



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



Рабочая программа дисциплины
БИОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

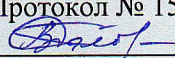
Научная специальность: 1.5.8. Математическая биология, биоинформатика

Форма обучения: очная

Согласовано с УМК биолого-почвенного факультета
протокол № 7 от 20.05. 2024 г

Председатель УМК  Матвеев А.Н.

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

17.04.2024 г. Протокол № 15
Зав. кафедрой  Саловарова В.П.

Иркутск 2024 г.

Содержание

1. Цели и задачи дисциплины (модуля)
2. Требования к результатам освоения дисциплины (модуля)
3. Объем дисциплины (модуля) и виды учебной работы
4. Содержание дисциплины (модуля)
 - 4.1 Содержание разделов и тем дисциплины (модуля)
 - 4.2 Разделы и темы дисциплин (модулей) и виды занятий
 - 4.3 Перечень семинарских, практических занятий и лабораторных работ.
5. Примерная тематика рефератов (при наличии)
6. Учебно-методическое и информационное обеспечение дисциплины (модуля):
 - 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).
8. Образовательные технологии
9. Фонды оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации
 - 9.1 Оценочные средства текущего контроля
 - 9.2 Оценочные средства для промежуточной аттестации

1. Цели и задачи дисциплины: приобретение аспирантами профессиональных умений и навыков в области использования компьютерных, вычислительных и биоинформационных технологий для обработки больших массивов геномных и протеомных данных с целью извлечения информации о закономерностях протекания биологических процессов на организменном, популяционном и экситонном уровнях.

2. Требования к результатам освоения дисциплины (модуля):

В результате изучения дисциплины аспирант должен:

Знать: методы получения и анализа геномных и протеомных данных с помощью специализированного программного обеспечения и скриптовых языков программирования, методы анализа больших массивов текстовой информации, кластеризации данных, методы машинного обучения, применимые в геномных и протеомных исследованиях.

Уметь: использовать персональные компьютеры, принципы распараллеливания вычисления и высокопроизводительные вычислительные системы для анализа полногеномных и протеомных данных в эволюционных, биомедицинских, экологических, эпидемиологических исследованиях, а также использовать международные базы данных для таксономической и функциональной аннотации геномных и метагеномных данных, предсказания функциональной активности белковых молекул.

Владеть: навыками построения вычислительных схем и комбинации программ для построения конвейеров для обработки больших массивов геномной и протеомной информации, навыками использования скриптовых языков программирования (R или Phyton) и специализированных пакетов для них при анализе геномных и протеомных данных.

3. Объем дисциплины (модуля) и виды учебной работы

Вид учебной работы	Всего академических часов	Курсы			
		2			
Аудиторные занятия (всего)	16	16			
В том числе:					
Лекции	8	8			
Практические занятия (ПЗ)	8	8			
Самостоятельная работа (всего)	18	18			
В том числе:					
Реферат (при наличии)					
Контактная работа					
Подготовка к зачету	18	18			
Промежуточная аттестация (всего)	2	2			
В том числе:					
Контактная работа во время промежуточной аттестации	2	2			
Форма промежуточной аттестации (зачет, экзамен)	зачет	зачет			
Общая трудоемкость	часы	36	36		
	зачетные единицы	1	1		

4. Содержание дисциплины (модуля)

4.1. Содержание разделов и тем дисциплины (модуля).

№	Наименование раздела	Содержание раздела дисциплины
1.	Механизмы хранения и реализации наследственной информации.	Рассматриваются основные положения молекулярной биологии, связанные с кодированием информации в нуклеотидных последовательностях геномов организмов, и реализации этой информации через процессы трансляции транскрипции. Рассматриваются основные механизмы регуляции экспрессии генетической информации на клеточном и организменном уровнях. С практической точки зрения рассматривается возможность изучения совокупности процессов реализации генетической информации с помощью анализа геномов, транскриптомов и протеомов.
2.	Механизмы трансформации геномов на организменном и популяционном уровнях, популяционная биология.	Рассматриваются различные типы мутационных процессов и их роль в трансформации геномов организмов на организменном и популяционном уровнях. Изучаются возможности сравнительной геномики для исследования эволюционных процессов на организменном, популяционном и видовом уровнях. С практической точки зрения рассматривается вопрос об исследовании механизмов формирования популяционного и видового разнообразия методами сравнительной геномики.
3.	Полимеразная цепная реакция, амплификация ДНК, расшифровка ДНК методом Сенгера.	Изучается процесс реализации полимеразной цепной реакции (ПЦР), изучаются различные варианты ПЦР. Рассматривается вопрос практического применения ПЦР в популяционной и эволюционной биологии, биомедицине. Изучается метод Сенгера для расшифровки нуклеотидных последовательностей. На практике изучается вопрос о первичной обработке и анализе результатов расшифровки ДНК методом Сенгера.
4.	Технологии расшифровки ДНК нового поколения (NGS) расшифровка и сборка полногеномных данных.	Изучаются различные технологии расшифровки геномной ДНК нового поколения (NGS – next generation sequencing) такие, как пироксигенирование, Illumina, Pacific Biosciences, нанопоровое секвенирование. Изучаются на практике форматы данных для хранения результатов NGS. Рассматривается программное обеспечение для контроля качества NGS данных. Изучается вопрос сборки полных геномов на основе данных NGS.
5.	Базы и банки данных геномной и протеомной информации.	Рассматривается практическая работа с базой данных NCBI (National Center for Biotechnology Information) – хранилище данных полногеномной информации. Рассматриваются практические принципы работы с базой данных NCBI RefSeq – хранилище данных полных геномов и полных протеомов организмов. Изучается программное обеспечение для работы с NCBI SRA (sequence read archive) архивами – хранилище первичной информации данных, полученных с помощью NGS технологий. Формируются навыки использования базы данных UNUPROT – база данных аминокислотных последовательностей белков с известной функциональной активностью
6.	Методы многомерной математической статистики, методы машинного обучения, применяемые для таксономической и функциональной аннотации полногеномных данных.	Рассматривается спектр методов математической статистики и машинного обучения таких, как кластерный анализ, наивный Байесовский классификатор, скрытые Марковские модели. Изучается программное обеспечение и интернет-ресурсы, позволяющие использовать эти методы для таксономической и функциональной классификации геномных данных. Даются практические примеры использования данных методов для аннотации геномов микроорганизмов и вирусов.
7.	Применение NGS технологий в	Рассматривается возможность применения геномных

	популяционных и экологических исследованиях, метагеномика.	данных в экологической генетике для контроля за состоянием популяций различных организмов. Изучаются алгоритмы получения и анализа метагеномных данных, позволяющих исследовать таксономическое и функциональное разнообразие различных сообществ микро- и микроорганизмов. Рассматривается вопрос практического анализа данных природной ДНК (eDNA - <i>Environmental DNA</i>).
8.	Применение NGS технологий в биомедицине, молекулярная эпидемиология.	Рассматриваются вопросы, связанные с анализом полного генома человека, для предсказания возможности возникновения наследственных заболеваний. Изучаются методы геномики для анализа патогенной микрофлоры человека. Рассматриваются методы эволюционной геномики для изучения истории возникновения, контроля и прогнозирования развития эпидемий патогенных организмов, вызывающих инфекционных заболеваний человека.

4.2. Разделы и темы дисциплины (модуля) и виды занятий

№ п/п	Наименование раздела	Наименование темы	Виды занятий в часах			
			Лекции	Практические занятия	Самостоятельная работа	Всего
1.	Механизмы хранения и реализации наследственной информации	Механизмы хранения и реализации наследственной информации	1	1	2	4
2.	Механизмы трансформации геномов на организменном и популяционном уровнях, популяционная биология	Механизмы трансформации геномов на организменном и популяционном уровнях, популяционная биология	1	1	2	4
3.	Полимеразная цепная реакция, амплификация ДНК, расшифровка ДНК методом Сенгера	Полимеразная цепная реакция, амплификация ДНК, расшифровка ДНК методом Сенгера	1	1	2	4
4.	Технологии расшифровки ДНК нового поколения (NGS) расшифровка и сборка полногеномных данных	Технологии расшифровки ДНК нового поколения (NGS) расшифровка и сборка полногеномных данных	1	1	2	4
5	Базы и банки данных геномной и протеомной информации	Базы и банки данных геномной и протеомной информации	1	1	2	4
6	Методы многомерной математической статистики, методы машинного обучения, применяемые для таксономической и функциональной аннотации полногеномных	Методы многомерной математической статистики, методы машинного обучения, применяемые для таксономической и функциональной аннотации полногеномных данных	1	1	2	4

	данных					
7	Применение NGS технологий в популяционных и экологических исследованиях, метагеномика	Применение NGS технологий в популяционных и экологических исследованиях, метагеномика	1	1	3	5
8	Применение NGS технологий в биомедицине, молекулярная эпидемиология	Применение NGS технологий в биомедицине, молекулярная эпидемиология	1	1	3	5

4.3. Перечень семинарских, практических занятий

№ п/п	№ раздела и темы дисциплины (модуля)	Наименование семинаров, практических и лабораторных работ	Трудоемкость (часы)	Оценочные средства
1	2	3	4	5
1.	Тема 1	Механизмы хранения и реализации наследственной информации	1	Опрос КСР
2.	Тема 2	Механизмы трансформации геномов на организменном и популяционном уровнях, популяционная биология	1	Опрос КСР
3.	Тема 3	Полимеразная цепная реакция, амплификация ДНК, расшифровка ДНК методом Сенгера	1	Опрос КСР
4	Тема 4	Технологии расшифровки ДНК нового поколения (NGS) расшифровка и сборка полногеномных данных	1	Опрос КСР
5	Тема 5	Базы и банки данных геномной и протеомной информации	1	Опрос КСР
6	Тема 6	Методы многомерной математической статистики, методы машинного обучения, применяемые для таксономической и функциональной аннотации полногеномных данных	1	Опрос КСР
7	Тема 7	Применение NGS технологий в популяционных и экологических исследованиях, метагеномика	1	Опрос КСР
8	Тема 8	Применение NGS технологий в биомедицине, молекулярная эпидемиология	1	Опрос КСР

5. Примерная тематика устных докладов:

1. Тема 1. Молекулярные механизмы транскрипции и трансляции.
2. Тема 1. Устройство и элементы геномов прокариотических и эукариотических организмов.
3. Тема 1. Доказательство наследованной роли ДНК.
4. Тема 1. Виды белков, классификация белковых молекул по структуре и функции.
5. Тема 2. Мутационный процесс, точечные и протяженные мутации.
6. Тема 2. Ретротранспозоны и ретровирусы и их роль в модификации генома.
7. Тема 2. Механизмы фиксации мутации на популяционном и видовом уровнях.
8. Тема 2. Соматические мутации и их роль в возникновении онкологических заболеваний.
9. Тема 4. Применение технологии Сенгера для расшифровки полных геномов.

10. Тема 4. Технология Illumina для расшифровки полных геномов..
11. Тема 4. Технология Pacific Biosciences для расшифровки полных геномов.
12. Тема 4. Нанопоровое секвенирование для расшифровки полных геномов.
13. Тема 7. Метабаркодинг и метабокодинговые исследования сообществ организмов с применением NGS технологий.

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

а) перечень литературы

1. Стефанов, Василий Евгеньевич. Биоинформатика [Электронный ресурс] : учебник для вузов / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. - Электрон. текстовые дан. - Москва : Юрайт, 2022. - 252 с. - (Высшее образование). - Режим доступа: <https://urait.ru/bcode/489775>, <https://urait.ru/book/cover/D69A8182-EF23-44FC-BE4B-CADE8229EA09>. - ЭБС "Юрайт". - Неогранич. доступ. - ISBN 978-5-534-00860-9 : 939.00 р. URL: <https://urait.ru/bcode/489775> (дата обращения)
2. Леск А. (2009). Введение в биоинформатику. М.: БИНОМ. Лаборатория знаний, 2009. — 318 с. - ISBN 78-5-94774-501-6+

б) периодические издания

1. <http://www.matbio.org/about.php> - сайт журнала «Математическая биология и биоинформатика». Содержит большое количество статей в pdf – формате.
2. <https://journal.r-project.org/> - сайт журнала по статистическим методам на R, «*The R Journal*»
3. <https://bmcbioinformatics.biomedcentral.com/> - сайт журнала «*BMC Bioinformatics*», публикующим научные работы в области биоинформационных технологий

в) 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. <https://www.elibrary.ru> – электронная библиотека научных статей, монографии и материалов конференций, выпущенных Российскими учеными.
5. <https://pubmed.ncbi.nlm.nih.gov/> - международная база данных научных статей и монографий, посвященная различным вопросам биологии.
6. <https://apps.webofknowledge.com> – международная база данных, индексирующая научные публикации в высокорейтинговых изданиях
7. <https://www.r-project.org/> - сайт проекта по R статистике
8. <http://qsar4u.com/pages/rtutorial.php> - Краткий курс "Введение в R и моделирование с R" (Pavel Polishchuk)
9. <http://www.ncbi.nlm.nih.gov/> - сайт международной базы данных NCBI (National Center Biotech Information)
10. <https://www.ncbi.nlm.nih.gov/home/tutorials/> - полное руководство по использованию базы данных NCBI.
11. <https://www.uniprot.org/> - международная база данных протеомной информации «UNIPROT»
12. <https://www.uniprot.org/docs/userman.htm> - полное руководство по использованию базы данных «UNIPROT»
13. <https://bioedit.software.informer.com/7.2/> - сайт и руководство к программе «BIOEDI» для анализа биологических последовательностей
14. <http://doua.prabi.fr/software/seaview> - сайт и руководство к программе «Seaview» для анализа биологических последовательностей

15. <http://ape-package.ird.fr/> - сайт по описанию пакета APE для эволюционного анализа нуклеотидных последовательностей.
16. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> - сайт программы «FSTQC» для контроля качества данных полногеномного NGS секвенирования
17. <http://www.usadellab.org/cms/?page=trimmomatic> – сайт программы «Trimmomatic» для фильтрации и первичной подготовки NGS данных
18. <https://cab.spbu.ru/software/spades/> - официальный сайт сборщика полных геномов «SPADES»
19. <https://narrative.kbase.us/> - онлайн сервер для запуска различных программных средств для биоинформационного анализа данных (в том числе «SPADES»).
20. <https://kaiju.binf.ku.dk/> - онлайн инструмент для таксономической классификации коротких NGS прочтений в метагеномике.
21. <https://www.beast2.org/> - сайт программы «BEAST 2» по реконструкции эволюционной и популяционной истории организмов на основе геномных данных

7. Материально-техническое обеспечение дисциплины

7.1. Учебно-лабораторное оборудование:

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)экран ScreenVtdiaEcot

7.2. Программное обеспечение:

DreamSpark Premium Electronic Software Delivery (3 years) Renewal (Windows 10 Education 32/64-bit (Russian) - Microsoft Imagine, Windows 7 Professional with Service Pack 1 32/64-bit (English) - Microsoft Imagine, Windows Server 2008 Enterprise and Standard without Hyper-V with SP2 32/64-bit (English) - Microsoft Imagine, Access 2016 32/64-bit (Russian) - Microsoft Imagine, Access 2010 32/64-bit (Russian) - Microsoft Imagine). Договор №03-016-14 от 30.10.2014г.

Kaspersky Endpoint Security для бизнеса - Стандартный Russian Edition. 250-499. 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.

7.3. Технические и электронные средства:

Презентации по всем темам курса.

8. Образовательные технологии:

При реализации различных видов учебной работы дисциплины используются как стандартные методы обучения, так и интерактивные формы проведения занятий.

Стандартные методы обучения:

1. Информационная лекция
2. Практические занятия, предназначенные для освоения студентами базовых методов анализа геномных и протеомных данных.
3. Самостоятельная работа студентов.
4. Консультации преподавателя.
5. Подготовка ответов на контрольные вопросы.
6. Обучения с применением интерактивных форм образовательных технологий.
7. Кейс-метод – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия научной деятельности (разбор конкретных ситуаций).
8. Работа с молекулярно-биологическими базами данных (genbank, UNIPROT.), обслуживающими их приложениями (BLAST и др) и компьютерными программами молекулярного моделирования.
9. Информационно-коммуникационные образовательные технологии – лекция-визуализация, представление результатов деятельности (рефератов и отчетов) с использованием специализированных программных сред.
10. Интернет-технология – задействование образовательного портала ИГУ - educa.isu.ru для предоставления письменных отчетов по домашним работам.

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

9.1 Оценочные средства текущего контроля:

В рамках дисциплины «Биоинформационные технологии» используются следующие формы текущего контроля:

- устный опрос;
- устный доклад по теме;
- письменная работа по решению домашних заданий;

Фонд оценочных средств включает:

- перечень тем докладов по темам дисциплины (см.5);
- вопросы для самостоятельного изучения (срс);
- задачи для самостоятельного домашнего решения;
- вопросы для зачета.

Перечень домашних задания для самостоятельного выполнения

1. Тема 3. Анализ первичных данных секвенирования по Сегеру с помощью программы «Bioedit».
2. Тема 3. Кластеризация последовательностей ДНК дистанционными методами в программе «Seaview».
3. Тема 5. Поиск в базе данных NCBI SRA первичных данных по расшифровке вирусов породе *Sarbecovirus*, контроль качества секвенирования в программах «FSTQC» и «Trimmomatic».
4. Тема 5. Сборка геномных данных в программе «Spades».
5. Тема 6. Аннотация полных геномов вирусов подрода с помощью онлайн средств базы данных <http://www.ncbi.nlm.nih.gov/>.
6. Тема 7. Поиск в базе данных NCBI SRA первичных данных по метагеномным исследованиям вирусных сообществ из оз. Байкал, контроль качества секвенирования в программах «FSTQC» и «Trimmomatic».
7. Тема 7. Таксономическая аннотация коротких прочтений NGS метагеномных данных вирусных сообществ из оз. Байкал с помощью онлайн сервиса <https://kaiju.binf.ku.dk/>.

8. Тема 8. Реконструкции эволюционной истории SARS-CoV-2 на территории США и Великобритании на основе анализа полных геномов в программе «BEAST 2» (<https://www.beast2.org/>).
9. Тема 8. Реконструкции динамики эпидемического процесса вируса SARS-CoV-2 на территории США и Великобритании на основе анализа полных геномов, сравнение реконструированной истории динамики пандемии с реальными данными об эпидемическом процессе.

Примерный список вопросов к зачету

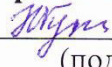
1. ДНК, РНК их биологическая роль, цикл реализации наследственной информации.
2. Опыты по определению биологической роли ДНК.
3. Доказательство полуконсервативной дупликации ДНК.
4. Регуляция экспрессии генетической информации.
5. Мутационные процесс, точечные и генные мутации.
6. Механизмы фиксации мутационных изменений в популяциях
7. ПЦР полимеразная цепная реакция
8. Расшифровка ДНК по методу Сенгера, секвенирующий ПЦР.
9. Первичная обработка результата секвенирования по Сегеру
10. Технология расшифровки полного генома с помощью метода Сенгера, ПЦР со случайными праймерами, технология клонирования генов.
11. Пиросеквенирование (roche 454) – технология секвенирования полных геномов второго поколения.
12. Технология секвенирования полных геномов второго поколения Illumina
13. Технология секвенирования полных геномов третьего поколения Pacific Biosciences (PacBio).
14. Нанопоровое секвенирование полных геномов (MinION) – технология секвенирования третьего поколения.
15. Сборка полных геномов графы Дебрейна.
16. Поиск функциональных элементов в полных геномах, скрытые Марковские модели
17. Обработка больших массивов текстовой информации, k-мерный анализ.
18. Эволюция ДНК, модели накопления нуклеотидных замен ДНК
19. Молекулярные часы, реконструкция истории эволюционных событий с использованием молекулярных часов.

Критерии оценки:

Оценка «Зачтено» выставляется аспиранту, если на вопросы даны правильные и полные ответы, раскрывающие суть рассматриваемой проблемы, ее основных факторов, теоретические положения и пути решения; допускается: ответ правильный, но аргументации недостаточно или даны недостаточно точные ответы.

Оценка «Не зачтено» выставляется аспиранту, если ответ неправильный или не дан вовсе.

Разработчик:

 _____ доцент Букин Ю.С.
(подпись)

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.