PLP 6905 INTRODUCTION TO APPLIED BIOINFORMATICS IN PLANT PATHOLOGY

(INTRO APPLIED BIOINFO, PLP6905 section 1771)

Jose C. Huguet-Tapia
Fifield 1501

Lecture: Wed 3-3:50 pm (Fifield 2564)
Lab: Wed 4:05 – 4:55 pm (Fifield 2564)

Quick notes and links on twitter: #PLP6905

Introduction to bioinformatics

Bioinformatics or Computational Biology?

- Bioinformatics is a field that develops and improves methods for storing, retrieving, organizing and analyzing biological data
- Computational Biology is the study of biology using computational techniques and tools

Russ B. Altman
Stanford University

- **"When I build a method (usually as software, and with my staff, students, post-docs–I never unfortunately do it myself anymore), I am engaging in an engineering activity: I design it to have certain performance characteristics, I build it using best engineering practices, I validate that it performs as I intended, and I create it to solve not just a single problem, but a class of similar problems that all should be solvable with the software. I then write papers about the method, and these are engineering papers. This is bioinformatics."**

- **"When I use my method (or those of others) to answer a biological question, I am doing science. I am learning new biology. The criteria for success has little to do with the computational tools that I use, and is all about whether the new biology is true and has been validated appropriately and to the standards of evidence expected among the biological community. The papers that result report new biological knowledge and are science papers. This is computational biology".**

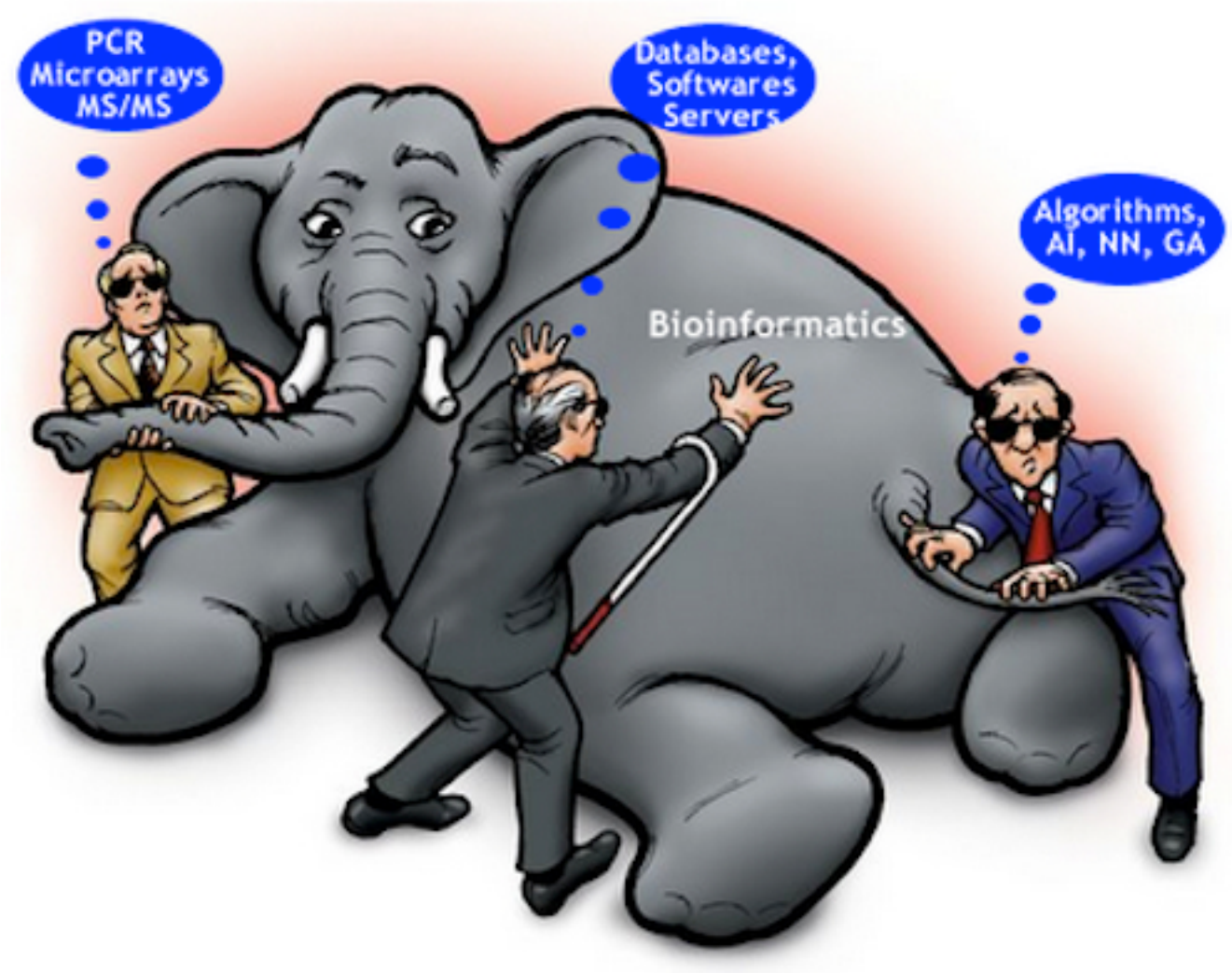
Bioinformatics or Computational Biology?

- Bioinformatics is a field that develops and improves methods for storing, retrieving, organizing and analyzing biological data
- Computational Biology is the study of biology using computational techniques and tools
- In this course we will learn how to organize, store and present biological data (bioinformatics) and use the output to study biological examples (computational biology)

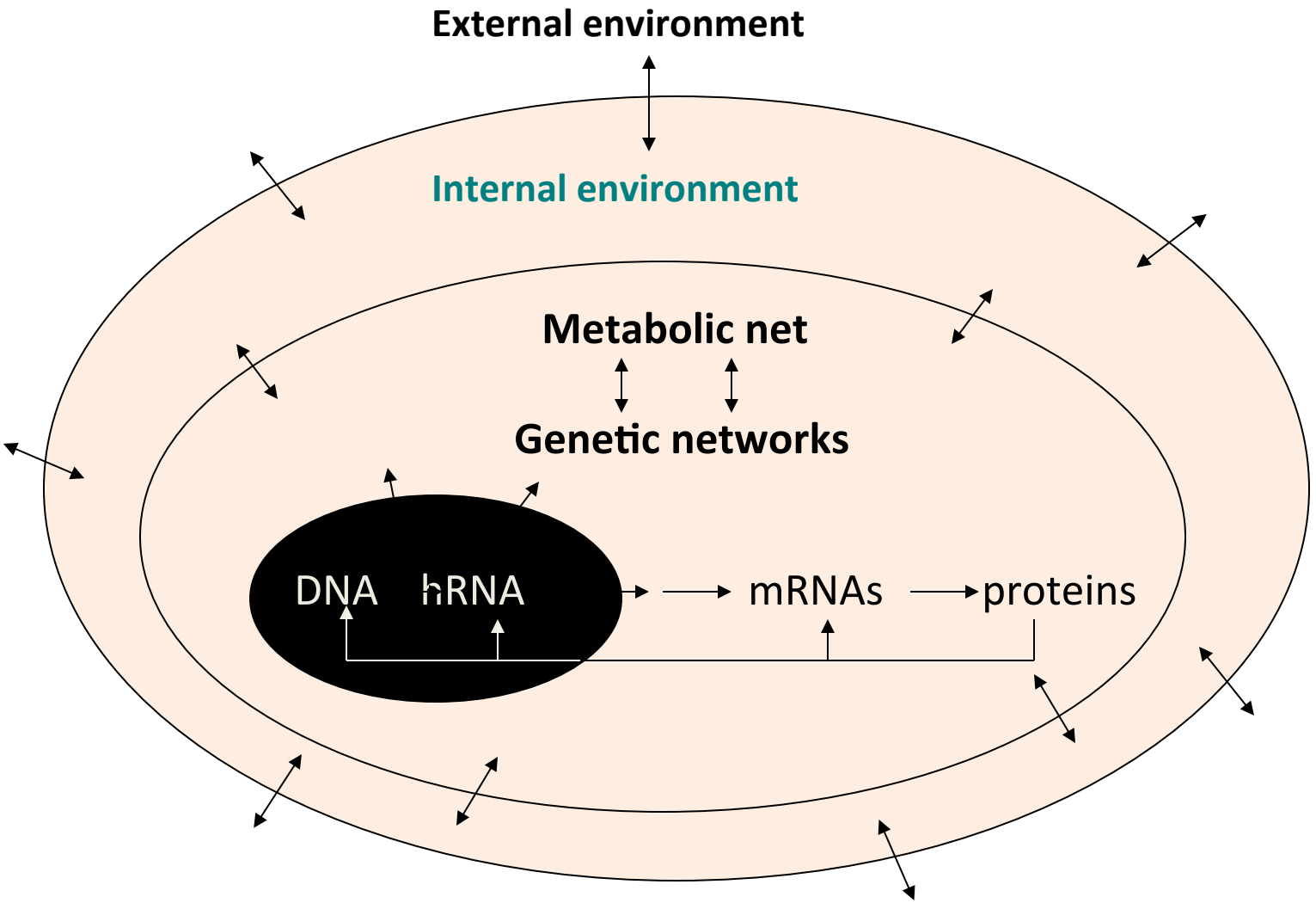
General Plan for PLP6905

Week 1: Wednesday Jan-4th
Introduction to bioinformatics to linux
Week 2: Wednesday Jan-11th
Introduction to linux, basic commands and text manipulation
Week 3: Wednesday Jan-18th
Sequence analysis and databases search
Blast and HMMER
Week 4: Wednesday Jan-25th
Genome data analysis 1: Introduction of high-throughput sequencing data analysis.
Fastq manipulation
Week 5: Wednesday Feb-1st
Genome data analysis 2: Mapping
Short read aligners: Bowtie
Week 6: Wednesday Feb-8th
Genome data analysis 3: de novo Assembly
De novo assemblers: Velvet and Spades
Week 7: Wednesday Feb-15th
Genome data analysis 4: Gene prediction and annotation
RAST and introduction to Artemis
Week 8: Wednesday Feb-22th
Genome data analysis 5: Post assembly and genome alignments
Mummer, *in silico* DNA hybridization ANI
Week 6: Wednesday Mar-1th
Class project presentations

Bioinformatics and data

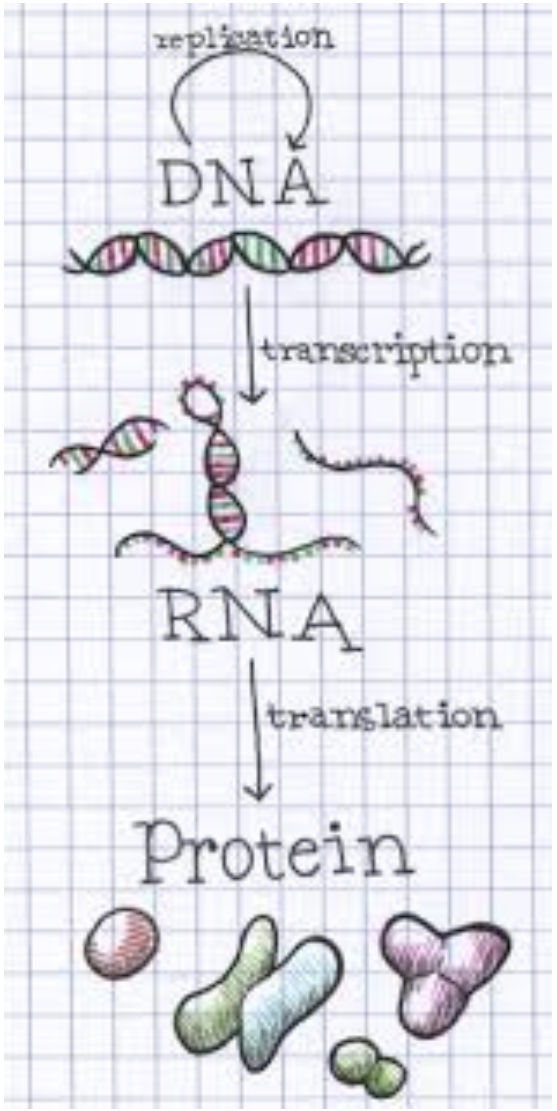


We do not know yet whether the information in the genome is sufficient to reconstruct an entire biological system. Information on building blocks not enough, information on their interactions is essential.



Basic concepts

Central dogma



DNA sequencing, genome assembly, gene prediction and annotation, comparative genomics and phylogenetics

RNA sequencing, gene expression and transcriptome assembly, global transcription and regulation gene prediction and annotation

Gene expression and cell response, protein structure, function discoery

Data types

primary data	**sequence**
AATGCGTATAGGC	**DNA**
DMPVERILEALAVE	**amino acid**
secondary data	**secondary protein structure**
“motifs”: regular expressions, blocks, profiles, fingerprints	**e. g., alpha-helices, beta-strands**
tertiary data	**tertiary protein structure**
atomic co-ordinates	**domains, folding units**

Dealing with data

Software designers and programmers

Expert analysts

Expert users

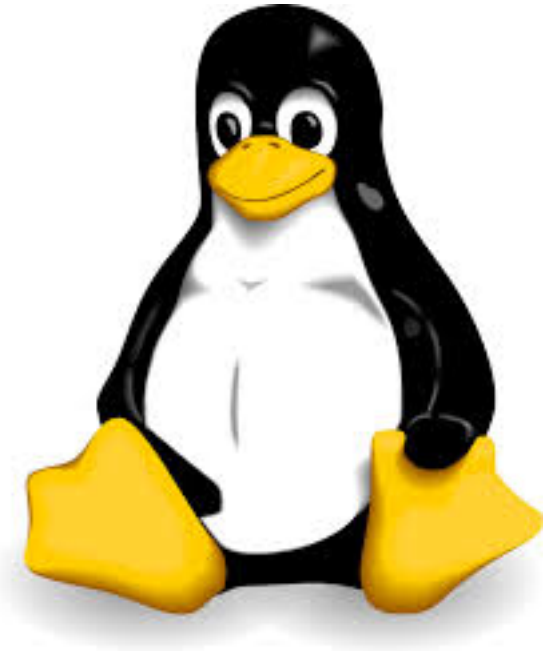
Users

UNIX/LINUX

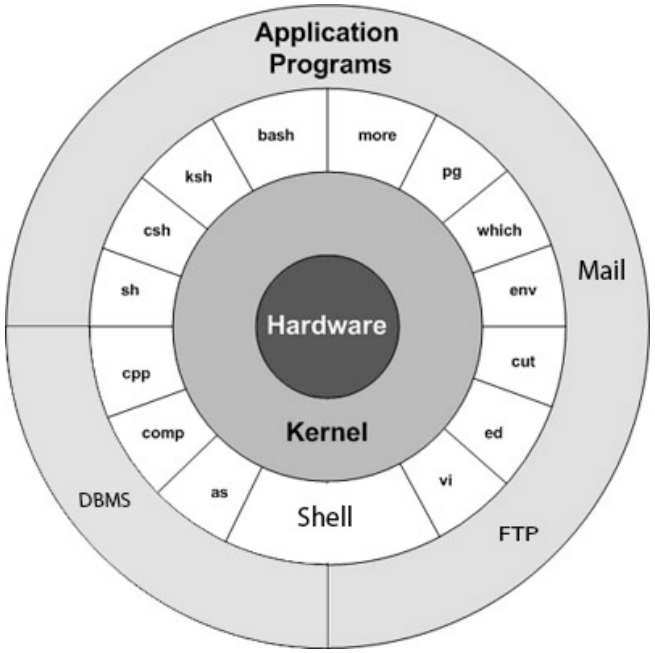
- UNIX is the operating system of choice for engineering and scientific computing.

Differences between Unix and Linux

- Unix was developed in the late 1960s and Linux in the early 1990s based on Unix-like system MINIX
- Linux is a UNIX clone

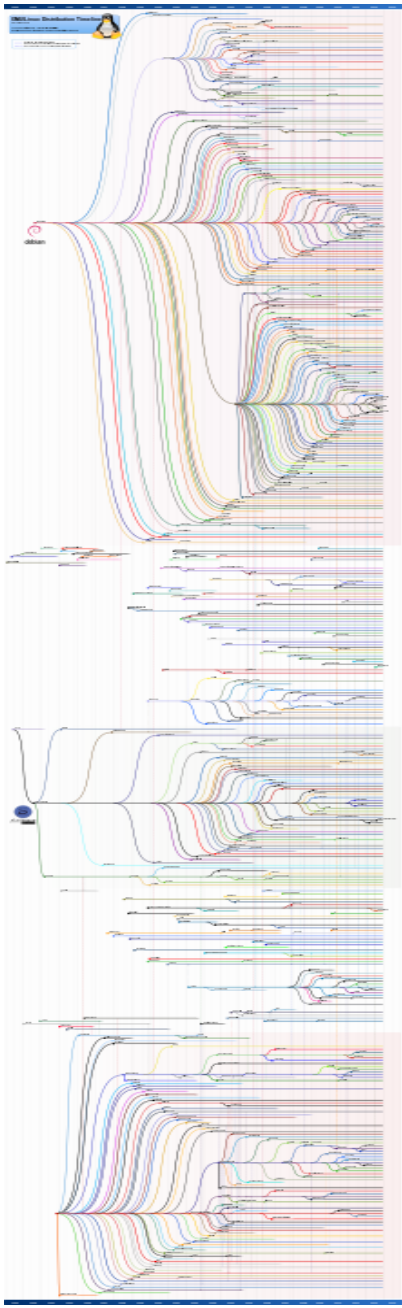


Linux Architecture



- **Kernel:** The heart of the operating system
  - ✓ It interacts with hardware.
  - ✓ Memory management, task scheduling and file management.
- **Shell:** The utility that processes your requests.
  - ✓ the shell interprets the command and calls the program that you want.
- **Commands and Utilities:**
  - ✓ Eg: **cp, mv, cat** and **grep** etc.
- **Files and Directories:**
  - ✓ All data in LINUX is organized into files.
  - ✓ All files are organized into directories.
  - ✓ These directories are organized into a tree-like structure called the filesystem.

Linux has many distributions



http://en.wikipedia.org/wiki/List_of_Linux_distributions

http://en.wikipedia.org/wiki/Comparison_of_Linux_distributions

Biolinux

Biolinux is not a distribution, it is version of linux/Ubuntu with more than 500 bioinformatic programs installed

NERC
Environmental
Bioinformatics
Centre

http://nebc.nerc.ac.uk/tools/bio-linux/