

«УТВЕРЖДАЮ»
Декан биологического факультета МГУ
Академик **М.П.Кирпичников**
«__» _____ 2015 г.



Рабочая программа дисциплины (модуля)

1. Код и наименование дисциплины (модуля): **«Биоинформатика»**
2. Уровень высшего образования – подготовка научно-педагогических кадров в аспирантуре.
3. Направление подготовки – **06.06.01 Биологические науки**. Направленность (профиль) программы – **Молекулярная биология**.
4. Место дисциплины (модуля) в структуре ООП: вариативная часть ООП (осенний семестр), спецкурс по выбору (читается на кафедре молекулярной биологии)
5. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции (код компетенции)	Планируемые результаты обучения по дисциплине (модулю)
<p>УК-1: Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях</p>	<p>Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях Код В1 (УК-1) Владеть: навыками критического анализа и оценки современных научных</p>

6. Объем дисциплины (модуля) составляет 2 зачетных единицы, всего 72 академических часа, из которых 24 часа составляет контактная работа аспиранта с преподавателем (24 часа занятий лекционного типа) и 48 часов составляет самостоятельная работа аспиранта (самостоятельное изучение научной литературы по проблеме и написание аналитического обзора).

7. Входные требования для освоения дисциплины (модуля), предварительные условия:

ЗНАТЬ: неорганическую и органическую химию, физическую химию, биохимию, молекулярную биологию, основы клеточной биологии (на уровне программ специалиста/магистра), теоретические и методологические основы биологических научных исследований

УМЕТЬ: вырабатывать на основе рационального анализа литературных данных и экспериментальных результатов свою точку зрения в вопросах эпигенетики и отстаивать ее во время дискуссии со специалистами и неспециалистами; читать и реферировать научную литературу в области эпигенетики, в том числе на английском языке, при условии соблюдения научной этики и авторских прав.

ВЛАДЕТЬ: современными информационно-коммуникационными технологиями, английским языком.

8. Образовательные технологии: классические лекционные технологии.

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

	<p>достижений и результатов деятельности по решению исследовательских и практических задач, в том числе в междисциплинарных областях Код В2 (УК-1)</p> <p>Знать: методы научно-исследовательской деятельности Код 31 (УК-2)</p>
<p>УК-2 <i>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.</i></p> <p>УК-3: <i>Готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач</i></p> <p>УК-4: <i>Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языке</i></p>	<p>Владеть: технологиями оценки результатов коллективной деятельности по решению научных и научно-образовательных задач, в том числе ведущейся на иностранном языке Код В2 (УК-3)</p> <p>Владеть: навыками анализа научных текстов на государственном и иностранном языках Код В1 (УК-4)</p> <p>Знать: стилистические особенности представления результатов научной деятельности в устной и письменной форме на государственном и иностранном языках Код 32 (УК-4)</p>
<p>ОПК-1 <i>Способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий</i></p>	<p>Уметь: собирать, отбирать и использовать необходимые данные и эффективно применять количественные методы их анализа</p>

Оценочные средства для промежуточной аттестации по дисциплине (модулю) приведены в Приложении.

6. Объем дисциплины (модуля) составляет 2 зачетных единицы, всего 72 академических часа, из которых 24 часа составляет контактная работа аспиранта с преподавателем (24 часа занятий лекционного типа) и 48 часов составляет самостоятельная работа аспиранта (самостоятельное изучение научной литературы по проблеме и написание аналитического обзора).

7. Входные требования для освоения дисциплины (модуля), предварительные условия:

ЗНАТЬ: неорганическую и органическую химию, физическую химию, биохимию, молекулярную биологию, основы клеточной биологии (на уровне программ специалиста/магистра), теоретические и методологические основы биологических научных исследований

УМЕТЬ: вырабатывать на основе рационального анализа литературных данных и экспериментальных результатов свою точку зрения в вопросах эпигенетики и отстаивать ее во время дискуссии со специалистами и неспециалистами; читать и реферировать научную литературу в области эпигенетики, в том числе на английском языке, при условии соблюдения научной этики и авторских прав.

ВЛАДЕТЬ: современными информационно-коммуникационными технологиями, английским языком.

8. Образовательные технологии: классические лекционные технологии.

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

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы					Самостоятельная работа обучающегося, часы			
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости коллоквиумы, практические контрольные занятия и др)*	Всего	Выполнение домашних заданий	Подготовка рефератов и т.п.	Всего
Выравнивания. Основные алгоритмы, поиск гомологии в биоинформатике. Основы теории алгоритмов. Поиск и сортировка. Редакционное расстояние. Гомология. Оценка выравниваний, статистическая значимость. Матрицы замен. Парные и множественные выравнивания. Алгоритмы Smith-Waterman, FASTA, BLAST. Множественное выравнивание, ClustalW2, MAFFT, SAGA, T-Coffee. Филогенетический анализ.	18	8					8	6	5	11
Анализ данных NGS. Обзор технологий	24	8					8	24	5	29

NGS (Illumina, 454, Solid). Первичная обработка данных: тримминг, картирование. Variant calling. Сборка. Аннотация кодирующих и некодирующих областей. Дифференциальная экспрессия. Методы анализа структуры генома: ChipSeq, Hi-C, поиск хромосомных транслокаций.										
Работа в командной строке ОС Linux. Основные понятия. Работа с файловой системой и потоками ввода\вывода. Утилиты grep, sed, awk. Переменные среды. Работа с биоинформатическими программами (bowtie2, STAR, samtools, bedtools, augustus, velvet и др.). Обзор основных форматов хранения данных в биоинформатике.	12	4					4	14	5	19
Программирование. Основные понятия. Язык Python. Работа с файлами. Переменные, функции, обзор библиотек. Biopython. Работа с базами данных ncbi при помощи biopython.	18	8					8	16	5	21
Промежуточная аттестация - зачет										
Итого:	108	28					28	60	20	80

10. Учебно-методические материалы для самостоятельной работы аспирантов.

Конспекты лекций, аудио- и видеозаписи лекций, файлы презентаций лекций, основная и дополнительная учебная литература (см. п.11)

11. Ресурсное обеспечение:

Основная литература

Jean-Michel Claverie, Cedric Notredame. «Bioinformatics For Dummies, 2nd Edition».

Гасфилд Дэн. «Строки, деревья и последовательности в алгоритмах». СПб.: Невский Диалект; БХВ-Петербург, 2003.

Артур Леск. «Введение в биоинформатику».

Дополнительная литература

Р.Дурбин, Ш.Эдди, А.Крог и Г.Митчисон. «Анализ биологических последовательностей»

Евгений Кунин. «Логика случая. О природе и происхождении биологической эволюции»

Жуан Сетубал, Жуан Мейданис. «Введение в вычислительную молекулярную биологию»

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

<https://www.coursera.org/>

(курсы по специализации «биоинформатика» <https://www.coursera.org/courses?query=bioinformatics>)

Подборка статей по алгоритмам из Wikipedia.

Подборка статей из Pubmed, посвященных основным темам курса.

Группа курса <https://vk.com/bioinformatics2015>

Перечень используемых информационных технологий, используемых при осуществлении образовательного процесса, включая программное обеспечение, информационные справочные системы (при необходимости):

Базы данных PubMed (NCBI, <http://www.ncbi.nlm.nih.gov/pubmed>)

Группа ВКонтакте <https://vk.com/bioinformatics2015> (презентации лекций, рекомендуемые дополнительные материалы).

Доступ к вычислительным ресурсам (кластеру с предустановленным необходимым ПО) для выполнения домашних заданий.

Описание материально-технической базы.

Кафедра молекулярной биологического факультета МГУ располагает необходимым аудиторным фондом, компьютерами, проекторами и экранами, аудиоаппаратурой.

12. Язык преподавания: русский

13. Преподаватель (преподаватели): научный сотрудник кафедры молекулярной биологии Герасимов Е.С.

(Егор Герасимов Е. С.)

Приложение

**Оценочные средства для промежуточной аттестации по дисциплине (модулю) «хроматин и регуляция транскрипции»
на основе карт компетенций выпускников**

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модулю)	КРИТЕРИИ и ПОКАЗАТЕЛИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ по дисциплине (модулю), баллы БРС					ОЦЕНОЧНЫЕ СРЕДСТВА
	1,	2	3	4	5	
Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях Код В1 (УК-1)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, зачет
Владеть: навыками критического анализа и оценки современных научных достижений и результатов деятельности по решению исследовательских и практических задач, в том числе в междисциплинарных областях Код В2 (УК-1)	0	1-29	30-59	60-89	90-100	-- индивидуальное собеседование, реферат, зачет
Знать: методы научно-исследовательской деятельности Код 31(УК-2)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, зачет
Владеть:	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат,

технологиями оценки результатов коллективной деятельности по решению научных и научно-образовательных задач, в том числе ведущейся на иностранном языке Код В2(УК-3)						<i>зачет</i>
Знать: стилистические особенности представления результатов научной деятельности в устной и письменной форме на государственном и иностранном языках Код З2(УК-4)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>
Владеть: навыками анализа научных текстов на государственном и иностранном языках Код В1(УК-4)	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>
Уметь: собирать, отбирать и использовать необходимые данные и эффективно применять количественные методы их анализа	0	1-29	30-59	60-89	90-100	- индивидуальное собеседование, реферат, <i>зачет</i>

Фонды оценочных средств, необходимые для оценки результатов обучения

Примеры вопросов к промежуточному контролю (темы рефератов, вопросы для индивидуального собеседования):

Вопросы зачета:

- 1) Какие алгоритмические подходы применяются для картирования коротких чтений?
- 2) В чем технические особенности платформы Illumina в сравнении с 454 Roche?
- 3) Какие основные принципы работы геномных сборщиков на примере Newbler (GS denovo Assembler) и Velvet.
- 4) Опишите принципиальную схему алгоритма BLAST.
- 5) Алгоритм Смита-Ватермана, примеры выравниваний.
- 6) Методы построения филогенетический деревьев.
- 7) Какие методы оценки качества сборки можно предложить?
- 8) Что такое ортологи и паралоги?
- 9) В чем отличие paired-end от mate-pair чтений?
- 10) Какие основные эвристические подходы применяются для получения множественных выравниваний?
- 11) В чем основная суть алгоритмов сборки генома? Какие трудности им приходится решать?
- 12) В чем задача аннотации? Какие основные подходы к ее решению?
- 13) Назовите принципы действия программ-картировщиков.
- 14) Как оценивают качество выравнивания? О чём оно говорит?

Практические задачи:

- 1) Построить филогенетическое дерево имеющихся последовательностей, охарактеризовать дерево
- 2) Указать наиболее оптимальную пару праймеров для амплификации заданного фрагмента ДНК
- 3) Охарактеризовать фрагмент неизвестной молекулы ДНК (возможная видовая принадлежность, кодирующая\некодирующая и т.д.)
- 4) По имеющейся аннотации построить гистограммы распределения характерных длин гена, интрона, экзона.
- 5) Постройте выравнивание двух последовательностей алгоритмом Смита-Ватермана (не используя компьютер).
- 6) Провести контроль качества полученных данных секвенирования, при необходимости провести процедуры тримминга по качеству.

ПРОГРАММА зачета по спецкурсу «Биоинформатика»

ВЫРАВНИВАНИЯ.

Биологический смысл выравниваний, парные выравнивания. Локальное и глобальное выравнивание. Поиск оптимального выравнивания методом динамического программирования. Алгоритмы Нидлмана-Вунш и Смита-Ватермана. Эвристические алгоритмы парного выравнивания: FASTA и BLAST. Применение вариаций алгоритма BLAST (blastn, blastp, blastx, tblastn, tblastx) для поиска информации на серверах NCBI. Оценка выравниваний. Другие применения алгоритмов выравнивания. Множественные глобальные и локальные выравнивания. Средняя генетическая дистанция. Применение множественных выравниваний. Программы T-Coffee, MUSCLE, ClustalW, MSA, MAFFT. Принципы построения филогенетических деревьев. Укорененные и неустановленные деревья. Outgroup. Форматы, связанные с представлением множественных выравниваний. Программа SeaView. Филогенетика. Пакет программ PHYLIP. Построение филогенетических деревьев с помощью пакета программ MEGA5. Анализ филогенетического дерева, терминология, особенности. Метод bootstrap.

АНАЛИЗ ДАННЫХ NGS.

Анализ данных секвенирования нового поколения. Формат FASTQ. Основные понятия, связанные с технологией NGS (типа библиотек, качество прочтения, принципы первичной обработки данных). Тrimming по качеству, обрезание адаптера. Картрирование коротких чтений на типовой (референсный) геном. Алгоритмы на основе суффиксных деревьев. Картрировщики STAR и bowtie2. Формат sam\bam. Пакет samtools.

SNP calling - поиск полиморфизмов. Поиск крупомасштабных перестроек в геноме (фьюженов).

Методы изучения взаимодействий на полногеномном уровне (ChIPSeq, Hi-C).

Анализ дифференциальной экспрессии генов, альтернативный сплайсинг. Особенности работы с РНК-данными.

Сборка геномов. Сборка транскриптомов. Алгоритмы, лежащие в основе геномных сборщиков SGA, GS DeNovo Assembler, Velvet. Графы Дебрейна. Эйлеров и Гамильтонов пути. k-мерные и olce сборщики.

Аннотация. Принципы поиска генов в геноме. ORF finder. Аннотация ab initio. Augustus, GeneMark, GeneWise. Аннотация при наличии референсного генома. Синтезия. Перенос аннотации с помощью RATT и MUMMER.

ОС LINUX

Основные понятия. Принципы работы операционных систем. Файл, файловая система. Работа с файлами и папками. Работа с потоками ввода и вывода. Потоки и процессы. Устройство памяти, иерархия памяти. Основные утилиты командной строки. Поиск паттернов с помощью grep. Способы фильтрации данных. Язык awk. Использование awk-скриптов.

ПРОГРАММИРОВАНИЕ

Способы визуализации данных. Python. Работа с файлами. Функции, переменные, типы данных. Списки, словари, кортежи. Использование библиотек. Biopython. Использование библиотеки biopython для решения ряда биоинформатических задач.