

✨ биоинформатика ✨

лекция 1, вводная



биоинформатика VS вычислительная биология

- связана с биологическими макромолекулами
- анализ последовательности, структуры, функций ДНК, РНК, белков

- связана со всем на свете (биологическим)

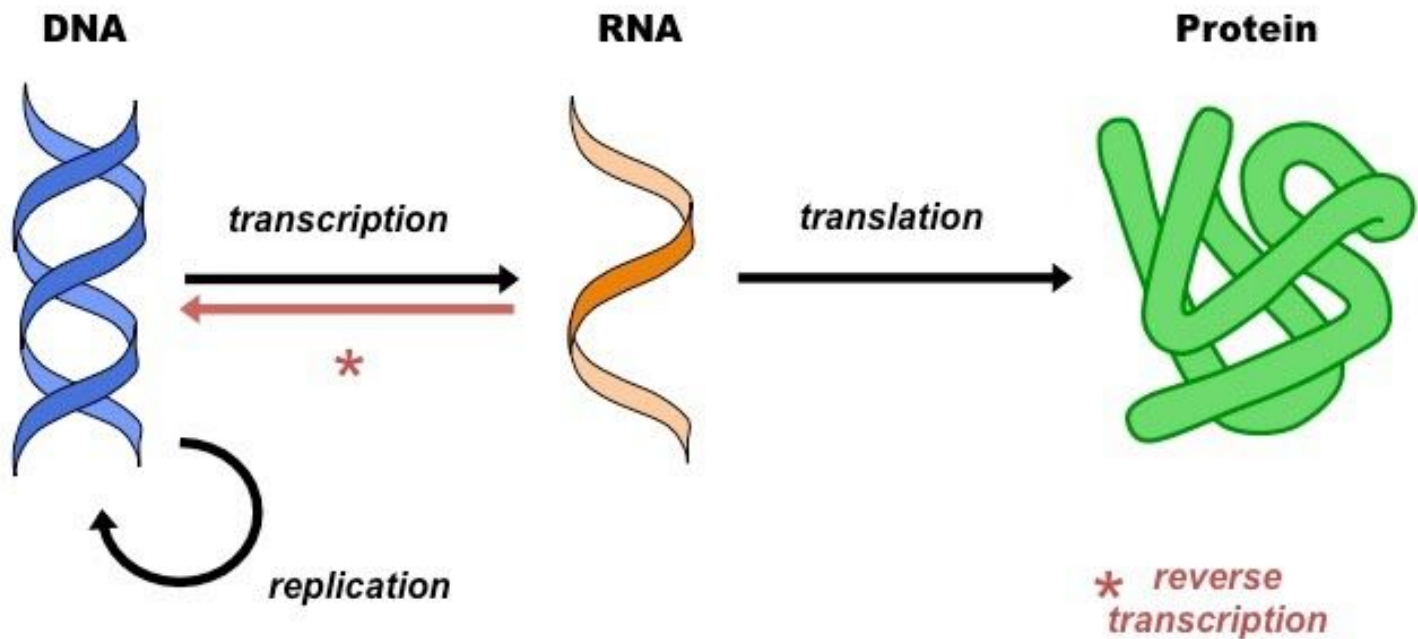


<https://ncase.me/trust/>

почему биоинформатика это здорово ?

1. молодая наука
2. в России цветет и пахнет !
3. на пике технологий

центральная догма молекулярной биологии

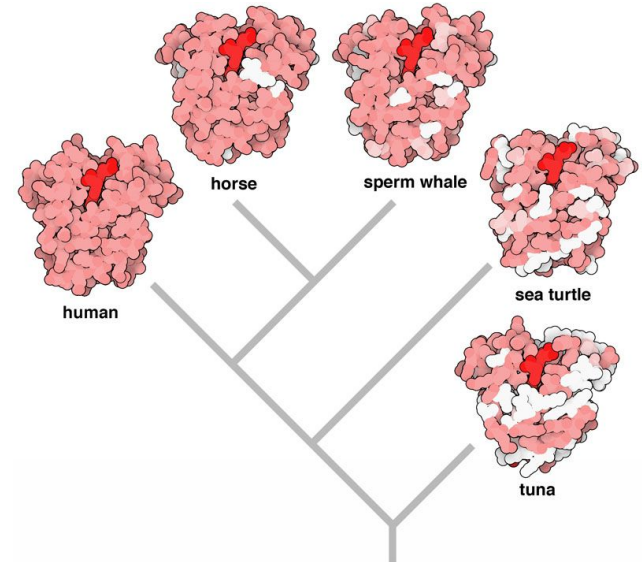
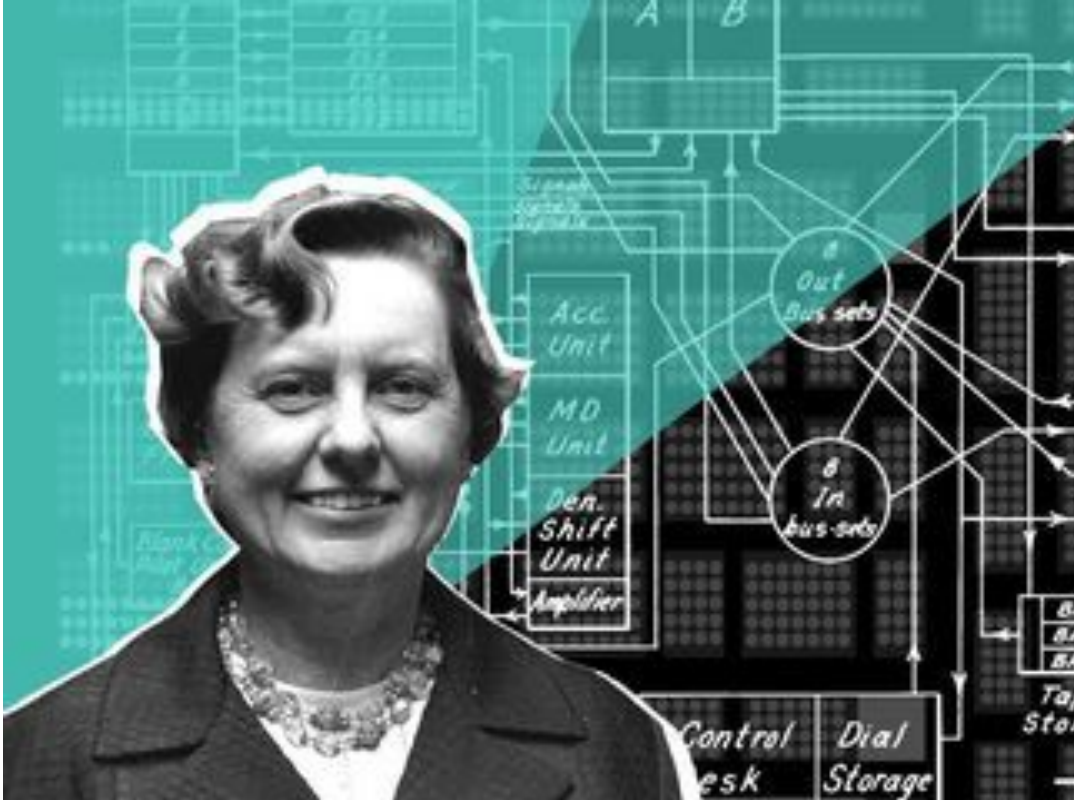


краткий экскурс в историю

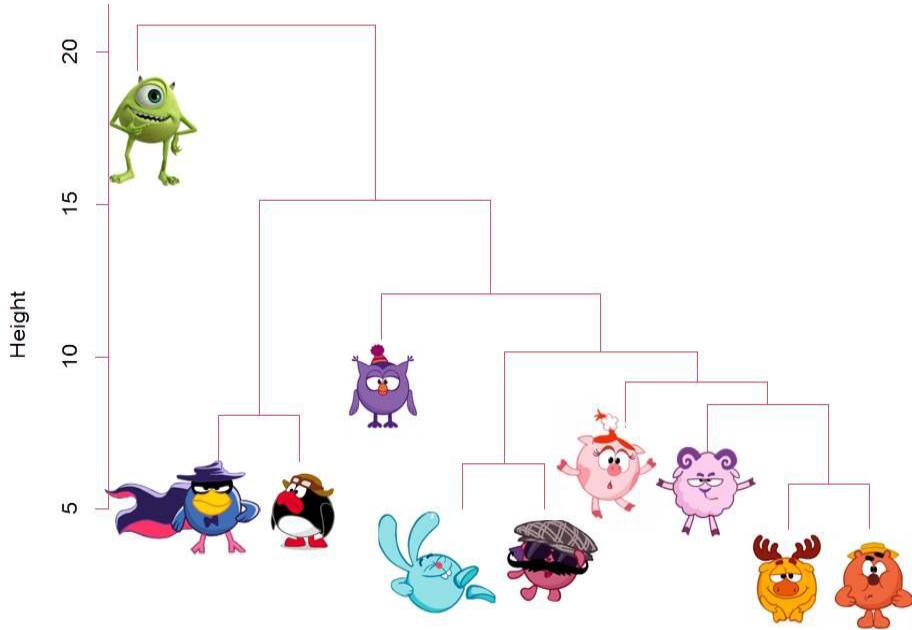
“The culture of statistical analysis, let alone of digital computing, were completely foreign to most biochemists. Some even prided themselves in not being ‘theorists,’ which is how they understood data analysis using mathematical models.”

- Bruno Strasser

1965 - Margaret Dayhoff - Atlas of Protein Sequence and Structure

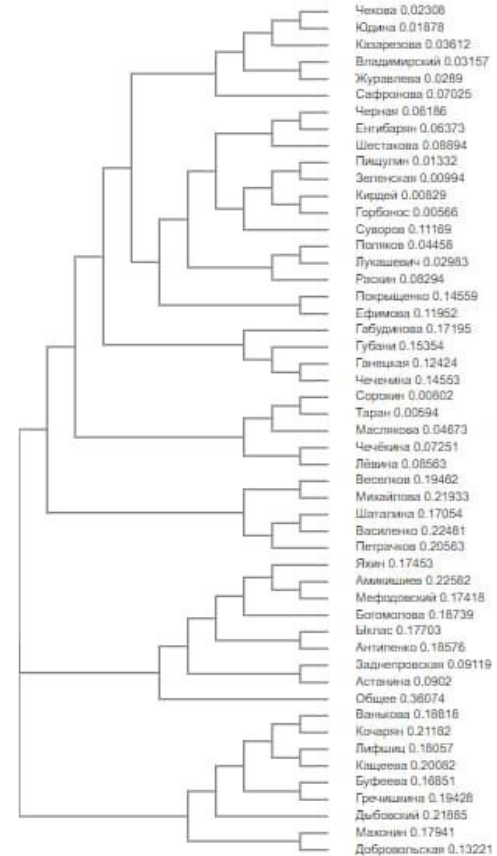


дерево, построенное по фенотипу



https://vk.com/biochemistry_mems?w=wall-143120001_315032&z=photo-143120001_457286103%2Falbum-143120001_00%2Frev

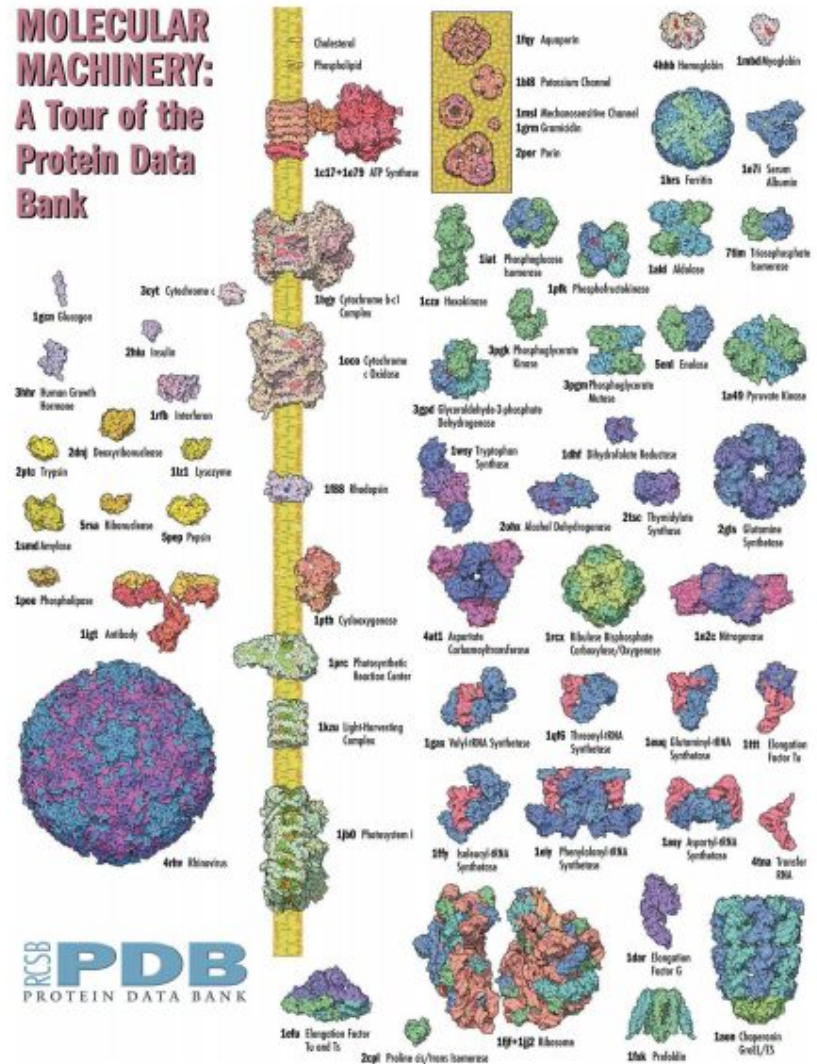
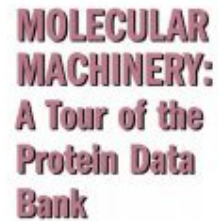
дерево, построенное по последовательностям



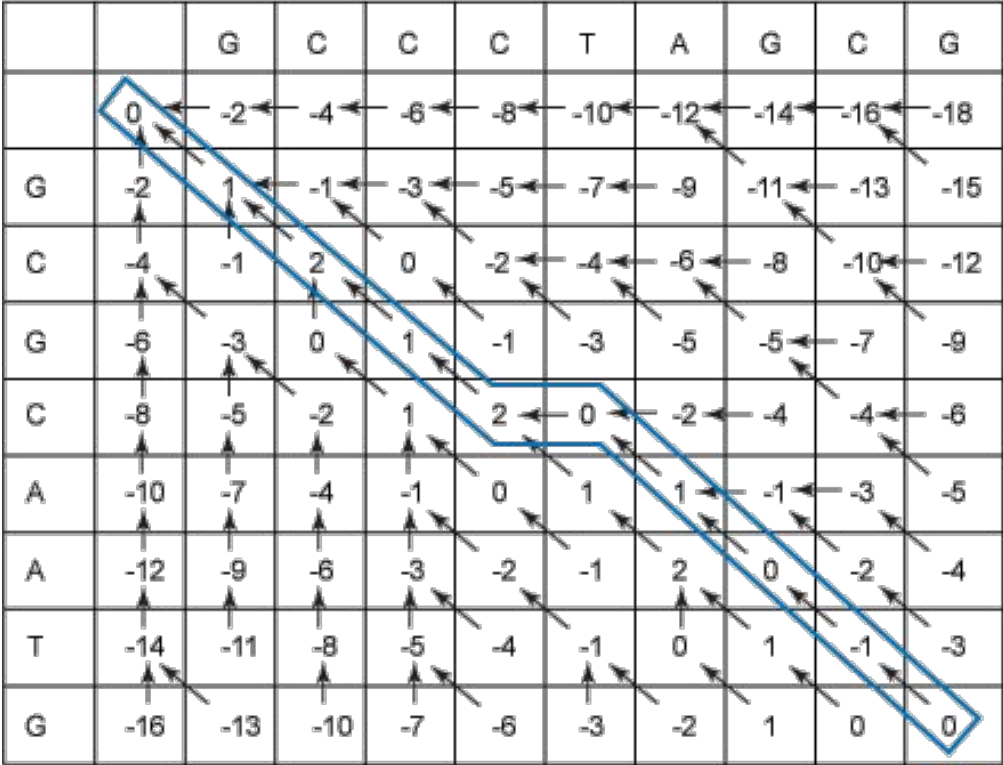
RCSB PDB Deposit Search

RCSB PDB
PROTEIN DATA BANK

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatR Unified Data Resource



1970 - Needleman & Wunch - first alignment algorithm

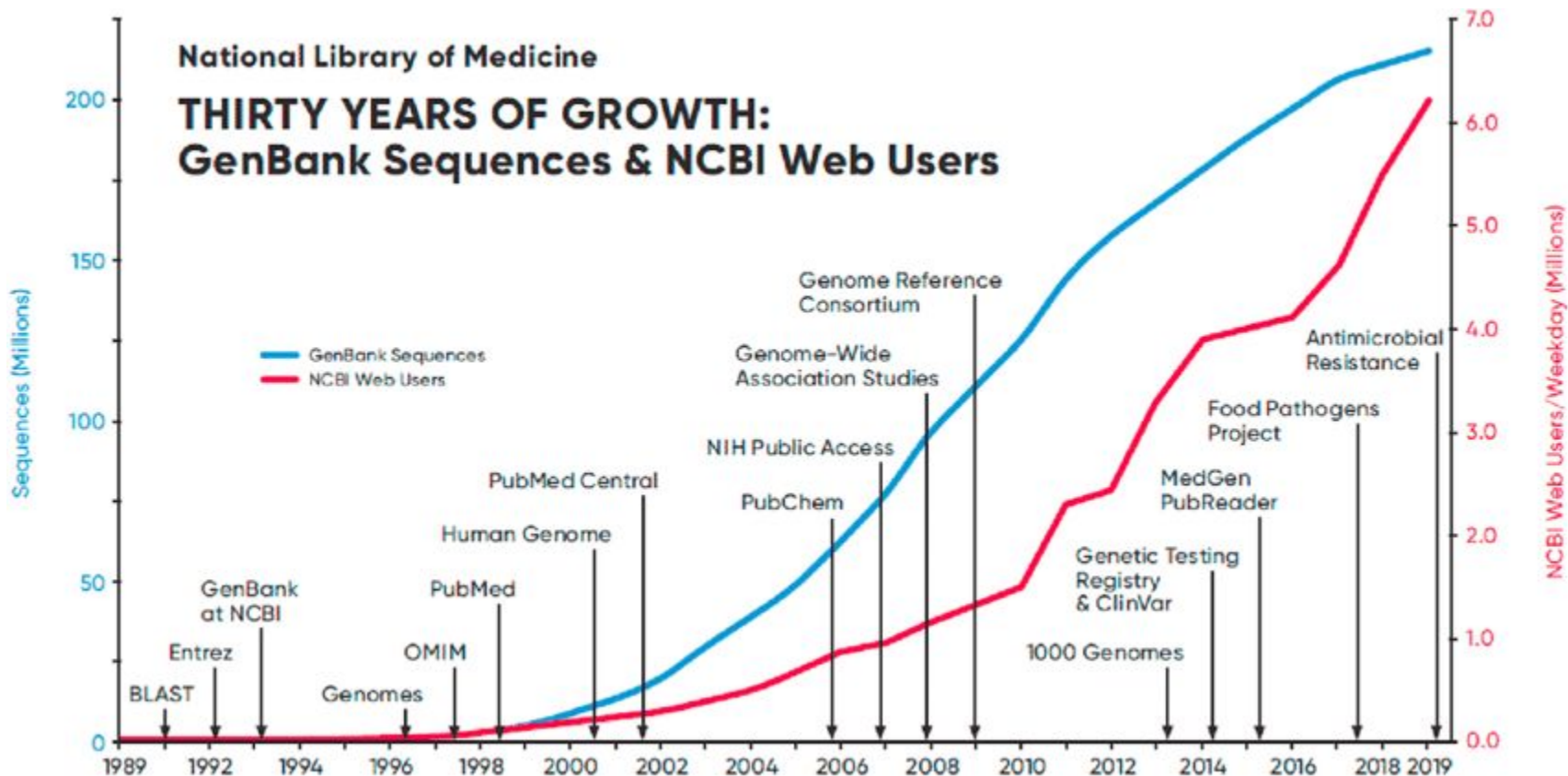


1975 - Chou & Fasman - protein structure prediction algorithm

Chou-Fasman Parameters

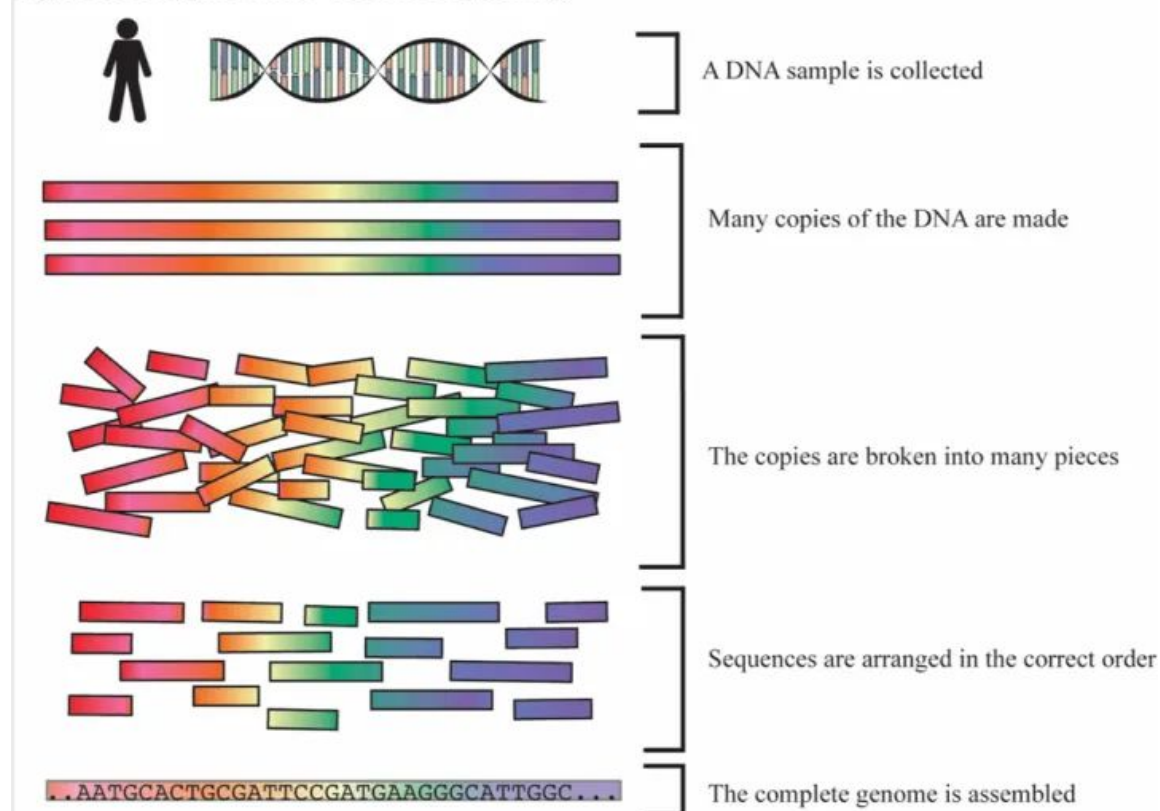
Residue	P α	Residue	P β	Residue	P t
Glu	1.51	Val	1.70	Asn	1.56
Met	1.45	Ile	1.60	Gly	1.56
Ala	1.42	Tyr	1.47	Pro	1.52
Leu	1.21	Phe	1.38	Asp	1.46
Lys	1.16	Trp	1.37	Ser	1.43
Phe	1.13	Leu	1.30	Cys	1.19
Gln	1.11	Cys	1.19	Tyr	1.14
Trp	1.08	Thr	1.19	Lys	1.01
Ile	1.08	Gln	1.10	Gln	0.98
Val	1.06	Met	1.05	Thr	0.96
Asp	1.01	Arg	0.93	Trp	0.96
His	1.00	Asn	0.89	Arg	0.95
Arg	0.98	His	0.87	His	0.95
Thr	0.83	Ala	0.83	Glu	0.74
Ser	0.77	Ser	0.75	Ala	0.66
Cys	0.70	Gly	0.75	Met	0.60
Tyr	0.69	Lys	0.74	Phe	0.60
Asn	0.67	Pro	0.55	Leu	0.59
Pro	0.57	Asp	0.54	Val	0.50
Gly	0.57	Glu	0.37	Ile	0.47

1980 - появление GenBank

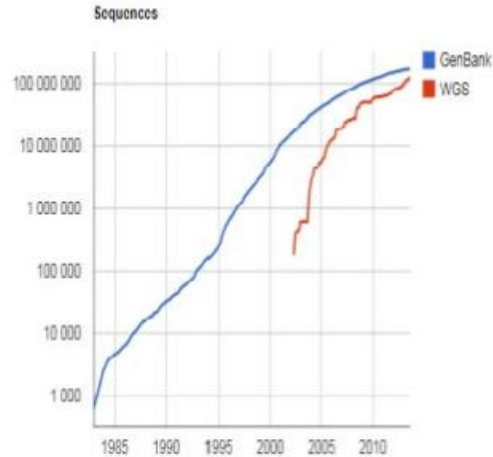
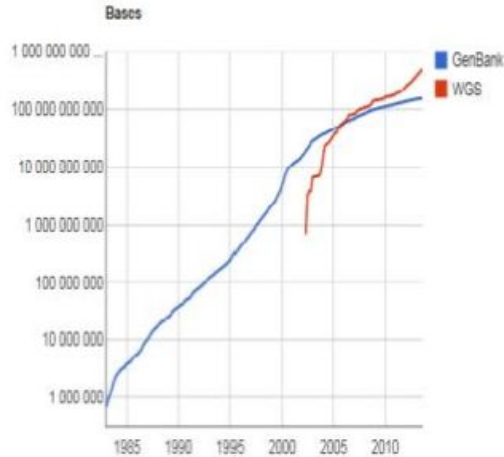


late 1980s - Human Genome Project

Figure 2: Shotgun Whole-Genome Sequencing



зачем нужна биоинформатика?



M. Kchouk and F. Mhamdi, "New Online Hierarchical Feature Extraction Algorithm for Classification of Protein," *2014 25th International Workshop on Database and Expert Systems Applications*, 2014, pp. 10-14, doi: 10.1109/DEXA.2014.20.

How many GB is a human genome?

1.5 Gigabytes

In order to represent the entire diploid human genome in terms of bytes, we can perform the following calculations: 6×10^9 base pairs/diploid genome \times 1 byte/4 base pairs = 1.5×10^9 bytes or **1.5 Gigabytes**, about 2 CDs worth of space! 16 мар. 2012 г.

биоинформатика в работе с ДНК

Основная технология — секвенирование ДНК (DNA-seq):

- весь геном — whole genome sequencing (WGS)
- весь экзом (то есть кодирующие участки ДНК) — whole exome sequencing (WES)
- несколько генов — targeted DNA-seq

Решаемые задачи:

- поиск и аннотация мутаций;
- copy number variation (CNV);
- поиск геномных ассоциаций с заболеваниями.

Пример: мутация в гене BRCA1 ассоциирована с риском развития рака молочной железы.

биоинформатике в работе с РНК

Основная технология — секвенирование РНК (RNA-seq);
(сводится к DNA-seq обратной транскрипцией)

Решаемые задачи:

- профилирование экспрессии генов в образце;
- поиск посттранскрипционных модификаций РНК (сплайсинг, редактирование РНК);
- поиск ассоциаций с заболеваниями.

Пример: тест-система OncotypeDX позволяет спрогнозировать рецидив рака молочной железы используя экспрессию 21 гена в первичной опухоли.

биоинформатика в работе с белками

- массовая количественная оценка белков в образце (proteomics);
- предсказание третичных структур (folding);
- предсказание констант и сайтов взаимодействий (docking);
- открытие и разработка лекарств (drug discovery).

биоинформатика последовательностей

- сборка геномов
- аннотация и анализ геномов новых организмов с использованием уже известных;
- наблюдение за эволюцией отдельных генов и их групп;

повседневные задачи: оценить специфичность праймеров ПЦР

