

# Metagenomics needs computers!

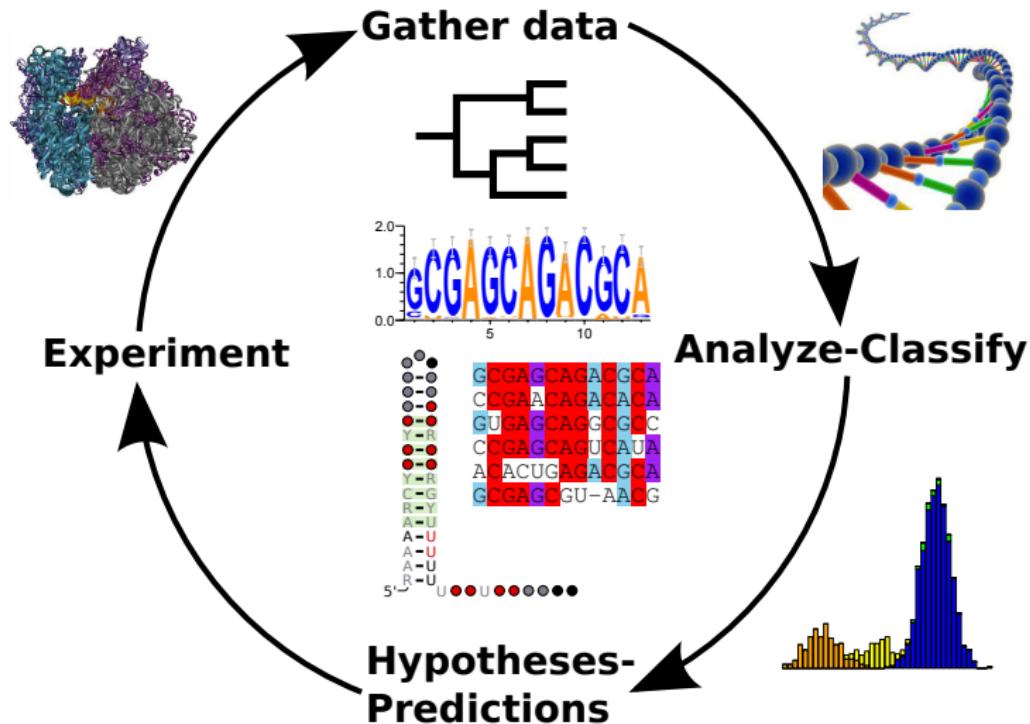
Paul Gardner

September 28, 2013

# What is computational biology?

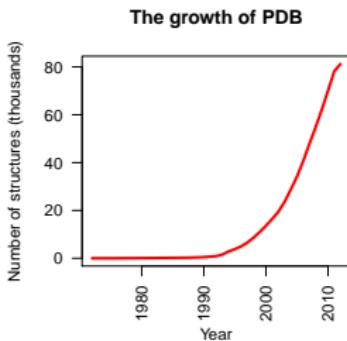
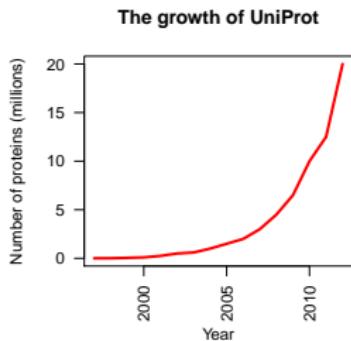
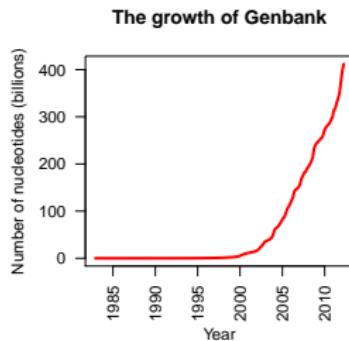


# How did the Geeks get so important?



# The data deluge and the rise of computational biology

- ▶ There is a deluge of data being generated by new techniques in sequencing and structure determination
- ▶ Computational tools are the only way to store and analyse the amount of data that now drives a lot of biological discoveries



# We need giant computers



# We need giant computers



- ▶ UC Biology cluster currently runs a 152 fast CPU cluster
- ▶ Access to BlueGene computer facility,  $\approx$  1,000 CPUs

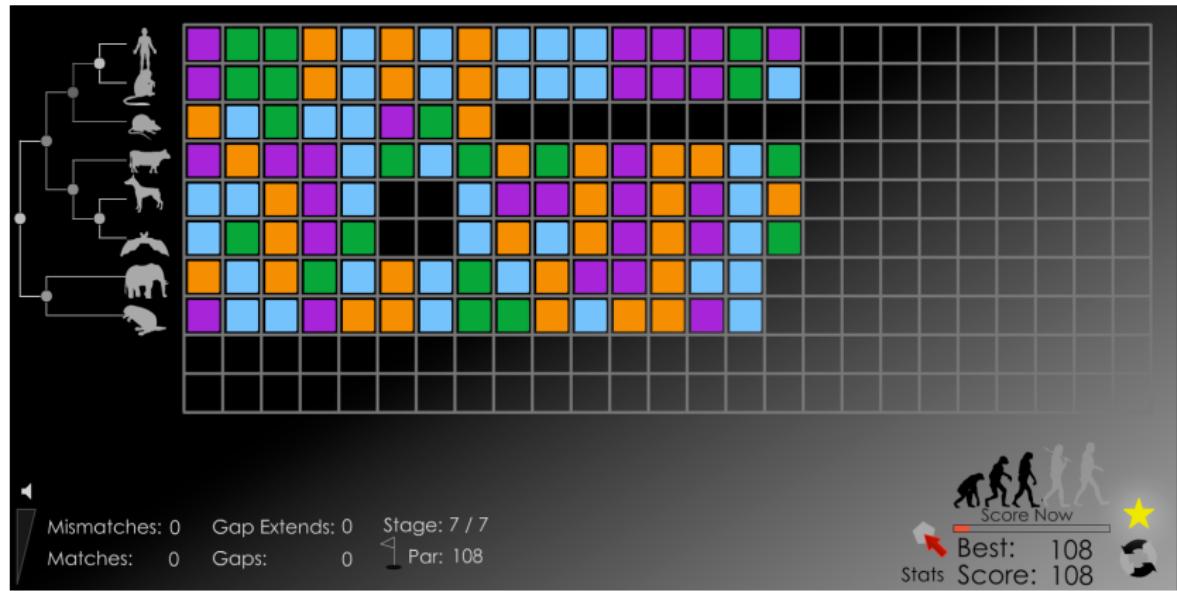
# New ways to build “computers” ...

- ▶ Who likes gaming?



# Games with a purpose

- ▶ Games with purpose
  - ▶ “We spend 3 billion hours a week as a planet playing videogames.” – Jane McGonigal
- ▶ Phylo, EteRNA, Fold.it

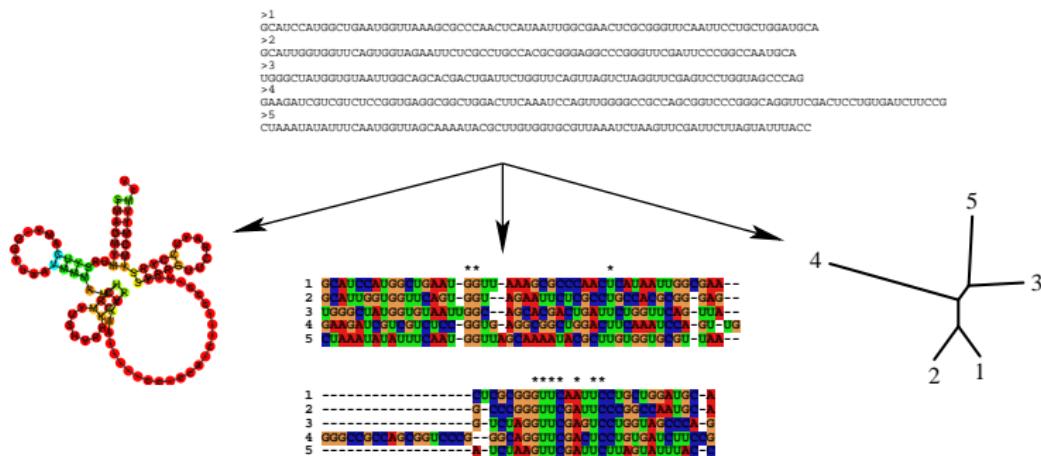


# Where can I find biological data?

- ▶ National Center for Biotechnology Information
  - ▶ PubMed
  - ▶ Genbank
  - ▶ Gene Expression Omnibus
  - ▶ dbSNP
  - ▶ RefSeq
  - ▶ Entrez Gene
  - ▶ OMIM
  - ▶ Taxonomy
- ▶ European Bioinformatics Institute
  - ▶ EMBL/ENA
  - ▶ ENSEMBL
  - ▶ InterPro
  - ▶ Reactome
  - ▶ UniProt
  - ▶ PRIDE
  - ▶ ArrayExpress
  - ▶ Gene Ontology
- ▶ The Ribosomal DB Project
- ▶ KEGG
- ▶ Pfam and Rfam
- ▶ UCSC genome browsers
- ▶ PDB
- ▶ GIYF and Wikipedia

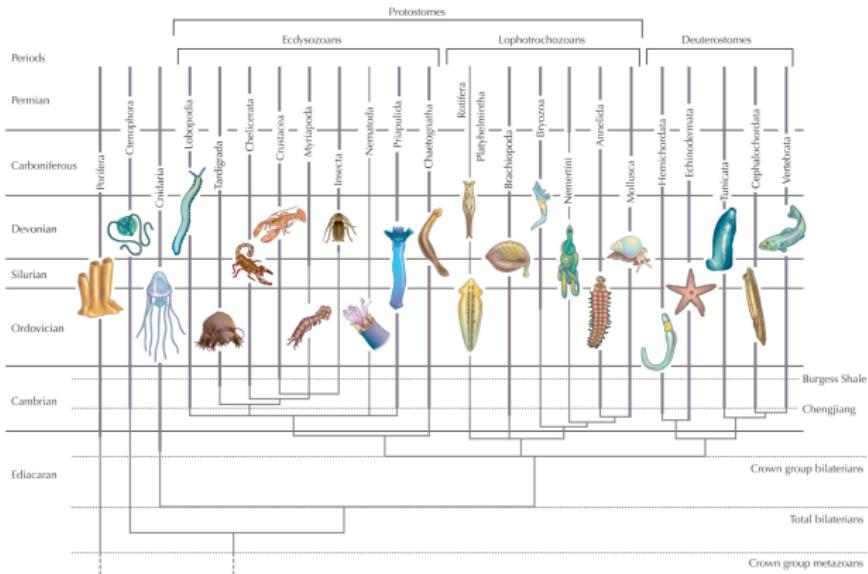
# What can computational biology do for me?

- ▶ Sequence alignment
- ▶ Phylogenetic analysis
- ▶ Analysis of gene expression
- ▶ Infer regulatory networks
- ▶ RNA and Protein structure prediction
- ▶ Gene prediction and annotation



# How can we measure evolutionary relatedness?

- ▶ Is a sponge more related to a jellyfish or the custard apple?
- ▶ How can we test this?



**FIGURE 10.16.** The fossil record and metazoan phylogeny. Dark lines represent the temporal range of phyla from their first appearance in the fossil record to the present. Extrapolation into the Ediacaran is based on molecular clock data. Some relationships, particularly among the arthropods, remain controversial (see Fig. 10.17).

10.16, adapted from original drawing by Susan Butts, based on design by Matthew Wills

Evolution © 2007 Cold Spring Harbor Laboratory Press

# The homology search problem

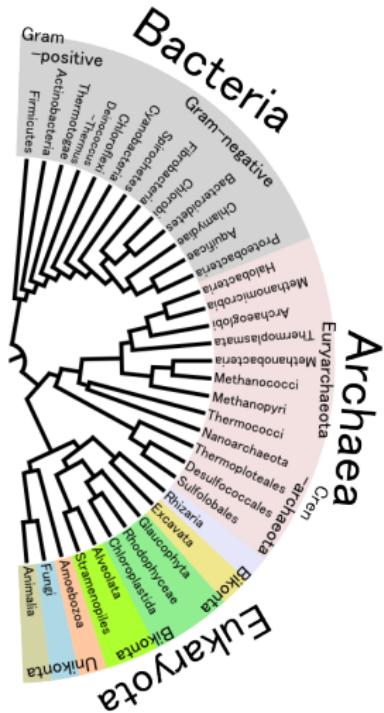


Image source: [www.wikimedia.org](http://www.wikimedia.org)

- ▶ How can we compare DNA sequences?

<i>H. pylori</i>	-	GVD	VALH	PPKQPF	GAANH	IIEGG	SITIIATALL	TIGGEMDDEV	-
<i>D. radiodurans</i>	-	GVD	TALLYPPKQPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>M. tuberculosis</i>	-	GVD	TALLYPPKQPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>C. pneumoniae</i>	-	GVD	SALIIPPPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>F. nucleatum</i>	-	GVD	SALIIPPPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>S. enterica</i>	-	GVD	SALIIPPPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>B. thetaiotaomicron</i>	-	GVD	SALIIPPPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>L. interrogans</i>	-	GVD	SALIIPPPF	GAANH	IIEGG	SITIIATAMV	TIGGEMDDEV	-	
<i>P. marinus</i>	-	GVDP	GYD	GYD	GYD	GYD	GYD	GYD	
<i>U. parvum</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>B. subtilis</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>C. difficile</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>S. griseus</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>F. nodosum</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>M. infernorum</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>T. yellowstonii</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>E. coli</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	
<i>H. volcanii</i>	-	GVD	GYD	GYD	GYD	GYD	GYD	GYD	

# Metagenomics

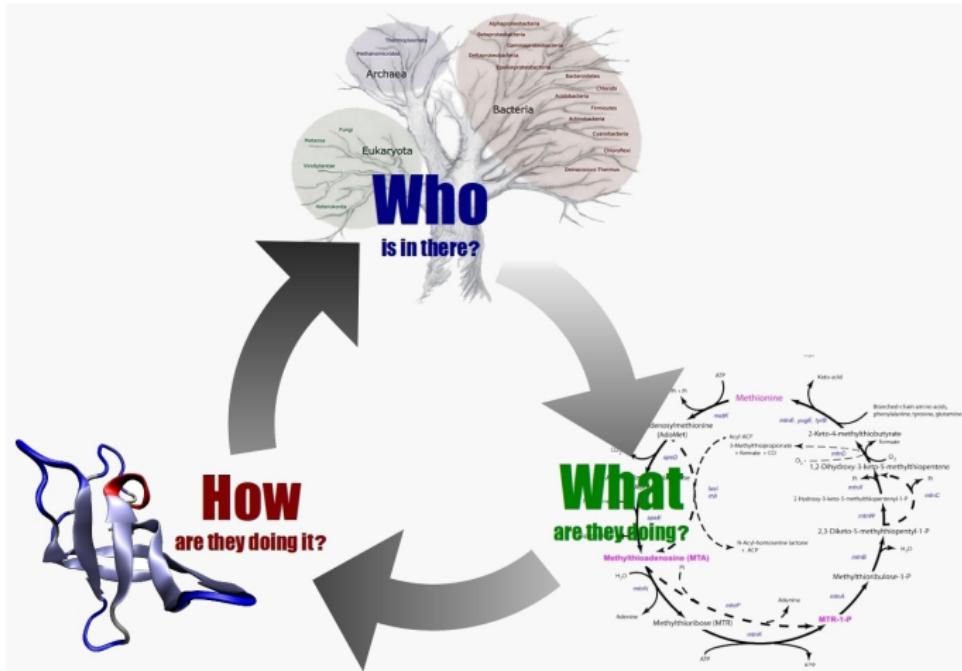


Image source: Center for Biological Sequence Analysis at the Technical University of Denmark

# Metagenomics

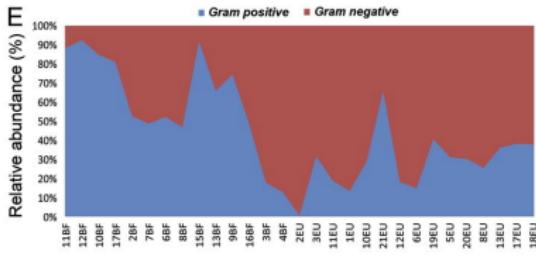
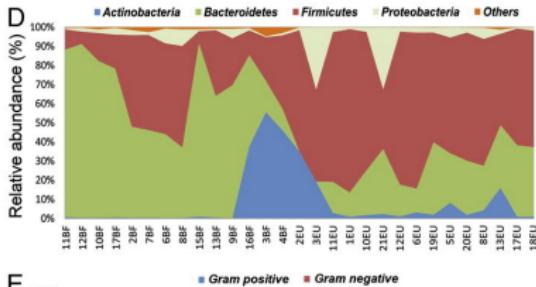
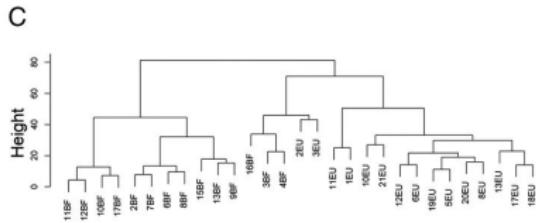
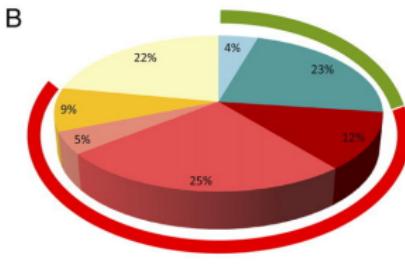
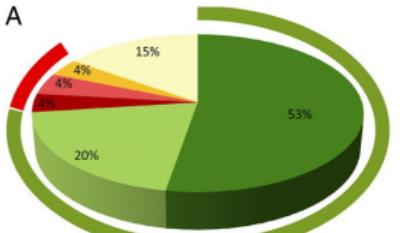


Image source: Filippo et al. (2010) PNAS.

# Computer lab

- ▶ biolc100 the password is saturday28 (additional usernames: biolc101-biolc131)
- ▶  $-36^{\circ} 53' 9''$ ,  $175^{\circ} 49' 21''$  Hot Water Beach
- ▶  $-36^{\circ} 47' 37.66''$ ,  $175^{\circ} 52' 37.82''$  Mutton bird burrow
- ▶  $-36^{\circ} 52' 29''$ ,  $174^{\circ} 46' 26''$  Epsom Girls High



