

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

**УТВЕРЖДЕНО**

**Директор физтех-школы  
биологической и медицинской  
физики**

**Д.В. Кузьмин**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Проблемы биоинформатики/ОМИКСы
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Биоинженерия и биоинформатика Физтех-школа Биологической и Медицинской Физики кафедра биоинформатики и системной биологии
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Подготовка к экзамену: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

И.В. Антонов, phd (канд. биол. наук)

Ю.А. Медведева, канд. биол. наук

Программа обсуждена на заседании кафедры биоинформатики и системной биологии 11.05.2020

## Аннотация

Цель данной дисциплины дать практические навыки применения методов биоинформатики для анализа и интерпретации биологических данных.

Студента обучат языкам программирования Perl и MySQL, а также основам работы в операционной системе UNIX, научат пользоваться основными биоинформатическими базами данных.

Студент ознакомится с базовыми алгоритмами и форматами данных биоинформатики.

## 1. Цели и задачи

### Цель дисциплины

дать практические навыки применения методов биоинформатики для анализа и интерпретации биологических данных.

### Задачи дисциплины

- обучить языкам программирования Perl и MySQL, а также основам работы в операционной системе UNIX;
- научить пользоваться основными биоинформатическими базами данных;
- ознакомить с базовыми алгоритмами и форматами данных биоинформатики;
- формирование у студентов основных биоинформатических навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области системной биологии.

## 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
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-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

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

В результате освоения дисциплины обучающиеся должны

знать:

- основные биоинформатические алгоритмы, базы данных, форматы файлов;
- 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:

- программировать на языках Perl и MySQL;
- имплементировать и отлаживать биоинформатические алгоритмы;
- проектировать и заполнять базы данных.

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	4			4
2	Основные понятия молекулярной биологии и алгоритмы биоинформатики	4			4
3	Аннотация генома. Методы предсказания генов.	4			4
4	Методы филогенетического анализа	2			2
5	Биоинформатические базы данных, язык программирования MySQL	2			2
6	Анализ больших объемов биологических данных	4			4
7	Анализ больших объемов биологических данных	2			2
8	Большой биоинформатический проект, работа группами	8			8
Итого часов		30			30
Подготовка к экзамену		30 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

###### 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.

### 7. Анализ больших объемов биологических данных

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

### 8. Большой биоинформатический проект, работа группами

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 заслушивать и обсуждать результаты, полученные каждой из групп за неделю. Возможный план работы – сборка прокариотического генома на основании данных Next Generation Sequencing, предсказание генов, функциональная аннотация генов, сравнительный и/или филогенетический анализ.

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

Оборудование, необходимое для лекций: аудитория, компьютер и мультимедийное оборудование (проектор, звуковая система), UNIX сервер с отдельным аккаунтом для каждого студента.

## 6. Перечень рекомендуемой литературы

### Основная литература

Базовая кафедра предоставляет литературу

1. Р. Дурбин, Ш. Эдди, А. Крог, Г. Митчисон. Анализ биологических последовательностей. Регулярная и хаотическая динамика, 2006.
2. Марк Бородовский, Светлана Екишева. Задачи и решения по анализу биологических последовательностей. Регулярная и хаотическая динамика, 2008

### Дополнительная литература

Базовая кафедра предоставляет литературу

1. 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)

Научно-библиографические и патентные базы данных в области физико-химической биологии, доступные по сети Интернет в бесплатном режиме - Science Citation Index (Web of Science), Medline (PubMed), Научная электронная библиотека (НЭБ), Российская патентная БД ФГУ ФИПС и американская патентная БД USPAFULL; электронные адреса крупных научных издательств, предоставляющих доступ к полным текстам текущих и архивным выпускам этих журналов.

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

Электронные ресурсы, включая доступ к базам данных, доступ к фондам библиотеки.

<http://www.ncbi.nlm.nih.gov/pubmed> — база данных статей

Для части занятий потребуется Zoom. Google Drive для доступа к материалам курса.

Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

## **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 дисциплины, 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>	Биоинженерия и биоинформатика Физтех-школа Биологической и Медицинской Физики кафедра биоинформатики и системной биологии
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

**Разработчики:**

И.В. Антонов, phd (канд. биол. наук)

Ю.А. Медведева, канд. биол. наук

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
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-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Проблемы биоинформатики/ОМИКСы» обучающийся должен:

### знать:

- основные биоинформатические алгоритмы, базы данных, форматы файлов;
- 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:

- программировать на языках Perl и MySQL;
- имплементировать и отлаживать биоинформатические алгоритмы;
- проектировать и заполнять базы данных.

### 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

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

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.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на экзамене.

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

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.

Примеры билетов:

Билет №1

Биоинформатические базы данных

Билет №2

Генетический код. Частота встречаемости кодонов.

Билет №3

Методы выявления мутаций

Билет №4

Методы филогенетического анализа.

Билет №5

SNP и ошибки секвенирования

#### **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**

При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать одного астрономического часа.