In the Name of God, the Merciful, the Compassionate

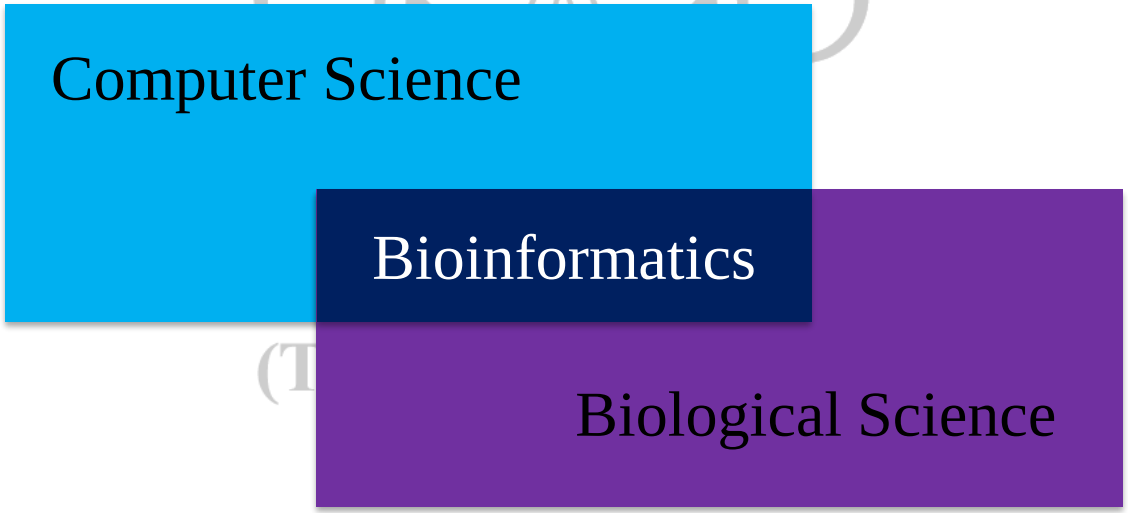
Introduction to Bioinformatics
01: Introduction

Instructor: Hossein Zeinali

Amirkabir University of Technology

What is Bioinformatics?

- Wikipedia
  - Bioinformatics: an **interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analyzing biological data**. A major activity in bioinformatics is to develop software tools to generate useful biological knowledge.



What is Bioinformatics?

- Wikipedia
  - Bioinformatics: an **interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analyzing biological data**. A major activity in bioinformatics is to develop software tools to generate useful biological knowledge.
- Differs from *computational biology*: Bioinformatics is limited to sequence, structural, and functional analysis of genes and genomes and their corresponding products and is often considered *computational molecular biology*. However, computational biology encompasses all biological areas that involve computation.

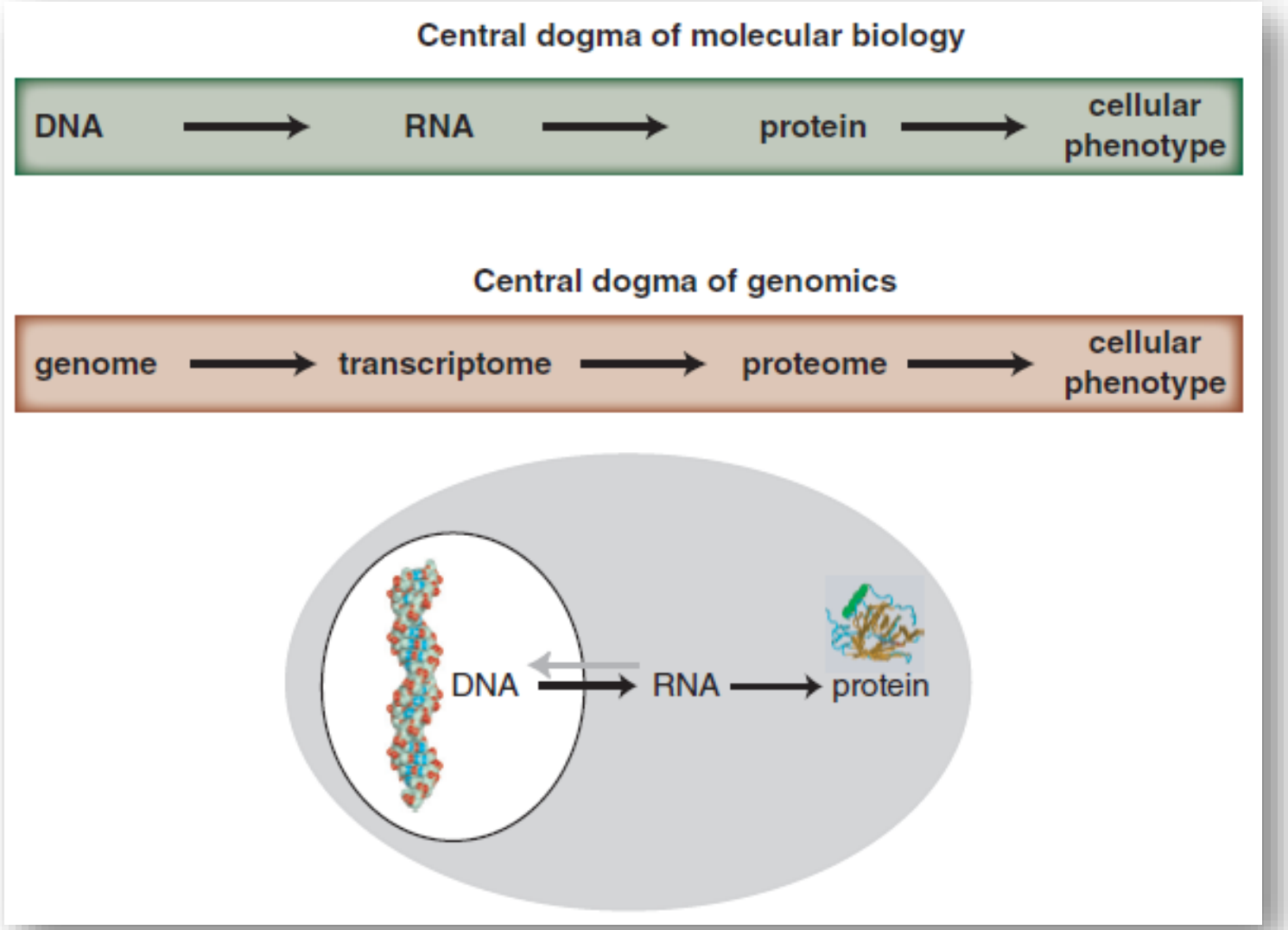
What is Bioinformatics? (Cont.)

- The ultimate goal of bioinformatics is to *better understand a living cell* and how it functions at the molecular level.
- Bioinformatics consists of two subfields:
  - Development of computational tools and databases.
  - Application of these tools and databases in generating biological knowledge to better understand living systems.

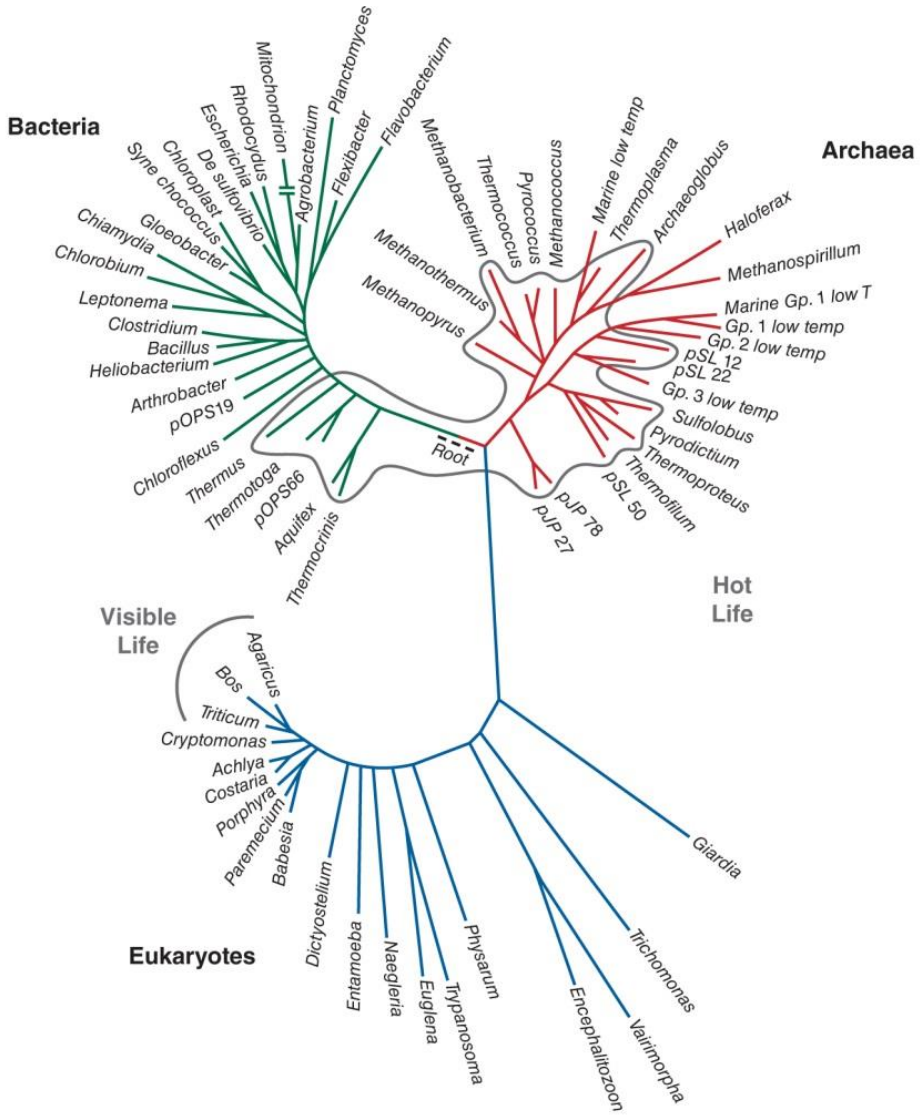
Another Definitions of Bioinformatics

- Bioinformatics is the use of computer databases and computer algorithms to analyze proteins, genes, and the complete collection of deoxyribonucleic acid (DNA) that comprises an organism (the genome).
- According to a *National Institutes of Health (NIH)* definition, bioinformatics is “research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral, or health data, including those to acquire, store, organize, analyze, or visualize such data.”

Central Dogma of Molecular Biology & Genomics

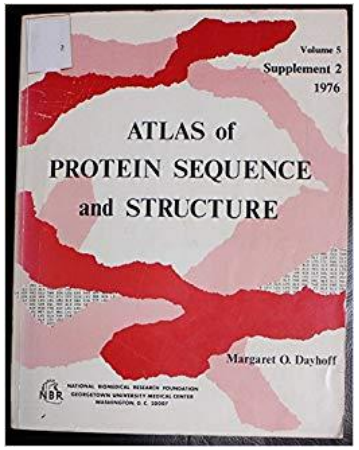


Three Domains of Life: Bacteria, Archaea, Eukaryotes



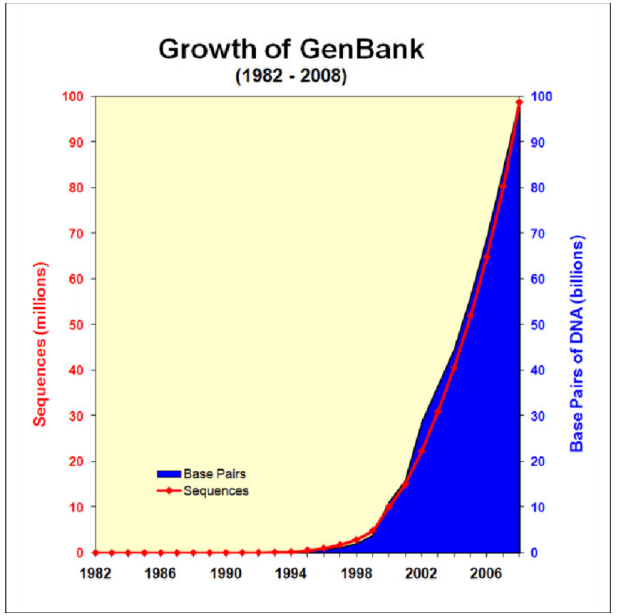
A Short History

- **1965**: the first major bioinformatics project was undertaken. Margaret Dayhoff developed a first protein sequence database called *Atlas of Protein Sequence and Structure*.
- **Early 1970s**: the Brookhaven National Laboratory established the Protein Data Bank for archiving three-dimensional protein structures.
- **1970**: the first sequence alignment algorithm was developed by Needleman and Wunsch.



A Short History (Cont.)

- **1974**: the first protein structure prediction algorithm was developed by Chou and Fasman.
- **1982**: the establishment of GenBank database.
- **1980s**: the development of fast database searching algorithms such as *FASTA* and *BLAST*.
- **1990-2003**: the Human Genome Project was started in the late 1980s. It provides a major boost for the development of bioinformatics.



Aspects of Molecular Biological Research

Molecular sequence analysis:

- Sequence alignment
- Sequence database searching
- Motif and pattern discovery
- Gene and promoter finding
- Reconstruction of evolutionary relationships
- Genome assembly and comparison

Molecular structural analysis:

- Protein and nucleic acid structure analysis
- Comparison, classification, and prediction

Molecular functional analysis:

- Gene expression profiling
- Protein–protein interaction prediction
- Protein subcellular localization prediction
- Metabolic pathway reconstruction, and simulation

Bioinformatics Applications

- Knowledge-based drug design: reducing development time and cost, fewer side effects, and less toxicity
- Forensic DNA analysis
- Individual medicine: developing personalized and customized medicine
- Agricultural biotechnology: development of new crop varieties that have higher productivity and more resistance to disease.

Limitations of Bioinformatics

- Bioinformatics has a number of inherent limitations:
  - Bioinformatics and experimental biology are independent, but complementary, activities.
  - Bioinformatics predictions are not formal proofs of any concepts.
  - The quality of bioinformatics predictions depends on the *quality of data* and the *sophistication of the algorithms* being used.
  - Sequence data from high throughput analysis often contain errors.
  - If the *sequences are wrong* or *annotations incorrect*, the results from the downstream analysis are misleading as well.
  - Bioinformatics is by no means a mature field.
  - There is a necessary trade-off between accuracy and computational feasibility.

More Information about our Coarse

سرفصل مطالب:

- **مقدمه**
  - تعریف بیوانفورماتیک
  - حوزه‌های اصلی بیوانفورماتیک
  - مرور مفاهیم پایه در زیست شناسی
- **تحلیل توالی‌های DNA، RNA و پروتئینی**
  - دسترسی به داده‌های توالی و اطلاعات مربوطه
  - تطابق جفت توالی
  - BLAST
  - جستجوی پیشرفته داده پایگاه‌ها
  - تطابق چند توالی
  - فیلوژنی مولکولی و تکامل
- **تحلیل ژنوم کامل توالی‌های DNA، RNA و پروتئینی**
  - DNA: کروموزوم یوکاریوتی
  - تحلیل داده‌های نسل بعد
  - روش‌های بیوانفورماتیکی مرتبط با RNA
  - بیان ژن: تحلیل داده‌های میکروآرایه و RNA-seq
  - تحلیل پروتئین و پروتئومیکس
  - ساختار پروتئین
  - ژنومیک عملکردی
- **مقدمه‌ای بر زیست شناسی سیستمی**

References

- Biology:
  - Alberts, Bruce, Dennis Bray, Karen Hopkin, Alexander D. Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. ***Essential cell biology***. Garland Science, 2013.
- Bioinformatics:
  - Pevsner, Jonathan. ***Bioinformatics and functional genomics***. John Wiley & Sons, 2015.
  - Xiong, Jin. ***Essential bioinformatics***. Cambridge University Press, 2006.
- R Programing:
  - Gentleman, Robert. ***R programming for bioinformatics***. Chapman and Hall/CRC, 2008.

Grading Policy

- Midterm exam: 25 %
- Final exam: 35 %
- Short quizzes: 10 %
- Assignments: 30 %
- All may include bonuses
- Plus 0.6 point for class attendance

Thanks for your attention