

**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

	Рабочая программа дисциплины (модуля)
по дисциплине:	Базы данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Алгоритмическая биология
	Физтех-школа Биологической и Медицинской Физики
	центр образовательных программ Физтех-школы биологической и медицинской физики
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Аудиторных часов: 45 всего, в том числе:

лекции: 30 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Количество контрольных работ, заданий: 2

Программу составил: В.Ю. Макеев, д-р физ.-мат. наук

Программа обсуждена на заседании центра образовательных программ Физтех-школы биологической и медицинской физики 12.07.2023

Аннотация

Дисциплина направлена на ознакомление слушателей с основными источниками данных, используемых в геномной и структурной биоинформатике. Студент после освоения курса будет понимать источники данных, используемых в биоинформатике; способы хранения геномных данных; способам анализа геномных данных, в частности выравнивания нуклеотидных последовательностей и построения филогенетических деревьев; способам хранения данных о молекулярных структурах; моделирование методом молекулярной динамики; подходы к виртуальному скринингу лигандов и разработки лекарств на основе структурных данных.

1. Цели и задачи

Цель дисциплины

- ознакомить слушателей с основными источниками данных, используемых в геномной и структурной биоинформатике, а также с базовыми алгоритмами и программными пакетами для анализа получаемых данных.

Задачи дисциплины

1. Сформировать актуальные знания по:
 - источникам данных, используемых в биоинформатике;
 - способам хранения геномных данных;
 - способам анализа геномных данных, в частности выравнивания нуклеотидных последовательностей и построения филогенетических деревьев;
 - способам хранения данных о молекулярных структурах;
 - подходам к анализу структурных данных;
 - подходам к визуализации структурных данных;
 - моделированию методом молекулярной динамики;
 - подходам к виртуальному скринингу лигандов и разработки лекарств на основе структурных данных.
2. Сформировать навыки использования общественно доступных баз данных с целью получения геномных и структурных данных; анализа последовательности генов, в частности выравнивание последовательностей и восстановление филогенетического дерева; визуализации структуры биологических молекул; анализа структуры биологических молекул, в частности выполнение структурного выравнивания, производить сравнение двух и более структур, производить расчеты геометрических характеристик.

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-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	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
	УК-4.1 Способен вести обмен деловой информацией в устной и письменной формах на государственном языке Российской Федерации и не менее чем на одном иностранном языке

УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.2 Владеет навыками, необходимыми для написания, письменного перевода и редактирования различных академических текстов (рефератов, эссе, обзоров, статей и т.д.)
	УК-4.3 Способен представлять результаты академической и профессиональной деятельности на различных научных мероприятиях, включая международные
	УК-4.4 Способен использовать современные средства информационно-коммуникационных технологий для академического и профессионального взаимодействия
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
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

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- источники данных, используемых в биоинформатике;
- способы хранения геномных данных;
- способы анализа геномных данных, в частности выравнивания нуклеотидных последовательностей и построения филогенетических деревьев;
- способы хранения данных о молекулярных структурах;
- подходы к анализу структурных данных;
- подходы к визуализации структурных данных;
- моделирование методом молекулярной динамики;
- подходы к виртуальному скринингу лигандов и разработки лекарств на основе структурных данных.

уметь:

- пользоваться общественно доступными базами данных с целью получения геномных и структурных данных;
- анализировать последовательности генов, в частности делать выравнивание последовательностей и восстановление филогенетического дерева;
- визуализировать структуры биологических молекул;
- анализировать структуры биологических молекул, в частности выполнять структурное выравнивание, производить сравнение двух и более структур, производить расчеты геометрических характеристик.

владеть:

- навыками выбора методов и средств решения задач исследования;
- методами теоретического и экспериментального исследования;
- навыками поиска (в том числе с использованием информационных систем и баз данных), обработки, анализа и систематизации информации;
- навыками критического анализа и оценки современных научных достижений.

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	Понятие базы данных	2	1		4
2	Аминокислотная последовательность белка	4	2		8
3	Знакомство с базами данных нуклеотидных последовательностей, их история	4	2		8
4	BLAST: краткое теоретическое введение	4	2		8
5	Биологические структуры, база Protein Data Bank	4	2		8
6	Краткие принципы биологической классификации и таксономии	4	2		8
7	Понятие метаболических путей, сигнальных путей	4	2		8
8	База данных OMIM	4	2		8
Итого часов		30	15		60
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 1 (Осенний)

1. Понятие базы данных

Какие биологические данные можно хранить в базах данных? Обзор основных существующих биологических баз данных. Обзор форматов биологических данных. Как конвертировать разные форматы биологических данных?

2. Аминокислотная последовательность белка

Знакомство с базой данных Uniprot, продвинутый поиск в Uniprot. База данных Pfam. Практическое задание: выжать как можно сведений о некотором белке с помощью упомянутых баз данных. Работа с выравниваниями, загруженными из Pfam, с помощью программы JalView. Понятие филогенетического дерева. Извлечение филогенетических деревьев из базы Pfam. Работа с филогенетическими деревьями в программе MEGA7.

3. Знакомство с базами данных нуклеотидных последовательностей, их история

Знакомство с различными базами данных NCBI. Работа с выравниваниями нуклеотидных последовательностей в программе JalView. Работа с геномным браузером.

4. BLAST: краткое теоретическое введение

Практические задания по работе с BLAST: разные виды BLAST, управление параметрами BLAST, переключение на разные базы данных.

5. Биологические структуры, база Protein Data Bank

Структура файла формата .pdb, основы работы в PyMol, построение пространственного выравнивания с помощью PDBFold.

6. Краткие принципы биологической классификации и таксономии

Краткие принципы биологической классификации и таксономии. Знакомство с некоторыми таксономическими базами данных (IT IS, NCBI, EOL).

7. Понятие метаболических путей, сигнальных путей

Понятие метаболических путей, сигнальных путей. Функциональные базы данных: GO, KEGG.

8. База данных OMIM

Базы мутаций. База данных PubMed. Как искать биологическую литературу с помощью PubMed и Google Scholar?

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

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

Данная литература предоставляется на базовой кафедре:

- Леск А. Введение в биоинформатику / пер. с англ. - М.: БИНОМ. Лаборатория знаний, 2009. - 318 с.
- Bhatt V.D., Patel M., Joshi C.G. (2018) An Insight of Biological Databases Used in Bioinformatics. In: Wadhwa G., Shanmughavel P., Singh A., Bellare J. (eds) Current trends in Bioinformatics: An Insight. Springer, Singapore. doi: 10.1007/978-981-10-7483-7_1

Дополнительная литература

1. Введение в системы баз данных [Текст] : [учебник для вузов] / К. Дж. Дейт ; [пер. с англ. К. А. Птицына] .— 8-е изд. — М. : Вильямс, 2008 .— 1328 с.
2. Основы баз данных [Текст] : Курс лекций : учеб. пособие для вузов / С. Д. Кузнецов .— М. : Интернет-Ун-т Информ. Технологий, 2005 .— 488 с.
3. Проектирование структур баз данных [Текст]. В 2 кн. Кн. 1/Т. Тиори, Дж. Фрай , пер. с англ. А. И. Роговского, В. И. Чучкина , -М., Мир, 1985
4. Проектирование структур баз данных [Текст]. В 2 кн. Кн. 2/Т. Тиори, Дж. Фрай , -М., Мир, 1985

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

- Курс Bioinformatics Methods I на Coursera рассказывает, помимо прочего, о биологических базах данных и BLAST, содержит практические задания.
- Курс Bioinformatics Capstone: Big Data in Biology на Coursera.

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

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.

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Алгоритмическая биология Физтех-школа Биологической и Медицинской Физики центр образовательных программ Физтех-школы биологической и медицинской физики
курс:	<u>1</u>
квалификация:	магистр
Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен	
Разработчик:	В.Ю. Макеев, д-р физ.-мат. наук

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-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	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
УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1 Способен вести обмен деловой информацией в устной и письменной формах на государственном языке Российской Федерации и не менее чем на одном иностранном языке
	УК-4.2 Владеет навыками, необходимыми для написания, письменного перевода и редактирования различных академических текстов (рефератов, эссе, обзоров, статей и т.д.)
	УК-4.3 Способен представлять результаты академической и профессиональной деятельности на различных научных мероприятиях, включая международные
	УК-4.4 Способен использовать современные средства информационно-коммуникационных технологий для академического и профессионального взаимодействия
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
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

2. Показатели оценивания компетенций

В результате изучения дисциплины «Базы данных» обучающийся должен:

знать:

- источники данных, используемых в биоинформатике;
- способы хранения геномных данных;
- способы анализа геномных данных, в частности выравнивания нуклеотидных последовательностей и построения филогенетических деревьев;
- способы хранения данных о молекулярных структурах;
- подходы к анализу структурных данных;
- подходы к визуализации структурных данных;
- моделирование методом молекулярной динамики;
- подходы к виртуальному скринингу лигандов и разработки лекарств на основе структурных данных.

уметь:

- пользоваться общественно доступными базами данных с целью получения геномных и структурных данных;
- анализировать последовательности генов, в частности делать выравнивание последовательностей и восстановление филогенетического дерева;
- визуализировать структуры биологических молекул;
- анализировать структуры биологических молекул, в частности выполнять структурное выравнивание, производить сравнение двух и более структур, производить расчеты геометрических характеристик.

владеть:

- навыками выбора методов и средств решения задач исследования;
- методами теоретического и экспериментального исследования;
- навыками поиска (в том числе с использованием информационных систем и баз данных), обработки, анализа и систематизации информации;
- навыками критического анализа и оценки современных научных достижений.

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

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

1. Обзор форматов биологических данных.
2. Понятие филогенетического дерева.
3. Работа с выравниваниями нуклеотидных последовательностей в программе JalView.
4. Биологические структуры.
5. Понятие метаболических путей, сигнальных путей. Функциональные базы данных: GO, KEGG.
6. База данных OMIM.
7. Работа с геномным браузером.
8. Краткие принципы биологической классификации и таксономии.
9. Функциональные базы данных.
10. Базы мутаций. База данных PubMed.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. 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. Какие биологические данные можно хранить в базах данных?
2. Обзор основных существующих биологических баз данных.
3. Как конвертировать разные форматы биологических данных?
4. База данных Pfam. Работа с выравниваниями, загруженными из Pfam, с помощью программы JalView.
5. Понятие филогенетического дерева. Извлечение филогенетических деревьев из базы Pfam.
6. Работа с филогенетическими деревьями в программе MEGA7.
7. Структура файла формата .pdb, основы работы в PyMol, построение пространственного выравнивания с помощью PDBFold.
8. Краткие принципы биологической классификации и таксономии.

9. Базы данных (IT IS, NCBI, EOL).
10. Как искать биологическую литературу с помощью PubMed и Google Scholar?
11. Краткие принципы биологической классификации и таксономии.
12. Функциональные базы данных.

Примеры билетов:

Билет 1.

База данных Pfam. Работа с выравниваниями, загруженными из Pfam, с помощью программы JalView.

Базы данных (IT IS, NCBI, EOL).

Билет 2.

Краткие принципы биологической классификации и таксономии.

Функциональные базы данных.

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 60 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать одного астрономического часа.