

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО
решением Ученого совета ННГУ
протокол № 10 от 02.12.2024 г.

Рабочая программа дисциплины

Введение в биоинформатику

Уровень высшего образования
Бакалавриат

Направление подготовки / специальность
02.03.02 - Фундаментальная информатика и информационные технологии

Направленность образовательной программы
Инженерия программного обеспечения

Форма обучения
очная

г. Нижний Новгород

2025 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.02.02 Введение в биоинформатику относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-3: Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники	<p>ПК-3.1: Знает методы анализа и исследования математических моделей в области фундаментальной информатики и информационных технологий</p> <p>ПК-3.2: Умеет определять ключевые свойства и ограничения системы</p>	<p>ПК-3.1: Знает фундаментальные и теоретические основы, необходимые для исследования научных проблем по предметной области курса.</p> <p>ПК-3.2: Умеет определять и профессионально реализовывать необходимые для решения прикладных задач нелинейной динамики вычислительные алгоритмы, анализировать полученные результаты.</p>	<p>Задачи</p> <p>Собеседование</p>	<p>Зачёт:</p> <p>Контрольные вопросы</p>

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	3
Часов по учебному плану	108
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	16
- занятия семинарского типа (практические занятия / лабораторные работы)	16
- КСР	1
самостоятельная работа	75
Промежуточная аттестация	0

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
0 0	0 0	0 0	0 0	0 0	
Биоинформатика: определение, цели и задачи.	19	2	2	4	15
Подходы и области применения биоинформатики. Биологические базы данных.	21	3	3	6	15
Общая характеристика живой клетки: типы биомолекул, основные компоненты, функции.	21	3	3	6	15
Компьютерные инструменты анализа биомолекул.	23	4	4	8	15
Молекулярная эволюция: вычислительные подходы	23	4	4	8	15
Аттестация	0				
КСР	1			1	
Итого	108	16	16	33	75

Содержание разделов и тем дисциплины

Биоинформатика: определение, цели и задачи.

Подходы и области применения биоинформатики. Биологические базы данных.

Общая характеристика живой клетки: типы биомолекул, основные компоненты, функции.

Компьютерные инструменты анализа биомолекул.

Молекулярная эволюция: вычислительные подходы.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для обеспечения самостоятельной работы обучающихся используются:

Электронные курсы, созданные в системе электронного обучения ННГУ:

Введение в биоинформатику (Иванченко М.В., Юсипов И.И.),

<https://e-learning.unn.ru/course/view?id=5412>.

Иные учебно-методические материалы:

Гергель В.П. Высокопроизводительные вычисления для многоядерных многопроцессорных систем. Учебное пособие – Нижний Новгород; Изд-во ННГУ им. Н.И.Лобачевского, 2010. (47 экз)

Лаптева Т.В., Иванченко М.В. Математические модели генной регуляции: Учебно-методическое пособие. Нижний Новгород: Нижегородский госуниверситет, 2017. 24 с. Рег. № 1646.17.06 http://www.unn.ru/books/met_files/LI.pdf

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-3:

Задание 1

Разработать приложение, выполняющее поиск обратно-комплементарной цепи ДНК.

Входные параметры приложения: цепь ДНК.

Выходные параметры приложения: цепь ДНК, обратно-комплементарная к данной.

Задание 2

Разработать приложение, выполняющее поиск стартовых позиций паттерна в геноме.

Входные параметры приложения: цепь ДНК, паттерн.

Выходные параметры приложения: массив стартовых позиций паттерна в цепи ДНК.

Задание 3

Разработать приложение, выполняющее подсчет числа вхождений паттерна в геном.

Входные параметры приложения: цепь ДНК, паттерн.

Выходные параметры приложения: число вхождений паттерна в геном.

Задание 4

Разработать приложение, выполняющее подсчет числа вхождений паттерна в геном с учетом возможных допустимых мутаций паттерна.

Входные параметры приложения: цепь ДНК, паттерн, число допустимых мутаций.

Выходные параметры приложения: число вхождений паттерна в геном.

Задание 5

Разработать приложение, выполняющее поиск наиболее часто встречаемых паттернов в геноме.

Входные параметры приложения: цепь ДНК.

Выходные параметры приложения: паттерн (или массив паттернов), имеющий (имеющие) наибольшее число вхождений в геном.

Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Выполнена основная часть задания, возможно с незначительными недочетами
не зачтено	Выполнено менее половины задания, есть существенные недочеты

5.1.2 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции ПК-3:

1. Биоинформатика, определение и задачи. Мультидисциплинарность, фундаментальные и прикладные аспекты.
2. Мишени новых лекарств: поиск и валидация.
3. Парное выравнивание последовательностей. Алгоритмы Смита-Уотермана и Нидлмана-Вунша (локальное и глобальное выравнивание). Матрицы замен.
4. Структурная организация белка, база данных PDB.
5. Теория распознавания образов, применение в биоинформатике. Обучение с учителем.
6. Множественное выравнивание последовательностей. Матрицы замен.
7. Предсказание вторичной структуры РНК и ДНК.

8. Анализ филогений нуклеотидных и аминокислотных последовательностей, методы построения филогенетических деревьев. Понятия SNP и SAP.
9. Алгоритм BLAST. Модификации и области применения.
10. Первичная аннотация геномов. Поиск кодирующих фрагментов.
11. Предсказание пространственной структуры белка.
12. Методы сравнительной геномики, предсказание функций и регуляции генов.
13. Молекулярный докинг. Принципы и области применения.
14. Протеомные технологии. Принципы, области применения возможности и перспективы.
15. Понятие марковского процесса. Марковская цепь. Применение в биоинформатике.
16. Построение пространственных моделей белков и низкомолекулярных соединений.
17. Масс-спектрометрия. Принципы работы, типы масс-спектрометров. Идентификация пептидов и белков.
18. Динамическое программирование. Применение в задачах выравнивания последовательностей.
19. Структурно-функциональные мотивы как молекулярная основа сходства и различия белков
20. Базы данных нуклеотидных и аминокислотных последовательностей. Форматы представления данных.
22. Проект "Геном человека".
23. Компьютерное конструирование лекарств на основе структуры белка-мишени.
24. Компьютерное конструирование лекарств на основе структуры лигандов. База данных PubChem.
27. Белок-белковые взаимодействия, способы регистрации.
28. Регуляция экспрессии генов в прокариотах и эукариотов.

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Студент дал развернутый ответ на все вопросы без существенных ошибок.
не зачтено	При ответе студент допускает грубые ошибки в основном материале.

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше

		предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-3

1. Дайте определение биоинформатики. Цели и задачи биоинформатики. Приведите пример подходов и областей применения.
2. Какова общая характеристика живой клетки. Какие типы биомолекул присутствуют в клетке. Перечислите основные компоненты клетки.
3. Опишите строение белков и основные свойства белков.
4. Какие компьютерные инструменты анализа свойств белков Вы знаете.
5. Перечислите форматы записи первичной структуры белков.
6. Приведите примеры баз данных белков.
7. Перечислите основы структур баз данных (записи, поля, объекты), классификации баз по способу заполнения (автоматические, архивные, курируемые). Дайте характеристику основных баз данных: GenBank, EMBL, SwissProt, TrEMBL, PIR, PDB.
8. Приведите пример баз данных, содержащие результаты глобальных экспериментов по анализу экспрессии, протеомике, и т.п.
9. Какие программы расчета физико-химических свойств белков Вы знаете.
10. Домены: строение и свойства.

11. Перечислите основные требования к банкам данных белковых семейств (SCOP, Prosite, ProDom, PFAM, InterPro).
12. Пространственная структура белков – методы и программные средства для анализа.
13. Методы получения трехмерных структур белков.
14. Инструменты визуализации 3D структур белков.
15. Дайте основные характеристики базы данных 3D структур белков PDB.
16. Перечислите основные функции белков.
17. Как происходит аннотирование функций в базах данных.
18. База данных GO (Gene Ontology).
19. Сравнение последовательностей.
20. Выравнивание двух последовательностей.
21. Глобальное и локальное выравнивание, вес выравнивания, матрицы аминокислотных замен.
22. Дот-матрицы.
23. Глобальное выравнивание: алгоритм Нидельмана-Вунша.
24. Локальное выравнивание: алгоритм Смита-Ватермана.
25. Другие алгоритмы локального выравнивания.
26. Статистическая значимость выравниваний и ее зависимость от вероятностной модели последовательности (в т.ч. сегменты малой сложности)
27. Зависимость выравнивания от параметров.
28. Множественное выравнивание.
29. Динамическое программирование.
30. Последовательное выравнивание (Clustal).
31. Поиск по сходству в базах данных.
32. Алгоритм Smith-Waterman.
33. Алгоритм BLAST.
34. Алгоритм FASTA.
35. Статистическая оценка значимости (E-value, P-value).

36. Марковские модели (HMM, Hidden Markov models).
37. Позиция-специфичные матрицы расчета (Position-specific scoring matrices, PSSM). Паттерны, Sequence logos, Psi-BLAST.
38. Анализ структурных особенностей белков.
39. Моделирование белков.
40. Предсказание структур белков разных уровней организации.
41. Программное обеспечение для компьютерного анализа структур белков.
42. Белок-белковые взаимодействия.
43. Базы данных белок-белковых взаимодействий (IntAct, BIND).
44. Программы анализа белковых комплексов.
45. Компьютерный расчет областей белок-белкового контакта.
46. Компьютерный расчет поверхности белка, доступной для растворителя.
47. Нуклеиновые кислоты. Строение и свойства. Структура ДНК.
48. Генетический код и его свойства.
49. Базы данных нуклеотидных последовательностей (TrEMBL, EMBL).
50. Репликация ДНК и биосинтез белка.
51. Компьютерное предсказание белков на основе данных о нуклеотидной последовательности.
52. РНК. Транскриптомика.
53. Молекулярная эволюция.
54. Эволюция молекул и организмов.
55. Ортологи и паралоги.
56. Горизонтальный перенос.
57. Предсказание структур белков разных уровней организации.
58. Программное обеспечение для компьютерного анализа структур белков.
59. Белок-белковые взаимодействия.
60. Базы данных белок-белковых взаимодействий (IntAct, BIND).

61. Программы анализа белковых комплексов.
62. Компьютерный расчет областей белок-белкового контакта.
63. Компьютерный расчет поверхности белка, доступной для растворителя.
64. Нуклеиновые кислоты. Строение и свойства. Структура ДНК.
65. Генетический код и его свойства.
66. Базы данных нуклеотидных последовательностей (TrEMBL, EMBL).
67. Репликация ДНК и биосинтез белка.
68. Компьютерное предсказание белков на основе данных о нуклеотидной последовательности.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Студент ответил на большую часть вопросов возможно с незначительными недочетами.
не зачтено	При ответе студент допускает грубые ошибки в основном материале и решении стандартных задач.

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

Основная литература:

1. Лаптева Т. В. Математические модели генной регуляции : учебно-методическое пособие / Лаптева Т. В., Иванченко М. В. - Нижний Новгород : ННГУ им. Н. И. Лобачевского, 2017. - 24 с. - Рекомендовано методической комиссией института информационных технологий, математики и механики для студентов ННГУ, обучающихся по направлению подготовки 01.03.02 и 01.04.02 "Прикладная математика и информатика". - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции ННГУ им. Н. И. Лобачевского - Математика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=730016&idb=0>.
2. Асанов М. О. Дискретная математика: графы, матроиды, алгоритмы : учебное пособие / Асанов М. О., Баранский В. А., Расин В. В. - 3-е изд., стер. - Санкт-Петербург : Лань, 2020. - 364 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции Лань - Математика. - ISBN 978-5-8114-4998-9., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=707849&idb=0>.

Дополнительная литература:

1. Гергель Виктор Павлович. Высокопроизводительные вычисления для многопроцессорных многоядерных систем : учеб. для студентов вузов, обучающихся по направлениям ВПО 010400 "Прикладная математика и информатика" и 010300 "Фундам. информатика и информ. технологии"

/ Б-ка ННГУ. - М. : Изд-во Моск. ун-та, 2010. - 544 с. : ил. - (Суперкомпьютерное образование). - На обл. кн.: Суперкомпьютерный консорциум университетов России. - ISBN 978-5-211-05937-5 : 220.00., 49 экз.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. Курс «Алгоритмы биоинформатики» Н.И.Вяххи: <https://www.lektorium.tv/course/22933>
2. <http://www.fbb.msu.ru/> - факультет биоинженерии и биоинформатики МГУ
3. <http://bioinf.spbau.ru/about> - лаборатория Алгоритмической биологии, Санкт-Петербургский академический университет РАН.
4. http://www.bionet.nsc.ru/kib/?page_id=28 - кафедра информационной биологии, НГУ
5. <http://biodynamics.ucsd.edu/> - лаборатория Биодинамики Университета Калифорнии в Сан-Диего.
6. <http://bioinformatics.ru/> - российский портал по биоинформатике
7. Библиотека масс-спектров для идентификации белков, алгоритм X!Tandem (The GPM) <http://www.thegpm.org/TANDEM/instructions.html> (в свободном доступе)
8. Программа DAVID (The Database for Annotation, Visualization and Integrated Discovery) (в свободном доступе)
9. Платформа Cytoscape (в свободном доступе)

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 02.03.02 - Фундаментальная информатика и информационные технологии.

Автор(ы): Иванченко Михаил Васильевич, доктор физико-математических наук, доцент.

Заведующий кафедрой: Иванченко Михаил Васильевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 02.12.2024, протокол № 5.