

# Analysis of Bacterial Gene Neighbourhoods

Andre Sim<sup>1</sup>, Thomas Gotwig<sup>2</sup>, Hannah Piepenbreier<sup>1</sup> and Georg Fritz<sup>1</sup>

<sup>1</sup>LOEWE-Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität, Marburg, Germany

<sup>2</sup>THM University of Applied Sciences, Gießen, Germany

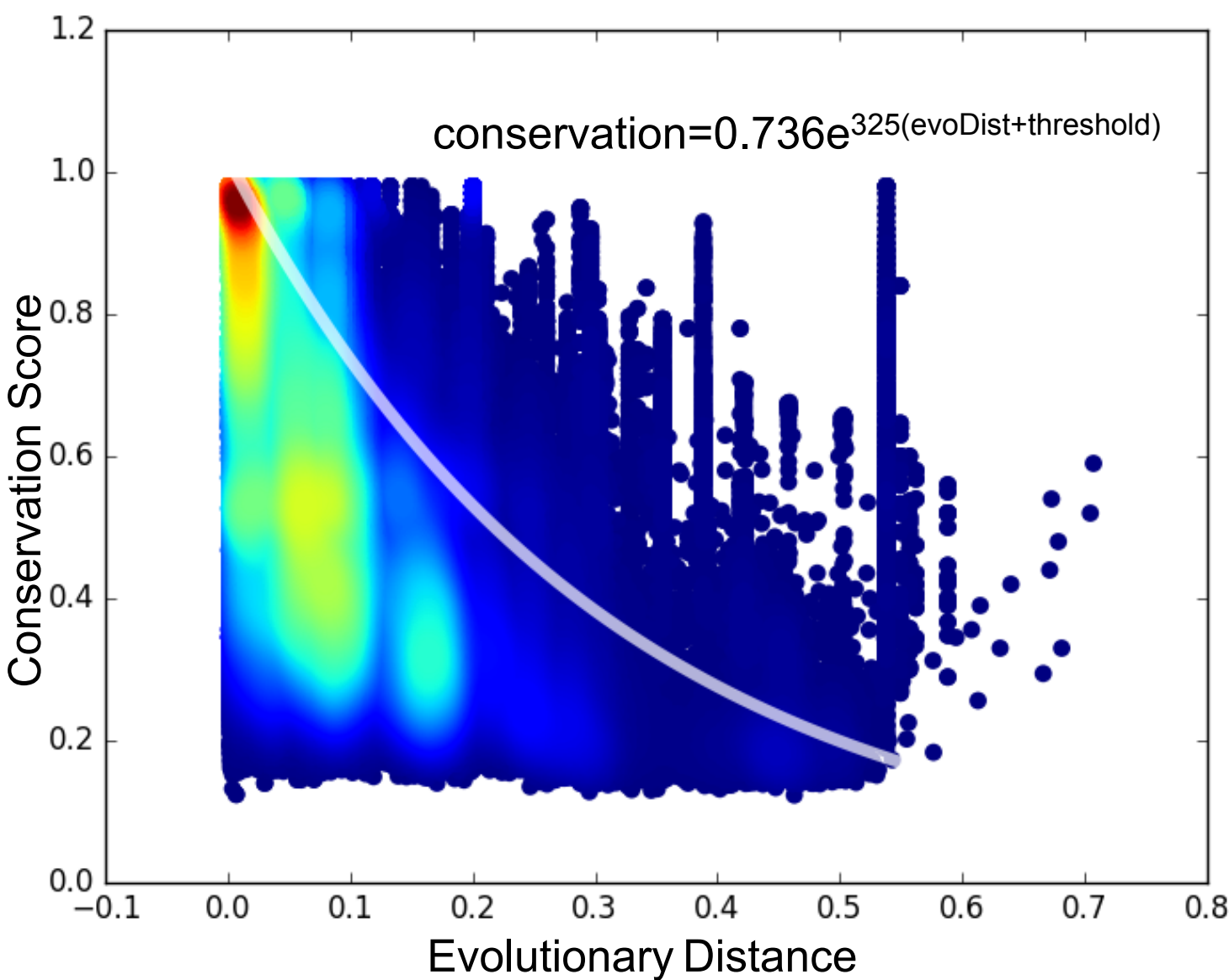
## Summary

Gene neighbourhoods describe groups of genes which are found in close proximity to each other across multiple different genomes. In prokaryotic organisms this is especially interesting as their genome is tightly compacted, and has a highly fluid gene order which is poorly conserved. To put the lack of conservation in perspective, gene order is conserved at a significantly lesser extent than protein amino acid sequences<sup>1</sup>. With the advance in high-throughput sequencing, many more genomes have been made available making research in this area more accessible. Therefore we took this opportunity to look at the evolution of gene neighbourhoods throughout the bacterial taxonomy at a quantitative level. This analysis revealed the pressure on maintaining and forming gene neighbourhoods varies across the bacterial kingdom. There was also a tendency for clusters to be maintained near either the origin or terminus of the genome. In order to facilitate future research in this area, we present the web application GECCO. With GECCO, users can ask quantitative driven research questions in order to drive hypothesis generation, find candidate interaction partners, predict possible gene function and much more.

**Figure 1** – A cartoon representation gene neighbourhood

Each line represents a different genome, arrows represent genes and colours represent ortholog groups (gene categories)

## Methods

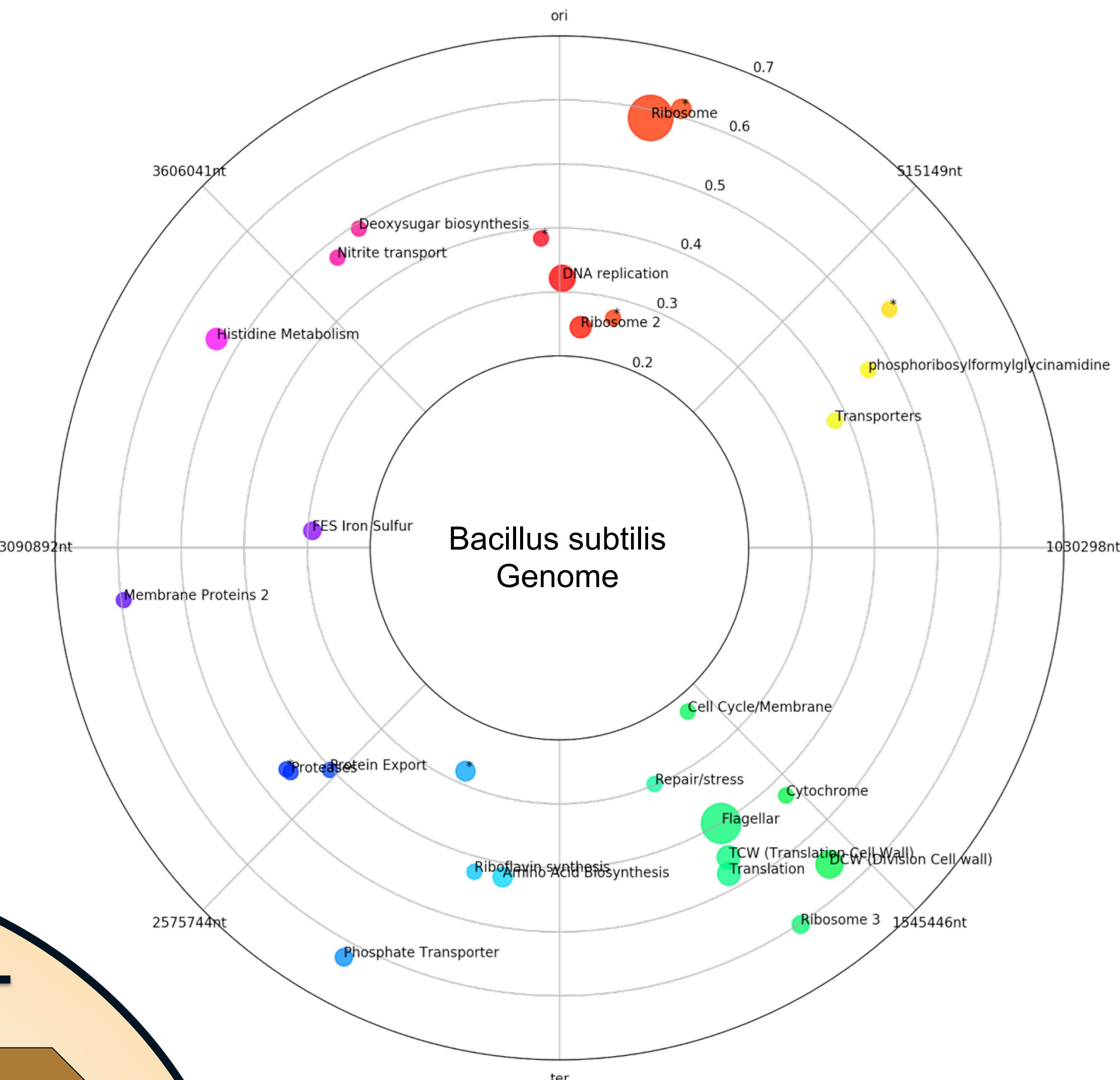


**Figure 2** – Neighbourhood conservation rapidly decreases with evolutionary distance

The regression line represents the used conservation significance threshold  
Evolutionary distance is the average patristic distance between all pairwise combinations of all included genomes from a 16S RNA built tree  
Conservation score represents the average conservation of the top 50 most conserved genes in a neighbourhood

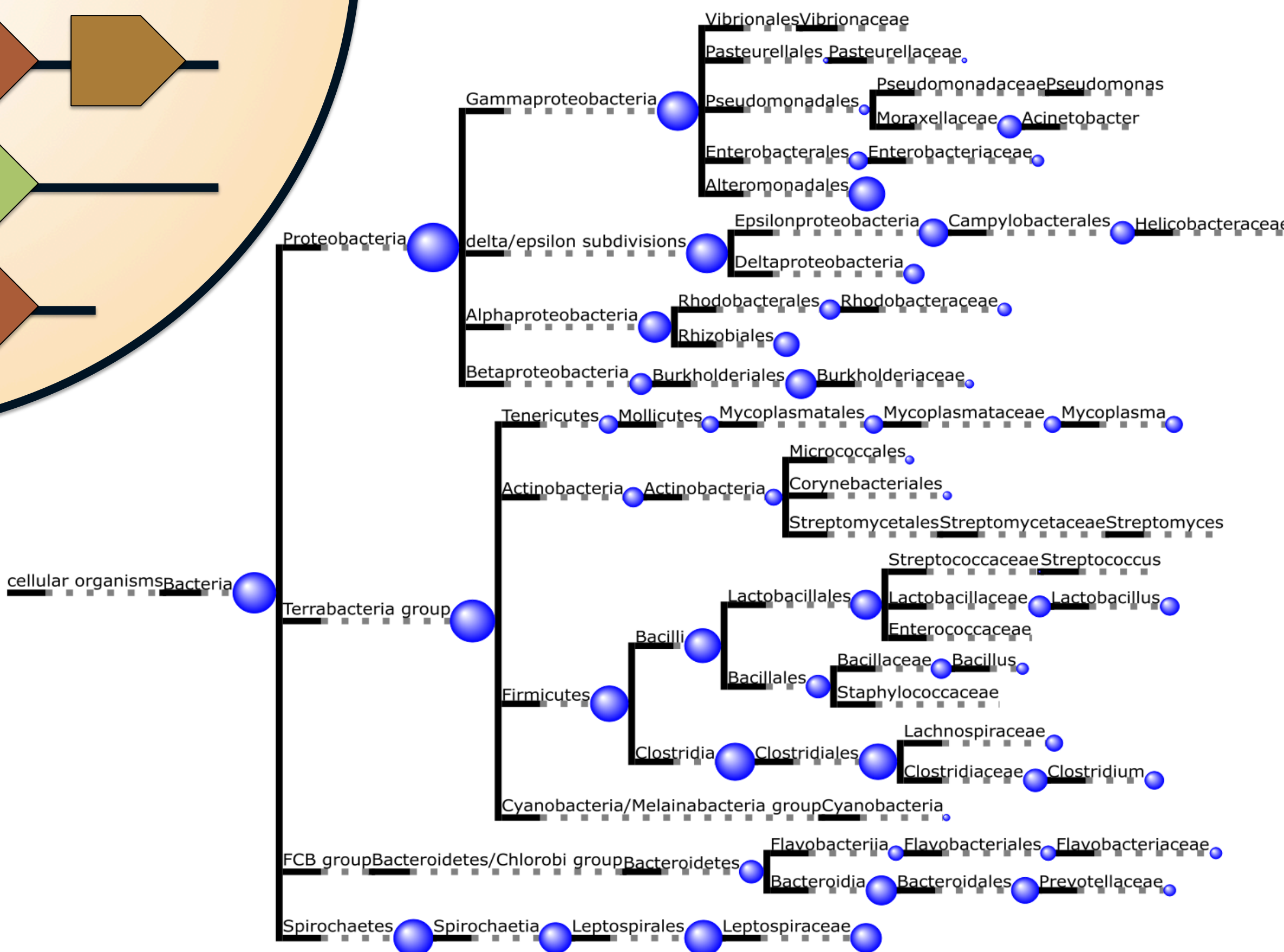
- Genes were grouped by the ortholog mapping of OrthoDB<sup>2</sup>
- For every ortholog group, the genomic neighbourhood surrounding each member was summed together (neighbourhood size = 50)
- Neighbourhoods were thresholded based on the evolutionary distance of the included genomes (i.e at different taxa)
- Clustering was done by finding isolated overlapping genomic neighbourhoods

## Results



**Figure 3** - Identified several conserved clusters

- Clusters were enriched near the origin of replication or the terminus
- Size of nodes represents the size of the cluster in genes
- Y-axis represents the conservation strength of the cluster across all genomes
- \* = cluster with unclear functions



**Figure 4** - Gene clustering occurs at varying frequencies across the bacterial kingdom

Spirochaetes and Proteobacteria had many phylum specific gene clusters  
Size of the nodes represents the number of clusters detected for this taxonomic group

## GECCO – GEnomic Context COservation Database

GECCO



Set the sensitivity of conservation

Filter by phylum

Search by:  
Ortholog Group ID  
Uniprot ID  
Refseq ID  
Genome  
Gene Name

See both positional and neighbourhood conservation

Display the conservation of biological processes in the neighbourhood

See how the neighbourhood evolves in different taxa

Export figures and data for your own research!

Gene	%
Cysteine Desulfurase	63.8%
Ferredoxin	61.8%
NIF system FeS cluster assembly	61.7%
Chaperone Protein HscA	61.5%
Iron-sulfur cluster assembly accessory protein	60.6%
Transcription regulator Rrf2-type	55.1%
FeS assembly protein IscX	38.2%
RNA methyltransferase	31.1%
Inositol monophosphatase	28.2%

**Table 1** – Conserved gene neighbours of co-chaperone HscB  
% - the neighbourhood conservation of the gene

## Outlook

Observe the expression behaviour of conserved gene neighbourhoods as well as wet lab experiments of interesting candidate neighbourhoods in the attempt to decipher the evolutionary forces keeping them together.

## References

1. Wolf, Y. I., Rogozin, I. B., Kondrashov, A. S. & Koonin, E. V. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res.* 11, 356–372 (2001).
2. Kriventseva, E. V. et al. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 43, D250–6 (2015).