

LSM2241 Lecture 1

Bioinformatics: First steps

Greg Tucker-Kellogg

12 August 2015

Outline

What is Bioinformatics

Bioinformatics and the cell

Bioinformatics and the tree of life

Afterward

Topic

What is Bioinformatics

Bioinformatics and the cell

Bioinformatics and the tree of life

Afterward

What is Bioinformatics?

Paulein Hogeweg (1970s) the study of informatic processes in biotic systems

US National Institutes of Health (2004) 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, archive, analyze, or visualize such data.

Dubitzky (2004) the development and application of techniques from computer science, mathematics and statistics to address biological problems

GTK (in the lift) Using computational approaches to study hard problems in the biological sciences

Why so many definitions?

- Thinking of DNA *as information* dates to the 1940s
 - ▶ dawn of computer age
 - ▶ dawn of molecular biology
- Systems dynamics dates from the 1960s
- Sequence analysis began to get more rigorous in the 1970s
- “Genomics” coined in 1986
- Human Genome Project initiated in 1990, completed in 2004
- Until 1980s, most sequence information was protein sequence, followed by RNA

Why Bioinformatics?

- Bioinformatics represents a convergence of modern disciplines: computing, statistics, and biology, to study biological problems
- Bioinformatics and computational biology make it possible to take biology to the next level
 - ▶ Huge data challenges in the post-genomic era
 - ▶ Thinking about living systems with information theory
- In 1990, incorporating computation into biology was unusual
- In 2013, incorporating computation into biology is *expected*

Where do we find it? In databases

- Many databases
- Some we use all the time
- Some we use once in a while
- Some we never use

Topic

What is Bioinformatics

Bioinformatics and the cell

Bioinformatics and the tree of life

Afterward

Bioinformatics from the perspective of the cell

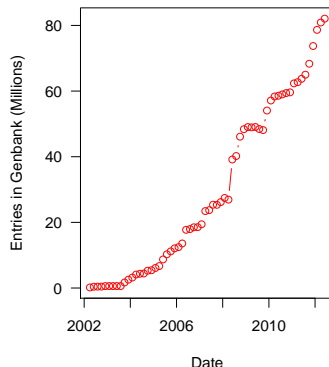
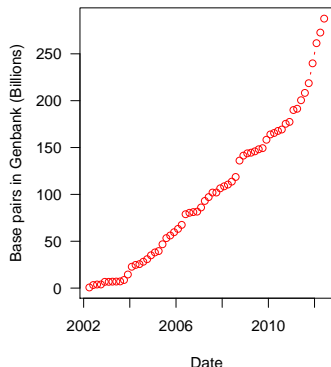
Central Dogma of Molecular Biology



Central Dogma in light of genomics



Explosive growth of sequence data

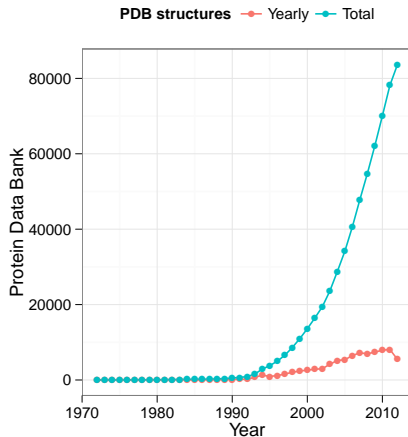


Growth of sequence data deposited in Genbank makes computer-based analysis essential

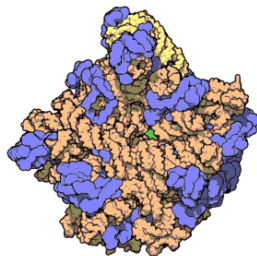
Bioinformatics: analysis of sequence information

- How to get sequences from databases
- How to compare two sequences
- How to compare a sequence to sequences in a database
- How to align sequences
 - ▶ in pairs
 - ▶ in groups
- How to view sequence alignment in phylogenetic trees

From sequence to structure



Growth of structures deposited in the Protein Data Bank



Large subunit of the ribosome, from Ban et al. 2000

The challenge: from sequence and structure to cellular function

- Can we understand the cell by simulation and experiment?
- *All* the genes
- *All* the proteins
- *All* the metabolites
- *All* the cellular compartments, intermediates, and functions
- Can we predict cellular behaviour?

Topic

What is Bioinformatics

Bioinformatics and the cell

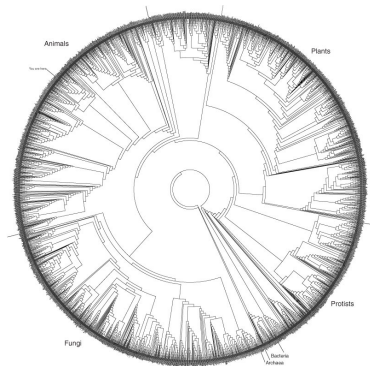
Bioinformatics and the tree of life

Afterward

The tree of life

A problem

- Most organisms leave no fossils
- The vast majority of species that have ever lived are now extinct
- **Q:** How do we understand the history of life?
- **A:** Bioinformatics!

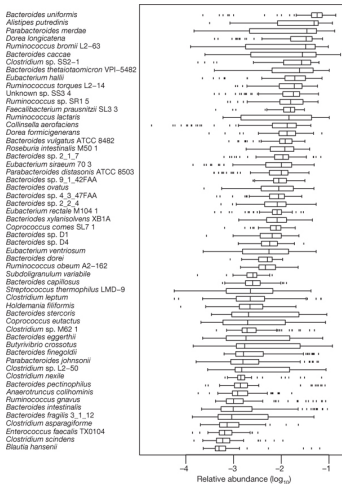


Model for universal tree of life,
from Pennisi 2003

No organism lives alone

- In humans, the vast majority of cells are not human cells
- Microorganisms are *pervasive* in your body, and essential players in your physiology
- “Metagenomics” of organismal pools, including the biota of human gut, is a rapidly growing area
- The vast majority of microscopic life is uncultured, so we know very little about it.

Example: the human gut microbiome



Abundance of organisms in the human gut microbiome from Qin et al. 2010

Assessing populations

- Faecal sequencing from 124 Europeans
- What are the differences between healthy individuals and patients suffering from IBD?

What we will cover

- How to construct and interpret phylogenetic trees
- Basics of whole genome analysis
- When different types of genomic sequencing and comparisons are appropriate

Topic

What is Bioinformatics

Bioinformatics and the cell

Bioinformatics and the tree of life

Afterward

The next step

- This module teaches you how to understand concepts and use tools provided to you
 - ▶ Algorithms
 - ▶ Web sites
 - ▶ Analysis software
- To go beyond this module, you need to provide your own tools
 - ▶ Writing scripts
 - ▶ Manipulating data directly
 - ▶ Designing pipelines
 - ▶ Advancing algorithms
- Upper level modules offer that opportunity

Bibliography



Ban, N et al. (2000). “The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution”. In: *Science* 289.5481, pp. 905–920 (cit. on p. 12).



Pennisi, Elizabeth (2003). “Modernizing the tree of life.” In: *Science* 300.5626, pp. 1692–1697. DOI: [10.1126/science.300.5626.1692](https://doi.org/10.1126/science.300.5626.1692) (cit. on p. 15).



Qin, Junjie et al. (2010). “A human gut microbial gene catalogue established by metagenomic sequencing.” In: *Nature* 464.7285, pp. 59–65. DOI: [10.1038/nature08821](https://doi.org/10.1038/nature08821) (cit. on p. 17).