Биоинформатика (Bioinformatics)

*Биоинформатика* — междисциплинарная область, объединяющая общую биологию, молекулярную биологию, кибернетику, генетику, химию, компьютерные науки, математику и статистику. Крупномасштабные биологические проблемы, требующие анализа больших объемов данных, решаются биоинформатикой с вычислительной точки зрения. Биоинформатика главным образом включает в себя изучение и разработку компьютерных методов и направлена на получение, анализ, хранение, организацию и визуализацию биологических данных.

**Основные направления биоинформатики:**

* Биоинформатика последовательностей

Это методы анализа аминокислотных и нуклеотидных последовательностей.

* Анализ экспрессий

Это анализ транскриптома, измерение транскрипционной активности гена с помощью определения количества его продукта, матричной РНК (мРНК), универсальной для большей части генов.

* Структурная биоинформатика

Это отрасль биоинформатики, которая связана с анализом и прогнозированием трехмерной структуры биологических макромолекул, таких как белки, РНК и ДНК.

* Изучение клеточной организации
* Системная биология

Это развивающаяся междисциплинарная область биологии, которая анализирует сложные биологические системы разного уровня исходя из их многокомпонентности, наличия прямых и обратных связей, разнородности экспериментальных данных, характеризующих системы.

Для каждого из перечисленных разделов можно выделить свои стандартные типы данных, способы их обработки, биоинформатические алгоритмы и базы данных. В биоинформатике используются методы прикладной математики, статистики и информатики. Биоинформатика используется в биохимии, биофизике, экологии и в других областях.

**Цели биоинформатики:**

Главная цель биоинформатики — способствовать пониманию биологических процессов. Отличие биоинформатики от других подходов состоит в том, что она фокусируется на создании и применении интенсивных вычислительных методов для достижения этой цели. Примеры подобных методов: распознавание образов, data mining, алгоритмы машинного обучения и визуализация биологических данных. Основные усилия исследователей направлены на решение задач выравнивания последовательностей, нахождения генов (поиск региона ДНК, кодирующего гены), расшифровки генома, конструирования лекарств, разработки лекарств, выравнивания структуры белка, предсказания структуры белка, предсказания экспрессии генов и взаимодействий «белок-белок», полногеномного поиска ассоциаций и моделирования эволюции.  
  
 Биоинформатика сегодня подразумевает создание и совершенствование баз данных, алгоритмов, вычислительных и статистических методов и теории для решения практических и теоретических проблем, возникающих при управлении и анализе биологических данных.

**Аннотация геномов:**

В контексте геномики аннотация — процесс маркировки генов и других объектов в последовательности ДНК. Первая программная система аннотации геномов была создана в 1995 году Оуэном Уайтом (англ. Owen White), работавшим в команде, секвенировавшей и проанализировавшей первый декодированный геном свободноживущего организма, бактерии Haemophilus influenzae. Доктор Уайт построил систему для нахождения генов, тРНК и других объектов ДНК и сделал первые обозначения функций этих генов. Большинство современных систем работают сходным образом, но эти программы постоянно развиваются и улучшаются.

**Вычислительная 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 более комплексные эволюционные события, такие как: дупликация генов, латеральный 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.

Область в компьютерных науках, которая использует генетические алгоритмы, часто путают с компьютерной эволюционной биологией. Работа в этой области использует специализированное программное обеспечение для улучшения алгоритмов и вычислений и основывается на эволюционных принципах, таких, как репликация, диферсификация через рекомбинацию или мутации, и выживании в естественном отборе.

**Оценка биологического разнообразия**

Биологическое разнообразие экосистемы может быть определено как полная генетическая совокупность определённой среды, состоящая из всех обитающих видов, была бы это биоплёнка в заброшенной шахте, капля морской воды, горсть земли или вся биосфера планеты Земля. Для сбора видовых имён, описаний, ареала распространения, генетической информации используются базы данных. Специализированное программное обеспечение применяется для поиска, визуализации и анализа информации, и, что более важно, предоставления её другим людям. Компьютерные симуляторы моделируют такие вещи, как популяционная динамика, или вычисляют общее генетическое здоровье культуры в агрономии. Один из важнейших потенциалов этой области заключается в анализе последовательностей ДНК или полных геномов целых вымирающих видов, позволяя запомнить результаты генетического эксперимента природы в компьютере и возможно использовать вновь в будущем, даже если эти виды полностью вымрут.

Часто из области рассмотрения биоинформатики выпадают методы оценки других компонентов биоразнообразия — таксонов (в первую очередь видов) и экосистем. В настоящее время математические основания биоинформационных методов для таксонов представлены в рамках такого научного направления как Фенетика, или численная таксономия. Методы анализа структуры экосистем рассматриваются специалистами таких направлений как системная экология, биоценометрия.

**Основные биоинформационные программы:**

Биоинформатика используется в биохимии, биофизике, экологии и в других областях. Наиболее часто используемыми инструментами и технологиями в этой области являются языки программирования Python, R, Java, C#, C++; язык разметки — XML; язык структурированных запросов к базам данных — SQL; программно-аппаратная архитектура параллельных вычислений — CUDA; пакет прикладных программ для решения задач технических вычислений и одноимённый язык программирования, используемый в этом пакете — MATLAB, и электронные таблицы.

* ACT (Artemis Comparison Tool) — геномный анализ
* Arlequin — анализ популяционно-генетических данных
* BioEdit — редактор множественного выравнивания нуклеотидных и аминокислотных последовательностей
* BioNumerics — коммерческий универсальный пакет программ
* BLAST — поиск родственных последовательностей в базе данных нуклеотидных и аминокислотных последовательностей
* ClustalW — множественное выравнивание нуклеотидных и аминокислотных последовательностей
* ClustalX — множественное выравнивание нуклеотидных и аминокислотных последовательностей
* DnaSP — анализ полиморфизма последовательностей ДНК
* FigTree — редактор филогенетических деревьев
* Genepop — популяционно-генетический анализ
* Genetix — популяционно-генетический анализ (программа доступна только на французском языке)
* JalView — редактор множественного выравнивания нуклеотидных и аминокислотных последовательностей
* MacClade — коммерческая программа для интерктивного эволюционного анализа данных
* MEGA — молекулярно-эволюционный генетический анализ
* Mesquite — программа для сравнительной биологии на языке Java
* Muscle — множественное сравнение нуклеотидных и аминокислотных последовательностей. Более быстрая и точная по сравнению с ClustalW
* PAUP — филогенетический анализ с использованием метода парсимонии (и других методов)
* PHYLIP — пакет филогенетических программ
* Phylo\_win — филогенетический анализ. Программа имеет графический интерфейс.
* PopGene — анализ генетического разнообразия популяций
* Populations — популяционно-генетический анализ
* PSI Protein Classifier — обобщение результатов, полученных с помощью программы PSI-BLAST
* Seaview — филогенетический анализ (с графическим интерфейсом)
* Sequin — депонирование последовательностей в GenBank, EMBL, DDBJ
* SplitsTree
* T-Coffee — множественное прогрессивное выравнивание нуклеотидных и аминокислотных последовательностей. Более чувствительное, чем в ClustalW/ClustalX.
* UGENE — свободный русскоязычный инструмент, множественное выравнивание нуклеотидных и аминокислотных последовательностей, филогенетический анализ, аннотирование, работа с базами данных.

## **Структурная биоинформатика**

К структурной биоинформатике относится разработка алгоритмов и программ для предсказания пространственной структуры белков. Темы исследований в структурной биоинформатике:

* Рентгеноструктурный анализ (РСА) макромолекул;
* Индикаторы качества модели макромолекулы, построенной по данным РСА;
* Алгоритмы вычисления поверхности макромолекулы;
* Алгоритмы нахождения гидрофобного ядра молекулы белка;
* Алгоритмы нахождения структурных доменов белков;
* Пространственное выравнивание структур белков;
* Структурные классификации доменов SCOP и CATH;
* Молекулярная динамика.

**ЛИТЕРАТУРА**

1. Jonathan Pevsner (2013) Bioinformatics and Functional Genomics
2. Jean-Michel Claverie Ph.D. (2007) Bioinformatics For Dummies. 2nd edition.
3. Дурбин Р, Эдди Ш, Крог А, Митчисон Г. «Анализ биологических последовательностей». — М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2006. — 480 с. — ISBN 5-93972-559-7
4. Бородовский М., Екишева С. «Задачи и решения по анализу биологических последовательностей». — М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2008. — 420 с. — ISBN 978-5-93972-644-3
5. Сетубал Ж, Мейданис Ж. «Введение в вычислительную молекулярную биологию». — М.-Ижевск: НИЦ «Регулярная и хаотичная динамика», 2007. — 420 с. — ISBN 978-5-93972-623-8
6. В. А. Таланов, Математические модели синтеза пептидных цепей и методы теории графов в расшифровке генетическиех текстов