Practical Bioinformatics for Biologists (BIOS 441/641)

- Course overview

Yanbin Yin

MO444

Room and computer access

- Room entry code:

2159

- Computer access:

Your ZID and password

Compared to BIOS 443/643 and 646

- Minimum theoretical intro

- A LOT of practical applications

- Goal: enhance the use of computational tools in molecular biology research

Half day on the web can save you half month in the lab

Questions

- How many have experience in Linux/Unix
- How many working in a research lab
- How many have programming experience of any kind language
- How many have used BLAST
- What bioinfo softwares you have ever used
- What softwares you will be using

Tools

- Two parts

Web-based resources

Stand-alone tools
  Windows-based (have a GUI)
  Linux-based (no GUI, command line)

Class rules:

- Attendance: students are required to attend all classes. Absences without notifying the instructor in advance will result in 5% reduction in final grade.

- Plagiarism: copying materials (figures, tables, sentences) directly from other people, literatures or internet without proper reference are considered as plagiarism and will lead to a fail of this course.

Grading:

- Attendance: 5%
- Home work: 40%
- Final report: 40%

  -Project 1: 20%

  -Project 2: 20%
- Final exam: 15%

Books:

Mainly uses slide notes, but may refer to:

- Practical Bioinformatics by Agostino, 2013 Garland Science

- Practical Computing for Biologists by Haddock and Dunn, 2011 Sinauer

- Developing Bioinformatics Computer Skills by Gibas and Jambeck, 2001 O'reilly

Class website

http://cys.bios.niu.edu/yyin/teach/PBB/

Slides, homeworks, reading materials, notes, handouts

Expectations

- Be focused inside class

- Spend at least 5 hours outside class per week

- Turn in homework on time

- Use what you learned in your research

What you can expect from this course

- How to find a particular gene in NCBI?
- How to find the homologs for that gene?
- How to get 100 protein sequences at a time?
- How to predict functions for an uncharacterized gene?
- How to perform a sequence alignment of 100 proteins?
- How to perform expression analysis?
- …

What you can expect from this course

- How to access a remote Linux machine from Windows?
- How to work with Linux terminals?
- How to install bioinformatics softwares using terminals?
- How to run bioinfo tools e.g. BLAST in the Linux terminal?
- How to do basic sequence analysis using bioinfo tools such as EMBOSS commands?
- How to extract 100 protein sequences from a large protein database, given the IDs of the proteins, using a combination of Linux commands and perl one-liner script?
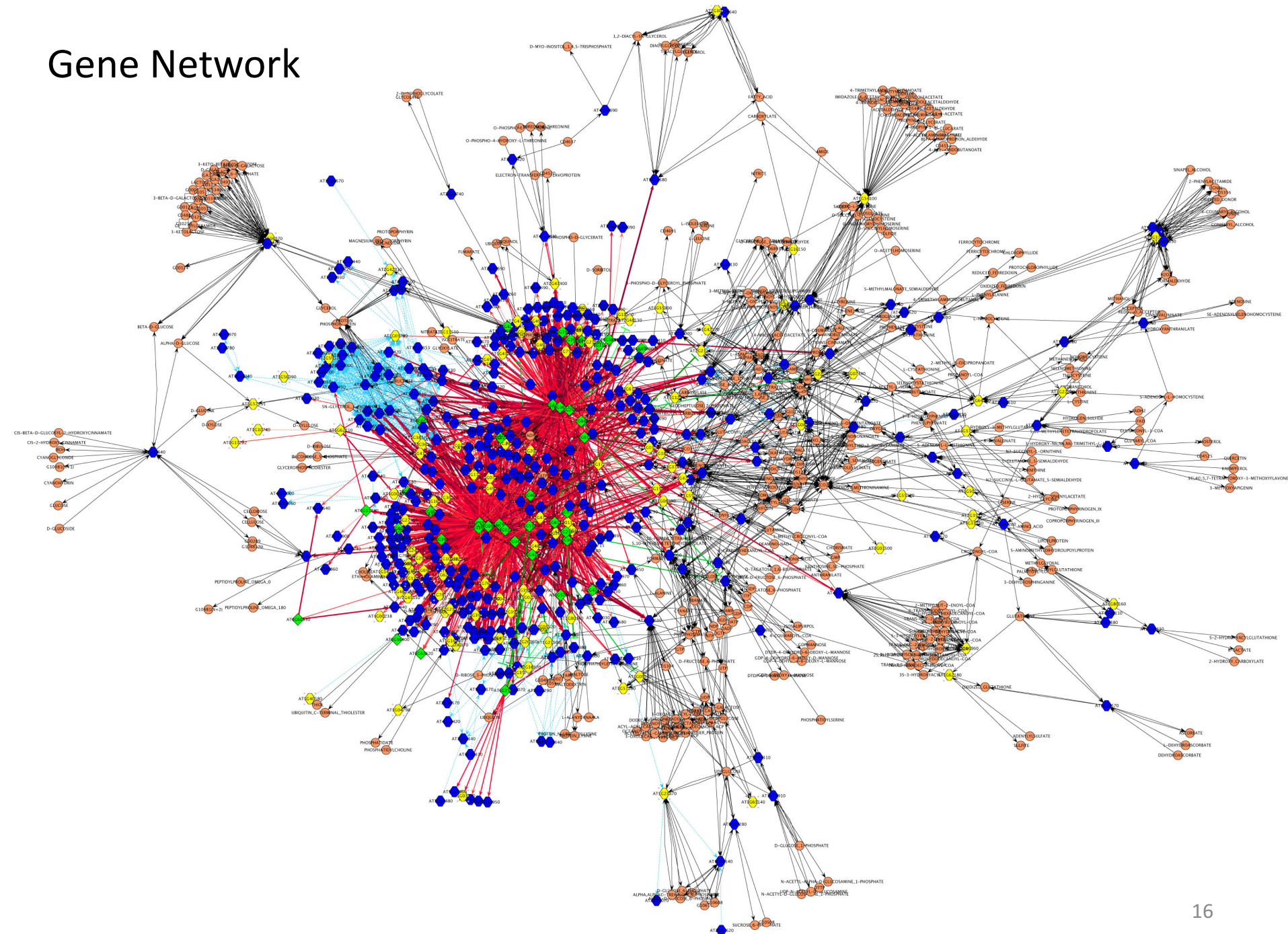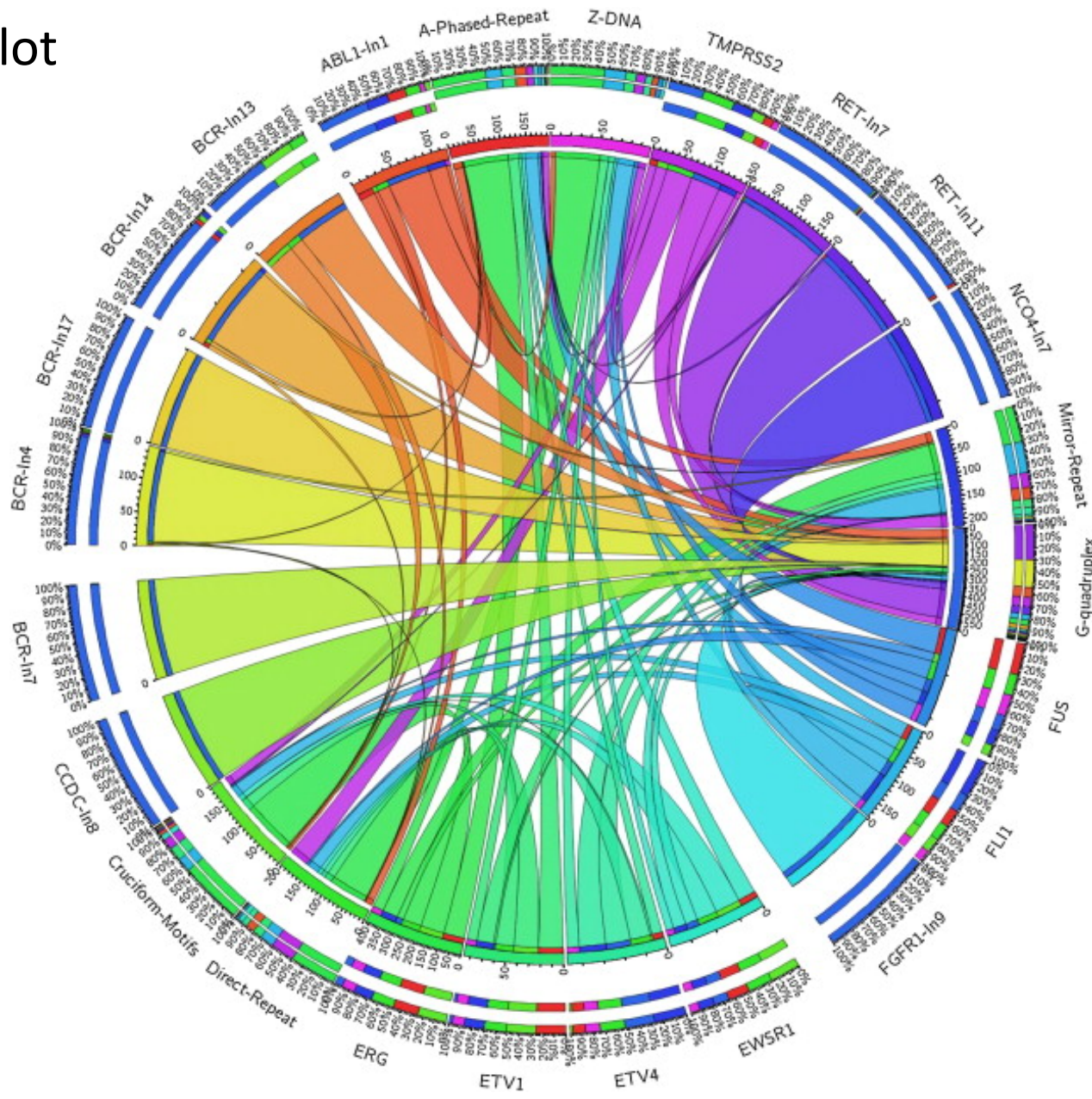- and many more …

Multiple sequence alignment
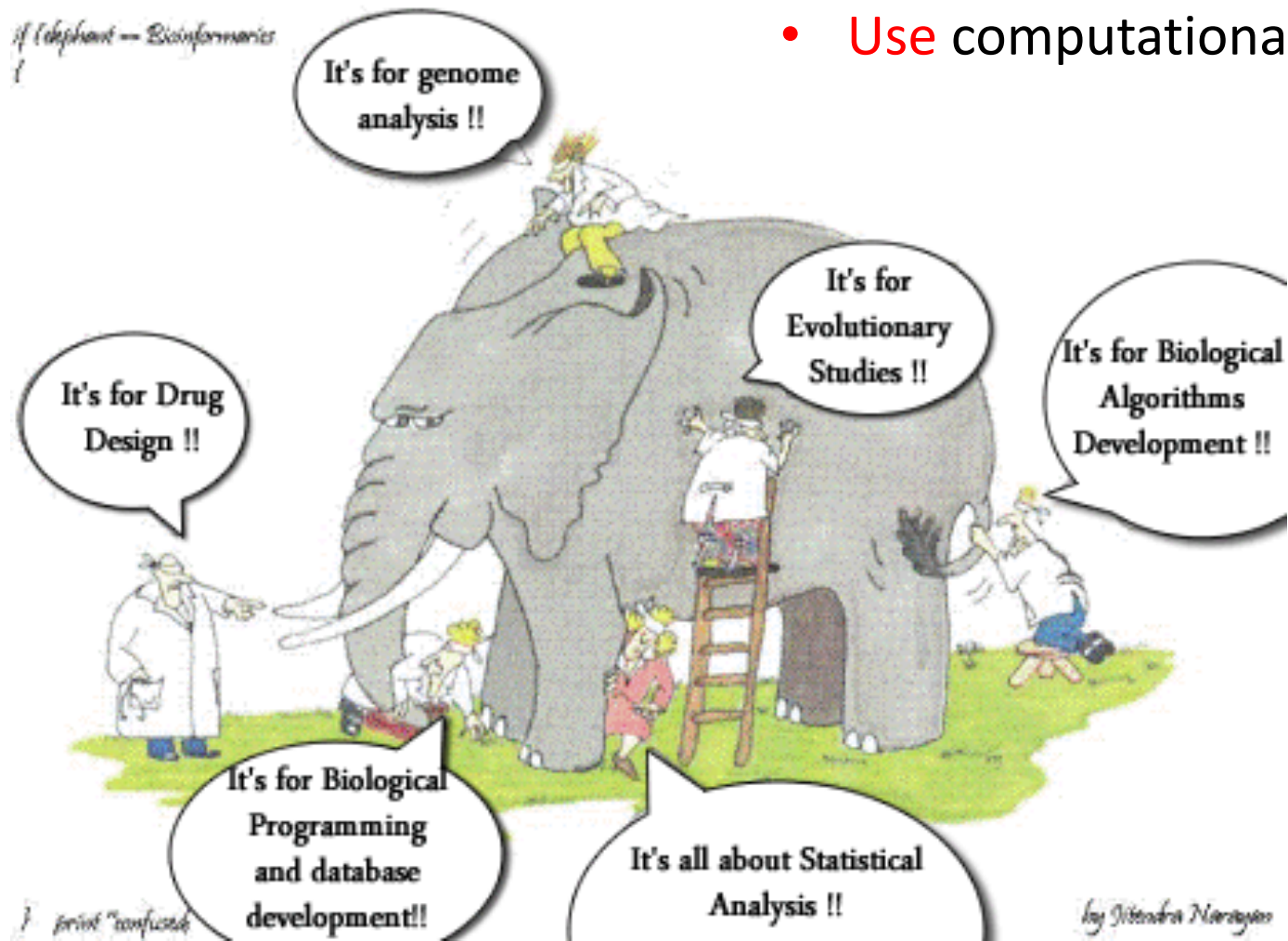
Sequence logo

Phylogenetic tree
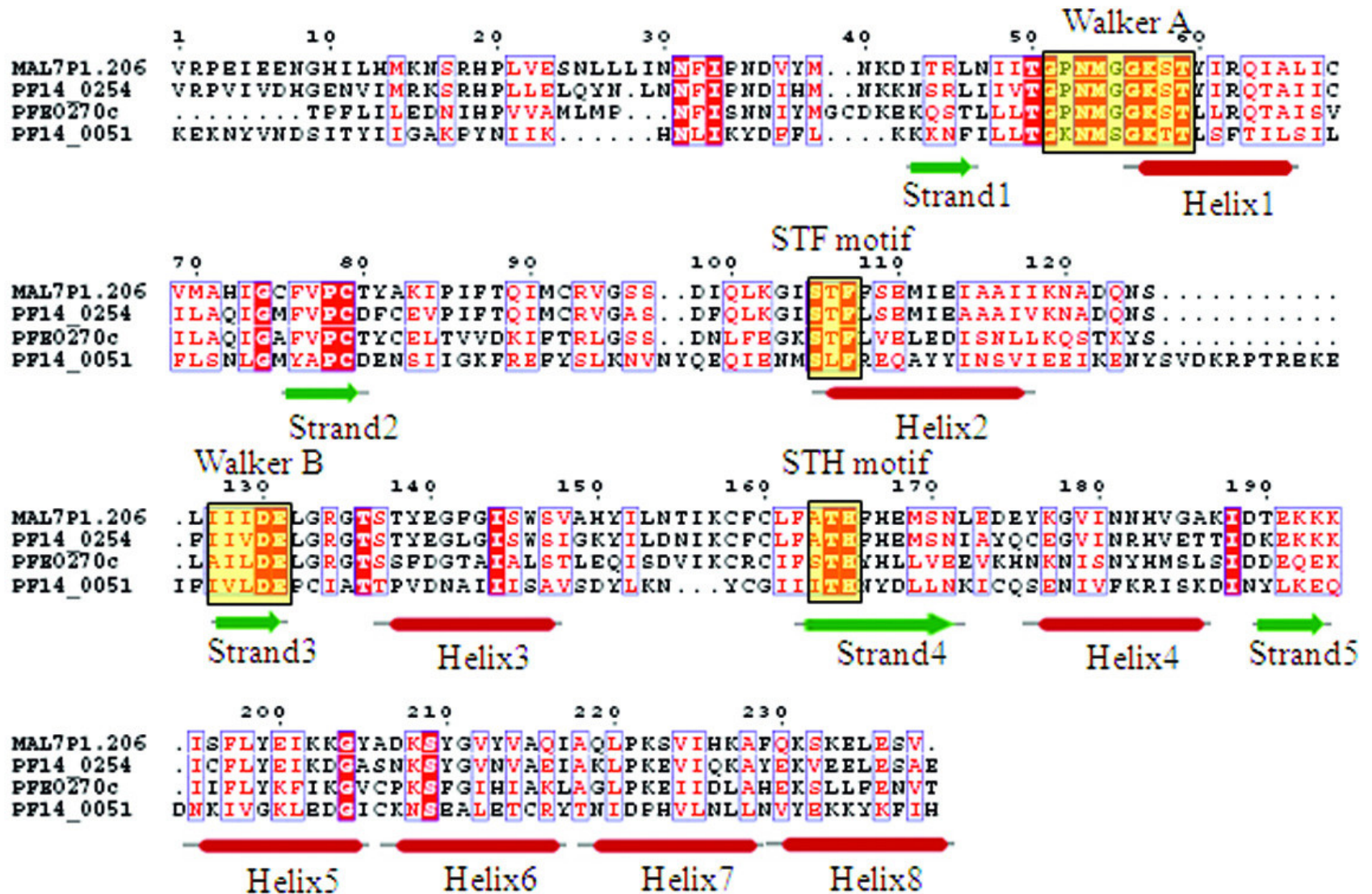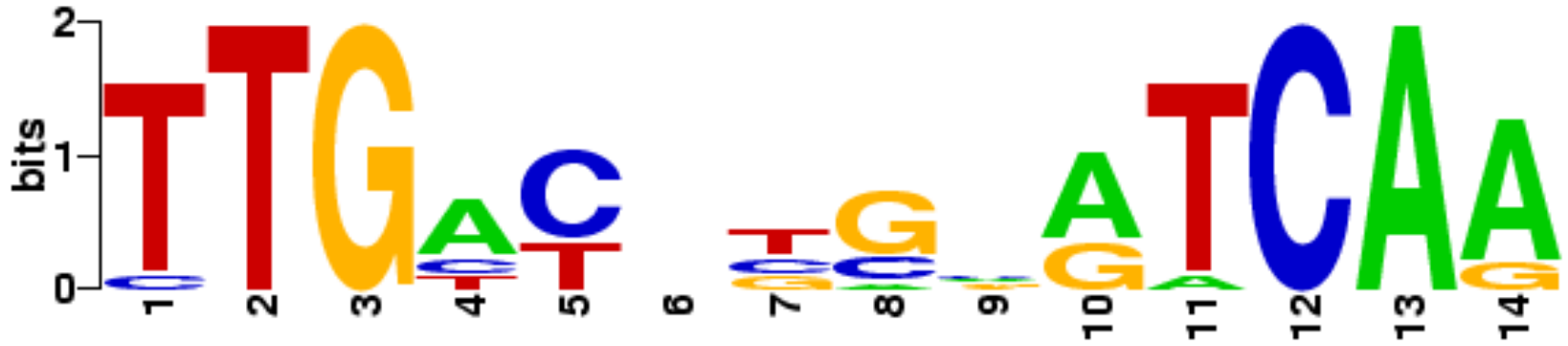
Gene Network

Circos plot

What is bioinformatics?

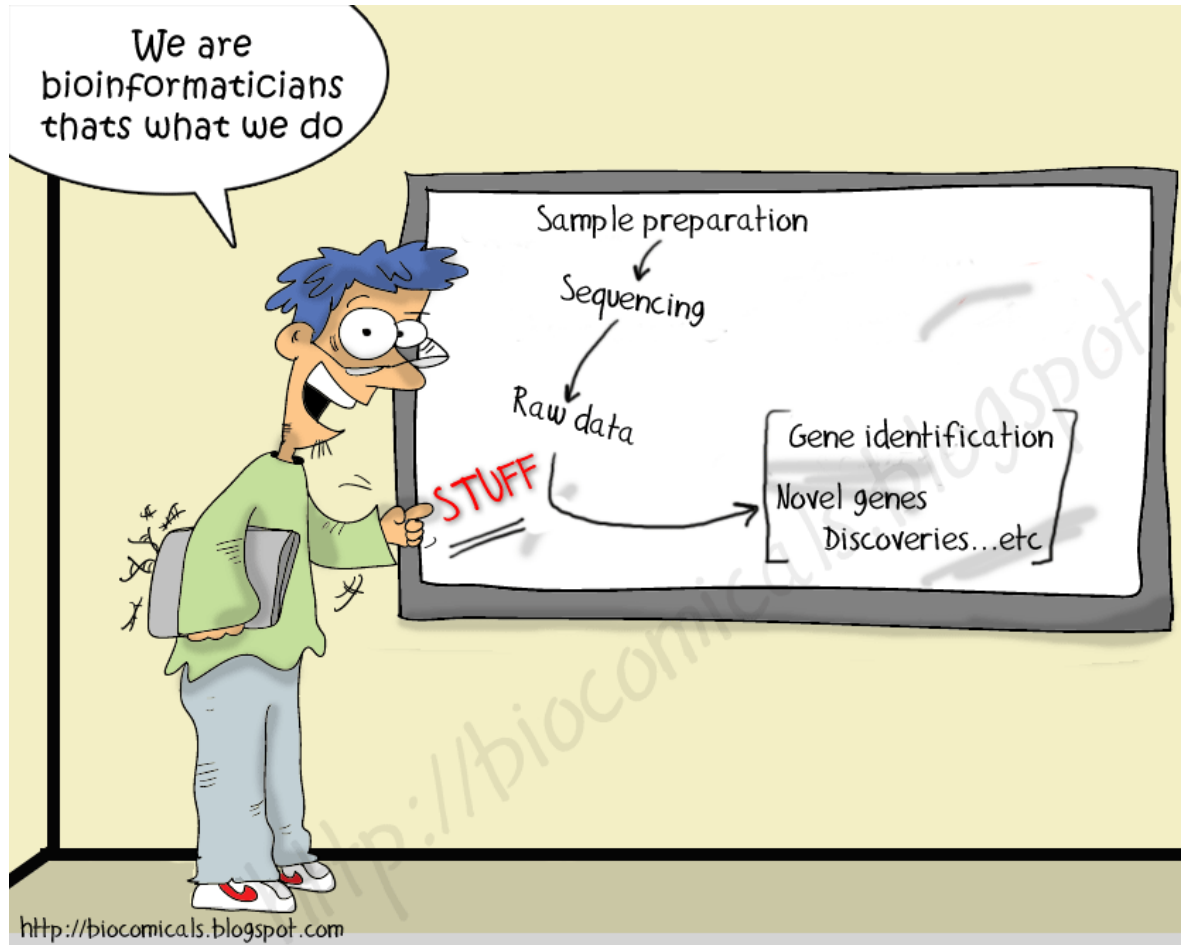- Develop computational tools
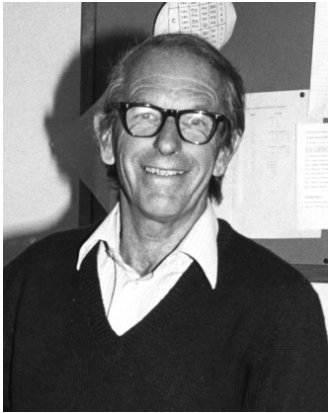- Use computational tools

Three themes in molecular biology

- DNA sequence determines protein sequence

- Protein sequence determines protein structure

- Protein structure determines protein function

  The study of how to represent, store, search, retrieve and analyze DNA/RNA/protein data

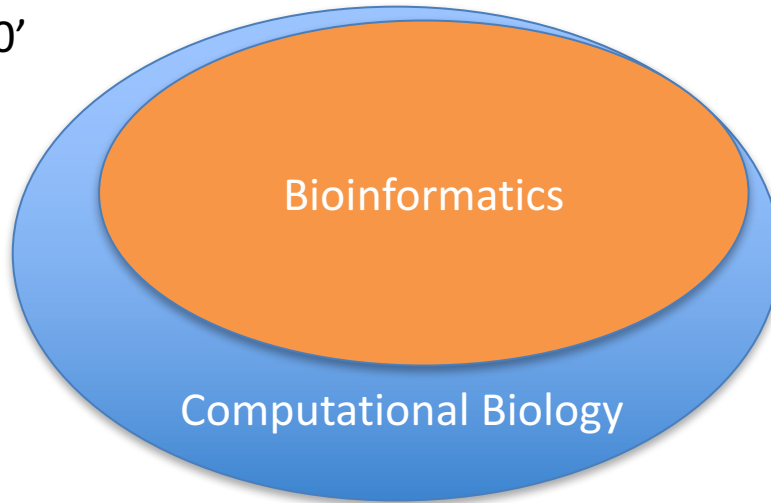# What are most bioinformaticians doing?

Bioinformatics pioneers

The study of how to represent, store, search, retrieve and analyze biological data

Sanger   50'

Dayhoff   60'

Bioinformatics

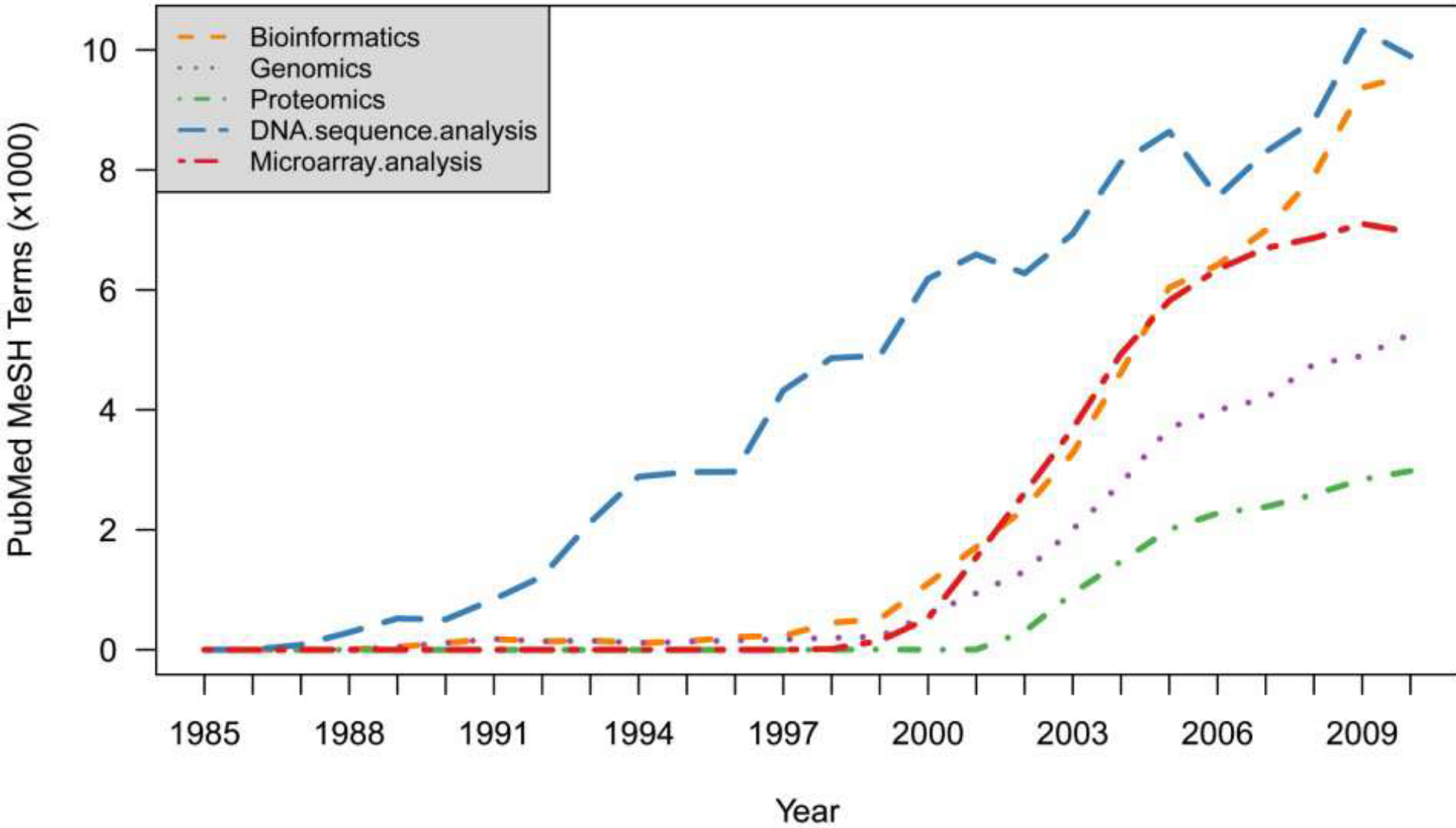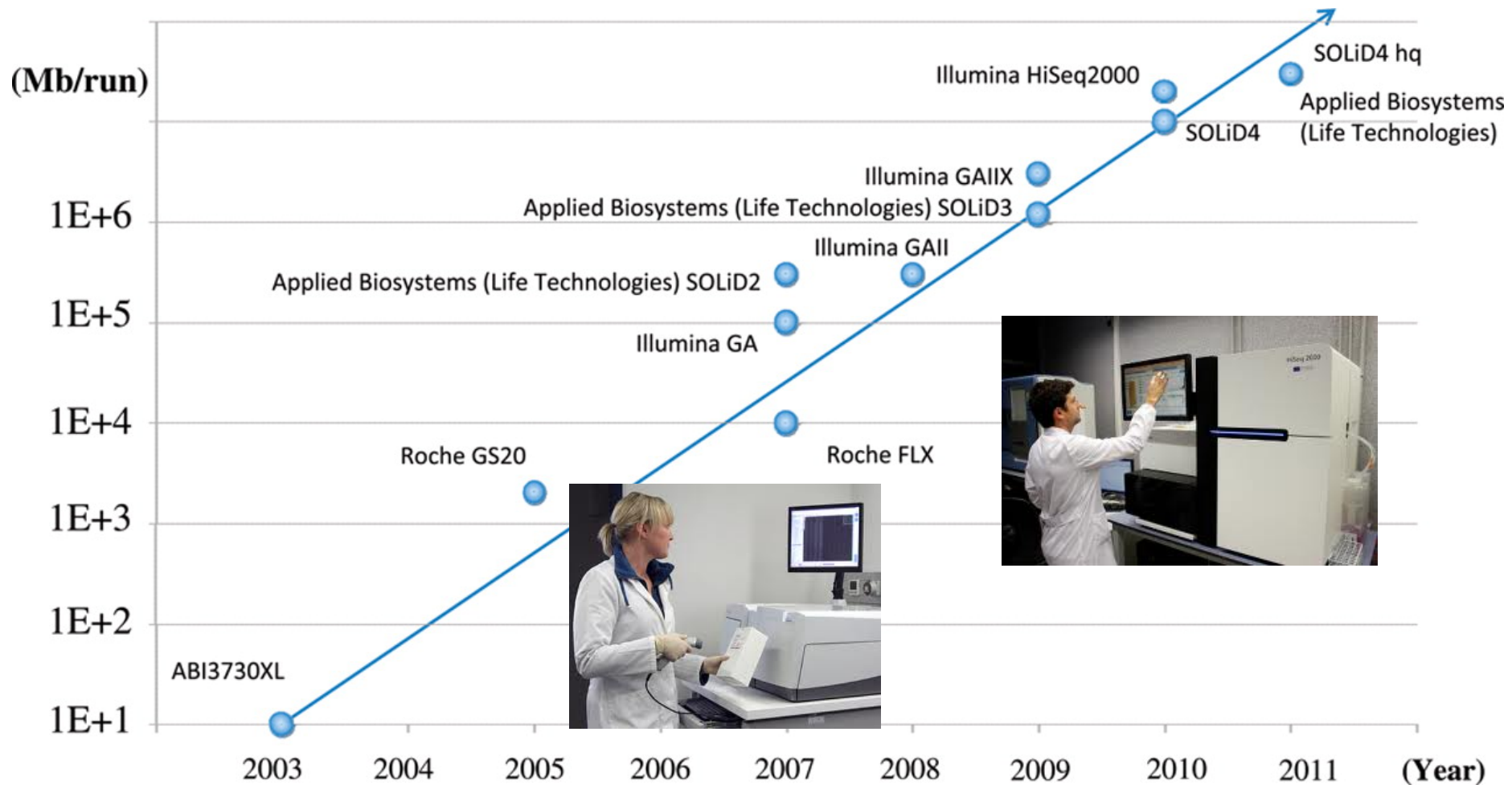Computational Biology

Smith and Waterman   80'

90'

Altschul   Lipman

PubMed MeSH Terms Evolution

Driven by big (sequence) data

Tsuji S, 2010, Hum. Mol. Genet

https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/
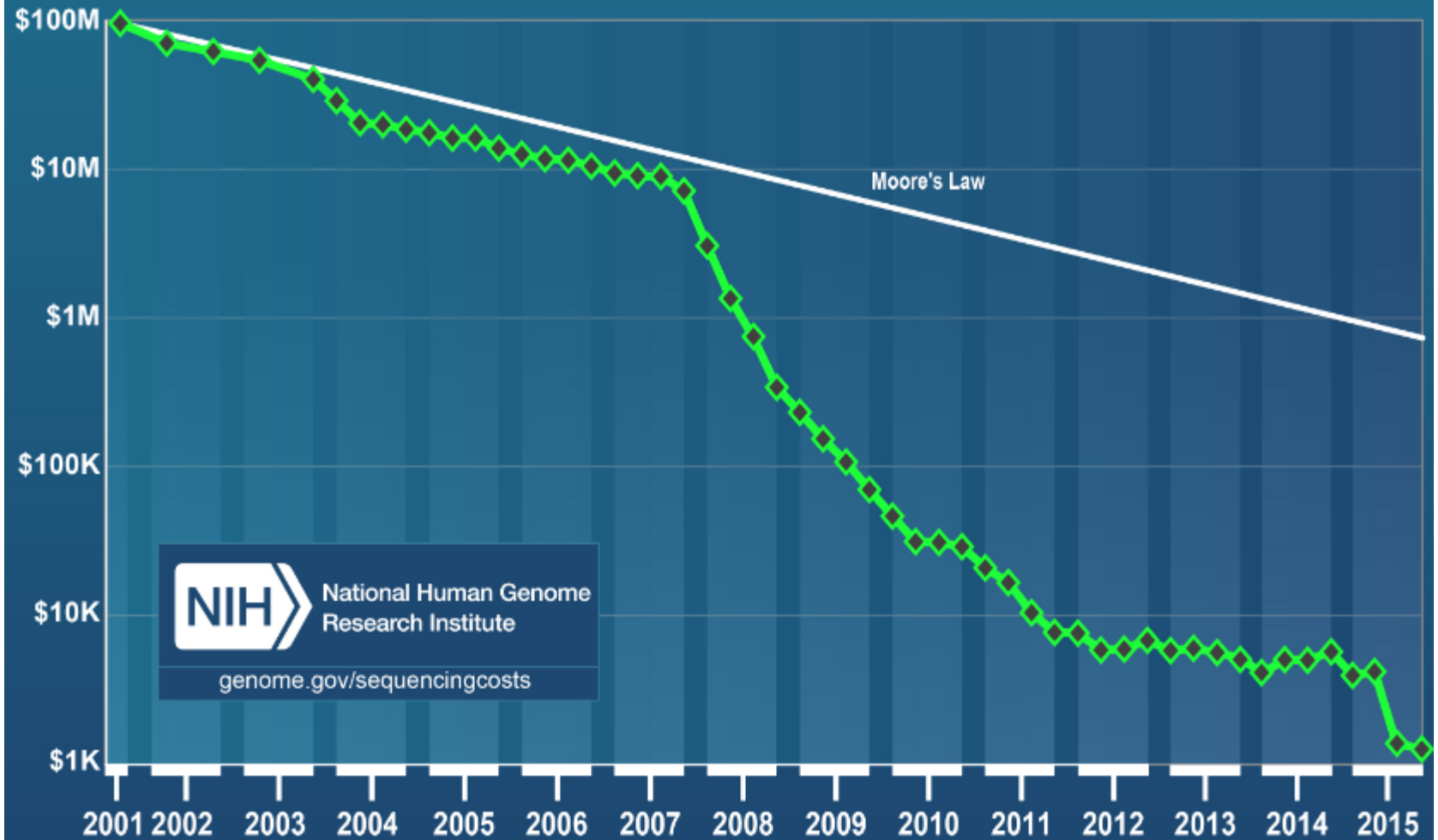
Human Genome Sequencing

https://www.genome.gov/images/illustrations/sequencing.pdf

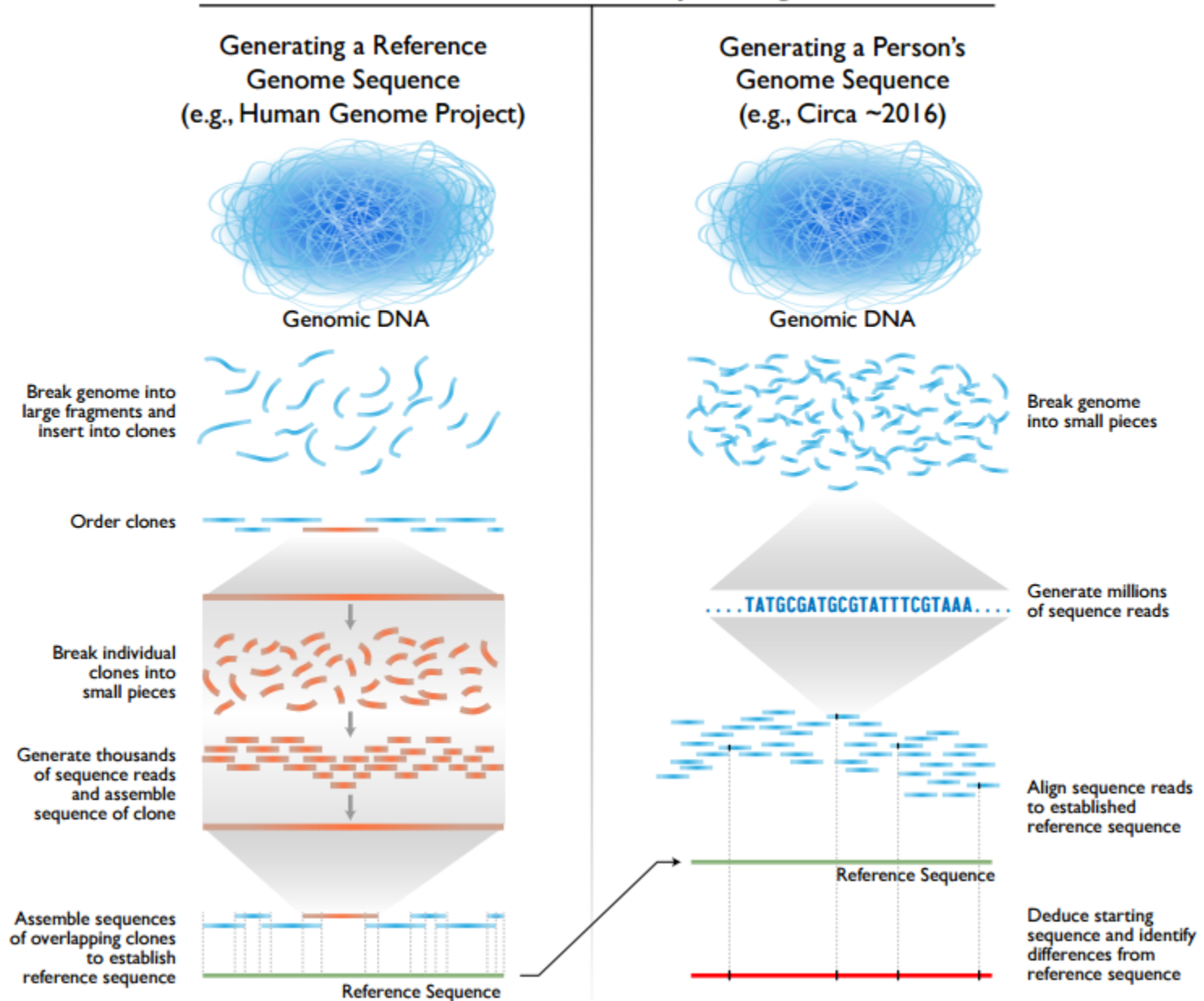Stein's current projects include Reactome,[3] WormBase,[4] BioPerl,[5] Gramene,[6] ENCODE,[7][8][9] the Generic Model Organism Database,[10] the Sequence Ontology[11] and Cloud computing.[12] In 2004, Lincoln was the winner of the Benjamin Franklin Award in Bioinformatics.[13]

Stein is also the original developer of CGI.pm and a contributor to mod_perl, both widely used in the Perl language programming for web applications, as well as many other Perl modules and associated books.[14][15][16][17][18]

Opinion

Bioinformatics: alive and kicking

Lincoln D Stein*†

Addresses: *Ontario Institute for Cancer Research, Toronto, ON, M5G 0A3, Canada. †Cold Spring Harbor Laboratory, NY 11724, USA.
Email: lincoln.stein@gmail.com

Bioinformaticians: gone by 2012. Bioinformatics: stronger than ever.

Abstract

Bioinformatics has become too central to biology to be left to specialist bioinformaticians. Biologists are all bioinformaticians now.

Bioinformaticians: gone by 2012.
Bioinformatics: stronger than ever.

In February 2003 I gave a keynote address for the second annual O'Reilly Bioinformatics Technology Conference called 'Bioinformatics: Gone in 2012' in which I predicted that bioinformatics as a discipline separate from mainstream biology would be gone in ten years.

In 2008 as it happens, my predictions were quite wrong.

My prediction was that bioinformatics would become one of a series of core courses taught in undergraduate and graduate biology programs, and that there would be a vanishing market for researchers who focus solely on biological data management.

Example

The graduate programs in bioinformatics are within the Institute of Bioinformatics at University of Georgia, which offer two different focuses of study:

(1) Bioinformatics applications designed for students with background and strengths in the biological sciences; and

(2) Bioinformatics Methods Development is designed for students with background and strengths in the computer science, mathematics, and statistics.

Bioinformaticians: gone by 2012. Bioinformatics: stronger than ever.

So bioinformatics isn't disappearing. But who is giving these bioinformatics talks, and making and analyzing these large databases? By and large these are not people who call themselves bioinformaticians. Instead, we are witnessing the rise of a new generation of computational biologists who spend part of their time at the bench and part of their time at the computer.

"If I were a senior or first-year graduate student interested in biology, I would migrate as fast as I could into the field of computational biology." -- Francis Collins

On July 8, 2009, President Barack Obama nominated Collins as Director of the National Institutes of Health

Next lecture: basic molecular biology and introduction to bioinformatics web resources