



МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Иркутский государственный университет»
(ФГБОУ ВО «ИГУ»)
Биолого-почвенный факультет

УТВЕРЖДАЮ
Декан биолого-почвенного
факультета А.Н. Матвеев/
« 14 » сентября 2023 г.

Рабочая программа дисциплины

2.1.1.3 МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ, БИОИНФОРМАТИКА

Научная специальность: 1.5.8. Математическая биология, биоинформатика

Форма обучения: очная

Согласовано с УМК биолого-почвенного
факультета
протокол № 6 от « 14 » 06.09 2023 г.

Председатель УМК, проф. Матвеев А.Н.

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

« 6 » 06 2023 г. Протокол № 20
Зав. кафедрой, проф. Саловарова В.П.

Иркутск 2023 г.

Содержание

1.	Цели и задачи дисциплины (модуля)	3
2.	Требования к результатам освоения дисциплины (модуля)	3
3.	Объем дисциплины (модуля) и виды учебной работы	4
4.	Содержание дисциплины (модуля)	5
4.1	Содержание разделов и тем дисциплины (модуля)	
4.2	Разделы и темы дисциплин (модулей) и виды занятий	
5.	Примерная тематика рефератов (при наличии)	8
6.	Учебно-методическое и информационное обеспечение дисциплины (модуля):	9
	а) основная литература;	
	б) дополнительная литература;	
	в) программное обеспечение;	
	г) интернет-ресурсы, базы данных, информационно-справочные и поисковые системы	
7.	Материально-техническое обеспечение дисциплины (модуля).	10
8.	Образовательные технологии	11
9.	Фонды оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации	11
9.1	Оценочные средства текущего контроля	
9.2	Оценочные средства для промежуточной аттестации	

1. Цели и задачи дисциплины (модуля): Целью освоения учебной дисциплины «Математическая биология, биоинформатика» является:

- Изучить базовые методы биоинформатики, применяемы при обработке и извлечения информации из биологических последовательностей (ДНК, РНК, белки), научиться применять полученные знания навыки в практической работе, провести углубленное изучение методов математического моделирования биологических процессов.

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 процедура, в том числе с использованием высокопроизводительных пероральных вычислительных технологий
- изучить технологии ДНК и белкового К-мерного анализа, применяемую для сравнения больших массивов биологических последовательностей без процедуры выравнивания
- ознакомиться с технологиями секвенирования нового поколения (NGS), применяемых 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, применимых для анализа и систематизации биологических последовательностей
- изучить методы анализа и аннотации полногеномных данных.
- Освоить методы математического моделирования биологических процессов и биосистем.

2. Требования к результатам освоения дисциплины (модуля):

В результате изучения дисциплины аспирант 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;
- классификации математических моделей биологических процессов и биологических систем.

Уметь:

- работать с различными форматами хранения данных, применимых 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;

	раздела	
1.	Регулярные выражения для обработки тестов анализ нуклеотидных аминокислотных последовательностей с помощью регулярных выражений.	В разделе рассматриваются созданные с обработкой строк с помощью системы регулярных выражений для анализа больших объемов 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, содержащий аннотацию полных геномов и протеомов.
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.
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.
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.
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 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.
6.	Анализ данных метагеномных исследований, проведенных с помощью технологий расшифровки нуклеотидных последовательностей	В разделе подробно с практическими примерами рассматриваются различные варианты анализа таксономического состава сообществ микро и макро организмов на основе расшифровки ампликонов стандартных генетических маркеров.

	тей нового поколения.	
7.	Сборка полных геномов, методы и подходы.	Изучаться различные алгоритмы для сборки полных геномов организмов из данных NGS секвенирования. Рассматриваться анализ метагеномных данных путем сборки фрагментов геном из ДНК и РНК смеси организмов. На практике проводится сборка полного митохондриального генома по результатам NGS секвенирования.
8.	Машинное обучение в биоинформатике.	В данном разделе рассматриваться различные варианты кластерного анализа, Байесовского наивного классификатора, нейронные сети для поиска закономерностей формирования первичной структуры нуклеотидных и аминокислотных последовательностей, применяемые при аннотации и анализе полных геномов.
9.	Аннотация полных геномов на основе методов машинного обучения.	Рассматриваться различные варианты аннотации полных геномов с использованием методов выравнивания биологических последовательностей и методов машинного обучения. На практике рассматривается аннотация полного генома природного бактериофага.
10.	Структурная биоинформатика, реконструкция трехмерных структур биополимеров	В рамках раздела рассматриваться различные биоинформационные методы, применяемые для реконструкции и предсказания трехмерных структур белков и РНК. На практике изучается работа в различных программных пакетах и онлайн серверах для реконструкции, анализа и сравнения трехмерных структур белков и РНК.
11.	Дифференциальные уравнения и системы уравнений, описание сложных биологических систем в виде дифференциальных уравнений.	Рассматриваются линейные и нелинейные дифференциальные уравнения, и системы дифференциальных уравнений. Осваиваются методы аналитического и численного интегрирования дифференциальных уравнений. Рассматриваться различные классы задач по сдавлению дифференциальных уравнений для описания биологических систем.
12.	Методы теории случайных процессов для описания биологических систем.	В данной теме рассматриваются основы понятия теории случайных процессов. Приводятся приемы задач, связанных с исследованием биологических систем и биологических процессов методами теории случайных процессов

4.2. Разделы и темы дисциплины (модуля) 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	2				7	9

		выражений.						
2.		Тема 2. Выравнивание аминокислотных и нуклеотидных последовательностей.					5	5
3.		Тема 3. Методы ДНК и аминокислотных К-мер для анализа биологических последовательностей.					5	5
4.		Тема 4. Технологии расшифровки – секвенирования нуклеотидных последовательностей нового поколения (NGS), первичный анализ данных секвенирования.					5	5
5.		Тема 5. Различные варианты использования технологии расшифровки нуклеотидных последовательностей нового поколения в биологических исследованиях					7	7
6.		Тема 6. Анализ данных метабаркодинговых исследований, проведенных с помощью технологий расшифровки нуклеотидных последовательностей нового поколения					5	5
7.		Тема 7. Сборка полных геномов, методы и подходы.					7	7
8.		Тема 8. Машинное обучение в биоинформатике.					5	12
9.		Тема 9. Аннотация полных геномов на основе методов машинного обучения.					5	5
10		Тема 10. Структурная биоинформатика, реконструкция трехмерных структур биополимеров					5	5

11		Тема 11. Дифференциальные уравнения и системы уравнений, описание сложных биологических систем в виде дифференциальных уравнений.					7	7
12		Тема 11. Методы теории случайных процессов для описания биологических систем.					7	7

5. Примерная тематика рефератов, докладов, проектов (при наличии):

Темы докладов для самостоятельной подготовки:

- 1) Применение методов Байесовской статистики при анализе биологических данных.
- 2) Методы теории случайных процессов в генетики популяций.
- 3) Машинное обучение при реконструкции трехмерных структур белков.
- 4) Матричные модели в популяционно динамике.

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

а) основная литература

1. Дурбин, Р., Эдди, Ш., Крог, 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 «Python»

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/tutorial.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.

7. Материально-техническое обеспечение дисциплины (модуля):

Материально-техническое обеспечение дисциплины «Математическая биология, биоинформатика» базируется на следующих ресурсах:

- **Аудитория для проведения лекционных и семинарских занятий (компьютерный класс)** 664003, Иркутская область, г. Иркутск, ул. Сухэ-Батора, д.5, аудитория 209, (площадь - 30,2 м², номер помещения – 27):

Посадочных мест – 20, Стол преподавателя -1, Стул преподавателя – 1, Стол – 20, Стул -20,

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, МФУ Kyocera Taskalfa – 1 шт.

Комната для хранения и ремонта оборудования (серверная).

Аудитория для проведения лекционных, семинарских, практических и лабораторных занятий (664003, Иркутская область, г.Иркутск, ул. Сухэ-Батора, д.5, аудитория 117 (площадь - 46,4 м², номер помещения – 59)):

Посадочных мест-28, Столы – 14, Стол преподавателя – 1, Стулья – 29, Мобильная доска меловая- маркерная – 1, Проектор Epson EB-X03 – 1, Экран ScreenMedia – 1, Шкаф – 1, Сейф – 1,

Помещение для хранения оборудования – 2 шт., Наглядные пособия, демонстрационный материал.

8. 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.

Стандартные методы обучения:

- занятия, предназначенные для освоения математических методов в биологии;
- Самостоятельная работа аспирантов;
- 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);
- информационно-коммуникационные образовательные технологии – лекция-визуализация, представление результатов деятельности (рефератов и отчетов) с использованием специализированных программных сред.

9. Фонды оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации

9.1 Оценочные средства текущего контроля:

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
23. Методы реконструкции трёхмерных структур белков
24. Методы реконструкции пространственных структур РНК
25. Классификация математических моделей применимых в анализе биологических процессов и биологических систем
26. Методы численного интегрирования обыкновенных дифференциальных уравнений для анализа, применяемые при анализе биологических процессов и биологических систем
27. Методы математического описания случайных процессов

9.2. Оценочные средства для промежуточной аттестации:

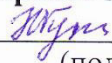
Примерный перечень вопросов к экзамену

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 с помощью NGS технологий.
13. Полногеномное секвенирование с помощью NGS технологий.
14. Метагеномное секвенирование для исследования сообществ микроорганизмов.
15. NGS технологий в популяционно-генетических исследованиях.
16. Сборка полных геномов, методы сборки полных геномов.
17. Сборщик полных геномов SPADES? Запуск сборщик 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.
24. Реконструкция трёхмерных структур белков и РНК
25. Описание биологических процессов и биологических систем с помощью дифференциальных уравнений.
26. Аналитические и численные методы исследования систем дифференциальных уравнений.
27. Уравнение Колмогорова для описания случайных процессов в биологии.

Критерии оценки:

- Оценка «отлично» ставится, если полностью раскрыта тематика вопросов, проанализировано их современное состояние; аспирант излагает логичный, последовательный и лаконичный ответ, соблюдая основные правила культуры речи, а также дает исчерпывающие, аргументированные, корректные ответы на дополнительные вопросы.
- Оценка «хорошо», если раскрыта суть вопросов, но при этом материал в недостаточной степени проанализирован автором. Ответы аспиранта не на все вопросы являются исчерпывающими и аргументированными.
- Оценка «удовлетворительно», если ответ на вопросы не полный, аспирант показывает поверхностные знания и не всегда дает правильные, исчерпывающие ответы на задаваемые дополнительные вопросы.
- Оценка «неудовлетворительно», если тема вопросов не раскрыта, на дополнительные вопросы аспирант не дает ответы или они не соответствуют заданным вопросам

Разработчик:

 _____ доцент Букин Ю.С.
(подпись)

Программа рассмотрена на заседании кафедры физико-химической биологии, биоинженерии и биоинформатики 06.06.2023 г. протокол № 20

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

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