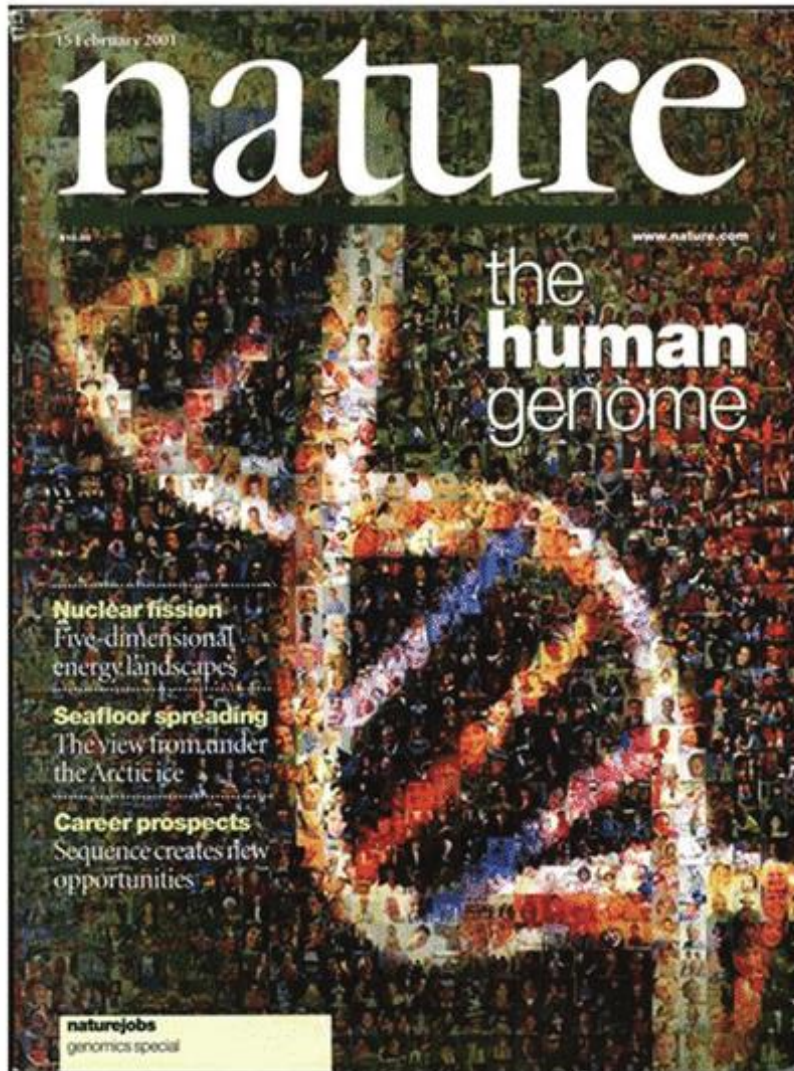


Биоинформатика: почему сегодня биологу нужен компьютер

Лиходеевский Георгий

**м.н.с. Лаборатории МБИ Уральского ГАУ
м.н.с. СТЦ в области картофелеводства**



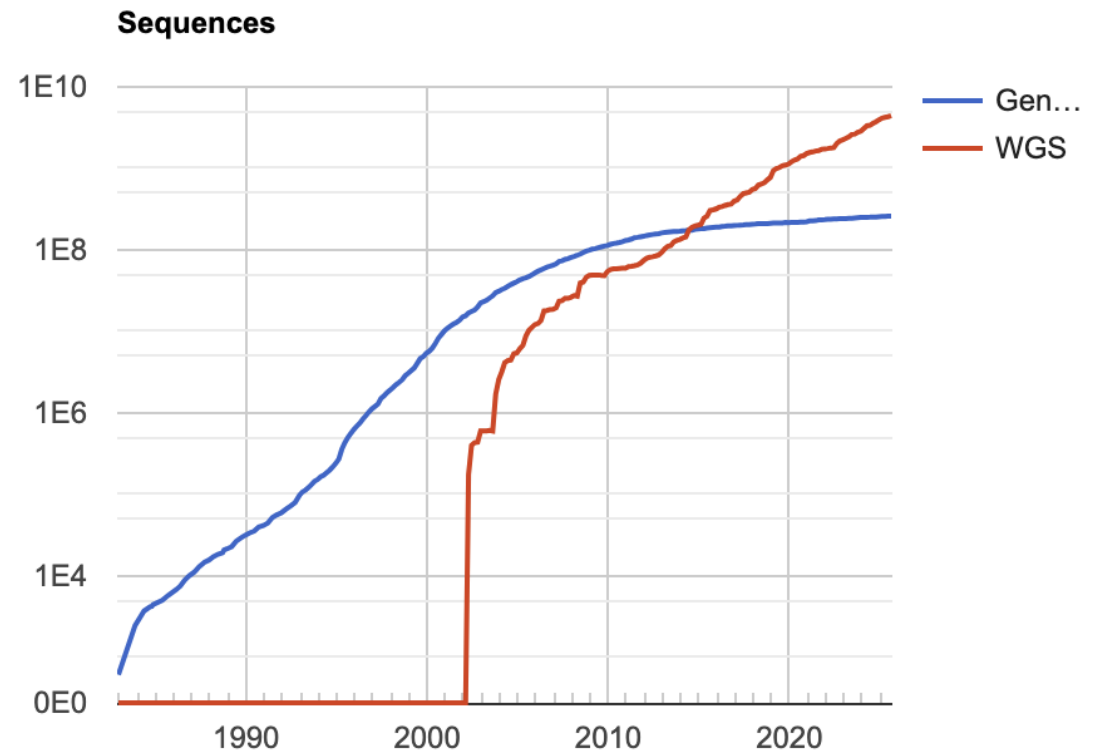
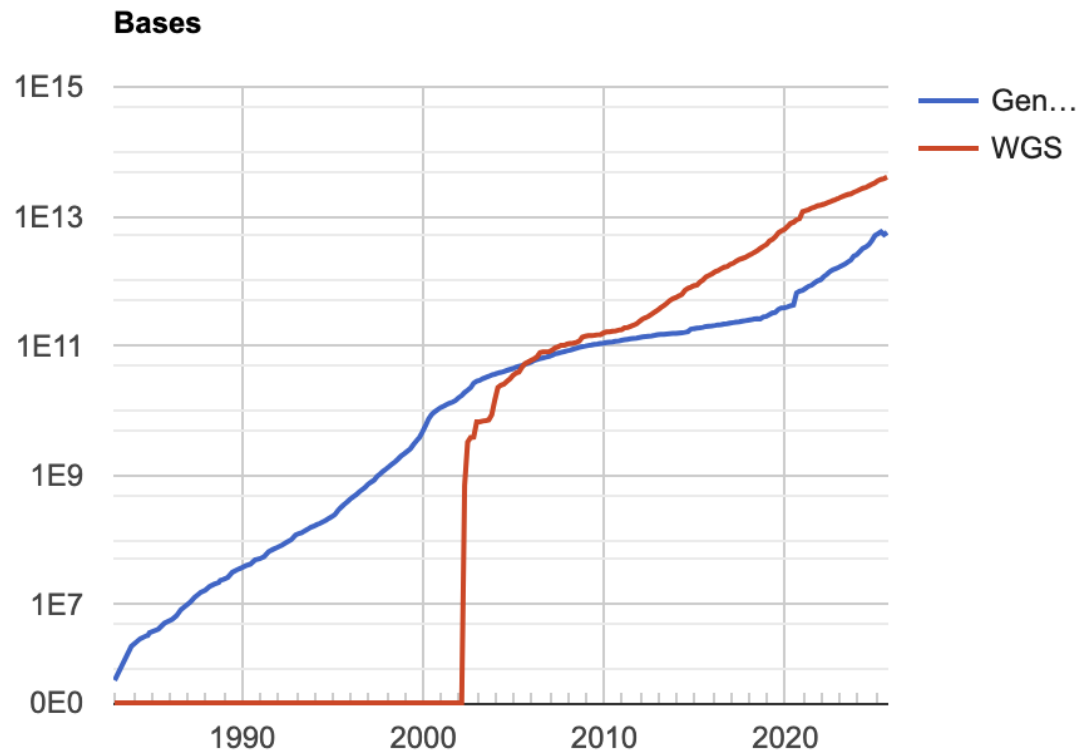
- В 2000 году секвенировали первый геном человека за 13 лет и \$3 млрд.
- Сегодня это делают за 1 день и \$200.
- В 2023 году произведено больше генетических данных, чем за всю историю до 2015.

Что такое биоинформатика

– это междисциплинарная область знаний, которая объединяет:

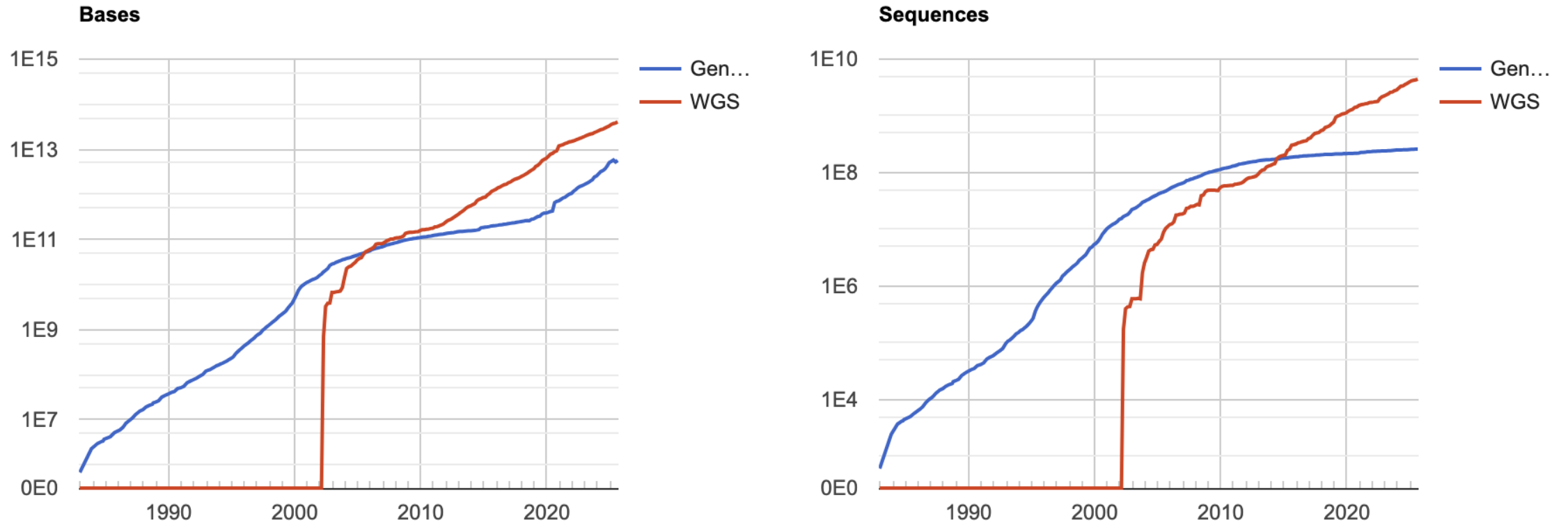
- Математику и математическую статистику
- Информатику
 - Кибернетику
 - Машинное обучение
- Общую биологию
 - Молекулярную биологию
 - Генетику
- Химию

Проблема объёма данных



40 3903483406298 оснований и 44413337 387 последовательностей в базе WGS WGS
 5 67560676778413 оснований 2558 320 620 последовательностей в базе BankBank

Проблема объёма данных



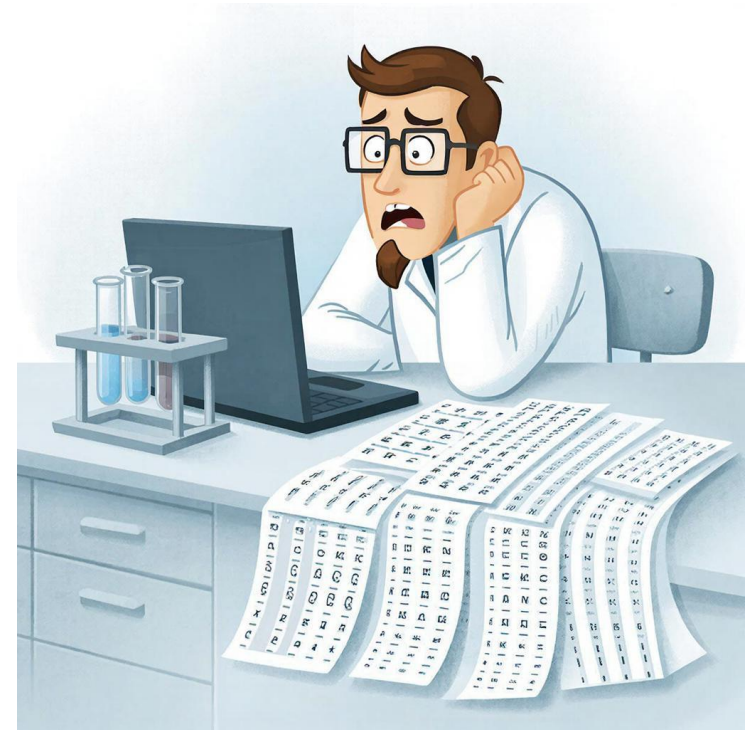
Хранилище NCBI (GenBank) удваивается каждые 18 месяцев –
закон Мура для ДНК

Проблема объёма данных

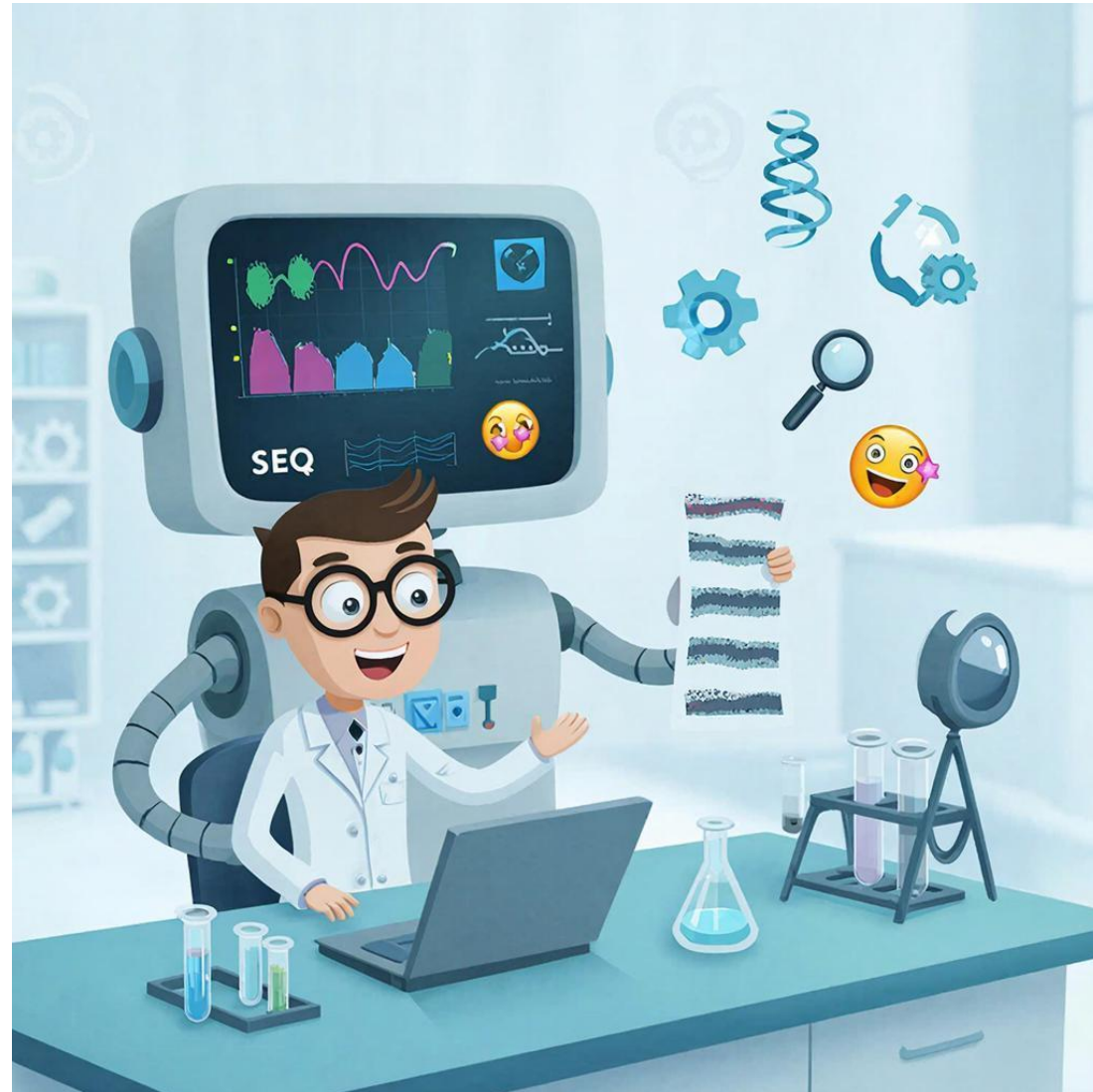
- Один секвенатор (например, Illumina NovaSeq) выдает 20 терабайт данных за сутки — это как:
 - ~ 7 000 000 экземпляров «Война и мир»
 - ~ 136 трилогий фильмов «Властелин колец» в 4K
 - ~ 64 всех частей фильмов «Гарри Поттер» в 4K



Проблема объёма данных



Проблема объёма данных



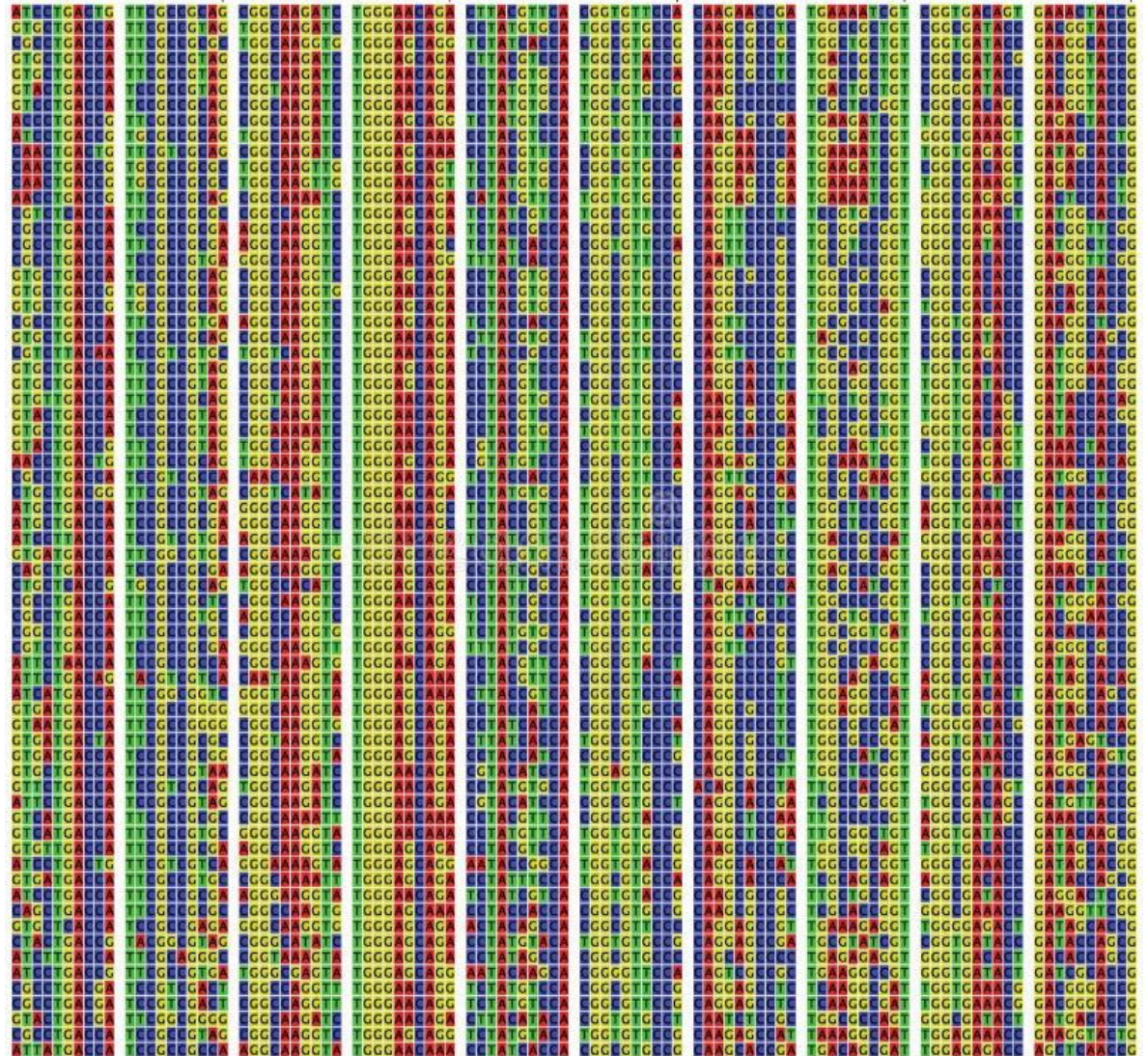
Суперсила №1: Поиск родственников

- **BLAST** — «Гугл для ДНК».
- Берём неизвестный ген → за секунды находим похожие гены у мыши, банана, бактерии.
- Например так нашли ген, отвечающий за устойчивость к засухе у пшеницы, сравнив с геном риса. И функции многих генов, как раз определяются при сравнении с генами других организмов.

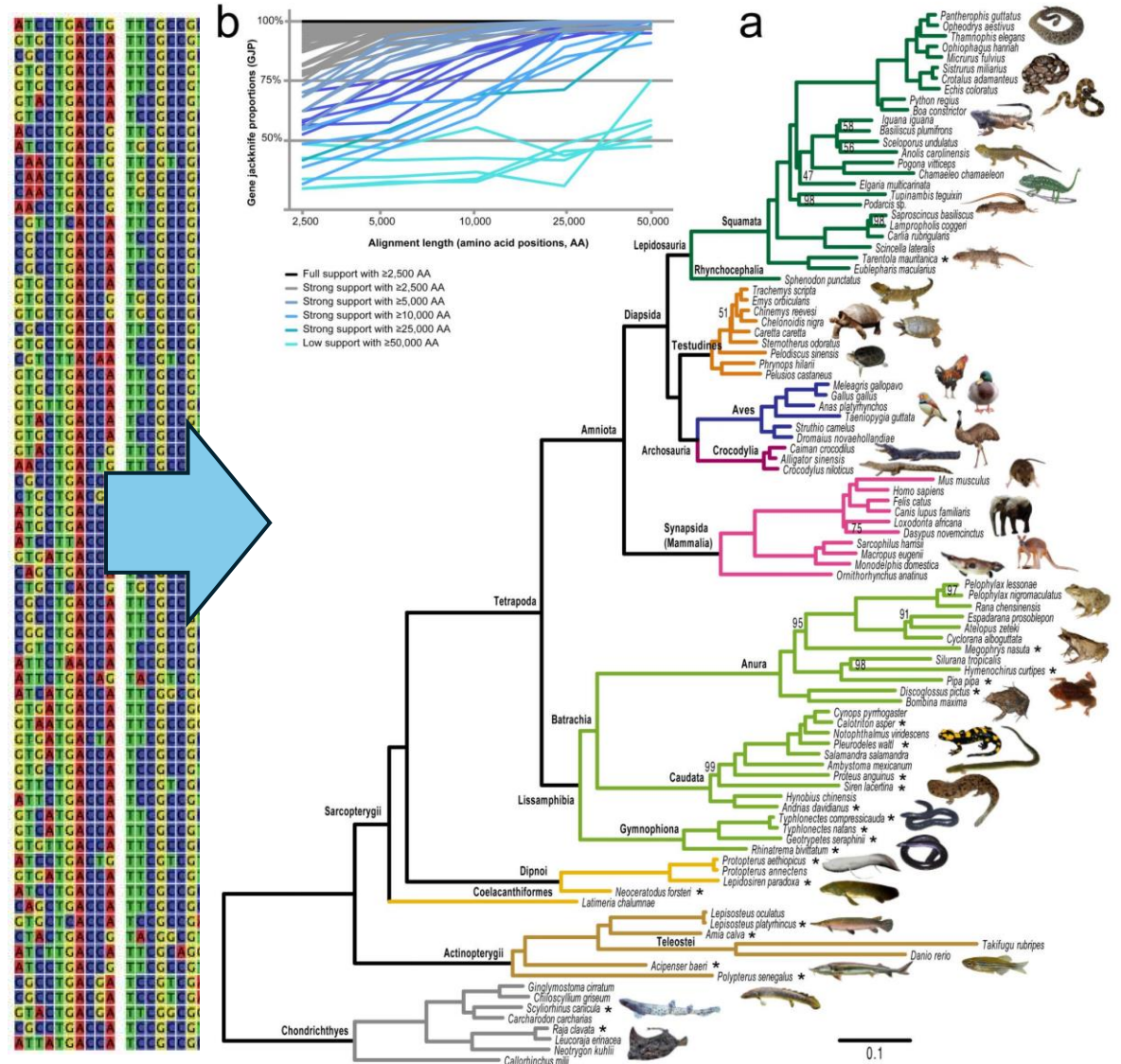
The screenshot displays the BLAST website interface. At the top, the NIH National Library of Medicine logo is visible, along with a 'Log in' button. The main heading is 'BLAST®' with navigation links for 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. The central section is titled 'Basic Local Alignment Search Tool' and includes a brief description of the tool's function. A 'NEWS' sidebar on the right highlights a recent update: 'Mon, 21 Jul 2025' and 'Download BLAST+ 2.17.0 now!'. Below this, the 'Web BLAST' section offers three main options: 'Nucleotide BLAST' (nucleotide to nucleotide), 'blastx' (translated nucleotide to protein), and 'Protein BLAST' (protein to protein). At the bottom, the 'BLAST Genomes' section features a search input field with a 'Search' button and a list of organism categories: Human, Mouse, Rat, and Microbes.

Суперсила №2: Сравнение геномов

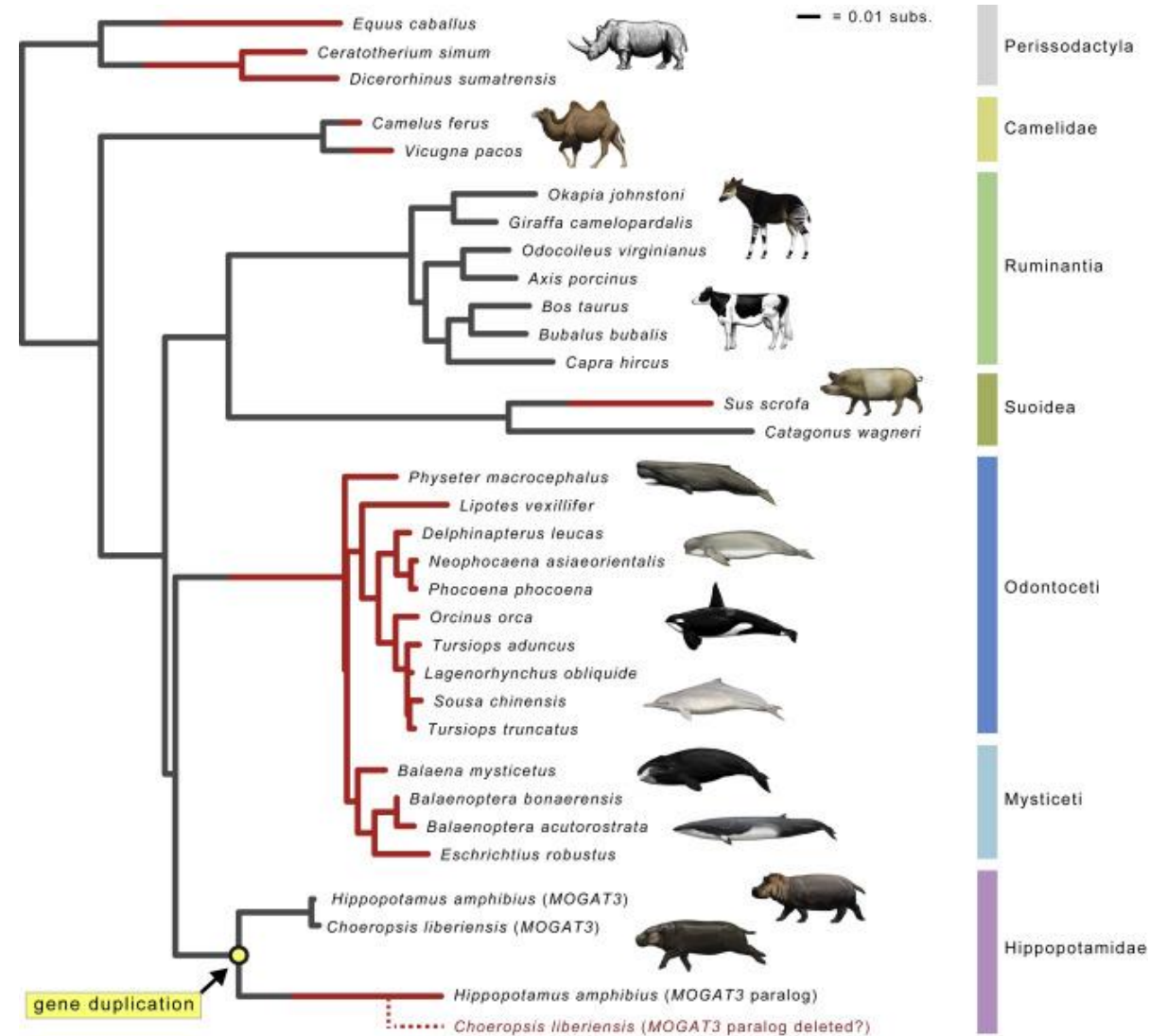
- Выравнивание последовательностей.
- Поиск различий (опечаток) в двух почти одинаковых текстах «Войны и мира» (но текстов 10 000 раз больше).
- Мы ищем и находим мутации, вставки, удаления. Так можно определить, почему один человек болеет, а другой нет.



Суперсила №2: Сравнение геномов

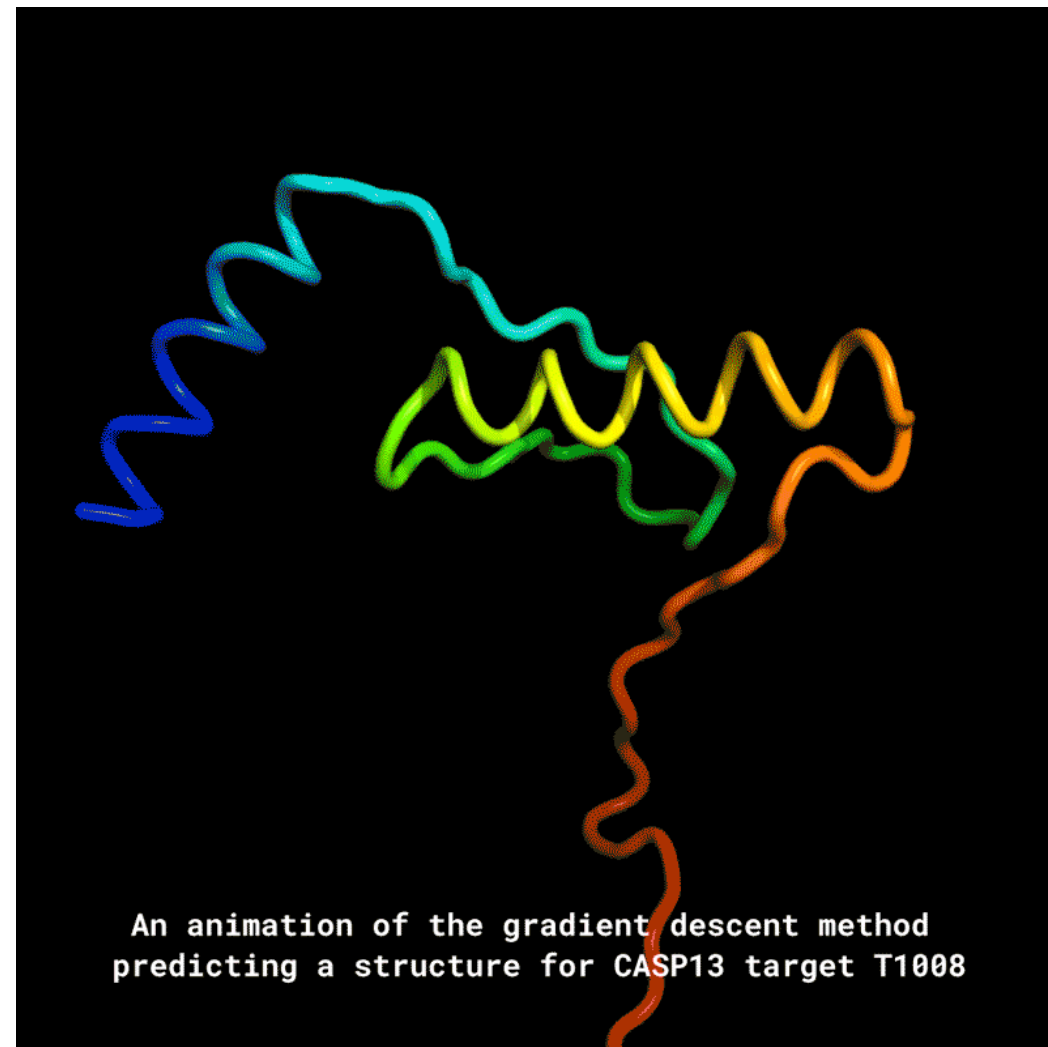


Суперсила №2: Сравнение геномов



Суперсила №3: Предсказание 3D-структуры белков

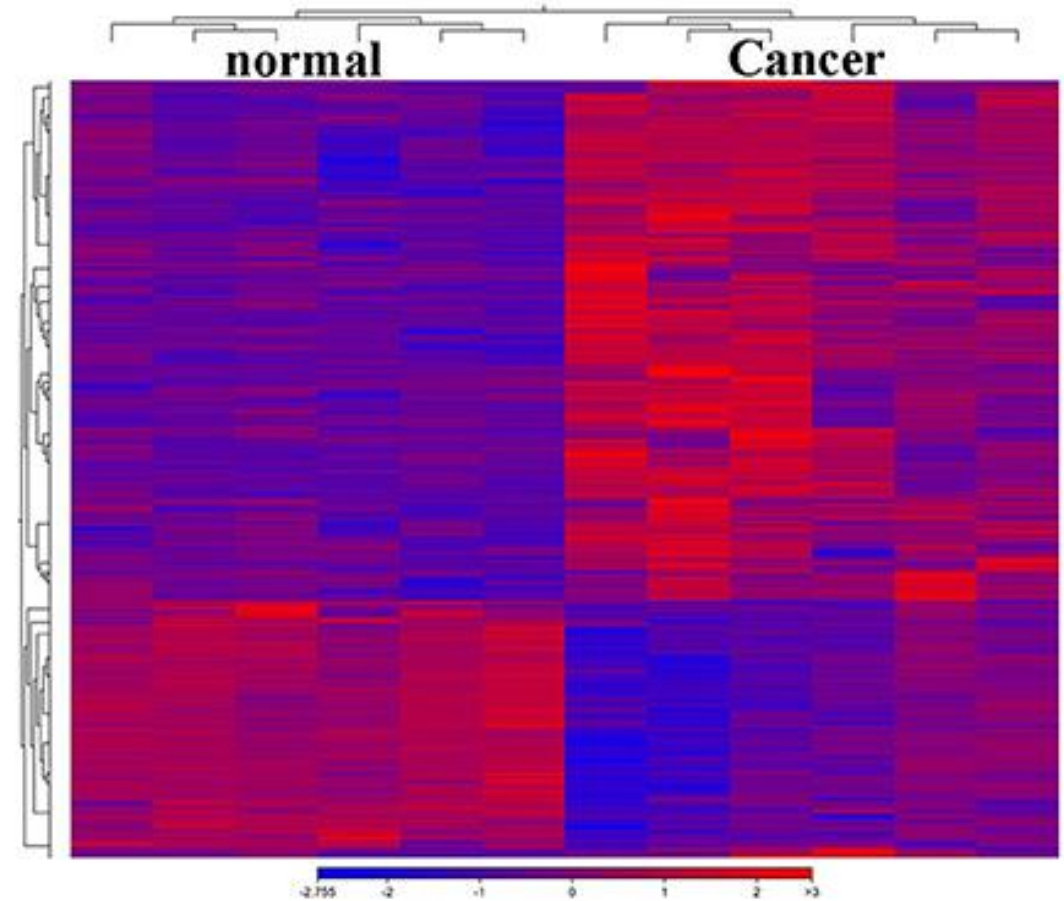
- Белок – цепочка аминокислот, которая сворачивается в сложную фигуру.
- Возможных вариантов сворачивания – больше, чем атомов во Вселенной.
- Революция: ИИ AlphaFold (2021) предсказал структуру 200 млн белков. Раньше на один белок уходили годы в лаборатории.



Суперсила №4: Анализ «Включено / Выключено»

- RNA-seq – измеряем, какие гены работают в клетке.
- Тепловая карта: видно, что раковая клетка отличается от здоровой по профилю активности сотен генов.
- Без компьютера это не проанализировать – слишком много данных.

A



Найди мутацию

АСТГАСТГАСТГАСТ**Г**АСТГАСТ**Г**АСТГАСТГАСТГ
АСТГАСТГАСТГАСТ**А**АСТГАСТ**С**АСТГАСТГАСТГ

Онкология и персонализированная медицина

- У рака легких пациента берут биопсию, секвенируют опухоль.
- Программа ищет драйверные мутации (например, в гене EGFR).
- Находит известную мутацию → назначают таргетный препарат (например, ирлотиниб), а не универсальную химиотерапию.
- Результат: эффективность с 30% до 80%, меньше побочных эффектов.



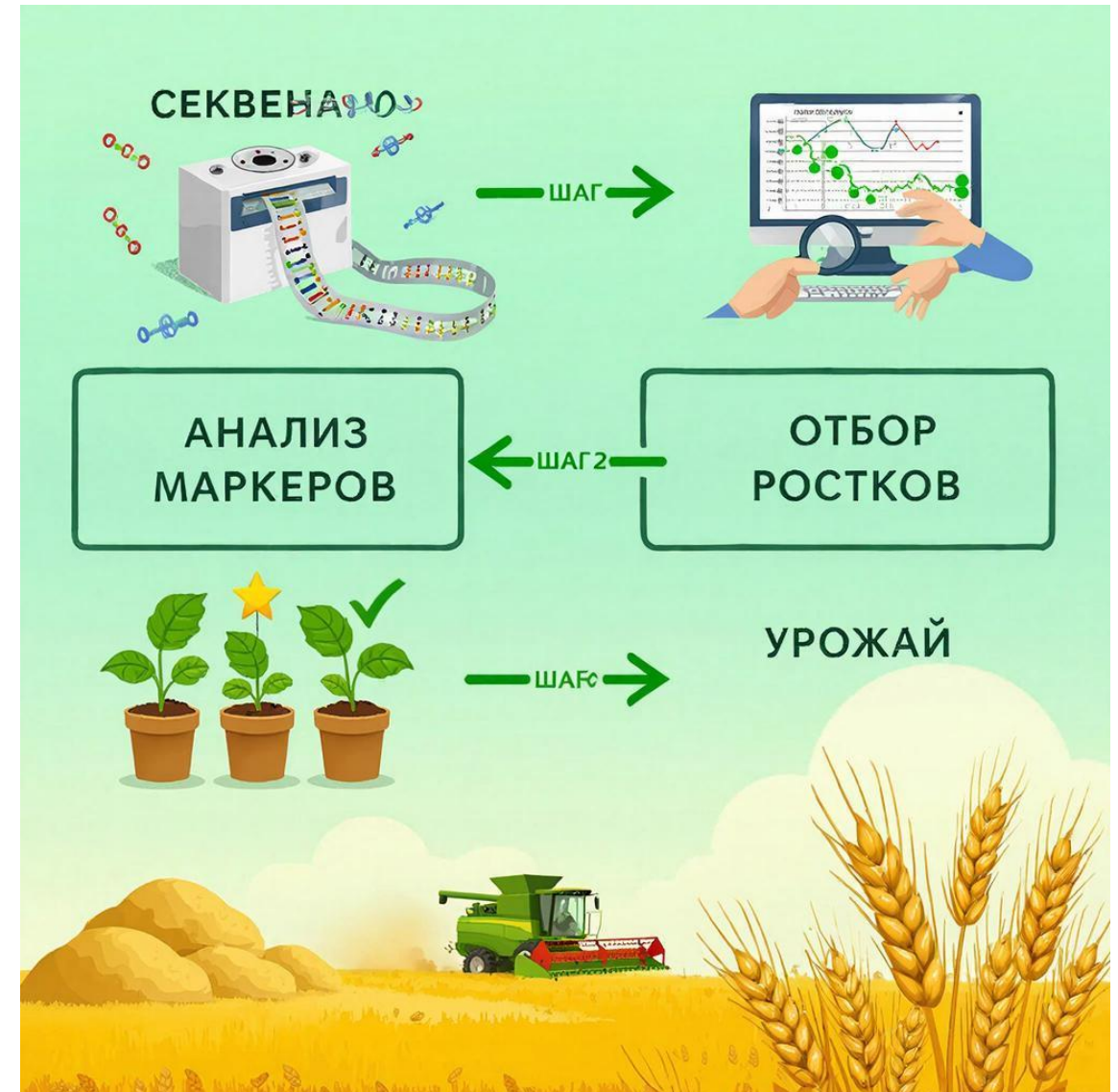
Поиск новых антибиотиков

- Традиционный метод: 10 лет и \$1 млрд.
- Новый метод: ИИ обучается на тысячах известных молекул, предсказывает, какие пептиды убьют бактерию, но не тронут наши клетки.
- Учёные Ставропольского государственного медицинского университета с помощью ИИ создали более 200 потенциальных антибиотических пептидов из которых 5 были отобраны в ходе дальнейшего биоинформатического анализа, а 2 прошли проверку *in vivo*
- Пример: в 2022 году создан антибиотик цилагицин



Сельское хозяйство – супер-культуры без ГМО-спор

- Традиционная селекция — 10–15 лет на новый сорт. Нужно скрещивать, сеять тысячи растений, ждать урожая. А изменения климата требуют быстрых решений
- Решение No1: Маркер-ориентированная селекция (MAS)
- Решение No2: Геномная селекция (GWAS)
- Решение No3: Поиск генов диких предков



Чем реально занимается биоинформатик на работе

- Поставить ПЦР
- Написать скрипт на Python или R, который обрабатывает 10 000 файлов с ДНК.
- Визуализировать данные (графики в R или Python).
- Покушать
- Работать в Linux на сервере (или в облаке).
- Скачать данные из баз (NCBI, ENA, PDB).
- Сделать статистический тест, чтобы отличить настоящий сигнал от шума.



Инструменты

Unipro



GENE



e! **Ensembl**

 **Lincoln**

Как стать биоинформатиком

- Можно прийти с двух сторон:
 - Биолог, подучивший программирование (курсы Stepik по Python и биоинформатике).
 - Программист, выучивший биологию (сложнее, но тоже возможно).
 - Математик, выучивший биологию.
- Совет: Сейчас заглянуть на сайт Rosalind.info – задачи по биоинформатике с автопроверкой (от простых к сложным).
- ВУЗы: биоинформатика есть в МГУ, ИТМО, ВШЭ, Сколтехе, УрФУ и география дисциплины ширится.

stepik Каталог Моё обучение Преподавание

Уровень сложности

- Для начинающих
- Для продолжающих
- Для профи
- Для всех

Цена

От 0 Р До 250000 Р

Бесплатно до 5 000 до 10 000

Только со скидкой

Только с сертификатом

Язык

- Любой
- Русский
- Английский

Тип обучения

Программирование на Python
Bioinformatics Institute
Курс посвящен базовым понятиям и элементам языка программирования Python (операторы, числовые и строковые переменные, списки, условия и циклы). Курс является вводным и наиболее подойдет слушателям, не имеющим опыта написания программ ни на одном из языков программирования.
93/100

Введение в биоинформатику: метагеномика
Санкт-Петербургский государственный университет
Курс посвящен введению в метагеномику и затрагивает основные теоретические и практические подходы анализа видового состава, генетического материала и функционала сообществ микроорганизмов, обитающих в окружающей среде.
33/43

Medical Genomics and Bioinformatics
Benjamin Rodriguez Santiago
Module of the Master of Advanced Genetics. Practical afternoon sessions using web resources to manage and analyse genomics data.
302 11 ч

Bioinformatics I: Finding Hidden Messages in DNA
This course is the first part of the Bioinformatics Specialization. Here is a series of classes illustrating the power of

Rosalind About Problems Statistics Glossary search f t Log in Register

Locations

Rosalind is a platform for learning bioinformatics and programming through problem solving. [Take a tour](#) to get the hang of how Rosalind works.

If you don't know anything about programming, you can start at the [Python Village](#). For a collection of exercises to accompany Bioinformatics Algorithms book, go to the [Textbook Track](#). Otherwise you can try to storm the [Bioinformatics Stronghold](#) right now.



If you are completely new to programming, try these initial problems to learn a few basics about the Python programming language. You'll get familiar with the operations needed to start solving bioinformatics challenges in the Stronghold.



Discover the algorithms underlying a variety of bioinformatics topics: computational mass spectrometry, alignment, dynamic programming, genome assembly, genome rearrangements, phylogeny, probability, string algorithms and others.



Ready-to-use software tools abound for bioinformatics analysis. Whereas in the Stronghold you implement algorithms on your own, in the Armory you solve similar problems by using existing tools.

Bioinformatics Textbook Track

A collection of exercises to accompany Bioinformatics Algorithms: An Active-Learning Approach by Phillip Compeau & Pavel Pevzner. A full version of this text is hosted on [stepik.org](#)



Algorithmic Heights

A collection of exercises in introductory algorithms to accompany "Algorithms", the popular textbook by Dasgupta, Papadimitriou, and Vazirani.

«Большие вызовы» и «Иннагрика»

Ген коровьего каппа-казеина небольшой по размеру, что делает его хорошим кандидатом для экспрессии в бактериях.

Разработайте плазмиду

для продукции каппа-казеина в *E. coli* и методы её верификации, выполнив нижеперечисленные задания.

Найдите последовательность гена,

кодирующего каппа-казеин в геноме коровы, и решите, какими рестриктазами его извлечь.

Подберите исходную плазмиду

для включения целевого гена. Обоснуйте свой выбор.

Сформируйте карту плазмиды

с включенным целевым геном.

Предложите протокол

рестрикционного анализа для подтверждения включения целевого гена в плазмиду.

Выберите вектор для экспрессии.

Найдите ген ксиланазы

в данном микроорганизме (если это не удастся, то выберите ближайшую по вашему мнению доступную последовательность ксиланазы).

Опишите, как и с помощью каких праймеров вы будете клонировать нуклеотидную последовательность данного фермента.

Заключение: биолог 2030 года

- Это человек, который утром делает ПЦР, днём пишет код на Python или R, а вечером интерпретирует данные с секвенатора.
- Главная мысль: компьютер не заменяет биолога, он дает суперсилы. С ним один человек делает работу целой лаборатории 90-х годов.
- Цитата: «Биология становится наукой о данных. Тот, кто умеет программировать, увидит то, что другие пропустят».



Спасибо за внимание!