

Tree of Life

Inferring species phylogenies from entire proteomes

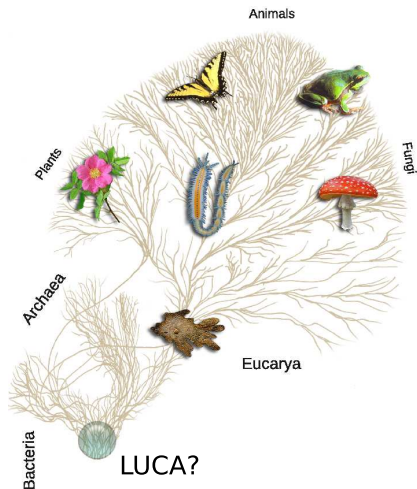
Arnold Kuźniar and Hannes Schabauer

Department of Ecology and Evolution

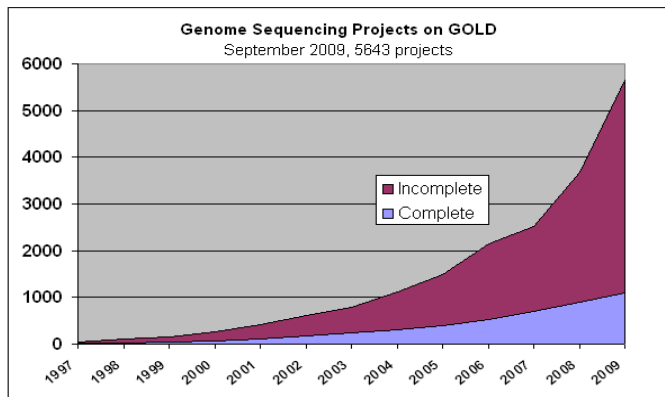
February 25, 2011



Darwin's big idea...



Evolutionary genomics (phylogenomics)



(<http://genomesonline.org>)

Bioinformatics infrastructure for *in silico* analyses



Figure: “Milky Way”, world’s fastest supercomputer [Xinhuanet, 2010]

Project description

- **Problem:** How do species relate to each other in evolution?

Project description

- **Problem:** How do species relate to each other in evolution?
- **Data set:** Seven fully sequenced plant genomes

Project description

- **Problem:** How do species relate to each other in evolution?
- **Data set:** Seven fully sequenced plant genomes
- **Goals:**
 - Infer species tree(s) using entire proteomes
 - Compare the results to single-gene and “supertree” approaches

Project description

- **Problem:** How do species relate to each other in evolution?
- **Data set:** Seven fully sequenced plant genomes
- **Goals:**
 - Infer species tree(s) using entire proteomes
 - Compare the results to single-gene and “supertree” approaches
- **What can you learn?**
 - To get familiar with domain-specific terminology
 - To use some bioinformatics tools
 - To write simple programs/scripts
 - To improve your math and writing skills

Methods

- Single-gene based phylogeny

Methods

- Single-gene based phylogeny
- Whole proteome-based phylogeny
 - Pair-wise proteome comparisons
 - Orthology clustering
 - Cluster-based phyletic patterns or profiles
 - Compare profiles (need some math)

Methods

- Single-gene based phylogeny
- Whole proteome-based phylogeny
 - Pair-wise proteome comparisons
 - Orthology clustering
 - Cluster-based phyletic patterns or profiles
 - Compare profiles (need some math)
- Compare phylogenetic trees (need some math)

Methods

- Single-gene based phylogeny
- Whole proteome-based phylogeny
 - Pair-wise proteome comparisons
 - Orthology clustering
 - Cluster-based phyletic patterns or profiles
 - Compare profiles (need some math)
- Compare phylogenetic trees (need some math)
- Linux/MacOSx commands

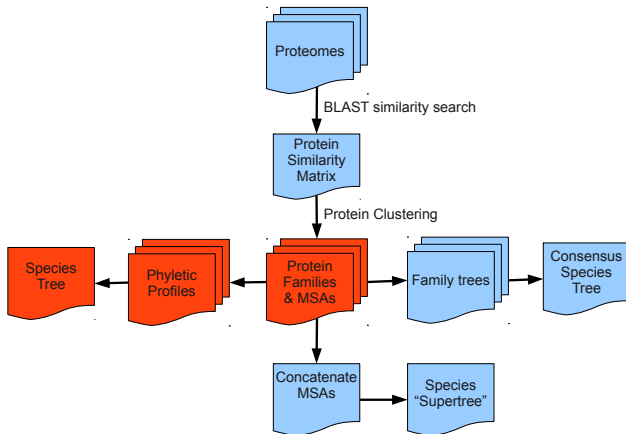
Methods

- Single-gene based phylogeny
- Whole proteome-based phylogeny
 - Pair-wise proteome comparisons
 - Orthology clustering
 - Cluster-based phyletic patterns or profiles
 - Compare profiles (need some math)
- Compare phylogenetic trees (need some math)
- Linux/MacOSx commands
- Programming in Perl, R or Octave

Example formulas

- Manhattan distance $d(p, q) = \sum_{i=1}^n |p_i - q_i|$
- Euclidean distance $d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$
- Minkovski distance $d(p, q) = \left(\sum_{i=1}^n |p_i - q_i|^p\right)^{\frac{1}{p}}$

Work flow summarized



Deliverables

- Intermediate and final report

Deliverables

- Intermediate and final report
- Final presentation

Deliverables

- Intermediate and final report
- Final presentation
- Documented program source codes/scripts