

INTRODUCTION TO BIOINFORMATICS

Moi University workshop

24th March 2023

Practical exercise – unmute your audio

Alphabet

Instructions

RAPRAB

Measure

CLAP ONCE

Language

CRAP

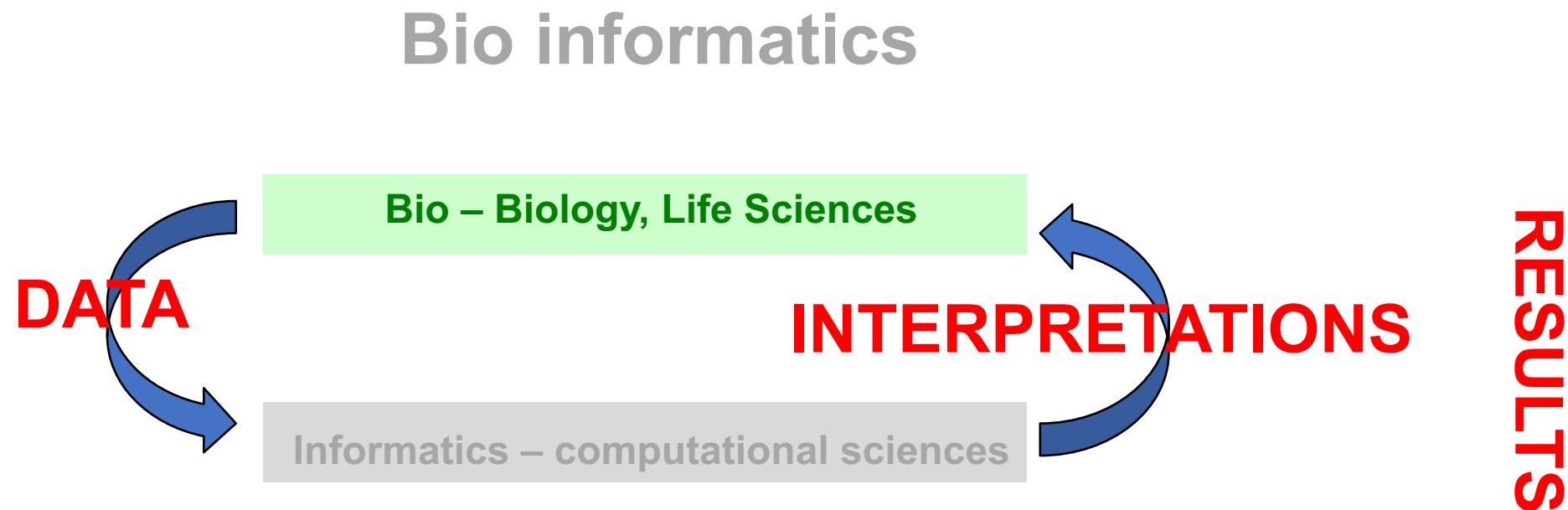
Typo/Rhythm Variation

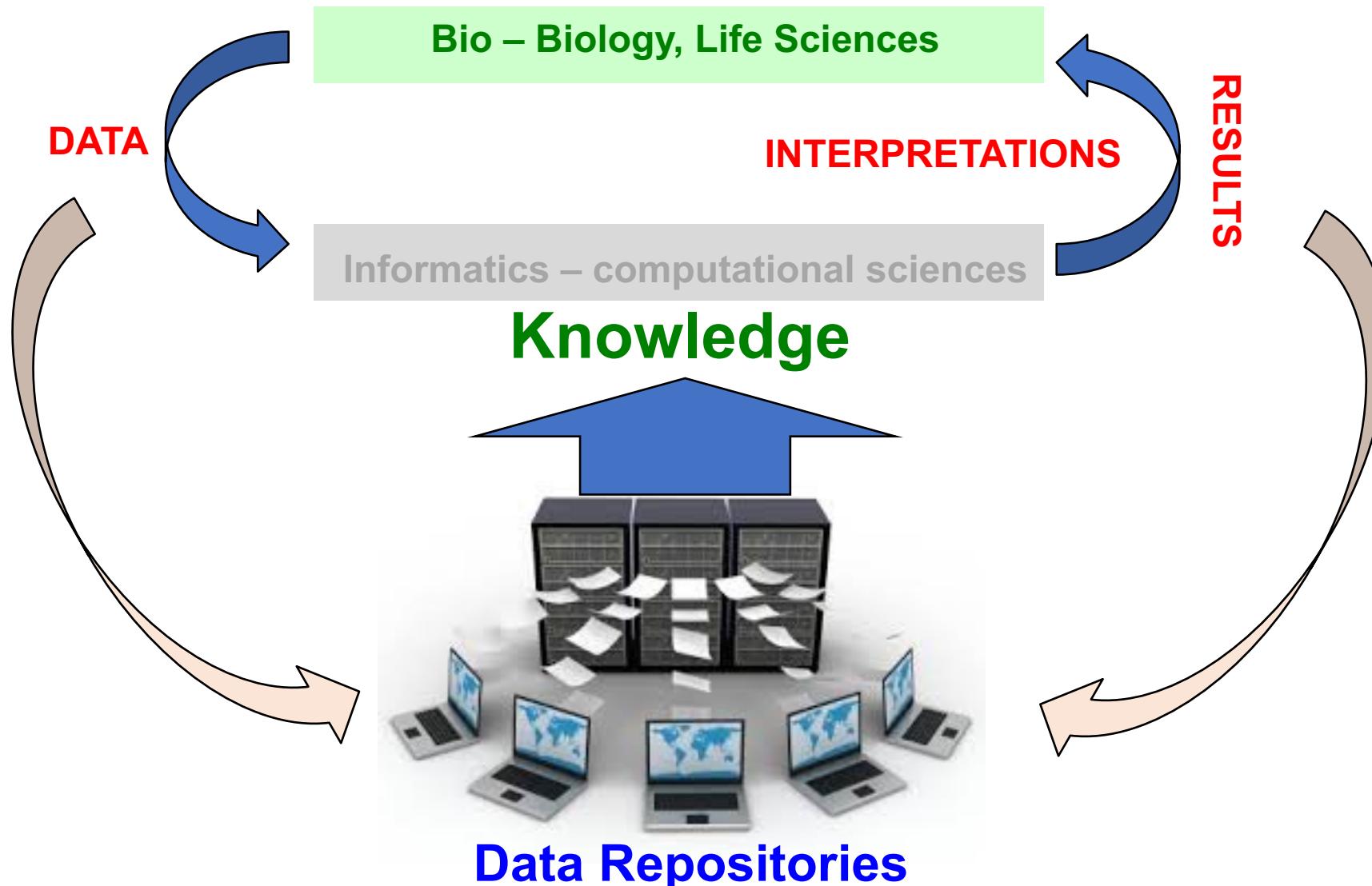


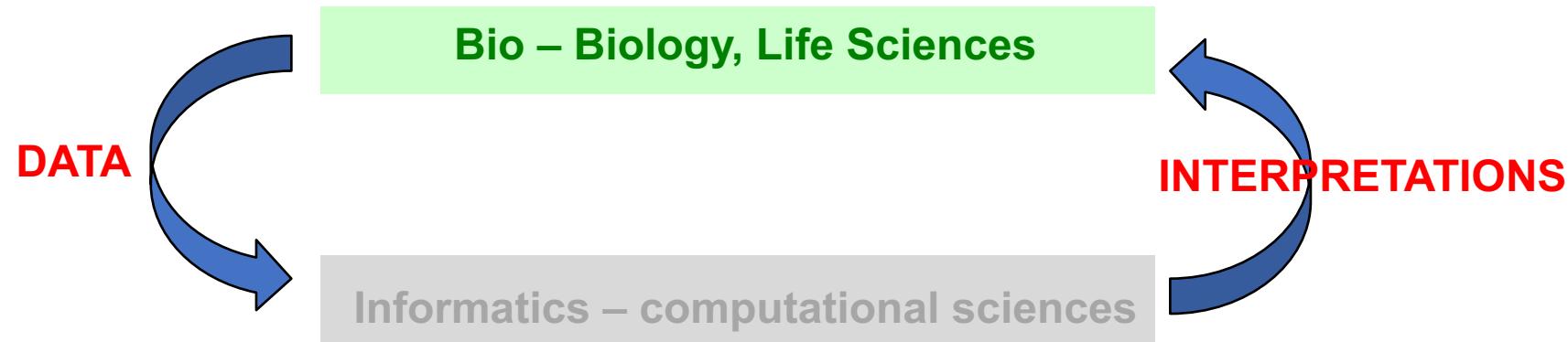
Overview

- Background of Bioinformatics
- Trends in Bioinformatics
- Fields of application of Bioinformatics

Background and definition





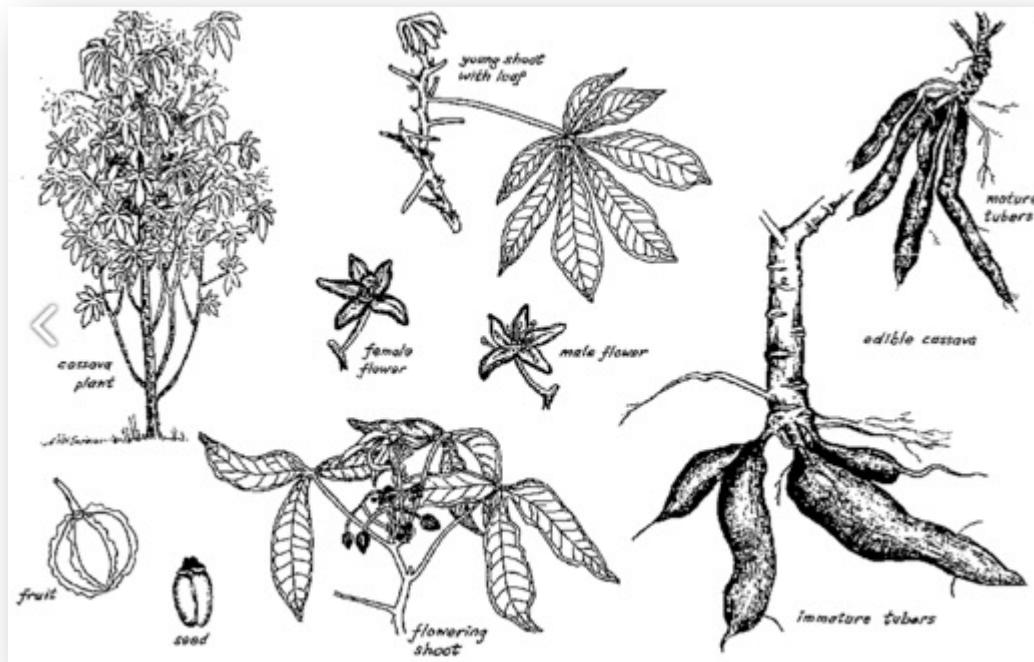


Bioinformatics is an interdisciplinary science that develops and improves on methods of analyzing biological data and storing, retrieving, organizing, and visualizing them.

This is in order to support to solve biological problems and discover the wealth of biological information hidden in biological data.

Types of Data

- Descriptions
 - Pictures

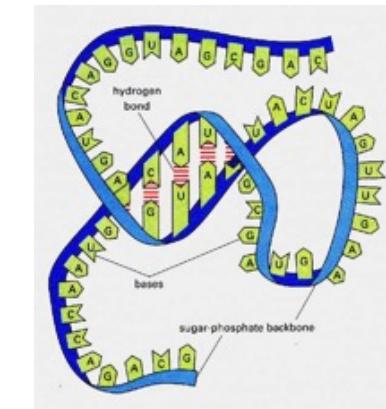
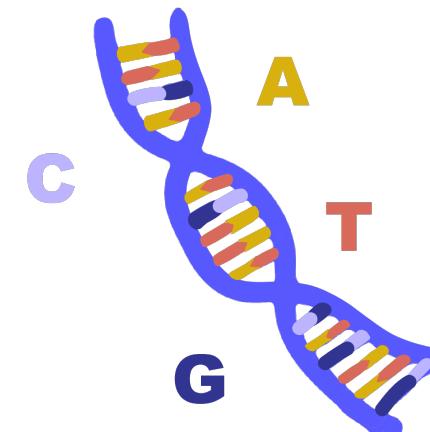


Types of Data

- Descriptions
- Pictures
- Sequences

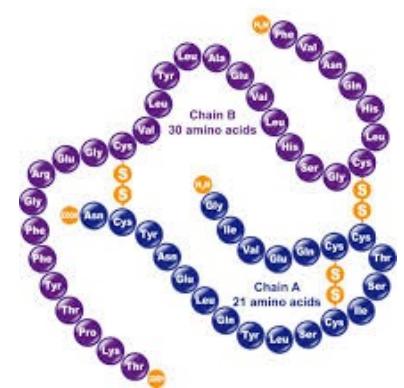
Types of Data

- Descriptions
- Pictures
- Sequences
 - DNA
 - RNA
 - Proteins



CENTRAL DOGMA

Human Insulin

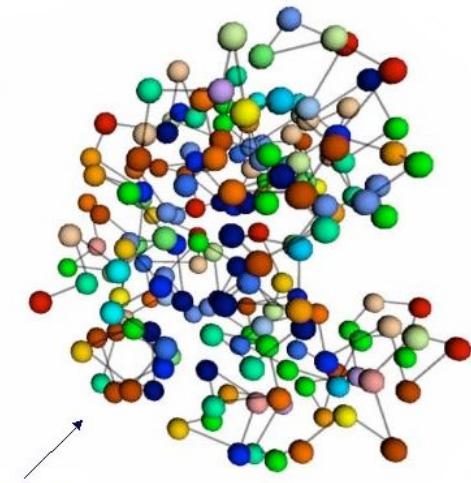


Types of Data

- Descriptions
- Pictures
- Sequences (DNA, RNA, and proteins)
- Structures (3D structures)

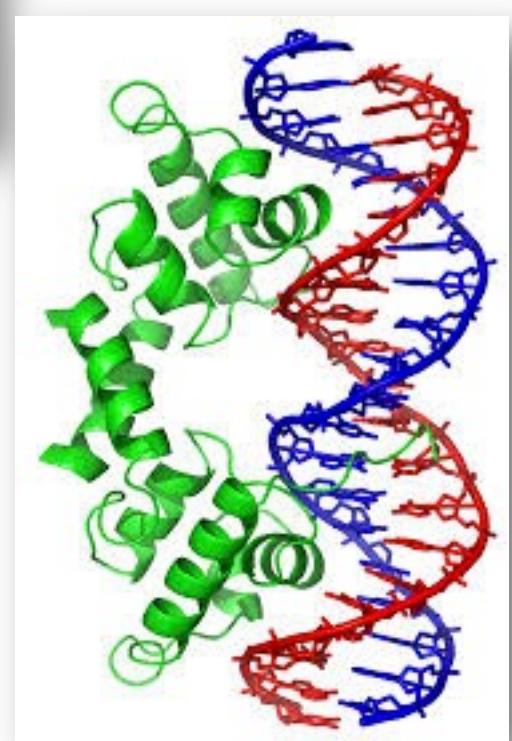
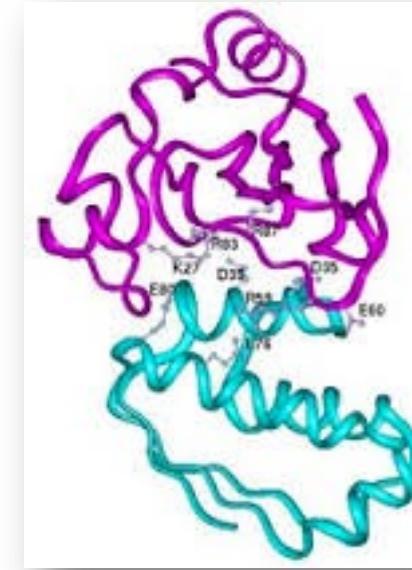
R T D C Y G N V N R I D T T G
A S C K T A K P E G L S Y C G
V S A S K K I A E R D L Q A M
D R Y K T I I I K K V G E K L C
V E P A V I A G I I S R E S H
A G K V L K N G W G D R G N G
F G L M Q V D K R S H K P Q G
T W N G E V H I T Q G T T I L
I N F I K T I Q K K F P S W T
K D Q Q L K G G I S A Y N A G
A G N V R S Y A R M D I G T T
H D D Y A N D V V A R A Q Y Y
K Q H G Y

↑
Primary Sequence 3D Structure



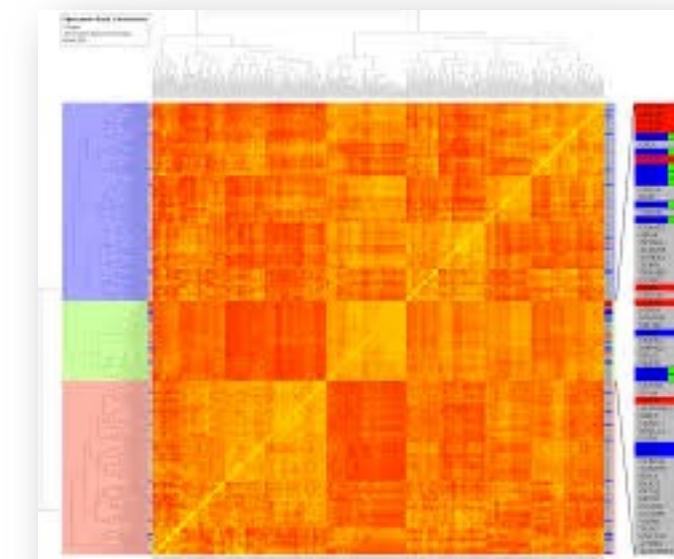
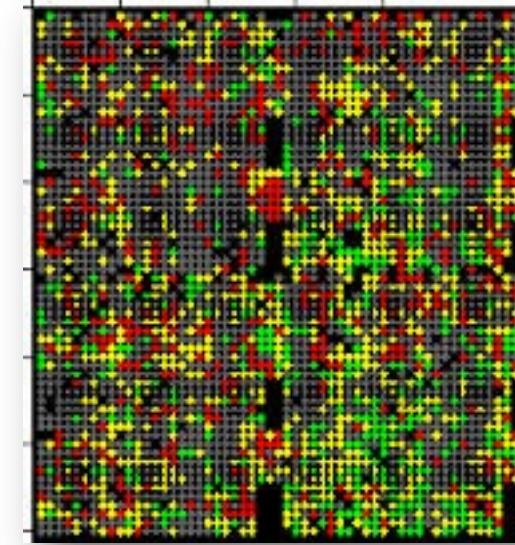
Types of Data

- Descriptions
- Pictures
- Sequences (DNA, RNA, and proteins)
- Structures (Protein and RNA)
- Interactions



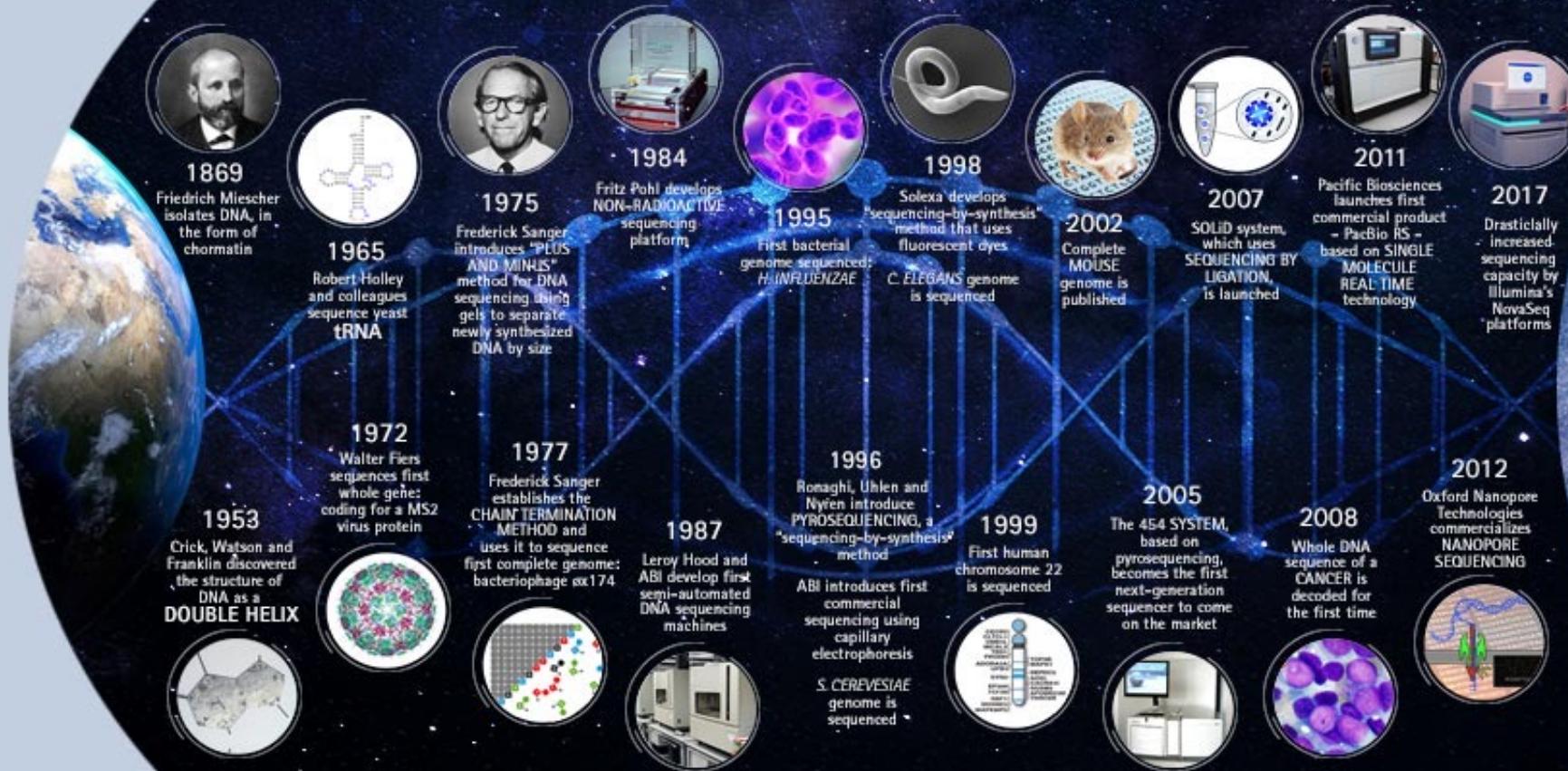
Types of Data

- Descriptions
- Pictures
- Sequences (DNA, RNA, and proteins)
- Structures (Protein and RNA)
- Interactions
- Expressions
- ...

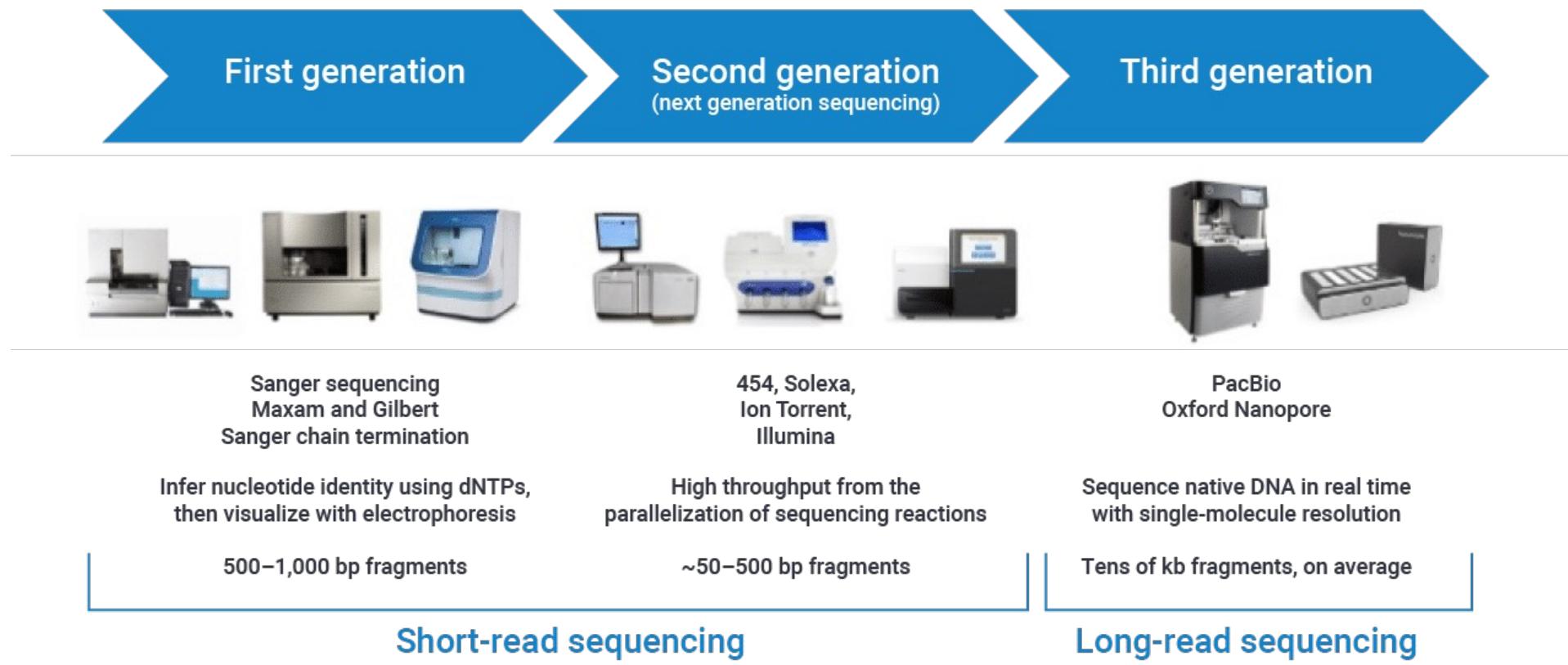


- Background of Bioinformatics
- Trends in Bioinformatics
- Fields of application of Bioinformatics

A JOURNEY THROUGH THE HISTORY OF DNA SEQUENCING



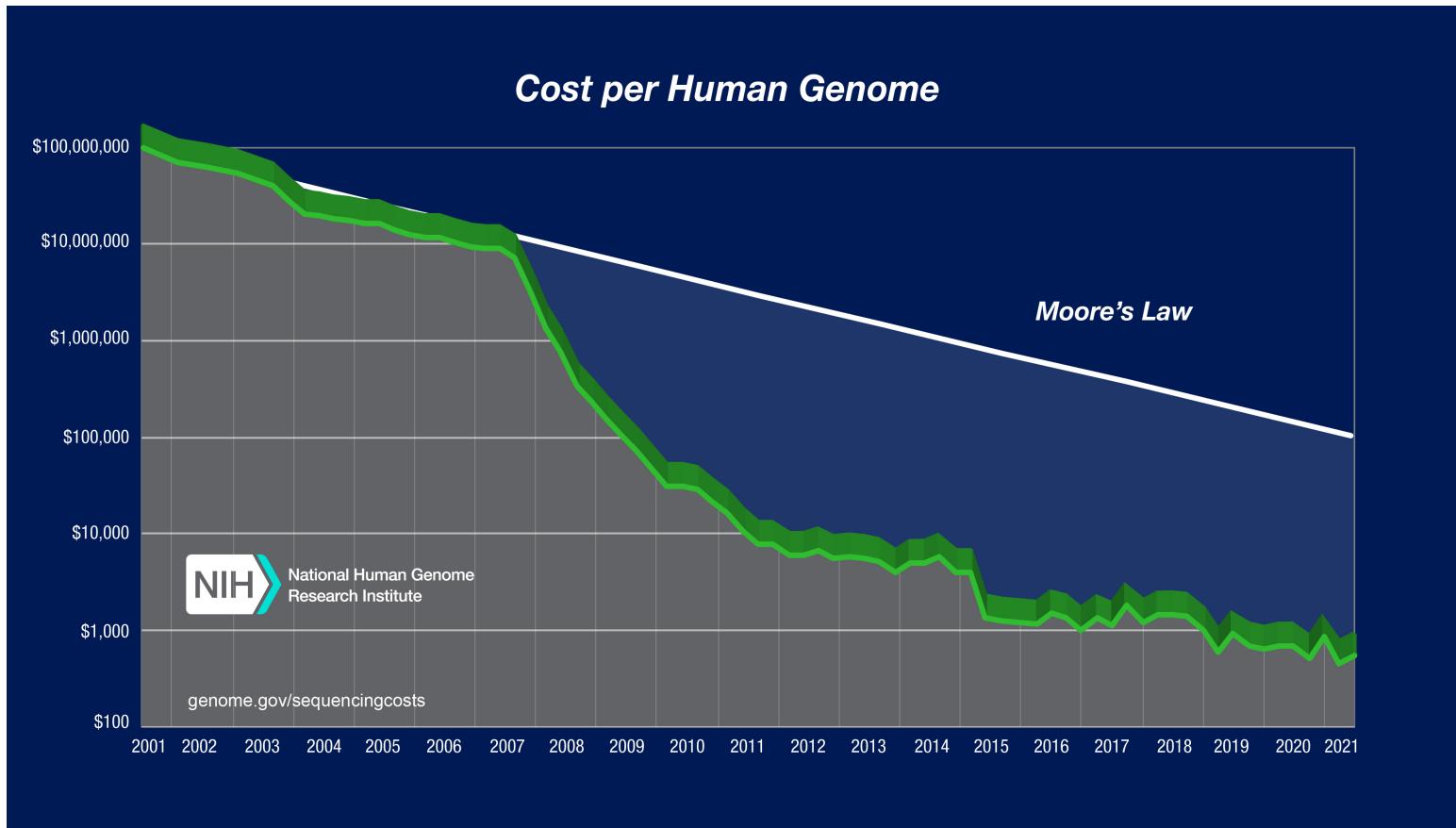
Technological advances



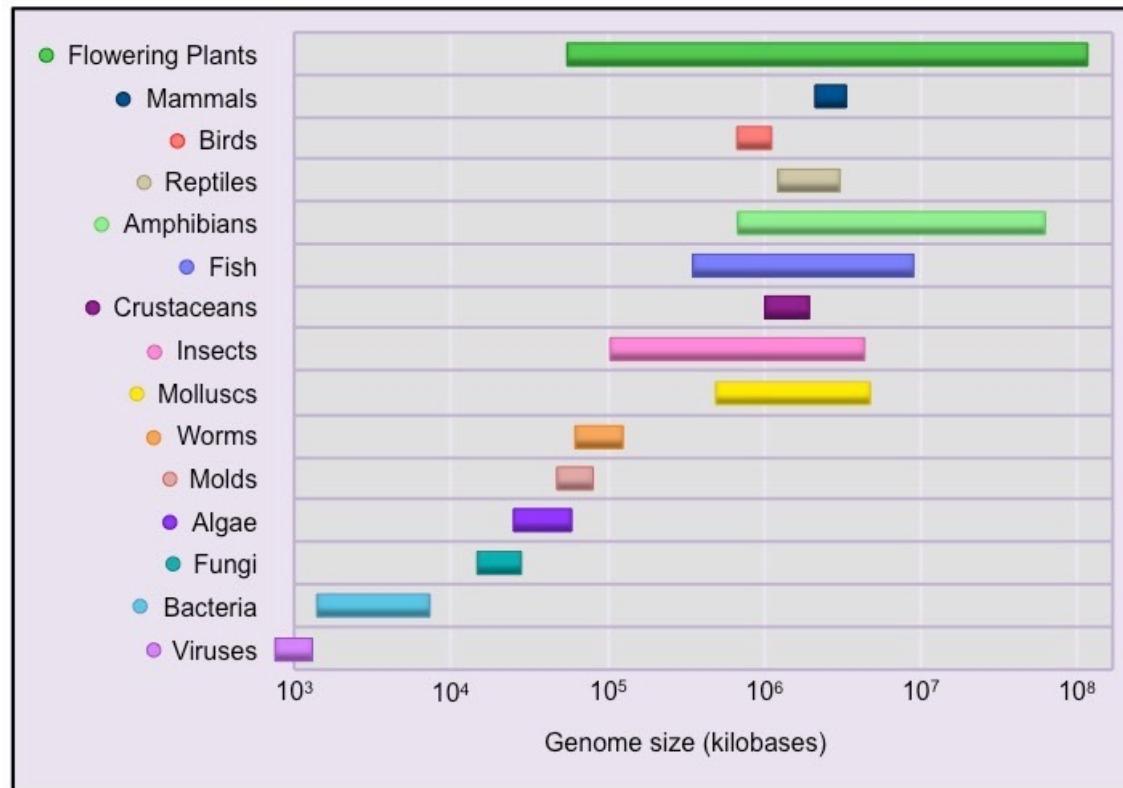
Fastest fingers quiz

- Which year did the Human Genome Project begin?
- How long did the project take?
- What was the estimated cost of the project?
- What is the size of the human genome?

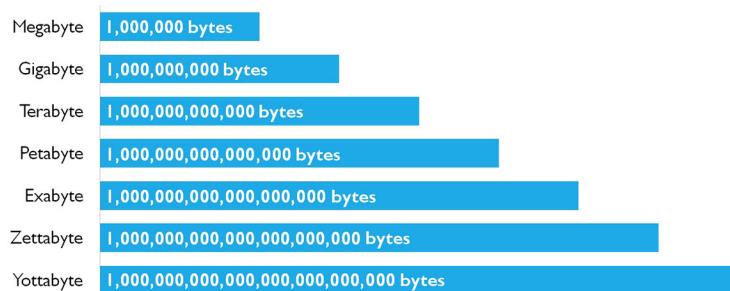
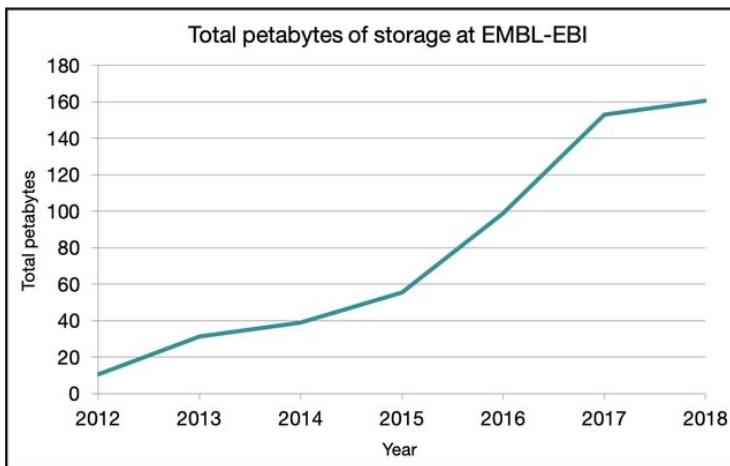
Technological advances



Genomes (entire set of instructions)



Big Data!



HOW BIG ARE THEY?

Sequence without knowledge connected to it, is meaningless!

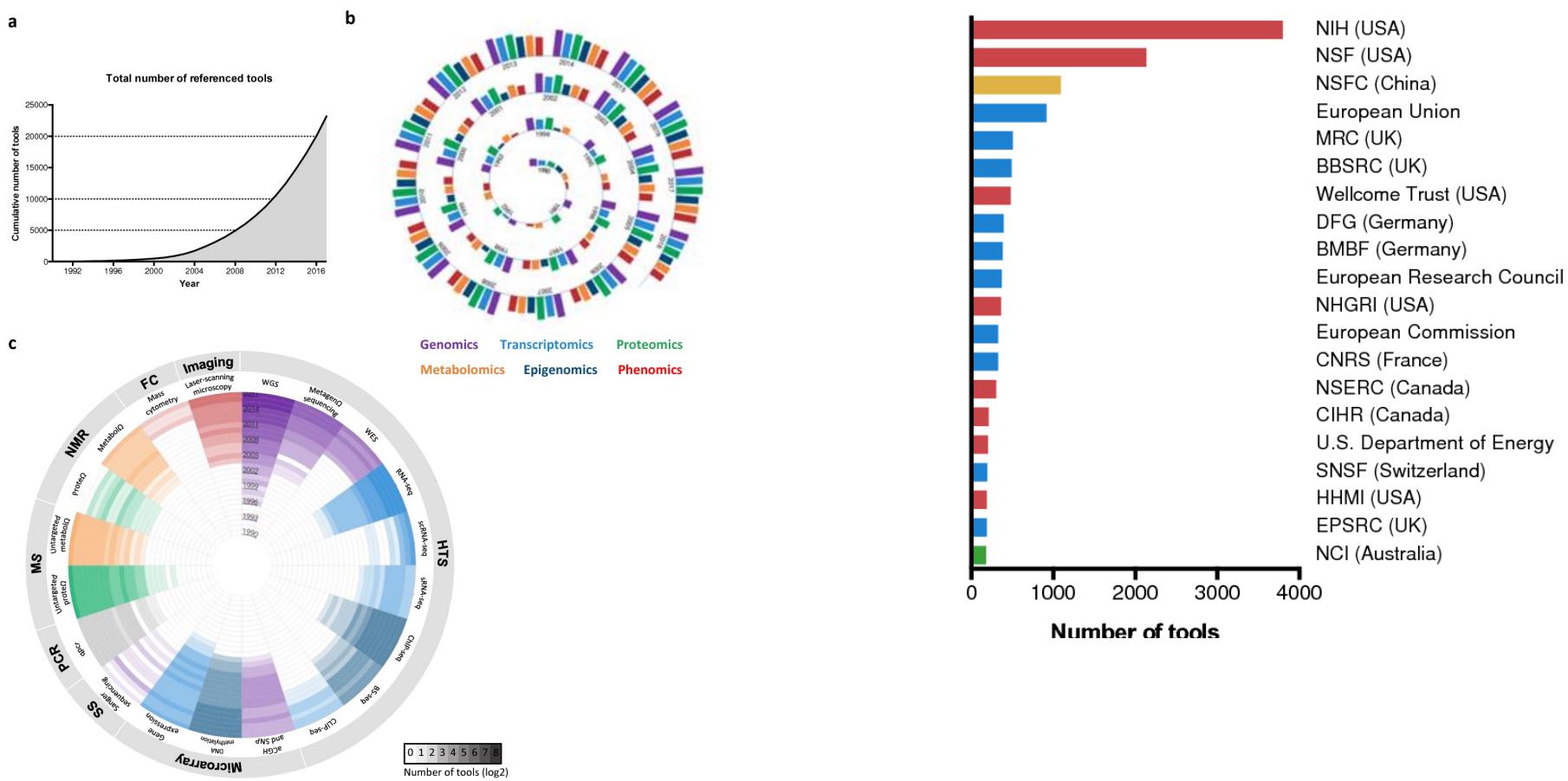
What to do?

- Sequence similarity
- Finding genes and regulatory elements
- Functional analysis of genes
- Homology
- Polymorphism
-

BIOINFORMATICS

And That's where Data Science Tools are Needed

Tool and technology development

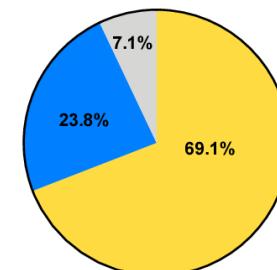


<https://arxiv.org/pdf/1807.06808.pdf>

Tool and technology development

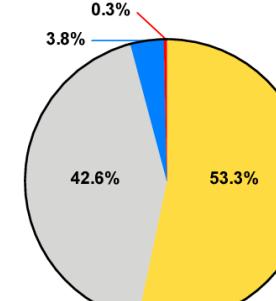
a

Tools specifications



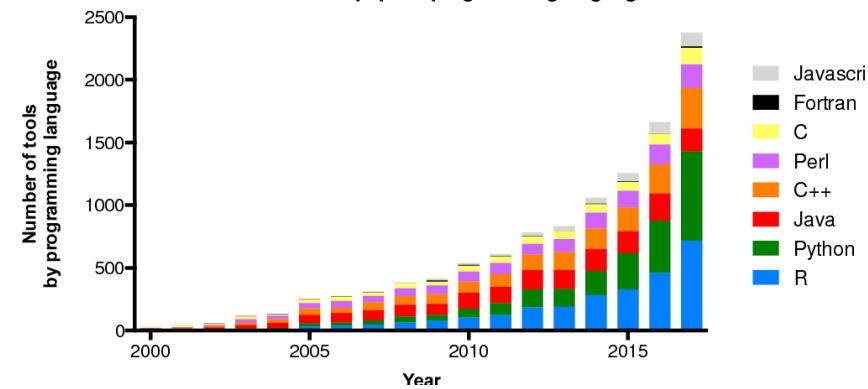
b

Tools operating systems



c

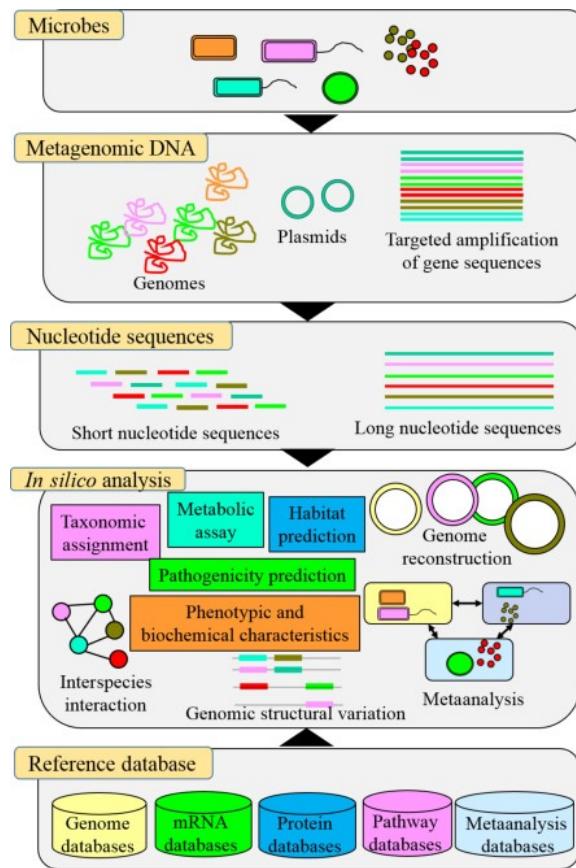
Evolution of some popular programming languages



<https://arxiv.org/pdf/1807.06808.pdf>

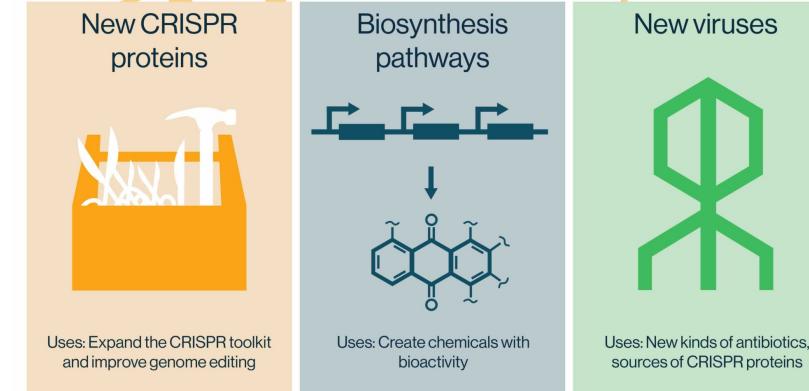
- Background of Bioinformatics
- Trends in Bioinformatics
- Fields of application of Bioinformatics

Metagenomics

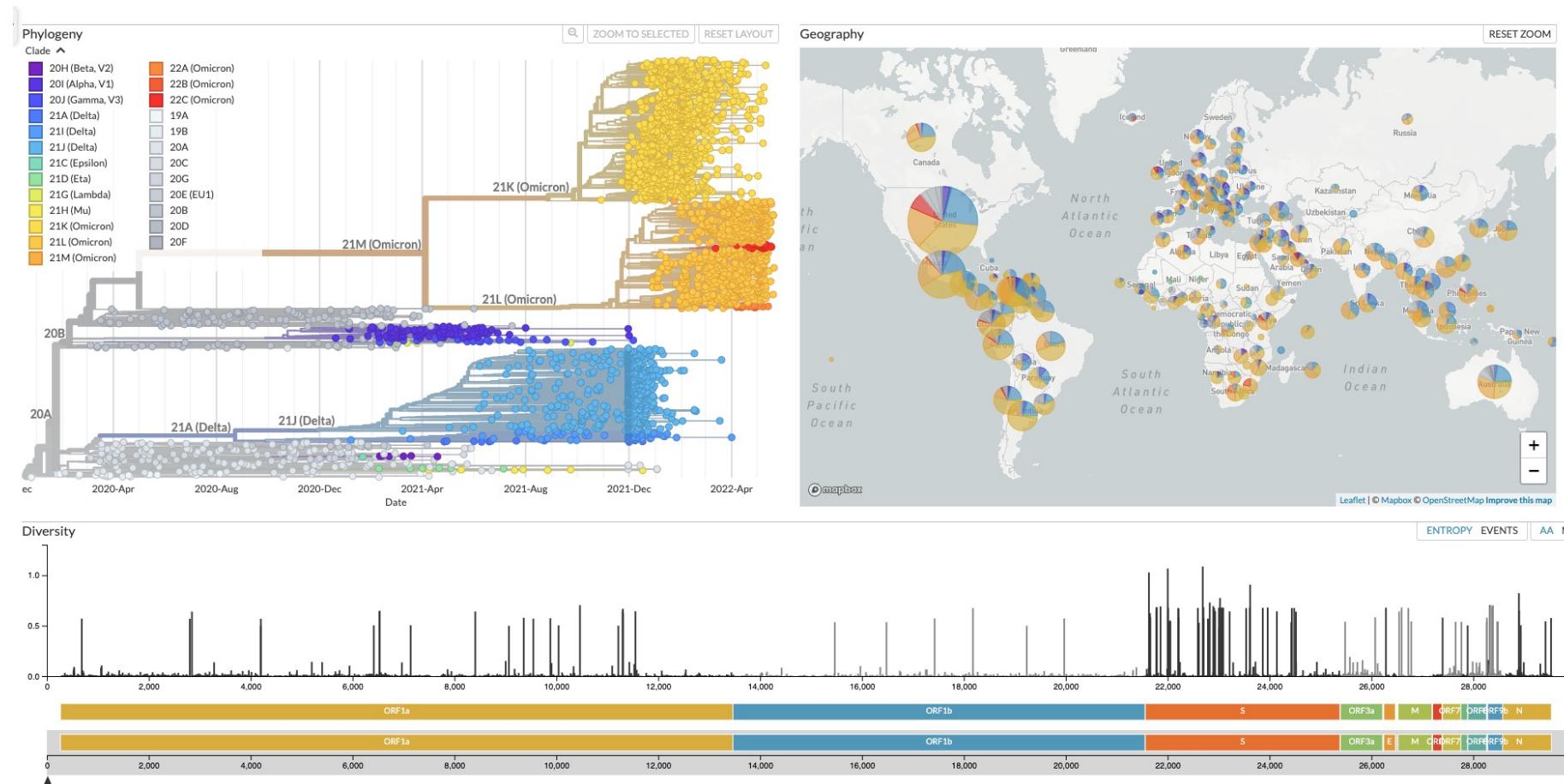


Discoveries from the field of metagenomics

Only a small fraction of the world's microorganisms have been cultured. Yet, we do have large databases of DNA sequences from a wide array of organisms. In "metagenomics," researchers use computational tools to search through these databases and find candidate DNA sequences that encode new kinds of proteins, groups of proteins, or even new organisms. Using metagenomics, we can find many new tools with biotechnological applications.

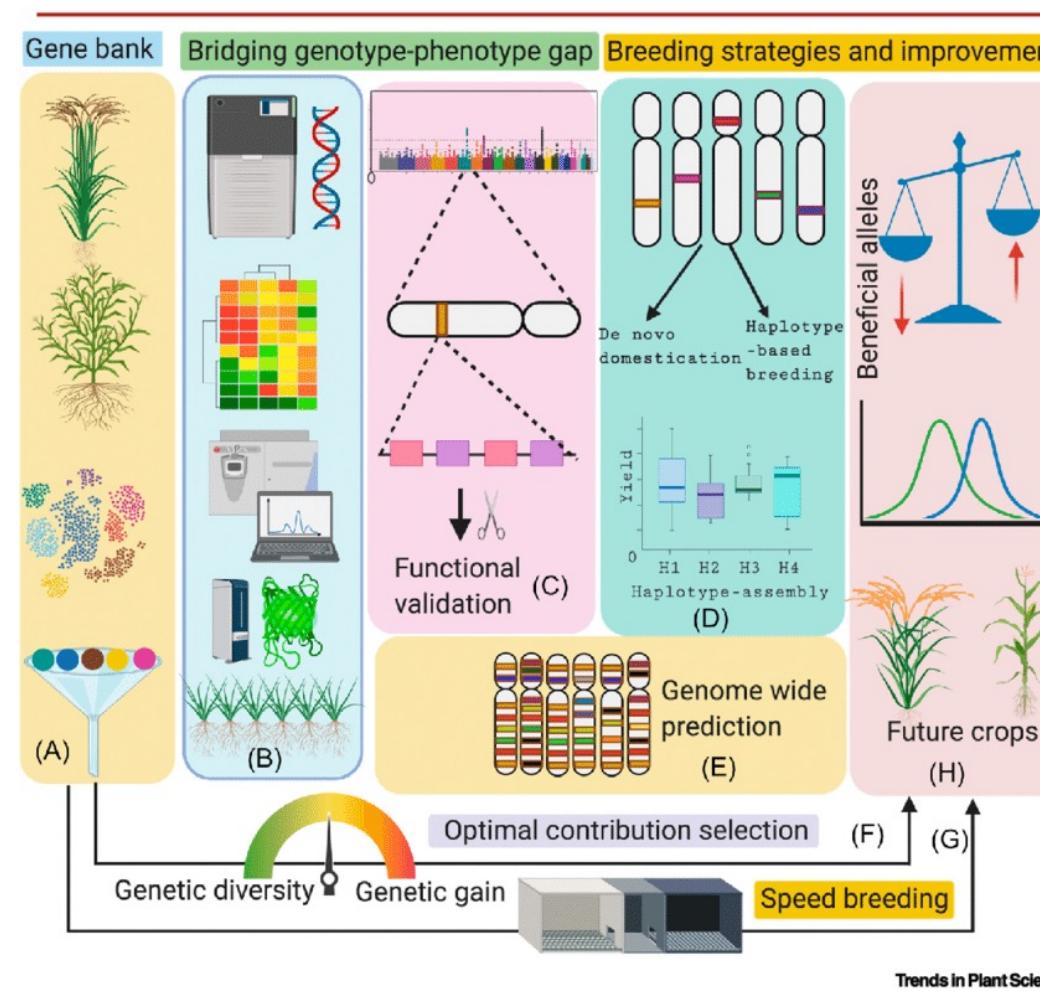


SARS-CoV2 (Pandemic) – Epidemiology, Vaccine, Drug Design



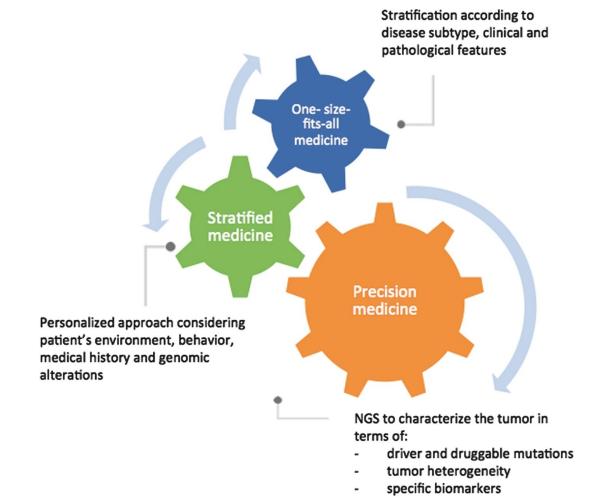
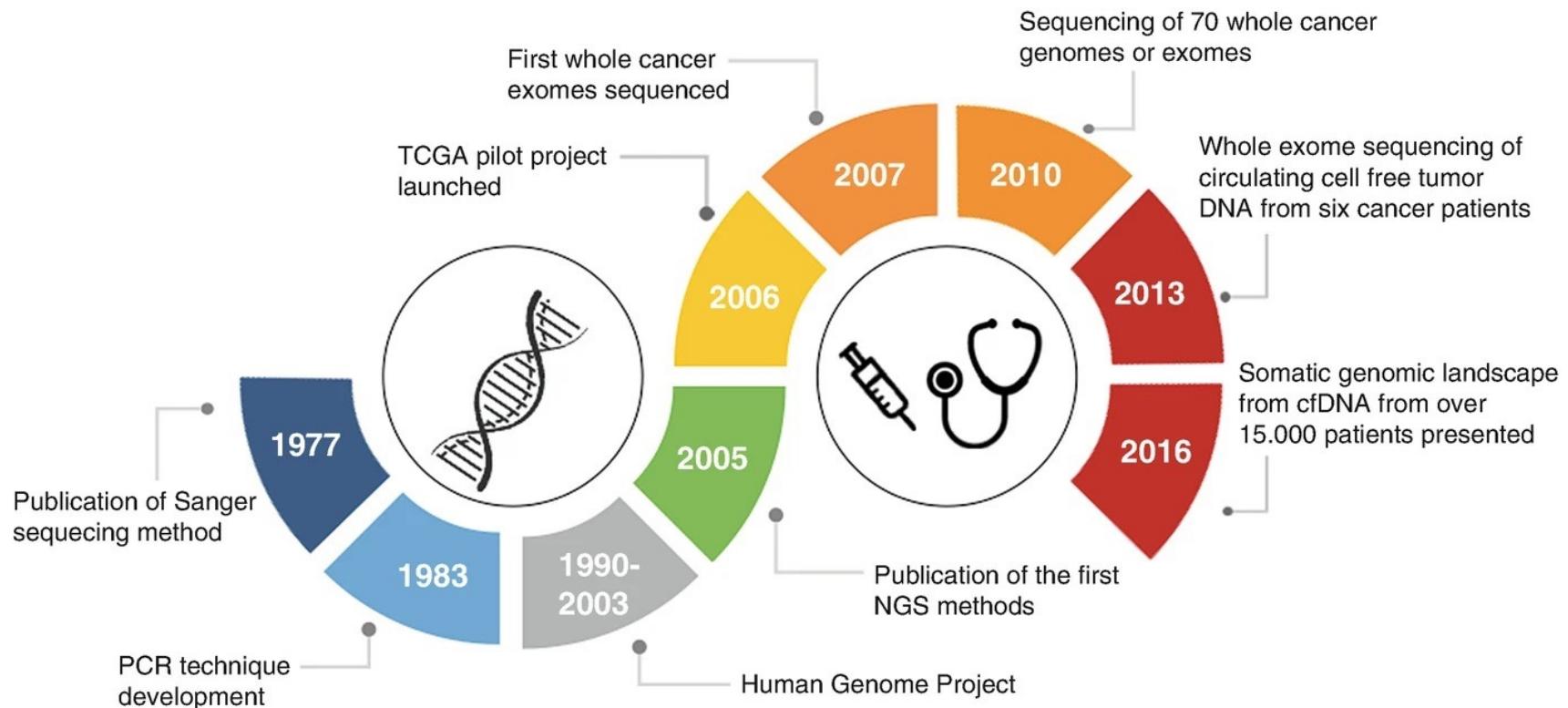
<https://nextstrain.org/ncov/gisaid/global/6m>

Genomic Assisted Breeding



<https://doi.org/10.1016/j.tplants.2021.03.010>

Precision (personalised) Medicine



Personalised



Find out what your DNA says about you and your family.

- See how your DNA breaks out across 2000+ regions worldwide
- Discover DNA relatives from around the world
- Share reports with family and friends
- Learn how your DNA influences your facial features, taste, smell and other traits

[order now](#)

USD\$99

AncestryDNA®



Discover where your family is from without even leaving your living room.

Only \$99

[Buy now](#)

AncestryDNA® + Family Tree Package



Uncover your whole family story with an AncestryDNA® kit and a 3-month World Explorer Membership.

Only \$100 *

[Buy now](#)

AncestryDNA® Traits + All Access Package



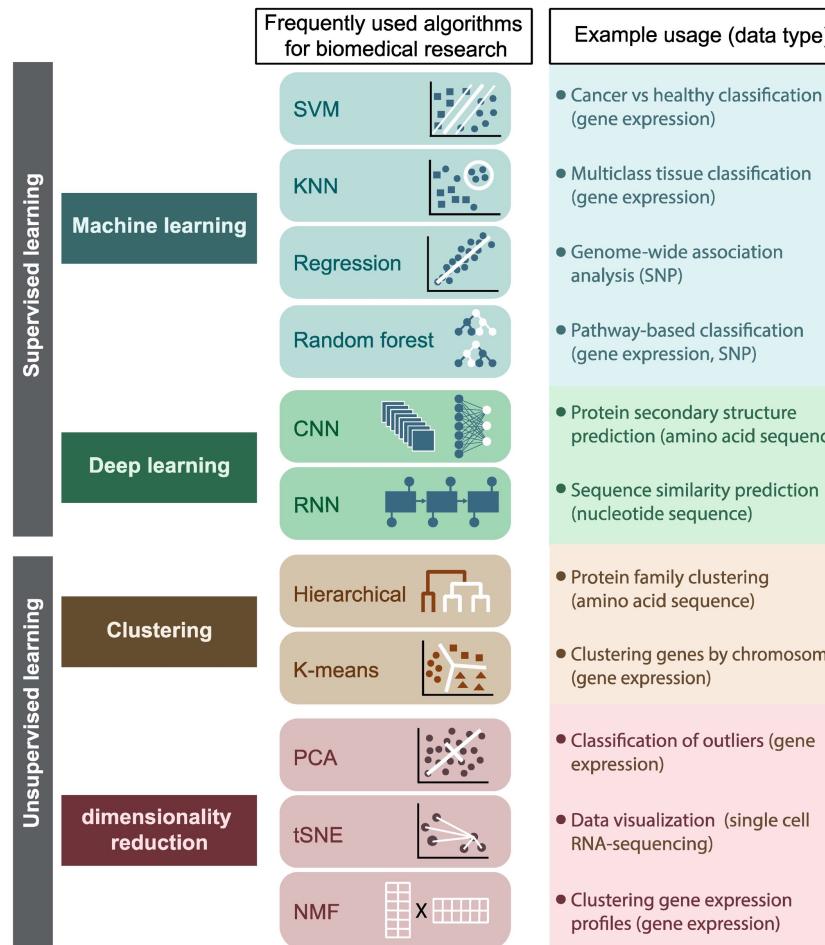
Get all that AncestryDNA® offers, plus personal traits and a 3-month All Access Membership.

Only \$199 *

[Buy now](#)

<https://www.nytimes.com/wirecutter/reviews/best-dna-test/>

Machine learning, AI and Bioinformatics – “Data Science”



<https://www.mdpi.com/1422-0067/22/6/2903>

Conclusion

Store – in the right format, stored properly, compressed



Infrastructure – Computing infrastructure (CPU, RAM, Storage)



Tools – Programs that process your data (shared – open source)



Reference data – Databases for existing data

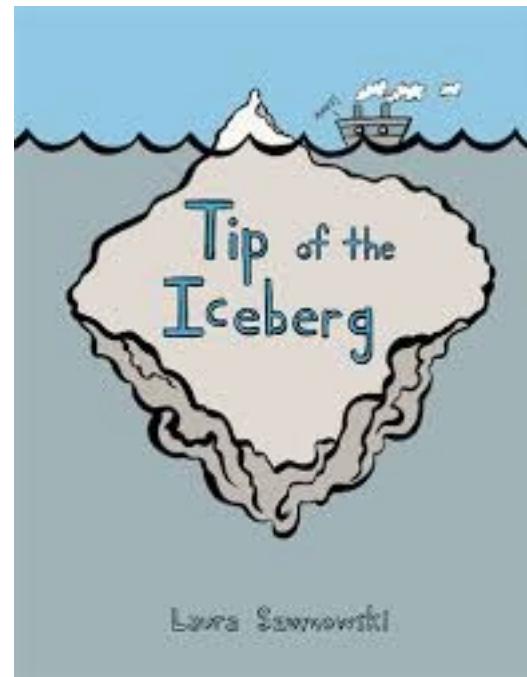


Visualise



INTERNET – connection to external Databases

and this is only the



... enjoy bioinformatics!!