

Г.П. Несговорова

БИОИНФОРМАТИКА: ПУТИ РАЗВИТИЯ И ПЕРСПЕКТИВЫ

ВВЕДЕНИЕ

Со времени своего появления, во второй половине 20-го века, наука информатика начала широко внедряться и сотрудничать (и продолжает это и в наши дни) с другими науками: физико-математическими, техническими, гуманитарными и пр. Перефразируя всем известные слова М.В. Ломоносова, можно сказать, что «широко простирает информатика руки свои в дела человеческие». Сейчас трудно найти такую область науки, которая бы обходилась без методов информатики. Не избежали этого и естественные науки.

В конце 60-х – начале 70-х годов прошлого века ЭВМ стали активно применяться в биологии: к этому времени возросла их память, увеличилась скорость операций, уменьшились размеры. И к тому же накопилось большое количество экспериментальных данных по биологии, требующих осмысления и обработки. Например, уже к 2003 г. объединенными усилиями ученых многих стран был в общих чертах прочитан геном человека.

Таким образом, на стыке XX-го и XXI-го веков появилась бурно развивающаяся область биомедицинской науки биоинформатика, которой к настоящему времени насчитывается уже около 30 лет. Биоинформатика обязана своим появлением накоплению обширных экспериментальных данных. Особенно заметно это накопление стало проходить в 70-х гг. XX-го века. Информации, получаемой в биологических экспериментах, было значительно больше, чем возможности человека к запоминанию фактов и их анализу. Возникла необходимость хранения все быстрее увеличивающегося объема информации. Первые несколько сотен расшифрованных последовательностей белков были опубликованы в виде книги-атласа. Однако уже в начале 70-х г. число расшифрованных последовательностей возросло настолько, что из-за их объема оказались невозможными публикации в виде книг. Стало понятно, что нужны специальные программы для сравнения последовательностей, поскольку мозг человека не справляется с анализом такой информации.

В 90-е годы происходит расцвет геномики. К настоящему времени расшифрованы полные геномные последовательности человека, мыши, цып-

ленка, лягушки, отдельных видов рыб, круглых червей, нескольких сотен вирусов и бактерий и т.д. Прочтение генома бактерии – это теперь задача, посильная для группы из 2–3 исследователей за время, меньшее, чем один год. Геном человека составляет около 3-х миллиардов букв, что эквивалентно 15000 книжных томов. А значение факта его «прочтения» для биологов сравнимо с открытием Д.И. Менделеевым периодического закона для химиков.

Поэтому для обработки такой обширной биологической информации на помощь пришли компьютерные технологии. Первый алгоритм выравнивания генов последовательностей был предложен еще в 1970 г. Компьютеры позволили хранить информацию в виртуальных банках данных и оперировать ею с большой скоростью. Биоинформатика, как и многие другие современные науки, развивается на стыке разных наук: молекулярной биологии, генетики, математики и компьютерных технологий. Основная ее задача – разработка вычислительных алгоритмов для анализа и систематизации данных о структуре и функциях биологических молекул, прежде всего нуклеиновых кислот и белков.

Объем генетической информации, накапливаемый в банках данных, начал увеличиваться с возрастающей скоростью после того, как были разработаны быстрые методы секвенирования (расшифровки) нуклеотидных последовательностей ДНК. Достижения информатики, лингвистики и теории информации сделали возможным анализ генетических текстов. Взаимосвязанное развитие биоинформатики с другими областями науки позволяет рассчитывать на формирование нового уровня понимания биологических процессов, происходящих в клетках и организмах.

Если учесть, что первый персональный компьютер появился в 1981 г., а Интернет (World Wide Web) – в 1991 г., т.е. совсем недавно, то это имело колоссальное значение для развития не только предмета науки, но и ее организации. Один из главных принципов биоинформатики (как, впрочем, и других наук) – принцип единого мирового информационного пространства, объединяющего усилия ученых, работающих по всему миру.

Цель данной статьи – дать определение и понятия для широкого круга лиц, что такое биоинформатика, не вдаваясь в глубинные процессы, доступные для понимания лишь узким специалистам-биоинформатикам. В конце статьи в качестве приложения приводится словарь специфических биологических терминов, не всегда известных читателям других специальностей.

1. ЧТО ТАКОЕ БИОИНФОРМАТИКА

Истоки зарождения биоинформатики восходят к давним временам, к 13 веку. Молодой итальянец Леонардо из Пизы, вошедший в историю математики под именем Фибоначчи и знаменитый одноименными числами, описал решение задачи о размножении кроликов, построив тем самым первую математическую модель биологического процесса. А в 20-е годы XX века другой итальянский математик, Вито Вольтерра, создал модель совместного существования двух биологических популяций типа «хищник–жертва». После Второй мировой войны, в конце 40-х годов прошлого столетия, в биологию пришли математики и физики. Современная история биологии начинается в 1953 году, когда американскими учеными Уотсоном и Криком была открыта двойная спираль ДНК.

К настоящему времени существуют разные определения биоинформатики, но в основном под биоинформатикой понимают любое использование компьютеров для обработки разного рода биологической информации. Сегодня поле термина «биоинформатика» значительно расширилось и включает все реализации математических алгоритмов, связанных с биологическими объектами, и информационно-коммуникационных дисциплин, применяемых в биологических исследованиях. В биоинформатике, помимо самой информатики, используются методы прикладной математики, статистики и других точных наук. Биоинформатика используется также в биохимии, биофизике, экологии, генетике и в ряде других областей естественных наук.

Биоинформатика включает в себя:

- 1) математические методы компьютерного анализа в сравнительной геномике (геномная биоинформатика);
- 2) разработку алгоритмов и программ для предсказания пространственной структуры белков (структурная биоинформатика);
- 3) исследование стратегий соответствующих вычислительных методологий, а также общее управление биологическими системами информационной сложности.

Модное среди современных ученых слово «биоинформатика» – это уже почти отделившаяся ветвь молекулярной биологии. На сегодняшний день существует множество его определений и интерпретаций. Пока это еще не совсем устоявшийся термин. Рассмотрим это подробнее.

Одни говорят, что биоинформатика – это раздел современной науки, занимающейся раскодированием человеческого генома.

Другие утверждают, что биоинформатика – это системная биология, позволяющая рассмотреть, изучить и систематизировать глобальную картину биологии.

Третьи считают, что биоинформатика – это аналог молекулярной биологии с той лишь разницей, что молекулярная биология занимается научными исследованиями в пробирке, а биоинформатика – при помощи мощных компьютерных систем.

Четвертые верят, что биоинформатика – это возможность по структуре макромолекулы очень быстро найти гены-мишени и создать новые лекарственные препараты.

И каждый из них по-своему прав.

Существуют такие биологические утверждения, которые сначала были предсказаны, а потом проверены, и оказалось, что так и есть. Отсюда появляются такие ветви биоинформатики, как предсказательная, аналитическая, эволюционная.

В практическом смысле биоинформатика – это прикладная наука, обслуживающая интересы биологов. К области технической биоинформатики относится первичная обработка данных. Полученные данные надо где-то хранить, обеспечивать к ним удобный доступ и т.д. Более сложное и интересное занятие биоинформатиков – получать на основе данных о геноме конкретные утверждения: белок А обладает такой-то функцией, ген В включается в таких-то условиях и т.п. В этом и состоит практическое применение науки биоинформатики.

Биоинформатика находит свое применение в следующих направлениях биологической науки:

- геномика, транскриптомика и протеомика;
- компьютерное моделирование в биологии развития;
- компьютерный анализ генных сетей;
- моделирование в популяционной генетике.

Биоинформатика легко вписалась также в фармакологию, позволяя снизить срок проектирования препарата с 5-6 лет до нескольких месяцев, а также интегрировалась во многие другие медицинские и биологические науки.

На сегодняшний день существуют следующие разделы биоинформатики:

- биоинформатика в целом;
- клиническая биоинформатика;
- структурная геномика;
- функциональная геномика;

- фармакогеномика;
- клиническая протеомика;
- функциональная протеомика;
- структурная протеомика.

С помощью методов биоинформатики возможно не просто обрабатывать огромный массив различных биологических данных, но и выявлять закономерности, которые не всегда можно заметить при обычном эксперименте, предсказывать функции генов и зашифрованных в них белков, строить модели взаимодействия генов в клетке, конструировать лекарственные препараты.

2. ОСНОВНЫЕ ОБЛАСТИ ИССЛЕДОВАНИЙ В БИОИНФОРМАТИКЕ

Рассмотрим подробнее основные области исследований в биоинформатике.

2.1. Анализ генетических последовательностей

С тех пор как в 1977 г. был секвенирован фаг Phi-X 174, последовательности ДНК всё большего числа организмов были дешифрованы и сохранены в базах данных. Эти данные используются для определения последовательностей белков и регуляторных участков. С ростом количества данных уже давно стало невозможным анализировать последовательности вручную. В наши дни для поиска по геномам тысяч организмов, состоящих из миллиардов пар нуклеотидов, используются компьютерные программы.

Сборка фрагментов ДНК может быть довольно сложной задачей для больших геномов. Сейчас этот метод применяется практически для всех геномов, и алгоритмы сборки геномов являются одной из острейших проблем биоинформатики на сегодняшний день. Обработка гигантского количества данных, получаемых при секвенировании, является одной из важнейших задач биоинформатики. Другим примером применения компьютерного анализа генетических последовательностей является автоматический поиск генов и регуляторных последовательностей в геноме.

2.2. Аннотация геномов

В контексте геномики аннотация – это процесс маркировки генов и других объектов в последовательности ДНК. Первая программная система аннотации геномов была создана еще в 1955 году Оуэном Уайтом.

2.3. Вычислительная эволюционная биология

Эволюционная биология исследует происхождение и появление видов, их развитие с течением времени. Информатика помогает эволюционным биологам в нескольких аспектах:

- 1) изучать эволюцию большого числа организмов, измеряя изменения всех ДНК;
- 2) сравнивать целые геномы, что позволяет изучать более комплексные эволюционные события;
- 3) строить компьютерные модели популяций, чтобы предсказывать поведение системы во времени;
- 4) отслеживать появление публикаций, содержащих информацию о большом количестве видов.

2.4. Оценка биологического разнообразия

Биологическое разнообразие экосистемы может быть определено как полная генетическая совокупность определенной среды, состоящая из всех обитающих видов, будь это капля воды или горсть земли или вся биосфера планеты Земля. Специализированное программное обеспечение применяется для поиска, визуализации и анализа информации и, что особенно важно, для предоставления ее другим людям.

Помимо биоинформатики в современной научной биологической литературе встречается также термин «вычислительная биология». Считается, что вычислительная биология – это не область науки, а подход к использованию компьютеров для изучения биологических процессов. Термины «биоинформатика» и «вычислительная биология» пока еще часто употребляются как синонимы, хотя «вычислительная биология» чаще указывает на разработку алгоритмов и конкретные вычислительные методы. Считается, что не всё использование вычислительной биологии является биоинформатикой, например, математическое моделирование – это не биоинформатика, хотя и связано с биологическими задачами.

Существует еще и математическая биология, которая, как и биоинформатика, также решает биологические задачи, но используемые при этом методы не являются численными, и им не требуется реализация в программном или аппаратном обеспечении.

Отдельно выделяется структурная биоинформатика, к которой относится разработка алгоритмов и программ для предсказания пространственной структуры белков. Биоинформатика, таким образом, входит в разделы био-

логии наряду с анатомией, ботаникой, вирусологией, микробиологией, цитологией, палеонтологией, физиологией и т.д.

3. МЕСТО БИОИНФОРМАТИКИ В ЦЕПИ БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Биоинформатика анализирует факты, которые получены в экспериментальной биологии. Полученные данные экспериментатор сравнивает со всем набором, уже имеющимся в банке данных. Если он не нашел в этом банке открытой им последовательности, он заносит ее туда, пополняя тем самым банк. В функции банков включается хранение, систематизация, обновление информации и обеспечение доступа к ней. Эти операции требуют огромных компьютерных мощностей.

Существуют также банки научных публикаций по биологической тематике. Каждая статья, выходящая в номере любого научного журнала по биологии, помещается в банк и аннотируется так, чтобы любой другой ученый мог легко найти ее через Интернет. Крупнейшая on-line библиотека медико-биологических публикаций PubMed содержит более 16 млн. статей, вышедших в течение последних 50 лет.

Интегральные банки данных и энциклопедии выполняют важную функцию, объединяя в себе всю известную информацию о конкретном гене, белке, организме и т.д. Они обобщают информацию из большого числа других банков данных и постоянно обновляют ее.

Рассмотрим, какие исследовательские цели ставит перед собой биоинформатика.

3.1. Анализ геномов, поиск в них генов

Любой вновь прочитанный геном представляет собой огромную последовательность повторяющихся в разных комбинациях букв. Биоинформатика позволяет выделить гены в этом текстовом многообразии. Такая операция выделения гена из генома называется разметкой генома.

3.2. Предсказание функции генов

Выяснить функции всех генов опытным путем достаточно трудоемко. В этом случае биоинформатика помогает предсказывать их, опираясь на сравнение с теми генами, функции которых уже определены.

3.3. Оценка роли отдельных участков последовательности в функционировании белка

В молекуле белка есть участки, отвечающие за решение разного рода биологических задач. Распознавание этих участков с помощью методов биоинформатики расширяет весь спектр функций конкретного белка.

3.4. Построение молекулярных моделей белков на основе их последовательностей

Структуры белков определяют опытным путем, например, облучая микроскопический кристалл, состоящий из молекул белка, рентгеновскими лучами. Это достаточно долгий и дорогой процесс. Некоторые белки вообще невозможно анализировать так, поскольку они не имеют кристаллической структуры. Биоинформатика с помощью компьютерного моделирования помогает воссоздать пространственную модель белка, если известна структура белка с хотя бы отдаленно похожей последовательностью.

3.5. Исследование механизма функционирования макромоделей, исходя из их моделей

Зная пространственную структуру молекулы, полученную благодаря методам биоинформатики, можно предсказать, как она работает, и как на ее работу можно повлиять.

3.6. Компьютерное конструирование лекарств

Лекарства можно конструировать, моделируя в пространстве взаимодействие различных химических соединений с белком-мишенью. При этом надо перебрать огромное количество соединений и выбрать оптимальное.

3.7. Анализ информации

Соединение биологии, химии, физики, математики и информатики позволяет разносторонне описывать биологическую систему. Использование компьютерных ресурсов помогает многократно ускорить процесс анализа и повысить точность и скорость получения результатов.

3.8. Использование достижений информатики

Новые открытия в биологии с использованием технологий биоинформатики быстро находят свое применение в медицине, фармакологии, косметологии, биотехнологии, сельском хозяйстве, экологии и других областях. Биоинформатика самостоятельно дает результаты, имеющие практическую значимость, а также обеспечивает условия для работы разных областей биологии.

3.9. Основные технологии биоинформатики

Большая часть работ по биоинформатике сконцентрирована вокруг технологии использования баз данных для хранения биологической информации с последующей ее обработкой. Такие базы данных могут быть публичными или частными. Очень важным является предоставление публичного доступа к таким базам данных через открытые стандарты. Получает развитие использование онтологий и логических методов для обработки биологической информации, хотя этот подход менее распространен, чем использование баз данных.

4. БИОИНФОРМАТИКА НА СОВРЕМЕННОМ ЭТАПЕ РАЗВИТИЯ

4.1. Развитие биоинформатики в мире

На рубеже 20-21 веков биоинформатика превратилась в бурно развивающуюся область мировой биомедицинской науки. Наряду с исследователями, ведущими фундаментальные разработки, потребителями биоинформационных технологий являются медицинские, фармакологические, биотехнологические и учебные учреждения. Эта область науки определена в качестве приоритетной как в США, так и во всех других развитых странах.

Количество центров биоинформатики постоянно растет во всех странах Европы, Азии, США и Австралии. Наряду с государственными, академическими и образовательными центрами биоинформатики, в последние годы возникло значительное число организаций и проектов, ориентированных на коммерческое использование результатов исследований в области биоинформатики. Это прежде всего организации, деятельность которых ориентирована на структурный, функциональный и сравнительный анализ геномов, включая геном человека. Наряду с применением уже созданных ме-

тодов биоинформатики интенсивно развивается техническая и программная база для решения прикладных задач, особенно в фармакологии. Быстрыми темпами совершенствуется также и индустрия программного обеспечения для решения таких задач.

4.2. Состояние биоинформатики в России

Биоинформатика относится к интеллектуальным, высокотехнологичным разделам науки, где получаемые результаты в значительной степени зависят от развитого творческого мышления ученых, а не определяются в основном затратами на их техническую вооруженность. Таким образом, учитывая достаточно высокий интеллектуальный и образовательный уровень российских ученых и практическую невозможность больших финансовых затрат в современной экономической ситуации, биоинформатика имеет все основания стать одним из приоритетных направлений науки в Российской Федерации.

Уже к 1984 г. стало ясно, что в России (тогда бывшем СССР) образовалась достаточная «критическая масса» специалистов в области применения математических методов в биологии. К этому же времени скопилось достаточно большое количество биологической информации, обработка, осмысление и анализ которой стало невозможно проводить без компьютерной поддержки. Все это и явилось предпосылкой для возникновения новой отрасли биологической науки – биоинформатики.

Количество публикаций по биоинформатике в последние годы стремительно нарастает. Работы в области биоинформатики активно печатались в 90-е годы, но сами эксперименты, описываемые в статьях, были проведены ранее. База для реализации комплексного подхода к проблеме и написания статей создавалась в течение многих лет усилиями многочисленных исследователей. Первые работы по теоретическому анализу аминокислотных последовательностей белков были опубликованы в 50-х годах прошлого века.

В 1988 г. был начат проект по расшифровке генома человека, ставивший своей целью определение полной последовательности ДНК, составляющей хромосомы человека. Этот проект общими усилиями ученых разных стран был завершен к 2003 г.

Отметим ряд институтов, где в России ведутся работы по биоинформатике.

В Москве организован Институт математических проблем биологии РАН. Там занимаются решением разных биологических проблем при под-

держке информационных технологий, таких как структурная и сравнительная геномика, основные молекулярно-биологические механизмы, молекулярно-генетические системы управления, протеомика, метаболика, базы биологических данных, математическое обеспечение биологических экспериментов.

При Институте проблем передачи информации РАН (ИППИ РАН) в 2003 г. основан учебно-научный центр «Биоинформатика». Также к научным институтам такой тематики относятся следующие учреждения:

ВНИИ «Генетика» (Москва);

Институт белка РАН (Пушино);

Институт биоорганической химии РАН (Москва);

Институт молекулярной биологии РАН (Москва);

Институт биомедицинской химии РАМН (Москва);

Институт цитологии и генетики СО РАН (Новосибирск) и ряд других.

Web-сайты имеются у следующих институтов:

Институт биомедицинской химии РАМН (Москва) –

<http://www.ibmh.msk.su/bioinform>

Институт физико-химической биологии МГУ им. Белозерского –

<http://www.genebee.msu.su/>

Институт цитологии и генетики СО РАН (Новосибирск) –

<http://www.mgs.bionet.nsc.ru/> и другие.

Ряд оригинальных компьютерных программ и баз данных, созданных российскими учеными, приведен и на других сайтах. Ведется работа по подготовке отечественных специалистов в области биоинформатики. Во многих университетах и вузах биолого-медицинской тематики созданы кафедры биоинформатики и организованы учебные курсы по биоинформатике, такие как «Биоинформатика и компьютерное конструирование лекарств».

4.3. Создание Центра Коллективного Пользования по биоинформатике в СО РАН

В СО РАН (Новосибирский Академгородок) существует центр для разработки экспериментальных технологий в области биоинформатики. Для биологов НГУ и СО РАН создали единую информационную базу по биоинформатике. Совместными усилиями нескольких институтов СО РАН и Новосибирского государственного университета открыт Центр Коллективного Пользования (ЦКП) «БИОИНФОРМАТИКА» Суперкомпьютер центра один из самых быстрых в мире. Его производительность 16,5 терафлопс

(16,5 триллионов операций в секунду), по мощности он не уступает аналогичным, установленным в центре Vital-IT в Лозанне (Швейцария), система хранения данных рассчитана на 48 терабайт.

Центр «Биоинформатика», по сути, – это огромная вычислительная система, суперкомпьютер, с помощью которого можно моделировать лекарства от рака и СПИДа, расшифровывать геномы растений и живых организмов. Прочитать миллиарды и миллионы букв в геномах человека, пшеницы, вируса гепатита или энцефалита сегодня позволяет ДНК-секвенатор. В будущем ЦКП сможет помочь и специалистам смежных областей науки. Например, ученые планируют сравнить геном древнего гоминида из алтайской пещеры с геномом современного человека

В ЦКП более 400 пользователей, это около 20 академических институтов СО РАН. Есть и учебные институты, и фирмы, например, «Унипро», «Шлюмберже» и др.

Цель данного центра – создать сетевую инфраструктуру, которая объединит в технологический комплекс экспериментальные установки с вычислительными комплексами и хранилищами данных. Задачами самой структуры являются биоинформационная поддержка научно-исследовательских работ, подготовка специалистов в области компьютерной системной биологии и биоинформатики, разработка новых экспериментально-теоретических технологий.

Активное использование вычислительных мощностей ЦКП предполагается в рамках проекта СО РАН «Геномика, протеомика, биоинформатика». Организациями-учредителями проекта стали Институт цитологии и генетики СО РАН, Институт вычислительной математики и математической геофизики СО РАН, Международный Томографический центр СО РАН, Институт химической биологии и фундаментальной медицины СО РАН, Институт математики СО РАН, Институт вычислительных технологий СО РАН и Новосибирский государственный университет.

С 1998 г. в Институте цитологии и генетики СО РАН регулярно проводятся Международные конференции по биоинформатике регуляции и структуры геномов и системной биологии (Bioinformatics of Genome Regulation and Structure System Biology – BGRS/SB). На регулярно проходящем с недавних пор в Новосибирске Международном молодежном инновационном форуме Интерра также уделяется внимание вопросам биоинформатики.

Новое направление в науке требует и специалистов нового профиля. Поэтому в Академгородке осуществляется их подготовка: в школе N162 совместно с Институтом цитологии и генетики СО РАН создан профиль-

ный 10-й класс по направлению «биоинформатика» с углубленным изучением биологии, информатики, математики, физики на базе Института цитологии и генетики СО РАН и кафедры биоинформатики Новосибирского государственного университета.

В Сибирском государственном медицинском университете (СибГМУ, Новосибирск) в учебном плане появился новый предмет – биоинформатика. Со второго семестра студенты 5-го курса начинают изучать эту интересную и сложную науку, одно из направлений молекулярной биологии, которое исследует молекулярные процессы, но не *in vitro*, а *in silico*, т.е. не в пробирке, а при помощи компьютера.

Следует также отметить и заслуги сибирских ученых в создании новой науки. К ним относится А.А. Ляпунов, выдающийся русский ученый-математик. В 2011 г. отмечалось 100 лет со дня его рождения. Последние годы своей жизни он жил и работал в Академгородке. Надо сказать, что будучи «отцом отечественной кибернетики», он заинтересовался биологией еще в 30-е годы прошлого века. Позднее, в 50-е годы, он снова обратился к биологии. Сложность биологических систем и процессов их эволюции привлекала его как объект приложения методов исследований, характерных для дескриптивной теории множеств. Таким образом, он стоял у истоков создания новой науки на стыке математики и биологии.

5. ПРИМЕРЫ КОНКРЕТНЫХ ПРОГРАММНЫХ КОМПЛЕКСОВ БИОИНФОРМАТИКИ

В системной биологии важнейшую роль играют методы биоинформатики. Они дают возможность с помощью компьютеров

- накапливать и интегрировать в банки данных экспериментальную информацию;
- осуществлять ее компьютерный анализ;
- проводить математическое моделирование структурно-функциональной организации живых систем;
- предсказывать новые свойства живых систем;
- на этой основе планировать новые этапы экспериментальных исследований.

В этом разделе приводятся конкретные примеры баз и банков данных, базовых пакетов, программных средств для полного анализа макромолекул, а также других программных продуктов по биоинформатике с их ад-

ресами в Интернете, разработанных как отечественными, так и зарубежными учеными.

GenBank – <http://www.ncbi.nlm.nih.gov/GenBank>

Банк данных по нуклеотидным последовательностям (3400000000 пар оснований в 461000 последовательностей).

SWISS-PROT – <http://www.expasy.ch/sprot/sprot-top.html>

Аннотированный банк данных по аминокислотным последовательностям белков.

PIR – <http://www.nbrf.georgetown.edu/pir/searchdb.html>

Аннотированный банк данных по аминокислотным последовательностям белков, организованных в соответствии с гомологией и таксономией.

PDB – <http://www.rcsb.org/pdb/>

Банк данных по 3D структуре биологических макромолекул.

NDB – <http://ndbserver.rutgers.edu>

Банк данных по нуклеиновым кислотам. Включает структуры ДНК и РНК вместе с их 3-хмерными изображениями.

ProDom – <http://protein.toulouse.inra.fr/prodom.html>

Банк данных по доменам белков.

Web-серверы, предоставляющие пользователю генетическую информацию, оснащены комплексом программных средств для поиска информации в банках данных и анализа нуклеотидных и аминокислотных последовательностей. В качестве запросов при поиске последовательностей в банках данных могут использоваться номенклатурные названия генов, организмов, ключевые слова и др.

В качестве примера предложим программу Auto Dok, которая является программой для автоматического докинга. С ее помощью можно посмотреть, как молекулы лекарств или кандидатов на роль лекарств взаимодействуют в известной 3D-структуре. В частности, программа применяется для разработки лекарств, специфически связывающихся с тем или иным белком.

Здесь же приведем примеры основных программ сравнения аминокислотных и нуклеотидных последовательностей.

ACT – (Artemis Comparison Tool) – геномный анализ;

Arlequin – анализ популяционно-генетических данных;

Bio Edit – редактор множественного выравнивания аминокислотных и нуклеотидных последовательностей;

Bio Numerics – коммерческий универсальный пакет программ по биоинформатике;

BLAST – поиск родственных последовательностей в базе данных аминокислотных и нуклеотидных последовательностей;

ClustalW – множественное выравнивание аминокислотных и нуклеотидных последовательностей;

FASTA – набор алгоритмов определения схожести аминокислотных и нуклеотидных последовательностей;

Mesquite – программа для сравнительной биологии на языке Java;

Muscle – множественное сравнение аминокислотных и нуклеотидных последовательностей. Более быстрая и точная программа в сравнении ClustalW;

Pop Gene – анализ генетического разнообразия популяций;

Populations – популяционно-генетический анализ.

Примером интегрированного инструмента биолога является также Unipro UGENE. Это свободно распространяемое программное обеспечение для работы молекулярного биолога. Пользовательский интерфейс этого продукта обеспечивает:

- простую и удобную работу с последовательностями;
- визуализацию хроматограмм;
- использование редактора множественного выравнивания последовательностей;
- просмотр трехмерных моделей PDB и MMDB с поддержкой стереорежима;
- просмотр филогенетических деревьев;
- применение конструктора вычислительных схем, автоматизирующего процесс анализа;
- поддержку сохранения изображений в векторные форматы для удобства публикаций.

Ряд оригинальных компьютерных программ, баз и банков данных, созданных российскими учеными, можно также найти по разным поисковикам на других многочисленных сайтах по биоинформатике.

ЗАКЛЮЧЕНИЕ

Сегодня в биологии происходит глобальная смена модели науки. Еще недавно эксперимент подтверждал или опровергал те или иные гипотезы.

Теперь же биология становится наукой, которая оперирует с данными, получаемыми в промышленных масштабах. Использование методов биоинформатики позволяет ученым делать содержательные выводы, которые опираются на сопоставление разнородных данных или на объединение согласованных наблюдений.

Хотя биоинформатике как науке всего лишь около 30 лет, в ней уже существуют свои традиционные направления: компьютерный анализ ДНК, РНК и белковых последовательностей, реконструкция пространственных структур биополимеров, теоретический и компьютерный анализ структурно-функциональной организации геномов и белков. Сюда же относятся и развитие баз данных по молекулярно-генетической тематике, структурированию экспериментальных данных. Биоинформатика по своей сути – это интегративная наука, являющаяся инструментом в руках тех, кто занимается биологией, молекулярной биологией, молекулярной генетикой, медицинской генетикой, фармакологией, а также биохимией, биофизикой и т.д.

Подводя итоги, можно сказать что биоинформатика это широко применяющийся и быстро развивающийся способ заниматься биологией, не наблюдая живые существа, как зоологи, не делая опытов в пробирке, как в экспериментальной биологии, а анализируя результаты массовых данных или целых проектов, используя компьютерные технологии, можно делать совершенно конкретные, проверяемые биологические утверждения.

Биоинформатика, геномика и протеомика – это науки о жизни, определяющие начало 21 века, так же как молекулярная биология, иммунология и биотехнология определяли конец 20 века. Не так давно биохимики тратили всю жизнь, чтобы определить структуру одного белка и выявить его функцию. С помощью методов биоинформатики можно предсказать функции тысячи белков. Таким образом, в триаде геномика–протеомика–биоинформатика по сути заложены все новые подходы к созданию принципиально новой медицины будущего: новые методы диагностики, новые лекарства, так как биоинформатика – это путь от гена к лекарству через структуру макромолекулы.

Современное понимание биоинформатики приводит нас к мысли о том, что все те задачи, которые до недавнего времени решались биохимией и молекулярной биологией в реальных экспериментах, в будущем могут быть решены с той или иной степенью точности в виртуальных компьютерных экспериментах.

ПРИЛОЖЕНИЕ

Словарь биологических терминов

Биоинформатика, геномика, протеомика и прочие «-омы» и «-омики» – все эти слова прочно вошли в современную биологическую науку, она буквально пестрит красивыми названиями. При этом «-ом» обозначает множество однородных элементов (ген–геном), а «-омика» – методы изучения (геном–геномика). Посмотрим, что же означают основные термины современной биологии.

Аминокислота (аминокарбоновая кислота) – органическое соединение, в молекуле которого одновременно содержатся карбоксильные и аминные группы. К настоящему времени в различных объектах живой природы обнаружено до 200 аминокислот.

Аннотация – процесс маркировки генов и других и других объектов в последовательности ДНК.

Генетика изучает механизмы изменчивости и наследственности в живых организмах, определяет развитие современной науки о жизни. Изучает структуру и функцию генов, инвентаризирует гены, создавая геномные карты живых организмов.

Геном (нем. Genom, англ. genome) – совокупность генов, локализованных в одиночном наборе хромосом данного организма. Термин «геном» предложен Гансом Винклером в 1920 г.

Геномика – раздел молекулярной генетики, изучает геном и гены живых существ. Занимается созданием карты метаболических сетей.

Гомология (греч. homologia – соответствие) – сходство основных структур и органов организмов, основанных на общем генетическом наследстве.

Докинг – молекулярный докинг – это метод молекулярного моделирования, который позволяет предсказать наиболее выгодную для образования устойчивого комплекса ориентацию и положение одной молекулы по отношению к другой.

Метаболизм (греч. metabole – перемена, превращение) – совокупность химических реакций, протекающих в живых клетках и обеспечивающих организм веществами и энергией для его жизнедеятельности, роста и размножения.

- Метабономика** дает возможность понять, как происходит обмен веществ в клетке, изучить и смоделировать метаболизм, исследовать совместимость функций элементов биологической системы и, как следствие, ускорить процесс создания лекарственных препаратов.
- Нуклеотиды** – фосфорные эфиры нуклеозидов, нуклеозидфосфаты. Свободные нуклеотиды, в частности, АТФ, АДФ, играют важную роль в энергетических и информационных внутриклеточных процессах.
- Онтология** – раздел философии, в котором рассматриваются всеобщие основы, принципы бытия, его структура и закономерности.
- Пептиды** (греч. *peptos* – сваренный, переваренный) – органические вещества, состоящие из остатков одинаковых или различных аминокислот, соединенных пептидной связью. Другими словами, пептиды – это универсальные белки.
- Протеомика** изучает и получает данные о сотнях тысяч пептидов и белок-белковых взаимодействиях, иначе – занимается инвентаризацией белков, т.е. реально работающих молекулярных машин в клетке. Это новая наука, выросшая из геномики.
- Секвенирование** (англ. *sequence* – последовательность) – это определение первичной или аминокислотной, или нуклеотидной последовательности белков и нуклеиновых кислот: ДНК и РНК.
- Таксономия** (греч. *taxis* – расположение, порядок и *nomos* – закон) – теория классификации и систематизации сложноорганизованных областей действительности, имеющих обычно иерархическое строение.
- Транскриптомика** – изучает активность генов, получает данные о концентрации десятков тысяч матричных РНК.
- Фаг** (греч. *fagos* – пожиратель) – вирус, избирательно поражающий бактериальные клетки.
- Фармакогенетика** (греч. *pharmakon* – лекарство + генетика) – раздел медицинской генетики и фармакологии, изучающий характер реакций организма на лекарственные средства в зависимости от наследственных факторов.
- Филогенетика**, или филогенетическая систематика – область биологической систематики, которая занимается идентификацией и прояснением эволюционных взаимоотношений среди разных видов жизни на Земле, как современных, так и вымерших.

ССЫЛКИ ПО ТЕМЕ

1. <http://www.bioinformatics.ru>
2. <http://www.rusbiotech.ru>
3. <http://elementy.ru/lib/430895>
4. <http://www.sibai.ru>
5. <http://lake.baikal.ru>
6. <http://www.jcbi.ru>
7. http://www.rtcb.iitp.ru/index_r.html
8. <http://medi.ru/pbmc/8800101.htm>
9. <http://www.mgs.bionet.nsc.ru/mgs/>
10. <http://www.BioinforMatix.ru>