EECS 4425: **Introductory Computational Bioinformatics**

Suprakash Datta
Email: datta [at] eecs.yorku.ca
Course page: www.cse.yorku.ca/course/4425
Office: LAS 3043

Many of the slides have been taken from the book website

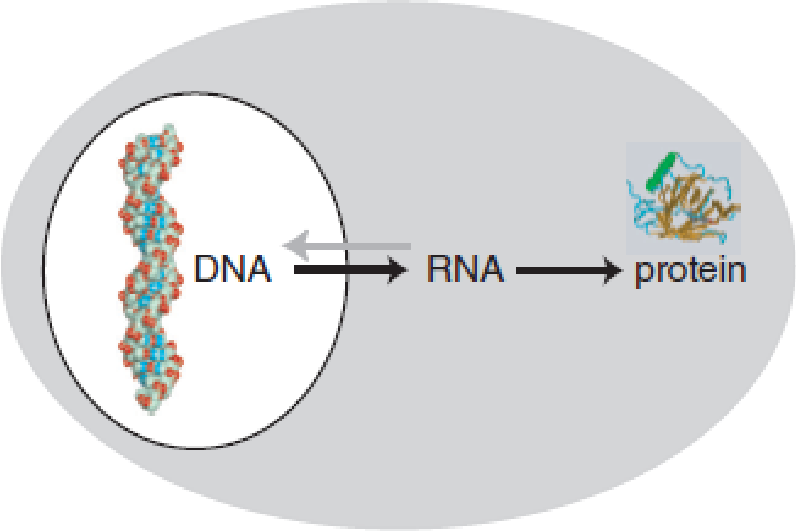
Central dogma of molecular biology & genomics

Central dogma of molecular biology

DNA → RNA → protein → cellular phenotype

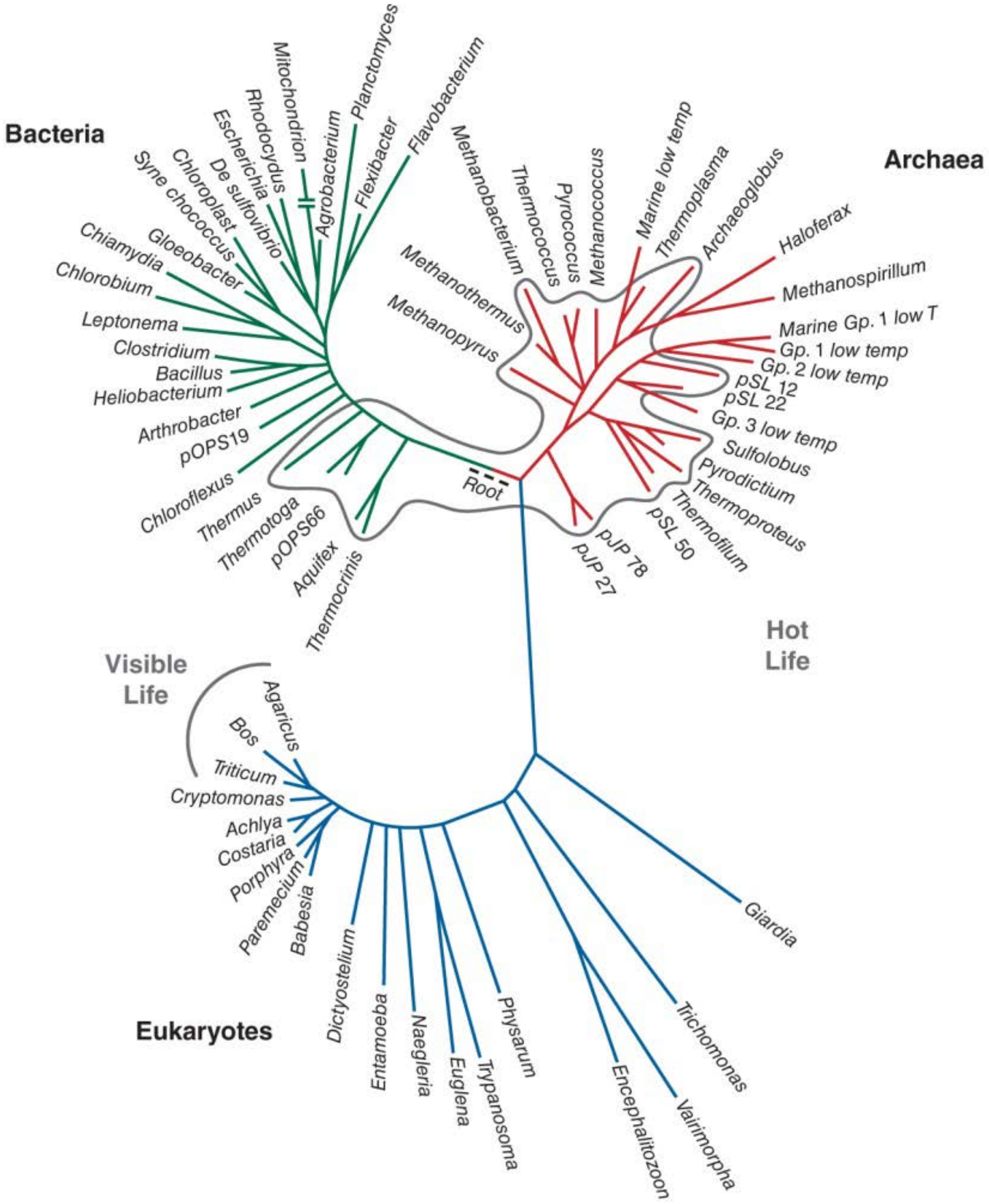
Central dogma of genomics

genome → transcriptome → proteome → cellular phenotype



B&FG 3e
Fig. 1-1
Page 4

Three domains of life: bacteria, archaea, eukaryotes



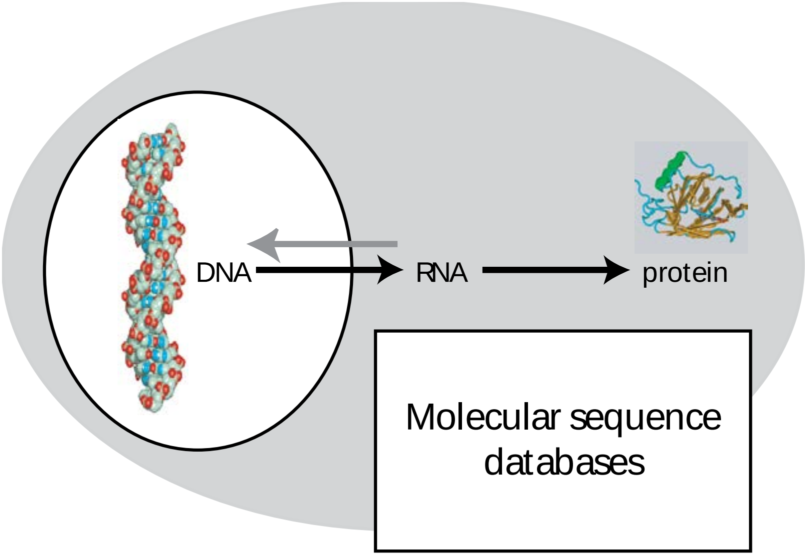
B&FG 3e
Fig. 1-3
Page 7

The role of Taxonomy

- Prokaryotes – single celled, no membrane enclosed nucleus
- Eukaryotes – may be unicellular (algae, yeast) or multicellular, membrane enclosed nucleus and organelles
- The role of evolution in complexity
- Problems: definition of species in prokaryotes

Outline

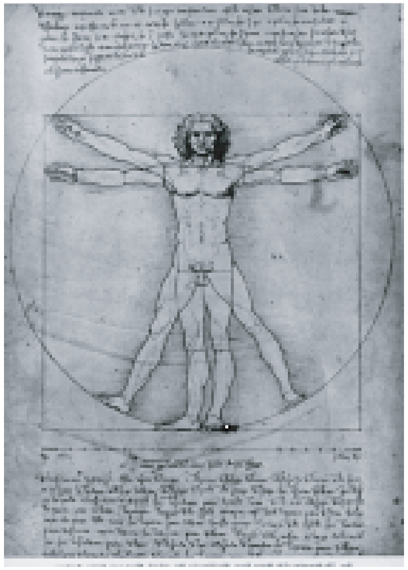
Organization of the book
Bioinformatics: the big picture
Organization of the chapters
Suggestions For Students and Teachers:
    Exercises, Find-a-Gene, Characterize-a-Genome
Bioinformatics software: two cultures
    Web-based software
    Command-line software
    Bridging the two cultures
    New paradigms for learning programming
Bioinformatics and other disciplines



Part 1: Bioinformatics: analyzing DNA, RNA, and protein

Chapter 1: Introduction
Chapter 2: How to obtain sequences
Chapter 3: How to compare two sequences
Chapters 4 and 5: How to compare a sequence
across databases
Chapter 6: How to multiply align sequences
Chapter 7: How to view multiply aligned sequences
as phylogenetic trees

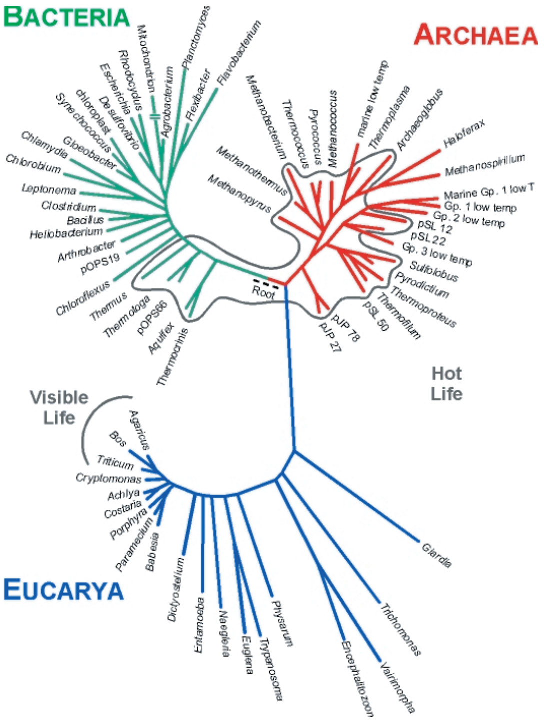
Figure 1.4
Bioinformatics and Functional Genomics (3rd ed., 2015)



Part 2: Functional genomics: from DNA to RNA to protein

Chapter 8: DNA: The eukaryotic chromosome
Chapter 9: DNA analysis: next-generation sequencing
Chapter 10: Bioinformatics approaches to RNA
Chapter 11: Microarray and RNA-seq data analysis
Chapter 12: Protein analysis and protein families
Chapter 13: Protein structure
Chapter 14: Functional genomics

Figure 1.4
Bioinformatics and Functional Genomics (3rd ed., 2015)



Part 3: Genomics

Chapter 15: The tree of life
Chapter 16: Viruses
Chapter 17: Bacteria and archaea
Chapter 18: Fungi
Chapter 19: Eukaryotes from parasites to plants to primates
Chapter 20: The human genome
Chapter 21: Human disease

Figure 1.4
Bioinformatics and Functional Genomics (3rd ed., 2015)

Outline

Organization of the book
Bioinformatics: the big picture
Organization of the chapters
Suggestions For Students and Teachers:
    Exercises, Find-a-Gene, Characterize-a-Genome
Bioinformatics software: two cultures
    Web-based software
    Command-line software
    Bridging the two cultures
    New paradigms for learning programming
Bioinformatics and other disciplines

Bioinformatics and genomics: two cultures

Many bioinformatics tools and resources are available on the internet, such as major genome browsers and major portals (NCBI, Ensembl, UCSC).

These are:

- accessible (requiring no programming expertise)
- easy to browse to explore their depth and breadth
- very popular
- familiar (available on any web browser on any platform)

B&FG 3e
Fig. 2-3
Page 22

Web-based or
graphical user interface (GUI)

Central resources
(NCBI,
EBI,)

Genome browsers
(UCSC, Ensembl)

GUI software
(Partek, MEGA,
RStudio,
BioMart,
IGV)

Galaxy
(web access
to NGS tools,
browser data)

Command line (often Linux)

Biopython,
Python, BioPerl, R:
manipulate data files

Data analysis
software: sequences,
proteins, genomes

Next generation
sequencing tools

Figure 1.5
Bioinformatics and Functional Genomics (3rd ed., 2015)

Bioinformatics and genomics: two cultures

Many bioinformatics tools and resources are available on the command-line interface (sometimes abbreviated CLI).

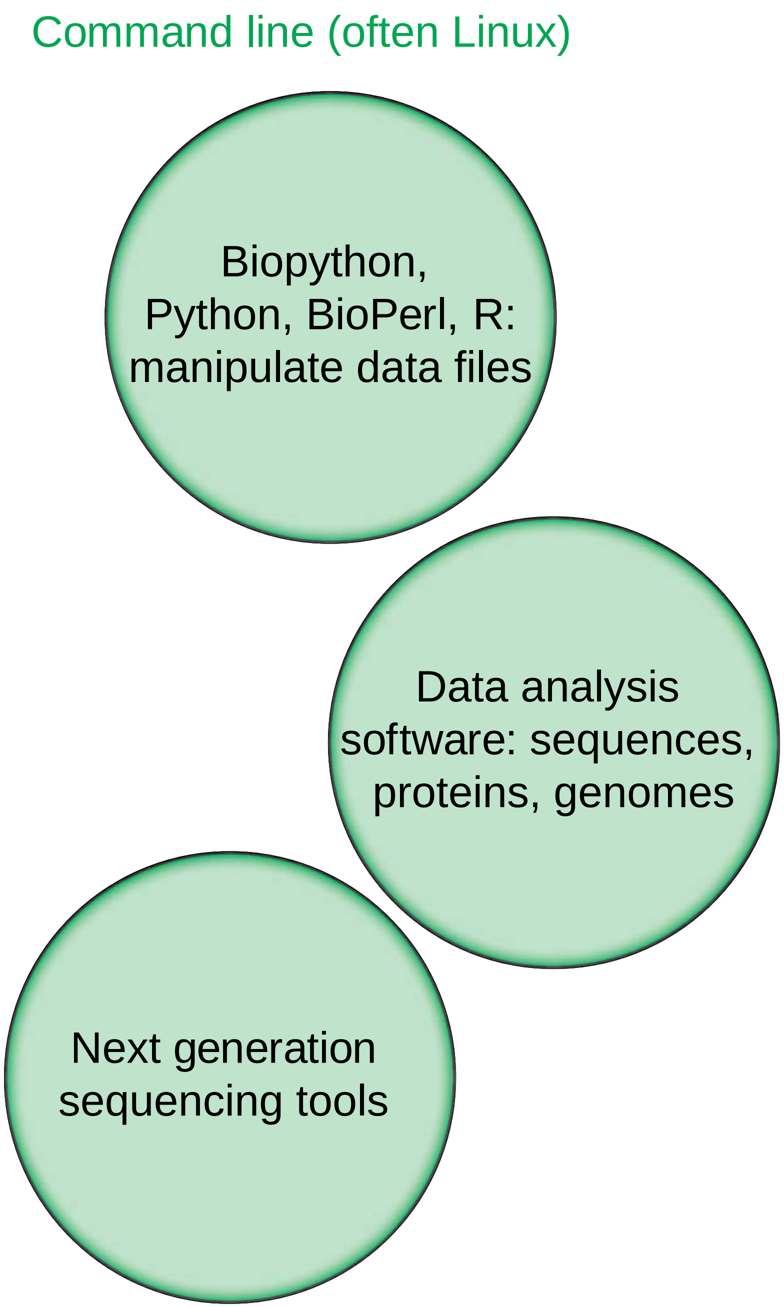
These are often on the Linux platform (or other Unix-like platforms such as the Mac command line). They are essential for many bioinformatics and genomics applications.

- Most bioinformatics software is written for the Linux platform.
- Many bioinformatics datasets are so large (e.g. high throughput technologies generate millions to billions or even trillions of data points) requiring command-line tools to manipulate the data.

Should **you** learn to use the Linux operating system? Yes, if you want to use mainstream bioinformatics tools.

Should **you** learn Python or Perl or R or another programming language? It’s a good idea if you want to go deeper into bioinformatics, but also, it depends what your goals are. Many software tools can be run in Linux on the command-line without needing to program.

Think of this figure like a map. Where are you now? Where do you want to go?



Command line (often Linux)

Biopython, Python, BioPerl, R: manipulate data files

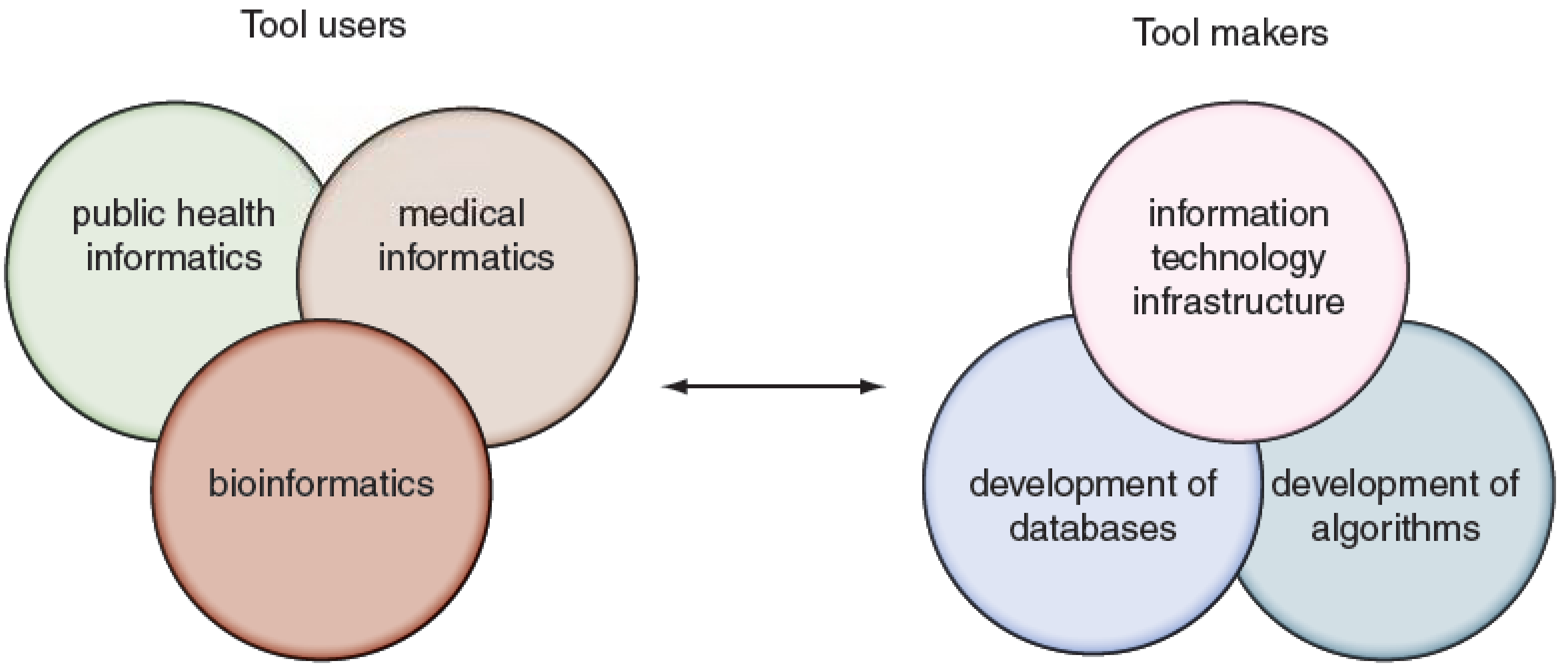
Data analysis software: sequences, proteins, genomes

Next generation sequencing tools

Outline

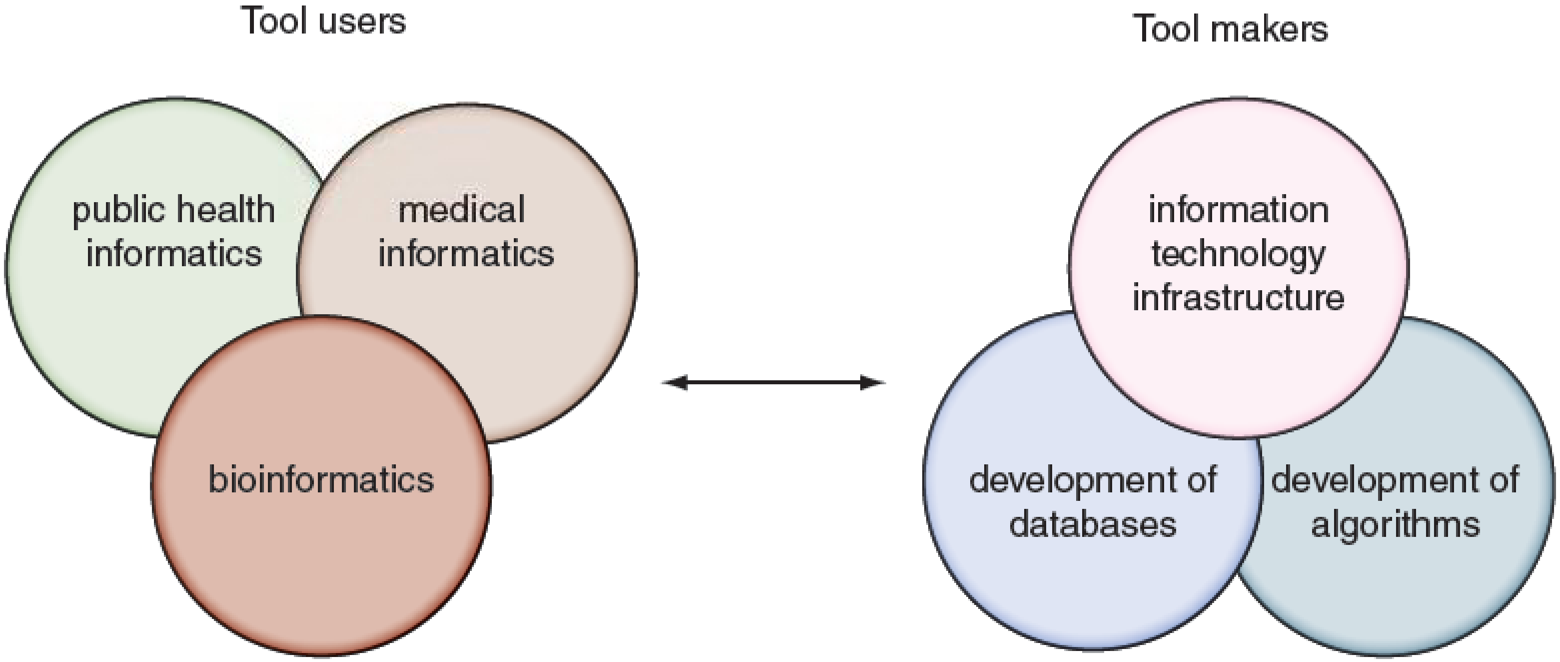
Organization of the book
Bioinformatics: the big picture
Organization of the chapters
Suggestions For Students and Teachers:
    Exercises, Find-a-Gene, Characterize-a-Genome
Bioinformatics software: two cultures
    Web-based software
    Command-line software
    Bridging the two cultures
    New paradigms for learning programming
Bioinformatics and other disciplines

Tool makers and tool users across informatics disciplines



B&FG 3e
Fig. 1.6
Page 15

Tool makers and tool users across informatics disciplines



Many informatics disciplines have emerged in recent years. Bioinformatics is distinguished by its particular focus on DNA and proteins (impacting its databases, its tools, and its entire culture).

B&FG 3e
Fig. 1.6
Page 15

Areas of Bioinformatics

- Genomics
- Proteomics
- Systems Biology
- Transcriptomics
- Metabolomics
- Epigenomics

What about Medical Image Processing?
Medical informatics?