой для анализа рассуждений, оценок.
- *Лекция-визуализация*. Учит студентов преобразовывать устную и письменную информацию в визуальную форму, что формирует у них профессиональное мышление за счет систематизации и выделения наиболее значимых, существенных элементов содержания обучения. Задача преподавателя использовать такие формы наглядности, которые не только дополняют словесную информацию, но и сами являются носителями информации (схемы, рисунки, слайды-презентации, и т.п.). Этот вид лекции лучше всего использовать на этапе введения студентов в новый раздел, тему дисциплины.
- *Практические занятия* – это занятие, проводимое под руководством преподавателя в учебной аудитории, направленное на углубление научно-теоретических знаний и



овладение определенными методами самостоятельной работы, которое формирует практические умения. Одной из форм практических занятий является коллоквиум.

- *Самостоятельная работа студентов* (см. п.4.4).
- *Дистанционные образовательные технологии.* Под дистанционными образовательными технологиями понимаются образовательные технологии, реализуемые в основном с применением информационно-телекоммуникационных сетей. При освоении дисциплины «Биоинформатика» используется *компьютерные сетевые технологии* (интернет-технологии) – способ дистанционной передачи информации, основанный на использовании глобальных и локальных компьютерных сетей для обеспечения доступа обучающихся к информационным образовательным ресурсам и для формирования совокупности методических, организационных, технических и программных средств реализации и управления учебным процессом независимо от места нахождения его субъектов. Для организации дистанционного обучения на основе этих технологий используется специализированное программное средство - образовательный портал ИГУ ([educa.isu.ru](http://educa.isu.ru)).

## **VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

### ***Оценочные материалы для входного контроля***

Для входного контроля оценки уровня знаний студентов используется тестирование по основным разделам математики, физики, химии, биохимии и молекулярной биологии.

#### ***Демонстрационный вариант теста для входного контроля***

1. Какие связи образуют  $\alpha$ -спираль во вторичной структуре белка? а) Вандер-Ваальса; б) гидрофобные; в) пептидные; г) водородные
2. Если содержание остатков тимина (от общего числа остатков) ДНК составляет 20%, то содержание гуанина составит: а) 40%; б) 35%; в) 25%; г) 30%
3. Операционная система представляет из себя: а) комплекс программ специального назначения; б) комплекс аппаратных средств; в) совокупность ресурсов компьютера; г) комплекс инструментальных программ.

### ***Оценочные материалы текущего контроля***

Оценочные материалы текущего контроля формируются в соответствии с ЛНА университета. В рамках дисциплины «Биоинформатика» используются следующие формы текущего контроля:

- устный опрос;
  - выполнение отчетов по практическим работам
  - контроль самостоятельной работы.
- Фонд оценочных средств включает:

- контрольные вопросы;
- практические задания;
- перечень тем докладов;
- вопросы для самостоятельного изучения (СРС);
- перечень экзаменационных вопросов.

Назначение оценочных средств: выявить сформированность компетенции ОПК-4, ОПК-5, ОПК-7 (см. п. III). Студенты, не выполнившие задания текущего контроля или получившие за них оценку «не удовлетворительно», до промежуточной аттестации не допускаются, пока не будут ликвидированы все задолженности.

### Темы докладов (Задание 2)

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, внутриспредметных, 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).

Оценка *«отлично»*. 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.

Оценка *«хорошо»*. 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.

Оценка *«удовлетворительно»*. 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.

Оценка *«неудовлетворительно»*. 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.

## Демонстрационные практические задания для текущего контроля

### Задание 1.

На конце гена  $\beta$ -гемоглобина человека имеется следующая последовательность:

...ctggccsacaagtatcactaa

- Какова аминокислотная последовательность, соответствующая представленной?
- Напишите нуклеотидную последовательность, в которой единичная замена нуклеотида приводит к «молчащей» мутации в этом фрагменте.
- Напишите нуклеотидную последовательность, в которой единичная замена основания может привести к миссенс-мутации в этом фрагменте.
- Напишите нуклеотидную последовательность, в которой единичная замена основания приводит к преждевременной остановке синтеза белка.
- Напишите нуклеотидную последовательность, в которой единичная замена нуклеотида приводит к ошибке терминации и продолжению синтеза цепи.

### Задание 3.

Найдите в БД UniProtKB документ, содержащий информацию о белке, указано 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 4.

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

### Задание 5.

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 6.**

Мутацией, приводящей к серповидно-клеточной анемии, является единственная замена основания А на Т, приводящая к замене Glu → 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 7.**

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 8.**

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 9.**

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 10.**

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 11.**

Хордовых относят к группе вторичноротых, 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 вторичноротых.

*Критерии оценивания заданий:*

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;

Оценка «хорошо» выставляется, если задание выполнено правильно, но студент затрудняется изложить и обосновать алгоритм решения и / или интерпретировать полученные результаты;

Оценка «удовлетворительно» выставляется, если задание выполнено неправильно, но студент демонстрирует верный подход к проблеме, поставленной в задаче;

Оценка «неудовлетворительно» выставляется, если задание выполнено неправильно или не выполнено совсем.

## Контрольные вопросы для текущей аттестации

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 экспрессируемых 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, геномы 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 CATH?
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 выравниваемости 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 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 филогении? 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.

Ответ оценивается на «отлично», если студент: полно излагает изученный материал, дает правильное определение 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.

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. «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. Оценка «неудовлетворительно» ставится, если ответ не удовлетворяет требованиям положительной оценки или студент отказывается отвечать на контрольные вопросы

### ***Оценочные средства для промежуточной аттестации***

*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.

Фонд оценочных средств для промежуточной аттестации включает:

- тестовые задания для экзамена.

Назначение оценочных средств: выявить сформированность компетенций ОПК-4, ОПК-5, ОПК-7 (см. п. III).

Тестовое задание включает два варианта по 20 вопросов по всем темам курса. К тесту допускаются студенты, задавшие все домашние задания и получившие по каждому заданию зачет.

### **Критерий оценивания тестового экзаменационного задания**

#### **Система получения баллов за тестирование**

<b>№</b>	<b>Тип задания</b>	<b>Критерии оценки</b>	<b>Результат оценивания</b>
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	Считается верным, если правильно указаны цифры (буквы) правильного ответа и	Полное совпадение с верным ответом – 1 балл Все остальные случаи

	выбора	приведены корректные аргументы, используемые при выборе ответа	– 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

#### Критерий оценки за тестирование

Оценка	критерий
отлично	18 и более баллов
хорошо	16 – 17 баллов
удовлетворительно	15 – 13 баллов
неудовлетворительно	12 баллов и менее



## Оценочные материалы для промежуточной аттестации (экзамена)

### Тестирование (Вариант 1).

Индекс и содержание формируемой компетенции	Индикаторы компетенций	Тестовые задания для промежуточной аттестации
<b>ОПК-4</b> 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.	<b>ИДК ОПК 4.1</b> Демонстрирует навыки использования методов биоинженерии и биоинформатики для получения новых фундаментальных знаний	<b>Задание комбинированного типа с выбором одного или нескольких верных ответов из четырех предложенных с аргументацией выбора</b>  <b>Вопрос 1.</b> Что является предметом изучения биоинформатики? а) Хранение и анализ биологических данных б) Разработка методов секвенирования ДНК in vitro в) Прогноз структуры и функций белков г) Проведение генетических экспериментов на клеточном уровне Ответ _____ Правильный ответ: а, в Аргументация: Биоинформатика связана с анализом данных и моделированием, а не с экспериментальными манипуляциями.  <b>Вопрос 2.</b> Какое открытие оказало наибольшее влияние на развитие биоинформатики? а) Изобретение микроскопа б) Расшифровка структуры ДНК в) Создание метода ПЦР г) Разработка метода КРИСПР-Cas9 Ответ _____ Правильный ответ: б, в Аргументация: Открытие двойной спирали ДНК (1953) и ПЦР (1983) стали основой для секвенирования и хранения данных.  <b>Вопрос 3.</b> Какие языки программирования традиционно применяются в биоинформатике? а) PERL б) Python в) Java г) COBOL Ответ _____ Правильный ответ: а, б, в
	<b>ИДК ОПК 4.2</b> Применяет методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами.	
	<b>ИДК ОПК 4.3</b> Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования	
<b>ОПК-5</b>	<b>ИДК ОПК 5.1</b>	

<p>Способен находить и использовать информацию, накопленную в базах данных по биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформатическими средствами анализа.</p>	<p>Использует информацию, накопленную в базах данных по структуре геномов, белков и другую биологическую информацию</p> <p><i>ИДК ОПК 5.2</i> Умеет применять биоинформатические методы и полученные знания для анализа геномной, структурной и иной информации</p> <p><i>ИДК ОПК 5.3</i> Демонстрирует навыки владения основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p>	<p>Аргументация: PERL, Python и Java — широко применяются для анализа последовательностей и биоинформатического софта. COBOL в этой области не используется.</p> <p><b>Вопрос 4.</b> Какой протокол используется для передачи гипертекста в Интернете? а) FTP б) POP в) HTTP г) TCP/IP Ответ _____ Правильный ответ: г Аргументация: HTTP (HyperText Transfer Protocol) — стандарт передачи гипертекстовых документов.</p> <p><b>Вопрос 5.</b> Метод Сэнгера для секвенирования ДНК основан на: а) Использовании дидезоксинуклеотидов б) Флуоресцентной детекции в) Рестриктазном расщеплении ДНК г) Автоматическом чтении последовательности Ответ _____ Правильный ответ: а, б, г Аргументация: Метод Сэнгера использует ddNTP, автоматическое определение и метки. Рестриктазы применяются в других методах.</p> <p><b>Вопрос 6.</b> Какая техника применяется для анализа экспрессии белков? а) Вестерн-блоттинг б) Нозерн-блоттинг в) EST г) ДНК-чипы Ответ _____ Правильный ответ: а Аргументация: Вестерн-блоттинг применяется для выявления белков, остальные методы — для анализа РНК или ДНК.</p>
<p><i>ОПК-7.</i> Способен понимать принципы работы современных информационных технологий и</p>	<p><i>ИДК ОПК 7.3</i> Демонстрирует теоретические и практические навыки использования современных информационных</p>	<p><b>Вопрос 7.</b> Какая база данных относится к первичным? а) GenBank б) UniProtKB в) SCOP</p>

использовать их для решения задач профессиональной деятельности	технологий в области профессиональной деятельности	<p>d) Pfam          Ответ _____          Правильный ответ: a          Аргументация: GenBank хранит первичные нуклеотидные последовательности.</p> <p><b>Вопрос 8.</b>          Какая СУБД наиболее распространена в биоинформатике?          a) Иерархическая          b) Реляционная          c) Сетевая          d) Древовидная          Ответ _____          Правильный ответ: b          Аргументация: В большинстве биобаз данных используется реляционная модель.</p> <p><b>Вопрос 9.</b>          Что является целью выравнивания последовательностей?          a) Сравнение структур белков          b) Поиск гомологичных последовательностей          c) Определение экспрессии генов          d) Определение степени сходства нуклеотидов          Ответ _____          Правильный ответ: b, d          Аргументация: Выравнивание применяется для поиска сходства и гомологии в последовательностях.</p> <p><b>Вопрос 10.</b>          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.</p> <p><b>Вопрос 11.</b>          Ортологи — это:          a) Гены, возникшие в результате дупликации в одном геноме          b) Гены, разделившиеся после видообразования          c) Гены, перенесённые горизонтальным переносом</p>
	<i>ИДК ОПК 7.2</i> Использует современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности.	

		<p>d) Гены, кодирующие идентичные белки          Ответ _____          Правильный ответ: b          Аргументация: Ортологи — гены у разных видов, происходящие от общего предка.</p> <p><b>Вопрос 12.</b>          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.</p> <p><b>Вопрос 13.</b>          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.</p> <p><b>Задание закрытого типа на установление соответствия</b></p> <p><b>Вопрос 14.</b>          Установите соответствие между учёным/методом и его вкладом в биоинформатику:          a) Ф. Сэнгер          b) Э. Эдман          c) Кэри Муллис          d) GenBank          1 — Метод ПЦР          2 — Секвенирование ДНК методом обрыва цепи          3 — Секвенирование белков          4 — Первая база данных нуклеотидных последовательностей          Ответ _____          Правильный ответ:          a — 2</p>
--	--	---

		<p>b — 3 c — 1 d — 4</p> <p>Аргументация: Сэнгер предложил метод ddNTP, Эдман — метод деградации белков, Муллис — изобрёл ПЦР, GenBank — база нуклеотидных последовательностей.</p> <p><b>Вопрос 15.</b> Установите соответствие между базой данных и её назначением:</p> <p>a) PDB b) UniProtKB c) GenBank d) SCOP</p> <p>1 — Структуры белков 2 — Белковые последовательности 3 — Нуклеотидные последовательности 4 — Классификация белков по структуре</p> <p>Ответ _____</p> <p>Правильный ответ:</p> <p>a — 1 b — 2 c — 3 d — 4</p> <p>Аргументация: PDB хранит 3D-структуры, UniProtKB — белковые последовательности, GenBank — ДНК/РНК, SCOP — иерархическая классификация белков.</p> <p><b>Задание закрытого типа на установление последовательности</b></p> <p><b>Вопрос 16.</b> Расположите этапы секвенирования по Сэнгеру в правильной последовательности:</p> <p>a) Добавление праймера b) Включение дидезоксинуклеотидов (ddNTP) c) Репликация ДНК d) Электрофоретическое разделение фрагментов</p> <p>Ответ _____</p> <p>Правильный ответ: Правильная последовательность: a → c → b → d</p> <p>Аргументация: сначала праймер, затем синтез цепи, встраивание ddNTP, а в конце электрофорез.</p> <p><b>Вопрос 17.</b> Установите порядок действий при построении филогенетического дерева:</p>
--	--	---

		<p>a) Выравнивание последовательностей  b) Выбор модели эволюции  c) Расчёт матрицы расстояний  d) Построение дерева  Ответ _____  Правильный ответ:  Правильная последовательность: <math>a \rightarrow c \rightarrow b \rightarrow d</math>  Аргументация: сначала выравниваются последовательности, затем создаётся матрица различий, выбирается модель, и только потом строится дерево.</p> <p><b>Задание открытого типа с развернутым ответом</b></p> <p><b>Вопрос 18.</b>  Перечислите известные базы данных геномов и нуклеотидных последовательностей?  Ответ:</p> <ul style="list-style-type: none"> <li>• <b>NCBI GenBank (National Center for Biotechnology Information GenBank):</b> Одна из самых крупных и всеобъемлющих баз данных нуклеотидных последовательностей, поддерживаемая NCBI (США). Включает GenBank, RefSeq и Trace Archive.</li> <li>• <b>EMBL-EBI ENA (European Nucleotide Archive):</b> Европейский аналог GenBank, поддерживаемый EMBL-EBI (Европейский институт биоинформатики). Включает базы данных EMBL, Trace Archive и Sequence Read Archive (SRA).</li> <li>• <b>DDBJ (DNA Data Bank of Japan):</b> Японская база данных нуклеотидных последовательностей, сотрудничающая с GenBank и ENA в рамках Международного сотрудничества баз данных нуклеотидных последовательностей (INSDC).</li> </ul> <p><b>Вопрос 19.</b>  Сформулируйте принцип работы алгоритмов локального выравнивания нуклеотидных и аминокислотных последовательностей?  Ответ:</p> <p>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:</p> <ol style="list-style-type: none"> <li>1. <b>Матрица подсчета баллов (scoring matrix):</b> <ul style="list-style-type: none"> <li>• Определяется матрица подсчёта баллов, которая задаёт баллы за совпадения (match), несовпадения (mismatch) и гэпы (пропуски).</li> <li>• Для нуклеотидных: Обычно используются простые схемы: +1 за совпадение, -1 за несовпадение, гап - штраф.</li> <li>• Для аминокислотных: Используются более сложные матрицы, такие как BLOSUM или PAM, отражающие эволюционные взаимосвязи между аминокислотами (замена консервативной</li> </ul> </li> </ol>
--	--	--

		<p>аминокислоты оценивается выше, чем замена на радикально отличающуюся).</p> <ol style="list-style-type: none"> <li><b>Создание матрицы выравнивания:</b> <ul style="list-style-type: none"> <li>Строится матрица выравнивания, размерностью (длина sequenceA + 1) * (длина sequenceB + 1).</li> <li>Первая строка и первый столбец заполняются нулями (это ключевое отличие от глобального выравнивания).</li> </ul> </li> <li><b>Заполнение матрицы выравнивания:</b> <ul style="list-style-type: none"> <li>Каждая ячейка матрицы (i, j) рассчитывается по следующей формуле: <ul style="list-style-type: none"> <li><math>Score(i, j) = \max \{</math> <ul style="list-style-type: none"> <li>0, // Важно! Если все варианты дают отрицательный результат, начинаем с нуля – это позволяет "начать" выравнивание заново.</li> <li><math>Score(i-1, j-1) + score(sequenceA[i], sequenceB[j])</math>, // Совпадение или несовпадение</li> <li><math>Score(i-1, j) + gap\_penalty</math>, // Гэп в sequenceA</li> <li><math>Score(i, j-1) + gap\_penalty</math> // Гэп в sequenceB</li> </ul> </li> <li><math>\}</math></li> <li></li> <li><math>score(sequenceA[i], sequenceB[j])</math> - балл из матрицы подсчёта баллов для выравнивания символов sequenceA[i] и sequenceB[j].</li> <li>gap_penalty - штраф за введение гэпа.</li> </ul> </li> </ul> </li> <li><b>Поиск максимального значения:</b> <ul style="list-style-type: none"> <li>После заполнения матрицы находится ячейка с максимальным значением. Это значение представляет собой наилучший счет локального выравнивания.</li> <li>Координаты этой ячейки (i, j) являются конечной точкой оптимального локального выравнивания.</li> </ul> </li> <li><b>Трассировка (backtracking):</b> <ul style="list-style-type: none"> <li>Начиная с ячейки с максимальным значением, выполняется трассировка в обратном направлении, чтобы определить путь, который привёл к этому значению.</li> <li>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.</li> <li>Трассировка продолжается до тех пор, пока не будет достигнута ячейка со значением 0.</li> <li>Путь трассировки определяет оптимальное локальное выравнивание.</li> </ul> </li> </ol> <p><b>Ключевые особенности локального выравнивания:</b></p> <ul style="list-style-type: none"> <li><b>Начало с нуля:</b> 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.</li> <li><b>Максимальный балл:</b> Ищется максимальный балл во всей матрице, а не только в последней строке или столбце, как в глобальном выравнивании.</li> </ul>
--	--	---

		<ul style="list-style-type: none"> <li>• <b>Трассировка до нуля:</b> Трассировка начинается с ячейки с максимальным значением и продолжается до ячейки со значением 0.</li> </ul> <p><b>Преимущества локального выравнивания:</b></p> <ul style="list-style-type: none"> <li>• Поиск консервативных доменов: Позволяет находить участки высокого сходства даже в дивергентных последовательностях.</li> <li>• Работа с фрагментами: Подходит для выравнивания фрагментов последовательностей.</li> <li>• Более гибкий подход: Не требует, чтобы выравнивание охватывало всю длину последовательностей.</li> </ul> <p>Алгоритм Смита-Вотермана является мощным инструментом для выявления локальных сходств в биомолекулярных последовательностях, что делает его широко используемым в биоинформатике.</p> <p><b>Вопрос 20.</b>          Принцип работы дистанционных методов молекулярной филогении?          Ответ:          Дистанционные методы (Distance-based methods) молекулярной филогении - 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.</p> <p><b>Основные этапы дистанционных методов:</b></p> <ol style="list-style-type: none"> <li>1. <b>Выравнивание последовательностей:</b> <ul style="list-style-type: none"> <li>○ 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.</li> </ul> </li> <li>2. <b>Расчет генетических дистанций:</b> <ul style="list-style-type: none"> <li>○ На основе MSA вычисляется матрица попарных генетических дистанций между всеми парами последовательностей.</li> <li>○ Генетическая дистанция - это мера генетического различия между двумя последовательностями. Существует множество способов оценки генетических дистанций, наиболее распространённые:                     <ul style="list-style-type: none"> <li>▪ <b>Пропорция несовпадающих сайтов (p-distance):</b> 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.</li> <li>▪ <b>Модель Джукса-Кантора (Jukes-Cantor model):</b> Учитывает множественные замены,</li> </ul> </li> </ul> </li> </ol>
--	--	---



		<p>предполагая, что все нуклеотидные замены равновероятны.</p> <ul style="list-style-type: none"> <li>▪ <b>Модель Кимуры 2-параметра (Kimura 2-parameter model):</b> Различает переходы (пурин-пурин или пиримидин-пиримидин) и трансверсии (пурин-пиримидин или наоборот), так как переходы происходят чаще.</li> <li>▪ <b>Более сложные модели:</b> 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.</li> </ul> <p>3. <b>Построение филогенетического дерева:</b></p> <ul style="list-style-type: none"> <li>○ На основе матрицы дистанций строится филогенетическое дерево. Существует несколько алгоритмов: <ul style="list-style-type: none"> <li>▪ <b>UPGMA (Unweighted Pair Group Method with Arithmetic Mean):</b> <ul style="list-style-type: none"> <li>▪ Это иерархический аггломеративный (снизу-вверх) метод.</li> <li>▪ Начинает с того, что каждая последовательность рассматривается как отдельный кластер.</li> <li>▪ На каждом шаге объединяются два ближайших кластера (с наименьшей дистанцией).</li> <li>▪ Дистанция между новым кластером и другими кластерами вычисляется как среднее арифметическое дистанций между последовательностями в новом кластере и последовательностями в других кластерах.</li> <li>▪ 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.</li> </ul> </li> <li>▪ <b>Метод соседнего соединения (Neighbor-Joining, NJ):</b> <ul style="list-style-type: none"> <li>▪ Более продвинутый метод, который также является аггломеративным.</li> <li>▪ Не предполагает постоянную скорость эволюции.</li> <li>▪ На каждом шаге выбирает два кластера, которые минимизируют общую длину ветвей дерева.</li> <li>▪ NJ является одним из наиболее популярных дистанционных методов, так как он относительно быстрый и дает хорошие результаты, особенно когда последовательности не слишком сильно дивергировали.</li> </ul> </li> <li>▪ <b>Метод наименьших квадратов (Least Squares):</b> <ul style="list-style-type: none"> <li>▪ Сравнивает матрицу дистанций дерева с исходной матрицей дистанций, и стремится построить дерево, которое минимизирует разницу между ними (сумма квадратов разностей).</li> <li>▪ Более вычислительно сложный, чем NJ, но может давать более точные результаты, особенно если используется правильная модель эволюции.</li> </ul> </li> </ul> </li> </ul> <p>4. <b>Оценка достоверности дерева:</b></p>
--	--	---

		<ul style="list-style-type: none"> <li>○ После построения дерева важно оценить его достоверность. Это можно сделать с помощью: <ul style="list-style-type: none"> <li>▪ <b>Бутстреп-анализа (Bootstrap analysis):</b> 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, &gt;70%) указывают на то, что ветвь хорошо поддержана данными.</li> <li>▪ <b>Другие методы:</b> Существуют и другие методы оценки достоверности дерева, такие как джекна이프 (jackknife).</li> </ul> </li> </ul> <p><b>Преимущества дистанционных методов:</b></p> <ul style="list-style-type: none"> <li>• <b>Скорость:</b> Относительно быстры, что делает их пригодными для анализа больших наборов данных. NJ особенно быстр.</li> <li>• <b>Простота:</b> Концептуально простые.</li> </ul> <p><b>Недостатки дистанционных методов:</b></p> <ul style="list-style-type: none"> <li>• <b>Потеря информации:</b> Сводят информацию о последовательностях к попарным дистанциям, что приводит к потере информации, содержащейся в исходных данных.</li> <li>• <b>Зависимость от выравнивания:</b> Качество дерева сильно зависит от качества выравнивания последовательностей. Ошибки в выравнивании могут привести к неправильным дистанциям и, как следствие, к неверному дереву.</li> <li>• <b>Чувствительность к выбору модели:</b> Выбор модели эволюции для расчета дистанций может сильно влиять на результат.</li> <li>• <b>Не позволяют учесть все возможные деревья:</b> 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.</li> </ul>
--	--	--

#### Тестирование (Вариант 2).

Индекс и содержание формируемой компетенции	Индикаторы компетенций	Тестовые задания для промежуточной аттестации
ОПК-4	ИДК ОПК 4.1	Задание комбинированного типа с выбором одного или нескольких верных ответов из

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.	Демонстрирует навыки использования методов биоинженерии и биоинформатики для получения новых фундаментальных знаний	<p><b>четырёх предложенных и аргументацией выбора</b></p> <p><b>Вопрос 1.</b> Какие задачи решает биоинформатика? а) Анализ геномных данных б) Сравнительный анализ последовательностей в) Прямое редактирование ДНК г) Прогноз структуры белков Ответ _____ Правильный ответ: а, б, г Аргументация: Биоинформатика работает с данными, а не с редактированием ДНК.</p> <p><b>Вопрос 2.</b> Ключевым историческим этапом в развитии биоинформатики стало: а) Создание GenBank б) Создание CRISPR-Cas9 в) Изобретение микроскопа г) Разработка метода Эдмана Ответ _____ Правильный ответ: а, г Аргументация: GenBank — база последовательностей, а метод Эдмана позволил секвенировать белки.</p> <p><b>Вопрос 3.</b> К операционным системам, применяемым в биоинформатике, относятся: а) Windows б) Unix в) BIOS г) Android Ответ _____ Правильный ответ: а, б Аргументация: Windows и Unix активно применяются, BIOS не является ОС, Android — не используется в биоинформатике.</p> <p><b>Вопрос 4.</b> Для получения почты в Интернете используется: а) POP б) FTP в) TELNET г) HTTP</p>
	<i>ИДК ОПК 4.2</i> Применяет методы биоинженерии и биоинформатики для получения биологических объектов с целенаправленно измененными свойствами.	
	<i>ИДК ОПК 4.3</i> Владеет методами анализа и интерпретации результатов исследования с целью определения практической значимости исследования	
<i>ОПК-5</i> Способен находить и использовать информацию, накопленную в базах данных по	<i>ИДК ОПК 5.1</i> Использует информацию, накопленную в базах данных по структуре геномов, белков и другую	

биологическим объектам, включая нуклеиновые кислоты и белки, владеть основными биоинформатическими средствами анализа.	биологическую информацию	<p>Ответ _____  Правильный ответ: а  Аргументация: POP (Post Office Protocol) предназначен для загрузки писем. емы.</p> <p><b>Вопрос 5.</b>  Какие методы относятся к анализу экспрессии генов?  а) Нозерн-блоттинг  b) Вестерн-блоттинг  c) SAGE  d) ДНК-чипы</p> <p>Ответ _____  Правильный ответ: а, с, d  Аргументация: Эти методы анализируют экспрессию генов. Вестерн — для белков.</p> <p><b>Вопрос 6.</b>  Какие методы применяются для изучения 3D-структуры белков?  а) Рентгеноструктурный анализ  b) ЯМР-спектроскопия  c) Вестерн-блоттинг  d) Электронная микроскопия</p> <p>Ответ _____  Правильный ответ: а, b  Аргументация: Эти методы позволяют получить пространственные структуры белков.</p> <p><b>Вопрос 7.</b>  К каким базам данных относится UniProtKB?  а) Первичные  b) Вторичные  c) Смешанные  d) Библиографические</p> <p>Ответ _____  Правильный ответ: b  Аргументация: UniProtKB — вторичная база данных белков.</p> <p><b>Вопрос 8.</b></p>
	<p><i>ИДК ОПК 5.2</i>  Умеет применять биоинформатические методы и полученные знания для анализа геномной, структурной и иной информации</p> <p><i>ИДК ОПК 5.3</i>  Демонстрирует навыки владения основными биоинформатическими средствами анализа геномной, структурной и иной информации и способен критически оценивать развитие научных идей.</p>	
ОПК-7. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной	<p><i>ИДК ОПК 7.3</i>  Демонстрирует теоретические и практические навыки использования современных информационных технологий в области профессиональной деятельности</p>	

<p>деятельности</p>	<p><i>ИДК ОПК 7.2</i> Использует современные информационные технологии в рамках освоения материала и реализации задач в области профессиональной деятельности.</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> <p><b>Вопрос 9.</b> Локальное выравнивание применяется для:</p> <p>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.</p> <p><b>Вопрос 10.</b> Какой инструмент основан на эвристических методах поиска сходства?</p> <p>a) BLAST b) Нидлман-Вунш c) Смит-Уотерман d) Muscle Ответ _____ Правильный ответ: a Аргументация: BLAST — быстрый эвристический инструмент для поиска сходства.</p> <p><b>Вопрос 11.</b> Паралоги — это:</p> <p>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.</p> <p><b>Вопрос 12.</b></p>
---------------------	--	---

		<p>Какой метод построения филогенетических деревьев является кластерным?</p> <p>a) Neighbor-Joining (объединение соседей)  b) Максимальная парсимония  c) Наибольшее правдоподобие  d) Байесовский метод</p> <p>Ответ _____</p> <p>Правильный ответ: a</p> <p>Аргументация: Neighbor-Joining — один из кластерных методов.</p> <p><b>Вопрос 13.</b></p> <p>Какая база данных чаще всего используется для анализа белковых структур?</p> <p>a) PDB  b) GenBank  c) Ensemble  d) DDBJ</p> <p>Ответ _____</p> <p>Правильный ответ: a</p> <p>Аргументация: PDB (Protein Data Bank) хранит данные о пространственной структуре белков.</p> <p><b>Задание закрытого типа на установление соответствия</b></p> <p><b>Вопрос 14.</b></p> <p>Установите соответствие между алгоритмом и его применением:</p> <p>a) Нидлман–Вунш  b) Смит–Уотерман  c) BLAST  d) ClustalW</p> <p>1 — Глобальное выравнивание  2 — Локальное выравнивание  3 — Быстрый поиск сходных последовательностей  4 — Множественное выравнивание</p> <p>Ответ _____</p> <p>Правильный ответ:</p> <p>a — 1  b — 2  c — 3  d — 4</p> <p>Аргументация: Нидлман–Вунш — глобальное выравнивание, Смит–Уотерман — локальное, BLAST — быстрый поиск, ClustalW — множественное выравнивание.</p>
--	--	---

		<p><b>Вопрос 15.</b> Установите соответствие между термином и его определением:</p> <p>a) Ортологи b) Паралоги c) Ксенологи d) Фенетический метод</p> <p>1 — Гены, возникшие после дупликации в геноме 2 — Гены, разделившиеся после видообразования 3 — Гены, перенесённые горизонтальным переносом 4 — Метод построения деревьев на основе сходства признаков</p> <p>Ответ _____</p> <p>Правильный ответ:</p> <p>a — 2 b — 1 c — 3 d — 4</p> <p>Аргументация: ортологи — у разных видов, паралоги — дупликация, ксенологи — горизонтальный перенос, фенетика — анализ сходства.</p> <p><b>Задание закрытого типа на установление последовательности</b></p> <p><b>Вопрос 16.</b> Расположите основные шаги ПЦР в правильном порядке:</p> <p>a) Денатурация ДНК b) Отжиг праймеров c) Элонгация цепи d) Анализ продуктов</p> <p>Ответ _____</p> <p>Правильный ответ:</p> <p>Правильная последовательность: a → b → c → d</p> <p>Аргументация: стандартный цикл ПЦР: денатурация, отжиг, элонгация, после чего анализируется результат.</p> <p><b>Вопрос 17.</b> Расположите этапы при создании базы данных последовательностей:</p> <p>a) Секвенирование образца b) Аннотация полученных данных c) Загрузка данных в GenBank d) Использование информации исследователями</p> <p>Ответ _____</p>
--	--	--

		<p>Правильный ответ:  Правильная последовательность: <math>a \rightarrow b \rightarrow c \rightarrow d</math>  Аргументация: сначала секвенируют, затем аннотируют, далее загружают в базу и предоставляют доступ.</p> <p><b>Задание открытого типа с развернутым ответом</b></p> <p><b>Вопрос 18.</b>  Перечислите известные белковые базы данных?  Ответ:</p> <ul style="list-style-type: none"> <li>• <b>Protein Data Bank (PDB):</b> Хотя PDB в основном содержит информацию о структурах белков, она также может включать информацию о нуклеотидных последовательностях, связанных с этими белками.</li> <li>• <b>UniProt (Universal Protein Resource):</b> База данных белковых последовательностей и аннотаций.</li> <li>• <b>Pfam:</b> база данных семейств белков, представленных в виде множественных выравниваний последовательностей и скрытых марковских моделей.</li> <li>• <b>InterPro:</b> база данных, объединяющая информацию о семействах, доменах и функциональных сайтах белков из нескольких других баз данных.</li> </ul> <p><b>Вопрос 19.</b>  Сформулируйте принцип работы алгоритмов глобального выравнивания нуклеотидных и аминокислотных последовательностей?  Ответ:</p> <p>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:</p> <ol style="list-style-type: none"> <li>1. <b>Матрица подсчета баллов (scoring matrix):</b> <ul style="list-style-type: none"> <li>○ Определяется матрица подсчёта баллов, которая задаёт баллы за совпадения (match), несовпадения (mismatch) и гэпы (пропуски).</li> <li>○ Как и в локальном выравнивании: <ul style="list-style-type: none"> <li>▪ Для нуклеотидных: Обычно используются простые схемы: +1 за совпадение, -1 за несовпадение, гэп - штраф.</li> <li>▪ 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).</li> </ul> </li> <li>○ Выбор матрицы сильно влияет на результат, и, как правило, требует подбора под конкретную задачу.</li> </ul> </li> </ol>
--	--	---



		<p>2. <b>Создание матрицы выравнивания (alignment matrix):</b></p> <ul style="list-style-type: none"> <li>○ Строится матрица выравнивания (также называемая матрицей динамического программирования) размерностью (длина sequenceA + 1) * (длина sequenceB + 1).</li> <li>○ Первая строка и первый столбец инициализируются таким образом, чтобы отражать штрафы за введение гэпов с самого начала выравнивания. Например, значение ячейки (i, 0) равно <math>i * \text{gap\_penalty}</math>, а значение ячейки (0, j) равно <math>j * \text{gap\_penalty}</math>. <i>Это основное отличие от локального выравнивания.</i></li> </ul> <p>3. <b>Заполнение матрицы выравнивания:</b></p> <ul style="list-style-type: none"> <li>○ Каждая ячейка матрицы (i, j) рассчитывается по следующей формуле:</li> <li>○ <math>\text{Score}(i, j) = \max \{</math></li> <li>○ <math>\text{Score}(i-1, j-1) + \text{score}(\text{sequenceA}[i], \text{sequenceB}[j]),</math> // Совпадение или несовпадение</li> <li>○ <math>\text{Score}(i-1, j) + \text{gap\_penalty},</math> // Гэп в sequenceA</li> <li>○ <math>\text{Score}(i, j-1) + \text{gap\_penalty}</math> // Гэп в sequenceB</li> <li>○ <math>\}</math> <ul style="list-style-type: none"> <li>▪ <math>\text{score}(\text{sequenceA}[i], \text{sequenceB}[j])</math> - балл из матрицы подсчёта баллов для выравнивания символов sequenceA[i] и sequenceB[j]. Другими словами - "награда" за совпадение, или штраф за несовпадение.</li> <li>▪ <math>\text{gap\_penalty}</math> - штраф за введение гэпа. Этот штраф обычно отрицательный.</li> </ul> </li> <li>○ При расчете каждой ячейки выбирается тот вариант, который дает максимальный балл.</li> </ul> <p>4. <b>Трассировка (backtracking):</b></p> <ul style="list-style-type: none"> <li>○ После заполнения матрицы выбирается ячейка в <i>правом нижнем углу</i> матрицы (последняя строка и последний столбец). Значение в этой ячейке представляет собой наилучший счет глобального выравнивания.</li> <li>○ Начиная с этой ячейки, выполняется трассировка в обратном направлении, чтобы определить путь, который привёл к этому значению.</li> <li>○ 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.</li> <li>○ Трассировка продолжается до достижения ячейки (0, 0).</li> <li>○ Путь трассировки определяет оптимальное глобальное выравнивание.</li> </ul> <p><b>Ключевые особенности глобального выравнивания:</b></p> <ul style="list-style-type: none"> <li>• <b>Инициализация первой строки и столбца:</b> Первая строка и столбец инициализируются штрафами за гэпы, чтобы гарантировать, что выравнивание начинается с начала обеих последовательностей.</li> <li>• <b>Оптимальный путь:</b> Алгоритм ищет оптимальный путь через матрицу, приводящий к наилучшему глобальному выравниванию.</li> </ul>
--	--	---

		<ul style="list-style-type: none"> <li>• <b>Трассировка до начала:</b> Трассировка начинается с правого нижнего угла матрицы и продолжается до верхнего левого угла.</li> </ul> <p><b>Преимущества и недостатки глобального выравнивания:</b></p> <ul style="list-style-type: none"> <li>• <b>Преимущества:</b> <ul style="list-style-type: none"> <li>○ Оптимальное выравнивание по всей длине: Гарантирует наилучшее возможное выравнивание, учитывающее всю длину последовательностей.</li> <li>○ Подходит для гомологичных последовательностей: Особенно полезен для последовательностей, которые, как ожидается, имеют сходное происхождение и структуру.</li> </ul> </li> <li>• <b>Недостатки:</b> <ul style="list-style-type: none"> <li>○ Чувствительность к различиям в длине: Плохо работает с последовательностями сильно отличающимися по длине, так как вынужден вводить большое количество гэпов.</li> <li>○ Не подходит для поиска локальных сходств: Не предназначен для поиска консервативных доменов внутри более дивергентных последовательностей.</li> </ul> </li> </ul> <p><b>Вопрос 20.</b>          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.</p> <p><b>Основные этапы метода максимального правдоподобия:</b></p> <ol style="list-style-type: none"> <li>1. <b>Выравнивание последовательностей (Multiple Sequence Alignment - MSA):</b> <ul style="list-style-type: none"> <li>• Как и в других филогенетических методах, первым шагом является получение множественного выравнивания последовательностей (MSA). Качество выравнивания напрямую влияет на результат филогенетического анализа.</li> </ul> </li> <li>1. <b>Выбор модели эволюции:</b> <ul style="list-style-type: none"> <li>• <b>Ключевой аспект ML:</b> Необходимость выбора модели эволюции, описывающей процесс изменения последовательностей со временем.</li> <li>• Модель определяет вероятности различных типов замен нуклеотидов (или аминокислот).</li> <li>• Существуют различные модели, различающиеся по сложности и реалистичности:                     <ul style="list-style-type: none"> <li>○ <b>Модели нуклеотидной эволюции:</b> <ul style="list-style-type: none"> <li>▪ <b>Jukes-Cantor (JC69):</b> Самая простая модель, предполагающая, что все нуклеотидные замены равновероятны.</li> <li>▪ <b>Kimura 2-parameter (K80):</b> Различает переходы (пурин-пурин или пиримидин-</li> </ul> </li> </ul> </li> </ul> </li> </ol>
--	--	--

		<p>пиримидин) и трансверсии (пурин-пиримидин или наоборот), так как переходы происходят чаще.</p> <ul style="list-style-type: none"> <li>▪ <b>Hasegawa-Kishino-Yano (HKY85):</b> Учитывает разные частоты нуклеотидов.</li> <li>▪ <b>General Time Reversible (GTR):</b> Самая общая обратимая модель, позволяющая задавать разные скорости для всех типов замен.</li> <li>○ <b>Модели аминокислотной эволюции:</b> <ul style="list-style-type: none"> <li>▪ 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.</li> </ul> </li> <li>• <b>Оценка параметров модели:</b> Модель эволюции может содержать параметры (например, частоты нуклеотидов, скорости замен), которые необходимо оценить на основе данных.</li> <li>• <b>Выбор подходящей модели:</b> 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.</li> <li>1. <b>Вычисление правдоподобия для каждого дерева:</b> <ul style="list-style-type: none"> <li>• <b>Определение правдоподобия:</b> Правдоподобие (Likelihood, L) - это вероятность наблюдаемых данных (т.е. MSA) при заданном дереве, модели эволюции и параметрах этой модели.</li> <li>• <b>Расчет для каждого сайта:</b> 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.</li> <li>• <b>Суммирование по всем возможным состояниям внутренних узлов:</b> 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.</li> <li>• <b>Необходимость перебора деревьев:</b> 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.</li> </ul> </li> <li>1. <b>Поиск дерева с максимальным правдоподобием:</b> <ul style="list-style-type: none"> <li>• <b>Эвристические алгоритмы:</b> Используются для поиска дерева, максимизирующего функцию правдоподобия. <ul style="list-style-type: none"> <li>○ <b>Nearest Neighbor Interchange (NNI):</b> Начинается с произвольного дерева, затем ветви меняются местами, чтобы найти дерево с более высоким правдоподобием.</li> <li>○ <b>Subtree Pruning and Regrafting (SPR):</b> Поддеревья обрезаются и переносятся в другое место дерева, чтобы улучшить правдоподобие.</li> </ul> </li> </ul> </li> </ul>
--	--	--

		<ul style="list-style-type: none"> <li>○ <b>Tree Bisection and Reconnection (TBR):</b> Дерево делится на две части, а затем соединяется снова разными способами в поисках лучшего дерева.</li> <li>• <b>Множественные запуски (Multiple runs):</b> Часто алгоритм запускается несколько раз с разными начальными деревьями, чтобы увеличить вероятность нахождения глобального максимума правдоподобия.</li> <li>• <b>Выбор лучшего дерева:</b> Из всех найденных деревьев выбирается дерево с наибольшим значением правдоподобия.</li> </ul> <p>1. <b>Оценка достоверности дерева:</b></p> <ul style="list-style-type: none"> <li>• <b>Бутстреп-анализ (Bootstrap analysis):</b> 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 (&gt;70%) указывают на то, что ветвь хорошо поддержана данными.</li> <li>• <b>Байесовский анализ:</b> Иногда метод максимального правдоподобия комбинируют с байесовским анализом для получения апостериорных вероятностей ветвей.</li> </ul> <p><b>Преимущества метода максимального правдоподобия:</b></p> <ul style="list-style-type: none"> <li>• <b>Статистически обоснованный:</b> ML основан на строгих статистических принципах.</li> <li>• <b>Использование всей информации:</b> Использует всю информацию, содержащуюся в выравнивании последовательностей, а не только попарные дистанции.</li> <li>• <b>Гибкость:</b> Позволяет использовать сложные модели эволюции, которые лучше отражают реальные процессы изменения последовательностей.</li> </ul> <p><b>Недостатки метода максимального правдоподобия:</b></p> <ul style="list-style-type: none"> <li>• <b>Вычислительная сложность:</b> ML требует больших вычислительных ресурсов, особенно для больших наборов данных.</li> <li>• <b>Зависимость от модели:</b> Результат сильно зависит от выбора модели эволюции. Неправильный выбор модели может привести к неверным результатам.</li> <li>• <b>Эвристический поиск:</b> Эвристические алгоритмы не гарантируют нахождения глобального максимума правдоподобия. Существует риск "застрять" в локальном максимуме.</li> <li>• <b>Сложность в понимании и интерпретации:</b> Требуется хорошего понимания статистических принципов и моделей эволюции.</li> </ul>
--	--	--


Разработчик:

 доцент Приставка А.А.  
(подпись)

Программа составлена в соответствии с требованиями ФГОС ВО по направлению 06.05.01 «Биоинженерия и биоинформатика».

Программа рассмотрена на заседании кафедры физико-химической биологии, биоинженерии и биоинформатики 17.04.2024 г. протокол № 15.

Зав. кафедрой, д.б.н., профессор В.П. Саловарова



***Настоящая программа, не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.***