Genomic Supertrees of Life

Davide Pisani
Bioinformatics Laboratory
The National University of Ireland Maynooth
We are in the genomic era

Haemophilus influenzae

- 460 Published Complete Genomes
- 1054 Ongoing Prokaryotic Genome Projects
- 631 Ongoing Eukaryotic Genome Projects
- 460 genomes are already available and 2145 will be available within the next few years

Data From: http://www.genomesonline.org/
Challenges to the tree of life

“...the history of life cannot properly be represented as a tree” Doolittle W.F “...but not everybody agrees”


31 Genes-based tree of life


This is the tree of 1%!

Dagan & Martin (2006) Genome Biology
Supertrees

Any method for amalgamating trees on partially overlapping leaf sets is a supertree method.

(Modified from: Semple and Steel, 2000)

A Supertree is “a dendrogram from which each of the original [input] trees can be regarded as a sample”

Gordon (1986)
221 Complete genomes downloaded from COGENT and partitioned in four data sets

Single gene families defined (clusters of putatively orthologous sequences)

Misaligned sites removed (using GBlocks) & PTP test performed

Maximum likelihood analysis of each protein family (using Multiphylog -distributed ML)

Selection of trees for supertree construction

Excluding trees where Eukarya & Cyanobacteria formed a monophyletic group

Excluding ALSO all the trees were the Eukarya & the Alpha - Proteobacteria formed a monophyletic group
The genomic supertrees of life

5741 genes based supertree of life

Some Conclusions

- The phylogeny of life is a weakly connected ring-like network, not a tree.
- There are only two primary lineages of life.
- Eukaryotes originated from the symbiosis of an $\alpha$-proteobacterium with a Thermoplasmatales - like euryarchaeote.

Acknowledgments: James McInerney, James Cotton, Bill Martin & The Marie Curie Fellowships Scheme