

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ВЕРХНЕВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ
АГРОБИОТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «Верхневолжский ГАУ»)**

ИНСТИТУТ ВЕТЕРИНАРНОЙ МЕДИЦИНЫ И БИОИНЖЕНЕРИИ

УТВЕРЖДЕНА
протоколом заседания
методической комиссии института
№ 2 от «14» мая 2026 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

**«Современные методы анализа биологических данных
с основами биоинформатики»**

Направление подготовки / специальность	36.03.02. «Зоотехния»
Направленность(и) (профиль(и))	«Генетика, генетические методы в практике животноводства»
Уровень образовательной программы	Бакалавриат
Форма(ы) обучения	Очная, заочная
Трудоемкость дисциплины, ЗЕТ	3
Трудоемкость дисциплины, час.	108

Разработчик:

Доцент кафедры зооинженерии

Колганов А.Е.

(подпись)

СОГЛАСОВАНО:

Заведующий кафедрой зооинженерии

Колганов А.Е.

(подпись)

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

Цель дисциплины «Современные методы анализа биологических данных с основами биоинформатики» заключается в формировании у обучающихся системы фундаментальных знаний о вычислительных, статистических и информационных методах обработки биологических данных и прикладных профессиональных компетенций в области применения современных биоинформатических инструментов, языков программирования (R, Python) и алгоритмов машинного обучения для анализа геномных, транскриптомных, фенотипических и зоотехнических данных, направленных на принятие научно обоснованных решений в селекции, генетике и биотехнологии сельскохозяйственных животных.

Задачи дисциплины:

Теоретические основы биоинформатики: изучение принципов организации биологических баз данных, форматов хранения и передачи информации (FASTA, FASTQ, SAM/BAM, VCF, GFF), а также алгоритмов поиска, выравнивания и аннотации биологических последовательностей. Статистические методы и программирование: освоение основ статистического анализа биологических данных, языков программирования R и Python, работы с библиотеками (tidyverse, ggplot2, Bioconductor, pandas, scikit-learn) для обработки, визуализации и интерпретации данных. Анализ геномных и транскриптомных данных: формирование навыков работы с данными высокопроизводительного секвенирования (NGS) - контроль качества, выравнивание на референсный геном, вызов вариантов, дифференциальная экспрессия генов (RNA-seq), функциональная аннотация и анализ обогащения путей (GO, KEGG). Многомерная статистика и машинное обучение: изучение методов снижения размерности (PCA, t-SNE, UMAP), кластеризации, классификации и регрессии, применения алгоритмов машинного обучения (случайный лес, метод опорных векторов, нейронные сети) для прогнозирования продуктивности, здоровья и племенной ценности животных. Работа с зоотехническими и фенотипическими данными: освоение методов анализа больших массивов данных программ управления стадом, интеграции геномных и фенотипических данных, построения предиктивных моделей и цифровых двойников животных. Визуализация и воспроизводимость исследований: изучение принципов создания публикационной графики, интерактивных дашбордов, а также обеспечения воспроизводимости биоинформатических анализов (контейнеризация, системы контроля версий, Jupyter notebooks).

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В соответствии с учебным планом

дисциплина относится к*

Часть, формируемой участниками образовательных отношений

Статус дисциплины**

по выбору

Обеспечивающие (предшествующие) дисциплины

кормление животных, зоотехнический анализ кормов

Обеспечиваемые (последующие) дисциплины

направленное выращивание ремонтного молодняка, НИР, ГИА

* базовой / вариативной

** обязательная / по выбору / факультативная

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ) (ХАРАКТЕРИСТИКА ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ)

Шифр и наименование компетенции	Индикатор(ы) достижения компетенции / планируемые результаты обучения	Номер(а) раздела(ов) дисциплины (модуля), отвечающего(их) за формирование данного(ых) индикатора(ов) достижения компетенции
ПК-9 Способен формировать и решать задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации	ИД-1ПК9 Знать: Понятия и задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации ИД-2 ПК9 Уметь: использовать в практической деятельности методы решения производственных задач связанных с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации ИД-3 ПК9 Владеть: практическими навыками формирования и решения задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации	все разделы
ПК-10 Способен владеть методами селекции, кормления, содержания и использования различных видов животных	ИД-1ПК10 Знать: Методы селекции, кормления, содержания и использования различных видов животных ИД-2 ПК10 Уметь применять методы селекции, кормления, содержания и использования различных видов животных ИД-3 ПК10 Владеть: Способностью владеть методами селекции, кормления, содержания и использования различных видов животных	все разделы
ПК-11 Способен применять современные методы исследований в области животноводства	ИД-1ПК11 Знать: Современные методы исследований в области животноводства ИД-2 ПК11 Уметь: Применять современные методы исследований в области животноводства ИД-3 ПК11 Владеть: Способностью применять современные методы исследований в области животноводства	все разделы

ПК-12 Способен использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных	ИД-1ПК12 Знать: Методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных ИД-2 ПК12 Уметь: Использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных ИД-3 ПК12 Владеть: Способностью использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных	все разделы
ПК-1 Способен выбирать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных	ИД-1ПК1 Знать: Понятия, теоретические основы выбора и соблюдения режимов содержания животных, составления рационов кормления, прогнозирования последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных ИД-2 ПК1 Уметь: Применять в практической деятельности выбирать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных ИД-3 ПК1 Владеть: Способен осознано использовать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных	все разделы

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

4.1. Содержание дисциплины (модуля)

4.1.1. Очная форма:

№ п/п	Темы занятий	Виды учебных занятий и трудоёмкость, час.				Кон- троль знаний*	Применяемые ак- тивные и интерак- тивные техноло- гии обучения
		лекции	лабора- торные	практиче- ские	самостоя- тельная работа		
1.	Основы биоинформатики и биологические базы данных						
1.1	Предмет и зада- чи биоинформа- тики. Организа- ция биологиче- ских баз данных (NCBI, Ensembl, EMBL-EBI). Форматы дан- ных: FASTA, GenBank, GFF/GTF. Алгоритмы по- иска и выравни- вания последова- тельности:	2	2	-	8	УО, Т	Лекция- визуализация; Ра- бота с интерфей- сом NCBI: поиск генов, последова- тельности, за- грузка данных в формате FASTA
1.2	BLAST, FASTA, множественное выравнивание (ClustalW, MUSCLE).	2	2	-	8	УО, ПР	Практикум в ком- пьютерном классе: BLAST-поиск, по- строение множе- ственного выравни- вания, интер- претация E-value
2.	Статистические методы и основы программирования в R и Python						
2.1	Основы стати- стики для биоло- гов: описатель- ная статистика, проверка гипотез (t-тест, ANOVA, хи-квадрат), кор- реляция и ре- грессия.	2	2	-	8	УО, Т	Лекция с разбором примеров; Реше- ние задач в R: рас- чет средних, дис- персий, построе- ние гистограмм и box-plot
2.2	Введение в про- граммирование на R и Python: синтаксис, структуры дан- ных, работа с па- кетами (tidyverse, pandas, numpy).	2	4	-	10	УО, ПР	Лабораторная ра- бота: загрузка данных из CSV, фильтрация, групп- пировка, базовая визуализация (ggplot2, matplotlib)
3.	Анализ геномных данных NGS						
3.1	Контроль каче-	2	2	-	8	УО, ПР	Анализ реальных

№ п/п	Темы занятий	Виды учебных занятий и трудоёмкость, час.				Контроль	Применяемые активные и интерактивные
	ства данных NGS: форматы FASTQ, метрики Phred, инструменты FastQC, Trimmomatic. Выравнивание на референсный геном (BWA, Bowtie2). Вызов генетических вариантов (Variant Calling): форматы SAM/BAM, VCF. Инструменты GATK, SAMtools, bcftools. Аннотация вариантов.						FASTQ-файлов: интерпретация отчетов FastQC, фильтрация прочтений, запуск BWA
3.2		2	2	-	8	УО, ПР, 3	Лабораторная работа: конвертация форматов, вызов SNP, фильтрация по качеству, аннотация с помощью SnpEff
4.	Анализ транскриптомных данных (RNA-seq) Дизайн эксперимента RNA-seq: биологические и технические повторности, глубина секвенирования. Количественная оценка экспрессии генов (featureCounts, HTSeq). Дифференциальная экспрессия генов: нормализация данных (TPM, FPKM), пакеты DESeq2, edgeR. Визуализация результатов (heatmaps, volcano plots).						
4.1		2	2	-	8	УО, Т	Лекция-дискуссия; Расчет необходимого объема секвенирования, анализ матрицы экспрессии
4.2		2	4	-	10	УО, ПР, 3	Лабораторная работа: запуск DESeq2 в R, построение volcano plot, анализ списков дифференциально экспрессируемых генов
5.	Многомерная статистика и машинное обучение в биологии Методы снижения размерности:						
5.1	PCA (метод главных компонент), t-SNE, UMAP. Класте-	2	2	-	8	УО, ПР	Визуализация многомерных биологических данных: PCA-plot для образцов RNA-seq, дендрограммы

№ п/п	Темы занятий	Виды учебных занятий и трудоёмкость, час.				Контроль	Применяемые активные и интерактивные кластеризации
5.2	ризация: k-means, иерархическая кластеризация. Основы машинного обучения: классификация и регрессия. Алгоритмы: случайный лес (Random Forest), метод опорных векторов (SVM). Применение для прогнозирования продуктивности.	2	4	-	8	УО, ПР, 3	Обучение модели на зоотехнических данных (удой, жир, белок, экстерьер), оценка точности (ассурасу, ROC-кривая)
6.	Интеграция данных, визуализация и воспроизводимость						
6.1	Интеграция геномных, транскриптомных и фенотипических данных. Функциональная аннотация и анализ обогащения путей (GO, KEGG). Воспроизводимость биоинформатических исследований: Jupyter	2	2	-	4	УО, ПР	Анализ обогащения GO-терминов для списка генов, построение сетей взаимодействий (STRING)
6.2	notebooks, R Markdown, системы контроля версий (Git), контейнеризация (Docker).	2	2	-	6	УО, 3	Создание интерактивного отчета в R Markdown/Jupyter, защита мини-проекта по анализу биологических данных
	Итого:	20	20	-	68		
	Форма контроля: Зачет (3)						

* Указывается форма контроля. Например: УО – устный опрос, КЛ – конспект лекции, КР – контрольная работа, ВЛР – выполнение лабораторной работы, ВПР – выполнение практической работы, К – коллоквиум, Т – тестирование, Р – реферат, Д – доклад, ЗКР – защита курсовой работы, ЗКП – защита курсового проекта, Э – экзамен, З – зачет.

4.1.2. Заочная форма:

№ п/п	Темы занятий	Виды учебных занятий и трудоёмкость, час.				Контроль знаний*	Применяемые активные и интерактивные технологии обучения
		лекции	практические (семинарские)	лабораторные	самостоятельная работа		
1.1.							
1.2.							
1.3.							
1.4.							
1.5.							
2.1.							
2.2.1.							
2.2.2.							
2.2.3.							
2.2.4.							

4.2.1. Очная форма:

Вид занятий	1 курс		2 курс		3 курс		4 курс		5 курс	
	1 сем.	2 сем.	3 сем.	4 сем.	5 сем.	6 сем.	7 сем.	8 сем.	9 сем.	10 сем.
Лекции							20			
Лабораторные							20			
Практические										
Итого контактной работы							40			
Самостоятельная работа							68			
Форма контроля							3			

4.2.2. Заочная форма:

Вид занятий	1 курс	2 курс	3 курс	4 курс	5 курс	6 курс
Лекции						
Лабораторные						
Практические						
Итого контактной работы						
Самостоятельная работа						
Форма контроля						

5. ОРГАНИЗАЦИЯ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Организация самостоятельной работы студентов основана на ПВД-12 «О самостоятельной работе обучающихся»

5.1. Содержание самостоятельной работы по дисциплине (модулю)

Темы выносимые на самостоятельную проработку

1. История и этапы развития биоинформатики: от первых баз данных до эпохи больших биологических данных.
2. Предмет, цели и задачи биоинформатики в современной биологии и зоотехнии.
3. Архитектура международных биологических баз данных: NCBI, EMBL-EBI, DDBJ.
4. Специализированные базы данных по сельскохозяйственным животным: Ensembl, Animal QTLdb, FAANG.
5. Форматы хранения биологических данных: FASTA, GenBank, GFF/GTF.
6. Алгоритмы поиска гомологичных последовательностей: BLAST, FASTA.
7. Множественное выравнивание последовательностей: ClustalW, MUSCLE, MAFFT.
8. Основы молекулярной филогенетики: методы построения деревьев (NJ, ML, Bayesian).
9. Дескриптивная статистика для биологов: меры центральной тенденции, дисперсии, визуализация.
10. Проверка статистических гипотез: t-тест, ANOVA, хи-квадрат, непараметрические тесты.
11. Корреляционный и регрессионный анализ в биологии.
12. Введение в язык R: синтаксис, структуры данных, пакеты tidyverse.
13. Экосистема R для биоинформатики: Bioconductor, управление проектами.
14. Введение в Python для биологов: pandas, numpy, matplotlib, seaborn.
15. Сравнение R и Python для анализа биологических данных.
16. Работа с табличными данными: импорт, фильтрация, группировка, слияние.
17. Визуализация данных в R (ggplot2): слои, эстетика, публикационная графика.
18. Визуализация данных в Python (matplotlib, seaborn).
19. Основы высокопроизводительного секвенирования (NGS): платформы Illumina, Ion Torrent, PacBio.
20. Форматы данных NGS: FASTQ, метрики качества Phred.
21. Контроль качества NGS-данных: FastQC, Trimmomatic, fastp.
22. Предобработка прочтений: удаление адаптеров, фильтрация по качеству.
23. Выравнивание прочтений на референсный геном: BWA-MEM, Bowtie2, STAR.
24. Форматы SAM/BAM: структура, конвертация, SAMtools.
25. Вызов генетических вариантов: GATK, bcftools, FreeBayes.
26. Формат VCF: структура, фильтрация SNP и InDel.
27. Аннотация генетических вариантов: SnpEff, VEP.
28. Дизайн эксперимента RNA-seq: повторности, глубина секвенирования.
29. Количественная оценка экспрессии: featureCounts, HTSeq, TPM, FPKM.
30. Нормализация данных RNA-seq: методы DESeq2, edgeR.
31. Дифференциальная экспрессия генов: DESeq2, edgeR, limma-voom.
32. Визуализация результатов RNA-seq: volcano plot, heatmap, PCA-plot.
33. Функциональная аннотация и анализ обогащения: GO, KEGG, clusterProfiler.
34. Анализ белок-белковых взаимодействий: база STRING.
35. Метод главных компонент (PCA): математические основы, интерпретация.
36. Нелинейные методы снижения размерности: t-SNE, UMAP.
37. Кластеризация биологических данных: k-means, иерархическая кластеризация, DBSCAN.
38. Основы машинного обучения: обучение с учителем vs без учителя.

39. Алгоритмы классификации: логистическая регрессия, SVM, Random Forest, k-NN.
40. Оценка качества моделей: accuracy, precision, recall, F1-score, ROC-кривая, AUC.
41. Применение машинного обучения в зоотехнии: прогнозирование продуктивности и здоровья.
42. Интеграция геномных, транскриптомных и фенотипических данных.
43. Основы системной биологии: сети регуляции генов, метаболические пути.
44. Воспроизводимость исследований: принципы FAIR-данных.
45. Интерактивные отчёты: Jupyter notebooks, R Markdown, Quarto.
46. Системы контроля версий: Git, GitHub/GitLab.
47. Контейнеризация в биоинформатике: Docker, Singularity.
48. Облачные вычисления: Galaxy, Terra, AWS, Google Cloud.
49. Анализ single-cell RNA-seq: технологии 10x Genomics, пакеты Seurat, Scanpy.
50. Пространственная транскриптомика: технологии Visium, MERFISH.
51. Метагеномика и анализ микробиома: 16S rRNA, QIIME2, MetaPhlAn.
52. Искусственный интеллект и глубокое обучение в биологии: нейронные сети, AlphaFold2.
53. Языковые модели в биоинформатике: ESM, ProtTrans.
54. Биоинформатика в селекции: интеграция с геномной селекцией, расчёт GEBV.
55. Цифровые двойники животных: концепция, источники данных, предиктивные модели.
56. Этика и правовые аспекты работы с биологическими данными.
57. Современные тренды биоинформатики: пангеномика, multi-omics.
58. Карьерные возможности в биоинформатике.

5.2. Контроль самостоятельной работы

Оценка результатов самостоятельной работы организуется следующим образом:

- Оценка результатов самостоятельной работы по итогам всех форм контроля принятых в рамках дисциплины: устного опроса, коллоквиумов, реферата зачёта.

Оценка результатов самостоятельной работы проводится в соответствии с ПВД-07 «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся».

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Основная учебная литература, необходимая для освоения дисциплины (модуля)

Селекция животных : учебно-методическое пособие / составитель О. В. Абрампальская. — Тверь : Тверская ГСХА, 2016. — 30 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/134140> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

Разведение сельскохозяйственных животных / А. Х. Хайитов, С. А. Брагинец, У. Ш. Джуроева [и др.] ; под редакцией А. Х. Хайитов. — 2-е изд., стер. — Санкт-Петербург : Лань, 2023. — 248 с. — ISBN 978-5-507-47109-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/328505> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

6.2. Дополнительная учебная литература, необходимая для освоения дисциплины (модуля)

Контроль качества данных в животноводстве на основе биологических норм и интеллектуального анализа / В. Ф. Ахмедли, V. F. Akhmedli, С. Ю. Бакоев [и др.] // Аграрный вестник Урала. — 2025. — № 10. — С. 1589-1598. — ISSN 1997-4868. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/journal/issue/375276> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

Секвенирование одной молекулы как инновационный способ изучения эпигенетики и генетического разнообразия сельскохозяйственных животных / А. П. Дысин, А. Р. Dysin, Н. В. Дементьева, N. V. Dement'eva // Молочнохозяйственный Вестник. — 2024. — № 4 (56). — С. 63-82. — ISSN 2225-4269. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/journal/issue/367994> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

Яковенко, А. М. Биометрические методы анализа качественных и количественных признаков в зоотехнии : учебное пособие / А. М. Яковенко, Т. И. Антоненко, М. И. Селионова. — Ставрополь : СтГАУ, 2013. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/45734> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

6.3. Ресурсы сети «Интернет», необходимые для освоения дисциплины (модуля)

- 1) Научная электронная библиотека e-library.ru / <http://e-library.ru>.
- 2) Единое окно доступа к образовательным ресурсам. Электронная библиотека / <http://window.edu.ru>

6.4. Методические указания для обучающихся по освоению дисциплины (модуля)

Кадиев, А. К. Генетика. Руководство к практическим занятиям : учебное пособие для вузов / А. К. Кадиев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2025. — 252 с. — ISBN 978-5-507-50759-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/462710> (дата обращения: 25.06.2026). — Режим доступа: для авториз. пользователей.

6.5. Информационные справочные системы, используемые для освоения дисциплины (модуля) (при необходимости)

1. Научная электронная библиотека eLIBRARY.R (<http://eLIBRARY.RU>);
2. ЭБС издательства «ЛАНЬ» (<http://e.lanbook.ru>);
3. «Гарант-Студент» (<http://www.edu.garant.ru>);
4. ЭБС «ЦНСХБ» (<http://cnshb.ru/terminal/>);

6.6. Программное обеспечение, используемое для освоения дисциплины (модуля) (при необходимости)

1. Операционная система типа Windows

2. Пакет программ общего пользования Microsoft Office
3. Интернет-браузеры
4. Учебная версия «Рационы» ИАС РЦ «Плино», Расчет кормовых рационов (Договор № 257/37 от 27.04.2017)

6.7. Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)

- 1) Лекции-презентации
- 2) LMSMoodle

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

№ п/п	Наименование специализированных аудиторий, кабинетов, лабораторий и пр.	Краткий перечень основного оборудования
1.	Учебная аудитория для проведения занятий лекционного типа	укомплектована специализированной (учебной) мебелью, набором демонстрационного оборудования и учебно-наглядными пособиями, обеспечивающими тематические иллюстрации, соответствующие рабочей программе дисциплины, а также техническими средствами обучения (стационарным мультимедийным проектором, портативным компьютером типа «Ноутбук», стационарным раздвижным экраном), служащие для представления учебной информации большой аудитории.
2.	Учебная аудитория для проведения занятий семинарского типа, для групповых и индивидуальных консультаций, для текущего контроля и промежуточной аттестации	укомплектована специализированной (учебной) мебелью, переносными техническими средствами обучения (мультимедийным проектором, портативным компьютером типа «Ноутбук», переносным раздвижным экраном, DVD-плеером, телевизором), служащими для представления учебной информации.
3.	Учебная аудитория для выполнения курсовых работ	укомплектована специализированной (учебной) мебелью, переносными техническими средствами обучения (мультимедийным проектором, портативным компьютером типа «Ноутбук», переносным раздвижным экраном, DVD-плеером, телевизором), служащими для представления учебной информации.
4.	Учебная аудитория для проведения занятий семинарского типа, для групповых и индивидуальных консультаций.	укомплектована специализированной (учебной) мебелью и лабораторным оборудованием (Электронные весы, технические весы ВЛТК-500, аналитические весы, шкафы сушильные, насос вакуумный, иономер цифровой, дистиллятор ДЭ - 4-2, фотоэлектрокалориметр, нитратомер, Аппарат Сокслетта, Аппарат Кьельдаля, муфельная печь, плитки электрические, вытяжной шкаф, комплект лабораторной химической посуды, коллекция кормов.

5.	Учебная аудитория для проведения занятий семинарского типа, для групповых и индивидуальных консультаций.	укомплектована специализированной (учебной) мебелью, переносными техническими средствами обучения (мультимедийным проектором, портативным компьютером типа «Ноутбук», переносным раздвижным экраном, телевизором), служащими для представления учебной информации и персональными компьютерами IntelPentiumCP 4 G 3220, сетью Internet.
6.	Аудитория для самостоятельной работы	укомплектовано специализированной (учебной) мебелью, оснащено компьютерной техникой (10 ПК) с возможностью подключения к сети "Интернет" и обеспечено доступом в электронную информационно-образовательную среду организации.

Приложение № 1
к рабочей программе по дисциплине
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

**«Современные методы анализа биологических данных
с основами биоинформатики»**

1.1. Очная форма:

Шифр и наименование компетенции	Индикатор(ы) достижения компетенции / планируемые результаты обучения	Форма контроля*	Оценочные средства
1	2	3	4
ПК-9 Способен формировать и решать задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации	ИД-1ПК9 Знать: Понятия и задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации ИД-2 ПК9 Уметь: использовать в практической деятельности методы решения производственных задач связанных с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации ИД-3 ПК9 Владеть: практическими навыками формирования и решения задачи в производственной деятельности связанные с реализацией и разработкой технологий отраслей в животноводстве и рационального использования средств механизации и автоматизации	УО, Т,Р, К, 3	Вопросы для подготовки к устному опрос коллоквиумам у, темы рефератов, вопросы для подготовки к зачёту

ПК-10 Способен владеть методами селекции, кормления, содержания и использования различных видов животных	ИД-1ПК10 Знать: Методы селекции, кормления, содержания и использования различных видов животных ИД-2 ПК10 Уметь применять методы селекции, кормления, содержания и использования различных видов животных ИД-3 ПК10 Владеть: Способностью владеть методами селекции, кормления, содержания и использования различных видов животных	УО, Т,Р, К, 3	Вопросы для подготовки к устному опрос коллоквиумам у, темы рефератов, вопросы для подготовки к зачёту
ПК-11 Способен применять современные методы исследований в области животноводства	ИД-1ПК11 Знать: Современные методы исследований в области животноводства ИД-2 ПК11 Уметь: Применять современные методы исследований в области животноводства ИД-3 ПК11 Владеть: Способностью применять современные методы исследований в области животноводства	УО, Т,Р, К, 3	Вопросы для подготовки к устному опрос коллоквиумам у, темы рефератов, вопросы для подготовки к зачёту
ПК-12 Способен использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных	ИД-1ПК12 Знать: Методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных ИД-2 ПК12 Уметь: Использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных ИД-3 ПК12 Владеть: Способностью использовать методы оценки качества и стандартизации кормов и продукции, сертификации племенных животных	УО, Т,Р, К, 3	Вопросы для подготовки к устному опрос коллоквиумам у, темы рефератов, вопросы для подготовки к зачёту

ПК-1 Способен выбирать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных	ИД-1ПК1 Знать: Понятия, теоретические основы выбора и соблюдения режимов содержания животных, составления рационов кормления, прогнозирования последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных ИД-2 ПК1 Уметь: Применять в практической деятельности выбирать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных ИД-3 ПК1 Владеть: Способен осознано использовать и соблюдать режимы содержания животных, составлять рационы кормления, прогнозировать последствия изменений в кормлении, разведении и содержании животных и на этом основании проводить зоотехническую оценку животных	УО, Т,Р, К, З	Вопросы для подготовки к устному опрос коллоквиумам у, темы рефератов, вопросы для подготовки к зачёту
--	--	--------------------------	---

*УО – устный опрос, КЛ – конспект лекции, КР – контрольная работа, ВЛР – выполнение лабораторной работы, ВПР – выполнение практической работы, К – коллоквиум, Т – тестирование, Р – реферат, Д – доклад, ЗКР – защита курсовой работы, ЗКП – защита курсового проекта, Э – экзамен, З – зачет.

1.3. Заочная форма:

Шифр и наименование компетенции	Индикатор(ы) достижения компетенции / планируемые результаты обучения	Форма контроля	Оценочные средства
1	2	3	4

2. Показатели и критерии оценивания сформированности компетенций на данном этапе их формирования

Показатели	Критерии оценивания*			
	неудовлетворительно не зачтено	удовлетворительно	хорошо	отлично
Полнота знаний	Уровень знаний ниже минимальных требо-	Минимально допустимый уровень зна-	Уровень знаний в объеме, соответству-	Уровень знаний в объеме, соответству-

	ваний, имели место грубые ошибки	ний, допущено много негрубых ошибок	ющем программе подготовки, допущено несколько негрубых ошибок	ющем программе подготовки, без ошибок
Наличие умений	При решении стандартных задач не продемонстрированы основные умения, имели место грубые ошибки	Продemonстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме
Наличие навыков (владение опытом)	При решении стандартных задач не продемонстрированы базовые навыки, имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов
Характеристика сформированности компетенции	Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству практических задач	Сформированность компетенции в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных практических (профессиональных) задач	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных практических (профессиональных) задач
Уровень сформированности компетенций	Низкий	Ниже среднего	Средний	Высокий

* Преподаватель вправе изменить критерии оценивания в соответствии с ФГОС ВО и особенностями ОПОП.

3. Оценочные средства

3.1. Вопросы для подготовки к устным опросам (коллоквиумам), тестированиям и зачету

3.1.1. Вопросы для подготовки к коллоквиумам по разделам

Коллоквиум № 1. Основы биоинформатики и биологические базы данных

- Выбор базы данных:** Вам необходимо найти информацию о гене DGAT1 у крупного рогатого скота: его нуклеотидную последовательность, положение в геноме, известные SNP-маркеры и ассоциации с молочной продуктивностью. Опишите пошаговый алгоритм работы с базами NCBI (Gene, dbSNP), Ensembl и Animal QTLdb. В какой базе вы найдёте каждый тип информации и почему?
- Интерпретация BLAST-поиска:** После секвенирования неизвестного фрагмента ДНК (650 п.н.) из образца мяса вы выполнили BLAST-поиск и получили результат: максимальное сходство 99% с *Equus caballus* (лошадь), E-value = 1e-150, Query Cover = 100%. Что означает каждый из этих показателей? Почему именно E-value является ключевым критерием статистической значимости совпадения? Можно ли на основании этих данных утверждать, что продукт содержит конину?

3. **Множественное выравнивание и филогенетика:** Вам необходимо установить эволюционные отношения между 5 аборигенными породами КРС России и 3 зарубежными породами. Опишите алгоритм: выбор маркера (митохондриальная ДНК vs ядерные гены), выравнивание (ClustalW, MUSCLE), построение дерева (Neighbor-Joining, Maximum Likelihood), оценка bootstrap. Как интерпретировать длину ветвей и значение bootstrap >70%?

Коллоквиум № 2. Статистические методы и программирование в R и Python

4. **Выбор статистического теста:** Вы сравниваете удои коров трёх разных пород (голштинская, черно-пестрая, холмогорская) — по 30 голов в каждой. Данные не имеют нормального распределения (проверено тестом Шапиро-Уилка). Какой статистический тест вы примените для сравнения групп: параметрический (ANOVA) или непараметрический (Краскелла-Уоллиса)? Обоснуйте выбор. Как в R (пакет `stats` или `rstatix`) выполнить этот тест и провести post-hoc сравнение?

5. **Работа с данными в R и Python:** Вам поступил CSV-файл с данными 1000 коров: ID, порода, удои, жир, белок, сервис-период, дата отела. Опишите код на R (с использованием `dplyr` из `tidyverse`) и на Python (с использованием `pandas`) для решения задачи: отфильтровать коров с удоем >8000 кг, сгруппировать по породе, рассчитать средние удои и стандартное отклонение в каждой группе, визуализировать результат с помощью `box-plot`.

6. **Корреляция и регрессия:** В хозяйстве собрана информация по 200 быкам-производителям: EBV по удою, EBV по жиру, EBV по белку. Как рассчитать матрицу корреляций (Пирсона/Спирмена) и построить корреляционную тепловую карту (heatmap) в R (`corrplot`, `ggcorrplot`) или Python (`seaborn.heatmap`)? Почему высокая корреляция между удоем и белком может создавать проблему при построении селекционного индекса?

Коллоквиум № 3. Анализ геномных данных NGS

7. **Контроль качества FASTQ:** После секвенирования 50 образцов ДНК КРС на платформе Illumina вы получили FASTQ-файлы. Отчёт FastQC показал: (а) падение качества к концу прочтений (per base sequence quality); (б) высокое содержание дубликатов (sequence duplication levels); (в) обнаружены адаптерные последовательности (adapter content). Объясните причины каждой проблемы. Опишите пайплайн предобработки: какие инструменты (Trimmomatic, fastp, fastp) и параметры вы используете?

8. **Variant Calling и фильтрация:** После выравнивания прочтений на референсный геном (BWA-MEM) и конвертации в BAM-формат (SAMtools) вы выполнили вызов вариантов с помощью GATK HaplotypeCaller. В VCF-файле обнаружено 5 миллионов SNP. Опишите алгоритм жёсткой фильтрации (hard filtering) по параметрам VQSR или стандартным порогам (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5). Зачем нужна фильтрация и как она влияет на downstream-анализ?

9. **Аннотация вариантов:** После фильтрации у вас осталось 500 000 SNP. Как с помощью SnpEff или Ensembl VEP определить, какие из них находятся в кодирующих регионах (missense, nonsense, synonymous), какие — в интронах, а какие — в межгенных пространствах?

ствах? Как отобрать 10 наиболее вероятных кандидатных мутаций, влияющих на продуктивность?

Коллоквиум № 4. Анализ транскриптомных данных (RNA-seq)

10. **Дизайн RNA-seq эксперимента:** Вы планируете изучить влияние теплового стресса на экспрессию генов в ткани вымени коров. Опишите дизайн эксперимента: сколько биологических повторностей необходимо (минимум 3, оптимум 6), какая глубина секвенирования нужна (20–30 млн прочтений на образец), длина прочтений (75–150 п.н., single-end или paired-end). Почему биологические повторности критически важны, а технические — нет?

11. **Дифференциальная экспрессия:** После количественной оценки экспрессии (featureCounts) вы загрузили матрицу counts в R-пакет DESeq2. Опишите пошаговый пайплайн: создание объекта DESeqDataSet, нормализация (median-of-ratios), запуск DESeq(), получение таблицы результатов (results), поправка на множественное тестирование (Benjamini-Hochberg FDR). Как отобрать гены с $|\log_2\text{FoldChange}| > 1$ и $\text{padj} < 0.05$?

12. **Интерпретация результатов RNA-seq:** По результатам анализа вы получили список из 500 дифференциально экспрессируемых генов (250 up-regulated, 250 down-regulated) при тепловом стрессе. Опишите, как выполнить функциональный анализ обогащения (GO, KEGG) с помощью пакета [clusterProfiler](#). Какие биологические пути (heat shock proteins, apoptosis, immune response) вы ожидаете увидеть и как это связано с физиологией теплового стресса?

Коллоквиум № 5. Многомерная статистика и машинное обучение

13. **PCA для визуализации образцов:** У вас есть матрица экспрессии 20 000 генов $\times$ 30 образцов RNA-seq (15 контрольных, 15 стрессовых). Как выполнить PCA в R ([prcomp](#), [ggplot2](#), [factoextra](#)) или Python ([sklearn.decomposition.PCA](#))? Как по PCA-plot оценить, разделяются ли группы образцов, и выявить выбросы (outliers)? Что означают PC1 и PC2 и какой процент дисперсии они объясняют?

14. **Классификация с помощью Random Forest:** Вы хотите построить модель для прогнозирования мастита у коров на основе 20 признаков: удои, SCC, активность, руминация, температура, данные датчиков, возраст, месяц отела. Опишите алгоритм обучения случайного леса в R ([randomForest](#), [ranger](#)) или Python ([sklearn.ensemble.RandomForestClassifier](#)). Как оценить важность признаков (variable importance), избежать переобучения (кросс-валидация) и интерпретировать ROC-кривую и AUC?

15. **Кластеризация и выявление подтипов:** У вас есть данные по 1000 коров: удои, жир, белок, SCC, сервис-период, продуктивное долголетие. Как с помощью k-means или иерархической кластеризации (в R: [kmeans](#), [hclust](#), [factoextra](#); в Python: [sklearn.cluster](#)) выделить 3–4 производственных типа коров? Как определить оптимальное число кластеров (метод локтя, silhouette score) и охарактеризовать каждый кластер по средним значениям признаков?

Коллоквиум № 6. Интеграция данных, визуализация и воспроизводимость

16. **Интеграция multi-omics данных:** У вас есть данные геномики (SNP-чип 50K), транскриптомики (RNA-seq печени) и фенотипики (удой, жир, белок) по 100 быкам. Как интегрировать эти данные для выявления кандидатных генов, влияющих на молочную продуктивность? Какие методы (eQTL-анализ, WGCNA, multi-omics factor analysis — MOFA) вы примените?

17. **Воспроизводимость анализа:** Вы выполнили сложный биоинформатический анализ в R, включающий 50 шагов. Как обеспечить его воспроизводимость: использование R Markdown или Jupyter notebooks, фиксация версий пакетов ([renv](#), [packrat](#)), документирование параметров, публикация кода на GitHub? Почему это требование является стандартом современной науки (принципы FAIR)?

18. **Контейнеризация и облачные вычисления:** Объясните, зачем нужны Docker и Singularity в биоинформатике. Как создать Docker-контейнер с установленными инструментами (R, Python, GATK, DESeq2) для обеспечения одинаковой вычислительной среды на любом компьютере? В каких случаях целесообразно использовать облачные платформы (Galaxy, Terra, AWS) вместо локальных вычислений?

3.1.2. Тестовые задания (с ключами ответов)

Блок 1. Основы биоинформатики и базы данных

1. **Формат FASTA используется для хранения:** а) Качества прочтений секвенатора; б) Нуклеотидных или аминокислотных последовательностей с заголовком; в) Выровненных прочтений на референсный геном; г) Генетических вариантов (SNP).
2. **Показатель E-value в BLAST-поиске означает:** а) Процент идентичности последовательностей; б) Ожидаемое число случайных совпадений с таким же или лучшим score в базе данных данного размера; в) Длину выравнивания; г) Количество несовпадений (mismatches).
3. **База данных Animal QTLdb содержит информацию о:** а) Только последовательностях геномов; б) Количественных признаков локусах (QTL), ассоциированных с хозяйственно полезными признаками сельскохозяйственных животных; в) Белковых структурах; г) Метаболитах.

Блок 2. Статистика и программирование

4. **В языке R пакет [dplyr](#) из экосистемы [tidyverse](#) используется преимущественно для:** а) Построения филогенетических деревьев; б) Манипуляции табличными данными: фильтрации, группировки, агрегации, слияния; в) Выравнивания последовательностей ДНК; г) Статистического моделирования.
5. **Непараметрический тест Краскела-Уоллиса применяется, когда:** а) Данные имеют нормальное распределение и сравниваются две группы; б) Данные не имеют нормального распределения и сравниваются три и более независимых группы; в) Данные имеют нормальное распределение и сравниваются три и более группы; г) Данные парные и имеют нормальное распределение.
6. **Библиотека [ggplot2](#) в R основана на концепции:** а) Грамматика графики (grammar of graphics), где график строится послойно; б) Базовой графики R; в) Интерактивных дашбордов; г) Трёхмерной визуализации.

Блок 3. Анализ геномных данных NGS

7. **Метрика качества Phred Q30 означает, что:** а) Вероятность ошибки при определении нуклеотида составляет 1 к 1000 (точность 99,9%); б) Вероятность ошибки составляет 1 к 100 (точность 99%); в) Вероятность ошибки составляет 1 к 10 (точность 90%); г) Прочтение имеет длину 30 п.н.
8. **Формат VCF (Variant Call Format) используется для хранения:** а) Сырых прочтений секвенатора; б) Информации о генетических вариантах (SNP, InDel) у одного или нескольких образцов; в) Выровненных прочтений; г) Экспрессии генов.
9. **Инструмент BWA-MEM применяется для:** а) Контроля качества FASTQ-файлов; б) Выравнивания (mapping) прочтений на референсный геном; в) Вызова генетических вариантов; г) Аннотации генов.

Блок 4. Анализ транскриптомных данных

10. **Метрика TPM (Transcripts Per Million) в RNA-seq используется для:** а) Оценки качества прочтений; б) Нормализованной количественной оценки экспрессии генов, позволяющей сравнивать образцы между собой; в) Выравнивания прочтений; г) Вызова SNP.
11. **Пакет DESeq2 в R предназначен для:** а) Выравнивания RNA-seq прочтений; б) Анализа дифференциальной экспрессии генов с использованием отрицательного биномиального распределения; в) Построения филогенетических деревьев; г) Контроля качества FASTQ.
12. **Volcano plot в RNA-seq анализе отображает:** а) PCA образцов; б) Зависимость статистической значимости ($-\log_{10} \text{padj}$) от величины изменения экспрессии ($\log_2 \text{FoldChange}$) для каждого гена; в) Кластеризацию генов; г) Корреляцию между образцами.

Блок 5. Многомерная статистика и машинное обучение

13. **Метод главных компонент (PCA) используется для:** а) Классификации образцов; б) Снижения размерности данных и визуализации многомерных выборок; в) Выравнивания последовательностей; г) Вызова генетических вариантов.
14. **Алгоритм Random Forest (случайный лес) представляет собой:** а) Один метод линейной регрессии; б) Ансамбль решающих деревьев, обученных на случайных подвыборках данных и признаков; в) Метод кластеризации; г) Метод снижения размерности.
15. **ROC-кривая и AUC используются для оценки качества:** а) Модели регрессии; б) Модели бинарной классификации; в) Кластеризации; г) Выравнивания последовательностей.

Блок 6. Визуализация и воспроизводимость

16. **R Markdown и Jupyter notebooks используются для:** а) Только написания кода; б) Создания интерактивных и воспроизводимых отчётов, объединяющих код, результаты и текст; в) Только визуализации данных; г) Хранения FASTQ-файлов.
17. **Docker в биоинформатике применяется для:** а) Секвенирования ДНК; б) Контейнеризации вычислительной среды для обеспечения воспроизводимости анализов на любых компьютерах; в) Статистического анализа; г) Управления базами данных.
18. **Принципы FAIR для научных данных означают:** а) Fast, Accurate, Intelligent, Reproducible; б) Findable, Accessible, Interoperable, Reusable; в) Free, Anonymous, Independent, Reliable; г) Formal, Applied, Integrated, Robust.

3.1.3. Практические кейсовые задачи

Кейс 1. Анализ и предобработка NGS-данных **Условие:** После секвенирования 20 образцов ДНК КРС на Illumina NovaSeq получены FASTQ-файлы (paired-end, 150 п.н.). Отчёт FastQC показал: среднее качество по всем образцам Q30 = 85%, но у 5 образцов обнаружено высокое содержание адаптеров (adapter content > 10%) и падение качества к концу прочтений (per base quality < Q20 после 130 п.н.). **Задание:**

1. Напишите команду для [fastp](#) или [Trimmomatic](#), которая выполнит: удаление адаптеров, обрезку прочтений по качеству (sliding window 4:20), отбраковку прочтений короче 50 п.н.
2. После предобработки выполните выравнивание с помощью [BWA-MEM](#) на референсный геном ARS-UCD1.2. Приведите команды для конвертации SAM → BAM, сортировки и индексирования (SAMtools).
3. Рассчитайте, какой процент прочтений должен быть uniquely mapped (карта на один локус) для качественных данных WGS КРС (норма >90%). Что может означать низкий процент mapping (<70%)?
4. Опишите дальнейший пайплайн вызова SNP с помощью GATK HaplotypeCaller и фильтрации (VQSR или hard filtering).

Кейс 2. Дифференциальная экспрессия генов (RNA-seq) **Условие:** Проведён RNA-seq анализ печени быков с высокой (группа H, n=6) и низкой (группа L, n=6) конверсией корма. После выравнивания (STAR) и количественной оценки (featureCounts) получена матрица raw counts (25 000 генов × 12 образцов). PCA-анализ показал чёткое разделение групп H и L по PC1 (объясняет 45% дисперсии). **Задание:**

1. Напишите код на R с использованием пакета [DESeq2](#) для: создания объекта DESeqDataSet из матрицы counts и метаданных, запуска пайплайна [DESeq\(\)](#), получения таблицы результатов с поправкой на множественное тестирование.
2. Отберите гены с $|\log_2\text{FoldChange}| > 1$ и $\text{padj} < 0.05$. Постройте volcano plot (пакет [EnhancedVolcano](#) или [ggplot2](#)) и heatmap топ-50 дифференциально экспрессируемых генов ([pheatmap](#)).
3. Выполните функциональный анализ обогащения (GO, KEGG) с помощью [clusterProfiler](#). Какие биологические пути (метаболизм липидов, синтез белка, детоксикация) вы ожидаете увидеть различающимися между группами?
4. Как валидировать результаты RNA-seq методом ПЦР в реальном времени (qPCR)? Какие гены референса (housekeeping) вы выберете?

Кейс 3. Построение предиктивной модели машинного обучения **Условие:** В хозяйстве собраны данные по 1500 коров: 30 признаков (удой за 305 дней, жир, белок, SCC, продуктивное долголетие, сервис-период, возраст при первом отёле, живая масса, данные датчиков активности и руминации за первые 60 дней лактации). Целевая переменная: «выбыла до 100 дней лактации» (1 — да, 0 — нет). В выборке 10% коров выбыли рано. **Задание:**

1. Опишите предобработку данных: обработка пропущенных значений (NA), масштабирование признаков (standardization), разделение на обучающую (70%) и тестовую (30%) выборки, стратификация по целевому классу.
2. Обучите модель случайного леса (Random Forest) в R ([ranger](#)) или Python ([sklearn.ensemble](#)). Оцените качество: accuracy, precision, recall, F1-score, AUC-ROC. Почему при несбалансированных классах (10% vs 90%) accuracy — плохая метрика?

3. Постройте график важности признаков (variable importance). Какие 5 признаков оказались наиболее значимыми для прогнозирования ранней выбраковки?
4. Как улучшить модель: подбор гиперпараметров (grid search, random search), использование SMOTE для балансировки классов, применение градиентного бустинга (XGBoost, LightGBM)?

Кейс 4. Филогенетический анализ и ДНК-штрихкодирование **Условие:** Студент получил последовательности гена COI (650 п.н.) от 10 особей разных аборигенных пород КРС России (якутская, калмыцкая, тагильская, бестужевская, аулиеатинская) и 3 зарубежных (голштинская, абердин-ангусская, герефордская). Последовательности сохранены в формате FASTA.

Задание:

1. Опишите пайплайн в MEGA или R (пакет [ape](#), [phangorn](#)): импорт FASTA, множественное выравнивание (MUSCLE), выбор модели нуклеотидных замен (Model Selection, например, Tamura-Nei), построение дерева методом Neighbor-Joining и Maximum Likelihood, bootstrap-анализ (1000 реплик).
2. Как интерпретировать полученное филогенетическое дерево? Какие породы образуют отдельные кластеры? О чём свидетельствует длина ветвей?
3. Можно ли на основе гена COI достоверно различить все 10 пород? Почему для внутривидовой филогенетики и идентификации пород часто используют микросателлиты или SNP-чипы, а не один ген?
4. Как использовать эти данные для ДНК-штрихкодирования и защиты от фальсификации продукции «от аборигенных пород»?

Кейс 5. Многомерный анализ зоотехнических данных **Условие:** В племенном хозяйстве по 500 коров голштинской и черно-пестрой пород собраны данные: удой, жир, белок, SCC, сервис-период, продолжительность лактации, живая масса, высота в холке, глубина тела, угол крестца (линейная оценка). Необходимо выявить скрытые производственные типы коров. **Задание:**

1. Опишите предобработку: масштабирование данных (функция [scale\(\)](#) в R), удаление коррелирующих признаков (корреляция $> 0,9$).
2. Выполните PCA. Сколько главных компонент необходимо для объяснения 80% дисперсии? Как по biplot (пакет [factoextra](#)) интерпретировать вклад каждого признака в PC1 и PC2?
3. Примените метод k-means кластеризации для $k = 3$. Как определить оптимальное число кластеров (метод локтя, silhouette)? Охарактеризуйте каждый кластер по средним значениям признаков.
4. Как результаты кластеризации можно использовать для дифференцированного менеджмента и кормления разных типов коров?

Кейс 6. Интеграция данных и воспроизводимость анализа **Условие:** Вы выполнили сложный биоинформатический анализ: RNA-seq (DESeq2) + GWAS + функциональная аннотация (GO, KEGG). Весь анализ занял 2 недели и включает 50 шагов в R. Заказчик (племенное хозяйство) просит предоставить не только результаты, но и возможность повторить анализ на новых данных. **Задание:**

1. Опишите, как перевести весь анализ в формат R Markdown или Jupyter notebook, чтобы код, результаты и текст были в одном документе. Какие разделы должен содержать отчёт?
2. Как обеспечить воспроизводимость среды: фиксация версий пакетов ([renv::snapshot\(\)](#)), описание зависимостей, использование Docker-контейнера?

3. Как опубликовать код на GitHub: структура репозитория (README, data/, code/, results/), лицензия (MIT, GPL), документация.
4. Какие принципы FAIR-данных (Findable, Accessible, Interoperable, Reusable) вы соблюдёте при публикации результатов и кода? Как это повышает цитируемость и доверие к исследованию?

Кейс 7. Анализ обогащения GO-терминов и сетей взаимодействий **Условие:** По результатам RNA-seq (Кейс 2) получен список из 300 дифференциально экспрессируемых генов (150 up, 150 down). Необходимо понять, какие биологические процессы и молекулярные функции затронуты. **Задание:**

1. Опишите алгоритм анализа обогащения GO-терминов с помощью `clusterProfiler` в R: функции `enrichGO()`, `gseGO()` (для gene set enrichment analysis). Как выбрать background (фон) — все гены генома или только экспрессируемые?
2. Постройте dotplot и cnetplot (concept network) для топ-10 значимых GO-терминов. Как интерпретировать размер точек (gene ratio) и цвет (p-value)?
3. Выполните анализ KEGG pathways. Какие пути (например, PPAR signaling, fatty acid biosynthesis) вы ожидаете увидеть при различиях в конверсии корма?
4. Загрузите список генов в базу STRING и постройте сеть белок-белковых взаимодействий (PPI). Как выявить ключевые гены-хабы (hubs) с высокой степенью связности (degree)?

3.1.4. Полный перечень вопросов к зачету

Раздел 1. Основы биоинформатики и биологические базы данных

1. Предмет, цели и задачи биоинформатики. Междисциплинарный характер дисциплины.
2. Архитектура международных биологических баз данных: NCBI, EMBL-EBI, DDBJ.
3. Специализированные базы данных по сельскохозяйственным животным: Ensembl, Animal QTLdb, FAANG.
4. Форматы хранения данных: FASTA, GenBank, GFF/GTF.
5. Алгоритм BLAST: принципы поиска, интерпретация E-value, bit score.
6. Множественное выравнивание последовательностей: ClustalW, MUSCLE, MAFFT.
7. Основы молекулярной филогенетики: методы построения деревьев (NJ, ML, Bayesian).

Раздел 2. Статистические методы и программирование 8. Дескриптивная статистика для биологов: меры центральной тенденции, дисперсии. 9. Проверка статистических гипотез: t-тест, ANOVA, хи-квадрат, непараметрические тесты. 10. Корреляционный и регрессионный анализ в биологии. 11. Введение в язык R: синтаксис, структуры данных, пакеты tidyverse. 12. Введение в Python для биологов: pandas, numpy, matplotlib. 13. Работа с табличными данными: импорт, фильтрация, группировка, слияние. 14. Визуализация данных: ggplot2 (R), matplotlib/seaborn (Python).

Раздел 3. Анализ геномных данных NGS 15. Основы высокопроизводительного секвенирования: платформы Illumina, Ion Torrent, PacBio. 16. Форматы данных NGS: FASTQ, метрики качества Phred. 17. Контроль качества NGS-данных: FastQC, Trimmomatic, fastp. 18. Выравнивание прочтений на референсный геном: BWA-MEM, Bowtie2, STAR. 19. Форматы

SAM/BAM: структура, конвертация, SAMtools. 20. Вызов генетических вариантов: GATK, bcftools, FreeBayes. 21. Формат VCF: структура, фильтрация SNP и InDel. 22. Аннотация генетических вариантов: SnpEff, VEP.

Раздел 4. Анализ транскриптомных данных 23. Дизайн эксперимента RNA-seq: повторности, глубина секвенирования. 24. Количественная оценка экспрессии: featureCounts, HTSeq, TPM, FPKM. 25. Нормализация данных RNA-seq: методы DESeq2, edgeR. 26. Дифференциальная экспрессия генов: DESeq2, edgeR, limma-voom. 27. Визуализация результатов RNA-seq: volcano plot, heatmap, PCA-plot. 28. Функциональная аннотация и анализ обогащения: GO, KEGG, clusterProfiler. 29. Анализ белок-белковых взаимодействий: база STRING.

Раздел 5. Многомерная статистика и машинное обучение 30. Метод главных компонент (PCA): математические основы, интерпретация. 31. Нелинейные методы снижения размерности: t-SNE, UMAP. 32. Кластеризация: k-means, иерархическая кластеризация, DBSCAN. 33. Основы машинного обучения: обучение с учителем vs без учителя. 34. Алгоритмы классификации: логистическая регрессия, SVM, Random Forest, k-NN. 35. Оценка качества моделей: accuracy, precision, recall, F1-score, ROC-кривая, AUC. 36. Применение машинного обучения в зоотехнии: прогнозирование продуктивности и здоровья.

Раздел 6. Интеграция данных, визуализация и воспроизводимость 37. Интеграция геномных, транскриптомных и фенотипических данных. 38. Основы системной биологии: сети регуляции генов, метаболические пути. 39. Воспроизводимость исследований: принципы FAIR-данных. 40. Интерактивные отчёты: Jupyter notebooks, R Markdown, Quarto. 41. Системы контроля версий: Git, GitHub/GitLab. 42. Контейнеризация: Docker, Singularity. 43. Облачные вычисления: Galaxy, Terra, AWS. 44. Анализ single-cell RNA-seq: технологии, пакеты Seurat, Scanpy. 45. Метагеномика и анализ микробиома: 16S rRNA, QIIME2. 46. Искусственный интеллект и глубокое обучение в биологии: нейронные сети, AlphaFold2. 47. Биоинформатика в селекции: интеграция с геномной селекцией, расчёт GEBV. 48. Цифровые двойники животных: концепция, источники данных, предиктивные модели. 49. Этика и правовые аспекты работы с биологическими данными. 50. Современные тренды и перспективы биоинформатики: пангеномика, multi-omics.

Методическая рекомендация для проведения зачета:

Для комплексной проверки компетенций рекомендуется использовать **комплексную форму зачета**, включающую:

- | | |
|----|---|
| 1. | Устный ответ по теоретическому вопросу (из Разделов 1–4) — 5 минут. |
| 2. | Решение практической задачи на компьютере (работа в R или Python: анализ данных, построение графика, запуск статистической модели) — 20 минут. |
| 3. | Защита мини-проекта (по желанию студента или как обязательный элемент) — 7 минут |

3.6. Методические материалы

Устные опросы проводятся на усмотрение преподавателя в начале практического занятия в течении 10-15 минут и оцениваются по четырёх бальной шкале.

Порядок проведения коллоквиумов, устных опросов, практических задач дан в Приложении № 2 к положению ПВД-07 «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся». Контроль за успеваемостью обучающихся осуществляется в соответствии с ПВД-20 «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся. Условия и порядок проведения зачета даны в Приложении № 2 к положению ПВД-07 «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»