MBI 3100: Introduction to medical bioinformatics
Lecture # 1
Course Syllabus and Introduction to Bioinformatics

Amjad Khan
https://mathbioinfo.github.io/amjadkhan/

Department of Pathology and Laboratory Medicine
Schulich School of Medicine & Dentistry
Western University

September 2023

Syllabus and introduction

We acknowledge the Anishinaabek (Ah-nish-in-a-bek), Haudenosaunee (Ho-den-no-show-nee), Lūnaapéewak (Len-ahpay-wuk) and Attawandaron (Add-a-won-da-run) peoples, whose traditional lands we are gathered upon today.

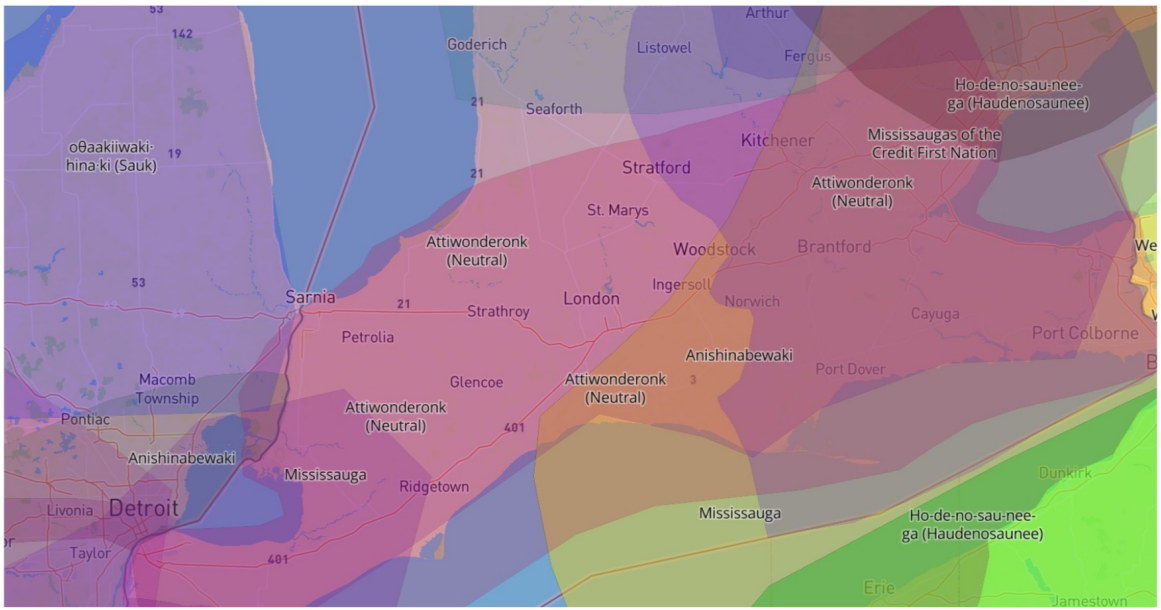


Image source: Native Land Digital, https://native-land.ca/.

History of Pandemics

The Worst Pandemic in History?

- Covid-19
- HIV/AIDS
- Smallpox
- Black Death (Bubonic Plague)

History of Pandemics

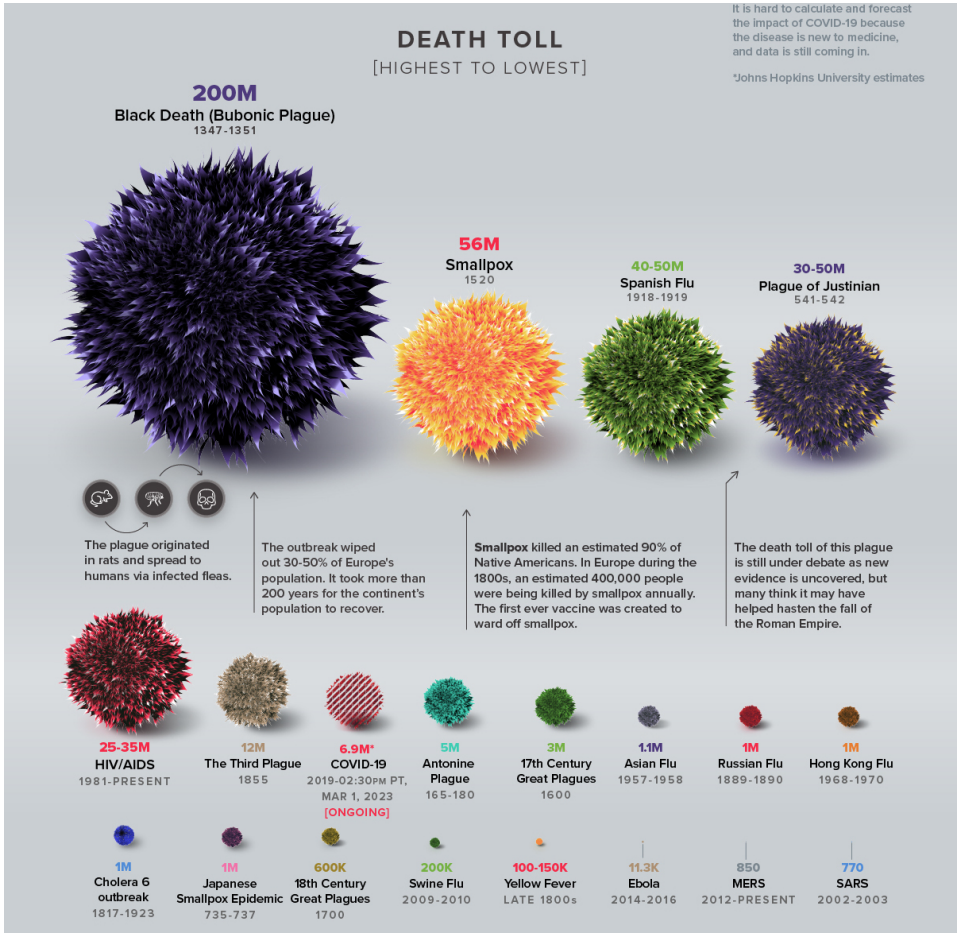


Image source:https://www.visualcapitalist.com/history-of-pandemics-deadliest/.

Yersinia pestis

Plague is a disease that affects humans and other mammals. It is caused by the bacterium, Yersinia pestis.

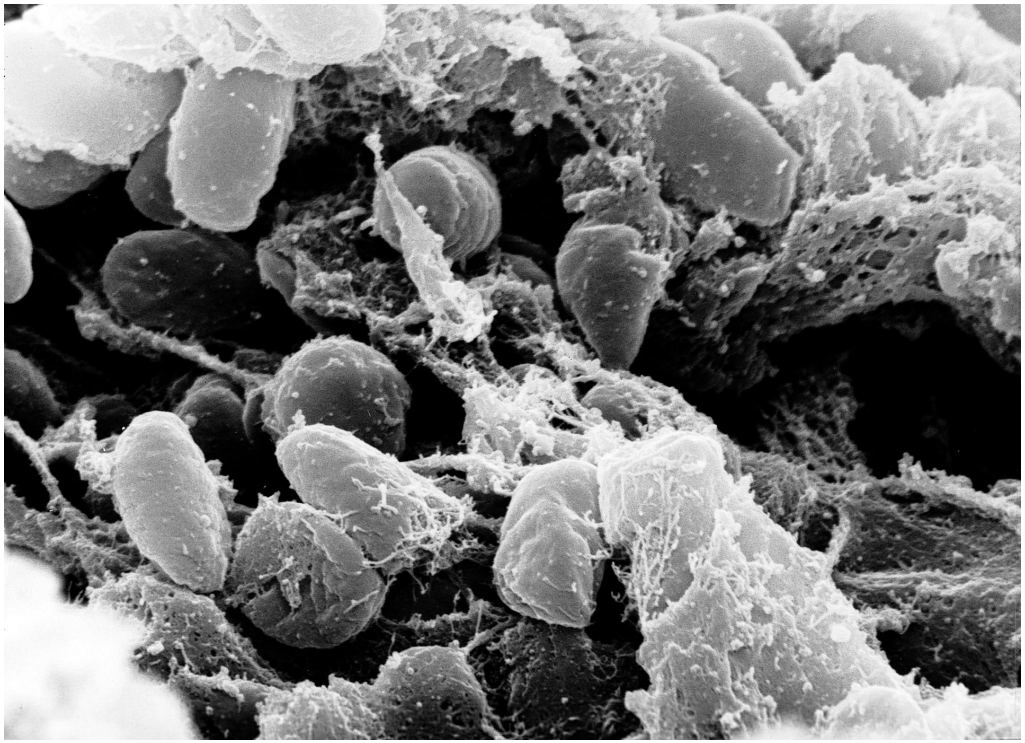


Image source:https://en.wikipedia.org/wiki/Yersinia_pestis.

Questions about Yersinia pestis

Where did Yersinia pestis come from?

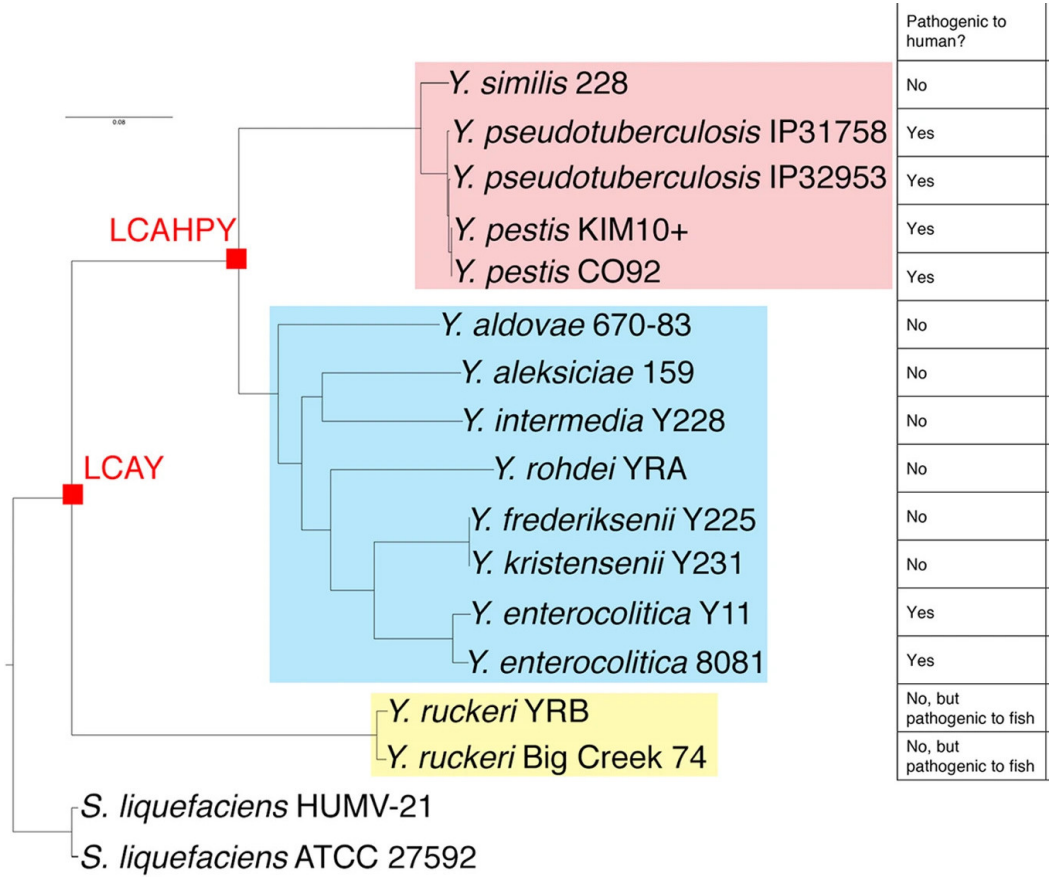


Image source:https://www.nature.com/articles/srep36116.

Questions about Yersinia pestis

How old is Yersinia pestis?

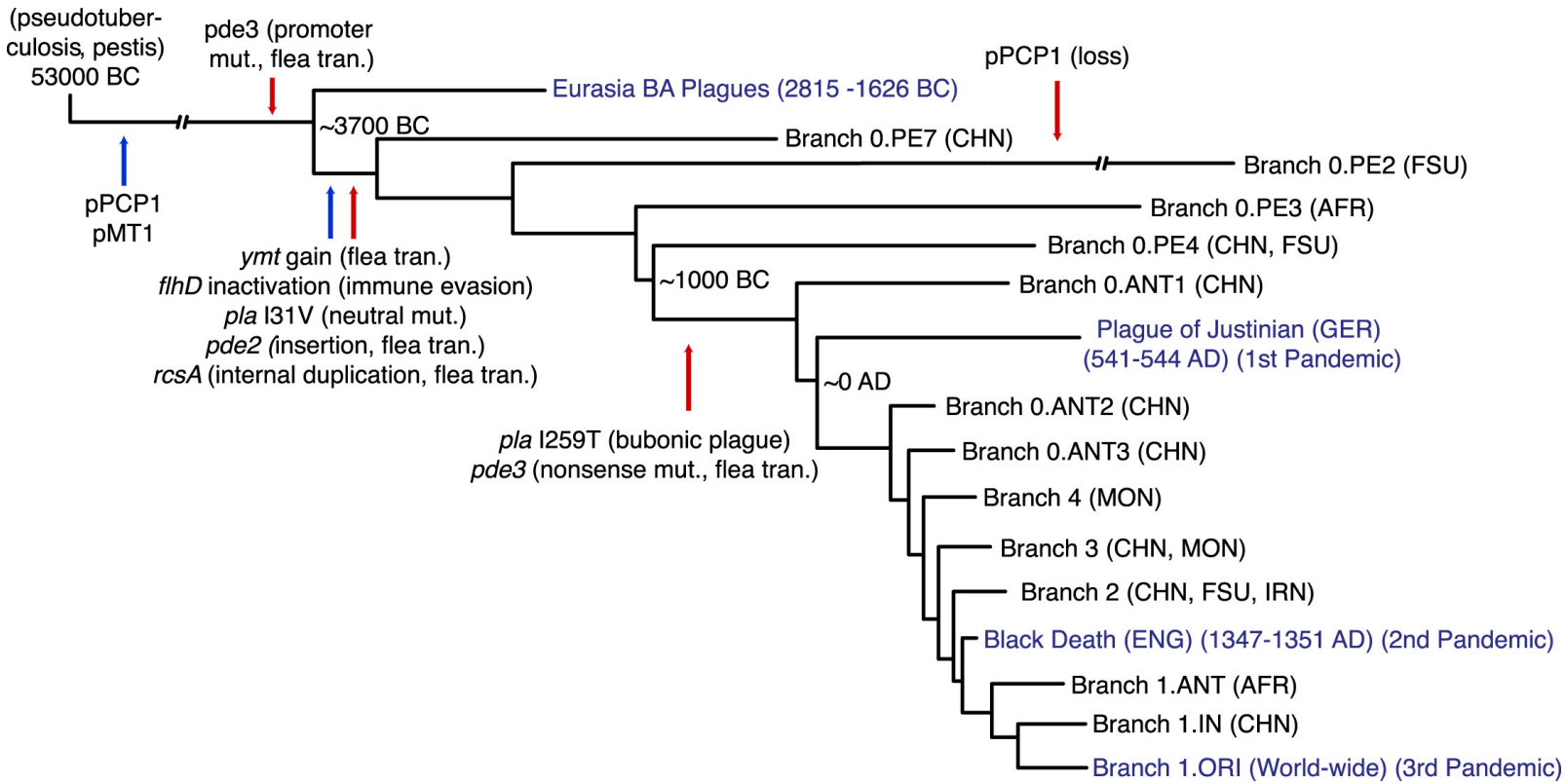
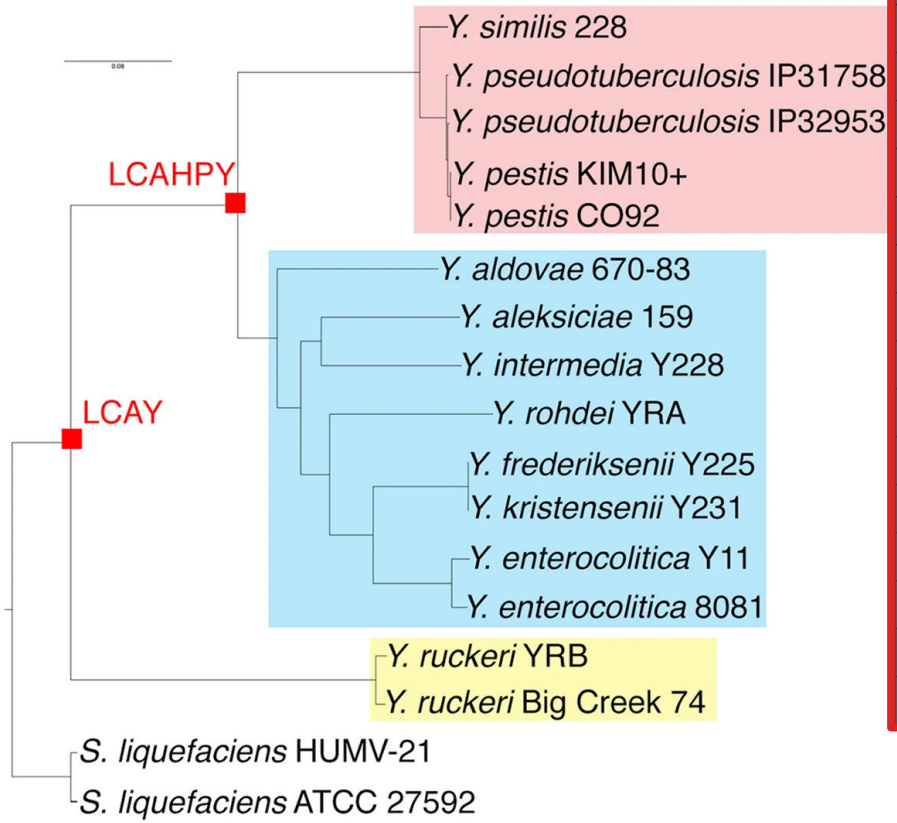


Image source:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4644222/.

Questions about Yersinia pestis

What makes Yersinia pestis virulent?



Pathogenic to human?	Presence of pYV plasmid?	Presence of CRISPR spacer to become immune to pYV or pYE854 plasmid?	Presence of *inv* homolog and N-terminal?	Number of copy of *ail* homolog?
No	No	Yes	Yes	4
Yes	Yes	No	Yes	4
Yes	Yes	Immune to pYV of *Y. enterocolitica*	Yes	4
Yes	Yes	Immune to pYV of *Y. enterocolitica*	No	4
Yes	Yes	Immune to pYV of *Y. enterocolitica*	No	4
No	No	No	No	1
No	No	No	No	1
No	No	No	No	1
No	No	No	No	1
No	No	Yes	No	1
No	No	Yes	No	1
Yes	Yes	No	Yes	2
Yes	Yes	No	Yes	2
No, but pathogenic to fish	No	No	No	1
No, but pathogenic to fish	No	No	No	1

Image source:https://www.nature.com/articles/srep36116.

Questions about Yersinia pestis

How did Yersinia pestis become so deadly?

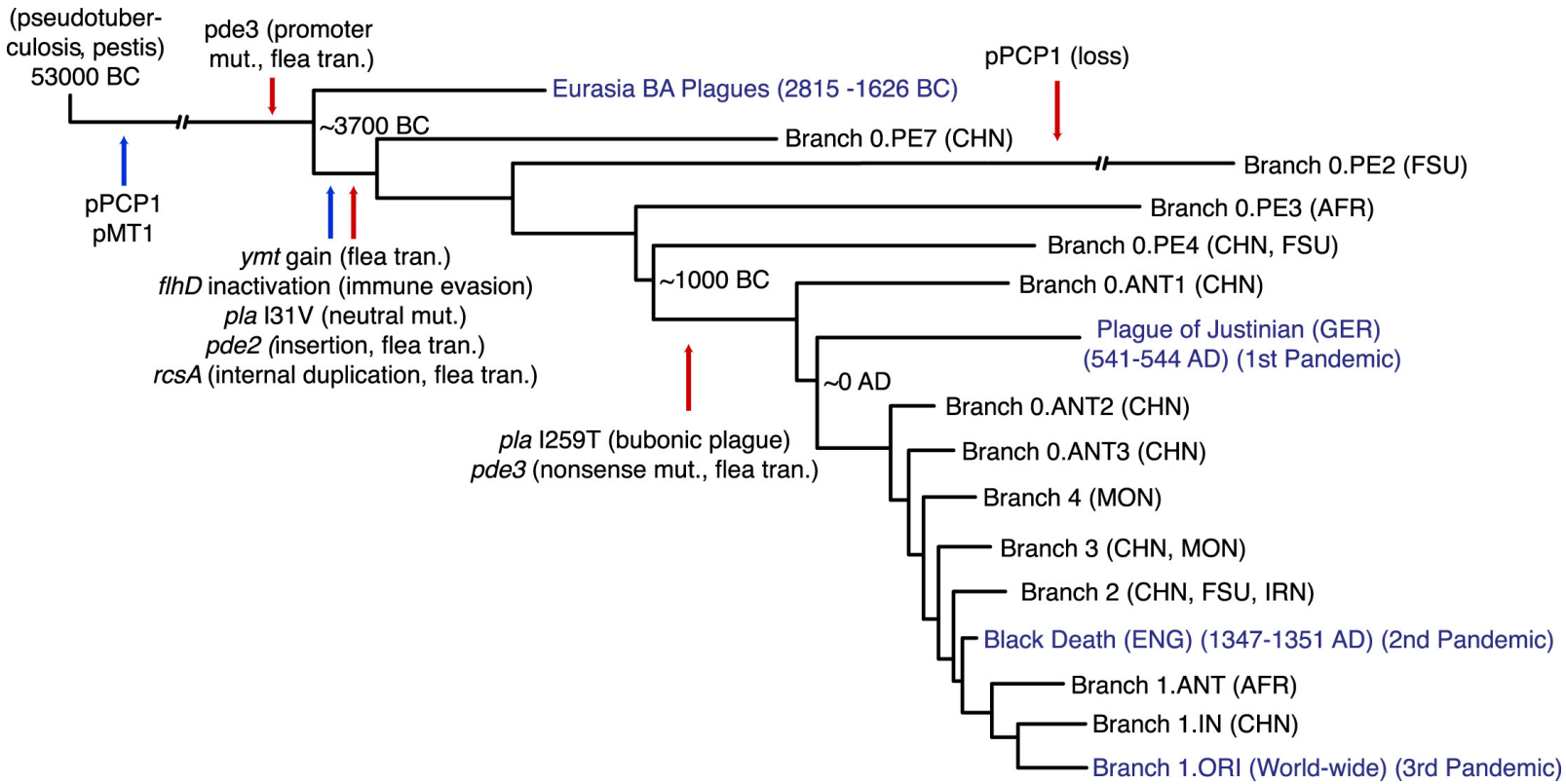
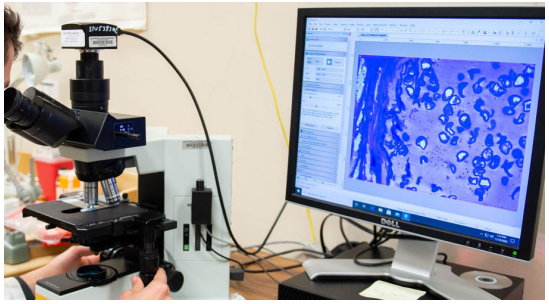


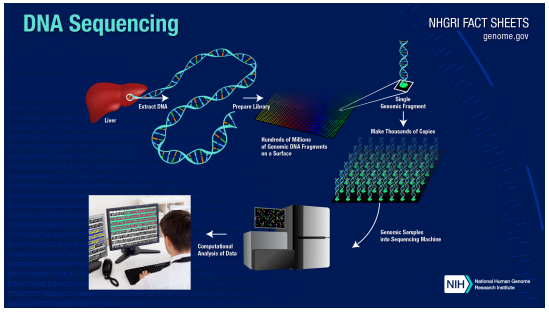
Image source:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4644222/.

Questions about Yersinia pestis

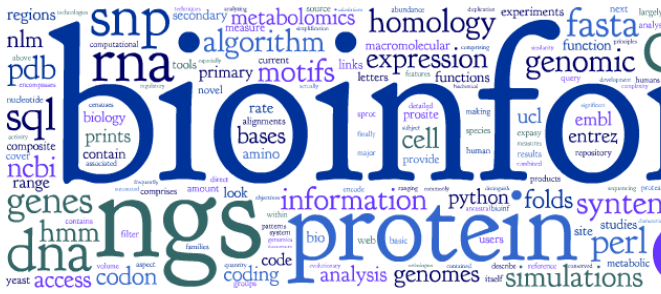
How do we know all this?



Molecular Biology



DNA Sequencing



Bioinformatics

Further reading

What is Bioinformatics?

Bioinformatics is an interdisciplinary field of science that develops
- methods and
- software tools

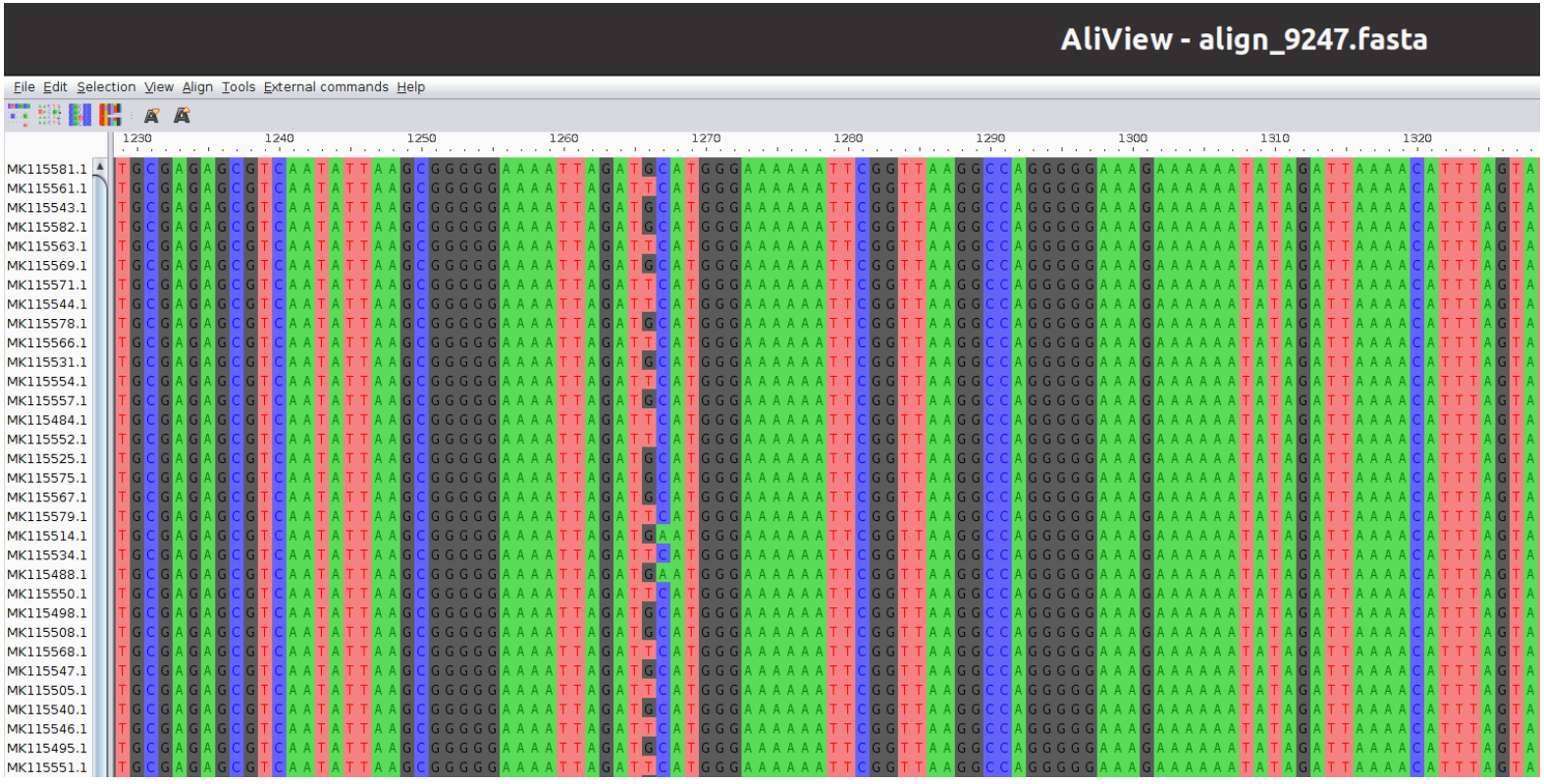
for understanding biological data, especially when the data sets are large and complex.

What is Bioinformatics?

Bioinformatics uses biology, chemistry, physics, computer science, computer programming, information engineering, mathematics and statistics to analyze and interpret biological data. (From Wikipedia).

What does a Bioinformatician do?

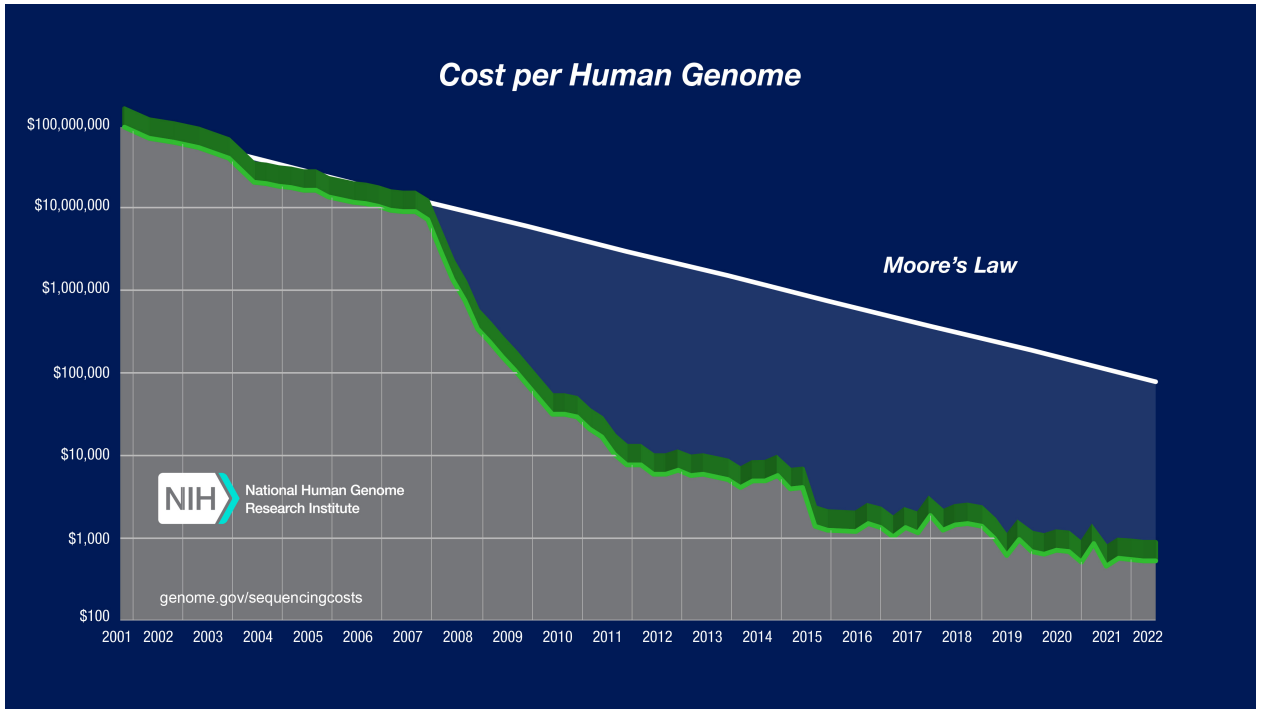
- set behind comupter and analyse biological data



?

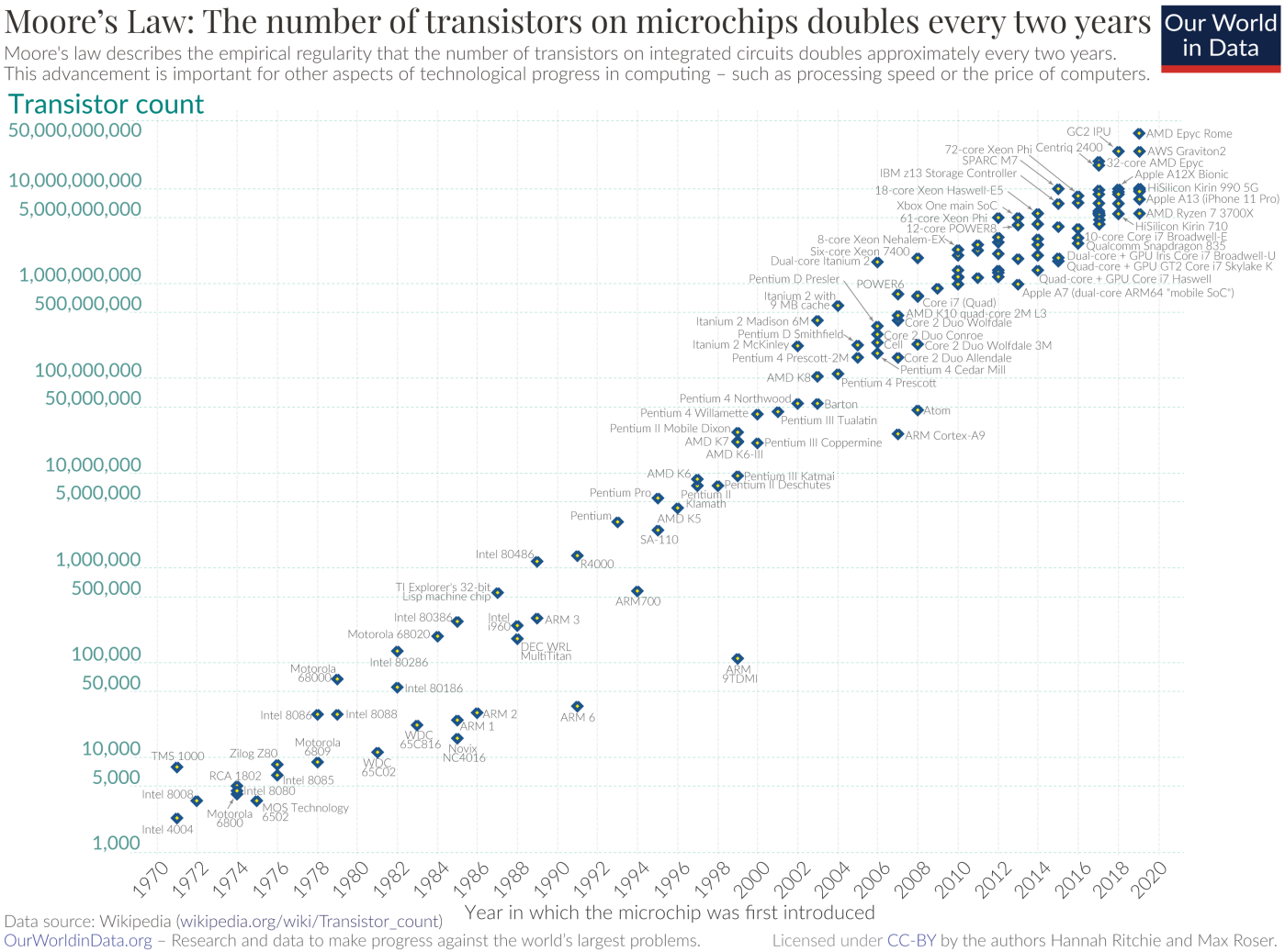
Why do we need bioinformatics?

- new technologies are coming to light
- more and more data is generated



Why do we need bioinformatics?

Schulich MEDICINE & DENTISTRY



Course materials

- My slides are available at OWL

Course syllabus

- In-person lectures (Wednesdays 2:30pm - 4:30pm) at WIRB-1170
- In-person lab practicals (Thursdays 2:30pm - 4:30pm) at HSB-16
- Team taught (Dr. Parisa Shooshtari, Dr. Christina Castellani, Dr. Roux-Cil Ferreira)

Getting help

- Please use OWL Forum to post questions!
- Teaching Assistant
  - Zi Huai (Matthew) Huang
  - Elly Shin

Evaluation

- Lab assignments (60%)
  Practical application of topics covered in class
- Literature review (10%)
  - Students are expected to work independently.
  - Submit a short written report (due Oct 30, 2023)
- Oral presentation (10%) (Dec 4-8, 2023)
- Oral examination (20%) (Exam period)
  - Individual assessment of basic knowledge by course instructors.

Topics

- Data formats
- Working on the command line (Linux/Unix)
- Sequence alignment, analysis
- Building trees
- Genome-wide association studies (GWAS)
- Metagenomics
- Epigenomics
- Transcriptomics
- An introduction to mathematical modeling

Basics of Biology, Central Dogma of life

- DNA makes RNA makes protein

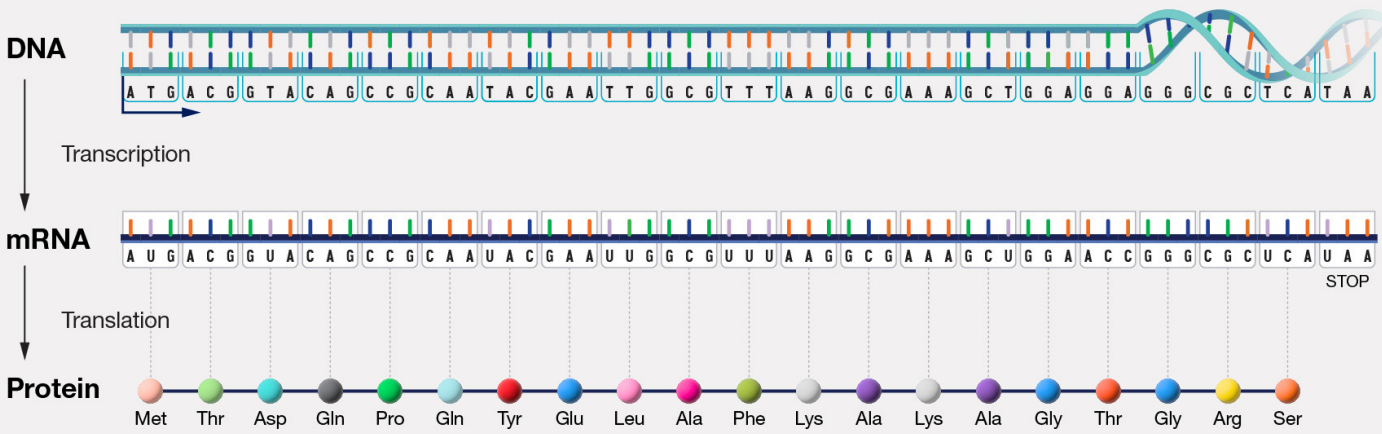


Image source: https://www.genome.gov/genetics-glossary/Central-Dogma.

Basics of Biology, Central Dogma

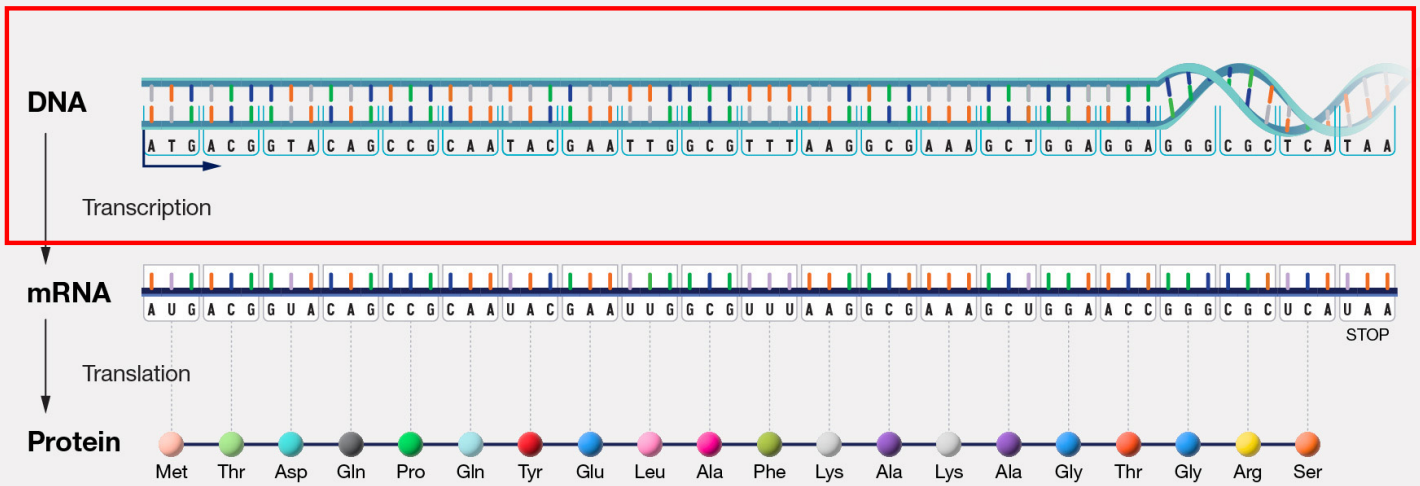


Image source: https://www.genome.gov/genetics-glossary/Central-Dogma.

The deoxyribonucleic acid (DNA)

- The deoxyribonucleic acid (DNA) codes all information of life (with some viral exceptions where information is coded in RNA).
- Composed of small molecules called nucleotides
  – four different nucleotides distinguished by the four bases: adenine (A), cytosine (C), guanine (G) and thymine (T)

The deoxyribonucleic acid (DNA)

- DNA molecules usually consist of two strands arranged in the double helix.

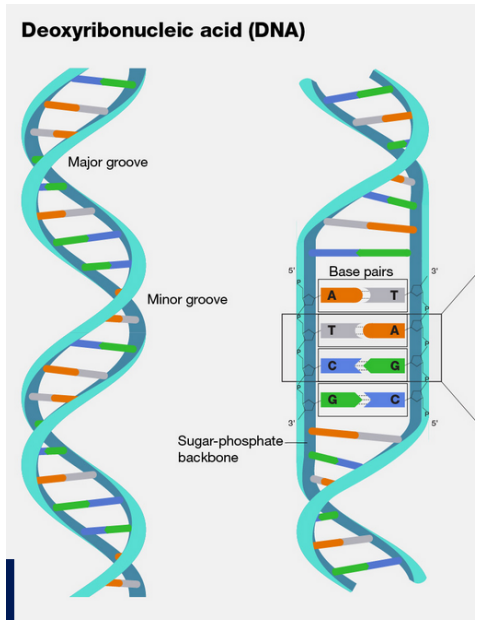


Image source: https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid.

The deoxyribonucleic acid (DNA)

- In canonical Watson–Crick base pairing in DNA,
- adenine (A) forms a base pair with thymine (T),
- and guanine (G) forms a base pair with cytosine (C)

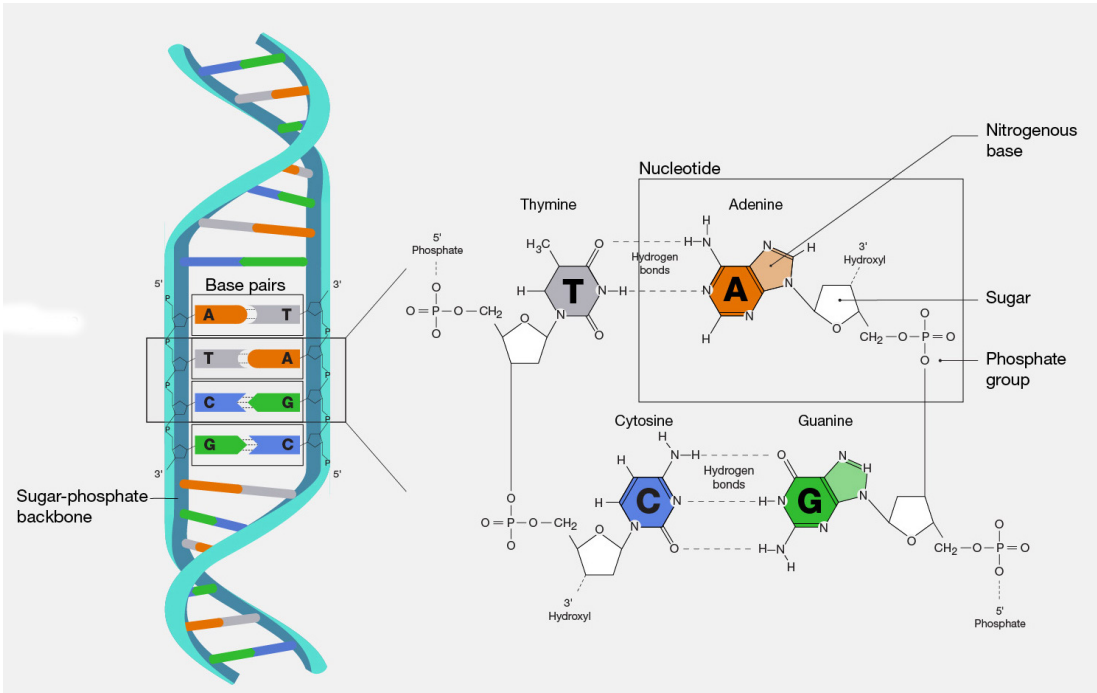


Image source: https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid.

The deoxyribonucleic acid (DNA)

- DNA strands has a "direction"
  - at one end, the terminal carbon atom in the backbone is the 5' carbon atom of the terminal sugar
  - at the other end, the terminal carbon atom is the 3' carbon atom of the terminal sugar
- therefore we can talk about the 5' and the 3' ends of a DNA strand
- in a double helix, the strands are antiparallel (arrows drawn from the 5' end to the 3' end go in opposite directions)

Basics of Biology, Central Dogma of life

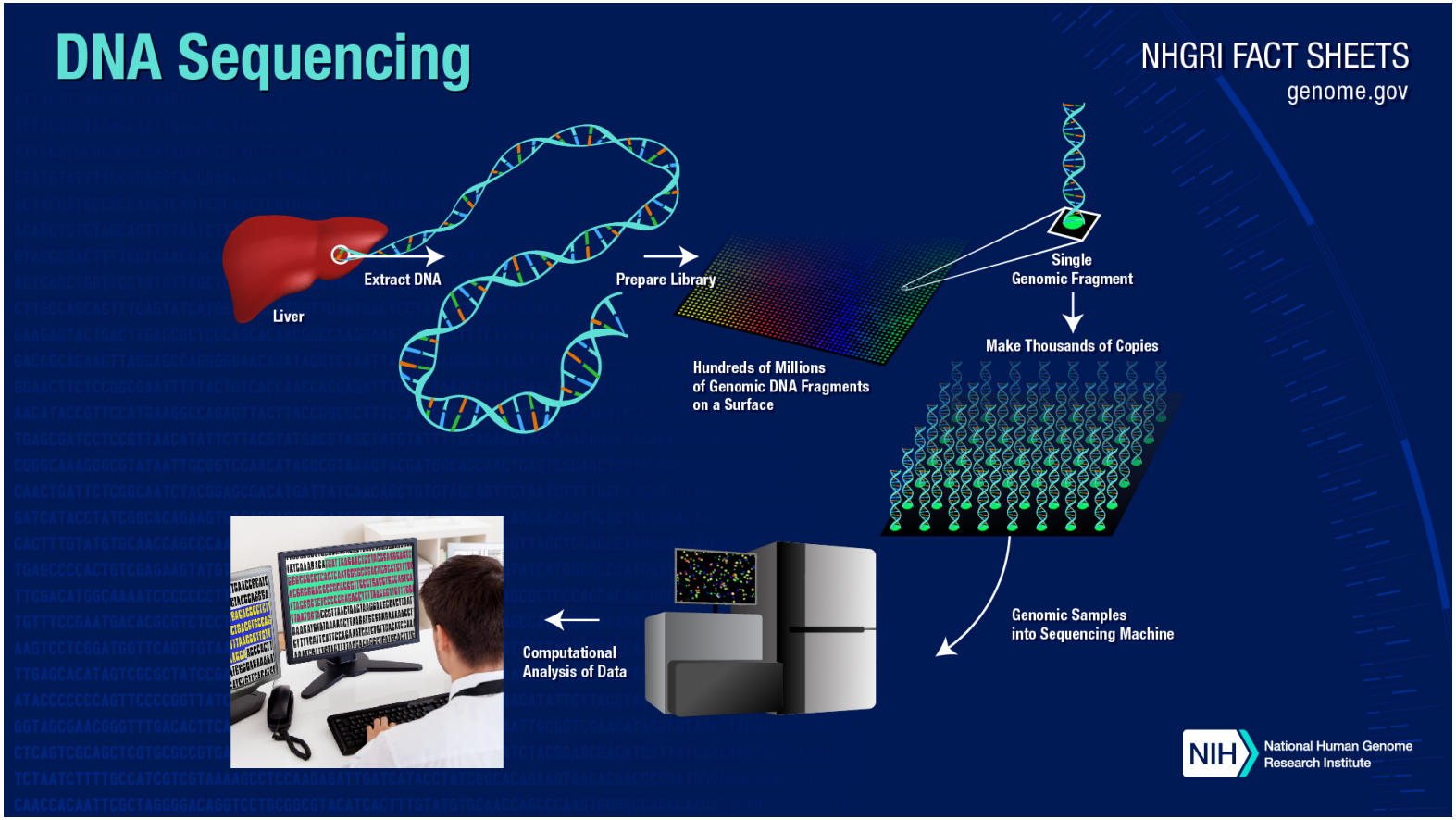


Image source: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet.

Basics of Biology, Central Dogma of life

- Whats is a DNA sequence

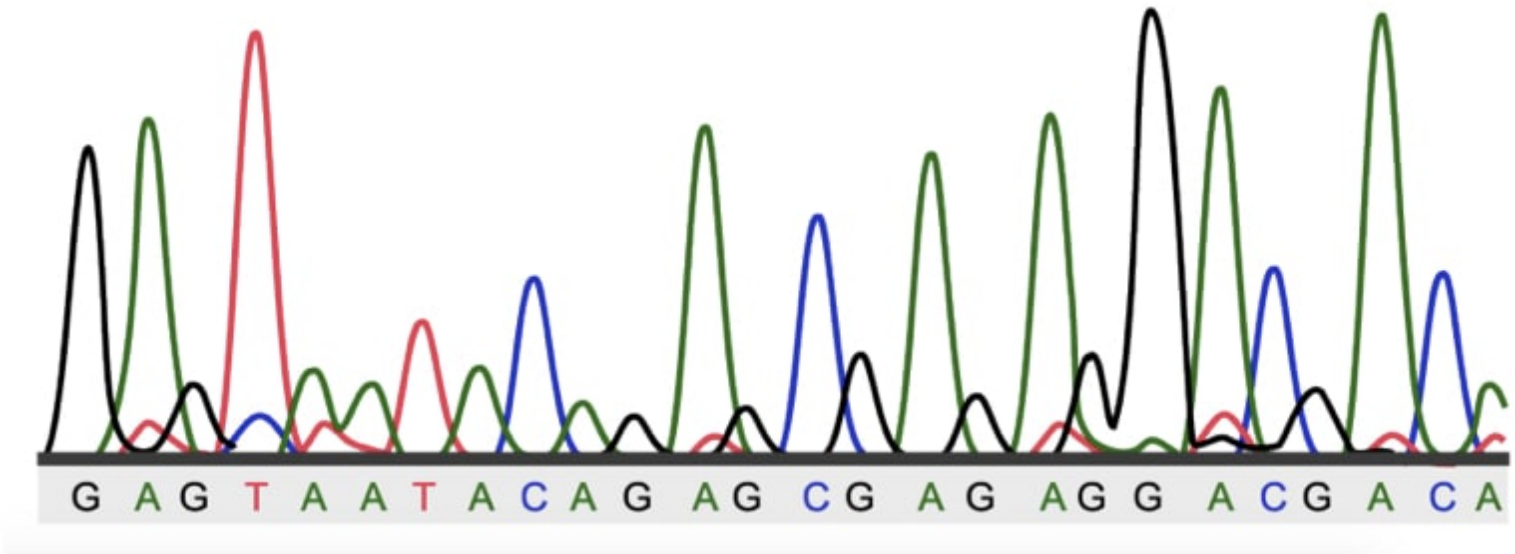
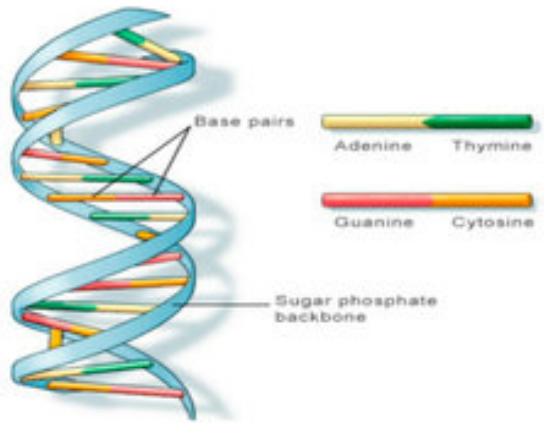


Image source: https://geneticeducation.co.in/what-is-electropherogram-how-to-read-it/.

DNA

- DNA sequence information is stored in different formats




```
CTGGGGCTTTACTGATGTCATACCGTCTTGCACGGGGATAGAAT
ATTTTCTGAAAGTTACAGACTTCGATTAAAAAGATCGGACTGCG
TTTTTCGACGTGTCAAGGACTCAAGGGAATAGTTTGGCGGGAGC
CGATAAAATTCAACTACTGGTTTCGGCCTAATAGGTCACGTTTT
CCCTGGGTGTTCTATGATAAGTCCTGCTTTATAACACGGGGCGG
ATCCAAGCGCCCGCTAATTCTGTTCTGTTAATGTTCATACCAAT
AGCCCAGTCGCAAGGGTCTGCTGCTGTTGTCGACGCCTCATGTT
GGTTAAGGCGTGTGATCGACGATGCAGGTATACATCGGCTCGGA
TCGCGGTTCGGCGCGTAGTTGAGTGCGATAACCCAACCGGTGGC
AGACAACCTAACTAATAGTCTCTAACGGGGAATTACCTTTACCA
CAATGATATCGCCCACAGAAAGTAGGGTCTCAGGTATCGCATAC
GACAGTAGAGAGCTATTGTGTAATTCAGGCTCAGCATTCATCGA
```

Image source: Improving Pattern Matching performance in Genome sequences using Run Length Encoding in Distributed Raspberry Pi Clustering Environment by Pratik Kanani and Mamta Padole

Chromosomes

DNA is packaged into individual chromosomes

- prokaryotes (single-celled organisms lacking nuclei) typically have a single circular chromosome
  – examples: bacteria, archea
- eukaryotes (organisms with nuclei) have a species-specific number of linear chromosomes
  – examples: animals, plants, fungi

Genomes

- the term genome refers to the complete complement of DNA for a given species
- the human genome consists of 23 pairs of chromosomes
  – mosquitos have 3 pairs
  – camels have 35 pairs
- every cell (except sex cells and mature red blood cells) contains the complete genome of an organism

Genomes

- genes are the basic units of heredity
- a gene is a sequence of bases that carries the information required for constructing a particular protein (more accurately, polypeptide)
- such a gene is said to encode a protein
- the human genome comprises 25,000 protein-coding genes

Gene Density

Gene Density

- not all of the DNA in a genome encodes protein:
  bacteria 90% coding gene/kb
  human 1.5% coding gene/35kb

Basics of Biology, Central Dogma

Transcription is the process of making an RNA copy of a gene's DNA sequence.

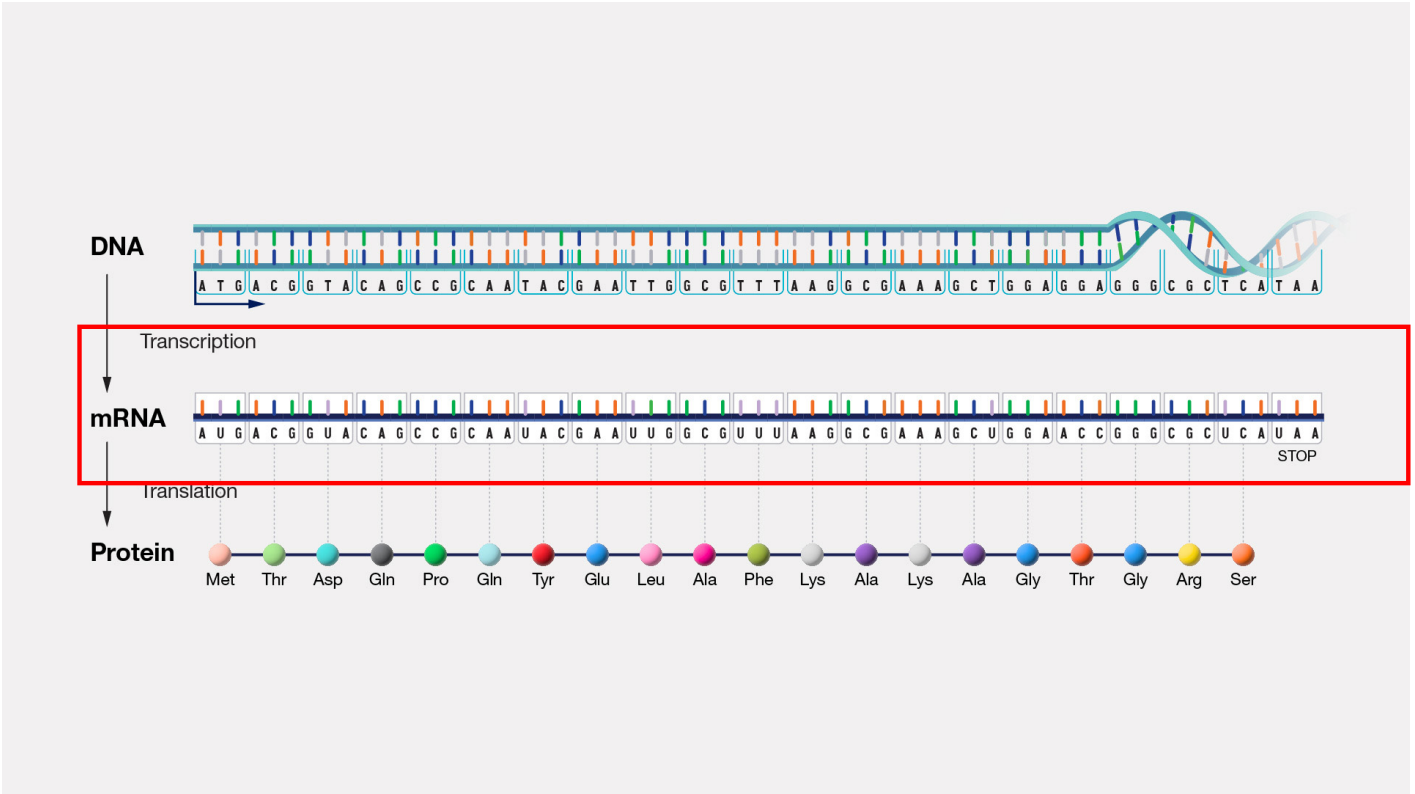


Image source: https://www.genome.gov/genetics-glossary/Central-Dogma.

RNA

- RNA is like DNA except:
  – often single stranded
  – the base uracil (U) is used in place of thymine (T)
- a strand of RNA can be thought of as a string composed of the four letters: A, C, G, U

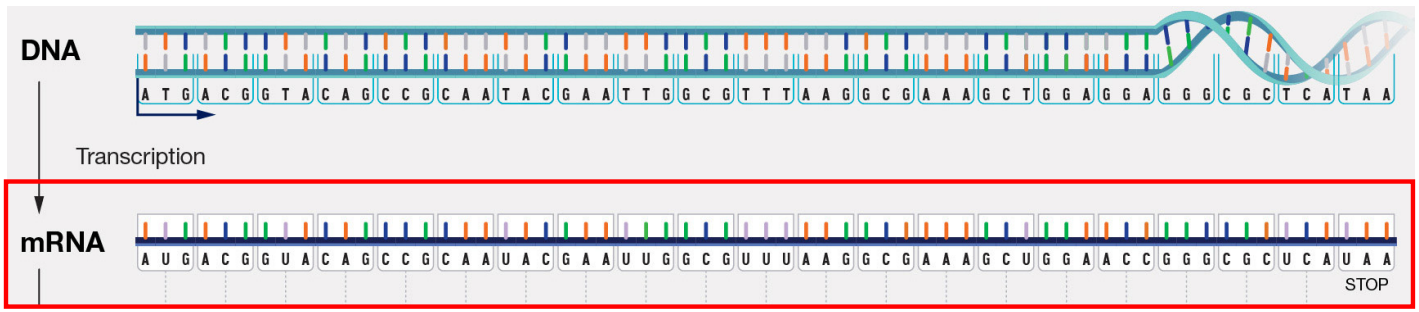


Image source: https://www.genome.gov/genetics-glossary/Central-Dogma.

RNA

- RNA is like DNA except:
  – often single stranded
  – the base uracil (U) is used in place of thymine (T)
- a strand of RNA can be thought of as a string composed of the four letters: A, C, G, U

Basics of Biology, Central Dogma

Translation is the process by which a protein is synthesized from the information contained in a molecule of messenger RNA (mRNA).

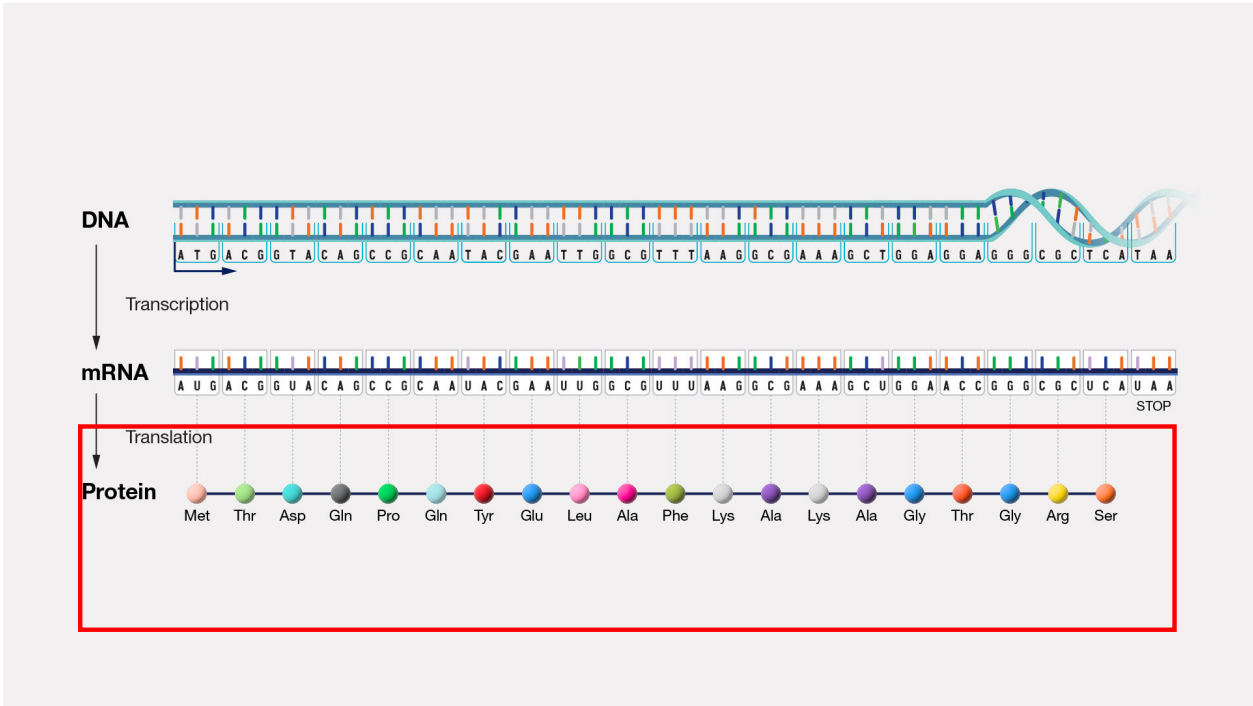


Image source: https://www.genome.gov/genetics-glossary/Central-Dogma.

Amino Acid codes

Amino acid	Three letter symbol	One letter symbol*
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

source: https://link.springer.com/article/10.1007/s00500-010-0624-9.