

Федеральное бюджетное учреждение науки  
«ГОСУДАРСТВЕННЫЙ НАУЧНЫЙ ЦЕНТР ВИРУСОЛОГИИ И  
БИОТЕХНОЛОГИИ «ВЕКТОР»  
Федеральной службы по надзору в сфере защиты прав потребителей и  
благополучия человека  
(ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора)



УТВЕРЖДАЮ  
Генеральный директор  
ФБУН ГНЦ ВБ «Вектор»  
Роспотребнадзора  
Р.А. Максютов

*«08» сентября 2017 г.*

**Рабочая программа дисциплины  
Основы биоинформатики**

Направление подготовки:  
06.06.01 – Биологические науки

Направленность (профиль)  
03.01.03 – Молекулярная биология

Квалификация выпускника  
Исследователь. Преподаватель-исследователь

Форма обучения: очная

**Кольцово 2017**

## СОДЕРЖАНИЕ

1.	Цель освоения дисциплины.....	3
2.	Место дисциплины в структуре образовательной программы.....	3
3.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.....	4
4.	Объем дисциплины и виды учебной работы.....	4
5.	Содержание дисциплины и трудоемкость по видам учебных занятий.....	4
6.	Самостоятельная работа обучающихся.....	9
7.	Формы проведения занятий.....	10
8.	Фонд оценочных средств.....	10
	8.1. Паспорт фонда оценочных средств .....	10
	8.2. Промежуточная аттестация (зачет).....	12
	8.3. Критерии оценивания.....	14
9.	Учебно-методическое и информационное обеспечение дисциплины.....	14
	9.1. Основная литература.....	14
	9.2. Дополнительная литература.....	14
	9.3. Перечень ресурсов информационно-телекоммуникационной сети Интернет, необходимых для освоения дисциплины.....	14
10.	Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине, в т.ч. программное обеспечение.....	15
11.	Материально-техническое обеспечение дисциплины.....	15

## 1. Цель освоения дисциплины

Целью освоения дисциплины является подготовка специалистов высшей квалификации для фундаментальной и прикладной науки в области клеточной биологии, молекулярной биологии, биотехнологии, иммунологии и генной инженерии, обладающих современными теоретическими знаниями в области биоинформатики, способных формулировать научные и прикладные задачи и предлагать пути их решения, нацеленных на непрерывное совершенствование и развитие своего научного потенциала.

Целью курса является приобретение навыков самостоятельной работы с современными инструментами решения биоинформатических задач, способов поиска и получения дополнительной информации из открытых источников биологических данных; изучение основных публичных источников биоинформатических данных; освоение современных инструментов для работы с нуклеотидными и аминокислотными последовательностями с геномными и протеомными данными, включая интернет-ресурсы и программное обеспечение для серверов и персональных компьютеров; изучение способов визуализации данных биологических исследований с целью анализа, интерпретации и публикации получаемых результатов; изучение базовых методов обработки результатов высокопроизводительного секвенирования нового поколения; изучение основ программирования, обработки и визуализации данных с помощью высокоуровневых языков программирования Python и R с использованием специализированных библиотек.

## 2. Место дисциплины в структуре образовательной программы

Дисциплина отнесена к вариативной части Блока 1 «Дисциплины (модули)» Б1.В.ДВ.3 программы подготовки научно-педагогических кадров в аспирантуре по направлению подготовки 06.06.01 Биологические науки.

Для освоения данной дисциплины обучаемый должен:

**знать:** основы биохимии, молекулярной биологии и генетики, строение нуклеиновых кислот, строение белков, принципы матричного синтеза биополимеров, процессы репликации, транскрипции и трансляции, устройство генетического кода, принципы организации и функционирования генов, понятие генома, основы структурно-функциональной организации геномов, принципы регуляции активности генов и передачи информации в живой клетке;

**уметь:** ориентироваться в научной литературе, отечественной и зарубежной; критически оценивать методы для решения научных задач;

**владеть:** основами компьютерной грамотности; навыками поиска научной информации в сети интернет, участия в научной дискуссии, принятия независимых суждений и самостоятельных решений.

Дисциплина Основы биоинформатики изучается в третьем семестре второго курса аспирантуры. Изучение дисциплины опирается на знания в области биологии, химии, физики, математики и специальных дисциплин направления подготовки Биологические науки, освоенных аспирантами на предшествующих этапах обучения.

### 3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы аспирантуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

ОПК-1 — способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий.

### 4. Объем дисциплины и виды учебной работы

Объем дисциплины – 2 зачетных единицы (ЗЕ) или 72 академических часов.

Вид учебной работы		Всего часов
Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)		48
Аудиторные занятия (всего)		48
в том числе:		
лекции (Л)		24
практические занятия (ПЗ), семинары (С)		24
лабораторные работы (ЛР)		
практикумы (ПР)		
Внеаудиторная работа (всего)		
в том числе:		
индивидуальная работа обучающихся с преподавателем		
Консультации		
Самостоятельная работа обучающихся (СР) (всего)		24
в том числе: реферат		
Вид промежуточной аттестации зачет (З), экзамен (Э)		зачет
Общая трудоемкость	Часов	72
	зачётных единиц	2

### 5. Содержание дисциплины и трудоемкость по видам учебных занятий

Содержание раздела		
№ п/п	Наименование темы	Объем, час
1	2	3
Лекции		
1.	Введение в биоинформатику. Предмет биоинформатики, биоинформатика как наука и прикладная дисциплина. История биоинформатики. Разделы биоинформатики. Биоинформатические методы и их связь с другими методами биологических исследований. Примеры биоинформатических задач.	2

2.	<p>Работа с последовательностями. Способы кодирования и записи нуклеотидных и аминокислотных последовательностей. Алфавиты IUPAC. Основные форматы записи последовательностей.</p> <p>Понятнее о выравнивании последовательностей. Парное и множественное выравнивание последовательностей. Локальное и глобальное выравнивание. Меры близости произвольных последовательностей символов. Меры качества выравнивания.</p> <p>Понятие референсной последовательности. Программы FASTA и BLAST. Применение различных вариаций алгоритма BLAST (blastn, blastp, blastx, tblastn, tblastx). Программы T-Coffee, MUSCLE, ClustalW, MSA, MAFFT. Форматы представления множественных выравниваний. BioEdit.</p> <p>Понятие о филогении и филогенетике. Филогенетические деревья, их устройство, виды и топологические свойства. Способы кодирования и отображения филогенетических деревьев. Методы обработки и сравнения филогенетических деревьев. Пакет MEGA.</p>	2
3.	<p>Современные инструменты реконструкции филогенетических деревьев: PhyML и MrBayes. Способы формальной записи филогенетических деревьев, форматы файлов. Способы визуализации филогенетических деревьев. Современные инструменты визуализации филогенетических деревьев: Archaeopteryx, TreeGraph 2, ETE Toolkit.</p> <p>Понятие эволюционного расстояния. Параметрические и эмпирические модели нуклеотидных и аминокислотных замен. Основные модели нуклеотидных замен: GTR, JC, HKY. Основные модели аминокислотных замен: BLOSUM, PAM, JTT. Гипотеза молекулярных часов, гипотеза нейтральности. Проверка эволюционных гипотез. Гомология последовательностей. Ортологи и паралоги.</p>	2
4.	<p>Секвенирование РНК и ДНК. История методов секвенирования. Обзор современных инструментов для секвенирования.</p> <p>Современные платформы секвенирования данных нового поколения.</p> <p>Понятие сборки генома, уровни сборки геномов. Сравнительная геномика. База данных геномов: NCBI Genome, UCSC. Способы идентификации, описания и хранения геномов. Интерфейс UCSC Genome Browser. Понятие геномных координат, способы их записи и преобразования. Поиск и скачивание геномов и геномных данных.</p> <p>Способы идентификации и кодирования генов и белков. Текстовые описания генов. Основные способы идентификации генов модельных организмов. Идентификация генов немодельных и неизученных организмов. Инструменты преобразования идентификаторов генов.</p> <p>Типы данных секвенирования нового поколения: Exomes, WGS,</p>	2

	RNA-Seq, ChIP-Seq. Задачи, связанные с данными секвенирования нового поколения. Анализ дифференциальной экспрессии генов, альтернативный сплайсинг.	
5.	<p>Форматы файлов для хранения информации о геномных элементах. Аннотация геномных элементов. Работа с биоинформатическими программами bedtools.</p> <p>Примеры пайплайнов обработки данных секвенирования нового поколения. Визуализация генов и геномов в Integrative Genomics Viewer (IGV). Высокоуровневая визуализация геномов с помощью инструмента Circos.</p>	2
6.	<p>Уровни организации белковых структур. Способы кодирования информации о пространственной структуре белковых молекул и соответствующие форматы файлов. RSCB PDB Инструменты для отображения и редактирования структур белков. Swiss PDB Viewr, NOCh, PyMOL, Chimera. Пример анализа структур с помощью PyMOL. Предсказание вторичной структуры белков. PSIPRED. Построение моделей пространственных структур белков по гомологии (Modeller, Swiss-Model). Моделирование пространственных структур белков de novo. Rosetta, I-TASSER. GeneSilico Metaserver. Задачи, связанные со структурами белков. Молекулярный докинг. Autodock Tools. Молекулярная динамика. VTK и NAMD. GROMAX.</p>	2
7.	<p>Основы поиска информации в сети интернет. Работа с базами данных. Базы данных National Center for Biotechnology Information (NCBI). Инструменты NCBI. EBI. UniProt. Иммунологическая база данных IEDB. Анализ текстовой информации, аннотаций, чтение различных форматов данных. 1000 геномов. Информация о нуклеотидных полиморфизмах, их распространенности и функциональной значимости: dbSNP. dbNSFP, ClinVar, SNPedia.</p>	2
8.	<p>Понятие о формальных онтологиях. Примеры онтологий для биомедицинских исследований: Gene Ontology, Uberon, Chemical Entities of Biological Interest (ChEBI), Cell Type Ontology (CL), Human Phenotype Ontology, RxNorm.</p> <p>Организация формальной онтологии Gene Ontology. Инструмент AmiGO 2 для доступа к Gene Ontology. Поиск биологических процессов, клеточных компонент, молекулярных функций и других терминов в Gene Ontology. Поиск и визуализация информации о генах. Понятие об обогащении терминами в онтологиях. Алгоритмы и инструменты обогащения терминами в Gene Ontology.</p> <p>Математическое понятие графа. Некоторые виды и свойства графов. Понятие сети в системной биологии, примеры сетей. Регуляторные сети. Сети белок-белковых взаимодействий. Метаболические сети. База данных Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways как источник информации о</p>	2

	метаболических и генных сетях. Интерфейс KEGG Pathways. Виды метаболических путей в KEGG Pathways. Примеры метаболических путей и способов их отображения. Введение в KEGG Atlas. Другие базы данных: Reactome, STRING.	
9.	Веб-сервисы для решения биоинформатических задач (инструменты NCBI, EBI). Сервисы для анализа пространственных структур. Инструменты для расчета олигонуклеотидных праймеров: Oligo, FastPCR, Primer3. Базы данных векторных последовательностей для клонирования (AddGene, SnapGene, PLASMID). Программы с графическим пользовательским интерфейсом для анализа последовательностей и клонирования. BioEdit, MEGA, UGENE (UniPro), SnapGene Viewer, GenBeans.	2
10.	Основы работы в командной строке. Работа с файловой системой и потоками ввода/вывода. Переменные среды. Работа с программами с интерфейсом командной строки. Понятие пайплайна. Устройство и реализация биоинформатических пайплайнов. Примеры пайплайнов и инструментов для их реализации.	2
11.	Язык программирования Python. Переменные, структуры данных (списки, кортежи, словари, массивы), функции. Основные модули и научные библиотеки Python: numeric, scipy, pandas, statsmodel, Biopython. Обзор основных форматов хранения данных в биоинформатике. Anaconda. Среда разработки Spyder. Интерактивная работа в среде Jupyter.	2
12.	Язык программирования R. Среда разработки RStudio. CRAN и Bioconductor. Основы работы с данными в R. Визуализация данных. Библиотека ggplot2.	2
	<b>Итого</b>	24
<b>Практические занятия и семинары</b>		
1.	Работа с последовательностями. Анализ нуклеотидных последовательностей. Построение (обратно) комплементарных последовательностей. Транскрипция и трансляция нуклеотидных последовательностей. Аминокислотные последовательности. Подбор олигонуклеотидных праймеров для ПЦР. Способы записи последовательностей. Построение парных и множественных выравниваний. Построение филогенетических деревьев. Используем UGENE.	2
2.	Работа с сервисами NCBI, EBI, ExPASy. GenBank. UniProt. Работа с BLAST. Запуск BLAST и поиск гомологичных последовательностей на NCBI. Локальная установка и запуск BLAST. Используем UGENE.	2
3.	Работа с данными секвенирования нового поколения. Выравнивание, поиск нуклеотидных полиморфизмов, анализ дифференциальной экспрессии генов. Работа в UGENE. Обзор	2

	различных методов работы с данными секвенирования нового поколения.	
4.	Работа с GO, KEGG, Reactome. Анализ групп генов на обогащение терминами (GSEA - gene set enrichment analysis). Используем UGENE.	2
5.	Работа по анализу пространственных структур белков. Визуализация пространственных структур (Swiss PDB Viewer, PyMOL). Поиск гомологичных структур. Структурное выравнивание. Работа с RSCB PDB. Моделирование структуры белка по гомологии (Modeller). Оценка качества моделей (PROCHECK, PDBsum).	2
6.	Основы поиска биомедицинской информации в сети интернет. Поиск литературы. Обзор информационных ресурсов. Системы для поиска реактивов (BioCompare.com). Основы работы с менеджерами ссылок (Mendeley Desktop).	2
7.	Основы работы в командной строке. Работа с файловой системой и потоками ввода/вывода. Переменные среды. Работа с программами с интерфейсом командной строки. Понятие пайплайна. Устройство и реализация биоинформатических пайплайнов. Примеры пайплайнов и инструментов для их реализации.	2
8.	Язык программирования Python. Переменные, структуры данных (списки, кортежи, словари, массивы), функции. Основные модули и научные библиотеки Python: numeric, scipy, pandas, statsmodel, Biopython. Обзор основных форматов хранения данных в биоинформатике. Anaconda. Среда разработки Spyder. Интерактивная работа в среде Jupyter.	2
9.	Работа с последовательностями в Biopython. Нуклеотидные последовательности и (обратно) комплементарные. Транскрипция. Трансляция. Таблицы трансляции. Доступ к базам данных Entrez от NCBI с помощью Biopython. Поиск и анализ записей Medline, GEO, UniGene, GenBank. Чтение файлов с последовательностями. Разбор записей в формате GenBank, в формате SwissProt. Запись файлов с последовательностями. Работа с выравниваниями в Biopython. Запуск BLAST через Интернет. Запуск BLAST локально. Анализ результатов BLAST.	2
10.	Работа с модулем Bio.Restriction – поиск сайтов рестрикции. Поиск ферментов по сайтам разрезания. Поиск изошизомеров. Подбор олигонуклеотидных праймеров для ПЦР. Планирование генноинженерных работ in silico.	2
11.	Основы языка программирования R. Знакомство со средой разработки RStudio. CRAN и Bioconductor. Основы работы с данными в R. Визуализация данных. Библиотека ggplot2.	2
12.	Использование методов машинного обучения для решения биологических задач. Обзор специализированного программного	2

	обеспечения: RapidMiner, Orange. Использование специализированных программных библиотек scikit-learn (Python), caret (R).	
	<b>Итого</b>	24
	<b>Всего</b>	48

### 6. Самостоятельная работа обучающихся

Аспиранты могут выполнять необходимую при изучении дисциплины самостоятельную работу в читальных залах ГПНТБ СО РАН, в читальном зале библиотеки ГНЦ ВБ «Вектор» Роспотребнадзора, в учебных кабинетах, на рабочих местах и на дополнительно оборудованных стационарных местах с выходом в Интернет, а также в домашних условиях.

№ п/п	Наименование вида самостоятельной работы	Трудоемкость, часы
1	2	3
1.	Поиск нуклеотидных и аминокислотных последовательностей в GenBank. Поиск гомологичных последовательностей с помощью BLAST. Построение филогенетического дерева.	2
2.	Подбор олигонуклеотидных праймеров для клонирования заданного целевого фрагмента. Планирование экспериментальных работ по встройке целевого фрагмента в плазмидный вектор.	2
3.	Построение модели пространственной структуры белка по гомологии. Визуализация и анализ полученной модели.	4
4.	Работа с тренировочным набором транскриптомных данных. Построение выравнивания. Построение таблицы экспрессии генов.	4
5.	Работа с тренировочным набором транскриптомных данных (продолжение). Поиск дифференциально экспрессированных генов. Анализ списка генов на обогащение функциональными GO-категориями.	4
6.	Практическое закрепление навыков работы в командной строке.	2
7.	Практическое закрепление навыков программирования на языке Python и работы с библиотекой Biopython.	2
8.	Подготовка к зачету	4
	<b>ИТОГО:</b>	<b>24</b>

Для обеспечения самостоятельной работы аспиранта наиболее рациональным ресурсом является сеть интернет, поскольку на сайтах постоянно идет обновление информации, и пользователь (аспирант) может получить актуальную информацию по интересующему его вопросу. Вся информация, необходимая для выполнения домашних заданий дана в рамках лекций и семинаров и содержится в презентациях к данному курсу.

Самостоятельная работа выполняется аспирантами по заданию преподавателя.

## 7. Формы проведения занятий

Преподавание курса ведется в виде лекций и семинаров. Начиная со второго занятия, в его начале проводится 5-минутный тест на знание материала предыдущей лекции. Тест состоит из двух - трех теоретических вопросов, (обычно это определения терминов и понятий, разобранных на предыдущей лекции).

Аудиторные занятия проводятся в интерактивной форме с использованием мультимедийного обеспечения (компьютер, проектор) и технологии проблемного обучения. Презентации позволяют качественно иллюстрировать практические занятия схемами, формулами, рисунками. Кроме того, презентации позволяют четко структурировать материал занятия. Электронная презентация позволяет отобразить процессы в динамике, что позволяет улучшить восприятие материала.

Обратная связь с аудиторией обеспечивается тем, что лектор отвечает на все вопросы, возникшие при прослушивании лекции. Такая форма преподавания позволяет гибко подходить к модификации лекционного курса. В случае возникновения каких-то трудностей в усвоении материала со стороны студентов лектор посвятит время более детальному разбору возникших по лекции вопросов. Каждая лекция содержит элементы диалога преподавателя со студентами, поскольку каждый из участников – студенты или преподаватель имеют право задавать вопросы в ходе лекции и участвовать в ее обсуждении.

Семинары проводятся с применением технологии проблемного обучения: каждый семинар начинается с постановки определенной задачи, последовательному разбору решения данной задачи и интерпретации полученных результатов. Такая форма проведения семинаров полностью отвечает целям освоения дисциплины и формирует необходимые компетенции; решаемые проблемные задачи стимулируют познавательную деятельность и научно-исследовательскую активность аспирантов, формируют критическое мышление.

## 8. Фонд оценочных средств

### 8.1. Паспорт фонда оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины (результаты по разделам)	Код контролируемой компетенции (или её части)	Наименование оценочного средства
1	2	3	4
1	<b>Работа с последовательностями.</b> Способы кодирования и записи нуклеотидных и аминокислотных последовательностей. Основные форматы записи последовательностей. Понятнее о выравнивании последовательностей. Парное и множественное выравнивание последовательностей.	ОПК-1	Собеседование, устный опрос, домашнее задание

2	<b>Филогенетический анализ.</b> Понятие о филогении и филогенетике. Филогенетические деревья, их устройство, виды и топологические свойства. Способы кодирования и отображения филогенетических деревьев. Современные инструменты реконструкции филогенетических деревьев. Основные модели нуклеотидных замен: GTR, JC, HKY. Основные модели аминокислотных замен: BLOSUM, PAM, JTT. Гомология последовательностей. Ортологи и паралоги.	ОПК-1	Собеседование, устный опрос, домашнее задание
3	<b>Работа с данными секвенирования нового поколения.</b> Секвенирование РНК и ДНК. История методов секвенирования. Обзор современных инструментов для секвенирования. Современные платформы секвенирования данных нового поколения. Базы данных геномов: NCBI Genome, UCSC. Понятие геномных координат, способы их записи и преобразования. Анализ дифференциальной экспрессии генов.	ОПК-1	Собеседование, устный опрос, домашнее задание
4	<b>Структурная биоинформатика.</b> Уровни организации белковых структур. Способы кодирования информации о пространственной структуре белковых молекул и соответствующие форматы файлов. RSCB PDB. Инструменты для отображения и редактирования структур белков. Построение моделей пространственных структур белков по гомологии.	ОПК-1	Собеседование, устный опрос, домашнее задание
5	<b>Биомедицинские ресурсы и поиск информации.</b> Основы поиска информации в сети интернет. Работа с базами данных. Базы данных National Center for Biotechnology Information (NCBI). Инструменты NCBI. EBI. UniProt. Иммунологическая база данных IEDB. Анализ текстовой информации, аннотаций, чтение различных форматов данных. 1000 геномов. Информация о нуклеотидных полиморфизмах, их распространенности и функциональной значимости: dbSNP. dbNSFP, ClinVar, SNPedia.	ОПК-1	Собеседование, устный опрос
6	<b>Специализированные веб-сервисы и программное обеспечение.</b> Веб-сервисы для решения биоинформатических задач	ОПК-1	Собеседование, устный опрос, домашнее

	(инструменты NCBI, EBI). Сервисы для анализа пространственных структур. Инструменты для расчета олигонуклеотидных праймеров: Oligo, FastPCR, Primer3. Базы данных векторных последовательностей для клонирования (AddGene, SnapGene, PLASMID). Программы с графическим пользовательским интерфейсом для анализа последовательностей и клонирования. BioEdit, MEGA, UGENE (UniPro), SnapGene Viewer, GenBeans.		задание
7	<b>Основы работы в командной строке.</b> Работа с файловой системой и потоками ввода/вывода. Переменные среды. Работа с программами с интерфейсом командной строки. Понятие пайплайна. Устройство и реализация биоинформатических пайплайнов. Примеры пайплайнов и инструментов для их реализации.	ОПК-1	Собеседование, устный опрос, домашнее задание
8	<b>Основы программирования.</b> Язык программирования Python. Переменные, структуры данных (списки, кортежи, словари, массивы), функции. Основные модули и научные библиотеки Python: numeric, scipy, pandas, statsmodel, Biopython. Обзор основных форматов хранения данных в биоинформатике. Anaconda. Среда разработки Spyder. Интерактивная работа в среде Jupyter.	ОПК-1	Собеседование, устный опрос, домашнее задание

## 8.2. Промежуточная аттестация (зачет)

Вопросы для подготовки к зачету:

Способы кодирования и записи нуклеотидных и аминокислотных последовательностей. Алфавиты IUPAC. Основные форматы записи последовательностей.

Понятнее о выравнивании последовательностей. Парное и множественное выравнивание последовательностей. Локальное и глобальное выравнивание. Меры близости произвольных последовательностей символов. Меры качества выравнивания. Понятие референсной последовательности.

Примеры специализированных программ для построения локальных и глобальных выравниваний, для парных и множественных выравниваний. Форматы представления множественных выравниваний.

Понятие о филогении и филогенетике. Филогенетические деревья, их устройство, виды и топологические свойства. Способы визуализации филогенетических деревьев. Интерпретация филогенетических деревьев.

Современные инструменты построения филогенетических деревьев.

Понятие эволюционного расстояния. Параметрические и эмпирические модели нуклеотидных и аминокислотных замен. Основные модели нуклеотидных замен: GTR, JC, HKY. Основные модели аминокислотных замен: BLOSUM, PAM, JTT.

Современные платформы секвенирования данных нового поколения. Принципы секвенирования нового поколения. Особенности различных современных платформ секвенирования нового поколения. Сферы применения методов секвенирования нового поколения.

Базы данных геномов. Понятие геномных координат, способы их записи и преобразования. Форматы файлов для хранения информации о геномных элементах. Аннотация геномных элементов.

Способы идентификации и кодирования генов и белков. Текстовые описания генов. Основные способы идентификации генов модельных организмов. Идентификация генов немодельных и неизученных организмов.

Анализ дифференциальной экспрессии генов.

Визуализация генов и геномов в Integrative Genomics Viewer (IGV). Высокоуровневая визуализация геномов с помощью инструмента Circos.

Уровни организации белковых структур. Способы кодирования информации о пространственной структуре белковых молекул и соответствующие форматы файлов. Инструменты для отображения и редактирования структур белков.

Построение моделей пространственных структур белков по гомологии. Моделирование пространственных структур белков *de novo*. Привести примеры специализированного программного обеспечения и веб-серверов.

Биомедицинские информационные ресурсы сети интернет. Базы данных National Center for Biotechnology Information (NCBI). Инструменты NCBI, EBI, ExPASy. Иммунологическая база данных IEDB.

Анализ текстовой информации, аннотаций, чтение различных форматов данных.

Понятие о формальных онтологиях. Примеры онтологий для биомедицинских исследований. Понятие об обогащении терминами.

Математическое понятие графа. Использование теории графов для решения биологических задач.

Понятие сети в системной биологии, примеры сетей. Регуляторные сети. Сети белок-белковых взаимодействий. Метаболические сети.

Расчет олигонуклеотидных праймеров. Инструменты для расчета олигонуклеотидных праймеров.

Базы данных векторных последовательностей для клонирования. Программы с графическим пользовательским интерфейсом для анализа последовательностей и клонирования.

Основы работы в командной строке. Работа с файловой системой и потоками ввода/вывода. Переменные среды. Работа с программами с интерфейсом командной строки. Понятие пайплайна.

Язык программирования Python. Переменные, структуры данных, функции.

### 8.3. Критерии оценивания:

Оценка «зачтено» выставляется обучающемуся, показавшему систематизированные знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач.

Оценка «не зачтено» выставляется обучающемуся, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

## 9. Учебно-методическое и информационное обеспечение дисциплины

Аспирантам ГНЦ ВБ «Вектор» Роспотребнадзора обеспечен полный доступ к обслуживанию в библиотеке ГНЦ ВБ «Вектор» Роспотребнадзора, в т.ч. библиотечное обслуживание, обслуживание по межбиблиотечному абонементу, справочно-библиографическое и информационное обслуживание.

Учебно-методическое и информационное обеспечение дисциплины включает в себя основную и дополнительную учебную и научную литературу, ресурсы информационно-телекоммуникационной сети Интернет, конспекты лекций.

### 9.1. Основная литература:

1. Артур М. Леск. Введение в биоинформатику. М: Бином. Лаборатория знаний, 2009.
2. Лукашов В. В. Молекулярная эволюция и филогенетический анализ. М.: Бином. Лаборатория знаний. 2009.
3. Дурбин Р., Эдди Ш., Крог А., Митчисон Г. Анализ биологических последовательностей. М: НИЦ «Регулярная и хаотическая динамика», 2006.

### 9.2. Дополнительная литература:

1. Кунин Е. Логика случая. О природе и происхождении биологической эволюции. М: Центрполиграф, 2014.
2. Бородовский М., Екишева С. Задачи и решения по анализу биологических последовательностей. РХД, 2008.
3. Структура и функционирование белков: применение методов биоинформатики. Под ред. Д. Дж. Ригдена. УРСС, 2014.

### 9.3. Перечень ресурсов информационно-телекоммуникационной сети Интернет, необходимых для освоения дисциплины

1. Научная электронная библиотека elibrary.ru: <http://elibrary.ru/defaultx.asp>
2. Электронные ресурсы удаленного доступа ГПНТБ России: <http://www.gpntb.ru/elektronnye-resursy-udalennogo-dostupa.html>
3. Электронные каталоги и базы данных ГПНТБ СО РАН: [http://webirbis.spsl.nsc.ru/irbis64r\\_01/cgi/cgiirbis\\_64.exe?C21COM=F&I21DBN=CAT&P21DBN=CAT](http://webirbis.spsl.nsc.ru/irbis64r_01/cgi/cgiirbis_64.exe?C21COM=F&I21DBN=CAT&P21DBN=CAT)
4. Электронная библиотека ГПНТБ СО РАН: <http://www.spsl.nsc.ru/win/nelbib/index-new1.html>

#### **10. Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине, в т.ч. программное обеспечение**

В рамках курса подготовки предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий.

В процессе обучения будут использованы:

- операционная система Microsoft Windows;
- офисная программа Microsoft Office;
- электронные презентации лекций.

#### **11. Материально-техническое обеспечение дисциплины**

- Конференц-зал корпуса №12а ГНЦ ВБ «Вектор» Роспотребнадзора, оснащенный презентационной техникой;
- средства мультимедиа: проектор, экран, компьютер/ноутбук;
- доска учебная маркерная;
- рабочее место аспиранта с выходом в Интернет.

Составители:

д.б.н., зав. отделом

к.б.н., с.н.с.

к.б.н., с.н.с.



С.И. Бажан

Д.В. Антоненц

А.Ю. Бакулина



Согласовано:

Зав. аспирантурой ФБУН ГНЦ ВБ «Вектор»

Роспотребнадзора



Т.А. Косонова

Рабочая программа рассмотрена и одобрена на заседании ученого совета ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора.

Протокол от «07» сентября 2017 г. № 8.