SevenBridges

Applied Bioinformatics

Agenda

- Introductions
- Course overview
- Bioinformatics intro
- Platform registration

Applied Bioinformatics

Introductions: Lecturers

Different backgrounds - bioinformatics engineers

- Sanja Mijalković, MATF
- Marko Matić, Faculty of Physics
- Boris Majić, ETF
- Dajana Panovic, MATF (Statistics)
- Milan Kovačević, PMF

Applied Bioinformatics

Course overview

Course logistics (1/2)

- 2 classes each week
- A mixture of lectures and hands-on exercises
- Exercises will be done in IPython notebooks on the CGC platform
    - We will provide help with the Python syntax if needed
- The course is not covered by a single textbook
- Lessons are mostly linked

Course logistics (2/2)

- Attending classes is not mandatory, but highly recommended
- Practice test in the end of semester
- For all course related questions contact
  - milan.kovacevic@sbgenomics.com
  - dajana.panovic@sbgenomics.com

Course topics

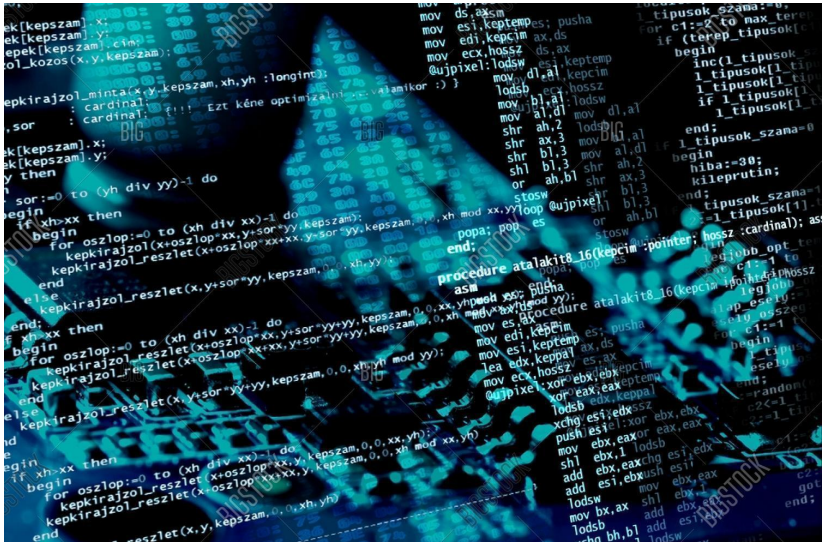
- Introduction to biological background and sequencing (2 weeks)
- DNA analysis (4 weeks)
- RNA analysis (2 weeks)
- Structural variation detection (1 week)
- Cancer genomics (1 week)
- Methylation analysis (1 week)
- Test practise (1 week)

Applied Bioinformatics

Bioinformatics intro

What is bioinformatics?

bio + informatics



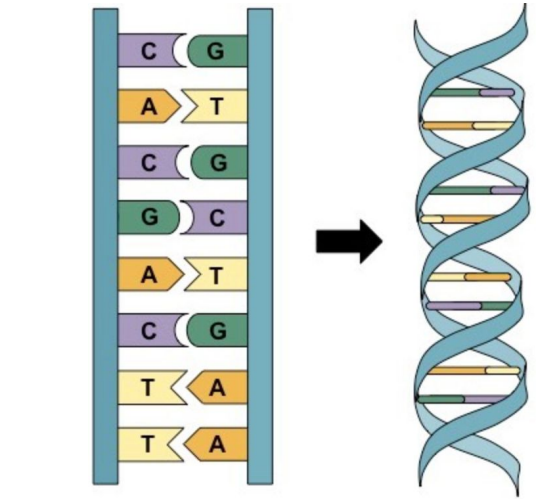
Bioinformatics intro

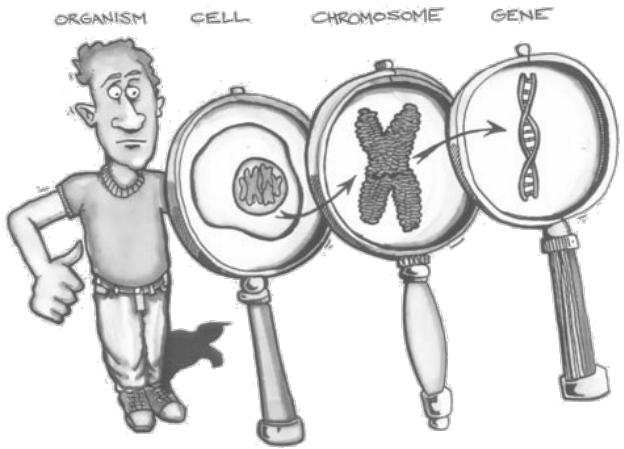
During this course you will learn/hear:

- More about
  - Human cells
  - DNA
  - RNA
- Why is the DNA so important?
- What can we find out from the DNA?
- More about certain diseases, like cancer…
- Genome digitalisation
- How to analyse genomic data
- Interesting algorithms
- Optimisations for huge data

Bioinformatics intro

- A, T, C, G
- 3.000.000.000 letters in a DNA molecule
- 46 chromosomes
- Almost everything in the body is predefined by the order of those letters

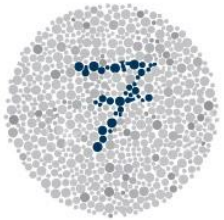




Bioinformatics intro

Some interesting characteristics:

LONGER EYELASHES



DALTONISM

LESS SLEEPING

SUPER STRENGTH

And some not that interesting things:

- Cancer
- Rare diseases
- Autoimmune diseases
- Neurological diseases

DNA digital representation

- DNA is stored as a string
- Human genome project and reference genome
- 3 billion As, Ts, Cs and Gs translates into ~ 3 GB of data
  - font 4

Bioinformatics intro

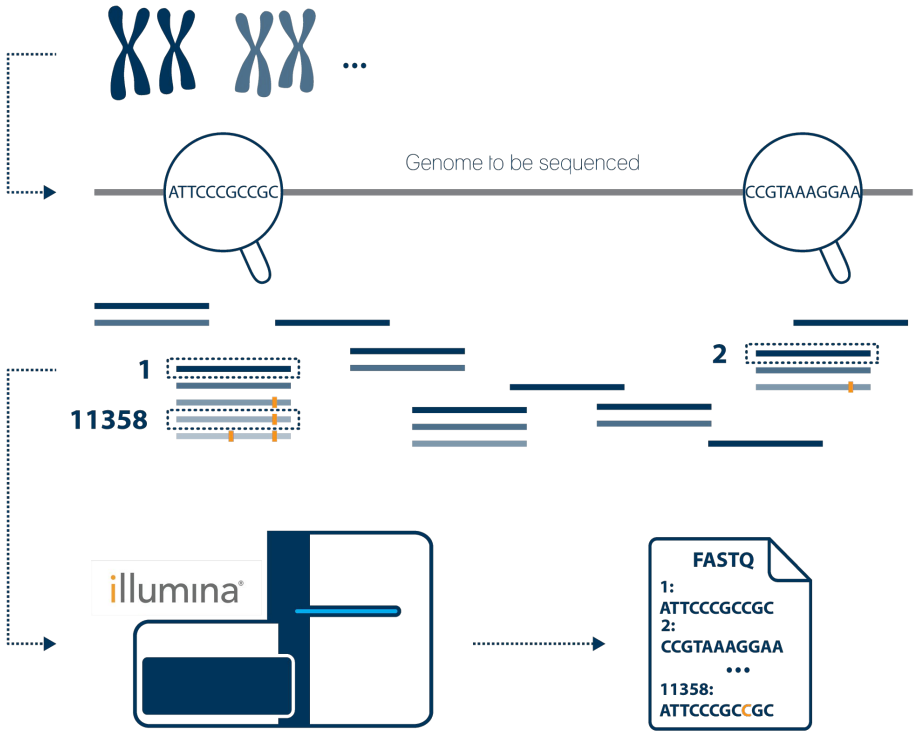
BiX starts with raw data

Typical sample (WGS):

- 200-500 millions of reads
- Each read is 50-150 chars (A, C, G, T)
- + Assessed qualities of sequencing
- 30x coverage ~ 150-300GB

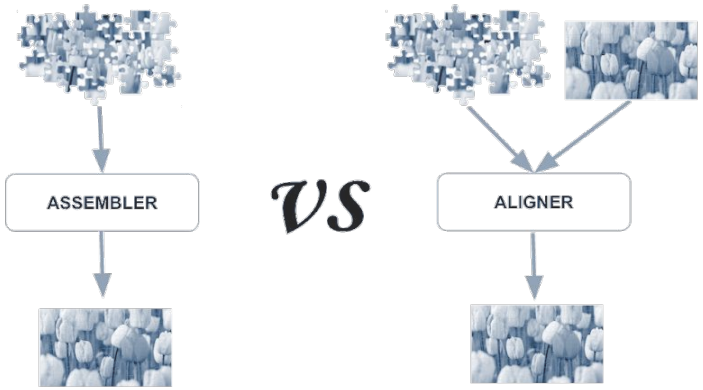
Various data:

- Digitized DNA, RNA, proteins
- Different techniques (lab prep)
- Comes from institutes, pharma companies, ...

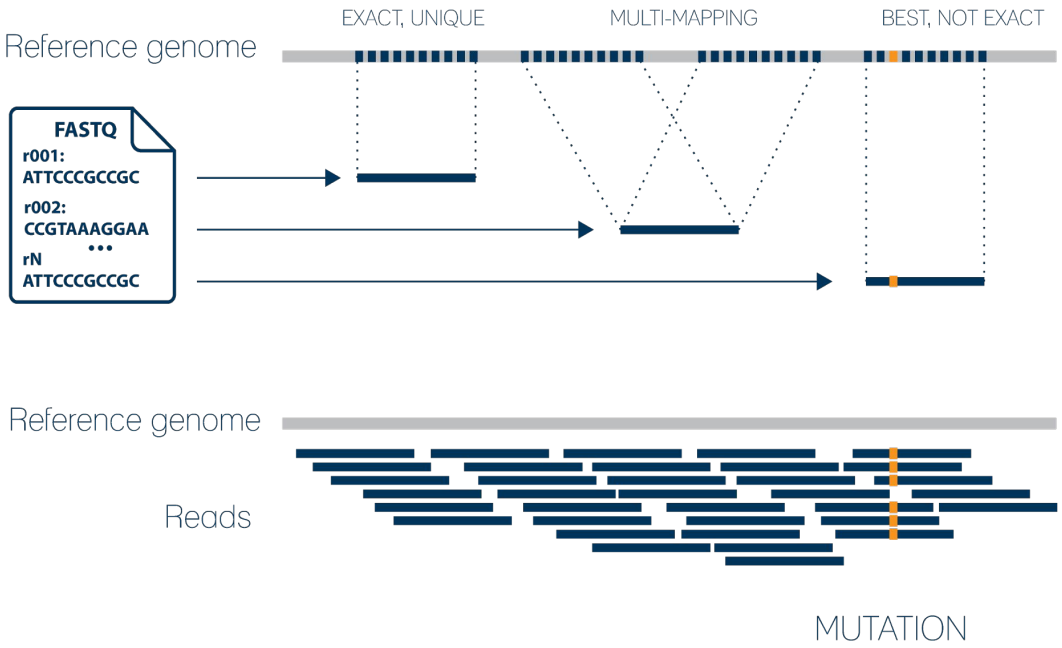


Bioinformatics intro

Common 1st step:
Reconstruct personal genome



$$\Theta(n^2) \text{ vs } \Theta(nm)$$



Bioinformatics intro

Single nucleotide variant

Copy number variant

Deletion

Translocation

Intra-chromosomal

Inter-chromosomal

Insertion

Whole genome duplication

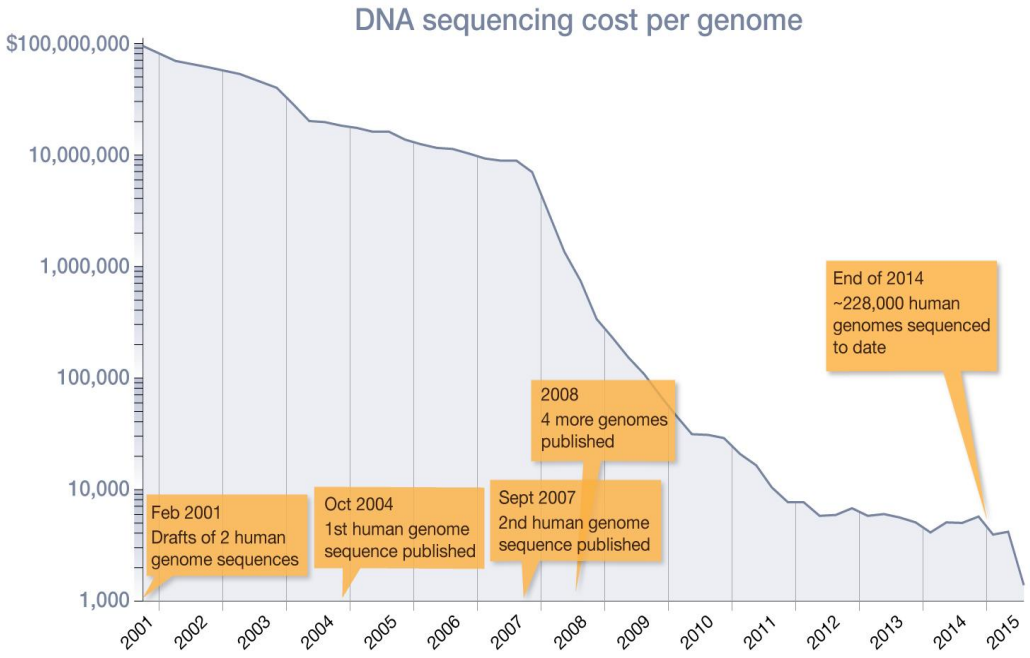
Inversion

Duplication

Tandem

Interspersed

Necessity of Cloud in Bioinformatics



- Python
- R
- Docker
- Amazon

My Full DNA: Whole Genome Sequencing with mtDNA

€259.00 EUR ~~€850.00 EUR~~ You save €591.00 EUR

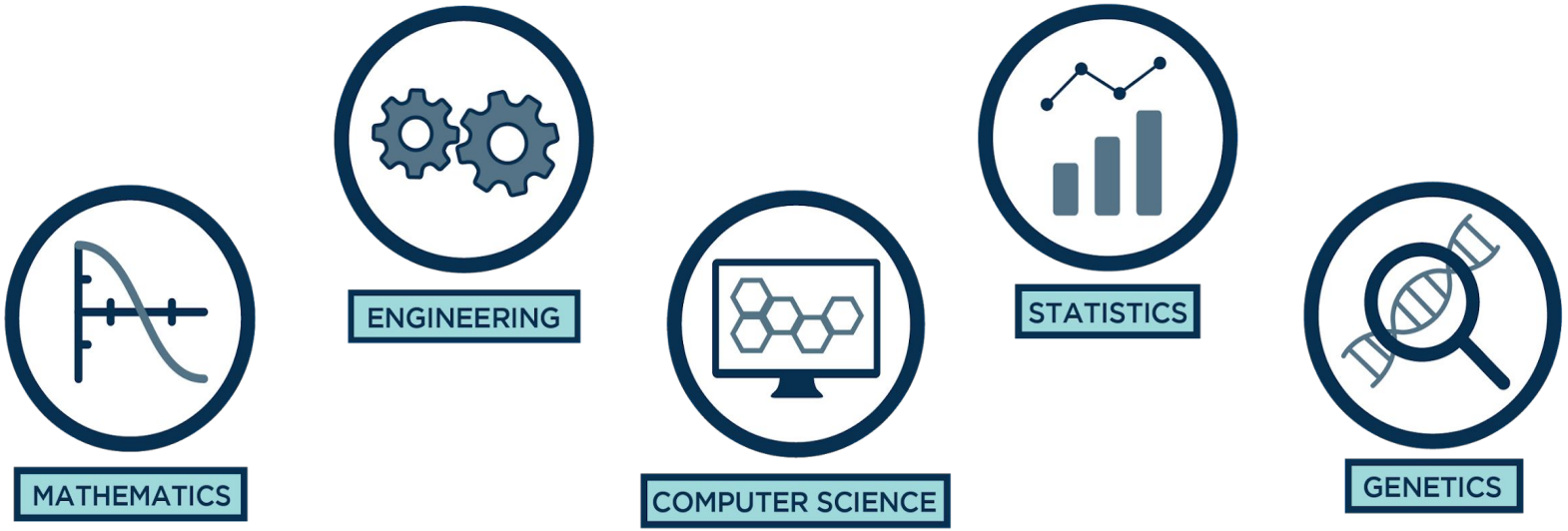
Type of report: My Full DNA... 1

ADD TO CART

DESCRIPTION

My Full DNA by Dante Labs gives you unparalleled access to your health, your family traits, and predispositions. Having your full genome sequenced means:

Bioinformatics engineer



Applied Bioinformatics

Platform registration

CGC registration

- Exercises are going to be done on CGC platform
- **cgc.sbgenomics.com**
- CGC = Cancer Genomics Cloud
  - Funded by NCI - National Cancer Institute (NIH)
  - Powered by Seven Bridges
  - For academic use
  - Many researchers/ institutes / labs are using it for their analysis
- More information about the project available here.
- Registration

Thank you!