REVIEW ARTICLE

The fate of new bacterial genes

Chih-Horng Kuo[1] & Howard Ochman[1,2]

[1]Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, AZ, USA; and [2]Department of Biochemistry & Molecular Biophysics, University of Arizona, Tucson, AZ, USA

Correspondence: Howard Ochman, Department of Biochemistry, 233 Life Sciences South, 1007 E. Lowell Street, University of Arizona, Tucson, AZ 85721, USA. Tel.: +1 520 626 8355; fax: +1 520 621 3709; e-mail: hochman@email.arizona.edu

Received 4 August 2008; revised 25 September 2008; accepted 4 October 2008.
First published online 19 November 2008.

DOI:10.1111/j.1574-6976.2008.00140.x

Editor: Michael Galperin

Keywords
genome evolution; lateral gene transfer; interaction networks; gene inventories.

Abstract

Bacteria experience a continual influx of novel genetic material from a wide range of sources and yet their genomes remain relatively small. This aspect of bacterial evolution indicates that most newly arriving sequences are rapidly eliminated; however, numerous new genes persist, as evident from the presence of unique genes in almost all bacterial genomes. This review summarizes the methods for identifying new genes in bacterial genomes and examines the features that promote the retention and elimination of these evolutionary novelties.

Introduction

Bacteria can differ greatly in genome size, with published genome sequences ranging from about 160 to 13 000 kb. On account of the very high coding potential of bacterial genomes – nearly 90% of the sequences in most bacterial genomes specify proteins – an increase or decrease in genome size will usually effect a change in gene number and in the functional capacity of the organism.

The basis for the variation in gene numbers among bacteria is fairly well understood. Bacteria gain genes by a variety of mechanisms (Davison, 1999), and this gene gain is offset by a pervasive mutational bias toward deletions (Mira *et al.*, 2001). The elucidation of these processes has been greatly facilitated by the availability of full genome sequences, which provide unequivocal evidence of the genes that are present in (as well as those that are absent from) a genome, and has thereby allowed in-depth comparisons of the complete gene inventories of numerous bacteria.

As a result from such analyses, it has been shown that (1) bacteria having the same or similar genome sizes can harbor different sets of genes; (2) relatively few genes (certainly < 100, Koonin, 2003; Charlebois & Doolittle, 2004) are shared among all bacteria; (3) closely related bacteria – even strains from the same species – can have very different genome sizes, gene contents and phenotypic capabilities (Welch *et al.*, 2002); (4) genes that are under no functional constraints will erode and eventually be eliminated from the genome; (5) those bacteria, such intracellular pathogens and symbionts, living in constant, nutrient-rich host environments have genomes that are reduced and derived from their free-living relatives (Moran & Mira, 2001; Silva *et al.*, 2001); and (6) the overall genome size and cellular replication rate is not correlated either within or across bacterial species (Mira *et al.*, 2001; Froula & Francino, 2007). These findings suggest that selection does not act on genome size *per se*; rather, it is likely that the efficacy of selection acting on individual genes largely determines the content of a genome, which in turn influences the genome size.

From these comparative analyses, it is evident that new genes crop up frequently in bacterial genomes, and that those genes serving a useful function can be retained whereas those that do not are often removed from the genome. Given the relatively brief interval (at least when considered on evolutionary timescales) between the appearance of a new gene and its inactivation/elimination, there are likely to be some specific properties that are common to those new genes that are maintained by or removed from a

genome. For example, Sorek *et al.* (2007) demonstrated that toxicity of an acquired gene can serve as a major barrier to the lateral transfer in *Escherichia coli*, especially in cases of single-copy genes already present in the host genome.

In this review, we ask how, why, and which new bacterial genes persist in bacterial genomes. We focus specifically on those new genes that are unique to a bacterial species (or strain) because the route by which such novel genes are recruited into cellular processes is intriguing and also because they are the most likely source of totally new traits. For example, it is easy to envision how a duplicated gene, which can arise fully functional in the genome, might complement or displace its paralog, and, similarly, for horizontally acquired genes with homologs already present in the recipient genome. Rather, we consider those sequences that originate by gene transfer but represent something entirely new (i.e. that lack homologs) in the resident genome, so they cannot immediately replace the function of existing genes.

Identification of new genes in genomes

The current methods for identifying newly acquired genes fall broadly into three categories: compositional analysis, detection of phylogenetic anomalies, and comparison of genome content (Ochman *et al.*, 2000; Nakamura *et al.*, 2004; Gogarten & Townsend, 2005). Compositional analysis is the only one of these methods that does not rely on the comparison of sequences from more than a single genome, but it can be unreliable if applied indiscriminately (Koski *et al.*, 2001; Wang, 2001). In short, this approach compares the sequence features such as nucleotide composition or codon usage of genes within a genome. Because base composition can vary widely among bacterial species but is relatively homogeneous within a bacterial genome, the detection of transferred genes is relatively straightforward if sequence features of the donor are different from those of the recipient. However, this approach has a limited sensitivity and selectivity when (1) the donor and the recipient species have similar nucleotide compositional properties; (2) the candidate genes are short such that anomalies might be attributable to stochastic factors; (3) the transfer event was ancient and the gene has mutated to the extent that it resembles the rest of the genes in the genome; or (4) the resident has biased compositional properties because its encoded proteins have unusual amino acid compositions.

Because genes that are inherited vertically have the same evolutionary history, the second approach to detect lateral gene transfer (LGT) is to identify genes that have discordant phylogenies (Beiko *et al.*, 2005). In this method, a gene is identified as being laterally transferred if it groups with other distantly related species in the phylogenetic tree or if the sequence similarity searches find its best hit with a phylogenetically distant species. Because this approach relies on the availability of homologous sequences from a collection of reference species, it cannot be applied to those genes that are unique to a particular lineage or genome. Although the relatively sparse sampling of available genomes might seem to make the chance of correctly identifying most laterally transferred genes rather low, numerous cases of LGT have been identified by this method. The shortcomings of this method are related to the intricacies of phylogenetic inference: factors such as poor taxon sampling, incomplete lineage sorting, and the inconsistency of phylogenetic methods can all create artifacts in the inferred gene trees and result in the erroneous identification of LGT (Felsenstein, 2004).

The final approach for identifying new genes is to directly align genomes and to examine the presence and absence of genes among organisms. This method entails comparisons of multiple related genomes, which are certainly available for many bacterial taxa, and can provide the most direct evidence of genes that are gained and lost from a lineage. However, the utility of this approach is limited by the level of syntenic conservation among the genomes examined.

Note that there is often substantial overlap between the sets of genes detected by each of these three approaches; for example, acquired genes with atypical sequence features are frequently unique to a genome, which would also allow them to be identified using genome alignments. Therefore, the use of a specific method depends largely on the availability of particular genome sequences. To limit any subsequent biases that might be introduced by the methods used to identify acquired genes, the best approach is to focus on the candidates that have multiple lines of independent support for their recent acquisitions.

The prevalence of novel genes

The observation that a very limited number of genes is common to all bacterial genomes implies that most genes in most genomes arose during the evolutionary diversification of bacteria. It is now possible to determine the processes by which most new bacterial genes originated by examining their occurrence and distribution within and among genomes. Although some very large multigene families, such as the ABC transporters (Tatusov *et al.*, 1997), have expanded through successive gene duplications (Teichmann & Babu, 2004), LGT is generally viewed as the predominant force in expanding the gene repertoires of microbial genomes (Pál *et al.*, 2005a; Dagan *et al.*, 2008). Furthermore, in an analysis of 13 gammaproteobacterial genomes, Lerat *et al.* (2005) determined that over half of the $> 14\,000$ gene families consist of a single gene present in a single genome.

It has long been known that all but the most reduced bacterial genomes contain large numbers of these so-called 'ORFan' genes, i.e. genes restricted to a single genome for which there are no known homologs in current databases (Fischer & Eisenberg, 1999), and an early survey estimated

the frequency of ORFans to be *c*. 14% of the total genes from 60 completely sequenced genomes (Siew & Fischer, 2003). ORFan genes were originally considered to be unique to each organism, but as the number of sequenced genomes increased, they are now recognized as belonging to a continuum of genes that originated over the entire history of a lineage. For this reason, 'lineage-specific ORFan genes' may be a more appropriate term for describing genes that are restricted to any monophyletic group of organisms (ranging from a single bacterial strain to a phylum). Using lineage specificity as a means to estimate the age of a gene, one can evaluate the rate and manner by which new genes are recruited into and maintained by a genome.

Pattern of sequence evolution subsequent to transfer

Upon introgression, laterally transferred genes often experience an amelioration process that gradually changes their sequence characteristics (e.g. GC content and codon usage) to that of the host genome (Lawrence & Ochman, 1997). It has been noted that newly acquired genes tend to evolve at faster rates than do native genes (Daubin & Ochman, 2004; Hao & Golding, 2006; Marri *et al*., 2007; Davids & Zhang, 2008), an observation that has been attributed both to the relaxation of selective constraints and to positive selection for new gene functions. In addition to changes in their coding regions, it has been found that newly acquired genes often experience accelerated sequence evolution in their *cis*-regulatory regions (Lercher & Pál, 2008).

Integration of new genes

Compared with many eukaryotes, the genomes of bacteria undergo deletional bias (Mira *et al*., 2001), such that genes that are not contributing to organismal fitness are eventually removed from the genome. Thus, for a new gene to persist, it must provide a relevant function soon after its arrival, and this function must operate within the native machinery of the host cell. To circumvent the difficulties involved in the integration of new genes into existing cellular networks (which may or may not have suitable interacting partners), multiple new genes can be acquired as a single self-contained functional unit (i.e. an operon), which can provide a function immediately upon introgression (Lawrence & Roth, 1996; Omelchenko *et al*., 2003; Pál *et al*., 2005a; Homma *et al*., 2007). In addition to the inclusion of functionally related genes as a single unit, operons are often acquired with their regulators, which further facilitate the integration process (Price *et al*., 2008). In the cases of single-gene acquisition, the fate of new genes depends largely upon the existing genes in its new host. For example, newly acquired genes are much more likely to persist in the presence of enzymes that catalyze a coupled metabolite flux (Pál *et al*., 2005b).

Because many newly acquired genes are unique to a bacterial lineage, assigning functions to these genes represents a particular challenge in understanding the integration process. Because of their lack of clear homologs, it is rarely possible to assign functions to such genes based on similarity to previously characterized sequences in other organisms. Even for those organisms that are experimentally tractable, the disruption of newly acquired genes rarely yields an observable phenotype. Currently, one of the best ways to understand the process by which new genes function and are integrated in cellular processes is to examine their associations and interactions with genes and proteins of known functions.

In the context of transcription networks, new genes are added to the periphery of a network as new targets for existing transcription factors (Cosentino Lagomarsino *et al*., 2007). Initially, newly acquired genes may not be regulated optimally and have low expression levels (Taoka *et al*., 2004), but the gradual integration of new genes into the cellular network leads to both an increase in the number of regulators and an improvement in the coexpression with their protein–protein interaction (PPI) partners (Lercher & Pál, 2008). Within PPI networks, new genes tend to have few interacting partners (Wellner *et al*., 2007; Lercher & Pál, 2008) and to preferentially interact with the central hubs that have many existing partners (Ochman *et al*., 2007; Davids & Zhang, 2008). The continuous influx of new genes appears be an important force that shapes the structure of cellular networks (Wolf *et al*., 2002; Light *et al*., 2005).

Ecological and evolutionary significance

Because of their impact on public health, the acquisition of pathogenicity (Groisman & Ochman, 1997; Hacker & Kaper, 2000; Wren, 2000) and antibiotic resistance determinants (Neu, 1992; Courvalin, 1994; Shoemaker *et al*., 2001; Springael & Top, 2003) represent some of the best-studied cases of LGT in bacteria. These examples, together with cases in which laterally acquired genes have enabled bacteria to develop symbiotic relationships (Sullivan & Ronson, 1998) or to survive in extreme environments (Kennedy *et al*., 2001; Campanaro *et al*., 2005; Martinez *et al*., 2006), highlight the adaptive role of new genes in allowing their hosts to exploit novel ecological niches.

In addition to conferring an ability to explore novel ecological niches, laterally acquired genes can also expand the metabolic capacities of their hosts (Boucher & Doolittle, 2000; Jahreis *et al*., 2002; Pál *et al*., 2005a). Taken together, new genes originating by LGT may be an important evolutionary force that promotes speciation in bacteria (Lawrence, 1999, 2002).

The misfits

So far, we have considered only the success stories, i.e. those genes that were ultimately integrated into the genome; but what fraction of newly acquired sequences achieves this status? It is likely that most acquisition (and subsequent deletion) events occur rapidly and do not leave a trace in the genome, so the actual fraction of acquired genes that become successfully incorporated is somewhat difficult to quantify, but it is presumably low (Lawrence & Ochman, 1998; Liu *et al.*, 2004; Hao & Golding, 2008). And while establishing the ancestry of each gene in a genome is a relatively straightforward procedure, determining which genes have been deleted is somewhat more problematic because this set is limited to those genes present in and ancestral to related lineages but missing from the genome in question. Nonetheless, by comparing the complete gene repertoires of contemporary genomes in a phylogenetic context, it is possible to obtain several clues about the manner in which genes are both gained and lost in bacterial genomes (Fuxelius *et al.*, 2008).

The first glimpse into the features of genes that are eliminated from bacterial genomes came from comparisons of the highly reduced gene repertoires of obligate endosymbionts compared with those of their free-living relatives (Moran & Mira, 2001; Dagan *et al.*, 2006). But because these symbionts are distantly related to their large-genomed ancestors, and they harbor an almost minimal set of genes and no recently acquired genes, they provide little information about the process by which new genes are integrated into and eliminated from genomes. However, it has been observed that a disproportionate number of genes purged from these highly reduced genomes were those having only a single interaction partner (Ochman *et al.*, 2007). This result suggests that the core of gene networks is maintained by a strong selection due to the requirement of these highly connected genes for the survival of organisms. In contrast, genes that have few interacting partners are less likely to be essential and thus prone to erosion and loss, due to a combination of high mutation rates and elevated levels of genetic drift in these obligate endosymbionts.

To elucidate the process of gene loss on more recent timescales, assessments of bacterial pseudogenes have improved our understanding about how newly acquired genes are maintained and purged. In bacteria, pseudogenes were originally thought to be limited to pathogens that were undergoing genome degradation and reduction (e.g. Cole *et al.*, 2001), but they are now known to be a regular feature of virtually every bacterial genome (Ochman & Davalos, 2006). As expected, the majority of genes that become inactivated are of unknown function (Lerat & Ochman, 2004, 2005); however, even these nonfunctional proteins have provided insights into the way in which acquired genes are removed from PPI networks. For example, of the genes acquired by *E. coli* since it diverged the common ancestor to all *Gammaproteobacteria*, over 10% are presently pseudogenes in *Shigella*. With regard to their numbers of interacting partners, genes that have been silenced in *Shigella* constitute a random sample of those acquired; such that most are partnered with a single highly connected protein (Ochman *et al.*, 2007).

Finally, by examining the distribution of genes along an evolutionary lineage, we have recently attempted to estimate the fraction of acquired genes that becomes dispensable as well as the features of those that are retained. Based on the numbers of recent arrivals present in the *E. coli* K12 genome compared with the number of acquired genes ancestral to and maintained by all members of this species, it seems that only 10–15% of acquired genes are retained over the long haul (Ochman & Davalos, 2006). Moreover, by establishing the origins of the genes acquired during the diversification of the *E. coli* and *Shigella*, we found that new genes having distant homologs in other bacteria are acquired much more frequently, but are not retained as often as acquired genes with no identifiable homologs (van Passel *et al.*, 2008). This offers the tantalizing prospect that totally novel genes, i.e. those that have never conferred a function in a cellular genome, are more likely to persist in bacterial genomes.

Conclusion

The acquisition of new genes is an ongoing process for most bacterial genomes (with the exception of some obligate endosymbionts). The continual incorporation of new genes provides the raw materials for selection to act upon: new genes that improve organismal fitness or facilitate adaptation to novel ecological niches are likely to reach fixation, whereas others are eventually eroded through a combination of mutation and drift. Although LGT is one of the primary sources of novel genetic material, the origin of most genes that have no known homologs remains an unsolved mystery. Future studies that aim at understanding the functional significance of new genes will shed light on the evolution and diversification of bacterial lineages; however, such studies involve formidable challenges even in model organisms.

References

Beiko RG, Harlow TJ & Ragan MA (2005) Highways of gene sharing in prokaryotes. *P Natl Acad Sci USA* **102**: 14332–14337.

Boucher Y & Doolittle F (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol* **37**: 703–716.

Campanaro S, Vezzi A, Vitulo N *et al*. (2005) Laterally transferred elements and high pressure adaptation in *Photobacterium profundum* strains. *BMC Genomics* **6**: 122.

Charlebois RL & Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res* **14**: 2469–2477.

Cole ST, Eiglmeier K, Parkhill J *et al*. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007–1011.

Cosentino Lagomarsino M, Jona P, Bassetti B & Isambert H (2007) Hierarchy and feedback in the evolution of the *Escherichia coli* transcription network. *P Natl Acad Sci USA* **104**: 5516–5520.

Courvalin P (1994) Transfer of antibiotic resistance genes between gram-positive and gram-negative bacteria. *Antimicrob Agents Ch* **38**: 1447–1451.

Dagan T, Blekhman R & Graur D (2006) The "Domino Theory" of gene death: gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. *Mol Biol Evol* **23**: 310–316.

Dagan T, Artzy-Randrup Y & Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *P Natl Acad Sci USA* **105**: 10039–10044.

Daubin V & Ochman H (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coil*. *Genome Res* **14**: 1036–1042.

Davids W & Zhang Z (2008) The impact of horizontal gene transfer in shaping operons and protein interaction networks – direct evidence of preferential attachment. *BMC Evol Biol* **8**: 23.

Davison J (1999) Genetic exchange between bacteria in the environment. *Plasmid* **42**: 73–91.

Felsenstein J (2004) *Inferring phylogenies*. Sinauer Assoc., Sunderland, MA.

Fischer D & Eisenberg D (1999) Finding families for genomic ORFans. *Bioinfomatics* **15**: 759–762.

Froula JL & Francino MP (2007) Selection against spurious promoter motifs correlates with translational efficiency across bacteria. *PLoS ONE* **2**: e745.

Fuxelius HH, Darby AC, Cho NH & Andersson SGE (2008) Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol* **2008**: R42.

Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**: 679.

Groisman E & Ochman H (1997) How *Salmonella* became a pathogen. *Trends Microbiol* **5**: 343–349.

Hacker J & Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641–679.

Hao W & Golding GB (2006) The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res* **16**: 636–643.

Hao W & Golding GB (2008) Uncovering rate variation of lateral gene transfer during bacterial genome evolution. *BMC Genomics* **9**: 235.

Homma K, Fukuchi S, Nakamura Y, Gojobori T & Nishikawa K (2007) Gene cluster analysis method identifies horizontally transferred genes with high reliability and indicates that they provide the main mechanism of operon gain in 8 species of gamma-proteobacteria. *Mol Biol Evol* **24**: 805–813.

Jahreis K, Bentler L, Bockmann J, Hans S, Meyer A, Siepelmeyer J & Lengeler JW (2002) Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. *J Bacteriol* **184**: 5307–5316.

Kennedy SP, Ng WV, Salzberg SL, Hood L & DasSarma S (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**: 1641–1650.

Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**: 127–136.

Koski LB, Morton RA & Golding GB (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* **18**: 404–412.

Lawrence JG (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol* **2**: 519–523.

Lawrence JG (2002) Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**: 449–460.

Lawrence JG & Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**: 383–397.

Lawrence JG & Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. *P Natl Acad Sci USA* **95**: 9413–9417.

Lawrence JG & Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843–1860.

Lerat E & Ochman H (2004) Ψ–Φ: exploring the outer limits of bacterial pseudogenes. *Genome Res* **14**: 2273–2278.

Lerat E & Ochman H (2005) Recognizing pseudogenes in bacterial genomes. *Nucleic Acids Res* **33**: 3125–3132.

Lerat E, Daubin V, Ochman H & Moran NA (2005) Evolutionary origins of genomic repertoires in bacteria. *PLoS Biology* **3**: e130.

Lercher MJ & Pál C (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol* **25**: 559–567.

Light S, Kraulis P & Elofsson A (2005) Preferential attachment in the evolution of metabolic networks. *BMC Genomics* **6**: 159.

Liu Y, Harrison PM, Kunin V & Gerstein M (2004) Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol* **5**: R64.

Marri PR, Hao W & Golding GB (2007) The role of laterally transferred genes in adaptive evolution. *BMC Evol Biol* **7** (Suppl 1): S8.

Martinez RJ, Wang Y, Raimondo MA, Coombs JM, Barkay T & Sobecky PA (2006) Horizontal gene transfer of PIB-type ATPases among bacteria isolated from radionuclide- and metal-contaminated subsurface soils. *Appl Environ Microb* **72**: 3111–3118.

Mira A, Ochman H & Moran NA (2001) Delectional bias and the evolution of bacterial genomes. *Trends Genet* **10**: 589.

Moran NA & Mira A (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* **2**: research 0054.1–0054.12.

Nakamura Y, Itoh T, Matsuda H & Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**: 760–766.

Neu HC (1992) The crisis in antibiotic resistance. *Science* **254**: 1064–1073.

Ochman H & Davalos LM (2006) The nature and dynamics of bacterial genomes. *Science* **311**: 1730–1733.

Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Ochman H, Liu R & Rocha EP (2007) Erosion of interaction networks in reduced and degraded genomes. *J Exp Zool Part B* **308**: 97–103.

Omelchenko MV, Makarova KS, Wolf YI, Rogozin IB & Koonin EV (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol* **4**: R55.

Pál C, Papp B & Lercher MJ (2005a) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37**: 1372–1375.

Pál C, Papp B & Lercher MJ (2005b) Horizontal gene transfer depends on gene content of the host. *Bioinformatics* **21** (Suppl 2): ii222–ii223.

Price MN, Dehal PS & Arkin AP (2008) Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol* **9**: R4.

Shoemaker NB, Vlamakis H, Hayes K & Salyers AA (2001) Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl Environ Microb* **67**: 561–568.

Siew N & Fischer D (2003) Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* **53**: 241–251.

Silva FJ, Latorre A & Moya A (2001) Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet* **17**: 615–618.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P & Rubin EM (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**: 1449–1452.

Springael D & Top EM (2003) Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol* **12**: 53–58.

Sullivan JT & Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe-tRNA* gene. *P Natl Acad Sci USA* **95**: 5145–5149.

Taoka M, Yamauchi Y, Shinkawa T, Kaji H, Motohashi W, Nakayama H, Takahashi N & Isobe T (2004) Only a small subset of the horizontally transferred chromosomal genes in *Escherichia coli* are translated into proteins. *Mol Cell Proteomics* **3**: 780–787.

Tatusov RL, Koonin EV & Lipman DJ (1997) A genomic perspective on protein families. *Science* **278**: 631–637.

Teichmann SA & Babu MM (2004) Gene regulatory network growth by duplication. *Nat Genet* **36**: 492–496.

van Passel MW, Marri PR & Ochman H (2008) The emergence and fate of horizontally acquired genes in *Escherichia coli*. *PLoS Comput Biol* **4**: e1000059.

Wang B (2001) Limitations of compositional approach to identifying horizontally transferred genes. *J Mol Evol* **53**: 244–250.

Welch RA, Burland V, Plunkett G III *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *P Natl Acad Sci USA* **99**: 17020–17024.

Wellner A, Lurie MN & Gophna U (2007) Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol* **8**: R156.

Wolf YI, Karev G & Koonin EV (2002) Scale-free networks in biology: new insights into the fundamentals of evolution? *BioEssays* **24**: 105–109.

Wren BW (2000) Microbial genome analysis: insights into virulence, host adaptation and evolution. *Nat Rev Genet* **1**: 30–39.