

Minireview

Shuffling bacterial metabolomes

Brendan Thomason and Timothy D Read

Address: Biological Defense Research Directorate, Naval Medical Research Center, 12300 Washington Avenue, Rockville, MD 20852, USA.

Correspondence: Timothy D Read. Email: readt@nmrc.navy.mil

Published: 27 February 2005

Genome Biology 2006, **7**:204 (doi:10.1186/gb-2006-7-2-204)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2006/7/2/204>

© 2006 BioMed Central Ltd

Abstract

Horizontal gene transfer (HGT) has a far more significant role than gene duplication in bacterial evolution. This has recently been illustrated by work demonstrating the importance of HGT in the emergence of bacterial metabolic networks, with horizontally acquired genes being placed in peripheral pathways at the outer branches of the networks.

The generation of bacterial genetic diversity through both vertical and horizontal gene transmission is well documented, but the relative contribution of each process has been a subject of much debate [1-7]. Now, because of the availability of many whole-genome sequences, biologists can perform unprecedented comparative analyses, looking at issues ranging from global genomic evolution to the roles of individual genes. A study published recently in *Nature Genetics* by Pál *et al.* [8] demonstrates how the powerful methodology of comparative genomics can be applied to the thorny issue of the evolution of bacterial metabolism.

The work by Pál *et al.* [8] is the first large-scale analysis performed so far to elucidate the effect of horizontal gene transfer (HGT) on bacterial metabolic networks. The subject of the study was the model organism *Escherichia coli* K12, which has the best-characterized metabolism for a bacterium, several tools for *in silico* metabolic reconstruction, and an abundance of proteobacterial relatives whose genomes have been fully sequenced. Whether *E. coli* is typical of all, or even most, free-living bacteria will have to be determined in subsequent work, but there is no reason at present to believe otherwise. Pál *et al.* [8] identified HGT events by determining the presence and absence of genes by comparison with a phylogenetic tree of conserved proteins in 51 proteobacterial species. Unlike the model eukaryote *Saccharomyces cerevisiae*, the genome of *E. coli* K12 contains only a few duplicated genes for enzymes in metabolic

pathways, and almost all these duplications appear to be ancient [8]. In fact, only one *E. coli* duplication event is predicted to have occurred since the divergence from the *Salmonella* lineage, whereas 15-32 genes are estimated to have been horizontally transferred during the same period. The likely reasons for this difference are the large number of mechanisms that facilitate gene transfer within and between bacteria (such as plasmids, phages, and other mobile DNA elements) and the bacterial bias for gene deletion [2] that tends to remove duplications before useful functional differences evolve.

Pál *et al.* [8] also investigated the effect of HGT on *E. coli* physiology, using simulated gene deletions in an *in silico* metabolic model under 136 simulated environments [9]. Only 7% of the horizontally transferred genes were essential under nutrient-rich conditions, compared with 23% of the 'native' genes. Instead, genes originating from HGT appear to be important in a smaller number of nutrient-limited conditions, and *E. coli* network topologies tended to place these genes at the outermost branches of the network; those responsible for nutrient uptake were more likely to undergo transfer, whereas enzymes involved in intermediate reactions within pathways remained relatively uniform [8]. Intuitively, this makes sense, as genes introduced at central points of the metabolic network are likely to affect multiple pathways unrelated to the primary function of the gene product. Furthermore, HGT events were not restricted to

isolated genes. Using flux-coupling analysis [10], Pál *et al.* [8] showed that both fully and directionally coupled enzymes are much more frequently gained or lost together than chance would allow, suggesting that whole pathways (or large portions, at least) are transferred simultaneously via HGT.

Because HGT genes were shown to be important in nutrient-limited conditions, the study by Pál *et al.* [8] demonstrates that the impetus for retention of a gene moved by horizontal transfer in *E. coli* is predicated on niche-specific necessity, whereas the genes that are invariant among proteobacteria contribute to fitness in most environments. This fits into an unfolding model for the genetic make-up of bacterial species [6] that is quite different from what we have come to expect in sexually reproducing eukaryotes. Sequencing of multiple strains from the same species of free-living prokaryotes has so far revealed a foundation of conserved genes and a potentially very much larger set of other, 'accessory', genes that, if expressed, probably confer transient advantages.

We are just beginning to understand the size of the accessory gene pool available to free-living bacteria that could provide the plug-in metabolic functions revealed by Pál *et al.* [8]. Tettelin *et al.* [11] have coined the concept of the 'pan-genome', which consists of genes present in all strains (the 'core genome') as well as genes absent in one or more strains and genes that are unique to each individual strain (the 'dispensable genome'). Even after sequencing eight group B *Streptococcus* genomes, the scale of the pan-genome could not be established. In fact, mathematical modeling predicted that even the data from hundreds of genomes might not be sufficient for such an endeavor. Using regression analysis, Tettelin *et al.* [11] found that the bacterial pan-genome for group B streptococci is immense; new genes were continually added on the completion of each fully sequenced genome. The analysis of their data in conjunction with previously generated data suggests that the environmental gene pool available for acquisition by methods such as HGT is much larger than previously estimated, allowing one to envisage a scenario of frequent flow of genetic material among bacteria inhabiting the same environment [11,12]. The pan-genome story for parasitic and endosymbiotic bacteria with smaller, reduced genomes, such as chlamydiae and rickettsiae, is likely to be different from that for free-living bacteria such as group B streptococci and *E. coli*, because millions of years of reductive evolution and an absence of opportunities for HGT have drastically pared the number of dispensable genes in the former.

The environmental specificity of genes acquired by HGT and their integration into the periphery of cellular transcription networks suggested by the results of Pál *et al.* [8] should be reflected in the abundance of their encoded proteins. Taoka *et al.* [13] have used liquid chromatography-based protein identification technology to analyze the proteome of *E. coli* K12, identifying 1,480 expressed proteins. Plotting this protein

inventory on a circular genome map demonstrated that protein-producing genes were infrequently found within K-loops, the genomic islands in K12 that most recently immigrated into the genome. The apparent under-representation of K-loop proteins was only observed in the proteome, as DNA microarray analysis demonstrated that K-loop genes were transcribed as efficiently as the native genes [13]. Although it is possible that K-loop genes are maintained for their function as cellular RNAs, it remains equally plausible that the recently acquired sets of proteins are retained for survival within a specific, unique environment, which was not met in this laboratory setting. The probable redundancy of many accessory genes in the genomes of free-living bacteria is also indicated by the rapid accumulation of pseudogenes and deletions in the genomes of pathogens undergoing reductive evolution, such as *Yersinia pestis* [14] and *Mycobacterium leprae* [15].

If the results of Pál *et al.* [8] are combined with other recent comparative genomic studies we arrive at the conclusion that large numbers of genes are available to a typical free-living bacterium via HGT to allow adaptation to specific environments, and that those genes fit into the outer branches of the metabolic networks. If this holds true across different species, how might one extrapolate from this information? We would predict that the physical presence of HGT-transferred genes is more closely linked to certain environments than that of core genes. This would happen because selection for retention of an accessory gene would be lost once an organism leaves the microenvironment where that gene is useful. Some early community DNA sequencing projects comparing different environments (for example [16]) have supported differing metabolic gene compositions. Thus, looking at the composition of accessory metabolic genes in a genome may give us more clues to the stresses faced by a bacterium, and possibly also to its recent evolutionary history. The study by Taoka *et al.* [13] illustrates the value of proteomic studies that verify which of the accessory gene products are actually produced *in vivo*. In the case of pathogens, the accessory metabolic pathways that are expressed in the host should point to *in vivo* survival mechanisms. We predict that many of the genes involved in these pathways encode membrane transporters and enzymes catalyzing initial reactions in metabolic pathways that have little supporting annotation other than gene model matches - the type of genes that are often overlooked in bioinformatic screens for virulence genes. It will be possible, therefore, to use knowledge of how HGT affects metabolic pathways to pick out novel virulence-associated proteins in pathogens.

References

1. Doolittle WF: **Lateral genomics.** *Trends Cell Biol* 1999, **9**:M5-M8.
2. Gogarten W, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19**:2226-2238.
3. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: quantification and classification.** *Annu Rev Microbiol* 2001, **55**:709-742.

4. Kurland CG, Canback B, Berg OG: **Horizontal gene transfer: a critical view.** *Proc Natl Acad Sci USA* 2003, **100**:9658-9662.
5. Lan R, Reeves PR: **Gene transfer is a major factor in bacterial evolution.** *Mol Biol Evol* 1996, **13**:47-55.
6. Lerat E, Daubin V, Ochman H, Moran NA: **Evolutionary origins of genomic repertoires in bacteria.** *PLoS Biol* 2005, **3**:e130.
7. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation.** *Nature* 2000, **405**:299-304.
8. Pál C, Papp B, Lercher MJ: **Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.** *Nat Genet* 2005, **37**:1372-1375.
9. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886-897.
10. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301-312.
11. Tettelin H, Maignani V, Cieslewics MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, et al.: **Genome analysis of multiple isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome".** *Proc Natl Acad Sci USA* 2005, **102**:13950-13955.
12. Hsiao WWL, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FSL: **Evidence of a large novel gene pool associated with prokaryotic genomic islands.** *PLoS Genet* 2005, **1**:e62.
13. Taoka M, Yamauchi Y, Shinkawa T, Kaji H, Motohashi W, Nakayama H, Takahashi N, Isobe T: **Only a small subset of the horizontally transferred chromosomal genes in *Escherichia coli* are translated into proteins.** *Mol Cell Proteomics* 2004, **3**:780-787.
14. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebahia M, James KD, Churcher C, Mungall KL, et al.: **Genome sequence of *Yersinia pestis*, the causative agent of plague.** *Nature* 2001, **413**:523-527.
15. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, et al.: **Massive gene decay in the leprosy bacillus.** *Nature* 2001, **409**:1007-1011.
16. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, et al.: **Comparative metagenomics of microbial communities.** *Science* 2005, **308**:554-557.