



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

УТВЕРЖДАЮ

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



Рабочая программа дисциплины

Наименование дисциплины: Б1.В.ДВ.04.02 «ПРАКТИЧЕСКАЯ
БИОИНФОРМАТИКА»

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 № 4 от 10 марта 2020 г.

Председатель _____
проф. Матвеев А.Н.

Протокол № 11 от 18 февраля 2020 г.

Зав. кафедрой Саловарова В.П.

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)	7
5.3 Разделы и темы дисциплин (модулей) и виды занятий	7
6. Перечень семинарских, практических занятий, лабораторных работ, план самостоятельной работы студентов, методические указания по организации самостоятельной работы студентов.	8
7. Примерная тематика курсовых работ (проектов) (при наличии)	10
8. Учебно-методическое и информационное обеспечение дисциплины (модуля) :	10
а) основная литература;	10
б) дополнительная литература;	10
в) программное обеспечение;	10
г) базы данных, поисково-справочные и информационные системы	11
9. Материально-техническое обеспечение дисциплины (модуля).	11
10. Образовательные технологии	12
11. Оценочные средства. (ОС).	13

1. Цели и задачи дисциплины (модуля):

Целью освоения учебной дисциплины «Практическая биоинформатика» является:

- Изучить базовые методы биоинформатики, применяемые при обработке и извлечения информации из биологических последовательностей (ДНК, РНК, белки), научиться применять полученные знания навыки в практической работе.

Задачи дисциплины:

- изучить методы автоматизированной обработки больших массивов текстовой информации, применяемые для обработки биологических последовательностей
- изучить методы выравнивания биологических последовательностей, основанные на разных алгоритмах и процедурах, в том числе с использованием высокопроизводительных пероральных вычислительных технологий
- изучить технологии ДНК и белкового К-мерного анализа, применяемые для сравнения больших массивов биологических последовательностей без процедуры выравнивания
- ознакомиться с технологиями секвенирования нового поколения (NGS), применяемыми для получения больших массивов данных расшифрованных нуклеотидных последовательностей, изучить варианты использования NGS технологии в биологических исследованиях
- изучить методы анализа данных, полученных с помощью технологий секвенирования нового поколения (NGS)
- ознакомиться с методами машинного обучения, применимыми для анализа и систематизации биологических последовательностей
- изучить методы анализа и аннотации полногеномных данных.

2. Место дисциплины в структуре ОПОП:

Дисциплина «Практическая биоинформатика» является дисциплиной по выбору вариативной части учебного плана подготовки бакалавров по направлению 06.04.01 «Биология».

Дисциплина рассматривает различные методы и подходы анализа биологических последовательностей (ДНК, РНК, белки) для извлечения информации об их структуре, функции и сравнительного анализа. Методы биоинформатики для анализа биологических последовательностей быстро развиваются, оказывая влияние на современное состояние геномики, метагеномики, транскриптомики и метаболомики. Используя методы практической биоинформатики можно извлечь большие объемы информации о структуре и функции биосистем от клеточного уровня до уровня сложных биологических сообществ.

Студент, приступающий к изучению дисциплины «Практическая биоинформатика», должен знать основные методы, теории и законы предшествующих естественнонаучных дисциплин:

- математика: матричное исчисление, теорию вероятностей;
- общая биология: основные понятия физиологии, биохимии, генетики, экологии эволюционной теории;
- основы программирования: алгоритмы, структурное программирование;
- генетика: законы наследственности и изменчивости, мутационный процесс;
- молекулярная биология: процессы реализации наследственной информации, биосинтез белка.
- информатика: навыки использования персонального компьютера для проведения вычислений, работы с электронными таблицами, работы с текстом и компьютерной графикой

- экология: понятия сообщества, взаимодействия организмов в сложных сообществах и с окружающей средой.

Данная дисциплина является необходимой основой для выполнении НИР и 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 методы автоматизированного анализа текстовой информации
- цель, основные задачи и области применения методов анализа биологических последовательностей;
- особенности методов анализа биологических последовательностей, возможные варианты их применения;
- классификацию методов геномного анализа и варианты их применения для анализа геномных данных.

Уметь:

- работать с различными форматами хранения данных, применяемых в биоинформационных исследованиях;
- формализовать задачу анализа большого массива геномных данных, разбить ее на отдельные подзадачи;
- выбирать адекватные методы для построения конвейера для анализа большого массива геномных данных;
- использовать языки программирования для автоматизации обработки геномных данных;
- осуществлять интерпретацию результатов анализа геномных данных.

Владеть:

- основными методами анализа бионических последовательностей, применяемыми в анализе больших массивов геномных данных;
- методами математической статистики и машинного обучения, применяемыми в сравнительном анализе геномных данных;
- методами распределенных параллельных вычислений и интернет-технологий, применяемыми в ускорении обработки геномных данных;
- навыками совершенствования своих профессиональных качеств в области анализа биологических последовательностей и геномных данных.

4. Объем дисциплины (модуля) и виды учебной работы

Вид учебной работы	Всего часов / зачетных единиц	Семестры
		7
Аудиторные занятия (всего)	56/1,5	56/1,5
Из них объем занятий с использованием электронного обучения и дистанционных образовательных технологий	11/0,31	11/0,31
В том числе:	-	-
Лекции	18/0,5	18/0,5
Практические занятия (ПЗ)	36/1,0	36/1,0
Семинары (С)	-	-
Лабораторные работы (ЛР)	-	-
КСР	2/0,06	2/0,06
Самостоятельная работа (всего)	88/2,44	88/2,44
В том числе:		
Курсовой проект (работа)	-	-
Расчетно-графические работы	88/2,44	88/2,44
Реферат (при наличии)	-	-
<i>Другие виды самостоятельной работы</i>	-	-
Вид промежуточной аттестации (зачет, экзамен)	зачет	зачет
Контактная работа (всего)	56/1,56	56/1,56
Общая трудоемкость	часы	144
	зачетные единицы	4

5. Содержание дисциплины (модуля)

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

Тема 1. Регулярные выражения для обработки тестов анализ нуклеотидных аминокислотных последовательностей с помощью регулярных выражений. В разделе рассматриваются связанные с обработкой строк с помощью системы регулярных выражений для анализа больших объёмов текстовой информации. Основой регулярных выражений является поиск подстроки и различных вариантов подстроки в строках. Данный подход позволяет находить различные мотивы в нуклеотидных и аминокислотных последовательностях, анализировать стандартные наборы текстовой информации, содержащей аннотацию полных геномов и протеомов.

Тема 2. Выравнивание аминокислотных и нуклеотидных

последовательностей. Рассматриваются алгоритмы локального выравнивания Смита-Ватермана и глобального выравнивания Нидлмана-Вунша. Изучается анализ и поиск гомологии в нуклеотидных и аминокислотных последовательностях с помощью BLAST алгоритмов и использование этих алгоритмов для анализа геномных данных.

Тема 3. Методы ДНК и аминокислотных К-мер для анализа биологических последовательностей. В разделе изучаются методы, основанные на анализе сходства между биологическими последовательностями, путем исследований сходства статистики встречаемости в них коротких фрагментов ДНК и аминокислотных последовательностей - К-мер рассматриваются различные варианты применения этих методов для сравнения геномов и протеомов различных организмов.

Тема 4. Технологии расшифровки – секвенирования нуклеотидных последовательностей нового поколения (NGS), первичный анализ данных секвенирования. В данной теме рассматриваются различные технологии и платформы для массового параллельного секвенирования ДНК и РНК нового поколения. Изучается формат данных fastq – стандарт для хранения и представления данных NGS секвенирования и различные алгоритмы коррекции ошибок секвенирования на основе первичных данных формата fastq.

Тема 5. Различные варианты использования технологии расшифровки нуклеотидных последовательностей нового поколения в биологических исследованиях. Рассматриваются различные варианты использования технологии NGS: для расшифровки полных геномов, исследование естественных и искусственных сообществ микроорганизмов путем расшифровки ампликонов стандартных генетических маркеров, метагеномный анализ – анализ смеси ДНК и РНК сообщества организмов, транскриптомный и метатранскриптомный анализ, применение NGS технологий в генетике популяций.

Тема 6. Анализ данных метабаркодиновых исследований, проведенных с помощью технологий расшифровки нуклеотидных последовательностей нового поколения. В разделе подробно с практическими примерами разбираются различные варианты анализа таксономического состава сообществ микро и макро организмов на основе расшифровки ампликонов стандартных генетических маркеров.

Тема 7. Сборка полных геномов, методы и подходы. Изучаются различные алгоритмы для сборки полных геномов организмов из данных NGS секвенирования. Рассматривается анализ метагеномных данных путем сборки фрагментов геномов из ДНК и РНК смеси организмов. На практике проводится сборка полного митохондриального генома по результатам NGS секвенирования.

Тема 8. Машинное обучение в биоинформатике. В данном разделе рассматриваются различные варианты кластерного анализа, Байесовского наивного классификатора, нейронные сети для поиска закономерностей формирования первичной структуры нуклеотидных и аминокислотных последовательностей, применяемых при аннотации и анализе полных геномов.

Тема 9. Аннотация полных геномов на основе методов машинного обучения. Рассматриваются различные варианты аннотации полных геномов с использованием методов выравнивания биологических последовательностей и методов машинного обучения. На практике рассматривается аннотация полного генома природного бактериофага.

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						
		1	2	3	5	6	7	9
1.	Научно-исследовательская работа	1	2	3	5	6	7	9
2.	Выпускная квалификационная работа	1	2	3	6	8	9	

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.		Тема 1. Регулярные выражения для обработки тестов, анализ нуклеотидных аминокислотных последовательностей с помощью регулярных выражений.	2	4			10	16
2.		Тема 2. Выравнивание аминокислотных и нуклеотидных последовательностей.	2	4			10	16
3.		Тема 3. Методы ДНК и аминокислотных К-мер для анализа биологических последовательностей.	2	4			10	16
4.		Тема 4. Технологии расшифровки – секвенирования нуклеотидных последовательностей нового поколения (NGS), первичный анализ данных секвенирования.	2	4			9	15
5.		Тема 5. Различные варианты использования технологии расшифровки нуклеотидных последовательностей нового поколения в биологических исследованиях	2	4			9	15
6.		Тема 6. Анализ данных метаброминовых исследований, проведенных с помощью технологий расшифровки нуклеотидных последовательностей нового поколения	2	4			10	16
7.		Тема 7. Сборка полных геномов, методы и подходы.	2	4			10	16
8.		Тема 8. Машинное обучение в биоинформатике.	2	4			10	16
9.		Тема 9. Аннотация полных геномов на основе методов машинного обучения.	2	4			10	16

6. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины	Наименование семинаров, практических и лабораторных работ	Трудо-емкость (часы)	Оценочные средства	Формируемые компетенции
1.	Тема 1	Решение задач в области программирования для работы с регулярными выражениями. Разработка алгоритмов анализа больших массивов текстовой информации	4	Контрольные вопросы и задачи	ПК-1; СПК-2; СПК-3
2.	Тема 2	Практическая работа по выравниванию аминокислотных и нуклеотидных последовательностей с помощью локальных программ и онлайн сервисов.	4	- « -	ПК-1; СПК-2; СПК-3
3.	Тема 3	Практическая работа по сравнению биологических последовательностей с помощью K-мерного анализа	4	- « -	ПК-1; СПК-2; СПК-3
4.	Тема 4	Анализ первичных данных NGS секвенирования в формате fsatq.	4	- « -	ПК-1; СПК-2; СПК-3
5.	Тема 5	Планирование эксперимента по использованию NGS при решении различных биологических задач	4	- « -	ПК-1; СПК-2; СПК-3
6.	Тема 6	Анализ наборов данных, полученных при метабаркодинговых исследованиях	4	- « -	ПК-1; СПК-2; СПК-3
7.	Тема 7	Сборка полных геномов с помощью программного обеспечения SPADES	4	- « -	ПК-1; СПК-2; СПК-3
8.	Тема 8	Решение практических задач по применению методов машинного обучения при сравнительном анализе биологических последовательностей	4	Контрольные вопросы, вопросы	ПК-1; СПК-2; СПК-3
9	Тема 9	Решение практических задач по аннотации сборок геномных данных с помощью онлайн приложений и сервисов.	4	Контрольные вопросы, вопросы	ПК-1; СПК-2; СПК-3

6.1. План самостоятельной работы студентов

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемая литература	Количество часов
1-2	Тема 1.	Решение задач Контрольные вопросы	Решение домашних заданий по теме регулярные выражения для анализа больших текстов	1, 2	2
3-4	Тема 2.	-«-	Решение домашних заданий по теме выравнивание биологических последовательностей	1, 2	2
5-6	Тема 3.	-«-	Решение домашних заданий по теме камерный анализ биологических последовательностей	1, 2	1
7-8	Тема 4.	-«-	Решение домашних заданий по анализу первичных данных NGS секвенирования.	1, 2	2
9-10	Тема 5.	-«-	Решение домашних задач по теме планирование биологического эксперимента с использованием NGS технологий	1, 2	2
11-12	Тема 6.	-«-	Решение домашних задач по теме анализ данных метабаркодинговых исследований	1, 2	1
13-14	Тема 7.	-«-	Решение задач по геномной сборке первичных данных NGS секвенирования	1, 2	1
15-16	Тема 8.	Подготовка докладов для выступления на практических работах	Подготовка к выступлению с докладом на практическом занятии по методам машинного обучения при анализе биологических последовательностей	1, 2	2
17-18	Тема 9.	Подготовка докладов для выступления на практических работах	Подготовка к выступлению с докладом на практическом занятии по теме – аннотация первичных сборок геномного секвенирования	1, 2	2

6.2. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студента преследует следующие цели:

- совершенствование навыков самообразовательной работы для повышения уровня образования и совершенствования профессиональных навыков;
- углубление и расширение знаний по предмету.

По дисциплине «Практическая биоинформатика» предлагаются следующие формы самостоятельной работы:

- а) решение расчетных задач
- б) анализ методической литературы для подготовки к контрольным вопросам;
- в) подготовка к докладам для выступления на практическом занятии;

Темы докладов для самостоятельных работ

1. Исследование и диагностика бактериальных патологий человека с помощью метабаркодинга на основе маркерного гена 16S рибосомальной РНК.

2. Исследование природных микробиомов с помощью метабокодинга на основе маркерного гена 16S рибосомальной РНК.
3. Транскриптомный анализ на основе NGS технология для исследования и диагностики метаболических заболеваний у человека.
4. Исследование природных вирусных сообществ с помощью метгеномного анализа на основе NGS технологий
5. Базы данных UNIPROT, KEGG, Pfam и PDB для хранения и анализа информации о первичных аминокислотных последовательностях белков и пространственных белковых структур.

Критерии оценки ответов на контрольные вопросы и задачи

- точность ответа на поставленный вопрос;
- раскрытие (определение) рассматриваемого понятия (определения, проблемы, термина);
- четкость структуры ответа;
- самостоятельность, логичность изложения;
- наличие выводов, сделанных самостоятельно.

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. Дурбин, Р., Эдди, Ш., Крэг, А., Митчисон, Г. Анализ биологических последовательностей. М.-Ижевск: РХД, 2006. – 480 с. - ISBN 5-93972-559-7
2. Игнасимуту, С., Чумичкин, А. А., Садовничий, В. А., Скулачев, В. П. Основы биоинформатики. Издательство: Регулярная и хаотическая динамика. 2007. – 320 с. - ISBN 978-5-93972-620-7

б) дополнительная литература

1. Бородовский М., Екишева С.. Задания и решения по анализу биологических последовательностей М.-Ижевск: РХД, 2008.- 440 с. - ISBN 978-5-93972-644-3
2. Леск А. (2009). Введение в биоинформатику. М.: БИНОМ. Лаборатория знаний, 2009. — 318 с. - ISBN 78-5-94774-501-6
3. Ребриков Д.В. NSG. Высокопроизводительное секвенирование. Издательство: БИНОМ. - 2014. – 230 с. - ISBN: 978-5-9963-1784-4.

в) 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. Форум 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.

Свободно распространяемый интерпретатор для языка программирования «R»

Свободно распространяемый интерпретатор для языка программирования «Phyton»

Свободно распространяемое программное обеспечение SPADES для сборки полных геномов.

Пакет программ BLAST для анализа биологических последовательностей.

г) базы данных, информационно-справочные и поисковые системы

1. <http://bioinformatics.ru/> - сайт сборник образовательных статей по биоинформатике
2. <https://habr.com/ru/post/403901/> - образовательные лекции по биоинформатике: от статистики до генетических конструкций
3. <http://www.biometrika.tomsk.ru/> - электронный журнал «Биометрика» для медиков и биологов – сторонников доказательной биомедицины. Содержит большое количество статей и иных материалов, посвященных математическим моделям в биологии.
4. <http://www.matbio.org/about.php> - сайт журнала «Математическая биология и биоинформатика». Содержит большое количество статей в pdf – формате.
5. <https://www.elibrary.ru> – электронная библиотека научных статей, монографии и материалов конференций, выпущенных Российскими учеными.
6. <https://pubmed.ncbi.nlm.nih.gov/> - международная база данных научных статей и монографий, посвященная различным вопросам биологии.
7. <https://apps.webofknowledge.com> – международная база данных, индексирующая научные публикации в высокорейтинговых изданиях
8. <https://www.r-project.org/> - сайт проекта по R статистики
9. <http://qsar4u.com/pages/rtutorial.php>- Краткий курс "Введение в R и моделирование с R" (Pavel Polishchuk)
10. <https://journal.r-project.org/> - сайт журнала по статистическим методам на R, «*The R Journal*»
11. <https://ru.stackoverflow.com/questions/506597/%D0%9A%D0%BD%D0%B8%D0%B3%D0%B8-%D0%B8-%D1%83%D1%87%D0%B5%D0%B1%D0%BD%D1%8B%D0%B5-%D1%80%D0%B5%D1%81%D1%83%D1%80%D1%81%D1%8B-%D0%BF%D0%BE-%D1%8F%D0%B7%D1%8B%D0%BA%D1%83-r> – список электронных ресурсов по языку программирования R
12. <https://rpubs.com/AllaT/rprog-ggplot2> - электронный учебник по визуализации научных данных с помощью пакета ggplot2 для языка программирования R.
13. https://ahmedushka7.github.io/R/scripts/hse_data_analysis/sem_10/ggplot2.html - руководство по визуализации данных в пакете ggplot2 для языка программирования R.
- 14.

9. Материально-техническое обеспечение дисциплины (модуля):

Материально-техническое обеспечение дисциплины «Практическая биоинформатика» базируется на следующих ресурсах:

- Аудитория для проведения занятий лабораторного типа. Компьютерный класс (учебная аудитория). 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. *учебно-наглядными пособиями*, обеспечивающими тематические иллюстрации по дисциплине «Практическая биоинформатика» в количестве 8 шт., презентации по каждой теме программы.

- 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.; 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. *учебно-наглядными пособиями*, обеспечивающими тематические иллюстрации по дисциплине «Практическая биоинформатика» в количестве 8 шт., презентации по каждой теме программы.

- Помещения для хранения и профилактического обслуживания учебного оборудования. Аудитория оборудована: специализированной мебелью на 11 посадочных мест; Шкаф для документов - 3 шт.; Сейф – 1 шт.; Шкаф-купе - 2 шт.; Принтер цв. Canon LBR-5050 Laser Printer; Принтер Canon LBP-3010; Ноутбук Lenovo G580 – 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. Лекционных занятий по дисциплине учебным планом не предусмотрено.

Стандартные методы обучения:

- Практические занятия, предназначенные для освоения математических методов в биологии;
- Самостоятельная работа студентов;
- Консультации преподавателя;
- Подготовка ответов на контрольные вопросы;

Обучение с применением интерактивных форм образовательных технологий:

- кейс-метод – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия научной деятельности (разбор конкретных ситуаций);
- 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 практических

занятий 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).

Вопросы и задания для входного контроля

1. Первичная и вторичная структура ДНК и РНК
2. Аминокислотные последовательности и белки, разновидности белков по структуре и функции.
3. Биосинтез белка в клетке.
4. Механизм дупликации ДНК.
5. Мутационный процесс, изменение первичной последовательности ДНК и белков
6. Репарация ДНК.
7. Генетические процессы в популяциях, влияющих на полиморфизм ДНК, РНК и белков.
8. Что такое вирусы, механизм размножения и распространения вирусов.
9. Определение понятия вероятностей.
10. Законы распределения случайных чисел.
11. Как можно рассчитать расстояние между двумя точками на координатной плоскости?

11.2 Контрольные вопросы для текущего контроля.

1. Определение понятия «регулярные выражения»
2. Реализация регулярных выражений в языке программирования R
3. Синтаксические конструкции в регулярных выражениях
4. Регулярные выражения для сопоставления двух строк
5. Определение понятия гомологичных последовательностей.
6. Меры расстояния между гомологичными последовательностями.
7. Процедура выравнивания двух последовательностей
8. Алгоритмы выравнивания двух последовательностей
9. Задачи процедуры выравнивания нескольких последовательностей (множественное выравнивание)
10. Принципы работы алгоритма BLAST для попарного выравнивания последовательностей
11. Принципы работы алгоритма MAFT для множественного выравнивания последовательностей
12. Структурное выравнивание РНК последовательностей
13. Структурное выравнивание аминокислотных последовательностей
14. Сравнение нуклеотидных последовательностей с помощью метода ДНК К-мер
15. Сравнение аминокислотных последовательностей с помощью метода белковых К-мер
16. Метод К-мерного анализа для кластеризации аминокислотных и нуклеотидных последовательностей
17. Технология секвенирования ДНК методом Сенгера
18. Технология секвенирования ДНК методом Illumina
19. Технология секвенирования ДНК PacBio
20. Нанопоровое секвенирование ДНК

21. Формат fastq для хранения первичных данных секвенирования ДНК нового поколения
22. Контроль качества секвенирования ДНК на основе поколения на основе fastq формата данных

Формат отчета по самостоятельной работе

По каждому практическому заданию, предназначенному для самостоятельного решения, студент предоставляет отчет в формате Microsoft Word. В отчете должны быть отражены следующие пункты:

1. Формулировка задания для самостоятельного решения
2. Список программ и интернет-ресурсов, используемых для решения задачи
3. Графическая визуализация конвертера для решения задачи в виде алгоритмической блок-схемы
4. Описание результата, полученного при выполнении задания
5. Графическая визуализация полученного результата.

11.3. Оценочные средства для промежуточной аттестации

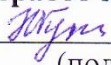
Промежуточная аттестация проходит в форме ответов на вопросы по темам и отчетам по решению домашних задач в течение семестра. Студенты, выполнившие в полном объеме аудиторную нагрузку, самостоятельную работу и успешно сдавшие промежуточную аттестацию и сделавшие доклад по выбранной теме на практическом занятии получают зачет автоматически. Студенты, имеющие задолженность, должны выполнить все обязательные виды деятельности, а затем сдают зачет по вопросам в виде собеседования.

Примерный перечень вопросов к зачету

1. Регулярные выражения и их использование в анализе биологических последовательностей.
2. Реализация регулярных выражений с помощью языка программирования R.
3. Попарное выравнивание нуклеотидных последовательностей.
4. Попарное выравнивание аминокислотных последовательностей.
5. Алгоритмы множественного выравнивания биологических последовательностей.
6. Метод К-мер для сравнительного анализа нуклеотидных и аминокислотных последовательностей.
7. Кластеризация последовательностей на основе результатов К-мерного анализа
8. Классическая технология Сенгера для расшифровки ДНК.
9. Технологии расшифровки ДНК нового поколения – NGS.
10. Первичная обработка данных технологий секвенирования ДНК нового поколения – контроль качества прочтения, удаление и фильтрация недостоверных данных.
11. Баркодинговые исследования – детектирование видов организмов на основе первичной последовательности маркерных генов.
12. Метабаркодинг – анализ видового состава сообществ организмов в помощью NGS технологий.
13. Полногеномное секвенирование с помощью NGS технологий.
14. Метагеномное секвенирование для исследования сообществ микроорганизмов.
15. NGS технологии в популяционно-генетических исследованиях.
16. Сборка полных геномов, методы сборки полных геномов.
17. Сборщик полных геномов SPADES? Запуск сборщика, опции командной строки для запуска.

18. Классификация методов машинного обучения, применимых для анализа биологических последовательностей.
19. Кластерный анализ для сравнения биологических последовательностей.
20. Наивный Байесовский классификатор для сравнения биологических последовательностей.
21. Нейронные сети для сравнения биологических последовательностей.
22. Методы аннотации полных геномов.
23. Использование методов машинного обучения для аннотации полных геномов

Разработчик:

 доцент Букин Ю.С.
(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.