

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Высшая школа управления и инноваций



УТВЕРЖДАЮ
и.о.декана
/В.В.Печковская /
«12» февраля 2019 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ВВЕДЕНИЕ В БИОИНЖЕНЕРИЮ И БИОИНФОРМАТИКУ

БАКАЛАВРИАТ

27.03.05 "ИННОВАТИКА"

Форма обучения:

очная

Рабочая программа рассмотрена и одобрена
Советом факультета

(протокол № 2, 12 февраля 2019 г.)

Москва 2019

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки / специальности 38.04.02 «Менеджмент» (программы магистратуры) в редакции приказа МГУ от 30 декабря 2016 г.

Год (годы) приема на обучение: 2016, 2017, 2018, 2019.

1. Цели и задачи освоения дисциплины

Цели. В результате освоения дисциплины обучающийся должен знать основные алгоритмы и понятия, используемые при работе с последовательностями и структурами биополимеров, а также основные понятия биоинженерии.

Задачи. Получение базовых теоретических знаний в геномной и структурной биоинформатике и биоинженерии. Умение использовать полученные базовые знания.

2. Место дисциплины в структуре образовательной программы

1. Дисциплина является обязательной.
2. Вариативная часть, естественно-научный блок.
3. Логическая и содержательно-методическая взаимосвязь с другими частями ОПОП (дисциплинами, модулями, практиками) состоит в следующем:
 - 3.1. Перед началом освоения дисциплины «Введение в биоинформатику и биоинженерию» студент должен изучить следующие дисциплины: «Высшая математика», «Химия», «Физика», «Биология», «Информационные технологии и компьютерное моделирование».
 - 3.2. Освоение данной дисциплины необходимо как предшествующее для следующих дисциплин и практикумов: «Биотехнологии с основами микробиологии».

3. Результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников, формируемые (полностью или частично) при реализации дисциплины

Способность осуществлять поиск, критический анализ и синтез информации (УК-1).

Способность в контексте профессиональной деятельности использовать знания об основных понятиях, объектах изучения и методах естествознания (УК-5).

Способность применять знания математики, физики и естествознания, химии и материаловедения, теории управления и информационные технологии в инновационной деятельности (ОПК-3).

Способность использовать информационно-коммуникационные технологии, управлять информацией с использованием прикладных программ деловой сферы деятельности, использовать компьютерные технологии и базы данных, пакеты прикладных программ управления проектами (ОПК-6).

Способность провести необходимый эксперимент (исследование) и оценить адекватность полученных результатов (ПК-7).

Способность использовать информационные технологии и инструментальные средства (пакеты прикладных программ) при разработке инновационных проектов (ПК-13).

Планируемые результаты обучения по дисциплине

ЗНАТЬ: современные проблемы биоинформатики; основные алгоритмы и методы, применяемые для анализа последовательностей и структур биомакромолекул; основные понятия биоинженерии

УМЕТЬ: применять биоинформатические и биоинженерные подходы при анализе живых систем

4. Формат обучения – очный.

5. Объем дисциплины

Общая трудоемкость дисциплины составляет 2 зачётных единицы (72 часа).

Вид работы	Семестр
	4
Общая трудоёмкость, акад. часов	72
Аудиторная работа:	32
Лекции, акад. часов	12
Семинары, акад. часов	20
Лабораторные работы, акад. часов	0
Самостоятельная работа, акад. часов	40
Вид итогового контроля (зачёт, зачёт с оценкой, экзамен)	Экзамен

6. Содержание и структура дисциплины:

Наименование и краткое содержание разделов и тем дисциплины Форма промежуточной аттестации по дисциплине	Всего (часы)	В том числе			
		Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, часы			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа	Всего	
<u>Введение. Предмет и задачи биоинформатики и биоинженерии</u>	4	1	1	2	2
<u>Тема 1. Элементы теории информации. Биологические базы данных</u>	7	1	2	3	4 Домашнее задание
<u>Тема 2. Методы секвенирования ДНК</u>	7	1	2	3	4 Домашнее задание
<u>Тема 3. Сравнение последовательностей</u>	7	1	2	3	4 Домашнее задание

<u>Тема 4. Множественные выравнивания и филогенетические деревья</u>	7	1	2	3	4 Домашнее задание
<u>Тема 5. Приложения методов секвенирования. Геномика</u>	4	1	1	2	2 Домашнее задание
<u>Тема 6. Основные вопросы генной инженерии</u>	4	1	1	2	2 Домашнее задание
<u>Тема 7. Основы структурной биологии. Структуры биополимеров</u>	7	1	2	3	4 Домашнее задание
<u>Тема 8. Объекты и методы структурной биоинформатики</u>	7	1	2	3	4 Домашнее задание
<u>Тема 9. Предсказание структуры белков</u>	7	1	2	3	4 Домашнее задание
<u>Тема 10. Методы исследования конформационного пространства биополимеров</u>	4	1	1	2	2 Домашнее задание
<u>Тема 11. Предсказание свойств низкомолекулярных соединений</u>	7	1	2	3	4
Итого:	72	12	20	32	40

6.1. Программа дисциплины

Введение

Задачи биоинформатики и биоинженерии. Поиск парных подобий нуклеотидных и аминокислотных последовательностей. Множественное выравнивание нуклеотидных и аминокислотных последовательностей. Проблема аннотации биологических последовательностей и поиска генов. Поиск регуляторных сигналов. Генетические сети и моделирование генетической регуляции клетки. Структурная биоинформатика. Базы данных в биоинформатике. Геномика, метаболомика и протеомика. Новые направления в биоинформатике. Методы и приложения генной и тканевой инженерии.

Тема 1. Элементы теории информации

Понятие информации. Способы измерения информации. Теория информации. Сжатие информации. Теорема Котельникова. Теорема Шеннона-Хартли. Шифрование информации. Хранение информации. Источники больших данных в биомедицине. Проблемы передачи больших данных. Типы баз данных. Биологические БД Реферативные базы данных и поиск научной литературы. Базы данных нуклеотидных последовательностей (в том числе GenBank, RefSeq). Базы данных белковых последовательностей. Базы данных трехмерных структур. Базы данных химических соединений. Геномные базы данных. Базы данных геном-фенотип. БД взаимодействий,

сигнальных путей. БД результатов секвенирования. БД заболеваний и медицинской информации. БД по экспрессии генов/гистологии. Таксономические БД.

Тема 2. Методы секвенирования ДНК

Методы секвенирования ДНК. Проект “Геном человека”. Общие принципы строения ДНК. Секвенирование по Сэнгеру. Принцип методов, основанных на терминации полимеразной реакции. Методы секвенирования второго поколения: Illumina, 454 Roche, SOLiD, полупроводниковое секвенирование. Секвенирование “третьего поколения”: Oxford Nanopore, PacBio. Ошибки секвенирования. Сравнение технологий секвенирования.

Тема 3. Сравнение последовательностей

Сравнение последовательностей. Точечная матрица сходства. Расстояние по Хэммингу и Левенштейну. Матрицы замен (PAM, Blosum, JTT и др.). Динамическое программирование, локальное и глобальное выравнивание, алгоритмы Нидлмана-Вунша и Смита-Уотермана.

Тема 4. Множественные выравнивания и Филогенетические деревья

Экспресс-методы сравнения последовательностей. Программа Blast. Значимость выравнивания. Множественное выравнивание аминокислотных и нуклеотидных последовательностей, динамическое программирование и его недостатки. Визуализация выравниваний: профили. Связь между числом мутаций и числом консервативных позиций в множественном выравнивании. Методы кластеризации (UPGMA, NJ и др.) и филогенетические деревья. Прогрессивное выравнивание - программа Clustal. Итеративное выравнивание - программа Muscle. Модели эволюции. Филогенетические деревья: переменная скорость эволюции и притяжение длинных ветвей, методы проверки. Скобочная формула.

Тема 5. Приложения методов секвенирования. Геномика

Методы Hi-C, ChipSeq, MNase, CLIP и т.п. Картирование ридов и алгоритмы, лежащие в его основе. Понятие геномики, структурная геномика, функциональная геномика, фармакогеномика, популяционная геномика, сравнительная геномика, эволюционная геномика. Структуры эукариотических и прокариотических геномов. Геном человека. Генетические вариации. Понятие гаплотипа и гаплогруппы. Геномные проекты: геном человека, 1000 Genomes, Epigenomics Roadmap, ENCODE, 4D nucleome, TCGA, ICGC. GWAS исследования, связь генотип-фенотип. Заболевания, наследуемые по Менделю и комплексные заболевания. Базы данных dbSNP, OMIM, ClinVar. Геномные браузеры. Демонстрация браузера Ensembl.

Тема 6. Основные вопросы генной инженерии

Получение изолированного гена. Введение гена в вектор для переноса в организм. Перенос вектора с геном в модифицируемый организм. Преобразование клеток организма. Отбор генетически модифицированных организмов (ГМО). Генная инженерия человека.

Тема 7. Основы структурной биологии. Структура биополимеров

Основы структурной биологии: рентгеноструктурный анализ биомолекул. Формат структурных файлов. Представление структуры белков и НК во внутренних координатах. Карты Рамачандрана. Другие примеры применения рентгеновского излучения в биологии. Структурная организация белков, нуклеиновых кислот, углеводов и т.д.

Тема 8. Объекты и методы структурной биоинформатики

Структурная биоинформатика: объекты и методы. Уровни структурной организации белков и НК. Типы вторичной структуры белков и НК. Структура РНК и теория узлов. Разнообразие структур биомолекул. Базы данных структур. Неупорядоченные белки. Структурные выравнивания: задание поворота в пространстве (матрицы, углы Эйлера и кватернионы), среднеквадратичное отклонение, методы минимизации функции многих переменных (наискорейший спуск и сопряженные градиенты). Матрицы расстояний. Эволюция и консервативность структуры белков, классификация структур белков. Структурная геномика.

Тема 9. Предсказание структуры белков

Предсказание вторичной структуры и искусственные нейронные сети. Программы PSIPRED и SABLE. Гидрофобность: энтропийная природа и методы определения. Профили гидрофобности и предсказание топологии белка. Моделирование структуры на основании гомологии: программы SwissModel и Modeller. RosettaCM. Библиотеки ротамеров. Веб-сервис ROSIE как пример универсальной платформы для моделирования. Алгоритмы распознавания пространственной укладки белка (фолда), протягивание. Радиальные функции распределения как способ оценки качества упаковки пространственных структур. Веб-сервис I-TASSER. Лейциновая застежка и предсказание суперспиралей, программа LOGICOIL. CASP. FoldIt. Прионные белки. AlphaFold.

Тема 10. Методы исследования конформационного пространства биополимеров

Молекулярная динамика: общая концепция, силовые поля, возможности и ограничения. Термостат, баростат. Метод Монте-Карло. Алгоритм Метрополиса.

Тема 11. Предсказание свойств низкомолекулярных соединений

Смежные вопросы хемоинформатики. Представление химических соединений: молекулярный граф, матрица смежности, линейные нотации SMILES, InChI, структурные файлы. Количественные отношения “структура-активность” (QSAR) и способы их предсказания. Обучающая и тестовая выборки. Метод наименьших квадратов и линейная регрессия. Методы проверки зависимостей: кросс-валидация, рандомизация. Молекулярные дескрипторы. Меры сходства выборок (коэффициент Жаккара, коэффициент Танимото). Предсказание биологических свойств. Веб-сервис PASS. Восприятие вкуса.

6.2. Методические указания для обеспечения самостоятельной работы студентов

Рекомендации по видам самостоятельной работы студентов.

Для овладения знаниями:

- чтение текста (учебника, первоисточника, дополнительной литературы);
- составление плана текста;
- графическое изображение структуры текста;
- конспектирование текста;
- выписки из текста

Для закрепления и систематизации знаний:

- работа с конспектом лекции (обработка текста);
- повторная работа над учебным материалом (учебника, первоисточника, дополнительной литературы);
- составление плана и тезисов ответа;
- составление таблиц для систематизации учебного материала;
- ответы на контрольные вопросы;

Для формирования умений:

- решение задач и упражнений по образцу;
- решение вариантных задач и упражнений.

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине

7.1. Критерии оценивания

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Введение в биоинформатику				
Оценка РО и соответ- ствующие виды	2	3	4	5

оценочных средств				
Знания (виды оценочных средств: устные и письменные опросы и контрольные работы)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания

7.2. Этапы формирования компетенций

Компетенции выпускников	Этап формирования	Типовые материалы для оценки знаний, умений, навыков
УК-1	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)
УК-5	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)
ОПК-3	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)
ОПК-6	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)
ПК-7	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)
ПК-13	частичное	См. типовые вопросы для текущего контроля успеваемости (п.п. 7.3.), типовые вопросы для проведения промежуточной аттестации (п.п. 7.5.)

7.3. Типовые вопросы для текущего контроля успеваемости

Образцы вопросов устного опроса и домашних заданий:

1. Записать в шестнадцатиричной системе двоичное число.
2. Рассчитать количество информации, необходимой для кодирования знаков в алфавите.
3. Найти публикацию по doi. Определить число процитированных в ней и число цитирующих её публикаций.
4. Найти последовательность белка по её коду в БД UniProt.
5. Посчитать долю нуклеотидов в заданной последовательности.
6. Построить разумное выравнивание и определите расстояние по Левенштейну между словами БИОИНФОРМАТИКА и КОНФОРМАЦИЯ.
7. Исходя из предположения о равновероятности замен, найти математическое ожидание числа консервативных позиций в выравнивании длиной 20, содержащем 10 замен.
8. Определить, какая аминокислотная замена (N \leftrightarrow D или L \leftrightarrow I) более вероятна по данным матрицы замен BLOSUM62?

9. Вероятность ошибочного прочтения нуклеотида в последовательности составила 0,32. Рассчитать качество прочтения этого нуклеотида.
10. В некой позиции множественного выравнивания имеются набор аминокислот AVAVLLLVLL. Рассчитайте высоту букв в этой позиции при построении профиля выравнивания.
11. По заданной матрице расстояний построить филогенетическое дерево методом UPGMA и представить его запись в формате Newick.
12. Во сколько раз число укоренённых деревьев с 12 листьями больше числа неукоренённых с таким же числом листьев?
13. Почему рентгеноструктурный анализ, как правило, не позволяет определить положение атомов водорода в молекулах белков?
14. Рассчитать координаты на плоскости вектора v_2 , получаемого из вектора v_1 путем поворота, описываемого матрицей M .
15. Для двух заданных пар точек на плоскости найти оптимальное соответствие и рассчитать среднеквадратичное отклонение при совмещении.
16. Построить профиль гидрофобности для заданной последовательности белка и определить вероятное число трансмембранных фрагментов.
17. Рассчитайте, на какое расстояние сместится атом массой m под действием силы F за время t .
18. Выберите запись в формате SMILES, соответствующую молекуле на рисунке
19. Среди веществ, чьи формулы приведены, выберите обладающие сладким вкусом.
20. Какое число акцепторов водородных связей содержится в молекуле с заданной структурной формулой?
21. Рассчитайте коэффициент Танимото по заданной библиотеке фрагментов для молекул с заданными структурными формулами.
22. Оцените температуру кипения (по шкале Цельсия) нормального алкана $C_{40}H_{82}$.

Образцы вопросов контрольных работ:

Не предусмотрены

7.4. Примерные темы докладов:

Не предусмотрены

7.5. Типовые вопросы для проведения промежуточной аттестации:

1. Однобуквенные обозначение нуклеотидов и аминокислот.
2. Биоинформатика, общее представление о дисциплине
3. Информация: измерение, передача, хранение
4. Биологические базы данных (классификация и примеры)
5. Меры сходства последовательностей. Точечная матрица сходства.
6. Матрицы замен: происхождение и применение
7. Метод динамического программирования. Алгоритм Нидлмана-Вунша и его вариации
8. Экспресс-методы сравнения последовательностей. Программы FASTA и BLAST. Значимость выравнивания
9. Множественное выравнивание и алгоритмы (Clustal, Muscle). Недостатки существующих методов множественного выравнивания.
10. Методы кластеризации: UPGMA и NJ.
11. Профиль множественного выравнивания
12. Филогенетические деревья, бутстреп. Проблема переменной скорости эволюции
13. Секвенирование по Сэнгеру, принцип метода
14. Метод секвенирования путем синтеза с обратимой терминацией, платформа Illumina, мостиковая амплификация
15. Пиросеквенирование, принцип метода

16. Отличительные особенности методов секвенирования второго и третьего поколения
17. Спектр вопросов, охватываемых геномикой.
18. Методы генной инженерии.
19. Уровни структурной организации белков. Типы вторичной структуры.
20. Способы задания поворота в пространстве. Преимущества и недостатки.
21. Три варианта постановки задачи структурного выравнивания. Меры структурного сходства.
22. Классификация структуры белков. Структурная геномика.
23. Принципы рентгеноструктурного анализа. Разрешение.
24. Принцип работы простейших нейронных сетей.
25. Предсказание вторичной структуры белков.
26. Гидрофобность и её использование для предсказания топологии белков.
27. Принципы моделирования по гомологии.
28. Распознавание фолда. Протягивание. Неупорядоченные белки.
29. Молекулярное моделирование как вид вычислительного эксперимента.
30. Моделирование молекулярной динамики: алгоритмы и применение.
31. Линейные нотации (SMILES) и матрицы смежности.
32. Количественные соотношения между структурой и свойствами. Методы проверки предсказывающих функций.
33. Меры сходства молекул. Молекулярные дескрипторы. Битовые строки. Коэффициент Танимото.
34. Виды виртуального скрининга. Понятие фармакофора.
35. Виды молекулярных поверхностей и их особенности.
36. Алгоритмы молекулярного докинга.

8. Ресурсное обеспечение:

Литература

Основная литература:

1. Биоинформатика : учеб. для акад. бакалавриата. / В. В. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. - М. : Юрайт, 2017
2. Молекулярная эволюция и филогенетический анализ : учеб. пособие для студентов. / В.В.Лукашов. - М. : БИНОМ. Лаб. знаний, 2009.
3. Задачи и решения по анализу биологических последовательностей / Марк Бородавский, Светлана Екишева ; пер. с англ. А. А. Чумичкина под ред. А. А. Миронова. - Ижевск : Ин-т компьютер. исслед. : R&C Dynamics, 2008.
4. Физика белка : Курс лекций с цв. стереоскоп. ил. и задачами с решениями: Учеб. пособие для студентов вузов. / А.В.Финкельштейн, О.Б.Птицын; Ин-т белка РАН. - М. : Университет, 2005.
5. Молекулярное моделирование : теория и практика. / Х.-Д. Хельтье [и др.] ; пер. с англ. А. А. Олиференко [и др.] под ред. В. А. Палюлина и Е. В. Радченко. - М. : БИНОМ. Лаб. знаний, 2013.
6. Патрушев Л. И. Искусственные генетические системы. — М.: Наука, 2004.
7. Щелкунов С. Н. Генетическая инженерия. — Новосибирск: Сиб. унив. изд-во, 2010.

Дополнительная литература:

1. Введение в биоинформатику / А. Леск ; пер. с англ. под ред. А. А. Миронова и В. К. Швядоса. - М. : БИНОМ. Лаб. знаний, 2017.
2. Pevzner, J. Bioinformatics and functional genomics, 3rd edition
3. Панчин А. Ю. Сумма биотехнологии. Руководство по борьбе с мифами о генетической модификации растений, животных и людей. — М.: АСТ. — 2015.

Интернет-ресурсы

Зарубежные журналы и библиографические базы данных, доступные через Интернет
<https://www.ncbi.nlm.nih.gov/>

Материалы на сайте «Центр развития электронных образовательных ресурсов»:
<https://distant.msu.ru/course/view.php?id=1654>

Материалы на сайте кафедры биоинженерии:
<https://intbio.org/bioinf2019-2020/>

Материально-техническое обеспечение

соответствует требованиям п. 5.3 ОС по направлению подготовки 27.03.05 "Инноватика".

9. Язык преподавания

Русский

10. Преподаватели

Кандидат физико-математических наук, доцент каф. биоинженерии биологического факультета Новоселецкий Валерий Николаевич

11. Авторы программы

Кандидат физико-математических наук, доцент каф. биоинженерии биологического факультета Новоселецкий Валерий Николаевич