The α -proteobacteria: the Darwin finches of the bacterial world

Thijs J. G. Ettema and Siv G. E. Andersson*

Department of Molecular Evolution, Evolutionary Biology Center,
752 36 Uppsala, Sweden

*Author for correspondence (siv.andersson@ebc.uu.se).

The α -proteobacteria represent one of the most diverse bacterial subdivisions, displaying extreme variations in lifestyle, geographical distribution and genome size. Species for which genome data are available have been classified into a species tree based on a conserved set of vertically inherited core genes. By mapping the variation in gene content onto the species tree, genomic changes can be associated with adaptations to specific growth niches. Genes for adaptive traits are mostly located in ‘plasticity zones’ in the bacterial genome, which also contain mobile elements and are highly variable across strains. By physically separating genes for information processing from genes involved in interactions with the surrounding environment, the rate of evolutionary change can be substantially enhanced for genes underlying adaptation to new growth habitats, possibly explaining the ecological success of the α -proteobacterial subdivision.

Keywords: adaptation; expansion; reduction; bacterial genomes; genomic islands

1. INTRODUCTION

The *Proteobacteria* represent the largest bacterial group that is currently recognized in the domain Bacteria. The name ‘proteobacteria’ is derived from the Greek God Proteus, which had the ability to change shape, to indicate that the proteobacterial species come in many forms, colours and shapes. However, the diversity of this subdivision is not merely restricted to shape, but extreme variation is also observed in lifestyle, metabolic capacity and ecological significance.

Over the past decade, the genome sequences of nearly 100 α -proteobacterial species have been determined, revealing an astonishing plasticity in genome size and architecture. In this paper, we will reflect on the diversity of the α -proteobacterial genomes and the coherence of the α -proteobacterial lineage in the light of all this genetic flexibility. We will argue that the α -proteobacteria represent an excellent model system for studying bacterial genome evolution because of the ease with which they adapt to new growth habitats. In this respect, the α -proteobacteria represent the bacterial equivalents of the Darwin finches.

One contribution of 11 to a Special Feature on ‘Whole organism perspectives on understanding molecular evolution’.

2. DIVERSITY OF THE α -PROTEOBACTERIA AND ECOLOGICAL RELEVANCE

Bacterial classification schemes are mostly based on the sequence diversity of the rRNA genes because of their universal presence and slow rate of evolution, enabling inferences of deep divergences. Likewise, core genes involved in information processing such as transcription and translation tend to evolve by vertical descent and are frequently used to delineate species relationships. Using the latter approach, a species tree has been inferred for all α -proteobacterial species with a sequenced genome (figure 1; Williams *et al.* 2007).

By mapping the growth habitat of each organism onto the species tree (figure 1), major environmental shifts can be inferred. This is of interest, since the α -proteobacteria are found in all imaginable habitats, ranging from the ocean floor to volcanic environments, in many of which they represent the most abundant bacterial group (Dutilh *et al.* 2008). For example, nitrogen-fixing members of the Rhizobiales are highly abundant in the soil where they interact with plant roots, whereas the SAR11 clade and the Rhodobacterales account for as much as 30–50% of all bacteria in the ocean surface waters (Giovannoni *et al.* 2005). Adaptations to vertebrate and invertebrate hosts have occurred several times independently (Batut *et al.* 2004), with up to 76 per cent of all arthropods and most of filarial nematodes being infected with *Wolbachia* (Bourtzis & Miller 2003). Since the species composition pattern is characteristically different for each environment, we conclude that the ecological abundance of the α -proteobacteria is not simply an effect of ‘everything is everywhere’, but rather a reflection of highly specific adaptation patterns.

3. A STORY OF GIANTS AND DWARFS: α -PROTEOBACTERIAL GENOME EVOLUTION

If every gene in the α -proteobacteria evolved by vertical descent, all genes would support the same tree topology. This is obviously not the case: only 33–97% of all genes have a most similar homologue that is also an α -proteobacterial species (Esser *et al.* 2007). Furthermore, these genomes show a 10-fold variation in genome size, ranging from 1 to more than 9 Mbp (figure 1), with species-specific genes counting in the thousands (Boussau *et al.* 2004). Thus, we may ask: are all genes equally flexible and how does this variability correlate with adaptations to specific growth niches?

Present in all species, including even the smallest genomes, are approximately 200 genes for DNA, RNA and protein syntheses and another 40 genes for nucleotide and cofactor biosyntheses (Boussau *et al.* 2004). Gene content statistics for different functional categories show weak increases in gene number with genome size for genes involved in basic *information processes*, such as transcription and translation (approx. 4 genes Mbp⁻¹; Boussau *et al.* 2004). Much more dramatic increases are observed for genes involved in *adaptability processes*, such as energy metabolism, transport and regulatory functions (approx. 80–100 genes Mbp⁻¹), with some of the smallest genomes having virtually no regulatory genes (Boussau *et al.* 2004).

Figure 1. The α -proteobacterial species phylogeny. Circular tree of the α -proteobacterial phylogeny (Williams *et al.* 2007). The growth habitat, GC content and genome size are depicted for each species (numbers taken from the Genomes Online database, <http://www.genomesonline.org>).

We anticipate these variations in gene content to reflect the quality and stability of the growth environment. Indeed, by mapping the fluctuations in genome size onto the species tree, periods of massive genome reduction and expansion can be identified with hundreds of genes lost and gained in direct association with important environmental shifts (Boussau *et al.* 2004). For example, bacteria adapted to the cytoplasm of eukaryotic host cells, which is a relatively static and rich growth environment, have evolved towards small genomes, whereas free-living bacteria adapted to soil, which is more variable in nutritional quality, have genomes of much larger sizes. Thousands of genes have been eliminated and thousands of genes have been acquired. Below, we discuss the mechanisms underlying these dramatic changes in gene content.

4. MECHANISMS THAT CONTRIBUTE TO GENOME REDUCTION

Gene loss has been a dominant process in the evolution of the intracellular members of Rickettsiales (figure 2a), which have single circular genomes in the 1–1.5 Mbp range (Darby *et al.* 2007). Among genes discarded in *Rickettsia* are those involved in the biosynthesis of metabolites that are instead imported from the nutritionally rich cytoplasm of the eukaryotic host cell, such as vitamins and amino acids (Darby *et al.* 2007; Fuxelius *et al.* 2007). This reductive mode of evolution is still ongoing, as manifested in high fractions of fragmented genes and non-coding contents of up to 30 per cent (Fuxelius *et al.* 2008). Although weakly mutated genes may still be expressed, a majority are severely degraded and probably non-functional (Darby *et al.* 2007).

Genomic streamlining is observed in the SAR11 clade, here represented by *Pelagibacter ubique* (figure 2a). This genome has also suffered from gene loss, but in contrast to the Rickettsiales genomes it contains neither pseudogenes nor remnants of mobile elements. With a genome size of 1.3 Mbp, and an intergenic distance of only 3 bp, this is the most compact bacterial genome sequenced to date (Giovannoni *et al.* 2005). It is suggested that genomic streamlining in *P. ubique* is the result of selection on genome size *per se* to increase the surface-to-volume ratio and thereby facilitate uptake of nutrients that are limiting in the oceans, such as phosphorus and nitrogen.

Gene transfer to the nuclear genome of the host may also potentially contribute to genome reduction in obligate host-associated bacteria (figure 2a). For example, bacterial genes from *Wolbachia* have been discovered in the nuclear genomes of their fly hosts; it is not yet known whether these transferred genes contribute bacterial functions (Hotopp *et al.* 2007). Such gene transfers underlie the reductive evolution of the mitochondrial genome, which is derived from an endosymbiotic α -proteobacterium (Kurland & Andersson 2000).

5. MECHANISMS THAT CONTRIBUTE TO GENOME EXPANSION

In addition to the chromosome, many species of the Rhodobacterales contain auxiliary replicons up to several megabase pairs in size. *Transfer of genes* located on auxiliary replicons may lead to rapid expansion of the gene pool (figure 2b). Some of these replicons, also called megaplasmids, contain plasmid-like replication systems and were presumably once derived from plasmids (Cevallos *et al.* 2008). The megaplasmids are

Figure 2. Reduction and expansion of α -proteobacterial genomes. Example of mechanisms that contribute to genome (a) reductions ((i) genome streamlining, (ii) gene loss and (iii) gene transfer to host) and (b) expansions ((i) auxiliary replicon transfer, (ii) auxiliary replicon integration, (iii) island transfer and (iv) gene transfer from host).

extremely variable among strains (Giuntini *et al.* 2005) and contain mostly genes that confer adaptive traits, thereby facilitating niche specificity. As such, they represent a playground for evolution, where selection may rapidly drive any gene circulating with the conjugation system to fixation.

The auxiliary replicons may also *integrate* into the main chromosome (figure 2b), as observed in *Bartonella* (Alsmark *et al.* 2004) and *Bradyrhizobium* (Viprey *et al.* 2000). Such integration events serve to stabilize the auxiliary gene pool, thereby preventing its loss from the population. Most α -proteobacterial genomes also contain genomic islands that carry genes encoding important adaptive traits that can integrate, excise and be *transferred* across species (Sullivan & Ronson 1998).

Finally, host-adapted bacterial genomes may expand in size by the acquisition of genes from the host (figure 2b). For example, 'eukaryotic-like' genes encoding proteins with tetratricopeptide and ankyrin (ANK) repeat domains are present in several members of the Rickettsiales (Wu *et al.* 2004; Cho *et al.* 2007; Klasson *et al.* 2008). In analogy with the secretion of ANK proteins in *Legionella* and *Coxiella* (Pan *et al.* 2008), it is thought that the ANK proteins in the Rickettsiales are secreted and used to modulate the functions of the host cell.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

The difference in size among α -proteobacterial genomes reflects adaptations to different environments, just as the different sizes of the beaks of the Darwin finches reflect adaptations to different food sources. By placing the genomic characteristics within a phylogenetic framework, the association between genotypic and phenotypic diversities can be explored (figure 1). Two opposing mechanisms explain most of the genome size variations observed: loss of genes in obligate host-adapted bacteria (figure 2a) and gain of genes in free-living bacteria (figure 2b). Most of the variability in gene content is associated with genes conferring adaptive traits, with bacteria living in isolated and rich environments having a smaller auxiliary gene pool than bacteria adapted to environments that frequently change in nutritional quality. Most of the variability is associated with the mobile gene pool, as a result of which genes for adaptive traits can be amplified and exchanged upon demand without compromising the stability of the vertically inherited core genes.

The acquisition of new adaptive traits in bacterial genomes is rarely due to single nucleotide polymorphisms. Much more dramatic changes, such as deletions, duplication–divergence and *de novo* acquisitions of genes, need to be invoked. These changes

are often associated with repetitive sequences, i.e. regions that are often not well resolved in eukaryotic draft genome sequences. Here, there may be an important lesson for future genomic studies of adaptive traits in eukaryotes, such as the evolution of different beak sizes in the Darwin finches.

T.J.G.E. acknowledges the support from a Marie Curie Intra-European fellowship by the European Union. S.G.E.A. acknowledges the support of the Swedish Research Council, the Swedish Foundation for Strategic Research, the European Union, the Göran Gustafsson Foundation, and the Knut and Alice Wallenberg Foundation.

- Alsmark, C. M. *et al.* 2004 The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc. Natl Acad. Sci. USA* **101**, 9716–9721. (doi:10.1073/pnas.0305659101)
- Batut, J., Andersson, S. G. & O’Callaghan, D. 2004 The evolution of chronic infection strategies in the alpha-proteobacteria. *Nat. Rev. Microbiol.* **2**, 933–945. (doi:10.1038/nrmicro1044)
- Bourtzis, K. & Miller, T. A. 2003 *Insect symbiosis*. Boca Raton, FL: CRC Press.
- Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. 2004 Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl Acad. Sci. USA* **101**, 9722–9727. (doi:10.1073/pnas.0400975101)
- Cevallos, M. A., Cervantes-Rivera, R. & Gutierrez-Rios, R. M. 2008 The repABC plasmid family. *Plasmid* **60**, 19–37. (doi:10.1016/j.plasmid.2008.03.001)
- Cho, N. H. *et al.* 2007 The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host–cell interaction genes. *Proc. Natl Acad. Sci. USA* **104**, 7981–7986. (doi:10.1073/pnas.0611553104)
- Darby, A. C., Cho, N. H., Fuxelius, H. H., Westberg, J. & Andersson, S. G. 2007 Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* **23**, 511–520. (doi:10.1016/j.tig.2007.08.002)
- Dutilh, B. E., Snel, B., Ettema, T. J. & Huynen, M. A. 2008 Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* **25**, 1659–1667. (doi:10.1093/molbev/msn115)
- Esser, C., Martin, W. & Dagan, T. 2007 The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol. Lett.* **3**, 180–184. (doi:10.1098/rsbl.2006.0582)
- Fuxelius, H. H., Darby, A. C., Min, C. K., Cho, N. H. & Andersson, S. G. 2007 The genomic and metabolic diversity of *Rickettsia*. *Res. Microbiol.* **158**, 745–753. (doi:10.1016/j.resmic.2007.09.008)
- Fuxelius, H. H., Darby, A. C., Cho, N. H. & Andersson, S. G. 2008 Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol.* **9**, R42. (doi:10.1186/gb-2008-9-2-r42)
- Giovannoni, S. J. *et al.* 2005 Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245. (doi:10.1126/science.1114057)
- Giuntini, E., Mengoni, A., De Filippo, C., Cavalieri, D., Aubin-Horth, N., Landry, C. R., Becker, A. & Bazzicalupo, M. 2005 Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* **6**, 158. (doi:10.1186/1471-2164-6-158)
- Hotopp, J. C. *et al.* 2007 Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753–1756. (doi:10.1126/science.1142490)
- Klasson, L. *et al.* 2008 Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol. Biol. Evol.* **25**, 1877–1887. (doi:10.1093/molbev/msn133)
- Kurland, C. G. & Andersson, S. G. 2000 Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820. (doi:10.1128/MMBR.64.4.786-820.2000)
- Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. 2008 Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**, 1651–1654. (doi:10.1126/science.1158160)
- Sullivan, J. T. & Ronson, C. W. 1998 Evolution of Rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe-tRNA* gene. *Proc. Natl Acad. Sci. USA* **95**, 5145–5149. (doi:10.1073/pnas.95.9.5145)
- Viprey, V., Rosenthal, A., Broughton, W. J. & Perret, X. 2000 Genetic snapshots of the *Rhizobium* species NGR234 genome. *Genome Biol.* **1**, RESEARCH0014.
- Williams, K. P., Sobral, B. W. & Dickerman, A. W. 2007 A robust species tree for the alphaproteobacteria. *J. Bacteriol.* **189**, 4578–4586. (doi:10.1128/JB.00269-07)
- Wu, M. *et al.* 2004 Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, E69. (doi:10.1371/journal.pbio.0020069)