

БАКАЛАВР. АКАДЕМИЧЕСКИЙ КУРС

В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко

БИОИНФОРМАТИКА

УЧЕБНИК



СООТВЕТСТВУЕТ
ПРОГРАММАМ
ВЕДУЩИХ НАУЧНО-
ОБРАЗОВАТЕЛЬНЫХ
ШКОЛ

Юрайт
ИЗДАТЕЛЬСТВО

УМО ВО рекомендует

biblio-online.ru

В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко

БИОИНФОРМАТИКА

УЧЕБНИК ДЛЯ АКАДЕМИЧЕСКОГО БАКАЛАВРИАТА

*Рекомендовано Учебно-методическим отделом высшего образования
в качестве учебника для студентов высших учебных заведений,
обучающихся по техническим и естественнонаучным направлениям*

**Книга доступна в электронной библиотечной системе
biblio-online.ru**

Москва ■ Юрайт ■ 2017

УДК 57.08(075.8)
ББК 30.16я73
С79

Авторы:

Стефанов Василий Евгеньевич — кандидат биологических наук, доцент, заведующий кафедрой биохимии биологического факультета Санкт-Петербургского государственного университета;

Тулуб Александр Александрович — доктор физико-математических наук, ведущий научный сотрудник кафедры биохимии биологического факультета Санкт-Петербургского государственного университета;

Мавропуло-Столяренко Григорий Ростиславович — старший преподаватель кафедры биохимии биологического факультета Санкт-Петербургского государственного университета.

Рецензенты:

Максимович Н. В. — доктор биологических наук, заведующий кафедрой ихтиологии и гидробиологии Санкт-Петербургского государственного университета;

Тимковский А. Л. — доктор физико-математических наук, профессор кафедры биофизики Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого.

Стефанов, В. Е.

С79 Биоинформатика : учебник для академического бакалавриата / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. — М. : Издательство Юрайт, 2017. — 252 с. — Серия : Бакалавр. Академический курс.

ISBN 978-5-534-00860-9

Приведены основы биоинформатики — науки, изучающей живые системы методом компьютерного эксперимента. Объем курса соответствует уровню академического бакалавриата, а четкое структурирование информации и обоснованный выбор разделов, описывающих ключевые концепции базовых дисциплин — молекулярной биологии, математики, квантовой физики, — делает книгу полезной для широкого круга читателей разного уровня подготовки и специализации. Содержит описания реальных и общедоступных ресурсов биологических баз данных и репозиториях биоинформатических инструментов.

Содержание учебника соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования.

Для студентов высших учебных заведений, обучающихся по инженерно-техническим и естественнонаучным направлениям.

УДК 57.08(075.8)
ББК 30.16я73



Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав. Правовую поддержку издательства обеспечивает юридическая компания «Дельфи».

ISBN 978-5-534-00860-9

© Стефанов В. Е., Тулуб А. А.,
Мавропуло-Столяренко Г. Р., 2016
© ООО «Издательство Юрайт», 2017

Оглавление

Предисловие	7
Глава 1. Предмет биоинформатики	10
1.1. Макромолекулы — носители информации в клетке	10
1.1.1. Информация на клеточном уровне	10
1.1.2. Структура ДНК. Азотистые основания	13
1.1.3. Структура РНК	18
1.1.4. Структура белков.....	20
1.2. Информационные матричные процессы на молекулярном уровне.....	26
1.2.1. Репликация	26
1.2.2. Транскрипция	29
1.2.3. Процессы модификации РНК. Сплайсинг	32
1.2.4. Трансляция	34
1.2.5. Формирование пространственных структур ДНК. Структура ДНК-белковых комплексов	37
1.2.6. Формирование пространственных структур РНК	43
1.2.7. Формирование пространственных структур белков.....	48
1.3. Междисциплинарный характер биоинформатики	53
1.3.1. Задачи биоинформатики.....	54
1.3.2. Геномика	55
1.3.3. Протеомика	56
1.3.4. Молекулярная филогения.....	57
<i>Вопросы и задания для самопроверки</i>	<i>58</i>
<i>Литература к главе 1.....</i>	<i>58</i>
Глава 2. Математический аппарат биоинформатики	60
2.1. Простые статистические модели в биоинформатике.....	60
2.1.1. Метод скользящего окна	60
2.1.2. Применение теоретико-информационного подхода	62
2.2. Алгоритмы выравнивания последовательностей.....	65
2.2.1. Алгоритм глобального выравнивания последовательностей.....	66
2.2.2. Алгоритм локального выравнивания последовательностей	69
2.2.3. Варианты реализации штрафующей функции за пропуски в выравнивании	71
2.2.4. Алгоритмы множественного выравнивания последовательностей	72
2.3. Марковские цепи и скрытые модели Маркова.....	74
2.3.1. Марковские цепи	74
2.3.2. Скрытые модели Маркова.....	76
2.3.3. Алгоритмы, применяемые при работе со скрытыми моделями Маркова.....	78

2.4. Искусственные нейронные сети	83
2.4.1. Одиночный нейрон	84
2.4.2. Обучение нейрона	86
2.4.3. Иерархические сети, многослойный пещептрон	88
2.4.4. Сеть Хопфилда	89
2.4.5. Самоорганизующиеся карты	91
2.5. Байесова статистика	93
2.5.1. Классическая и байесова вероятности. Байесов вывод	93
2.5.2. Формула Байеса для непрерывных распределений	96
2.6. Симуляция Монте-Карло и алгоритмы имитации отжига	97
2.6.1. Симуляция Монте-Карло	97
2.6.2. Методы Монте-Карло, использующие марковские цепи	99
2.6.3. Алгоритм имитации отжига	101
2.7. Эволюционные и генетические алгоритмы	103
2.7.1. Общие принципы	103
2.7.2. Эволюционные алгоритмы	103
2.7.3. Чередувание поколений в эволюционных алгоритмах	104
2.7.4. Оценка и ранжирование особей согласно степени их «приспособленности»	105
2.7.5. Способы отбора особей	107
2.7.6. Воспроизводство особей	110
2.7.7. Генетические алгоритмы	111
2.8. Алгоритмы кластерного анализа	112
2.8.1. Введение в кластерный анализ	112
2.8.2. Примеры алгоритмов	116
<i>Вопросы и задания для самопроверки</i>	118
<i>Задачи</i>	118
<i>Литература к главе 2</i>	118
Глава 3. Примеры моделей в биоинформатике	119
3.1. Общие модели	119
3.1.1. Сравнение последовательностей биологических полимеров	119
3.1.2. Методы быстрого поиска в базах данных	127
3.1.3. Статистическая оценка значимости значений оценочной функции при выравнивании последовательностей	129
3.1.4. Применение скрытых моделей Маркова для идентификации последовательностей	137
3.1.5. Представление входных данных и интерпретация результатов в моделях, основанных на искусственных нейронных сетях	140
3.1.6. Логотипы последовательностей. Информационно- теоретическая интерпретация консервативных участков последовательностей биологических полимеров	142
3.2. Геномика — примеры моделей	144
3.2.1. Поиск открытых рамок считывания. Поиск кодирующих фрагментов	144
3.2.2. Предсказание оперонной структуры прокариотических геномов ...	148
3.2.3. Предсказание вторичной структуры молекул РНК	151

3.3. Протеомика — примеры моделей	156
3.3.1. Профили аминокислотных последовательностей, позиционно- зависимые матрицы замещения	156
3.3.2. Предсказание наличия сигнальных пептидов в исследуемой последовательности	158
3.3.3. Применение сети Кохонена для ускорения поиска в базе	161
3.3.4. Предсказание вторичной структуры белков	162
3.3.5. Предсказание трансмембранных фрагментов белков.....	168
3.4. Моделирование структуры белка в свете информационных технологий .	170
3.4.1. Подходы к решению проблемы моделирования пространственной структуры белка. Задача сравнительного моделирования структуры белка по его аминокислотной последовательности.....	173
3.4.2. Альтернативные способы подбора матричной структуры.....	180
3.5. Эволюционные модели.....	188
3.5.1. Множественные выравнивания.....	189
3.5.2. Построение филогенетических деревьев.....	191
<i>Вопросы и задания для самопроверки</i>	195
<i>Литература к главе 3.....</i>	195
Глава 4. Биологические базы данных	197
4.1. Основы структур баз данных.....	197
4.2. Базы данных, содержащие информацию о биологических структурах и последовательностях	198
4.3. Основные форматы данных биоинформатики.....	202
4.4. Каталог известных баз данных.....	203
Глава 5. Биоинформатические программы и сервисы	208
5.1. Анализ последовательностей биологических полимеров	208
5.2. Расширенный поиск с применением алгоритмов семейства BLAST.....	211
5.3. Предсказание физико-химических свойств белков на основании аминокислотных последовательностей	214
5.4. Поиск белков, обладающих сходными физико-химическими свойствами	215
5.5. Предсказание особых типов структур и свойств белков	216
5.6. Каталог репозитория веб-сервисов	219
<i>Вопросы и задания для самопроверки к главам 4 и 5.....</i>	220
<i>Задачи</i>	220
<i>Литература к главам 4 и 5</i>	221
Глава 6. Квантовый мир и биология — две сущности одной и той же реальности	222
6.1. Информация и вычисления: от классического компьютеринга к квантовому.....	222
6.1.1. Что представляет собой информация — попытка определения.....	222
6.1.2. Классическая информация и вычисления	223
6.1.3. Квантовая информация и вычисления	224
6.2. Сознание как проявление квантового мира и человеческий мозг в качестве квантового компьютера.....	229

6.2.1. Квантовая информация и работа сознания находят друг друга.....	229
6.2.2. Одна проблема: как совместить локальность и нелокальность.....	233
6.2.3. Пути преодоления противоречий. Грани нелинейности: логика, множества, числа	233
6.2.4. Фрактальная p -адическая феноменология единства естественных систем.....	234
6.2.5. P -адическая система разбирается не только в работе сознания, но и в конформациях белков	235
6.3. Спин — носитель квантовой информации.....	237
6.3.1. Квантовый спин и фаза.....	237
6.3.2. Триплетные и синглетные состояния.....	238
6.3.3. Спин заставляет спирали ДНК закручиваться в спираль.....	241
<i>Вопросы и задания для самопроверки</i>	<i>244</i>
<i>Литература к главе 6.....</i>	<i>245</i>
Приложение. Примеры вычислений на p-адических числах	247
Новые издания по дисциплине «Биоинформатика» и смежным дисциплинам.....	252