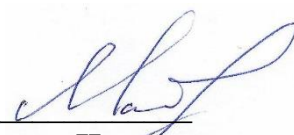




Министерство науки и высшего образования Российской Федерации
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»
Кафедра физиологии, молекулярной биологии и психофизиологии




Директор Института
Биологических наук
А. Н. Матвеев
«02» апреля 2026 г.

Рабочая программа дисциплины
Б1.В.ДВ.1.4 Элективный модуль «Физиология и биохимия»

Наименование дисциплины: Б1.В.ДВ.1.4.5 «**ОСНОВЫ ПРОГРАММИРОВАНИЯ В БИОЛОГИИ**»

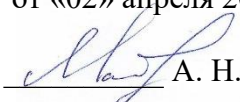
Направление подготовки: 06.03.01 «Биология»

Направленность (профиль) подготовки: «Биология»

Квалификация выпускника: Бакалавр

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

Согласовано с УМК Института
Биологических наук ИГУ
Протокол № 1 от «02» апреля 2026 г.

Председатель  А. Н. Матвеев

Рекомендовано кафедрой физиологии,
молекулярной биологии и психофизиологии
Протокол № 1 от «01» апреля 2026 г.

Зав. кафедрой  И.Н. Гутник

Иркутск 2026 г.

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

I. Цель и задачи дисциплины:

Цель: углубление знаний о методах работы с биологическими данными.

Задачи:

- получение знаний о языке программирования R;
- изучение баз данных и способах обработки биологических последовательностей;
- углубление знаний о биологических данных.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина Б1.В.ДВ.1.4.5 «Основы программирования в биологии» относится к части, формируемой участниками образовательных отношений.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами: Математика, Химия, Биохимия, Информатика.

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: Структура функции и синтез белков, Научно-исследовательская работа, Нейробиология, Основные метаболические пути и их регуляция, Большой практикум по физиологии и биохимии.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование компетенции:

ПК-2: Способен применять на практике основные методы и средства исследований биологических объектов, выбирать методы исследования в соответствии с поставленными задачами

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

Компетенция	Индикаторы компетенций	Результаты обучения
--------------------	-------------------------------	----------------------------

ПК-2 Способен применять на практике основные методы и средства исследований биологических объектов, выбирать методы исследования в соответствии с поставленными задачами	ИДК ПК 2.2 Проводит анализ и теоретическое обобщение научных данных, применяет на практике методы обработки экспериментальных данных, включая оценку достоверности результатов и биоинформатические алгоритмы; знает нормативные документы по организации и технике безопасности работ и принципы составления отчетности	Знать: Основные типы и структуры данных в языке R, синтаксис языка, базовые функции для работы с векторами, матрицами, списками и датафреймами. Методы описательной статистики, критерии проверки статистических гипотез (параметрические и непараметрические). Современные биоинформатические алгоритмы для анализа последовательностей, филогенетики, анализа сетей и обогатительного анализа (GSEA, GO). Принципы визуализации биологических данных и представления результатов в научной отчетности. Уметь: • Выбирать адекватные методы статистической обработки и биоинформатического анализа в зависимости от поставленной биологической задачи и типа данных. • Писать скрипты на языке R для автоматизации обработки экспериментальных данных, используя пакеты tidyverse, Biostrings, igraph, clusterProfiler. • Оценивать достоверность полученных результатов (интерпретировать p-value, доверительные интервалы, бутстреп-поддержки). • Визуализировать результаты анализа с помощью современных инструментов (ggplot2, ggtree, ComplexHeatmap), делая графики готовыми к включению в научные отчеты и публикации.
--	--	--

		Владеть: Навыками работы в среде RStudio для комплексного анализа биологических данных. Методами теоретического обобщения результатов анализа, формулировки выводов и подготовки отчетности по проделанной работе. Терминологией в области биоинформатики и статистического анализа данных. Подходами к воспроизводимым исследованиям (написание документированного кода, создание отчетов).
--	--	--

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 2 зачетных единицы, 72 часа.

Из них реализуется с использованием электронного обучения и дистанционных образовательных технологий 16 часов.

Форма промежуточной аттестации: зачёт.

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

№ п/п	Раздел дисциплины/тема	Семестр	Всего часов	Из них практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся , практическую подготовку и трудоемкость (в часах)			Самостоятельн ая работа	Форма текущего контроля успеваемости/ Форма промежуточной аттестации (по семестрам)
					Контактная работа преподавателя с обучающимися				
					Лекция	Семинар/ Практическое, лабораторное занятие/	Консультация		
1	2	3	4	5	6	7	8	9	10
1	Раздел 1. Синтаксис языка R. Тема 1.1. Введение в R и основы синтаксиса.	6	3		1	1		1	КСР
2	Тема 1.2. Работа с данными в R: векторы, индексация, функции.	6	4		1	1		2	КСР
3	Тема 1.3. Базовые структуры данных: матрицы, списки и датафреймы.	6	4		1	1		2	КСР
4	Тема 1.4. Основы визуализации в R. Базовые графики и ggplot2.	6	4		1	1	-	2	КСР
5	Тема 1.5. Обработка данных с tidyverse (dplyr, tidyr)	6	4		1	1	-	2	КСР
6	Раздел 2. Анализ данных. Тема 2.1. Статистический анализ в R: описательная статистика и корреляции.	6	4		1	1		2	КСР

7	Тема 2.2. Машинное обучение в R	6	4		1	1		2	KCP
8	Тема 2.3. Сравнение групп: параметрические и непараметрические тесты.	6	3,5		1	0,5		2	KCP
9	Тема 2.4. Реальный биоинформатический кейс: анализ данных ПЦР в реальном времени (qPCR).	6	3,5		1	0,5		2	KCP
10	Тема 2.5. Линейная регрессия и анализ выбросов.	6	4		1	1		2	KCP
11	Тема 2.6. Анализ больших данных - 1: Уменьшение размерности.	6	4		1	1		2	KCP
12	Тема 2.7. Анализ больших данных - 2: Кластерный анализ и введение в анализ сетей.	6	4		1	1		2	KCP
13	Раздел 3. Биоинформатика в R. Тема 3.1. Анализ сетей в биологии. GRN и кластеризация сетей.	6	4		1	1		2	KCP
14	Тема 3.2. Обоганительный анализ (Gene Set Enrichment Analysis, GSEA).	6	5		1	1		2	KCP
15	Тема 3.3. Работа с биологическими последовательностями в R.	6	5		1	1		2	KCP
16	Тема 3.4. Множественное выравнивание и филогенетический анализ.	6	5		1	1		2	KCP
17	Тема 3.5. Комплексный анализ генома	6	3		0	1		2	KCP

4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
6	Раздел 1. Синтаксис языка R. Тема 1.1. Введение в R и основы синтаксиса.	Установка R и RStudio. Выполнение базовых операций в консоли. Написание скрипта с арифметическими операциями и созданием переменных.	1 неделя	1	КСР	1. Леск А. Введение в биоинформатику, гл. 1. 2. Игнасимуту С. Основы биоинформатики, гл. 2. 3. Документация R (https://www.r-project.org/).
6	Тема 1.2. Работа с данными в R: векторы, индексация, функции.	Решение задач на создание векторов, индексацию и написание пользовательских функций (задачи t1-t3 из лекции).	2 неделя	1	КСР	1. Леск А., гл. 2. 2. Игнасимуту С., гл. 3. 3. R Cheatsheet (базовый R).
6	Тема 1.3. Базовые структуры данных: матрицы, списки и датафреймы.	Создание матрицы 10×10, выполнение операций по замене элементов. Создание и модификация датафрейма на примере iris.	3 неделя	2	КСР	1. Леск А., гл. 2. 2. Документация по data.frame.
6	Тема 1.4. Основы визуализации в R. Базовые графики и ggplot2.	Построение графиков plot() и ggplot2 на данных mtcars. Настройка тем и подписей.	4 неделя	2	КСР	1. Уилер Б. Статистический анализ и визуализация данных с помощью R, гл. 3. 2. Документация ggplot2 (https://ggplot2.tidyverse.org/).
6	Тема 1.5. Обработка данных с tidyverse (dplyr, tidyr).	Трансформация "сырого" датафрейма в "чистый" формат с использованием pivot_longer/pivot_wider. Фильтрация и группировка данных с dplyr.	5 неделя	2	КСР	1. Уилер Б., гл. 4. 2. Cheatsheet по dplyr и tidyr.

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
6	Раздел 2. Анализ данных. Тема 2.1. Статистический анализ в R: описательная статистика и корреляции.	Проведение теста Шапиро-Уилка на разных выборках. Расчет корреляций Пирсона и Спирмена для данных iris. Построение корреляционной матрицы.	6 неделя	2	КСР	1. Леск А., гл. 5. 2. Холмс С. Статистика для наук о данных, гл. 2.
6	Тема 2.2. Машинное обучение в R	Знакомство с пакетом caret или tidymodels. Проведение простой классификации (например, kNN) на данных iris.	7 неделя	2	КСР	1. Холмс С., гл. 7. 2. Документация caret.
6	Тема 2.3. Сравнение групп: параметрические и непараметрические тесты.	Проведение t-теста и ANOVA для данных penguins. Выбор критерия в зависимости от нормальности распределения.	8 неделя	2	КСР	1. Холмс С., гл. 4. 2. Лекционные материалы по статистике.
6	Тема 2.4. Реальный биоинформатический кейс: анализ данных ПЦР в реальном времени (qPCR).	Самостоятельное воспроизведение кода для расчета ddCt. Построение графиков (барплоты, тепловые карты) для визуализации экспрессии генов.	9 неделя	2	КСР	1. Материалы лекции (PCR equation). 2. Научные статьи по анализу qPCR.
6	Тема 2.5. Линейная регрессия и анализ выбросов.	Построение модели линейной регрессии для зависимости Petal.Width ~ Petal.Length. Проверка модели, интерпретация R ² , AIC. Детекция выбросов с rosnerTest.	10 неделя	2	КСР	1. Холмс С., гл. 5. 2. Документация EnvStats.
6	Тема 2.6. Анализ больших данных - 1: Уменьшение размерности.	Уменьшение размерности разных наборов больших данных линейными и нелинейными методами, подбор наилучшего метода	11 неделя	2	КСР	Документация FactoMiner, FactoInvestivage
6	Тема 2.7. Анализ больших данных - 2: Кластерный анализ и введение в анализ сетей.	Выбор и применение оптимального метода кластеризации для данных разной размерности	12 неделя	2	КСР	Документация FactoMiner, FactoInvestivage, cluster, igraph
6	Раздел 3. Биоинформатика в R. Тема 3.1. Анализ сетей в биологии. GRN и кластеризация сетей.	Построение генно-регуляторной сети на основе корреляций. Выделение кластеров подходящими алгоритмами. Визуализация с ggraph.	13 неделя	2	КСР	1. Лекционные материалы по сетям. 2. Уроки по ggraph.

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
6	Тема 3.2. Обогащительный анализ (Gene Set Enrichment Analysis, GSEA).	Запуск GSEA для набора дифференциально экспрессирующихся генов. Интерпретация графиков (dotplot, cnetplot).	14 неделя	2	КСР	1. Документация clusterProfiler. 2. База данных GO (http://geneontology.org/).
6	Тема 3.3. Работа с биологическими последовательностями в R.	Чтение FASTA-файла с помощью Biostrings. Расчет аминокислотного состава, длины последовательностей. Предсказание вторичной структуры.	15 неделя	2	КСР	1. Документация Biostrings. 2. Леск А., гл. 3.
6	Тема 3.4. Множественное выравнивание и филогенетический анализ.	Выравнивание последовательностей (MUSCLE/ClustalOmega/MAFFT). Построение дерева (NJ, ML). Визуализация и аннотация дерева с ggtree.	15 неделя	2	КСР	1. Леск А., гл. 6. 2. Документация ggtree.
6	Тема 3.5. Комплексный анализ генома	Полный анализ небольшого генома	15 неделя	2	КСР	
Общий объем самостоятельной работы по дисциплине (час) – 32						
Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий (час) – 12						

4.3 Содержание учебного материала

Раздел 1. Синтаксис языка R.

Тема 1.1. Введение в R и основы синтаксиса.

Язык R как среда для биологических исследований и статистической обработки данных. Установка и настройка R и RStudio. Интерфейс RStudio: консоль, скрипты, рабочая область, история команд. Базовые математические операции: +, -, *, /, ^. Комментирование кода. Понятие переменной. Оператор присваивания (<-, =). Типы данных: числовой (numeric), логический (logical), символьный (character). Функции: понятие аргументов, регистрозависимость, вызов встроенных функций (sqrt(), log(), paste()). Использование справки (?функция).

Тема 1.2. Работа с данными в R: векторы, индексация, функции.

Вектор как базовая структура данных. Создание векторов с помощью функции c(). Правило приведения типов (гомогенность векторов). Генерация последовательностей: оператор :, функции seq(), rep(). Именованные векторы (setNames()). Индексация векторов: обращение по номеру, по имени, по логическому условию. Исключение элементов с помощью отрицательного индекса. Специальные значения: NA, NULL. Проверка пропущенных значений (is.na()), их удаление. Векторные операции и правило переиспользования (recycling rule). Написание пользовательских функций.

Тема 1.3. Базовые структуры данных: матрицы, списки и датафреймы.

Матрицы: создание (matrix()), заполнение по строкам и столбцам. Индексация матриц: [строка, столбец]. Извлечение строк, столбцов и подматриц. Модификация элементов матрицы. Списки: создание (list()), гетерогенность. Именованные списки. Индексация списков: [] (возвращает список), [[]] (возвращает элемент), \$ (доступ по имени). Датафреймы (data.frame) как основная структура для хранения табличных данных. Создание датафреймов. Просмотр структуры (str()), основных статистик (summary()). Добавление новых столбцов. Индексация датафреймов: [строка, столбец], \$. Фильтрация строк по условию.

Тема 1.4. Основы визуализации в R. Базовые графики и ggplot2.

Философия грамматики графики (Grammar of Graphics). Базовые графические функции: plot(), hist(), boxplot(). Пакет ggplot2 как стандарт визуализации в R. Структура графика ggplot2: данные (data), эстетики (aes), геометрии (geom_). Построение scatter plot (geom_point), boxplot (geom_boxplot), гистограмм (geom_histogram). Добавление слоев: сглаживание (geom_smooth), точки (geom_jitter). Разделение на фасеты (facet_wrap, facet_grid). Настройка тем и элементов графика: theme(), theme_bw(), theme_light(). Работа с текстом: xlab(), ylab(), ggtitle(). Сохранение графиков (ggsave, pdf).

Тема 1.5. Обработка данных с tidyverse (dplyr, tidyr).

Концепция "чистых данных" (tidy data). Пакет dplyr для манипуляции данными. Оператор пайпа (%>%). Основные функции: filter() (фильтрация строк), select() (выбор столбцов), mutate() (создание новых столбцов), group_by() (группировка), summarize() (агрегация). Функция pull() для извлечения вектора. Пакет tidyr для преобразования форматов данных: pivot_longer() (из широкого в длинный), pivot_wider() (из длинного в широкий).

Раздел 2. Анализ данных.

Тема 2.1. Статистический анализ в R: описательная статистика и корреляции.

Описательные статистики: среднее (`mean`), медиана (`median`), стандартное отклонение (`sd`), квартили (`quantile`, `summary`). Проверка нормальности распределения: графические методы (`plot(density())`), тест Шапиро-Уилка (`shapiro.test`). Корреляционный анализ. Коэффициенты корреляции Пирсона, Спирмена и Кендалла. Функция `cor.test()`. Множественные корреляции: пакет `rstatix` (`cor_test`, `cor_mat`). Визуализация корреляционных матриц: `cor_plot()`, `chart.Correlation()` (пакет `PerformanceAnalytics`), `ggpairs()` (пакет `GGally`).

Тема 2.2. Машинное обучение в R.

Введение в машинное обучение для биологов. Задачи классификации и кластеризации. Пакет `caret` (Classification And REgression Training) как унифицированный интерфейс для моделей. Предобработка данных: центрирование, масштабирование. Разделение выборки на обучающую и тестовую. Модель k-ближайших соседей (kNN). Обучение модели, предсказание, оценка точности (confusion matrix). Понятие переобучения (overfitting).

Тема 2.3. Сравнение групп: параметрические и непараметрические тесты.

Выбор критерия в зависимости от нормальности распределения и количества групп. Параметрические критерии: t-тест (для двух групп), ANOVA (для трех и более групп). Непараметрические аналоги: критерий Манна-Уитни (Wilcoxon rank-sum test) для двух групп, критерий Краскела-Уоллиса (Kruskal-Wallis) для трех и более групп. Пост-хок тесты (post-hoc): поправка на множественные сравнения (`pairwise.t.test`, `pairwise.wilcox.test`). Визуализация результатов сравнения с помощью `ggboxplot`, `ggviolin` и `stat_compare_means()` из пакета `ggpubr`.

Тема 2.4. Реальный биоинформатический кейс: анализ данных ПЦР в реальном времени (qPCR).

Метод количественной ПЦР. Понятие порогового цикла (C_t). Метод относительной количественной оценки $2^{(-\Delta\Delta C_t)}$ ($\Delta\Delta C_t$). Нормировка на референсный ген (например, ACT-1). Загрузка и предобработка данных. Расчет ΔC_t , $\Delta\Delta C_t$ и уровня экспрессии (Fold Change). Логарифмическое преобразование ($\log_2$ Fold Change). Визуализация результатов: столбчатые диаграммы с error bars, тепловые карты (heatmaps) с `geom_tile`. Статистическая оценка значимости изменений экспрессии.

Тема 2.5. Линейная регрессия и анализ выбросов.

Понятие линейной регрессии. Простая и множественная линейная регрессия. Функция `lm()`. Интерпретация результатов: коэффициенты модели, p-value, R^2 (коэффициент детерминации), скорректированный R^2 . Информационные критерии AIC и BIC для сравнения моделей. Визуализация регрессионной модели: добавление линии регрессии и доверительного интервала. Детекция выбросов (outliers). Понятие выброса. Графический метод (boxplot). Статистические тесты для выявления выбросов: критерий Граббса (`grubbs.test`), критерий Рознера (`rosnerTest`) для множественных выбросов. Условия применимости тестов (нормальность распределения).

Тема 2.6. Анализ больших данных - 1: Уменьшение размерности.

Проблема "проклятия размерности" в биологических данных. Метод главных компонент (Principal Component Analysis, PCA). Геометрический смысл PCA: поиск направлений максимальной дисперсии. Вычисление PCA в R (`prcomp`). Анализ нагрузок (loadings) и собственных векторов. Визуализация результатов PCA: scree plot, biplot (`ggbiplot`). Интерпретация кластеров образцов на графике счетов (scores plot). Нелинейные методы снижения размерности: t-SNE (t-distributed Stochastic Neighbor Embedding). Настройка параметра perplexity. Визуализация t-SNE. Сравнение PCA и t-SNE.

Тема 2.7. Анализ больших данных - 2: Кластерный анализ и введение в анализ сетей. Задача кластеризации. Иерархическая кластеризация: агломеративные методы. Функция `hclust()`, метод Варда. Построение и визуализация дендрограммы. Метод k-средних (k-means). Выбор оптимального числа кластеров (метод локтя, силуэтный анализ). Индекс силуэта. Кластеризация на основе плотности: DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Подбор гиперпараметра ϵ . Введение в теорию графов. Граф, вершины (узлы), ребра. Взвешенные и невзвешенные графы. Пакет `igraph` для создания и анализа графов. Построение графа на основе корреляционной матрицы.

Раздел 3. Биоинформатика в R.

Тема 3.1. Анализ сетей в биологии. GRN и кластеризация сетей.

Генные регуляторные сети (Gene Regulatory Networks, GRN). Построение GRN на основе корреляций экспрессии генов (методы "по доверию", WGCNA-like подходы). Фильтрация ребер по порогу корреляции. Показатели центральности в графах: степень вершины (degree), посредничество (betweenness), близость (closeness). Выделение сообществ (кластеров) в сети. Алгоритмы кластеризации: Лувена (Louvain), fast-greedy, edge betweenness. Визуализация сетей с помощью `ggraph` (алгоритмы компоновки: "fr", "kk", "backbone"). Добавление аннотаций к узлам (цвет, размер, подписи).

Тема 3.2. Обоганительный анализ (Gene Set Enrichment Analysis, GSEA).

Понятие обогащения (enrichment). Базы данных функциональной аннотации генов: GO (Gene Ontology), KEGG. Анализ избыточной представленности (Over-Representation Analysis, ORA) с помощью `enrichGO`. Понятие "интересного" набора генов (например, дифференциально экспрессирующихся). Статистика гипергеометрического теста. Визуализация результатов ORA: `dotplot`, `cnetplot`. Анализ обогащения наборов генов (Gene Set Enrichment Analysis, GSEA) для ранжированных списков генов. Функция `gseGO`. Понятие нормализованного показателя обогащения (NES). Визуализация GSEA: `ridgeplot`, `gseaplot`.

Тема 3.3. Работа с биологическими последовательностями в R.

Пакет `Biostrings` для работы с ДНК и аминокислотными последовательностями. Чтение и запись FASTA-файлов (`readDNAStringSet`, `readAAStringSet`). Основные свойства последовательностей: длина, имена. Извлечение подпоследовательностей (`subseq`). Обратные и комплементарные последовательности (`reverseComplement`). Анализ состава последовательностей: частота нуклеотидов/аминокислот (`letterFrequency`, `alphabetFrequency`, `oligonucleotideFrequency`). Предсказание вторичной структуры белка (H, E, C) с помощью `PredictNEC` (пакет `DECIPHER`). Визуализация структуры белка (`plot_NEC`). Расчет физико-химических свойств белков (изоэлектрическая точка, заряд, гидрофобность) с пакетом `ampir`.

Тема 3.4. Множественное выравнивание и филогенетический анализ.

Необходимость множественного выравнивания последовательностей. Алгоритмы выравнивания: `ClustalOmega`, `MUSCLE`. Запуск выравнивания из R (`msaClustalOmega`, `msaMuscle`). Преобразование выравнивания в объект `AAStringSet`. Процент идентичности (PID). Построение дистанционной матрицы (`DistanceMatrix` из пакета `DECIPHER`). Филогенетические методы: Neighbor-Joining (NJ), Maximum Likelihood (ML). Построение NJ-деревя (`nj` из пакета `ape`). Бутстреп-поддержка узлов (`bootstrap.phyDat`). Визуализация и аннотация филогенетических деревьев с помощью пакета `ggtree`. Добавление bootstrap значений, окраска таксонов, работа с лейблами.

Тема 3.5. Комплексный анализ генома.

Анализ нуклеотидного состава генома: GC-состав. Скользящее окно (letterFrequencyInSlidingView). AT- и GC-скью (skew): расчет и биологический смысл (репликация, транскрипция). Визуализация состава генома (линейные графики, треки). Чтение аннотаций генома в формате GFF/GTF (read_gff3 из пакета plyranges). Визуализация аннотаций с ggbio. Оценка сложности последовательности (sequence complexity) методами WoottonFederhen, Trifonov, DUST. Кольцевые диаграммы генома (circos plots) в ggbio. Элементы non-B-ДНК (G-квадруплексы, Z-ДНК) и их предсказание. Параллельные вычисления в R: пакеты parallel, parabar. Запуск внешних программ (BLAST, MAFFT) из R с помощью system().

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

Исправить число часов в списке (сейчас 17, должно быть 16)

№ п/п	№ раздела и темы	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции (индикаторы) *
			Всего часов	Из них практич. подготовка		
1	2	3	4	5	6	7
1	Тема 1.1	Введение в R и основы синтаксиса. Знакомство с интерфейсом RStudio. Выполнение базовых арифметических операций. Создание переменных различных типов.	1		КСР	ПК-2 ИДК ПК 2.2
2	Тема 1.2	Работа с векторами в R. Создание векторов с помощью c(), seq(), rep(). Индексация векторов. Решение задач на поиск элементов по условию. Написание простейших пользовательских функций.	1		КСР	ПК-2 ИДК ПК 2.2
3	Тема 1.3	Матрицы, списки и датафреймы. Создание матриц и выполнение операций над ними. Работа со списками. Создание и модификация датафреймов. Фильтрация строк и выбор столбцов.	1		КСР	ПК-2 ИДК ПК 2.2

4	Тема 1.4	Визуализация данных в ggplot2. Построение scatter plot, boxplot, гистограмм. Добавление слоев и настройка тем. Работа с фасетами. Сохранение графиков.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
5	Тема 1.5	Обработка данных с dplyr и tidyr. Применение функций filter, select, mutate, group_by, summarize. Преобразование данных из широкого формата в длинный и обратно.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
6	Тема 2.1	Описательная статистика и корреляции. Расчет описательных статистик. Проверка нормальности распределения. Проведение корреляционного анализа (Пирсон, Спирмен). Построение корреляционной матрицы.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
7	Тема 2.2	Введение в машинное обучение. Знакомство с пакетом caret. Проведение классификации методом k-ближайших соседей на данных iris. Оценка точности модели.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
8	Тема 2.3	Сравнение групп. Выбор критерия в зависимости от распределения данных. Проведение t-теста, ANOVA и их непараметрических аналогов. Визуализация результатов сравнения.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
9	Тема 2.4	Анализ данных ПЦР (qPCR). Расчет ddCt метода. Построение столбчатых диаграмм и	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>

		тепловых карт для визуализации экспрессии генов.				
10	Тема 2.5	Линейная регрессия и анализ выбросов. Построение модели линейной регрессии. Интерпретация коэффициентов и R^2 . Детекция выбросов с помощью критерия Граббса и Рознера.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 2.6	Уменьшение размерности (PCA и t-SNE). Проведение PCA-анализа. Построение biplot и scree plot. Применение t-SNE для визуализации многомерных данных.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 2.7	Кластерный анализ и введение в сети. Иерархическая кластеризация и метод k-средних. Создание простого графа в igraph на основе корреляций.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 3.1	Анализ генных сетей (GRN). Построение сети на основе корреляций. Расчет центральностей узлов. Выделение кластеров алгоритмом Лувена. Визуализация сети с gggraph.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 3.2	Обогащительный анализ (GSEA). Проведение анализа избыточной представленности (ORA) с помощью enrichGO. Визуализация результатов. Знакомство с GSEA для ранжированных списков генов.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 3.3	Работа с биологическими последовательностями. Чтение FASTA-	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>

		файлов. Анализ аминокислотного состава. Предсказание вторичной структуры белка.				
	Тема 3.4	Филогенетический анализ. Множественное выравнивание последовательностей. Построение NJ-дерева. Визуализация и аннотация дерева с помощью ggtree.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>
	Тема 3.5	Комплексный анализ генома. Расчет GC-состава и skew. Построение кольцевых диаграмм (circo). Знакомство с параллельными вычислениями и запуском внешних программ из R.	1		КСР	ПК-2 <i>ИДК ПК 2.2</i>

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

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
1.	Тема 1.1. Введение в R и основы синтаксиса.	Установить R и RStudio. Выполнить базовые операции в консоли. Написать скрипт с арифметическими операциями и созданием переменных различных типов.	ПК-2	ИДК ПК 2.2
2.	Тема 1.2. Работа с данными в R: векторы, индексация, функции.	Решить задачи на создание векторов, индексацию и написание пользовательских функций (задачи t1-t3 из лекции).	ПК-2	ИДК ПК 2.2
3.	Тема 1.3. Базовые структуры данных: матрицы, списки и датафреймы.	Создать матрицу 10×10, выполнить операции по замене элементов (задачи из лекции).	ПК-2	ИДК ПК 2.2

		Создать и модифицировать датафрейм на примере данных <code>iris</code> .		
4.	Тема 1.4. Основы визуализации в R. Базовые графики и ggplot2.	Построить графики <code>plot()</code> и <code>ggplot2</code> на данных <code>mtcars</code> . Настроить темы и подписи. Сохранить результаты.	ПК-2	ИДК ПК 2.2
5.	Тема 1.5. Обработка данных с <code>tidyverse</code> (<code>dplyr</code> , <code>tidyr</code>).	Выполнить трансформацию "сырого" датафрейма в "чистый" формат с использованием <code>pivot_longer/pivot_wider</code> . Провести фильтрацию и группировку данных с <code>dplyr</code> .	ПК-2	ИДК ПК 2.2
6	Тема 2.1. Статистический анализ в R: описательная статистика и корреляции.	Провести тест Шапиро-Уилка на разных выборках. Рассчитать корреляции Пирсона и Спирмена для данных <code>iris</code> . Построить корреляционную матрицу.	ПК-2	ИДК ПК 2.2
	Тема 2.2. Машинное обучение в R.	Ознакомиться с пакетом <code>caret</code> или <code>tidymodels</code> . Провести простую классификацию (например, <code>kNN</code>) на данных <code>iris</code> . Оценить точность модели.	ПК-2	ИДК ПК 2.2
	Тема 2.3. Сравнение групп: параметрические и непараметрические тесты.	Провести <code>t</code> -тест и ANOVA для данных <code>penguins</code> . Обосновать выбор критерия в зависимости от нормальности распределения.	ПК-2	ИДК ПК 2.2
	Тема 2.4. Реальный биоинформатический кейс: анализ данных ПЦР в реальном времени (qPCR).	Самостоятельно воспроизвести код для расчета <code>ddCt</code> . Построить столбчатые диаграммы и тепловые карты для визуализации экспрессии генов.	ПК-2	ИДК ПК 2.2

	Тема 2.5. Линейная регрессия и анализ выбросов.	Построить модель линейной регрессии для зависимости <code>Petal.Width ~ Petal.Length</code> . Проверить модель, интерпретировать R^2 , AIC. Провести детекцию выбросов с помощью <code>rosnerTest</code> .	ПК-2	ИДК ПК 2.2
	Тема 2.6. Анализ больших данных - 1: Уменьшение размерности.	Применить PCA к данным экспрессии генов (или <code>iris</code>). Визуализировать результаты с <code>ggbiplot</code> и <code>autoplot</code> . Интерпретировать нагрузки на компоненты.	ПК-2	ИДК ПК 2.2
	Тема 2.7. Анализ больших данных - 2: Кластерный анализ и введение в анализ сетей.	Провести иерархическую кластеризацию (<code>hclust</code>) и кластеризацию методом k-средних. Построить дендрограмму. Создать простой граф в <code>igraph</code> на основе корреляций.	ПК-2	ИДК ПК 2.2
	Тема 3.1. Анализ сетей в биологии. GRN и кластеризация сетей.	Построить генно-регуляторную сеть на основе корреляций. Выделить кластеры алгоритмом Лувена. Визуализировать сеть с помощью <code>ggraph</code> .	ПК-2	ИДК ПК 2.2
	Тема 3.2. Обоганительный анализ (Gene Set Enrichment Analysis, GSEA).	Запустить GSEA для набора дифференциально экспрессирующихся генов. Интерпретировать графики (<code>dotplot</code> , <code>cnetplot</code>).	ПК-2	ИДК ПК 2.2
	Тема 3.3. Работа с биологическими последовательностями в R.	Прочитать FASTA-файл с помощью <code>Biostrings</code> . Рассчитать аминокислотный состав и длину последовательностей. Выполнить	ПК-2	ИДК ПК 2.2

		предсказание вторичной структуры.		
	Тема 3.4. Множественное выравнивание и филогенетический анализ.	Провести множественное выравнивание последовательностей (MUSCLE/ClustalOmega). Построить филогенетическое дерево (NJ, ML). Визуализировать и аннотировать дерево с помощью ggtree.	ПК-2	ИДК ПК 2.2
	Тема 3.5. Комплексный анализ генома (включая консультации)	Подготовить итоговый проект: выполнить полный анализ небольшого набора данных (от загрузки до построения сети и GO-обогащения). Оформить отчет и презентацию. Подготовиться к зачету: повторить все темы курса.	ПК-2	ИДК ПК 2.2

4.4. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов является составной частью учебного процесса и имеет целью закрепление и углубление полученных знаний и навыков, поиск и приобретение новых знаний, а также выполнение учебных заданий, подготовку к предстоящим занятиям, зачетам и экзаменам.

Для организации самостоятельной работы по дисциплине «**Основы программирования в биологии**» используются следующие формы самостоятельной учебной работы:

- Работа над конспектом лекции.
- Подбор, изучение, анализ рекомендованной литературы и документации.
- Самостоятельное изучение отдельных тем, разделов, не изложенных в лекции.
- Написание и отладка программного кода (скриптов) на языке R.
- Выполнение практических заданий и решение задач по программированию.
- Подготовка докладов и презентаций по результатам анализа данных.
- Подготовка итогового проекта.
- Подготовка к зачету.

1. Работа с теоретическим материалом и литературой

Для изучения тем, не изложенных в лекции, рекомендуется использовать основную и дополнительную литературу, а также источники, найденные при помощи

информационно-справочных и поисковых систем (например, официальную документацию пакетов R на сайтах CRAN, Bioconductor). Для закрепления материала рекомендуется делать краткие конспекты по теме, фиксируя основные функции, их аргументы и синтаксис.

При изучении материала особое внимание следует уделять:

- Синтаксису языка и типичным ошибкам (регистр, типы данных, индексация).
- Логике работы функций и пакетов.
- Интерпретации результатов статистических тестов и графиков.
- Биологическому смыслу полученных результатов.

2. Выполнение практических заданий по программированию

Программирование требует регулярной практики. Для успешного освоения дисциплины студенту необходимо:

1. **Выполнять задания к каждой теме.** Задания размещены в конце соответствующих разделов лекций (например, t1-t4 в первой лекции). Решение задач должно быть оформлено в виде R-скрипта (файл с расширением .R).
2. **Экспериментировать с кодом.** Не ограничиваться только решением задач. Модифицировать примеры из лекции, менять параметры функций, запускать код на собственных небольших данных.
3. **Анализировать ошибки.** При возникновении ошибок (error messages) внимательно читать текст сообщения, искать его значение в справке (?function) или в интернете. Это важнейшая часть обучения программированию.
4. **Комментировать код.** Все скрипты должны содержать комментарии (строки, начинающиеся с #), поясняющие логику написанного. Это облегчает проверку преподавателем и помогает самому студенту разобраться в коде позже.

3. Подготовка докладов и презентаций

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

Критерии оценивания устного доклада:

- **Оценка «отлично».** В докладе полностью раскрыта тема, проанализировано современное состояние вопроса; студент свободно владеет материалом, излагает его логично, последовательно, лаконично, соблюдая основные правила культуры речи. Доклад сопровождается презентацией, которая отражает основные положения доклада, презентация составлена грамотно с соблюдением общих требований, правил шрифтового оформления, подачи графического материала (графики должны быть информативны, подписаны оси, указаны заголовки), имеются ссылки на приведенные рисунки и источники, приводится список использованной литературы. При обсуждении доклада студент дает исчерпывающие, аргументированные, корректные ответы на вопросы.

- **Оценка «хорошо».** Тема раскрыта, приведено достаточное количество материала, но при этом материал в недостаточной степени проанализирован автором. Презентация не в полной степени соответствует общим требованиям (например, графики не подписаны, отсутствуют заголовки). Ответы студента не на все вопросы являются исчерпывающими и аргументированными.
- **Оценка «удовлетворительно».** Тема раскрыта не полно, материал приведен как простая констатация фактов, не проанализирован, студент показывает поверхностные знания. Презентация частично соответствует установленным требованиям, код в презентации нечитаемый или отсутствует. При обсуждении доклада студент не всегда дает правильные, исчерпывающие ответы на задаваемые вопросы.
- **Оценка «неудовлетворительно».** Тема доклада не раскрыта, скудный объем приведенных материалов; презентация отсутствует или не соответствует теме. При обсуждении доклада студент не дает ответы или они не соответствуют заданным вопросам.

4. Подготовка итогового проекта

Итоговый проект является формой промежуточной аттестации и выполняется в течение последних недель семестра. Цель проекта — продемонстрировать владение всем комплексом навыков, полученных в курсе.

Структура проекта:

1. **Постановка задачи.** Краткое описание биологической проблемы и цели анализа.
2. **Данные.** Описание набора данных (откуда взят, какие переменные, объем выборки). Данные могут быть предоставлены преподавателем или взяты из открытых источников (например, NCBI, UCI Machine Learning Repository).
3. **Анализ.** Основная часть, включающая:
 - Загрузку и первичную обработку данных.
 - Разведочный анализ (визуализация, описательные статистики).
 - Применение статистических методов (корреляции, сравнение групп, регрессия) или методов машинного обучения.
 - Применение биоинформатических методов (работа с последовательностями, построение сетей, обогатительный анализ) — в зависимости от выбранных данных.
4. **Результаты и выводы.** Четкая формулировка полученных результатов и их биологическая интерпретация.
5. **Приложение.** Полный R-скрипт с подробными комментариями.

Проект оформляется в виде письменного отчета и защищается устно с презентацией (5-7 минут). Критерии оценивания аналогичны критериям для устного доклада, но с повышенными требованиями к полноте анализа и качеству кода.

5. Подготовка к зачету

Подготовка к зачету включает в себя повторение всех тем курса, просмотр лекций и собственных конспектов, а также повторное выполнение ключевых практических заданий. Особое внимание следует уделить:

- Пониманию базовых структур данных в R (векторы, списки, датафреймы).
- Умению применять функции пакетов `dplyr` и `ggplot2`.

- Умению выбирать и интерпретировать статистические тесты.
- Пониманию логики основных биоинформатических методов (РСА, кластеризация, GSEA).

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

V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) перечень литературы

1. Основная литература

1. Леск, Артур. Введение в биоинформатику : пер. с англ. / А. М. Леск ; ред.: А. А. Миронов, В. К. Швядаса. - М. : Бином. Лаборатория знаний, 2009. - 318 с. - Пер. изд.: Introduction to bioinformatics / Arthur M. Lesk. - Oxford, 2002. - ISBN 978-5-94774-501-6

2. Дополнительная литература

1. Щербаков, Дмитрий Юрьевич. Актуальные проблемы современной генетики. Биоинформационные методы анализа биоразнообразия : учеб. пособие / Д. Ю. Щербаков, Р. В. Адельшин, М. В. Коваленкова ; Иркут. гос. ун-т, Биол.-почв. фак. - Иркутск : Изд-во ИГУ, 2018. - 119 с. - ISBN 978-5-9624-1600-7

б) периодические издания

в) список авторских методических разработок:.

г) базы данных, информационно-справочные и поисковые системы

1. Научная Электронная Библиотека <http://www.e-library.ru>
2. Информационная система «Единое окно доступа к образовательным ресурсам» (<http://window.edu.ru>)
3. ЭБС «ЮРАЙТ». Адрес доступа: <https://www.biblio-online.ru/>
4. ЭБ Издательского центра «Академия». Адрес доступа: <http://www.academia-moscow.ru>
5. <http://www.fptl.ru/biblioteka/biotehnologiya.html>
6. <http://www.medbook.net.ru/010512.shtml>
7. Союз образовательных сайтов - Естественные науки
8. <http://tusearch.blogspot.com> - Поиск электронных книг, публикаций, законов, ГОСТов на сайтах научных электронных библиотек.
9. Google Scholar –Поисковая система по научной литературе.
10. Science Research Portal - Научная поисковая система, осуществляющая полнотекстовый поиск в журналах многих крупных научных издательств, таких как Elsevier, Highwire, IEEE, Nature, Taylor & Francis и др. Ищет статьи и документы в открытых научных базах данных: Directory of Open Access Journals, Library of Congress Online Catalog, Science.gov и Scientific News.

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование:

Специальные помещения:

Аудитория для проведения занятий лекционного типа оборудована:

специализированной (учебной) мебелью на 12 посадочных мест,

техническими средствами обучения: Доска аудиторная меловая, Проектор BenQ MS504, служащими для представления учебной информации большой аудитории по дисциплине «Методы анализа молекулярно-генетических данных»;

Аудитория для проведения занятий семинарского типа оборудована:

специализированной (учебной) мебелью на 12 посадочных мест, биохимическая лаборатория (лабораторные столы - 4 шт.); раковина с тумбой - 1 шт., Деревянные тумбы для хранения реактивов - 2 шт., швытяжной ЛК-1500 ШВ - 2 шт., весы аналитические ГОСМЕТР Ленинград - 1 шт., фотоэлектроколориметр КФК-2 - 1 шт., аквадистиллятор электрический АЭ-14-«Я-ФП»-01 - 1 шт., термостат электрический суховоздушный ТС-1/80 СПУ - 1 шт.;

техническими средствами обучения: доска аудиторная меловая, проектор BenQ MS504, служащими для представления учебной информации большой аудитории по дисциплине «Методы анализа молекулярно-генетических данных»;

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.

Помещения для хранения и профилактического обслуживания учебного оборудования - Аудитория оборудована:

специализированной мебелью на 8 посадочных мест; Шкаф вытяжной ЛК-1500 ШВ+вентилятор - 2 шт., Стол двухтумбовый - 5 шт., Стол одностумбовый - 4 шт., Стол компьютерный - 1 шт., Металлические тумбы для хранения лабораторной посуды и оборудования - 4 шт., Деревянные тумбы для хранения лабораторной посуды и оборудования - 5 шт., Шкаф-купе двухдверный - 1 шт., Шкаф металлический - 1 шт., Холодильник NORD ДХ-241-0-010 - 1 шт., Электроплита Луч - 1 шт., Раковина с тумбой - 1 шт., Шкаф-купе трехдверный - 1шт., Шкаф книжный - 3 шт., Микроскоп Биомед 2 Led - 7 шт., Микроскоп Levenhuk D870T - 1 шт., Микроскоп Levenhuk D870T тринокуляр - 1 шт., Микроскоп Микромед Р-1-LED - 1 шт., Микроскоп МЛ-5-Б - 1 шт., Микроскоп биологический МБ-1600Б - 1 шт., Микроскоп Р-14 - 4 шт., Микроскоп Levenhuk 2L NG – 5 шт., Светитель ОИ-12 - 1 шт., Фазовый контраст КФ-3 - 1 шт., Фазовый контраст КФС - 1 шт., рН-метр иономер универсальный ЭВ-74 - 1 шт., Спектрофотометр ПЭ-5300 ВИ - 1 шт., Магнитная мешалка ММ-5 - 5 шт., Весы аналитические ВЛР-200 - 1 шт., Весы торсионные ВТП-500 - 4 шт., Весы торсионные WAGA TORSYJNA-WT - 3 шт., Проектор Оверхед GEHA OHP Ecovision 24/3 - 1 шт., Системный блок в комплекте ASUS - 1 шт., Монитор BenQ DL2215 - 1 шт., Ноутбук Lenovo G580 в комплекте - 1 шт., Мультифункциональное устройство SAMSUNG M2070 - 1 шт., Сканер HP Scanjet G2410 - 1 шт., Принтер Canon LBP 2900 – 1 шт.

6.2. Программное обеспечение:

Kaspersky Endpoint Security для бизнеса - Стандартный Russian Edition;

Foxit PDF Reader 8.0;

LibreOffice 5.2.2.2;

Ubuntu 14.0;

АСТ-Тест Plus 4.0 (на 75 одновременных подключений) и Мастер-комплект (АСТ-Maker и АСТ-Converter).

6.3. Технические и электронные средства:

Презентации по всем темам курса.

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Для освоения дисциплины «Методы анализа молекулярно-генетических данных» применяются следующие образовательные технологии:

- *Информационная лекция.* Лекция – это сжатое изложение основных научных фактов, что является базой для анализа рассуждений, оценок.

- *Лекция-визуализация.* Учит студентов преобразовывать устную и письменную информацию в визуальную форму, что формирует у них профессиональное мышление за счет систематизации и выделения наиболее значимых, существенных элементов содержания обучения. Задача преподавателя использовать такие формы наглядности, которые не только дополняют словесную информацию, но и сами являются носителями информации (схемы, рисунки, слайды-презентации, и т.п.). Этот вид лекции лучше всего использовать на этапе введения студентов в новый раздел, тему дисциплины.

- *Проблемная лекция.* В отличие от содержания информационной лекции, которое предлагается преподавателем в виде известного, подлежащего лишь запоминанию материала, на проблемной лекции новое знание вводится как неизвестное для обучающихся. Проблемная лекция начинается с вопросов, с постановки проблемы, которую в ходе изложения материала необходимо решить. Лекция строится таким образом, что познания обучающегося приближаются к поисковой, исследовательской деятельности. Здесь участвуют мышление обучающегося и его личностное отношение к усваиваемому материалу.

- *Лекция-беседа.* Предполагает непосредственный контакт преподавателя с аудиторией. Преимущество лекции-беседы состоит в том, что она позволяет привлекать внимание студентов к наиболее важным вопросам темы, определять содержание и темп изложения учебного материала с учетом особенностей студентов.

- *Практические занятия* – это занятие, проводимое под руководством преподавателя в учебной аудитории, направленное на углубление научно-теоретических знаний и овладение определенными методами самостоятельной работы, которое формирует практические умения. Одной из форм практических занятий в вузе является семинар.

- *Коллоквиумы* – вид учебного занятия, проводимого с целью проверки и оценивания знаний учащихся. Коллоквиум может проводиться в форме индивидуальной беседы преподавателя со студентом или как массовый опрос. В ходе группового обсуждения студенты учатся высказывать свою точку зрения по определенному вопросу, защищать свое мнение, применяя знания, полученные на занятиях по предмету. В ходе коллоквиума также проверяются рефераты, другие письменные работы студентов, проводится заслушивание докладов.

- *Самостоятельная работа студентов* (см. п.4.4).

- *Дистанционные образовательные технологии.* Под дистанционными образовательными технологиями понимаются образовательные технологии, реализуемые в основном с применением информационно-телекоммуникационных сетей. При освоении дисциплины «Методы анализа молекулярно-генетических данных» используются следующие технологии:

▪ кейсовая технология – форма дистанционного обучения, основанная на предоставлении обучающимся информационных образовательных ресурсов в виде специализированных наборов учебно-методических комплексов с использованием различных видов носителей информации (кейсов);

▪ интернет-технология – способ дистанционной передачи информации, основанный на использовании глобальных и локальных компьютерных сетей для обеспечения доступа обучающихся к информационным образовательным ресурсам и для формирования

совокупности методических, организационных, технических и программных средств реализации и управления учебным процессом независимо от места нахождения его субъектов. Используется Образовательный портал ИГУ - educa.isu.ru.

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Оценочные материалы для входного контроля

В качестве оценочных средств для входного контроля оценки уровня знаний студентов используется собеседование. В процессе собеседования оценивается уровень владения базовыми знаниями, умениями, навыками, необходимыми для начала обучения по дисциплине «Молекулярные механизмы гормональной регуляции», определяется степень владения новым материалом до начала его изучения.

Оценочные материалы текущего контроля формируются в соответствии с ЛНА университета

В рамках дисциплины «Молекулярные механизмы гормональной регуляции» используются следующие формы текущего контроля:

- устный опрос;
- письменная работа;
- тест;
- доклад;
- контроль самостоятельной работы.

Фонд оценочных средств включает:

- фонд тестовых заданий по дисциплине,
- тематика и материалы заданий,
- перечень тем докладов,
- вопросы для самостоятельного изучения (СРС)
- вопросы и билеты для экзамена,
- критерии оценки знаний студентов.

Назначение оценочных средств: выявить сформированность компетенции ПК-1 (см. п. III).

Оценочные материалы для промежуточной аттестации в форме зачёта

Форма промежуточной аттестации - **зачёт**. ОС этого типа должны выявлять степень освоения теоретических знаний как базу для формирования компетенций, умения их применять в ситуациях, моделирующих профессиональную деятельность, а также сформированность компетенции ПК-1, заявленной в п. III.

Примерный список вопросов к зачёту

1. С помощью пакета `rentrez` найти и скачать из GenBank нуклеотидную последовательность гена COI (цитохромоксидаза I) для трех видов рыб (например, *Clupea harengus*, *Engraulis encrasicolus*, *Gadus morhua*). Выровнять последовательности алгоритмом ClustalOmega, построить NJ-дерево, визуализировать его с помощью `ggtree` и сделать вывод о филогенетических отношениях.

2. Скачать полные геномы пяти штаммов бактерий *Escherichia coli* из базы NCBI Assembly (используя `rentrez` или вручную). Загрузить их в любой веб-сервис

полногеномной филогении (например, REALPHY, Gingr, или бактериальную пайплайн в BV-BRC). Описать полученное дерево, соотнести его с данными по серотипам или патогенности штаммов. Сопоставить результаты с современным представлением о филогении кишечной палочки.

3. Используя датасет DryBeanDataset (сорта фасоли), провести разведочный анализ данных (визуализация, описательная статистика). Применить три разных алгоритма машинного обучения (например, случайный лес, kNN, SVM) для классификации сортов по морфологическим признакам. Сравнить точность моделей, выбрать лучший алгоритм и обосновать выбор.

4. С помощью пакета coRdon проанализировать частоту использования кодонов в наборе CDS-последовательностей из пяти близкородственных видов бактерий. Выполнить снижение размерности (PCA или t-SNE) на основе частот кодонов и определить, кластеризуются ли виды в соответствии с их таксономией.

5. Скачать с помощью rentrez последовательности 16S рРНК для десяти видов бактерий из разных филумов (Firmicutes, Proteobacteria, Actinobacteria). Провести множественное выравнивание (MAFFT), построить максимально-правдоподобное дерево (ML), оценить бутстреп-поддержку. Сравнить полученную топологию с общепринятой классификацией бактерий.

6. Используя данные RNA-seq (или симулированные данные экспрессии генов), построить корреляционную генную сеть. Выделить кластеры генов с помощью алгоритма Лувена. Провести обогатительный анализ (GO, KEGG) для генов из двух самых крупных кластеров и интерпретировать их возможную биологическую роль.

7. Скачать из базы данных Arabidopsis Information Portal (или другой источник) последовательности промоторов (1000 п.н. выше старта транскрипции) для генов, связанных с ответом на засуху. С помощью пакетов для поиска мотивов (например, universalmotif или rGADEM) найти обогащенные цис-регуляторные элементы в этих промоторах. Визуализировать найденные мотивы.

8. Провести анализ GC-состава и AT/GC-skew для полного митохондриального генома человека. Определить регионы смены знака skew, сопоставить их с положением генов и ориентацией транскрипции. Визуализировать результаты в виде линейных графиков и круговой диаграммы (circos plot) с аннотацией генов.

9. Загрузить данные по экспрессии генов (например, из GEO) для двух условий (контроль/обработка). Провести дифференциальный анализ (например, с помощью limma или простого t-теста). Отобрать гены с $|\log_2FC| > 1$ и $p\text{-adj} < 0.05$. Выполнить для них обогатительный анализ (GSEA или ORA) и представить результаты в виде dotplot и cnetplot.

10. Написать R-функцию, которая принимает на вход идентификатор таксона (например, род *Drosophila*), скачивает все доступные последовательности гена *Adh* через rentrez, автоматически проводит их выравнивание, строит филогенетическое дерево методом neighbor-joining, раскрашивает ветви по видовой принадлежности и сохраняет результат в PDF-файл.

11. Используя набор данных по аминокислотным последовательностям неизвестного белка, предсказать его вторичную структуру (с помощью PredictNEC), рассчитать физико-химические свойства (изоэлектрическая точка, заряд, гидрофобность) и на основе полученных данных предложить гипотезу о его возможной функции (сравнить с известными белками через BLAST).

12. Скачать данные по микросателлитным маркерам (SSR) для популяции какого-либо вида растений или животных. Провести анализ главных компонент (PCA) на основе частот аллелей. Визуализировать распределение образцов в пространстве первых двух главных компонент и определить, наблюдается ли структурирование популяции (например, по географическому происхождению).

13. С помощью пакета `dbscan` выполнить кластеризацию данных по экспрессии генов (или любых многомерных биологических данных). Подобрать оптимальный параметр `epsilon` с помощью анализа силуэта. Визуализировать результаты кластеризации и интерпретировать выделенные кластеры (например, соответствуют ли они разным типам клеток или условиям эксперимента).

14. Скачать из PDB (через пакет `bio3d`) структуры гемоглобина человека в оксигенированной и дезоксигенированной форме. Провести сравнение структур, рассчитать RMSD между ними. Визуализировать конформационные изменения с помощью `ggraph` или `rgl` и описать, какие именно домены или участки изменили свое положение.

15. Используя данные по SNP-маркерам для набора образцов, провести анализ методом t-SNE. Интерпретировать полученные кластеры — соответствуют ли они популяциям, географическому происхождению или фенотипическим признакам. Сравнить результат t-SNE с результатом PCA на тех же данных.

Разработчик:


(подпись)

И.В. Горбенко

Программа составлена в соответствии с требованиями ФГОС ВО по направлению 06.03.01 «Биология» и элективному модулю «Физиология и биохимия».

Программа рассмотрена на заседании кафедры физиологии, молекулярной биологии и психофизиологии

«01» апреля 2026 г.

Протокол № 1 Зав. кафедрой  И.Н. Гутник

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