

## БИОИНФОРМАТИКА

**Анашкин В. С.**

Научный руководитель – Стрелюхин А. В.  
Белорусский национальный технический университет  
Минск, Беларусь

**Аннотация.** В работе представлены цели и задачи биоинформатики. Показаны базовые понятия, основные подходы и методы, применяемые в вычислительной биологии.

### **Введение**

Биоинформатика – междисциплинарная область, объединяющая общую биологию, молекулярную биологию, кибернетику, генетику, химию, информатику, математику и статистику. Основные биологические задачи, требующие анализа больших объемов данных, решаются биоинформатикой с вычислительной точки зрения. Биоинформатика в первую очередь включает в себя изучение и разработку вычислительных методов и фокусируется на сборе, анализе, хранении, организации и визуализации биологических данных.

Биоинформатика стала важной частью многих областей биологии. Методы биоинформатического анализа позволяют интерпретировать большие объемы экспериментальных данных, что было практически невозможно до развития этого направления. Например, экспериментальная молекулярная биология часто использует методы биоинформатики, такие как обработка изображений и сигналов. В генетике и геномике биоинформатика помогает в функциональной аннотации геномов, обнаружении и анализе мутаций. Важной задачей является изучение экспрессии генов и путей ее регуляции. Кроме того, инструменты биоинформатики позволяют сравнивать геномные данные, что является необходимым условием для изучения принципов молекулярной эволюции [1].

### **История**

История секвенирования биоинформатики начинается с 1950 годов. В феврале 1953 г. Уотсон и Крик предложили модель молекулы

ДНК, а в мае 1953 г. они опубликовали в журнале Nature статью, посвященную вопросу ДНК как носителя кода генетической информации. Кроме того, в конце 1950-х Сэнгер опубликовал первую последовательность белка, инсулина.

В 1970 году Нидлман и Вунш разработали первый алгоритм динамического программирования для попарного сопоставления белковых последовательностей. Алгоритмы выравнивания нескольких последовательностей появились позже: первый практический алгоритм был разработан в 1987 году Да-Фей Фенгом и Расселом Ф. Дулиттлом. Его упрощение, алгоритм CLUSTAL, используется до сих пор. Кроме того, в 1978 году группа ученых создала первую альтернативную модель, основанную на наблюдении за точечными мутациями (РАМ) в филогенетическом дереве 71 семейства белков с идентичностью более 85%. Результатом является матрица, содержащая значения вероятности аминокислотной замены.

Геномная эра биоинформатики началась с публикации генома человека в начале двадцать первого века. Проект начался в США в 1991 году и за 13 лет затраты на него превысили 2,7 миллиарда долларов. В 1998 году компания Celera Genomics организовала собственное исследование по секвенированию и сборке генома человека. Это исследование снизило затраты в десять раз и привело к разработке новых экспериментальных стратегий секвенирования, таких как Illumina. Стоимость секвенирования ДНК значительно снизилась, что привело к значительному увеличению количества последовательностей в общедоступных базах данных [2].

### **Цель исследований**

Основная цель биоинформатики – помочь понять биологические процессы. Отличие биоинформатики от других подходов заключается в том, что они сосредоточены на разработке и использовании интенсивных вычислительных методов для достижения этой цели. Примерами таких методов являются распознавание образов, сбор данных, алгоритмы машинного обучения и визуализация биологических данных. Основное внимание исследователей сосредоточено на секвенировании, генетических исследованиях (открытие генов, кодирующих ДНК), кодировании, разработке лекарств, прогнозировании синтеза белка, предсказания экспрессии генов, полногеномного поиска ассоциаций и моделирования эволюции.

Современная биоинформатика предполагает создание и совершенствование баз данных, алгоритмов, вычислительных, статистических и теоретических методов решения практических и теоретических задач, возникающих при обработке и анализе биологических данных [3].

Мнение Альперовича М., главы юнита биоинформатики ЕРАМ:

«Представьте, что у вас есть десять тысяч экземпляров «Войны и мира». Вы пропустили их через шредер, хорошенько перемешали, наугад вытащили из этой кучи ворох бумажных полосок и пытаетесь собрать из них исходный текст. Вдобавок у вас есть рукопись «Войны и мира». Текст, который вы соберете, нужно будет сравнить с ней, чтобы отловить опечатки (а они обязательно будут). Примерно так же читают ДНК современные машины-секвенаторы. ДНК выделяют из клеточных ядер и делят на фрагменты по 300–500 пар нуклеотидов (мы помним, что в ДНК нуклеотиды связаны друг с другом попарно). Молекулы дробят, потому что ни одна современная машина не может прочитать геном от начала до конца. Последовательность слишком длинная, и по мере ее прочтения накапливаются ошибки.

Вспоминаем «Войну и мир» после шредера. Чтобы восстановить исходный текст романа, нам нужно прочитать и расположить в правильном порядке все кусочки романа. Получается, что мы читаем книгу несколько раз по крошечным фрагментам. То же с ДНК: каждый участок последовательности секвенатор прочитывает с многократным перекрытием – ведь мы анализируем не одну, а множество молекул ДНК.

Полученные фрагменты выравнивают – «прикладывают» каждый из них к эталонному геному и пытаются понять, какому участку эталона соответствует прочитанный фрагмент. Затем в выровненных фрагментах находят вариации – значащие отличия прочтений от эталонного генома (опечатки в книге по сравнению с эталонной рукописью). Этим занимаются программы – вариант-коллеры. Это самая сложная часть анализа, поэтому различных программ – вариант-коллеров много и их постоянно совершенствуют и разрабатывают новые.

Подавляющее большинство найденных мутаций нейтральны и ни на что не влияют. Но есть и такие, в которых зашифрованы предрас-

положенность к наследственным заболеваниям или способность откликаться на разные виды терапии» [4].

Секвенсор ДНК (секвенатор) – научный прибор или устройство, с помощью которого выполняется автоматизированное определение последовательности нуклеотидов в цепи ДНК – секвенирования. В секвенатор загружается образец ДНК, результатом его работы является набор последовательностей оснований аденина, тимина, гуанна, цитозина.

Одной из важнейших и интереснейших ветвей биоинформатики является эволюционная биология.

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

- изучать эволюцию большого числа организмов, измеряя изменения в их ДНК, а не только в строении или физиологии;
- сравнивать целые геномы, что позволяет изучать комплексные эволюционные события, такие как: дупликация генов, горизонтальный перенос генов, и предсказывать бактериальные специализирующие факторы;
- строить компьютерные модели популяций, чтобы предсказать поведение системы во времени;
- отслеживать появление публикаций, содержащих информацию о большом количестве видов [5].

### **Заключение**

Биоинформатика является молодой, но перспективной областью биологической науки, основной задачей которой является работа с большими массивами молекулярно-генетических данных, а также моделирование и предсказание свойств отдельных биологических молекул и целых биологических систем. В будущем можно ожидать интенсивного развития биоинформатики, особенно в таких направлениях, как моделирование биологических систем, предсказание функций генов, систематика биологических объектов и выявление механизмов эволюции органического мира.

### **Литература**

1. Бородовский М., Екишева С. «Задачи и решения по анализу

биологических последовательностей». – М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2008. – 420 с.

2. Сетубал Ж, Мейданис Ж. «Введение в вычислительную молекулярную биологию». – М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2007. – 420 с.

3. Дурбин Р, Эдди Ш, Крог А, Митчисон Г. «Анализ биологических последовательностей». – М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2006. – 480 с.

4. Блог компании ЕРАМ [Электронный ресурс]. - Режим доступа: [https://habr.com/ru/company/epam\\_systems/blog/337892/](https://habr.com/ru/company/epam_systems/blog/337892/). – Дата доступа: 13.09.2017.

5. Албертс Б., Брей Д., Льюис Дж., Рэфф М., Роберте К. Уотсон Дж. Д. «Молекулярная биология клетки». – В 3-х т. 2-е изд. перераб. и доп. Т. 2.: Пер. с англ. – М.: Мир, 1993. – 539 с.