41. Exploring the evolution and the diversity of bacterial genes

Zara Ghazoui

zg105@york.ac.uk
University of York, UK.

Alpha proteobacteria species share orthologous proteins that include mostly housekeeping (basic) and some accessory genes. Basic genes are strongly over represented on the main chromosome, but accessory genes are mostly on secondary chromosomes or plasmids. Most orthologs, especially on the chromosome, support the ribosomal RNA phylogeny, but an interesting minority support other trees.

Although the 16S rRNA phylogeny has become a universal paradigm in bacteria, we believe it is critical to use complete genome sequences to investigate the evolutionary history of all genes. So our aim is to firstly assess the extent of horizontal gene transfer and build more accurate picture of bacterial evolutionary relatedness. Secondly, to distinguish between the basic and accessory genes in bacteria.

To address these issues, we chose the group of alpha proteobacteria because it contains species with a wide variety of niches and lifestyles, including plants symbionts and pathogens (Agrobacterium, Sinorhizobium, Bradyrhizobium), obligate intracellular (Rickettsia) and facultative intracellular (Bartonella, Brucella). Moreover the alpha proteobacteria species have an amazing genome size variation, from 1 Mb in Rickettsia spp. to 9 Mb in Bradyrhizobium japonicum. Furthermore, some alpha proteobacteria are very unusual because of the presence of multiple replicons within the same bacterial strain (e.g. the Agrobacterium tumefaciens linear chromosome over 1Mb) and/or one or more large plasmids (100 kb - 1 Mb, e.g. the Agrobacterium tumefaciens plasmid AT).

In this study, we have compared eight completely sequenced alpha proteobacteria genomes: Agrobacterium tumefaciens, Mesorhizobium loti, Sinorhizobium meliloti, Bradyrhizobium japonicum, Rhodopseudomonas palustris, Brucella melitensis and Caulobacter crescentus. Sets of putative orthologs were identified as reciprocal best BlastP hits in pairwise genome comparisons. From these, quartets of orthologous proteins (QuartOPs) were assembled by taking four genomes at a time in various combinations. For each QuartOP, the relative support for each of the three possible unrooted 4-taxon trees was estimated by a Bayesian method using MrBayes2. The phylogenetic support for every QuartOP was then visualised by mapping the three probabilities corresponding to the three possible topologies onto an equilateral triangle1. The QuartOPs were further explored to identify the set of universal genes between all of the selected genomes. The sequences of these genes were used to construct their multigene phylogeny and their evolution was further investigated.

Our results show that alpha proteobacterial genomes share a set of core genes that are mostly on the chromosomes. The majority of these genes support a consensus phylogeny, that is congruent with the 16s rDNA phylogeny. However a significant number of genes, mainly on the plasmids, support other phylogenies. Moreover, the common chromosomal genes that support the 16s rDNA phylogeny, are consistently organised along the genomes, whereas the plasmid genes show a random distribution.

The striking differences revealed in this study between the evolution of chromosomal and the plasmid

genes support the notion that alpha proteobacterial genomes contain two different gene pools, basic and accessory DNA. On the one hand, the chromsomes carry shared set of genes with consistent phylogeny and organisation. On the other hand, the majority of the plasmid genes often absent from related species and when present they are randomly distributed along the genomes, and often have phylogenies that imply a history of horizontal gene transfer.

References

1. Strimmer K, von Haeseler A (1997). “Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment” Proc Natl Acad Sci U S A 94(13):6815-9.

2. Zhaxybayeva O, G. J. (2002). “Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses.” BMC Genomics 3(1).