

**Федеральное агентство научных организаций (ФАНО России)  
Федеральное государственное бюджетное учреждение науки  
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ  
им. академиков М.М. Шемякина и Ю.А. Овчинникова  
Российской академии наук  
(ИБХ РАН)**

СОГЛАСОВАНО:  
Ученый совет ИБХ РАН  
Протокол № от « » 2021 г.

УТВЕРЖДАЮ:  
Директор ИБХ РАН

Ученый секретарь  
д.ф.-м.н. В.А.Олейников

академик А.Г.Габибов

от « » 2021 г.

от « » 2021 г.

**РАБОЧАЯ ПРОГРАММА  
по дисциплине  
БИОИНФОРМАТИКА**

**Направление подготовки:**

1.5. Биологические науки

**Направленность (профиль) программы:**

1.5.4. Биохимия

1.5.6. Биотехнология

1.5.3. Молекулярная биология

**Направление подготовки:**

1.4. Химические науки

**Направленность (профиль) программы:**

1.4.9. Биоорганическая химия

**Уровень высшего образования:** подготовка научных и научно-педагогических кадров в аспирантуре

**Квалификация выпускника:** Исследователь. Преподаватель-исследователь.

**Форма обучения:** очная

**Разработчики:** к.б.н. Залевский А.О. д.ф-м.н Ефремов Р.Г.

*Рабочая программа составлена на основании федеральных государственных образовательных стандартов высшего образования (ФГОС ВО), разработанных для реализации основных профессиональных образовательных программ высшего образования - программ подготовки научных и научно-педагогических кадров в аспирантуре по направлению 1.5. Биологические науки, 1.4. Химические науки.*

Согласно ФГОС ВО по направлению подготовки 1.5. «Биологические науки», 1.4. «Химические науки» (уровень подготовки кадров высшей квалификации) и учебному плану аспирантов, разработанного на основе этих требований, дисциплина «Биоинформатика» является обязательной учебной дисциплиной обязательной части Блока 1 образовательной программы по направленности (профилю) 1.5.4. Биохимия, 1.5.6. Биотехнология, 1.5.3. Молекулярная биология и 1.4.9. Биоорганическая химия на изучение которых отведена 1 зачетная единица. Соответствующий этому объёму курс составляет 36 академических часов, из них 18 академических часов лекций, 4 академических часов практические занятия (семинары), 10 часов самостоятельной внеаудиторной работы аспирантов, включая подготовку к дифференцированному зачету и 4 часа на контроль знаний в форме зачета.

## **I. Цели и задачи изучения дисциплины.**

Биоинформатика — междисциплинарная область, объединяющая молекулярную биологию, генетику, компьютерные науки, математику и статистику. Для решения крупномасштабных биологических проблем, требующих анализа больших объемов данных, используется биоинформатика. Биоинформатика включает в себя изучение и разработку компьютерных методов и направлена на получение, анализ, хранение, организацию и визуализацию биологических данных. Биоинформатика играет важную роль в формировании у будущих исследователей и преподавателей научного мировоззрения и современного информационного мышления, достаточной теоретической базы для успешного усвоения аспирантами общепрофессиональных и специальных дисциплин. В процессе изучения курса «Биоинформатика» происходит ознакомление аспирантов с современной научной литературой, вырабатываются умение решать конкретные профессионально ориентированные задачи.

**1.1. Цель курса:** ознакомление с высокопроизводительными информационно-компьютерными технологиями для информационного сопровождения научных исследований.

**1.2. Задачи курса:** ознакомить аспирантов с основными инструментами и алгоритмами биоинформатики, такими как базовый филогенетический анализ, алгоритмы выравнивания последовательностей, белок структурный анализ, аннотации генов и геномов, а также с основными базами данных, в которых хранятся последовательности нуклеиновых кислот, белков и данные эпигенетики.

**1.3. Связь с другими дисциплинами:**

Курс «Биоинформатика» в той или иной степени имеет непосредственную связь практически со всеми дисциплинами, изучаемыми на протяжении всего времени овладения аспирантами образовательной программы по направлению подготовки 1.5. Биологические науки, 1.4. Химические науки и является базовым курсом при подготовке специалистов направленности (профилю) в биохимия, молекулярная биология, биотехнология и биоорганическая химия.

## II. Требования к уровню освоения дисциплины

В рамках данной дисциплины углубляются и развиваются следующие компетенции:

### **Универсальные компетенции (УК):**

- способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях (УК-1);
- способность планировать и решать задачи собственного профессионального и личностного развития (УК-5).

### **Общепрофессиональные компетенции (ОПК):**

- способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий (ОПК-1);
- готовность к преподавательской деятельности по основным образовательным программам высшего образования (ОПК-2).

### **Профессиональные компетенции (ПК):**

- способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности (профилю) (ПК-1);
- обладание представлениями о системе фундаментальных понятий и методологических аспектов биологии, форм и методов научного познания (ПК-2);
- способность приобретать новые знания с использованием современных научных методов и владение ими на уровне, необходимом для решения задач, возникающих при выполнении профессиональных функций (ПК-3);
- обладание опытом профессионального участия в научных дискуссиях, умение представлять полученные в исследованиях результаты в виде отчетов и научных публикаций (стендовые доклады, рефераты и статьи в периодической научной печати) (ПК-4);
- владение методами отбора материала, преподавания и основами управления процессом обучения фундаментальной биологии в вузе (ПК-5).

В результате освоения дисциплины «Биоинформатики» аспирант должен:

### **Знать:**

- основные базы данных для проведения поиска научно-библиографической, фактографической и патентной информации;
- принципы составления поискового предложения при поиске научно-библиографической, фактографической и патентной информации;
- методы критического анализа и оценки современных научных достижений в области биоинформатики.

### **Уметь:**

- осуществлять выбор методов и средств решения исследовательских задач с помощью компьютерных технологий;
- осуществлять поиск библиографической и фактографической информации с использованием информационных систем и баз данных;
- работать с научно-технической информацией;
- выделять и систематизировать основные идеи в научных текстах;
- критически оценивать любую поступающую информацию, вне зависимости от источника;
- при решении исследовательских и практических задач, генерировать новые идеи;
- использовать основные законы естественнонаучных дисциплин в профессиональной деятельности.

### **Владеть:**

- навыками поиска с использованием информационных систем и баз данных,

обработки, анализа и систематизации информации;  
 - навыками критического анализа и оценки современных научных достижений в области биоинформатики.

### III. Объем дисциплины и виды учебной работы:

Форма обучения – ОЧНАЯ

**Общий объем дисциплины:** 1 зачетная единица или 36 академических часов.

Всего часов	Аудиторные занятия (час), в том числе:			Самостоятельная Работа (час)	Контроль (час)
	лекции	практические занятия (семинары)	лабораторные работы		
36	18	4	-	10	4
	22				

### Распределение аудиторных часов по темам и видам учебной работы:

№	Наименование тем и разделов (час), (с развернутым содержанием курса в том числе: по каждой теме и разделу)	Аудиторные занятия (час) в том числе	
		Лекции	Семинары
1	Введение в биоинформатику. Биоинформатика: данные, методы и применение. Множественное выравнивание.	2	2
2	Множественное локальное выравнивание (motif identification).	2	2
3	Классификация белков. Семейства белков PFAM. Структура белков. Прогнозирование структуры белка, докинг.	2	
4	Филогенетические деревья. Максимум вероятность.	2	
5	Распознавание генов. Методы de novo. Функциональная аннотация геномов.	4	
6	Эволюция и разнообразие CRISPR систем.	2	
7	NGS в медицине глазами клинициста - методы, клиническое значение, интерпретация, эволюция методов. Онкогенетика.	2	
8	Секвенирование геномов и транскриптомов. Пайплайны для геномного-экзомного секвенирования, классификации вариантов по патогенности. Базы данных.	2	
9	Биоинформатика последовательностей. Введение в системную биологию.	2	
	Всего:	18	4
	<b>Итого:</b>	22	

### IV. Содержание курса «Биоинформатика»

#### Раздел 1.

#### Введение в биоинформатику

Биоинформатика: данные, методы и применение. Краткий обзор: что может быть достигнуто использованием различных вычислительных методов в биологии. Краткое введение в NGS. Различные типы данных, которые обычно используют биоинформатики. Обсуждение экспериментальных лабораторных методов, лежащие в основе этих данных.

## **Раздел 2.**

### **Множественное локальное выравнивание (motif identification)**

Множественное локальное выравнивание (motif identification). Концепция выравнивания. Алгоритмы. Приложение, лучшие практики. CLUSTAL, MUSCLE. Локальное выравнивание. Алгоритмы. Как это может быть применимо? Регуляторные мотивы. MEME.

## **Раздел 3.**

### **Классификация белков и методы прогнозирования структуры белков**

Классификация белков. Семейства белков PFAM. Как узнать функцию белка? Базы данных, содержащие информацию о функции белков. Структура белка: основные элементы. Можно ли предсказать структуру белка? Прогнозирование вторичной структуры белка, de novo предсказание третичной структуры белка, докинг. Работа с белковыми моделями.

## **Раздел 4.**

### **Филогенетические деревья**

Филогенетические деревья. Максимум вероятность. Основные алгоритмы. Приложения. Лучшие практики. Пакеты: MEGA, PHYLIP и др.

## **Раздел 5.**

### **Распознавание генов. Методы de novo Функциональная аннотация геномов**

Распознавание генов. Методы de novo. Открытые рамки для чтения. Предсказание генов, разные методы и краткий обзор алгоритмов. GeneMark, GLIMMER, GenScan и т.д. Использование данных транскриптома: AUGUSTUS др. Сравнительные методы. Функциональная аннотация геномов. Основные конвейеры и инструменты для аннотирования генома. Основные базы данных и инструменты. Основные источники данных. Базы данных с геномными, эпигеномными и белковыми данными. GenBank and BLAST. InterPro. Genome Browser and ENSEMBLE.

## **Раздел 6.**

### **CRISPR систем.**

Эволюция CRISPR систем. Разнообразие CRISPR систем. Основные ферменты, участники системы.

## **Раздел 7.**

### **NGS в медицине глазами клинициста**

NGS в медицине глазами клинициста - методы, клиническое значение, интерпретация, эволюция методов. Онкогенетика.

## **Раздел 8. Секвенирование геномов и транскриптомов**

Пайплайны для геномного-экзомного секвенирования, классификации вариантов по патогенности. Базы данных.

## **Раздел 9. Биоинформатика последовательностей**

Биоинформатика последовательностей. Анализ и каталогизация биохимических путей и сетей. Введение в системную биологию.

### **V. Самостоятельная работа**

Предусмотрено самостоятельное изучение отдельных вопросов лекционного курса дисциплины «Биоинформатика» в виде проработки лекционного материала и соответствующих разделов курса по учебникам.

### **VI. Итоговая проверка знаний**

Учебный план по дисциплине «Биоинформатика» предусматривает контроль знаний в форме дифференцированного зачета с выставлением оценок в пятибалльной системе.

#### **Вопросы для дифференцированного зачета:**

1. БД Science Citation Index (Web of Science), её тематическая направленность, принципы составления поискового предложения.
2. БД Medline (PubMed), её тематическая направленность, принципы составления поискового предложения.
3. Научная электронная библиотека (НЭБ), её тематическая направленность, принципы составления поискового предложения.
4. Российская патентная БД ФГУ ФИПС, принципы составления поискового предложения.
5. Американская патентная БД USPATFULL, принципы составления поискового предложения.
6. Базы данных, содержащие нуклеотидные последовательности. Формат документа.
7. Базы данных, содержащие аминокислотные последовательности. Формат документа.
8. База данных, содержащая двумерные электрофоретические карты белков в полиакриламидных гелях. Формат документа.
9. Базы данных, содержащие семейства белков.
10. Базы данных, содержащие семейства гомологичных доменов белков.
11. Базы данных, содержащие пространственные структуры биологических макромолекул: белков, нуклеиновых кислот, белково-нуклеиновых комплексов. Формат документа.
12. Базы данных, содержащие структуры углеводов и низкомолекулярных биорегуляторов.
13. Комбинированные банки, содержащие информацию о семействах, структурах, функциях, генах, геномах. Интегрированная система для поиска библиографической и фактографической информации.

#### **Критерии оценки ответа аспиранта на зачете:**

##### **Оценка «отлично»**

Ответы на поставленные вопросы излагаются логично, последовательно и не требуют дополнительных пояснений. Делаются обоснованные выводы. Соблюдаются нормы литературной речи. Оценка “отлично” ставится аспирантам, которые при ответе:

- обнаруживают всестороннее систематическое и глубокое знание программного материала;
- способны творчески применять знание теории к решению профессиональных задач;
- владеют понятийным аппаратом;
- демонстрируют способность к анализу и сопоставлению;
- различных подходов к решению заявленной в вопросе проблематики;
- подтверждают теоретические постулаты примерами из педагогической практики.

### **Оценка «хорошо»**

Ответы на поставленные вопросы излагаются систематизировано и последовательно. Материал излагается уверенно. Демонстрируется умение анализировать материал, однако не все выводы носят аргументированный и доказательный характер. Соблюдаются нормы литературной речи. Оценка “хорошо” ставится аспирантам, которые при ответе:

- обнаруживают твёрдое знание программного материала;
- способны применять знание теории к решению задач профессионального характера;
- допускают отдельные погрешности и неточности при ответе.

### **Оценка «удовлетворительно»**

Допускаются нарушения в последовательности изложения. Демонстрируются поверхностное знание вопроса. Имеются затруднения с выводами. Допускаются нарушения норм литературной речи. Оценка “удовлетворительно” ставится аспирантам, которые при ответе:

- в основном знают программный материал в объёме, необходимом для предстоящей работы по профессии;
- допускают существенные погрешности в ответе на вопросы экзаменационного билета;
- приводимые формулировки являются недостаточно четкими, в ответах допускаются неточности.

**Положительная оценка может быть поставлена при условии понимания аспирантом сущности основных категорий по основному и дополнительным вопросам.**

### **Оценка «неудовлетворительно»**

Материал излагается непоследовательно, сбивчиво, не представляет определенной системы знаний. Имеются заметные нарушения норм литературной речи. Оценка “неудовлетворительно” ставится аспирантам, которые при ответе:

- обнаруживают значительные пробелы в знаниях основного программного материала;
- допускают принципиальные ошибки в ответе на вопрос билета;
- демонстрируют незнание теории и практики.

## **VII. Учебно-методическое обеспечение дисциплины**

**Рекомендуемая литература для освоения теоретического курса.**

### **Основная литература:**

1. Н.Ю. Часовских. Биоинформатика. 2020.
2. Хаубольд Б., Вие Т. Введение в вычислительную биологию. Эволюционный подход. URSS, 2011.
3. Ригден Д.Дж. Структура и функционирование белков: Применение методов биоинформатики. URSS. 2014.
4. Введение в биоинформатику / А. Леск; пер. с англ. —М.: БИНОМ. Лаборатория знаний, 2009.
5. Гельфанд М. Биоинформатика и геномика. <https://postnauka.ru/courses/42433>.

6. Методы молекулярной биологии. Биоинформатика. 2008.

**Дополнительная литература:**

7. Д.Леон, С.Маркел, Sequence Analysis in a Nutshell - справочник от O'Reilly по наиболее часто используемым в биоинформатике базам данных и программам по анализу последовательностей. <http://bioinformatics.ru/Ssylki.html>