



Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»
Кафедра физико-химической биологии, биотехнологии и биоинформатики

УТВЕРЖДАЮ
Декан биолого-почвенного факультета
Матвеев А.Н.
«10» марта 2020 г.

Рабочая программа дисциплины

Наименование дисциплины: Б1.В.13 «ОСНОВЫ БИОИНФОРМАТИКИ»

Направление подготовки: 06.03.01 «Биология»

Тип образовательной программы: академический бакалавриат

Направленность (профиль) подготовки: «Физико-химическая биология и биотехнология»

Квалификация выпускника: Бакалавр

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

Согласовано с УМК биолого-почвенного факультета

Рекомендовано кафедрой:
Протокол № 11 от 18 февраля 2020 г.
Зав. кафедрой Саловарова В.П.

Протокол № 4 от 10 марта 2020 г.
Председатель проф. Матвеев А.Н.

Иркутск 2020 г.

Содержание

	стр.
1. Цели и задачи дисциплины (модуля)	3
2. Место дисциплины (модуля) в структуре ОПОП	3
3. Требования к результатам освоения дисциплины (модуля)	4
4. Объем дисциплины (модуля) и виды учебной работы	5
5. Содержание дисциплины (модуля)	5
5.1 Содержание разделов и тем дисциплины (модуля)	5
5.2 Разделы дисциплины (модуля) и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами (модулями)	8
5.3 Разделы и темы дисциплин (модулей) и виды занятий	9
6. Перечень семинарских, практических занятий, лабораторных работ, план самостоятельной работы студентов, методические указания по организации самостоятельной работы студентов.	9
7. Примерная тематика курсовых работ (проектов) (при наличии)	16
8. Учебно-методическое и информационное обеспечение дисциплины (модуля):	
а) основная литература	16
б) дополнительная литература	16
в) программное обеспечение	17
г) базы данных, поисково-справочные и информационные системы	17
9. Материально-техническое обеспечение дисциплины (модуля).	18
10. Образовательные технологии	19
11. Оценочные средства (ОС)	19

1. Цели и задачи дисциплины (модуля):

Целью освоения учебной дисциплины «Основы биоинформатики» является:

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

Задачи дисциплины:

- 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.

2. Место дисциплины в структуре ОПОП:

Предмет «Основы биоинформатики» является обязательной дисциплиной вариативной части учебного плана подготовки бакалавров по профилю «Физико-химическая биология и биотехнология» направления 06.03.01 – Биология.

Актуальность курса обусловлена стремительным нарастанием молекулярно-биологических данных, связанных с массовой расшифровкой полных геномных последовательностей. Осмысливание этих данных невозможно без привлечения современных информационных технологий, создания и развития эффективных методов и алгоритмов анализа данных и моделирования биологических систем и процессов. На стыке биологии, математики, информатики возникла новая наука – информационная биология (биоинформатика). Дисциплина основана на методах организации информации; методах вычислительной математики и статистики, адаптированных к анализу последовательностей и пространственных структур биополимеров.

Содержание курса базируется на результатах, полученных в области математического анализа, информатики, различных разделов химии и биологии, поэтому его основные положения разрабатывались с учетом знаний и умений, полученных при изучении предшествующих дисциплин бакалавриата: «Математика», «Органическая химия», «Биохимия и молекулярная биология», «Общая биология», «Физико-химические методы в биологии», «Биофизика», «Генетика». Студент, приступающий к изучению дисциплины «Основы биоинформатики», должен знать основные методы, теории и законы предшествующих естественнонаучных дисциплин:

- физика: физические методы исследования строения структуры вещества, включая рентгеноструктурный анализ и спектральные методы.
- химия: строение и физико-химические свойства основных классов соединений; закономерности протекания химических реакций; ферментативный катализ, химические методы исследования биоорганических молекул.
- биология: молекулярные основы теории эволюции; строение и свойства белков и нуклеиновых кислот; матричные процессы в клетке и их регуляция; биологическая номенклатура и таксономия.
- математика: статистический анализ; основы математического моделирования.

- информатика: структура и общие свойства информации; ее сбор, хранение, поиск, переработка, преобразование, распространение и использование.

Данная дисциплина является необходимой основой при изучении последующих курсов, рассматривающих вопросы взаимосвязи биологии и информатики («Нанобиотехнологии», «Новейшие технологии в биомедицине»), а так же для прохождения преддипломной практики и выполнения выпускной квалификационной работы.

3. Требования к результатам освоения дисциплины (модуля):

Процесс изучения дисциплины (модуля) направлен на формирование следующих компетенций:

- способность использовать основные средства анализа геномной, структурной и другой биологической информации и способностью использовать основные биологические базы данных, в том числе содержащие геномную, структурную и другую информацию, в научно-исследовательской работе и практической деятельности (СПК-2);
- способность получать и грамотно использовать информацию, накопленную в базах данных по структуре геномов, белков и другой биологической информации (СПК-3);
- способность эксплуатировать современную аппаратуру и оборудование для выполнения научно-исследовательских полевых и лабораторных биологических работ (ПК-1)

В результате изучения дисциплины студент должен:

Знать:

- основные аппаратные и программные средства реализации информационных технологий, используемых в биоинформатике;
- принципы работы с молекулярно-биологическими базами данных и с обслуживающими их приложениями;
- методы эффективного поиска и обработки информации о последовательностях и структурах биополимеров;
- основные принципы молекулярного моделирования;
- новейшие достижения в области биоинформатики и перспективы их практического и теоретического использования.

Уметь:

- использовать полученные знания и навыки для решения профессиональных задач.
- организовывать поиск информации в базах данных и использовать возможности программных средств и сетевых технологий для молекулярно-биологических исследований;
- осуществить выбор наиболее оптимального информационно-вычислительного метода исследования в зависимости от поставленной задачи;
- использовать основные технологии и методы молекулярной визуализации.

Владеть:

- методами и средствами анализа молекулярно-биологической информации;
- навыками работы с биологическими базами данных и обслуживающими их приложениями;

- базовыми пакетами прикладных программ для анализа структуры и последовательной макромолекул.

4. Объем дисциплины и виды учебной работы

Вид учебной работы	Всего часов / зачетных единиц	Семестры			
		7			
Аудиторные занятия (всего)	54/1,5	54/1,5			
Из них объем занятий с использованием электронного обучения и дистанционных образовательных технологий	10/0,28	10/0,28			
В том числе:					
Лекции	18/0,5	18/0,5			
Практические занятия (ПЗ)	36/1	36/1			
Семинары (С)	-	-			
Лабораторные работы (ЛР)	-	-			
КСР	4/0,11	4/0,11			
Самостоятельная работа (всего)	59/1,64	59/1,64			
В том числе:					
Курсовой проект (работа)	-	-			
Расчетно-графические работы	36/1	36/1			
Реферат	-	-			
<i>Другие виды самостоятельной работы</i>					
Письменные работы	23/0,64	23/0,64			
Вид промежуточной аттестации (экзамен)	27/0,75	27/0,75			
Контактная работа (всего)	58/1,61	58/1,61			
Общая трудоемкость	часы	144	144		
	зачетные единицы	4	4		

5. Содержание дисциплины (модуля)

5.1. Содержание разделов и тем дисциплины (модуля). Все разделы и темы нумеруются.

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. Ярлыки экспрессируемых последовательностей (*Expressed Sequence Tags – EST*). 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 предсказания.

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 (Basic Input-Output System), DOS (Disk Operating System), 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 - модем, ISDN (Integrated Services Digital Network), ADSL (Asymmetric Digital Subscriber Line), оптоволоконная и спутниковая связь.

Всемирная паутина (World Wide Web) как информационная система, построенная на принципе гиперсреды. Веб-страницы и веб-узлы. Объектная сеть. Object Request Broker (ORB). Common Object Request Broker Architecture (CORBA). Interface Definition Language (IDL).

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).

EMBnet — European Molecular Biology net. Центры и узлы. Система выборки последовательностей (SRS — Sequence Retrieval System) как сетевой обозреватель баз данных EMBnet. Национальный центр биотехнологической информации (National Center for Biotechnology Information, NCBI) и сервисная программа Entrez. Выборка и применение информации. Номер GI (GenInfo Identifiers). Зеркала и Интернет.

Тема 3. Секвенирование и анализ ДНК и белков

Цели анализа последовательностей. 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. Теоретические методы - эмпирические статистические методы; методы, опирающиеся на физико-химические критерии; алгоритмы, основанные на гомологии структур.

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 и объектно-

ориентированные 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 (Structured Query Language - SQL).

Обзор основных БД. Первичные базы данных. Базы данных последовательностей нуклеиновых кислот. EMBL (European Molecular Biology Laboratory). DDBJ (DNA DataBank of Japan). GenBank. GSDB (Genome Sequence DataBase). Ensemble. Специализированные БД: SGD (Saccharomyces Genome Database), TDB (TIGR DataBase), EST. Базы данных белковых последовательностей. PIR (International Protein Sequence Database). SWISS-PROT. TrEMBL (TRanslated from EMBL). Protein Research Foundation. 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). FSSP (Fold classification based on Structure-Structure alignment of Proteins). Вторичные базы данных. *PROSITE*. *PRINTS*. *BLOCKS*. *IDENTIFY*. *KEGG* (Kyoto Encyclopedia of Genes and Genomes). Библиографические базы данных. *PubMed*. *AGRICOLA* (Agricultural Online Access). Виртуальная библиотека.

Специализированные средства анализа БД. Пакет GCG (Genetics Computer Group). EGCG. Staden. Lasergene. Sequencher. Vector, NTI. MacVector. SYNERGY. Pangea System. EMBOSS. Alfresco. DALI (Distance matrix Alignment). Современные тенденции в структурировании БД. База данных Uniprot-Swiss-Prot.

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 Смита-Уотермана. Матрица переходов. Метод k-кортежей. Алгоритмы программ FASTA (Fast Alignment) и BLAST (Basic Local Alignment Search Tool).

Множественное выравнивание последовательностей. Цели и задачи множественного выравнивания. Основные этапы множественного выравнивания. Методы множественного выравнивания – Profile, Blocs, Indicator, PSI-BLAST (Position Specific Iterated-BLAST), Скрытые марковские модели. Программы автоматического выравнивания – CLUSTAL, CINEMA (Color INteractive Editor for Multiple Alignments), Muscle, T-COFFEE.

Алгоритмы распознавания доменов в белковых структурах. PPUU (Parser for protein Unfolding Units). DOMAK. Detective. Strudl (STRUctural Domain Limits). Алгоритмы сравнения структур. Метод двойного динамического программирования и программа DALI.

Тема 6. Предсказание генов структур и функций

Аннотирование. Структурная и функциональная аннотация. Цели и задачи аннотирования. Предсказание генов. Стратегии отыскания генов – семантические,

позиционные и сравнительные. Программы предсказания генов - GRAIL (Gene Recognition and Analysis Internet Link), GENSCAN, PROCRUSTES и др. Предсказание вторичной структуры РНК.

Предсказание структуры белка. Стратегии предсказания. Предсказание вторичной структуры, основанное на склонности аминокислот к формированию определенной вторичной структуры. Метод Чоу-Фасмена. Метод GOR. Методы, основаны на принципе обучающейся машины. Нейронные сети. Поиск ближайшего соседа. Предсказание трехмерной структуры. Сравнительное моделирование (моделирование гомологии). Последовательность шагов алгоритма. Метод протягивания. Энергетический подход. Сравнительные достоинства и недостатки методов. Программы, позволяющие предсказывать структуру белка: AAccompIdent, AAccompSim, SAPS (Statistical Analysis of Protein Sequences), NNPREPREDICT, LINUS (Local Independently Nucleated Units of Structures). Предсказание функций белков.

Тема 7. Гомология, филогения и эволюционные деревья

Гомология и подобие. Ортологи, Паралоги и Ксенологи. Изучение ортологических и паралогических белков. Модульные белки. Филогения и родство. Эволюционное дерево. Основные подходы к филогенетическому анализу: фенетический, кладистический и эволюционный. Клада, таксон и узел. Критерии филогенетического анализа: морфологические характеристики, биохимические свойства и последовательности нуклеиновых кислот и белков. Этапы филогенетического анализа. Самосовершенствование как оценочный метод переборки дерева.

Теория графов, ее значение для изучения филогенетических отношений. Граф, узлы, дуги, путь. Ориентированные и взвешенные графы. Корневые и некорневые деревья. Характерные свойства филогенетических деревьев.

Методы построения деревьев. Методы расстояний. Групповой метод невзвешенных пар с вычислением среднего арифметического. Алгоритм объединения соседей. Метод Фитча-Маргобиаша. Метод минимальной эволюции. Методы подобия. Метод максимальной экономичности. Метод максимального правдоподобия. Эволюционные модели. Таблицы подобия и таблицы расстояний. Выравнивание по последовательности и структуре – сложность и точность вычислений.

Молекулярные подходы к определению филогении. Типы макромолекул, используемых для филогенетического анализа. Митохондриальная и рибосомная ДНК. Эволюционная значимость и область применения в филогении.

Базы данных филогенетического анализа. Универсальные средства: PAUP (Phylogenetic Analysis Using Parsimony) и PHYLIP (Phylogenetic Inference Package). PALI (Phylogeny and ALignment of homologous protein structures).

5.2 Разделы дисциплины и междисциплинарные связи с обеспечиваемыми (последующими) дисциплинами

№ п/п	Наименование обеспечиваемых (последующих) дисциплин	№ № разделов и тем данной дисциплины, необходимых для изучения обеспечиваемых (последующих) дисциплин								
		3	5	6						
1.	Нанобиотехнологии	3	5	6						
2.	Новейшие технологии в биомедицине	3	4	5	6	7				
3.	Преддипломная практика	1	2	3	4	5	6	7		
4.	Выпускная квалификационная работа	3	4	5	6	7				

5.3. Разделы и темы дисциплин (модулей) и виды занятий

№ п/п	Наименование раздела	Наименование темы	Виды занятий в часах					
			Лекц.	Практ. зан.	Сем.	Лаб. зан.	СРС	Всего
1.		Введение. История, предмет и значение биоинформатики	2	2	-	-	8	12
2.		Основные инструменты биоинформатики	2	6	-	-	8	16
3.		Секвенирование и анализ ДНК и белков	2	4	-	-	8	14
4.		Базы данных в биоинформатике	3	6	-	-	8	17
5.		Выравнивание последовательностей	3	6	-	-	9	18
6.		Предсказание генов структур и функций	3	6	-	-	9	18
7.		Гомология, филогения и эволюционные деревья	3	6	-	-	9	18

6. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины	Наименование семинаров, практических и лабораторных работ	Трудоемкость (часы)	Оценочные средства	Формируемые компетенции
1.	Тема 1	Информационные технологии в биологии	2	Контрольные вопросы.	СПК-2,3; ПК-1
2.	Тема 2	Инструментальная основа информационных технологий	2	- « -	- « -
3.	Тема 2	Визуализация молекул с использование компьютерных программ	4	Контрольные вопросы, отчеты	- « -
4.	Тема 3	Методы секвенирования нуклеиновых кислот и белков	2	Контрольные вопросы.	- « -
5.	Тема 3	Физико-химические методы анализа экспрессии генов	2	- « -	- « -
6.	Тема 4	Биологические базы данных. Поиск белков и нуклеотидных последовательностей.	3	Контрольные вопросы, отчеты	- « -
7.	Тема 4	Биологические базы данных. Работа в системе PubMed	3	- « -	- « -
8.	Тема 5	Выравнивание последовательностей	3	- « -	- « -

		программой BLAST.			
9.	Тема 5	Выравнивание последовательности запроса с последовательностью из базы данных и поиск гомологичных последовательностей	3	- « -	- « -
10.	Тема 6	Аннотирование информации. Методы предсказания функций генов	3	Контрольные вопросы.	- « -
11.	Тема 6	Методы предсказания пространственной структуры макромолекул	3	- « -	- « -
12.	Тема 7	Теоретические аспекты биоинформатики: множественное выравнивание и молекулярной филогении	3	Контрольные вопросы, отчеты	- « -
13.	Тема 7	Прикладные аспекты биоинформатики: медицина	3	- « -	- « -

6.1. План самостоятельной работы студентов

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
1-2	Тема 1. Введение. История, предмет и значение биоинформатики	Контрольные вопросы	1. История и актуальные проблемы биоинформатики. 2. Предпосылки для возникновения современной биоинформатики	1	8
3-4	Тема 2. Основные инструменты биоинформатики.	Контрольные вопросы, Практические задания	2. Визуализация биомолекул с использованием компьютерных программ RasMol и ACD ChemSketch. 2. Принцип работы и инфраструктура Internet	1, 2	8
5-7	Тема 3. Секвенирование и анализ ДНК и белков	Контрольные вопросы	1. Методы изучения последовательности и структуры белков и нуклеиновых кислот. 2. Методы исследования экспрессии генов.	1	8
8-10	Тема 4. Базы данных в биоинформатике	Контрольные вопросы, Практические задания Подготовка к КР	1. Биологические БД: виды и особенности организации информации. 2. Поиск заданных последовательностей белков и НК в БД. 3. Поиск последовательностей в БД по гомологии	1, 2, 3	8
11-13	Тема 5. Выравнивание	Контрольные вопросы,	1. Попарное выравнивание заданных	1, 2, 3	9

	последовательностей	Практические задания	последовательностей с помощью программы GeneDoc. 2. Парное выравнивание заданных последовательностей с использованием программы BLAST.		
14-16	Тема 6. Предсказание генов структур и функций	Контрольные вопросы, Подготовка к КР	1. Аннотирование информации в базах данных: проблемы, направления и перспективы. 2. Множественное выравнивание заданных последовательностей с использованием программы CLUSTAL W	1, 2	9
17-19	Тема 7. Гомология, филогения и эволюционные деревья	Контрольные вопросы, Практические задания	1. Построение филогенетических деревьев 2. Применение биоинформатики в медицинских исследованиях, сельском хозяйстве и пищевой промышленности.	1, 2, 3	9

6.2. 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 «Основы биоинформатики» предлагаются следующие формы самостоятельной работы:

- а) Углубленный анализ научно-методической литературы и изучение учебного материала, предусмотренного рабочей программой, но не изложенного в лекциях;
- б) подготовка к контрольному опросу на практических занятиях;
- в) выполнение практических заданий по овладению программными средствами биоинформатики;
- г) подготовка к контрольным работам по основным разделам дисциплины

Темы для самостоятельной работы

1. Проект «Геном человека»: цели, задачи, значение для становления и развития биоинформатики
2. Физико-химические методы изучения последовательности и структуры белков.
3. Физико-химические методы изучения последовательности и структуры нуклеиновых кислот.
4. Физико-химические и молекулярно-биологические методы исследования экспрессии генов.
5. Новейшие методы секвенирования нуклеиновых кислот
6. Принцип работы и инфраструктура Internet
7. Аннотирование информации в базах данных: проблемы, направления и перспективы.
8. Теоретические основы молекулярной филогении: теория молекулярной эволюции.
9. Поиск заданных белков и нуклеиновых кислот в базах данных.
10. Поиск последовательностей в банках данных по гомологии.
11. Парное выравнивание заданных последовательностей с использованием программы BLAST.

12. Попарное выравнивание заданных последовательностей с помощью программы GeneDoc.
13. Визуализация биомолекул с использованием компьютерных программ RasMol и ACD ChemSketch.
14. Применение биоинформатики в медицинских исследованиях.

Рекомендации по выполнению практических заданий

Тема 2.

Визуализация молекул с помощью программы ACD ChemSketch

1. В редакторе ACD нарисуйте химическую структуру пятой аминокислоты вашего белка (табл. 1). Изобразите как L, так и D форму аминокислоты, сделайте подписи. Проведите трехмерную оптимизацию, переведите рисунок в 3D формат и скопируйте полученные структуры в документ Microsoft Word.
2. В редакторе ACD создайте новую страницу и скопируйте туда L-форму вашей аминокислоты. Нарисуйте реакцию образования пептидной связи между четвертой и пятой аминокислотами вашего белка. Скопируйте полученный рисунок в файл *.doc.
3. На новой странице нарисуйте трипептид, образованный третьей, четвертой и пятой аминокислотами вашего белка. Выделите красным остов пептида и скопируйте рисунок в файл *.doc. В редакторе ACD сохраните файл со всеми полученными структурами.

Визуализация молекул с помощью программы RasMol

1. Найдите в БД изображение пространственной структуры своего белка (табл. 1). Для этого нужно войти на сервер NCBI/Structure (<http://www.ncbi.nlm.nih.gov/Structure/index.shtml>) и ввести в поисковой строке PDB_ID белка. В появившемся окне нажмите на иконку с изображением белка. Затем выберите кнопку «Structure View in RasMol» и загрузите файл на компьютер. Скопируйте в протокол образец описания состава PDB-структуры (табл. 5) и отредактируйте заголовок в соответствии со своим белком.
2. Откройте загруженный pdb-файл с помощью программы RasMol. Для этого необходимо запустить программу и через меню File загрузить изображение. Обратите внимание, что появляется два окна: с изображением белка и командное: "RasMol command Line". В командное окно можно вводить команды, позволяющие работать как со всей молекулой, так и с отдельными множествами атомов, входящих в ее структуру. Подробное описание команд представлено в приложении 2.
3. Создайте светлый фон (команда background white или по системе RGB: background [230,250,200] (числа не должны превышать 255)). Изобразите цепи белка в остовной модели, выбрав подходящую толщину линий: backbone [30, 40 и т.д.]
Раскрасьте разные цепи белка в разные цвета: color chain
4. Определите имена полипептидных цепей, входящих в состав молекулы белка. Щелкая по концевым C_{α} -атомам (в остовной модели C_{α} -атомы расположены в вершинах ломаной), получите в командной строке информацию об имени цепи, типе и номере остатка и др. Внесите информацию об именах цепей, типах и номерах концевых остатков в протокол.
5. Изобразите N-концевые азоты большими (400) синими шариками:

```
select <имя атома>:цепь (например, select arg58:a.n , где а – цепь, n – азот)
cpk 400
color blue
```

Тип остатка, номер остатка и имя цепи прочитайте в собственном протоколе. Аналогичным образом изобразите C-концевые кислороды красными шариками. Имейте в виду, что их по 2 на каждую цепь (-COOH !). Чтобы узнать как называется второй

кислород из карбоксильной группы выделите C-концевой остаток целиком (например, select 327A), изобразите его в проволоочной модели (wireframe 50), покрасьте по типам атомов (color cpk) и щелкая по атомам кислорода, определите их имена в командном окне.

6. Измените окраску цепей так, чтобы α -цепи гемоглобина отличались от β -цепей. Чтобы покрасить цепь с именем X:

```
select *:X  
color <цвет>
```

Например, А-цепь зеленая (green), В-цепь - оранжевая (orange). Имена цепей — в вашем протоколе.

7. Определите, какие небелковые компоненты присутствуют в структуре молекулы. Для этого их необходимо выделить (select not protein или select hetero), увеличить размер и покрасить по типам атомов. После их выделения, небелковые компоненты в командной строке будут иметь ID-шифр. Чтобы узнать, какое соединение зашифровано под условным обозначением, можно так же воспользоваться сервисом NCBI / Structure.

Заполните таблицу 5 в протоколе.

8. Вращая молекулу, подберите наиболее удачный ракурс и сохраните полученное изображение в графическом файле формата gif (меню Export). Творческий подход к выполнению заданий (подбор цветов объектов, толщины линий и способов изображения и т.п.) приветствуется!

Тема 4.

Поиск белков в базах данных

1. Найдите в БД UniProtKB документ, содержащий информацию о белке, указанном в таблице 1 напротив вашей фамилии. Для этого откройте главную страницу (<http://www.uniprot.org/uniprot/>). Наверху страницы появится поле для введения запроса для поиска по БД UniProtKB. Введите в него ID номер вашего белка.
2. Внимательно изучите полученный документ и заполните табл. 2.
3. Создайте текстовый файл с последовательностью Вашего белка в формате FASTA.
4. Определите, сколько документов в БД UniProtKB содержит белки с тем же кратким описанием, что и Ваш белок.

Работа с системой PubMed

1. Зайдите на сайт NCBI (<http://www.ncbi.nlm.nih.gov/>) и перейдите по гиперссылке "PubMed" (слева экрана в меню Popular Resources) Найдите ссылки на статьи, в которых упоминается полное название Вашего белка. Приведите в протоколе это название и количество найденных статей. Сохраните результаты поиска в виде текстового файла (используйте меню "Send to", в котором выберите "File").
2. Из списка найденных статей выберите одну (укажите в протоколе, какую). Сохраните её аннотацию ("Abstract") в текстовом формате.
3. Найдите ссылки на все публикации последнего автора этой статьи за последние 3 года. В протоколе укажите число статей. Чтобы найти публикации человека, который подписывает статьи как A.I. Petrov, следует написать в строке запроса: petrov ai [Author]. Для ограничения поиска последними тремя годами воспользуйтесь сервисом "Limits".

Тема 5

Выравнивание последовательностей с помощью программы GeneDoc

1. Скопируйте пару коротких последовательностей из таблицы 3 (против своей фамилии) в текстовый файл. Запустите программу GeneDoc и импортируйте этот файл. Для этого в меню File выбираете Import. В открывшемся окошке выбираете формат "Fasta (Pearson)" и нажимаете "Import". Выбираете файл, затем "Open" и "Done".

2. Выровняйте последовательности, стараясь, чтобы было сопоставлено максимальное число одинаковых букв. Редактировать выравнивание можно двумя способами: а) Включите режим "Grab and drag sequences". Это можно сделать либо в меню Arrange, либо комбинацией <Ctrl+A>. В этом режиме можно "цеплять" мышью участки последовательности и сдвигать их. б) Режимы "Insert gap into sequence" и "Delete gap from sequence" в меню Arrange. В этих режимах можно вставлять/убирать гэпы щелчком мыши. (Более детальное описание работы с программой представлено ниже).
3. Сохраните выравнивание под именем alignment1.ms. Посчитайте и занесите в протокол процент идентичности и процент сходства двух последовательностей. Процентом идентичности считается отношение числа колонок выравнивания, в которых стоят одинаковые буквы, к общему числу колонок (включая "гэповые"), умноженное на 100%. Процентом сходства будем считать отношение числа колонок со сходными буквами к общему числу, умноженное на 100%.

Выравнивание последовательностей с помощью программы BLAST

1. Последовательность seq1 из табл. 3 представляет собой фрагмент последовательности вашего белка (табл. 1). Чтобы определить, какой именно это фрагмент, выровняйте его с полной последовательностью белка. Для этого используйте программу "bl2seq", которая предназначена в первую очередь для выравнивания одинаковых или очень сходных последовательностей. Зайдите на сайт <http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi> (web-интерфейс к программе "bl2seq"). Определите с помощью этого сервиса координаты последовательности seq1 в полной последовательности вашего белка. Чтобы выравнивать аминокислотные последовательности, надо в меню Program выбрать "blastp". В поле "Sequence 1" скопируйте либо последовательность вашего белка в fasta-формате, либо просто его первый АС или ID. В поле "Sequence 2" положите фрагмент, координаты которого Вы хотите найти. Нажмите кнопку Blast и дождитесь результата. Занесите ответ в протокол.
2. Пользуясь тем же сервисом, выровняйте ваш белок с другим белком, указанным против вашего имени в таблице 4. Занесите в протокол: какие белки из каких организмов Вы выравнивали, проценты идентичности ("Identities") и сходства ("Positives"), число гэпов, координаты выровненного участка (участков, если выдано несколько выравниваний) в обеих последовательностях. Сохраните карту локального сходства в виде gif-файла; желательно, чтобы в имени файла фигурировали идентификаторы последовательностей.

Редактирование результатов выравнивания

1. Импортируйте последовательности, которые Вы выравнивали в предыдущем задании, в GeneDoc. Попробуйте вручную воспроизвести выравнивание, полученное программой bl2seq. Поскольку программа выдала одно или несколько частичных выравниваний, а вам нужно сделать полное, можно против всех остатков, не вошедших в частичные выравнивания, поставить гэпы. Ну а можно попытаться найти что-нибудь разумное и вне выданных выравниваний.
2. Поэкспериментируйте с программой bl2seq — попробуйте поменять матрицу или штрафы за гэпы (гиперссылка "Algorithm parameters"). Опишите результаты в протоколе.

Поиск белка по его гомологу

При выполнении задания пользуйтесь web-интерфейсом к BLASTP на сервере NCBI: <http://www.ncbi.nlm.nih.gov/blast/>. В разделе Basic BLAST необходимо перейти по гиперссылке "protein blast".

1. Подайте на вход программе BLASTP (Enter Query Sequence) последовательность своего белка из табл. 4. Проведите поиск гомологичных последовательностей в БД Swissprot. Внимание! Для этого надо изменить значение параметра database - по

умолчанию стоит банк "nr". Чтобы получить результат, надо нажать кнопку BLAST и подождать одну-две минуты.

В выдаче программы отыщите поданный на вход белок и ваш белок из таблицы 1. Занесите в протокол их порядковые номера в выдаче, Score, E-value.

Письменно ответьте на вопросы: Что за белок является последним в выдаче программы? Каков его порядковый номер, Score, E-value? Объясните, почему программа не выдала больше находок. Какие параметры программы в данном случае надо изменить, чтобы находок стало больше? Проверьте своё предположение.

2. Повторите поиск с той же входной последовательностью, указав в качестве банка pdb. Для первой последовательности в списке находки укажите: PDB-коды и идентификаторы, Score, E-value, начало и конец выравнивания во входной последовательности (Query) и в находке (Subject), процент совпадений (Identity).
3. Повторите поиск по Swiss-Prot, подав на вход не всю последовательность, а любую её треть. Ответьте на вопросы: Является ли исходная последовательность первой в списке находок? Как изменились Score и E-value? Объясните наблюдаемые изменения.

7. Примерная тематика курсовых работ (проектов) (при наличии)

Курсовых работ по дисциплине учебным планом не предусмотрено

8. Учебно-методическое и информационное обеспечение дисциплины (модуля):

а) основная литература

1. Леск А. Введение в биоинформатику : пер. с англ. / А. М. Леск ; ред.: А. А. Миронов, В. К. Шведаса. - 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 [Текст] : учеб.-метод. пособие : в 3 ч. / А. А. Приставка, В. П. Саловарова - Иркутск : Изд-во ИГУ, 2013. - Ч. 1 : Белки. - 2013. - 121 с. - ISBN 978-5-9624-0962-7 (69 экз.)
3. Белькова Н.Л. Большой практикум по биоинженерии и биоинформатике [Текст] : учеб.-метод. пособие : в 3 ч. / Н. Л. Белькова. - Иркутск : Изд-во ИГУ, 2013. - ISBN 978-5-9624-0956-6. Ч. 2 : Нуклеиновые кислоты. - 2014. - 155 с. - ISBN 978-5-9624-1184-2 (39 экз.)

б) дополнительная литература

1. Игнасимуту С. Основы биоинформатики / С. Игнасимуту ; пер. с англ. А. А. Чумичкин. - Ижевск : Регулярная и хаотическая динамика : Ин-т компьютер. исслед., 2007. - 316 с. - ISBN 978-5-93972-620-7 (1 экз.)
2. Каменская М.А. Информационная биология / М.А. Каменская. – М.: Академия, 2006. – 361 с. - ISBN 5-7695-2580-0 (8 экз.)
3. Компьютеры и суперкомпьютеры в биологии / Под ред. В.Д. Лахно, М.Н. Устинин. – Москва-Ижевск: Институт компьютерных исследований, 2002. – 528 с. - ISBN 5-93972-188-5 (2 экз.)
4. Математические методы для анализа последовательностей ДНК. / Под ред. М.С. Уотермена, перевод с англ. – М.: Мир, 1999. – 349 с. - ISBN 5030025200 (1 экз.)
5. Паун Г. ДНК-компьютер. Новая парадигма вычислений / Г. Паун, Г. Розенберг, А. Саломаа ; Пер. с англ. Д. С. Ананичева, И. С. Киселевой, О. Б. Финогеновой, ред. М. В. Волков. - М. : Мир, 2004. - 527 с. - ISBN 5-03-003480-3 (1 экз.)
6. Структура и функционирование белков: применение методов биоинформатики / пер. с англ.: В. Н. Новоселецкий, Е. Д. Балицкая, Т. В. Наumenкова ; ред. В. Н. Новоселецкий. - М. : УРСС : Ленанд, 2014. - 414 с. - ISBN 978-5-9710-0842-2. - ISBN 978-5-453-00057-9 (1 экз.)

Кроме этого, студентам рекомендуется изучение периодических научных изданий: «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) программное обеспечение

1. 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.
2. 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.№1B08161103014721370444.
3. Microsoft Office Enterprise 2007 Russian Academic OPEN No Level. Номер Лицензии Microsoft 43364238.
4. Microsoft Windows XP Professional Russian Upgrade Academic OPEN No Level. Номер Лицензии Microsoft 41059241.
5. Office 365 профессиональный плюс для учащихся. Номер заказа: 36dde53d-7cdb-4cad-a87f-29b2a19c463e.
6. GeneDoc – программа для выравнивания аминокислотных последовательностей;
7. RasMol – программа для визуализации молекул (формат *.pdb);
8. Cn3D - программа для визуализации молекул (формат *.cgi);
9. ACD/ChemSketch – программа для создания и представления химических структур.

г) базы данных, информационно-справочные и поисковые системы

1. <http://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.
2. <http://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.
3. <http://www.biengi.ac.ru/analyz.htm> - Биоинформатика в Центре «Биоинженерия» РАН
4. <http://www.bioinformatix.ru/> - Биоинформатика, геномика, протеомика, биософт, имейджинг — портал по биоинформатике, имейджингу и биософту.
5. <http://www.chem.qmul.ac.uk/iupac/misc/naabb.html> - номенклатура IUPAC
6. <http://www.ebi.ac.uk/> - база данных EMBL EBI (European Bioinformatics Institute).
7. <http://www.expasy.ch/> - система анализа белка ExPASy (Expert Protein Analysis System, SwissProt, TrEMBL)
8. <http://www.iscb.org/> - Международное сообщество вычислительной биологии.
9. <http://www.matbio.org/> - электронный журнал «Математическая биология и биоинформатика».
10. <http://www.ncbi.nlm.nih.gov/> - сайт NCBI (National Center Biotech Information)
11. <http://www.ncbi.nlm.nih.gov/blast> - программа выравнивания последовательностей BLAST (Basic Local Alignment Sequence Tool)

12. <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html> - программа выравнивания последовательностей BLAST 2 SEQUENCES
13. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi> - поисковая система по базам данных ENTREZ
14. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide> - поиск последовательностей ДНК
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/>

9. Материально-техническое обеспечение дисциплины (модуля):

Материально-техническое обеспечение дисциплины «Основы биоинформатики» базируется на следующих ресурсах:

- Аудитория для проведения занятий лекционного типа. Аудитория оборудована: *специализированной (учебной) мебелью* на 12 посадочных мест; оборудована *техническими средствами обучения*: Проектор Epson EB-X03, Экран ScreenMedia, Ноутбук Lenovo"-2 шт, 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. служащими для представления учебной информации большой аудитории по дисциплине «Основы биоинформатики». *учебно-наглядными пособиями*, обеспечивающими тематические иллюстрации по дисциплине «Основы биоинформатики»: презентации в количестве 5 шт.

- 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. служащими для представления учебной информации большой аудитории по дисциплине «Основы биоинформатики».

- 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 блок Athlon 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 Pentium G3250, Монитор Samsung 740N – 1 шт.; Проектор BenQ MX503; экран ScreenVtdiaEcot. 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.

10. Образовательные технологии:

При реализации различных видов учебной работы дисциплины используются как стандартные методы обучения, так и интерактивные формы проведения занятий, доля которых составляет не менее 25 % аудиторных занятий. Доля лекционных занятий по дисциплине составляет 33 % от аудиторной нагрузки.

Стандартные методы обучения:

- Информационная лекция
- Практические занятия, предназначенные для освоения студентами базовых методов биоинформатики;
- Самостоятельная работа студентов;
- Консультации преподавателя;
- Подготовка ответов на контрольные вопросы;

Обучения с применением интерактивных форм образовательных технологий:

- кейс-метод – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия научной деятельности (разбор конкретных ситуаций);
- работа с молекулярно-биологическими базами данных (EMBL, GenBank, PDB и т.д.), обслуживающими их приложениями (BLAST, ENTREZ и др) и компьютерными программами молекулярного моделирования (RasMol, ACD/ChemSketch);
- информационно-коммуникационные образовательные технологии – лекция-визуализация, представление результатов деятельности (рефератов и отчетов) с использованием специализированных программных сред.

Все разделы дисциплины обеспечены контрольными материалами для текущей и промежуточной аттестации, которые представлены в электронно-образовательной среде Educa. Предусмотрена возможность проведения лекционных и практические занятия с использованием on-line видеоконференций (на платформах Zoom, BigBlueButton).

11. Оценочные средства (ОС):

11.1. Оценочные средства для входного контроля (могут быть в виде тестов с закрытыми или открытыми вопросами).

Тестовые задания для входного контроля

1. Пептидная связь в белках является: а) одинарной; б) двойной; в) частично одинарной и частично двойной; г) тройной
2. Какие связи образуют α -спираль во вторичной структуре белка? а) Вандер-Ваальса; б) гидрофобные; в) пептидные; г) водородные
3. Из пуриновых оснований в нуклеиновых кислотах обнаружены: а) аденин; б) Тимин; в) урацил; г) цитозин
4. Специализированные концевые районы хромосомной ДНК эукариот, состоящие из многократно повторяющихся коротких нуклеотидных последовательностей, называются: а) теломеры; б) хромомеры; в) палиндромы; г) спейсерные участки
5. Нуклеотиды в молекуле ДНК связаны друг с другом: а) О-гликозидной связью; б) 3,5 –фосфодиэфирной связью; в) N – гликозидной связью; г) α –1,4 –гликозидной связью
6. На один виток двойной спирали ДНК, находящейся в В-форме, приходится следующее число пар оснований: а) 5; б) 10; в) 15; г) 20.
7. Если одна цепь ДНК содержит фрагмент Г-Ц-Ц-А-А-Т-Г-Ц-А-Ц, то вторая цепь: а) А-А-Ц-А-Т-Т-Г-Г-Т-Г; б) Ц-Т-Г-Т-А-А-Т-А-Т-Г; в) Ц-Ц-А-А-Т-Г-А-Т-Г-Т; г) Ц-Г-Г-Т-Т-А-Ц-Г-Т-Г
8. Если содержание остатков тимина (от общего числа остатков) ДНК составляет 20%, то содержание гуанина составит: а) 40%; б) 35%; в) 25%; г) 30%
9. Выберите все, что характерно для РНК (1) и для ДНК (2): а) молекулярная масса млн дальтон и выше; б) одноцепочечная; в) двуцепочечная; г) небольшая молекулярная масса; д) содержит урацил; е) содержит Тимин; ж) содержит рибозу; з) содержит дезоксирибозу
10. Структурная единица нуклеиновой кислоты является: а) моноклеотид; б) аминокислота; в) нуклеозид; г) пуриновое или пиримидиновое основание.
11. В процессе репликации участвуют все ферменты, кроме: а) ДНК-полимеразы; б) РНК-праймазы; в) ДНК-лигазы; г) ДНКазы
12. Укажите для процесса репликации матрицу: а) тРНК; б) ДНК; в) мРНК; г) рРНК
13. Промотор это: а) специфическая последовательность ДНК, определяющая начало синтеза РНК; б) затравка для ДНК-полимеразы; в) последовательность ДНК, определяющая куда должен присоединиться репрессор; г) последовательность ДНК, кодирующая рРНК
14. Вырожденный генетический код это: а) Неперекрывающийся код; б) Кодирование одной аминокислоты двумя и более триплетами; в) Кодирование одной аминокислоты одним триплетом; г) Кодирование двух разных белков одной и той же последовательностью ДНК
15. Перекрывающийся код это: а) Некодирующие фрагменты ДНК; б) Кодирование одной аминокислоты двумя и более триплетами; в) Кодирование одной аминокислоты одним триплетом; г) Кодирование двух разных белков одной и той же последовательностью ДНК
16. Процессинг мРНК это: а) Участие мРНК в процессе трансляции; б) Участие мРНК в процессе обратной транскрипции; в) Секвенирование мРНК; г) Дефрагментация генов первичного транскрипта
17. Специфичность генетического кода состоит в: а) кодировании аминокислот более чем двумя различными триплетами; б) кодировании каждым триплетом только

- одной аминокислоты; в) наличии единого кода для всех живущих на земле существ; г) различии кода между эукариотами и прокариотами
18. Вырожденность генетического кода – это: а) кодирование одним триплетом только одной аминокислоты; б) кодирование одним триплетом одной либо нескольких аминокислот; в) кодирование одной аминокислоты несколькими триплетами; г) кодирование аминокислоты иницирующим или терминирующим триплетом
 19. Число возможных триплетов: а) 64; б) 28; в) 72; г) 128
 20. Транскрипция – это: а) Процесс самокопирования ДНК с образованием двух идентичных дочерних молекул; б) Процесс переписывания информации, содержащейся в РНК, в форме ДНК; в) Процесс переписывания информации, содержащейся в ДНК, в форме РНК; г) Процессинг мРНК
 21. Отличие процессов репликации и транскрипции: а) при репликации материнская двойная спираль ДНК разрушается, а при транскрипции – сохраняется; б) для функционирования основного фермента репликации необходимы ионы Mg^{2+} , а транскрипции – Fe^{2+} ; в) в активном центре полимеразы транскрипции находятся ионы Zn , а репликации – Li ; г) В ходе транскрипции образуются фрагменты Оказаки, а в ходе репликации – нет
 22. В процессе транскрипции участвует: а) только одна из двух цепей материнской молекулы ДНК – смысловая; б) только одна из двух цепей материнской молекулы ДНК – антисмысловая; в) любая из двух цепей материнской молекулы ДНК; г) Одновременно две цепи материнской молекулы ДНК.
 23. Первичный транскрипт - это: а) соединение РНК с белком в цитоплазме; б) ДНК, синтезированная полуконсервативным методом; в) совокупность всех видов РНК, синтезируемых в стадии транскрипции; г) РНК, полученная в результате модификации концов молекулы
 24. Пространственное соответствие (дополнительность) азотистых оснований друг другу в молекулах нуклеиновых кислот осуществляется по принципу: а) кооперативности; б) комплементарности; в) копланарности; г) аддитивности
 25. Окончание полипептида, содержащее аминокислотную группу, называется: а) С – конец; б) 3'-конец; в) N – конец; г) 5'-конец
 26. Простые белки состоят: а) только из нуклеотидов; б) только из аминокислот; в) из аминокислот и небелковых соединений; г) из аминокислот и неорганических соединений
 27. За единицу измерения количества информации принят: а) 1 биткойн; б) 1 бит; в) 1 байт; г) 1 Кбайт
 28. Производительность работы компьютера (быстрота выполнения операций) зависит от: а) объема оперативной памяти; б) объема внешней памяти; в) частоты процессора; г) размера экрана
 29. Минимальным объектом, используемым в текстовом редакторе, является: а) слово; б) пиксел; в) абзац; г) символ
 30. Объем текстовой информации в сообщении на 40 страницах (на странице 40 строк по 80 символов в строке) равен: а) 1 Мбайт; б) 120 Кбайт; в) 12 Кбайт; г) 125 Кбайт
 31. Даны утверждения: 1) компакт - диск является долговременной памятью; 2) сканер не является устройством ввода информации в компьютер; 3) принтер является устройством вывода информации; 4) модем является устройством приема - передачи данных. Из них верными являются: а) 1, 2, 3; б) 2, 3; в) 1, 3, 4; г) 2
 32. Каталог (папка) – это: а) раздел файловой системы, содержащий имена файлов и каталогов и сведения о их размещении на носителе информации; б) команда операционной системы, обеспечивающая доступ к данным; в) группа файлов, объединенных общим именем; г) устройство для хранения группы файлов и организации доступа к ним; д) путь, по которому операционная система определяет место файла.

33. Реляционная база данных может быть представлена в форме: а) гипертекста; б) алгоритма; в) иерархического каталога; г) таблицы
34. Что означает запись `ren *.txt *.doc`? а) переименовать файл с *.txt в файл *.doc; б) скопировать все файлы с расширением .txt в файлы с расширением .doc; в) переименовать все файлы с расширением .txt в файлы с расширением .doc; г) нет правильного ответа
35. Что понимается под информацией в кибернетике: а) СУБД; б) автоматизированная обучающая система; в) любая совокупность сигналов, воздействий или сведений; г) килобайты
36. К прикладному программному обеспечению относятся: а) новые языки программирования компиляторы к ним, интерфейсные системы; б) системы обработки текстов, электронные процессоры, базы данных; в) решение вопросов об анализе потоков информации в различных сложных системах; г) поисковые системы, глобальные системы хранения и поиска информации.
37. Бит – это: а) состояние диода: открыт или закрыт; б) 8 байт; в) запись текста в двоичной системе; г) наименьшая возможная единица информации.
38. Система счисления – это: а) подстановка чисел вместо букв; б) способ перестановки чисел; в) принятый способ записи чисел и сопоставления этим записям реальных значений чисел; г) правила исчисления чисел.
39. Процедура преобразования сообщения из одного алфавита в другой называется: а) кодом; б) кодировщиком; в) перекодировщиком; г) перекодировкой.
40. Как называется графическое представление алгоритма: а) последовательность формул; б) блок-схема; в) таблица; г) словесное описание?
41. В состав программного обеспечения ЭВМ не входят: а) система программирования; б) операционная система; в) аппаратные средства; г) прикладные программы.
42. Операционная система представляет из себя: а) комплекс программ специального назначения; б) комплекс аппаратных средств; в) совокупность ресурсов компьютера; г) комплекс инструментальных программ.
43. Поименованная совокупность данных, хранимых во внешней памяти, – это: а) файловая система; б) директорий; в) файл; г) запись.
44. Основными компонентами в составе операционной системе являются: а) утилиты, командный процессор, ядро; б) резидентные программы, утилиты; в) утилиты, командный процессор, центральный процессор; г) резидентные программы, ядро, командный процессор.
45. Все существующие языки программирования делятся на: а) функциональные и логические; б) русско- и не русскоязычные; в) процедурные и не процедурные; г) языки низкого и высокого уровня.
46. Транслятор – это программа, которая: а) переводит текст программы в машинный код; б) предоставляет средства просмотра и изменения значений переменных; в) подключает к исходному объектному модулю объектные модули соответствующих подпрограмм; г) распознает и выполняет команды программ.
47. Текстовый редактор Word – это: а) прикладная программа; б) базовое программное обеспечение; в) сервисная программа; г) редактор шрифтов.
48. Массив – это: а) запись множества переменных разного типа; б) неупорядоченная совокупность отличных друг от друга однотипных элементов; в) последовательность, состоящая из фиксированного числа однотипных элементов; г) тип одномерных величин

11.2. Оценочные средства текущего контроля формируются в соответствии с Положением о балльно-рейтинговой системе университета. Назначение оценочных средств ТК - выявить сформированность компетенций СПК-1,2,3; ПК-2.

Контрольные вопросы для текущего контроля

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 биоинформатики?
9. Каковы задачи биоинформатики?
10. В каких областях Биоинформатика находит применение?
11. Каким образом защищается интеллектуальная собственность в биоинформатике?
12. Что такое компьютер?
13. Что такое программное обеспечение?
14. Приведите примеры языков программирования и назовите их отличительные особенности.
15. Что такое HTML и BIOML?
16. Что такое Internet?
17. Опишите принцип работы Internet.
18. Что такое «Всемирная паутина»?
19. Что такое браузер? Назовите наиболее популярные обозреватели сети.
20. Дайте описание сети «EMBNet».
21. Какую пользу приносит система выборки последовательностей в биоинформатике?
22. Какова роль NCBI в обслуживании баз данных последовательностей?
23. Перечислите основные функции системы “Entrez”
24. Из каких операций состоит основная реакция секвенирования ДНК?
25. Опишите полный процесс секвенирования ДНК.
26. Какова роль открытой рамки считывания?
27. Опишите метод определения последовательности клона.
28. Что такое ярлыки экспрессируемых последовательностей?
29. Каким образом секвенируют IST?
30. Какие методы используют для секвенирования белков?
31. Опишите суть процесса гибридизации микроматриц ДНК.
32. Назовите сетевые ресурсы, полезные в анализе экспрессии генов на микроматрицах.
33. В чем состоит анализ экспрессии белков?
34. Какие подходы к обнаружению генов вам известны?
35. Приведите пример организмов, геномы которых были успешно расшифрованы.
36. Назовите главные цели проекта «Геном человека».
37. Какие результаты были получены в ходе работ над проектом «Геном человека»? Какова их практическая значимость?
38. Что такое база данных?
39. Какие типы баз данных вы знаете? Какие их функции?
40. Приведите примеры баз данных последовательностей нуклеиновых кислот. Для каких целей они созданы?
41. Каковы функции баз данных белковых последовательностей? Назовите несколько ресурсов.
42. Какие базы данных белковых последовательностей поддерживает PIR?
43. Что такое базы данных структур? Приведите примеры.
44. Что такое библиографическая база данных. Приведите несколько примеров.

45. Назовите специализированные пакеты анализа содержимого баз данных и укажите их возможности.
46. Что такое СУБД? Какие типы СУБД вы знаете?
47. Что такое выравнивание последовательностей?
48. Каковы цели выравнивания последовательностей?
49. Какие типы выравнивания последовательностей вы знаете?
50. Опишите этапы анализа точечной диаграммы.
51. Опишите принцип назначения счета мутациям, выпадениям и заменам.
52. Какие программы применяются для попарного выравнивания последовательностей в базах данных?
53. Что такое матрица процентов точечных мутаций?
54. Что такое динамическое программирование?
55. Какие алгоритмы используются в динамическом программировании? Чем они отличаются?
56. Почему для оценки выравниваемости последовательностей нельзя использовать традиционные статистические критерии?
57. На каких критериях основано сравнение последовательностей?
58. Что такое Хеммингово расстояние?
59. Что такое расстояние Ливенштейна?
60. Что такое множественное выравнивание последовательностей?
61. Перечислите основные этапы в построении множественного выравнивания.
62. Какие программы применяют для множественного выравнивания?
63. В каких случаях необходимо применять методы предсказания структуры?
64. Какие стратегии используются в предсказании генов?
65. Какие методы предсказания структуры иРНК вам известны?
66. Приведите примеры некоторых из наиболее употребляемых методов предсказания генов.
67. Чем вызвана необходимость предсказания белковых структур?
68. Какие методы разработаны для предсказания вторичной структуры белка?
69. В чем заключается «стремление» аминокислот к формированию α и β -структур?
70. Какова разница между методами предсказания *ab initio* и методами, основанными на эмпирических знаниях?
71. Каким образом осуществляется сравнительное моделирование?
72. Назовите основные шаги сравнительного моделирования.
73. Что такое метод протягивания?
74. В чем заключается энергетический подход к предсказанию структур молекул?
75. В чем состоит предсказание функций белка? Приведите примеры программ, позволяющих предсказывать белки.
76. Какие преимущества дает визуальное отображение молекул?
77. Приведите примеры программ визуализации молекулярных структур.
78. Как вы понимаете разницу между понятиями «гомология» и «подобие»?
79. Чем отличаются ортологи, паралоги и ксенологи?
80. Что такое филогения?
81. В чем сущность фенетического подхода?
82. Назовите характерные особенности кладистического подхода.
83. Дайте определение следующим терминам: оргграф, узел (вершина), дуга, путь, контур.
84. Что такое филогенетическое дерево? Каковы его отличительные свойства?
85. Какие виды филогенетических деревьев вы знаете?
86. Какие допущения приняты в построении филогенетических деревьев?
87. Какие методы применяют в филогенетике?
88. В чем молекулярная филогенетика превосходит традиционную?

89. Какие базы данных используют в филогенетическом анализе?
90. Какие существуют методы поиска медицинских препаратов?
91. Каким способом можно опознать мишени для медикаментозного воздействия?
92. Перечислите желательные качества опытного препарата.
93. В чем заключается оптимизация опытного препарата?
94. Какие стадии должен пройти лекарственный препарат при клинических испытаниях?
95. Что такое фармакоинформатика?
96. Охарактеризуйте предметную область фармакоинформатики.
97. Каково назначение химических библиотек?
98. Приведите примеры программ поиска в химических базах данных.

Практические задания для текущей аттестации и контроля СРС

1. Визуализация молекул с помощью программы ACD/ChemSketch
2. Визуализация молекул с помощью программы RasMol
3. Работа с системой PubMed
4. Поиск белков в базах данных
5. Выравнивание последовательностей с помощью программы GeneDoc
6. Выравнивание последовательностей с помощью программы BLAST
7. Редактирование результатов выравнивания
8. Поиск белка по его гомологу
9. Множественное выравнивание
10. Построение филогенетических деревьев

Демонстрационный вариант контрольной работы №1

1. С использованием программы BLAST было идентифицировано 4 гомологичных последовательности. Какая из них более всего соответствует последовательности запроса?
а) E value = 0.1, % identical residues = 16; б) E value = 0.1, % identical residues = 30; в) E value = 0.001, % identical residues = 16; г) E value = 10, % identical residues = 16
2. Наиболее широко используемый метод выравнивания белковых последовательностей (BLOSUM матрицы) разработан на основе: а) Генетического кода; б) физико-химических свойств; в) простого тождества; г) статистики наблюдаемых замен в множественного выравнивания последовательностей

Демонстрационный вариант контрольной работы № 2

1. Скорость замены в определенной области ДНК длиной 1000 оснований оценивается как 10^{-9} оснований в год. Если два вида разошлись примерно 10 миллионов лет назад, доля сайтов, которые отличаются между ними должно быть примерно равно: а) 1%; б) 2%; в) 20%; г) 75%
- Имеется 2. две последовательности ДНК: CAGCAT и CGACAT. Эти последовательности выравниваются с использованием алгоритма глобального выравнивания, где за совпадение добавлялся 1 балл, 0 - за замену и штраф за гэп - 0,2. Какое из следующих утверждений является верным?

11.3. Оценочные средства для промежуточной аттестации

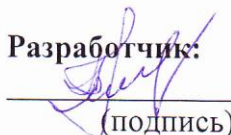
Промежуточная аттестация проходит в форме экзамена (7 семестр). К экзамену допускаются студенты, выполнившие в полном объеме аудиторную нагрузку, самостоятельную работу и успешно сдавшие промежуточную аттестацию. Студенты, имеющие задолженность, должны выполнить все обязательные виды деятельности, и только затем допускаются к сдаче экзамена.

Примерный перечень вопросов к экзамену

1. Биоинформатика: предмет, цели, задачи, прикладное значение.
2. Биоинформатика в историческом аспекте. Роль физико-химическим и молекулярно-биологических методов исследования биополимеров. Закон Мура и проект «Геном человека».
3. Геномика и протеомика.
4. Компьютер, компьютерная программа, программное обеспечение. Операционные системы (BIOS, DOS, Windows, Unix).
5. Языки программирования - HTML, Java Script, Java, PERL, BSML, BIOML.
6. Интернет: история возникновения и развития, современная структура.
7. Сетевые протоколы - UUCP, POP, FTP, TELNET, TCP/IP, HTTP.
8. Виды подключения к Интернет. IP-адреса и доменные номера.
9. Всемирная паутина (WWW). Веб-страницы и веб-узлы. Объектная сеть. ORB. CORBA. IDL.
10. Программы-обозреватели: Lynx, Mosaic, Netscape Navigator, Internet Explorer, Opera, Mozilla. Гиперссылки. URL. XML. DBMS. Зеркала и Интранет.
11. EMBnet. Центры и узлы. SRS как сетевой обозреватель баз данных EMBnet. NCBI и Entrez.
12. Методы картографирования генома – виды карт и общая характеристика методов их построения.
13. Основные технические подходы к секвенированию ДНК: гибридные молекулы ДНК, амплификация, кДНК-синтез, метод Сенгера.
14. Ярлыки экспрессируемых последовательностей (EST) – назначение и принцип метода.
15. Методы секвенирования белков – прямые и косвенные.
16. Методы определения пространственной структуры белка: практические и теоретические.
17. Методы анализа экспрессии генов: Нозерн- и Вестерн-блоттинг, ДНК-чипы. Анализ экспрессии белков.
18. Базы данных - функции и классификация. Записи базы данных. Современные тенденции в структурировании БД.
19. Система управления базами данных. Компоненты, функции и типы СУБД. Язык структурированных запросов (SQL).
20. Базы данных последовательностей нуклеиновых кислот (EMBL, DDBJ, GenBank, GSDB). Специализированные БД.
21. Базы данных белковых последовательностей (PIR, SWISS-PROT, TrEMBL, Protein Research Foundation).
22. Базы данных структур (PDB, MSD, SCOP, CATH, NDB, CSD, BMRB, FSSP).
23. Вторичные базы данных (PROSITE. PRINTS. BLOCKS. IDENTIFY. KEGG). Библиографические базы данных.
24. Специализированные средства анализа баз данных. Пакеты GCG, EGCG, Staden, Sequencher, SYNERGY, Pangea System, EMBOSS, DALI.
25. Выравнивание – цели, задачи, основные понятия и определения. Типы выравнивания - глобальное и локальное.
26. Общие принципы выравнивания. Расстояние как критерий меры различия последовательностей.
27. Счет выравнивания. Матрица процентов точечных мутаций. Матрицы блочных замен аминокислот.
28. Методы попарного выравнивания последовательностей. Точечная матрица. Динамическое программирование. Алгоритмы Нидлмана-Вунша и Смита-Уотермана.

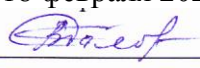
29. Методы попарного выравнивания последовательностей. Матрица переходов. Метод k-кортежей. Алгоритмы программ FASTA и BLAST.
30. Множественное выравнивание последовательностей – цели, задачи, основные этапы.
31. Методы множественного выравнивания (Profile, Blocs, Indicator, PSI-BLAST, Скрытые марковские модели). Программы автоматического выравнивания.
32. Алгоритмы выравнивания структур. Метод двойного динамического программирования. Программное обеспечение.
33. Аннотирование – цели, задачи, виды аннотаций.
34. Стратегии предсказания генов. Программы предсказания генов. Предсказание вторичной структуры РНК.
35. Предсказание структуры белка - стратегии и методы. Предсказание вторичной структуры.
36. Предсказание структуры белка - стратегии и методы. Предсказание трехмерной структуры - сравнительное моделирование, протягивание, энергетический подход.
37. Гомология и подобие. Виды гомологов. Филогения и родство. Основные подходы к филогенетическому анализу. Критерии филогенетического анализа
38. Эволюционное дерево и теория графов. Клада, таксон и узел. Свойства и виды филогенетических деревьев.
39. Основные этапы и методы филогенетического анализа. Методы расстояний. Групповой метод невзвешенных пар с вычислением среднего арифметического. Алгоритм объединения соседей. Метод Фитча-Маргобиаша.
40. Основные этапы и методы филогенетического анализа. Метод минимальной эволюции. Методы подобия. Метод максимальной экономичности. Метод максимального правдоподобия.
41. Молекулярные подходы к определению филогении. Типы макромолекул, используемых для филогенетического анализа.
42. Базы данных филогенетического анализа. RAUP, PHYLIP, PALI.

Разработчик:

 доцент Приставка А.А.
(подпись)

Программа рассмотрена на заседании кафедры физико-химической биологии, биоинженерии и биоинформатики

Протокол № 11 от 18 февраля 2020 г.

Зав.кафедрой  проф. Саловарова В.П.

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