

Poster presentation

Open Access

## Phylogenomic inference of functional divergence

Tom A Williams<sup>1</sup>, Brian E Caffrey<sup>1</sup>, Xiaowei Jiang<sup>1</sup>, Christina Toft<sup>1,2</sup>  
and Mario A Fares<sup>1</sup>

Address: <sup>1</sup>Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland and <sup>2</sup>Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

from Fifth International Society for Computational Biology (ISCB) Student Council Symposium  
Stockholm, Sweden 27 June 2009

Published: 19 October 2009

BMC Bioinformatics 2009, 10(Suppl 13):P4 doi: 10.1186/1471-2105-10-S13-P4

This article is available from: <http://www.biomedcentral.com/1471-2105/10/S13/P4>

© 2009 Williams et al; licensee BioMed Central Ltd.

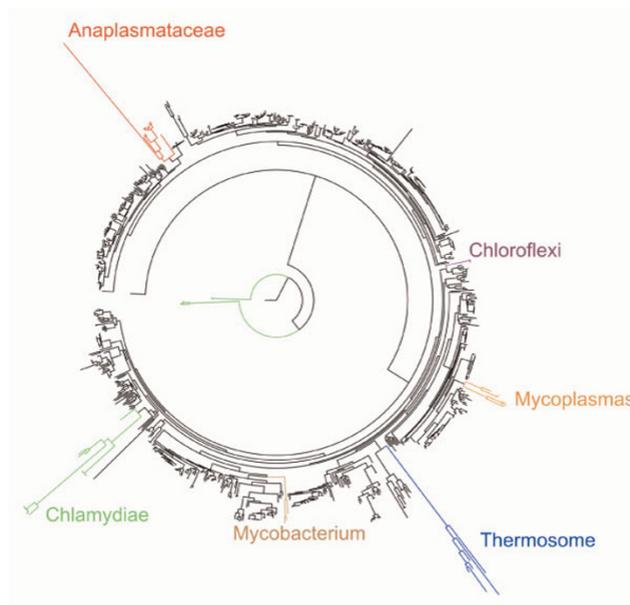
This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Background

The divergence of protein function following gene duplication – or the colonization of new ecological niches – is of central importance in the evolution of novelty. Changes in protein structure and function are reflected at the level of amino acid sequence. This principle suggests that lineage-specific functional divergence in proteins can be identified by the analysis of primary sequence data. However, many amino acid substitutions have a negligible effect on protein function. This means that a simple comparison of the sequence differences between two clusters of proteins will not reveal the subset of changes responsible for functional divergence. While several methods to identify these biologically important substitutions exist [1], they are not optimized for analyses of large numbers of protein sequences. Here, we present a fast new method for identifying these substitutions across a large phylogenetic tree.

### Materials and methods

Our method requires a bifurcating phylogenetic tree and a protein sequence alignment. Each node on the tree is defined by two downstream clades and one or more outgroup sequences. Using BLOSUM [2] scores to quantify how radical or conservative substitutions in each clade are relative to the outgroup, we assign a score to each column of the alignment at each tree node, which is then tested for



**Figure 1**  
**Bacterial lineages enriched for functional divergence in GroEL.** The thermosome-related sequences are found in certain extremophilic bacteria, perhaps as a result of horizontal gene transfer from archaea. The other highlighted lineages are intracellular pathogens, with the exception of Chloroflexi. The tree was produced by RAxML [6].

significance [3]. Here, we apply our method to a tree of the GroEL genes from 622 bacterial genomes.

## Results

GroEL is an important molecular chaperone which helps at least 250 client proteins fold in *Escherichia coli* [4]. Interestingly, we found that four out of the five bacterial lineages most enriched for functional divergence are intracellular pathogens (see Figure 1). Radical change in GroEL has previously been implicated in the adaptation of endosymbiotic bacteria to intracellular life [5], and these results suggest this may be a more general response to the population-genetic conditions of an intracellular lifestyle.

## References

1. Gu X and Velden Vander K: **DIVERGE: Phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics* 2002, **18**:500–501.
2. Henikoff S and Henikoff JG: **Amino acid substitution matrices from protein blocks.** *Proc Natl Acad Sci USA* 1992, **89**:10915–10919.
3. Toft C, Williams TA and Fares MA: **Genome-wide functional divergence after the symbiosis of Proteobacteria with Insects unraveled through a novel computational approach.** *PLoS Comput Biol* 2009, **5**:e1000344.
4. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang H-C, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M and Hartl FU: **Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*.** *Cell* 2005, **122**:209–220.
5. Fares MA, Moya A and Barrio E: **GroEL and the maintenance of bacterial endosymbiosis.** *Trends Genet* 2004, **20**:413–6.
6. Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**:2688–2690.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

