

А. Леск

# ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Перевод с английского  
под редакцией  
доктора биол. наук, профессора А. А. Миронова  
и доктора хим. наук, профессора В. К. Швядаса



Москва  
БИНОМ. Лаборатория знаний  
2009

УДК 577.3  
ББК 28.071я73  
Л50

**Леск А.**

Л50 Введение в биоинформатику / А. Леск ; пер. с англ. — М. : БИНОМ. Лаборатория знаний, 2009. — 318 с. : ил., [4] с. цв. вкл.

ISBN 978-5-94774-501-6 (русск.)

ISBN 0-19-925196-7 (англ.)

В учебном издании, написанном английским ученым — пионером в использовании приемов информатики в биологических исследованиях, ведущим преподавательскую работу в Кембриджском университете, изложены основы информационных технологий в применении к биологическим наукам. Приведены тексты некоторых программ, упражнения и задачи.

Для студентов университетов и научных работников.

**УДК 577.3**

**ББК 28.071я73**

Первый тираж издания осуществлен при финансовой поддержке  
Российского фонда фундаментальных исследований по проекту № 05-04-62033

---

*Учебное издание*

**Леск Артур**

**ВВЕДЕНИЕ В БИОИНФОРМАТИКУ**

Ведущий редактор канд. хим. наук *Т. И. Почкаева*

Редактор канд. хим. наук *Н. А. Анижин*

Художники *С. Инфантэ, Н. А. Новак*

Оригинал-макет подготовлен *О. Г. Лапко* в пакете  $\text{\LaTeX}$  2 $\epsilon$

Подписано в печать 02.06.09. Формат 70×100/16.

Усл. печ. л. 26,325. Тираж 1000 экз. Заказ

Издательство «БИНОМ. Лаборатория знаний»

125167, Москва, проезд Аэропорта, д. 3

Телефон: (499) 157-5272, e-mail: binom@Lbz.ru, http://www.Lbz.ru

Отпечатано в производственной фирме «Полиграфист»

160001, г. Вологда, ул. Челюскинцев, 3

---

ISBN 978-5-94774-501-6 (русск.)

ISBN 0-19-925196-7 (англ.)

© Arthur M. Lesk 2002

“Introduction to Bioinformatics” was originally published in English in 2002. This translation is published by arrangement with Oxford University Press.

Книга “Introduction to Bioinformatics” была впервые опубликована на английском языке в 2002 г. Этот перевод опубликован по договору с Oxford University Press

© БИНОМ. Лаборатория знаний, 2009

# Оглавление

---

Предисловие редакторов русского издания . . . . .	5
Предисловие . . . . .	8
<b>1. Введение . . . . .</b>	<b>15</b>
Сценарий . . . . .	17
Жизнь в пространстве и времени . . . . .	18
Догмы: основные и второстепенные . . . . .	19
Архивы данных и доступ к ним . . . . .	22
Курирование, аннотация и контроль качества . . . . .	25
Всемирная Паутина (The World Wide Web) . . . . .	26
Что такое URL? . . . . .	28
Электронные публикации . . . . .	29
Компьютеры и компьютерные науки . . . . .	29
Программирование . . . . .	31
Биологическая классификация и номенклатура . . . . .	34
Использование последовательностей для определения филогенетических взаимосвязей . . . . .	37
Использование SINE и LINE для установления филогенетического родства . . . . .	45
Поиск схожих последовательностей в базах данных: PSI-BLAST . . . . .	48
Структуры белков. Введение . . . . .	56
Иерархия в белковой архитектуре . . . . .	57
Классификация белковых структур . . . . .	59
Предсказание структур белков и белковая инженерия . . . . .	61
Критическая оценка предсказания структуры (CASP) . . . . .	68
Белковая инженерия . . . . .	68
Медицинские аспекты . . . . .	68
Будущее . . . . .	71
Упражнения, задачи и компьютерные задания . . . . .	73
<b>2. Организация генома и эволюция . . . . .</b>	<b>81</b>
Геномика и протеомика . . . . .	81
Гены . . . . .	82
Белки . . . . .	85
Протеомы . . . . .	86
Отслеживание передачи генетической информации . . . . .	89
Соответствие между картами . . . . .	91
Генетические карты высокого разрешения . . . . .	94
Локализация генов в геноме . . . . .	97
Геномы прокариот . . . . .	98
Геном бактерии <i>Escherichia coli</i> . . . . .	98
Геном архея <i>Methanococcus jannaschii</i> . . . . .	102
Геномы наиболее просто организованных организмов: <i>Mycoplasma genitalium</i> . . . . .	103
Геномы эукариот . . . . .	104
Геном <i>Saccharomyces cerevisiae</i> (пекарские дрожжи) . . . . .	108
Геном <i>Caenorhabditis elegans</i> . . . . .	110

Геном <i>Drosophila melanogaster</i> . . . . .	112
Геном <i>Arabidopsis thaliana</i> . . . . .	112
Геном <i>Homo sapiens</i> (геном человека) . . . . .	114
Белок-кодирующие гены . . . . .	114
Повторяющиеся последовательности . . . . .	116
РНК . . . . .	117
Однонуклеотидные полиморфизмы (SNP, СНП) . . . . .	118
Генетическое разнообразие в антропологии . . . . .	120
Генетическое разнообразие и идентификация личности . . . . .	121
Генетический анализ одомашнивания крупного рогатого скота . . . . .	122
Эволюция геномов . . . . .	123
Пожалуйста, передайте гены: горизонтальный перенос генов . . . . .	127
Сравнительная геномика эукариот . . . . .	128
Упражнения, задачи и компьютерные задания . . . . .	131
<b>3. Архивы и извлечение информации</b> . . . . .	<b>135</b>
Введение . . . . .	136
Оглавление базы данных и терминология поисковых систем . . . . .	136
Какие еще вопросы могут возникать . . . . .	137
Анализ полученных данных . . . . .	138
Архивы . . . . .	138
Базы данных последовательностей нуклеиновых кислот . . . . .	139
Ген ингибитора бычьего панкреатического трипсина (последовательность ДНК из базы данных EMBL) . . . . .	140
Геномные базы данных . . . . .	141
Базы данных белковых последовательностей . . . . .	142
Базы данных, близкие SWISS-PROT . . . . .	144
PIR и связанные с ним базы данных . . . . .	144
Базы данных структур . . . . .	146
Индикаторы качества структуры . . . . .	152
Ядерный магнитный резонанс (ЯМР) . . . . .	153
Классификации белковых структур . . . . .	153
Специализированные, или локальные, базы данных . . . . .	154
Базы данных по экспрессии и протеомике . . . . .	155
Банки данных метаболических путей . . . . .	158
Библиографические базы данных . . . . .	159
Обзоры баз данных и серверов по молекулярной биологии . . . . .	159
Вход в архивы . . . . .	160
Доступ к базам данных в молекулярной биологии . . . . .	161
Как приобрести навык работы в молекулярной биологии через Интернет? . . . . .	161
ENTREZ . . . . .	161
Поиск по базе данных белков ENTREZ . . . . .	162
Поиск в банке данных нуклеотидных последовательностей ENTREZ . . . . .	162
Поиск в банке данных геномов ENTREZ . . . . .	166
Поиск в банке данных структур ENTREZ . . . . .	166
Поиск по библиографической базе данных PubMed . . . . .	168
Интерактивный каталог «Менделевская (по Менделю) наследственность человека» (OMIM) . . . . .	169
Система поиска последовательностей (Sequence Retrieval System, SRS) . . . . .	170

Ресурс идентификации протеинов (Protein Identification Resource, PIR) . . . . .	173
ExPASy — экспертная система анализа белков. . . . .	177
Ресурс Ensembl. . . . .	178
Куда мы отправимся дальше? . . . . .	179
Упражнения, задачи и компьютерные задания. . . . .	181
<b>4. Выравнивания и филогенетические деревья . . . . .</b>	<b>184</b>
Выравнивание последовательностей. Введение. . . . .	184
Точечная матрица сходства. . . . .	185
Точечные матрицы и выравнивание последовательностей . . . . .	192
Мера сходства последовательностей. . . . .	198
Схемы оценки. . . . .	199
Получение матриц замен . . . . .	200
Матрицы BLOSUM. . . . .	201
Взвешивание вставок/делений. . . . .	201
Расчет выравнивания для двух последовательностей . . . . .	203
Вариации и обобщения . . . . .	204
Приближенные методы для быстрого поиска в базах данных . . . . .	204
Алгоритм динамического программирования для построения оптимального парного выравнивания последовательностей . . . . .	205
Значимость выравниваний. . . . .	211
Множественное выравнивание последовательностей . . . . .	215
Связь множественных выравниваний последовательностей и структур . . . . .	216
Программы для поиска множественного выравнивания последовательностей по базам данных . . . . .	218
Профили . . . . .	219
PSI-BLAST . . . . .	221
Скрытые марковские модели (HMM) . . . . .	224
Филогения . . . . .	226
Филогенетические деревья. . . . .	231
Методы кластеризации . . . . .	232
Кладистические методы . . . . .	235
Проблема переменной скорости эволюции . . . . .	236
Вычислительный анализ. . . . .	237
Упражнения, задачи и компьютерные задания. . . . .	238
<b>5. Структура белков и разработка лекарств . . . . .</b>	<b>247</b>
Введение . . . . .	247
Стабильность и сворачивание (фолдинг) белков . . . . .	249
Графические представления по Сасисекхаран—Рамакришнан—Рама- чандран для описания разрешенных конформаций основной цепи. . . . .	249
Боковые остатки. . . . .	252
Стабильность и денатурация белков . . . . .	253
Сворачивание (фолдинг) белков . . . . .	256
Применения гидрофобности . . . . .	258
Совмещение структур и структурные выравнивания. . . . .	263
Выравнивание матриц расстояний с помощью программы DALI. . . . .	266
Эволюция белковых структур. . . . .	267
Классификация структур белков . . . . .	270
База данных SCOP . . . . .	270

Предсказание и моделирование белковых структур . . . . .	271
Критическая оценка предсказаний структуры (CASP) . . . . .	274
Предсказание вторичной структуры . . . . .	275
Нейронные сети . . . . .	276
Моделирование по гомологии . . . . .	280
Распознавание фолда . . . . .	283
3D-профили . . . . .	283
Использование 3D-профилей для определения качества структур . . . . .	284
Трединг . . . . .	285
Распознавание фолда в CASP 2000 . . . . .	286
Вычисление конформационной энергии и молекулярная динамика . . . . .	287
Программа ROSETTA . . . . .	290
Программа LINUS . . . . .	292
Определение белковых структур в геномах . . . . .	293
Предсказание функции белка . . . . .	296
Дивергенция функций: ортологи и паралоги . . . . .	297
Открытие и разработка лекарств . . . . .	299
Лидерное соединение (Лид) . . . . .	300
Уточнение лида: количественное соотношение структура– активность (QSAR) . . . . .	302
Компьютерный дизайн лекарств . . . . .	304
Упражнения, задачи и компьютерные задания . . . . .	309
<b>Заключение . . . . .</b>	<b>314</b>