



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины «Научно-исследовательский семинар «Исследование медико-биологических данных»
для образовательной программы «Анализ данных в биологии и медицине» направления подготовки магистра 01.04.02 Прикладная математика и информатика

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет компьютерных наук

**Рабочая программа дисциплины «Научно-исследовательский семинар «Исследование медико-биологических данных»
2 курс**

для образовательной программы «Анализ данных в биологии и медицине» направления подготовки 01.04.02 Прикладная математика и информатика
уровень магистр

Разработчик программы

С.А.Спирин, доцент кафедры технологий моделирования сложных систем, sspirin@hse.ru

Одобрена на заседании кафедры технологий моделирования сложных систем
Зав. Кафедрой А.Н.Соболевский

Утверждена Академическим руководителем программы

Академический руководитель образовательной программы
М.С.Гельфанд

Москва, 2017

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.



1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих «Научно-исследовательский семинар «Исследование медико-биологических данных», учебных ассистентов и студентов направления подготовки 01.04.02 Прикладная математика и информатика, обучающихся по образовательной программе «Анализ данных в биологии и медицине».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ 01.04.02 Прикладная математика и информатика;
- Образовательной программой «Анализ данных в биологии и медицине».
- Объединенным учебным планом университета по образовательной программе «Анализ данных в биологии и медицине», утвержденным в 2017г.

2 Цели освоения дисциплины

Целями освоения дисциплины «Научно-исследовательский семинар «Исследование медико-биологических данных» являются

- Познакомить с основными понятиями, используемыми в вычислительной филогенетике;
- Познакомить с основными алгоритмами вычислительной филогенетики;
- Познакомить с публикациями по данной тематике;
- Научить практически выполнять простейший филогенетический анализ семейства биологических последовательностей.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

1. Понимание теоретических принципов вычислительной филогенетики;
2. Понимание русской и английской терминологии, относящейся к теме семинара, способность читать и понимать литературу по теме;
3. Владение основными программами для реконструкции филогенетических деревьев.

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к блоку обязательных дисциплин

Изучение данной дисциплины базируется на следующих дисциплинах:

- Научно-исследовательский семинар «Исследование медико-биологических данных», 1 курс
- «Молекулярная эволюция»

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Основные понятия эволюционной биологии
- Основы эволюционной медицины



5 Тематический план учебной дисциплины

№	Название раздела	Общих часов	Аудиторные часы	Самостоятельная работа
			Семинары	
1	Основные принципы молекулярной эволюции	24	4	20
2	Филогенетическое дерево как комбинаторный объект	24	4	20
3	Программы работы с филогенетическими деревьями. Форматы файлов	24	4	20
4	Алгоритмы реконструкции филогенетических деревьев	24	4	20
5	Современная литература по вычислительной филогенетике	28	8	20
6	Разбор статей и доклады студентов	130	24	106
	ВСЕГО	304	48	256

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	2 год				Параметры
		1	2	3	4	
Текущий	Контрольная работа	*				60 мин
	Доклад по статье/практическому заданию		*			Презентация
Итоговый	Экзамен		*			Устный экзамен

7 Критерии оценки знаний, навыков

Преподаватель оценивает работу студентов на семинаре, контрольную работу по теоретической части курса, качество презентации по статье/практическому заданию, оценки за которую выставляет в рабочую ведомость.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

8 Содержание дисциплины

Лекции

Тема 1. Комбинаторика филогенетических деревьев

Терминология (лист, ветвь, узел, клада, корень).

Топология дерева как набор разбиений множества листьев.

Смысл длин ветвей.

Число различных разрешённых деревьев с n листьями (укоренённых и неукоренённых).

Что такое «ультраметричность» (применительно к филогенетическому дереву)



Тема 2. Принципиальная схема реконструкции филогении
Дистанционные и символьные методы
Прямые и переборные методы

Тема 3. Оценка эволюционных расстояний между последовательностями
Формула Джукса – Кантора (Jukes-Cantor)
Оценка по принципу максимального правдоподобия

Тема 4. Поиск в пространстве деревьев
Выращивание (stepwise addition)
Nearest neighbor interchange (NNI)
Subtree pruning and re-grafting (SPR)
(*) Markov chain Monte Carlo

Тема 5. Критерии качества дерева
Максимальная экономия (maximum parsimony)
Максимальное правдоподобие (maximum likelihood)
Минимальная эволюция (minimum evolution)
Наименьшие квадраты (OLS = Ordinary least squares, Fitch – Margoliash)

Тема 6. Прямые методы реконструкции филогении
UPGMA
Neighbor-joining

Тема 7. Укоренение филогенетических деревьев
Укоренение в среднюю точку (midpoint)
Укоренение с помощью внешней группы (outgroup)

Тема 8. Поддержка ветвей дерева
Бутстреп (bootstrap)
Складной нож (jackknife)

Журнальный клуб

Каждый студент делает доклад по одной из статей из предложенного списка.

Практическая работа

Два варианта:

1. Взять статью, где описывается реконструкция филогении какой-либо группы организмов и повторить часть работы (15–30 организмов) другим способом (те же последовательности, но другие методы, или же другие последовательности из тех же видов или родов и т.п.). Сравнить результаты. Можно не организмы, а мультигенное семейство (ортологи и паралоги из нескольких организмов).
2. Взять набор организмов по своему усмотрению, подобрать ортологичные последовательности, реконструировать несколькими способами.

9 Порядок формирования оценок по дисциплине

Две контрольные работы по материалу лекций: 25% итоговой оценки



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины «Научно-исследовательский семинар «Исследование медико-биологических данных»
для образовательной программы «Анализ данных в биологии и медицине» направления подготовки магистра 01.04.02 Прикладная математика и информатика

Журнальный клуб:

- собственный доклад: 30%
- на экзамене: рассказ об одной из статей (по выбору экзаменатора), доклад о которой был на журнальном клубе: 20%
- Отчёт о практической работе 25%

9 Рекомендуемая литература

1. Лукашов В.В. Молекулярная эволюция и филогенетический анализ. – М.: БИНОМ. Лаборатория знаний, 2009.
2. Felsenstein J. Inferring Phylogenies. – Sunderland: Sinauer Associates, 2004.