

«ПРИНЯТО»

На заседании Ученого совета

ФИЦ Биотехнологии РАН

Протокол № 1 от «28» июля 2015 г.

«УТВЕРЖДАЮ»

Директор

ФИЦ Биотехнологии РАН

Член-корр. РАН

В.О. Попов



**РАБОЧАЯ ПРОГРАММА  
ПО ДИСЦИПЛИНЕ  
ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ БИОИНФОРМАТИКИ**

**Направление подготовки:** 06.06.01 Биологические науки

**Уровень образования:** высшее образование - подготовка кадров высшей квалификации

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

Москва

2015 г.

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

### **1.1. Цели дисциплины**

— дать аспирантам наиболее важные представления о математических основах современных алгоритмов, используемых для анализа последовательностей биополимеров, основных биологических задачах, в которых возникает потребность в этих алгоритмах, и об практике и ограничениях их применимости.

### **1.2. Задачи дисциплины**

— формирование базовых знаний 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;

— формирование у студентов основных вычислительных навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в биоинформатике анализа;

## **2. Место дисциплины в структуре ООП**

Дополнительные главы биоинформатики является дисциплиной по выбору вариативной части учебного плана основной образовательной программы.

В рамках данной дисциплины углубляются и развиваются следующие компетенции:

### **Универсальные компетенции**

— 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 проектировать и осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения с использованием знаний в области истории и философии науки (УК-2);

— Готовность участвовать в работе российских и международных

исследовательских коллективов по решению научных и научно-образовательных задач (УК-3);

— Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках (УК-4);

— Способность планировать и решать задачи собственного профессионального и личностного развития (УК-5);

### **Общепрофессиональные компетенции**

— 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**

— 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 (научной специальности) (ПК-1);

— Владение 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);

— Владение методами отбора материала, преподавания и основами управления процессом обучения фундаментальной биологии (ПК-5);

В результате освоения дисциплины обучающиеся должны

#### **знать:**

<b>З1(УК-1)</b>	методы критического анализа и оценки современных научных достижений, а также методы генерирования новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях
<b>З1(УК-2)</b>	методы научно- исследовательской деятельности
<b>З1(УК-3)</b>	особенности представления результатов научной деятельности в устной и письменной форме при

	работе в российских и международных исследовательских коллективах
<b>32(УК-4)</b>	стилистические особенности представления результатов научной деятельности в устной и письменной форме на государственном и иностранном языках
<b>31(УК-5)</b>	содержание процесса целеполагания профессионального и личностного развития, его особенности и способы реализации при решении профессиональных задач, исходя из этапов карьерного роста и требований рынка труда
<b>31(ОПК-1)</b>	основной круг проблем (задач), встречающихся в избранной сфере научной деятельности, и основные способы (методы, алгоритмы) их решения
<b>32(ОПК-1)</b>	основные источники и методы поиска научной информации
<b>31(ОПК-2)</b>	нормативно-правовые документы, регламентирующие организацию и содержание образовательного процесса
<b>31(ПК-1)</b>	современное состояние науки в области: (молекулярной биологии; биохимии; биотехнологии, микробиологии, биоинформатики)
<b>32(ПК-1)</b>	порядок организации, планирования и проведения научно-исследовательской работы с использованием современных научно-исследовательских, образовательных и информационных технологий
<b>33(ПК-2)</b>	методы поиска необходимой информации
<b>32(ПК-3)</b>	базовые принципы и основные приемы молекулярной биологии; биохимии; биотехнологии, бионанотехнологии, математической биологии, биоинформатики, микробиологии;
<b>31(ПК-5)</b>	современное состояние науки в области биологических наук

#### **уметь:**

<b>У1(УК-1)</b>	анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов
<b>У2(УК-1)</b>	при решении исследовательских и практических задач генерировать новые идеи, поддающиеся операционализации исходя из наличных ресурсов и ограничений
<b>У1(УК-2)</b>	использовать положения и категории философии науки для анализа и оценивания различных фактов и явлений
<b>У1(УК-3)</b>	следовать нормам, принятым в научном общении при работе в российских и международных исследовательских коллективах с целью решения научных и научно-образовательных задач
<b>У1(УК-4)</b>	следовать основным нормам, принятым в научном общении на государственном и иностранном языках
<b>У1(УК-5)</b>	формулировать цели личностного и профессионального развития и условия их достижения, исходя из тенденций развития области профессиональной деятельности, этапов профессионального роста, индивидуально-личностных особенностей.
<b>У1(ОПК-1)</b>	находить (выбирать) наиболее эффективные (методы) решения основных типов проблем (задач), встречающихся в избранной сфере научной деятельности
<b>У2(ОПК-1)</b>	обобщать и систематизировать передовые достижения научной мысли и основные тенденции хозяйственной практики
<b>У3(ОПК-1)</b>	анализировать, систематизировать и усваивать передовой опыт проведения научных исследований
<b>У4(ОПК-1)</b>	собирать, отбирать и использовать необходимые данные и эффективно применять количественные методы их анализа

<b>У2(ОПК-2)</b>	осуществлять отбор и использовать оптимальные методы преподавания оценивания успеваемости обучающихся в области биологических наук
<b>У1 (ПК-1)</b>	самостоятельно формулировать конкретные задачи научных исследований и проводить углубленную их разработку
<b>У-1(ПК-2)</b>	использовать современные методы сбора, анализа и обработки научной информации
<b>У-2(ПК-2)</b>	анализировать и систематизировать полученную информацию
<b>У1(ПК-3)</b>	проводить обработку результатов исследований
<b>У1(ПК-5)</b>	преподавать учебные предметы, курсы, дисциплины
<b>У2(ПК-5)</b>	разрабатывать научно- методическое обеспечение реализации курируемых учебных предметов, курсов, дисциплин

### **владеть:**

<b>В1(УК-1)</b>	навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях
<b>В2 (УК-1)</b>	навыками критического анализа и оценки современных научных достижений и результатов деятельности по решению исследовательских и практических задач, в том числе в междисциплинарных областях
<b>В1(УК-2)</b>	навыками анализа основных мировоззренческих и методологических проблем, в т.ч. междисциплинарного характера, возникающих в науке на современном этапе ее развития
<b>В2(УК-2)</b>	технологиями планирования в профессиональной деятельности в сфере научных исследований
<b>В1(УК-3)</b>	навыками анализа основных мировоззренческих методологических проблем, в.т.ч. междисциплинарного характера, возникающих при работе по решению научных и научно-образовательных задач в российских или международных исследовательских коллективах
<b>В4(УК-3)</b>	различными типами коммуникаций при осуществлении работы в российских и международных коллективах по решению научных и научно-образовательных задач
<b>В2(УК-4)</b>	навыками критической оценки эффективности различных методов и технологий научной коммуникации на государственном и иностранном языках
<b>В1(УК-5)</b>	приемами и технологиями целеполагания, реализации целей и оценки результатов деятельности по решению профессиональных задач
<b>В2(УК- 5)</b>	способами выявления и оценки индивидуально-личностных, профессионально-значимых качеств и путями достижения более высокого уровня их развития.
<b>В1(ОПК-1)</b>	современными методами, инструментами и технологией научно-исследовательской и проектной деятельности в определенных областях биологии
<b>В2(ОПК-2)</b>	методиками и технологиями преподавания и оценивания успеваемости обучающихся (биологические науки)
<b>В2 (ПК-1)</b>	методами и приемами экспериментальных исследований в области (молекулярной биологии; биохимии; биотехнологии, биоинформатике, микробиологии)
<b>В1(ПК-2)</b>	методами работы с основными базами данных биологической информации
<b>В1(ПК-3)</b>	навыками использования биологических Интернет-ресурсов
<b>В1(ПК-5)</b>	умениями разрабатывать научно-методическое обеспечение реализации курируемых учебных предметов, курсов, дисциплин

### **3. Объем дисциплины и виды учебной работы:**

Всего часов	Аудиторные	Самостоятельная	Всего
-------------	------------	-----------------	-------

	занятия (час):	работа	зачетных единиц
144	36	108	4

#### 4. Распределение аудиторных часов по темам и видам учебной работы

№ п/п	Наименование тем и разделов (с развернутым содержанием курса по каждой теме и разделу)	Аудиторные занятия (час), в том числе:
1	Основные структуры данных: хэш-таблица, суффиксное дерево, суффиксный массив	4
2	Быстрый поиск подстроки в строке	4
3	Индекс и преобразование Барроуза-Уиллера	4
4	BLAST	2
5	Мотивы в геномах	2
6	Алгоритмы динамического программирования	4
7	Приложения алгоритмов динамического программирования	4
8	Скрытые цепи Маркова	2
9	Основы Байесовской статистики	4
10	Методы оптимизации	2
11	Оценка параметров скрытой цепи Маркова	2
12	Методы функциональной аннотации генома	2
	Всего	36

#### 5. Содержание курса

##### Тема 1

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

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

##### Тема 2

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.

##### Тема 3

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.

#### **Тема 4**

##### **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). P-значение и E-значение. Битовый скор.

#### **Тема 5**

##### **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.

#### **Тема 6**

##### **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).

#### **Тема 7**

##### **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.

#### **Тема 8**

##### **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.

### **Тема 9**

#### **Основы Байесовской статистики**

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

### **Тема 10**

#### **Методы оптимизации**

Максимизация математического ожидания (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.

### **Тема 11**

#### **Оценка параметров скрытой цепи Маркова**

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

### **Тема 12**

#### **Методы функциональной аннотации генома**

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

## **6. Самостоятельная работа**

В процессе освоения дисциплины предусмотрено самостоятельное изучение отдельных вопросов лекционного курса с целью подготовки к устному опросу по теме.

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

### **7.1. Основная литература**

1. Льюин Б. Гены. М.: Бином. Лаборатория знаний. 2011.
2. Коничев А.С., Севастьянова Г.А. Молекулярная биология. М.: Академия, 2008.
3. Жимулев И.Ф. Общая и молекулярная генетика. – Новосибирск, Сибирское университетское издательство, 2003
4. Албертс Б. и др. Молекулярная биология клетки. В 3-х томах. М.: Мир, 1994.
5. Эллиот В., Эллиот Д. Биохимия и молекулярная биология. М., Академкнига, 2002.
6. Сингер М., Берг П. Гены и геномы. М., Мир, 1998.
7. Молекулярная биология. Структура и биосинтез нуклеиновых кислот (под ред. акад. Спирина А.С.). М., Высшая школа, 1990.
8. Патрушев Л.И. Экспрессия генов. М., Наука, 2000.
9. Antman S.S. Mathematical Physiology [electronic resource] / Antman S.S. и др. — New York, NY : Springer New York, 2009. — (Interdisciplinary Applied Mathematics, ISSN 0939-6047; v. 8/2). — DOI: 10.1007/978-0-387-79388-7. — ISBN 978-0-387-75846-6.

### **7.2. Дополнительная литература**

1. Патрушев Л.И. Искусственные генетические системы. В 2-х томах. Том 1. Генная и белковая инженерия. «Наука» Москва. 2004. 530 С.
2. Щелкунов С.Н. Генетическая инженерия. «Сибирское университетское издательство» Новосибирск. 2008. 514 С.
3. Писарчик А.В., Картель Н.А. Простые повторяющиеся последовательности и экспрессия генов. // Мол. биол. 2000. Т. 34. № 3. С. 357-362.
4. Эпигенетика. Под ред. С.Д. Эллиса, Т. Дженювейна, Д. Рейнберга - М.: Техносфера, 2010. – 496 с.
5. Разин С.В., Быстрицкий А.А. Хроматин: упакованный геном. - БИНОМ. Лаборатория знаний, 2009. – 176 с.
6. Коряков Д.Е., Жимулев И.Ф. Хромосомы. Структура и функции. - Новосибирск: Изд-во СО РАН, 2009. – 258 с.
7. Гасфилд, Д. Строки, деревья и последовательности в алгоритмах. Информатика и вычислительная биология, Издательство Невский диалект, Санкт-Петербург, 2003.
8. Иванов Петр Мацович. Алгебраическое моделирование сложных

систем. — М. : Наука. Физматлит, 1996. — 271 с. : ил. — Рез. англ. Библиогр. : с. 265-268. — ISBN 5-02-015218-8.

### 7.3. Электронные ресурсы

1. <http://www.benran.ru/> - Библиотека по естественным наукам Российской академии наук
2. <https://apps.webofknowledge.com/> - Научно-библиографическая база данных Web of Science.
3. <http://www.scopus.com/> - Научно-библиографическая база данных Scopus.
4. <http://elibrary.ru/> - Научная электронная библиотека НЭБ.
5. <http://www.rsl.ru/> - Электронная библиотека РГБ.
6. <http://www.diss.rsl.ru/> - Электронная библиотека диссертаций РГБ.
7. <http://www.sciencedirect.com/> - Журналы издательства Elsevier.
8. <http://link.springer.com/> - Журналы издательства Springer.
  - a) <http://www.springerprotocols.com> - SpringerProtocols
  - b) <http://www.springermaterials.com> - SpringerMaterials
  - c) <http://www.springerimages.com> - SpringerImages
  - d) <http://www.zentralblatt-math.org/zbm/zbmath/en> - Zentralblatt MATH
9. <http://link.springer.com/> - Архивные материалы на платформе Springer.
  - a) Журналы (Journals) 1832-1996 и 2002-2011 гг.
  - b) Журналы (Journals) 1997-2001 гг.
  - c) Книги (Books) 2005-2010 гг., включая книжные серии и справочники.
  - d) Книжные серии (Book Series) 1902-1996 гг.
  - e) Книжные серии (Book Series) 2005-2010 гг.
  - f) Электронные справочники (E-References) 2005-2010 гг.
10. <http://xlink.rsc.org?genre=journal&eissn=1364-548X&date=1996> - Chemical Communications (Cambridge)
11. <http://xlink.rsc.org?genre=journal&eissn=1460-4744&date=1972> - Chemical Society Reviews
12. <http://xlink.rsc.org?genre=journal&eissn=1477-9234&date=2003> - Dalton Transactions
13. <http://xlink.rsc.org?genre=journal&eissn=1364-5501&date=1991> - Journal of Materials Chemistry
14. <http://xlink.rsc.org?genre=journal&eissn=2050-7496&date=2012> - Journal of Materials Chemistry A
15. <http://xlink.rsc.org?genre=journal&eissn=2050-7518&date=2013> - Journal of Materials Chemistry B
16. <http://xlink.rsc.org?genre=journal&eissn=2050-7534&date=2013> - Journal

of Materials Chemistry C

17. <http://xlink.rsc.org?genre=journal&eissn=1463-9084&date=1999> - Physical Chemistry Chemical Physics
18. <http://pubs.rsc.org/en/journals/journalissues/ob#!recentarticles&all> - Organic & Biomolecular Chemistry
19. <http://journals.cambridge.org/> - Журналы издательства Cambridge University Press.
20. <http://www.oxfordjournals.org/en/> - Журналы издательства Oxford University Press.
21. <http://onlinelibrary.wiley.com/> - Журналы издательства Wiley.
22. <http://pubs.acs.org/> - American Chemical Society.
23. <http://www.nature.com/> - Журнал «Nature» (и другие журналы группы Nature).
24. [www.sciencemag.org](http://www.sciencemag.org) - Журнал «Science».
25. <http://www.fips.ru/> — Патентная база данных РФ (РОСПАТЕНТ).
26. <http://www.uspto.gov/> - Патентная база данных США (USPATFULL).
27. <http://arxiv.org> - arXiv.org/ - международный архив электронных научных статей.
28. <http://www.ccdc.cam.ac.uk/> - Кэмбриджская база структурных данных органических и металлоорганических соединений

#### **7.4. Лицензионное программное обеспечение**

1. Office Professional
2. Project Professional
3. Visio Professional
4. Windows
5. Exchange Server Standard CAL - Device CAL

#### **8. Составители программы:**

к.б.н. Ю.А. Медведева