

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**

**Директор физтех-школы физики  
и исследований им. Ландау  
А.В. Рогачев**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Современные методы анализа омиксных данных
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Вычислительная биоинформатика (on-line-программа) Физтех-школа физики и исследований им. Ландау кафедра биофизики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Дифференцированный зачет
- 2 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

- лекции: 30 час.
- семинары: 30 час.
- лабораторные занятия: 0 час.

Самостоятельная работа: 135 час.

Подготовка к экзамену: 30 час.

Всего часов: 225, всего зач. ед.: 5

Программу составили:

Ф.Е. Френкель, канд. биол. наук  
О.Ю. Плотникова, ассистент

Программа обсуждена на заседании кафедры биофизики 16.04.2023

## Аннотация

В рамках дисциплины «Современные методы анализа омиксных данных» обучающиеся познакомятся с базовыми знаниями об основных алгоритмах, применяемых в задачах функциональной аннотации геномов, математических конструкциях, лежащих в их основе, а также статистических методах оценки параметров этих алгоритмов из реальных биологических последовательностей. На занятиях обучающиеся освоят методы анализа биологических последовательностей путем создания оптимальных статистических моделей сегментов последовательностей биополимеров, принадлежащих к тем или иным функциональным классам и получат основные вычислительные навыки, а также практический опыт, необходимый для проведения самостоятельных научных исследований в биоинформатике анализа.

### 1. Цели и задачи

#### Цель дисциплины

Дать студентам наиболее важные представления о математических основах современных алгоритмов, используемых для анализа последовательностей биополимеров, и практические навыки применения методов биоинформатики для анализа и интерпретации биологических данных.

#### Задачи дисциплины

- Программирование на Perl и MySQL, а также основам работы в операционной системе UNIX
- Умение пользоваться основными биоинформатическими базами данных
- Знакомств с базовыми алгоритмами и форматами данных биоинформатики;
- Формирование основных биоинформатических навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области системной биологии.
- Базовые знания об основных алгоритмах, применяемых в задачах функциональной аннотации геномов, математических конструкциях, лежащих в их основе, а также статистических методах оценки параметров этих алгоритмов из реальных биологических последовательностей;
- Практическое освоение студентами методов анализа биологических последовательностей путем создания оптимальных статистических моделей сегментов последовательностей биополимеров, принадлежащих к тем или иным функциональным классам;
- Получение основных вычислительных навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в биоинформатике анализа.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений

ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные биоинформатические алгоритмы, базы данных, форматы файлов;
- быстрый поиск подстроки в строке — алгоритмы наивный, Кнута-Мориса-Пратта, Рабина-Карпа, алгоритм кенгуру;
- индекс и преобразование Барроуза-Уиллера;
- BLAST — индексирование, статистика Альтшуля-Карлина;
- основы Байесовской статистики, правдоподобие, метод наибольшего правдоподобия, маргинализация распределений и маргинальное правдоподобие;
- оценка параметров скрытой цепи Маркова, обучение Витерби, метод Баума-Велша;
- методы анализа генома, основанные на скрытых марковских цепях, поиск кодирующих последовательностей, поиск однородных доменов хроматина;
- современные молекулярно-биологические методы исследования процессов в клетке;
- физические и химические свойства нуклеиновых кислот;
- процессы с участием нуклеиновых кислот;
- методы исследования генома и эволюции живых организмов.

уметь:

- пользоваться Интернет и справочной литературой по биологии научного и прикладного характера для быстрого поиска необходимых данных и понятий;
- программировать на языках Perl и MySQL;
- имплементировать и отлаживать биоинформатические алгоритмы;
- проектировать и заполнять базы данных;
- находить оптимальные алгоритмы для решения задач анализа биологических последовательностей, уметь оценить трудоемкость алгоритмов.

владеть:

- навыками работы с большими объемами биологических данных;
- культурой планирования и осуществления многоступенчатого биоинформатического анализа.

#### 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

##### 4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в язык программирования Perl и операционную систему UNIX	1	1		4
2	Основные понятия молекулярной биологии и алгоритмы биоинформатики	2	2		8
3	Аннотация генома. Методы предсказания генов	2	2		8
4	Методы филогенетического анализа	2	2		8
5	Биоинформатические базы данных, язык программирования MySQL	1	2		8
6	Анализ больших объемов биологических данных	2	2		8
7	BLAST	2	2		8
8	Алгоритмы динамического программирования	2	1		4
9	Быстрый поиск подстроки в строке	1	1		4
10	Индекс и преобразование Барроуза-Уиллера	1	1		5
11	Методы оптимизации	2	2		10
12	Методы функциональной аннотации генома	1	1		5
13	Мотивы в геномах	2	2		10
14	Основные структуры данных: хэш-таблица, суффиксное дерево, суффиксный массив	1	1		5
15	Основы Байесовской статистики	2	2		10
16	Оценка параметров скрытой цепи Маркова	2	2		10
17	Приложения алгоритмов динамического программирования	2	2		10
18	Скрытые цепи Маркова	2	2		10
Итого часов		30	30		135
Подготовка к экзамену		30 час.			
Общая трудоёмкость		225 час., 5 зач.ед.			

##### 4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 1 (Осенний)

#### 1. Введение в язык программирования Perl и операционную систему UNIX

Понятие переменной, массива, хэша, ссылки. Циклы, работа со строками, регулярные выражения. Основные команды UNIX.

## 2. Основные понятия молекулярной биологии и алгоритмы биоинформатики

Прокариоты, эукариоты, геном, ген, хромосома, ДНК, РНК, транскрипция, трансляция, генетический код, рибосома, нуклеотиды, аминокислоты, комплементарность, экзоны, интроны. Алгоритмы выравнивания биологических последовательностей – Smith-Waterman, Needleman-Wunsch, BLAST, динамическое программирование. Sequence LOGO, ПЦР.

## 3. Аннотация генома. Методы предсказания генов

Разница между прокариотическими и эукариотическими генами, сайт связывания рибосомы, GC%, рамка считывания.

## 4. Методы филогенетического анализа

Археи. Понятие консервативности и ее использование для предсказания функционально-значимых событий – программируемый сдвиг рамки считывания, псевдогены, Ka/Ks.

## 5. Биоинформатические базы данных, язык программирования MySQL

Базы данных NCBI, GENCODE, Genome Browser.

## 6. Анализ больших объемов биологических данных

Методы исследования генома и экспрессии генов. Анализ данных микрочипов и Next Generation Sequencing. Методы выявления мутаций, SNP, ошибки секвенирования, пайплайн, API.

Уровни регуляции экспрессии генов, транскрипционные факторы, микроРНК, длинные некодирующие РНК, вторичная структура РНК, антисмысловые взаимодействия, биохимические и регуляторные пути, биологические сети, база данных KEGG.

## 7. BLAST

Индексирование, зависимость длины ключа от алфавита, использование BLAST индекса в задачах протеомики, сравнение подходов BLAST и Смита-Вотермана к поиску локальных выравниваний. Статистика Альтшуля-Карлина. Распределение экстремальных значений.

Распределение Гумбеля. Пути с высоким локальным весом (HSP). Р-значение и Е-значение. Битовый скор.

## 8. Алгоритмы динамического программирования

Алгоритмы динамического программирования для поиска кратчайшего пути между двумя вершинами в направленном ациклическом графе (Беллмана-Форда) и вычисления суммы весов по всем таким путям (статсумма).

## 9. Быстрый поиск подстроки в строке

Алгоритмы наивный, Кнута-Мориса-Пратта, Рабина-Карпа, алгоритм кенгуру. Оценки трудоемкости. Оптимальность для поиска мотивов разной длины. Учет замен (wildcards). Оптимальная реализация.

Семестр: 2 (Весенний)

## 10. Индекс и преобразование Барроуза-Уиллера

Индекс и преобразование Барроуза-Уиллера. Оценка трудоемкости поиска. Проблема с учетом вставок-делеций. Использование в программах BWA и Bowtie для картирования ридов на геномы.

## 11. Методы оптимизации

Максимизация матожидания (Expectation maximization). Задача разделения двух кластеров. Роль выбора начальных значений. Оценка сходимости. Использование для построения множественных локальных выравниваний (MEME). Метод сэмплирование Гиббса. Детальный баланс. Проблема оценки сходимости.

## 12. Методы функциональной аннотации генома

Методы функциональной аннотации, основанные на скрытых марковских цепях, поиск кодирующих последовательностей, поиск однородных доменов хроматина.

## 13. Мотивы в геномах

Мотивы в геномах, поиск и идентификация мотивов, множественное локальное выравнивание. Представления мотивов: консенсусная строка, матрица позиционных весов, байесовская сеть. Алгоритм Тузе-Варре вычисления вероятности встречи мотива в случайной последовательности. Алгоритмы построения множественных локальных выравниваний и идентификации мотивов: жадный алгоритм Штормо, MEME. Ансамбли мотивов, ChIPmunk.

## 14. Основные структуры данных: хэш-таблица, суффиксное дерево, суффиксный массив

Хэш-таблица, суффиксное дерево, суффиксный массив, трудоемкость поиска в каждом случае.

## 15. Основы Байесовской статистики

Правдоподобие, метод наибольшего правдоподобия, маргинализация распределений и маргинальное правдоподобие. Последовательное байесовское оценивание. Интеграл Дирихле. Смесь Дирихле. Сопряженные распределения. Роль априорного распределения. Состоятельной байесовских оценок.

## 16. Оценка параметров скрытой цепи Маркова

Обучение Витерби, метод Баума-Велша, роль динамического программирования и байесовского оценивания.

## 17. Приложения алгоритмов динамического программирования

Приложения алгоритмов динамического программирования. Алгоритм поиска локального выравнивания Смита-Вотермана. Матрица Смита-Вотермана и соответствующий граф. Примеры путей. Алгоритм оптимальной сегментации последовательности на домены, однородные по составу. Формулировка на языке графов.

## 18. Скрытые цепи Маркова

Понятие о скрытой марковской модели, переходные и эмиссионные вероятности, поиск оптимальной последовательности переходов между состояниями для последовательности, порожденной скрытой марковской моделью (алгоритм Витерби), вычисление вероятности перехода в данной точке (алгоритм туда-обратно), использование алгоритма динамического программирования для анализа скрытых цепей Маркова.

## **5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)**

Занятия по учебной дисциплине проводятся с использованием дистанционных образовательных технологий. Каждый обучающийся обеспечен доступом к образовательной платформе [bostongene.com](http://bostongene.com)

## **6. Перечень рекомендуемой литературы**

### **Основная литература**

1. Анализ биологических последовательностей. Вероятностные модели белков и нуклеиновых кислот [Текст]/Р. Дурбин [и др.] , -М. : Ижевск, Регулярная и хаотическая динамика, 2006
1. Durbin, R., Eddy, S., Krogh, A., Mitchison, G. Biological sequence analysis, Cambridge University Press, 1998.
- Перевод: Дурбин, Р., Эдди, Ш., Крог, А., Митчисон, Г. Анализ биологических последовательностей (перевод А. Миронова). Издательство: Институт компьютерных исследований, 2006.
2. Borodovsky, M., Ekisheva, S. Problems and solution in biological sequence analysis. Cambridge University Press, 2006.
3. Pevzner, P.A., Shamir, R. Bioinformatics for Biologists. Cambridge University Press, 2011
4. Марк Бородовский, Светлана Екишева. Задачи и решения по анализу биологических последовательностей. Регулярная и хаотическая динамика, 2008

### **Дополнительная литература**

1. Гасфилд, Д. Строки, деревья и последовательности в алгоритмах. Информатика и вычислительная биология, Издательство Невский диалект, Санкт-Петербург, 2003
2. Рэндал Л. Шварц, Том Феникс, Брайан Д Фой. Изучаем Perl. Символ-Плюс, 2009

## **7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

База данных Национального института стандартизации и технологии США по свойствам соединений: <http://webbook.nist.gov/chemistry/>

## **8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)**

Аккаунт в Zoom.

Персональные компьютеры с UNIX системой, Perl, MySQL, Python.

## **9. Методические указания для обучающихся по освоению дисциплины (модуля)**

Студент, изучающий дисциплину, должен, с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения и понятия, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.



**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Вычислительная биоинформатика (on-line-программа) Физтех-школа физики и исследований им. Ландау кафедра биофизики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Дифференцированный зачет
- 2 (весенний) - Экзамен

**Разработчики:**

Ф.Е. Френкель, канд. биол. наук  
О.Ю. Плотникова, ассистент

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
	ОПК-4.3 Способен аргументировано выбирать способ проведения научного исследования
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Современные методы анализа омиксных данных» обучающийся должен:

**знать:**

- основные биоинформатические алгоритмы, базы данных, форматы файлов;
- быстрый поиск подстроки в строке — алгоритмы наивный, Кнута-Мориса-Пратта, Рабина-Карпа, алгоритм кенгуру;
- индекс и преобразование Барроуза-Уиллера;
- BLAST — индексирование, статистика Альтшуля-Карлина;
- основы Байесовской статистики, правдоподобие, метод наибольшего правдоподобия, маргинализация распределений и маргинальное правдоподобие;
- оценка параметров скрытой цепи Маркова, обучение Витерби, метод Баума-Велша;
- методы анализа генома, основанные на скрытых марковских цепях, поиск кодирующих последовательностей, поиск однородных доменов хроматина;
- современные молекулярно-биологические методы исследования процессов в клетке;
- физические и химические свойства нуклеиновых кислот;
- процессы с участием нуклеиновых кислот;
- методы исследования генома и эволюции живых организмов.

#### **уметь:**

- пользоваться Интернет и справочной литературой по биологии научного и прикладного характера для быстрого поиска необходимых данных и понятий;
- программировать на языках Perl и MySQL;
- имплементировать и отлаживать биоинформатические алгоритмы;
- проектировать и заполнять базы данных;
- находить оптимальные алгоритмы для решения задач анализа биологических последовательностей, уметь оценить трудоемкость алгоритмов.

#### **владеть:**

- навыками работы с большими объемами биологических данных;
- культурой планирования и осуществления многоступенчатого биоинформатического анализа.

### **3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю**

Примеры вопросов:

Алгоритмы выравнивания биологических последовательностей — Smith-Waterman, Needleman-Wunsch, BLAST, динамическое программирование. Sequence LOGO, ПЦР.

BLAST: индексирование и поиск локально-выровненных участков. Веса локально выровненных участков, распределение Гумбеля, статистика Альтшуля-Карлина, Р-значение и Е-значение.

Пример самостоятельной работы:

Предполагается дать студентам биоинформатическую задачу, максимально приближенную к реальной исследовательской работе. Студенты будут разбиты на 3-4 группы и каждая группа будет отвечать за свою часть проекта. На лекциях будем выслушивать и обсуждать результаты, полученные каждой из групп за неделю.

### **4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся**

9 семестр:

1. Основные команды операционной среды UNIX
2. Основы языка программирования Perl, регулярные выражения
3. Основы языка программирования MySQL, архитектура базы данных
4. Основные понятия молекулярной биологии
5. Алгоритм выравнивания нуклеотидных последовательностей Smith-Waterman
6. Алгоритм выравнивания нуклеотидных последовательностей Needleman-Wunsch
7. Подходы к предсказанию генов в геноме. Разница между эукариотическими и прокариотическими генами.
8. Методы филогенетического анализа.
9. Биоинформатические базы данных, Genome Browser
10. Методы исследования генома и экспрессии генов.
11. Методы выяснения первичной структуры нуклеиновых кислот.

12. Полимеразная цепная реакция (ПЦР).
13. Репликация, транскрипция, трансляция.
14. Уровни регуляции экспрессии генов. Системная биология.
15. Трансляция.
16. Генетический код. Частота встречаемости кодонов.
17. Подходы к анализу больших объемов биологических данных.
18. Основные структуры данных: хэш-таблица, суффиксное дерево, суффиксный массив.
19. Алгоритмы поиска подстроки в строке: наивный, Кнута-Мориса-Пратта, Рабина-Карпа

10 семестр:

1. Индекс и преобразование Барроуза-Уиллера.
2. BLAST: индексирование и поиск локально-выровненных участков.
3. BLAST: веса локально выровненных участков, распределение Гумбеля, статистика Альтшуля-Карлина, Р-значение и Е-значение.
4. Представления мотивов в геномах: консенсусная строка, матрица позиционных весов, байесовская сеть.
5. Алгоритм Тузе-Варре вычисления вероятности встречи мотива в случайной последовательности.
6. Алгоритмы построения множественных локальных выравниваний и идентификации мотивов: жадный алгоритм Штормо, MEME. Ансамбли мотивов, ChIPmunk.
7. Алгоритм динамического программирования для поиска кратчайшего пути между двумя вершинами в направленном ациклическом графе (Беллмана-Форда).
8. Алгоритмы динамического программирования для вычисления сумм весов по всем путям между двумя вершинами в направленном ациклическом графе (Беллмана-Форда)
9. Модификации алгоритмов динамического программирования для поиска локально выравнивания и сегментации последовательностей на блоки, однородные по составу.
10. Понятие о скрытой марковской модели, переходные и эмиссионные вероятности.
11. Алгоритм Витерби поиска оптимальной последовательности переходов между состояниями для последовательности, порожденной скрытой марковской моделью.
12. Алгоритм «туда-обратно» вычисление вероятности перехода в скрытой цепи Маркова в данной точке
13. Основы Байесовской статистики. Априорное распределение вероятностей. Маргинализация..
14. Алгоритмы максимизации математического ожидания (Expectation maximization) и сэмплирования по Гиббсу для поиска максимального правдоподобия.
15. Оценка параметров скрытой цепи Маркова методом обучения Витерби.
16. Оценка параметров скрытой цепи Маркова с помощью алгоритма Баума-Велша.
17. Поиск кодирующих последовательностей с помощью скрытых Марковских цепей.
18. Программа GeneMark.
19. Поиск участков с конкретным состоянием хроматина с помощью скрытых марковских цепей. Алгоритм Эрнста-Келлиса

Пример экзаменационного билета в 10 семестре:

Билет 1.

1. BLAST: индексирование и поиск локально-выровненных участков.
2. Алгоритмы динамического программирования для вычисления сумм весов по всем путям между двумя вершинами в направленном ациклическом графе (Беллмана-Форда)

#### Критерии оценивания

Оценка отлично 10 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 9 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 8 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо 7 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо 6 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо 5 баллов - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно 4 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно 3 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно 2 балла - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно 1 балл - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

## **5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Дифференцированный зачет и экзамен проводятся в устной форме по билетам. В каждом билете представлено два теоретических вопроса. При проведении дифференцированного зачета и экзамена обучающемуся предоставляется 50 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.