Molecular Microbiology (2003) **50**(3), 739–749 doi:10.1046/j.1365-2958.2003.03778.x

MicroReview

Lateral gene transfer: when will adolescence end?

Jeffrey G. Lawrence* and Heather Hendrickson
Pittsburgh Bacteriophage Institute and Department of Biological Sciences, 352 Crawford Hall, University of Pittsburgh, Pittsburgh, PA 15260, USA.

Summary

The scope and impact of horizontal gene transfer (HGT) in Bacteria and Archaea has grown from a topic largely ignored by the microbiological community to a hot-button issue gaining staunch supporters (on particular points of view) at a seemingly ever-increasing rate. Opinions range from HGT being a phenomenon with minor impact on overall microbial evolution and diversification to HGT being so rampant as to obfuscate any opportunities for elucidating microbial evolution – especially organismal phylogeny – from sequence comparisons. This contentious issue has been fuelled by the influx of complete genome sequences, which has allowed for a more detailed examination of this question than previously afforded. We propose that the lack of common ground upon which to formulate consensus viewpoints probably stems from the absence of answers to four critical questions. If addressed, they could clarify concepts, reject tenuous speculation and solidify a robust foundation for the integration of HGT into a framework for long-term microbial evolution, regardless of the intellectual camp in which you reside. Here, we examine these issues, why their answers shape the outcome of this debate and the progress being made to address them.

Coming of age

The first complete genome sequence of a free-living organism (*Haemophilus influenzae*) was released in 1995 (Fleischmann *et al*., 1995) and, as of writing, more than 100 microbial genome sequences representing diverse lineages of Bacteria and Archaea have subsequently become available. The promise for new information held in complete genome sequences is vast and manifold, including (i) insight into previously unsuspected metabolic functions; (ii) elucidation of a microbe's underlying physiology even in the absence of a tractable genetic system or the ability to propagate the organism in pure culture; (iii) the identification of potential drug targets in pathogenic organisms; (iv) observations into the conservation of gene order, operon structure, variation in rates of evolution within and among genes, and so forth. The possibilities for mining novel answers to unasked questions also appear to be nearly endless, and the so-called 'post-genomic era' has indeed brought about the publication of clever investigations that have called attention to hitherto unimagined aspects of microbiology. However, judging by surveys of the literature, it also seems that complete genome sequences have generated more debate, speculation, discussion and publication of works – both those presenting objective analyses of new data and those proffering primarily interpretation and extrapolation of data according to one's point of view – regarding horizontal (lateral) gene transfer (HGT) than any other subject regarding the utilization of complete genome sequences. The availability of significant numbers of eukaryotic genome sequences has allowed the issue to be examined as a potent evolutionary force outside the prokaryotic domains.

This discussion of the scope and impact of HGT is not a young one, as the transmission of plasmid-borne antibiotic resistance genes between organisms has been recognized for decades (Davies, 1996). Yet, at the time, this phenomenon was not thought to be widespread. Owing to the nature of bacterial reproduction, genes were viewed as being inherited primarily by vertical transfer, transmitted faithfully from mother cell to daughter cell during binary fission. HGT was an idea in its infancy – new and cute, but of no impact on the weightier matters of overall microbial evolution. More contemporary genome analyses often reach the same conclusions (Snel *et al*., 1999), that is that vertical inheritance is the dominant mode of gene propagation, although the resolution becomes less staunch as more taxa are included for analysis and other evolutionary forces (gene loss and gene 'genesis') are examined in more detail (Snel *et al*., 2002). The trickle of DNA sequence data throughout the 1980s and early

Accepted 15 August, 2003. *For correspondence. E-mail jlawrenc@pitt.edu; Tel. (+1) 412 624 4204; Fax (+1) 412 624 4759.

1990s led to several compelling cases for HGT playing a role in the evolution of particular genes in some taxa (e.g. the *gapA* gene in proteobacteria; Doolittle *et al*., 1990). Even then there was no serious consideration of HGT as a major player in microbial evolution; vertical inheritance with periodic selection (Levin, 1981) was still the dominant perspective of microbial evolution, even when DNA transfer between closely related strains of the same 'species' was recognized (Dykhuizen and Green, 1991).

As a conceptual brick on the edifice of biological thought, HGT made its mark via numerous analyses of complete genome sequences; two general approaches were used. Phylogenetics could point out incongruent evolutionary histories of genes within the same genome (Gogarten *et al*., 1992; Gogarten, 1995), whereas parametric analyses found genes displaying sequence patterns that could be interpreted as telltale signs of long-term evolution in another mutational (and therefore, genomic) context (Médigue *et al*., 1991). Yet, despite the apparent surplus of data, HGT can still be considered to be an idea in its conceptual adolescence, so to speak. It has clearly shown promise in potentially changing directions of thought, allowing new insight into problems once thought to be tidily solved (Gogarten *et al*., 2002) and possibly offering new paradigms for interpreting microbial systematics, phylogeny and evolution. At its most dramatic interpretation by some readers, the apparently rampant and indiscriminate nature of HGT could dismantle the entire framework of bacterial phylogeny based on sequences of one or a few genes (Doolittle, 1999); this would occur primarily because multiple phylogenies would better represent the mosaic nature of bacterial chromosomes (Gogarten *et al*., 2002).

For all its promise, HGT has not really established itself in any of these areas of opportunity. It has not reached scientific adulthood, where it would be accepted as a cornerstone of microbial evolution with well-defined roles, boundaries, causes and consequences (Kurland, 2000). The 'genomic era' brought HGT to this point, and we propose four hurdles that must be passed for HGT to step out of the spotlight of debates between sceptics and champions – both often interpreting the same data from different viewpoints – and reach scientific maturity. We do not present here a comprehensive overview of the mechanism, elucidation, interpretation or impact of horizontal transfer (which have been reviewed extensively elsewhere, e.g. see Ochman *et al*., 2000; Koonin *et al*., 2001; Gogarten *et al*., 2002; Doolittle *et al*., 2003), or provide an overarching framework for its role in microbial genome evolution. Rather, we discuss these four questions and the progress being made towards answering them. With these data in hand, perhaps microbiologists could proceed to outline with rigour and confidence the roles of gene transfer in microbial evolution.

How does HGT impact the evolutionary history of different genes?

Perhaps nowhere has the HGT debate been more focused than on its relative influence on the evolutionary histories of different genes. No-one denies that certain classes of genes (e.g. those encoding antibiotic resistance) are associated with mobile genetic elements and can experience high rates of transfer (Hall, 1997). In contrast, the rRNA genes have long been considered relatively recalcitrant to transfer, allowing the foundations of bacterial phylogenetics (Woese, 1987). The phylogenies of other highly conserved genes, such as tRNA synthetases (Woese *et al*., 2000), primarily reflect that inferred from rRNA genes (Ludwig *et al*., 1998), although some notable transfers are evident among these phylogenies (Woese *et al*., 2000). These data support the view that a 'core' set of genes has been inherited by vertical descent and represent the 'true' phylogeny of the bacteria that harbour them. Along these lines, it has been proposed that genes whose products interact with a large number of other proteins and RNAs would be those least likely to be transferred (Jain *et al*., 1999). Newly introduced orthologues would be unlikely to express a product that could outperform one that had experienced long-term co-evolution with its cognate partners.

Implicit (but unstated) in the idea that highly conserved genes would be subject to less transfer is the verity that there would be a smaller subset of strains that could benefit from receiving the new genes. Clearly, most genomes would already contain a homologue of the transferred gene, and an orthologous replacement would have to occur. Among less highly conserved genes, many lineages may be naïve to the gene's product, and a selective advantage could arise by the newly acquired gene(s) providing a novel function (Lawrence and Roth, 1996; Lawrence, 1997; 1999). Among more highly conserved genes, orthologous replacement would occur at a rate of 50% at best, ignoring any detriments inherent in the retention of introgressed genes (which are discussed below) beyond their lack of co-evolution with potentially interacting partners.

Yet there have been cases where genes involved in information transfer (replication, transcription, translation) have been subject to HGT (for some examples, see Table 1). Indeed, even rRNA genes have been shown to experience HGT (Mylvaganam and Dennis, 1992; Yap *et al*., 1999); their ability to be transferred lies in many of the same features originally cited as reasons why they would probably not be: they are ubiquitous in distribution, are highly conserved and perform the identical function in all cells. Yet these properties actually promote exchange of all or parts of the rRNA molecule, fuelled by long regions of nucleotide identity (not encoding a protein, this

Table 1. Likely examples of HGT where genes participate in information transfer.

Protein	Phylogenetic incongruities	Reference
Ribosomal RNA (*rrn*)	(i) *Thermomonospora* contains *rrn* operon donated from *Thermobispora* (ii) *Haloarcula* contains *rrn* operon from a probable Halobacterial donor	Mylvaganam and Dennis (1992); Yap *et al.* (1999)
RNA polymerase	*Mycoplasma* branches at the bottom of the Bacterial domain	Klenk *et al.* (1999)
Ribosomal protein L32 (RpmF)	*Lactococcus lactis* groups within Proteobacteria	Makarova *et al.* (2001)
Ribosomal protein L33 (RpmG)	(i) *Deinococcus* groups with *Aquifex* instead of *Thermus* (ii) *Mycobacterium leprae* groups separately from *M. tuberculosis*	Makarova *et al.* (2001)
Ribosomal protein S14 (RpsN)	(i) Mycoplasmas are separate from other low-GC Gram-positive Bacteria (ii) *Deinococcus* is separated from *Thermus* and groups with some low-GC Gram-positive Bacteria	Brochier *et al.* (2000)
Ribosomal protein S18 (RpsR)	Three Mycoplasmatales species group with ε-proteobacteria	Makarova *et al.* (2001)
Lysyl-tRNA synthase	*Borrelia* groups with Archaea	Ibba *et al.* (1997)
Phenylalanyl-tRNA synthase	Spirochaetes group with Archaea	Woese *et al.* (2000)
Prolyl-tRNA synthase	(i) *Deinococcus*, *Mycoplasma* and *Borrelia* group with the Archaea (ii) *Borrelia* does not group with the spirochaete *Treponema*, which remains within the Bacterial clade	Gogarten *et al.* (1999); Woese *et al.* (2000)
Seryl-tRNA synthase	The Archaeon *Haloarcula* groups with Bacteria	Doolittle and Handy (1998); Woese *et al.* (2000)

gene lacks the variant bases that arise as a result of the degeneracy of the genetic code) and a high degree of conservation of function (Fig. 1). Moreover, surveys of genomes for atypical genes show that many other genes have been acquired recently, up to 25% of the genome (Hayes and Borodovsky, 1998; Karlin, 1998; Lawrence and Ochman, 1998; 2002; Nelson *et al.*, 1999; Garcia-Vallve *et al.*, 2000a; Ochman *et al.*, 2000; Ragan, 2001a).

Here is where one can interpret data in different ways. One position may be 'Look, no gene is immune to transfer, even if it is involved in a complex molecular machine with co-evolving parts. Therefore, no consortium of co-evolving genes defines the essence of a bacterial cell. As a result, one cannot simply deduce microbial evolution from molecular phylogenies as represented by a single, bifurcating tree; rather, this mosaicism is best represented by reticulation, where genomes contain genes with differing histories'. Such an argument has been made convincingly for bacteriophage genome evolution (Lawrence *et al.*, 2002); but in this case, the transferred fragments represent much larger fragments of the genome (up to 50%), and it is impossible to identify a common 'core' of genes shared among all bacteriophage lineages. By analogy, is it valid to extend this argument to bacterial genomes as well?

An alternative, and equally valid, viewpoint is that the transfer of highly conserved genes (Table 1, Fig. 1) is relatively rare, and therefore does not affect the robustness of the underlying organismal phylogeny in an analogous fashion. Instead, much of HGT would be limited to genes that affect bacterial lifestyle, but do not have a large impact on the 'core' set of genes involved in information transfer or central metabolism. Certainly no gene is

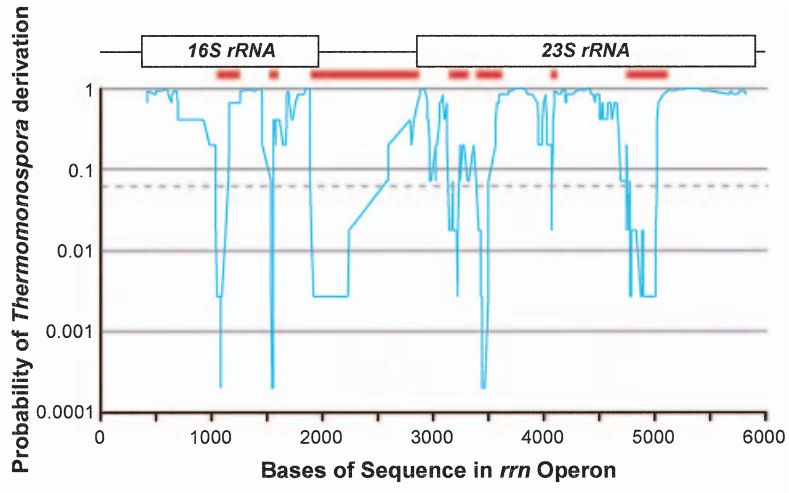

Fig. 1. Mosaicism within the *Thermomonospora chromogena rrnB* operon, which bears regions of identity to the *rrn* operons of *Thermobispora bispora*. Informative sites were identified as positions where (i) three full-length *T. chromogena rrn* operons were identical; (ii) two full-length *T. bispora rrn* operons were identical to each other, but differed from the *T. chromogena* sequences; and (iii) the *T. chromogena rrnB* base matched one of the two. Of the 478 informative sites, 202 sites (42%) paired the *T. chromogena rrnB* operon with *T. bispora rrn* operons, whereas 276 showed identity across all four *T. chromogena rrn* loci examined. A window of 10 informative sites was used to calculate the probability of the *T. chromogena rrnB* operon matching the other three *T. chromogena rrn* loci (blue line); $P = 0.0002$ indicates that all 10 sites within the window matched the *T. bispora rrn* loci. The red bars denote regions likely to be of *T. bispora* origin ($P < 0.05$). Figure adapted from Gogarten *et al.* (2002).

immune to HGT, and one can always identify the occasional transfer event among any set of genes. Yet, on the whole, these are exceptions to the rule of vertical inheritance of the as-yet-undesignated 'core' set of genes that encode the consortium of essential gene products enabling cellular life. The impact of HGT on these genes is constrained by its rarity in this arena, thus leaving organismal phylogeny – and all the biological inferences made from it – intact.

Recent analyses of orthologous sequences among diverse genomes supports their general congruence with the rRNA phylogeny, at least among the relatively closely related genomes of some clades (Makarova *et al*., 1999; Nesbo *et al*., 2001; Daubin *et al*., 2003). These data support the idea that there may be core sets of genes recalcitrant to frequent HGT, although their numbers may be small, and the composition of these sets may vary among bacterial lineages. Indeed, the same data used to infer high rates of gene transfer among genomes (Ochman *et al*., 2000) have been reanalysed to infer that not all classes of genes – here, using a functional classification scheme (Serres and Riley, 2000) – are found in proportional abundance among the newly acquired genes identified in numerous bacterial genomes (J. G. Lawrence, unpublished data). On the contrary, genes involved in 'information transfer' are rarely, if ever, identified as 'recently acquired', whereas genes encoding transporters and other more peripheral metabolic functions are highly represented in this group. These analyses support the idea that not all genes are transferred with equal likelihood among all lineages; this conclusion affects large-scale genomic analyses, such as supergene trees (Brown *et al*., 2001). In this example, the data set was reduced to 14 genes to generate a tree topology congruent with the rRNA phylogeny (Brown *et al*., 2001).

Yet the questions as to how many genes remain primarily recalcitrant to transfer, how many experience frequent HGT to escape loss, and the nature of the continuum between these two extremes remain unanswered in any quantitative fashion. More importantly, the issue as to how population structure and subdivision affect the likelihood of successful lateral transfer has only begun to be explored; a recent model shows that genes with low selective value are likely to be lost unless transferred into 'patchy' populations allowing local fixation (Berg and Kurland, 2002), consistent with the predictions of the neutral model of molecular evolution (Kimura, 1983). Because a gene's selection coefficient is a function of both its identity and its genomic (and hence ecological) context, assigning genes along a spectrum of 'readily transferred' to 'rarely transferred' becomes even more difficult, as is discussed below.

Lastly, one can add another layer of complexity by asking the transferred genes to provide a selective value during transit. Although this is not necessary if genes are introduced by transformation, many genes are introduced by transduction; indeed, the original conception of horizontally acquired 'pathogenicity islands' was intimately associated with bacteriophages (Barinaga, 1996). Bacteriophages are highly mosaic, and many contain genes typically thought to be 'bacterial' in origin (Pedulla *et al*., 2003). If a gene is in transit between bacterial genomes via a bacteriophage intermediate, it has a higher likelihood of completing the voyage successfully if it provides a useful function to the phage or to the prophage. Not all genes would satisfy this criterion. As a result, placing genes on an overall scale from 'nearly immobile' to 'highly transmissible' is a formidable task with a great number of variables to consider.

How does the role of HGT differ among different lineages?

As alluded to above, the rate of HGT of individual genes must also vary among bacterial lineages, owing to the different selective values they would impart in different genomic contexts. This constraint is obvious when examining the genomes of intracellular parasites and obligate pathogens, both of which are experiencing genome reduction (Andersson and Andersson, 1999a,b; Moran and Wernegreen, 2000; Cole *et al*., 2001). Here, organisms that experience strong declines in effective population size and/or rate of gene exchange by homologous recombination cannot retain the genes they currently possess, as their thresholds for effectively neutral mutations have increased (Lawrence, 2001). As a result, many of their ancestral genes cannot be retained as their benefits are insufficient to prevent their loss by mutation and genetic drift. In addition, their sheltered lifestyles limit access to the agents of HGT (bacteriophages, other bacteria with conjugative plasmids, etc.), also lowering the likelihood of gene acquisition. A comparison of insect endosymbionts shows remarkable genome stasis over 50 Myr (Tamas *et al*., 2002), including the lack of genes acquired by HGT.

Genome reduction can also play the opposite, more counterintuitive, role in affecting a lineage's propensity for participating in HGT. Although many of the lineages undergoing genome reduction will probably not give rise to descendents that undergo genome expansion, some will. For example, the *Mycoplasma pneumoniae* genome has significantly more DNA than its congener *Mycoplasma genitalium*. Much of the 'additional' DNA found in *M. pneumoniae* is atypical (Ochman *et al*., 2000), suggesting that a small genome has acquired new functions by HGT, and is thus experiencing genome expansion. Here, it is likely that the population size or recombination rate has increased so that the likelihood of retaining newly introduced genes has increased. More importantly, the

organism must be shifting into an ecological niche in which the newly acquired genes serve useful purposes. Therefore, genome reduction should not be viewed only as an indicator that a lineage will probably see reduced rates of HGT; some lineages may see enormous increases in their rates of HGT as they regain genes previously lost.

There are certainly other biological limitations to the free exchange of DNA between all taxa. First, transmissible agents have restricted ranges, e.g. bacteriophages have limited host ranges, as do many conjugative plasmids. Secondly, the apparatuses of transcription and translation become increasingly different with phylogenetic distance, imposing a barrier to facile gene exchange across large genetic distances. Only genes that provide large selective benefits would be retained after 'long-distance' transfer as their initial expression levels would be poor. This is perhaps most dramatically illustrated by the difficulty in transfer of bacterial genes into eukaryotes, where operons cannot be expressed by a native promoter at the site of insertion. Rather, the eukaryotic transcription and translation machineries require independent expression of each gene, thereby imposing a barrier to gene transfer beyond the necessity for transit of the DNA to the nucleus of a germline cell and its provision of a selectable function.

Data examining the effect of ecological niche on the propensity of gene exchange among cohabitants has had tantalising beginnings [e.g. among thermophilic Bacteria and Archaea (Nelson *et al*., 1999; Worning *et al*., 2000) or between Bacteria and fungi dwelling in the rumen (Garcia-Vallve *et al*., 2000b)], but remain largely unexplored; in addition, caveats can always be raised with regard to methods used to make these inferences (Logsdon and Fuguy, 1999). As gene exchange – by either transformation or transduction – does not require donors and recipients to cohabit, it is not clear how dwelling in the same physical environment increases gene flow by HGT. Moreover, the breadth of ecologies explored by individual 'species' is also a field of great interest but few data; preliminary work suggests that it may differ greatly among lineages (Gordon, 2001; Okada and Gordon, 2001; Gordon *et al*., 2002; Vogel *et al*., 2003), which is perhaps not unexpected. This variability makes taxon-to-ecology assignment difficult, if not infeasible, impractical and potentially misleading, e.g. witness the strong and well-documented ecological differences among strains of *Salmonella enterica* as a pathogen (Baumler *et al*., 2000; Rabsch *et al*., 2002).

Lastly, bacterial chromosomes themselves may have higher ordered structures that allow for proper replication termination and chromosome segregation. Such structures may be imparted by the asymmetric distribution of sequences arising naturally by strand-specific mutational biases (Lobry, 1996; Capiaux *et al*., 2001; Lobry and Sueoka, 2002; Lobry and Louarn, 2003). Unlike the factors discussed above, these sequence features can be examined quantitatively to test hypotheses in a rigorous fashion. We have examined such sequences in numerous taxa and have found that they are conserved only among phylogenetically related taxa (H. Hendrickson and J. G. Lawrence, unpublished results). The octomeric sequence shown in Fig. 2 displays a distribution indicative of participation in proper chromosome termination and segregation in its host *Mesorhizobium loti*, where it is counterselected from appearing on the 'improper' strand. This sequence, GGGCAGGG, has a similar distribution among closely related α-proteobacteria, but is found in

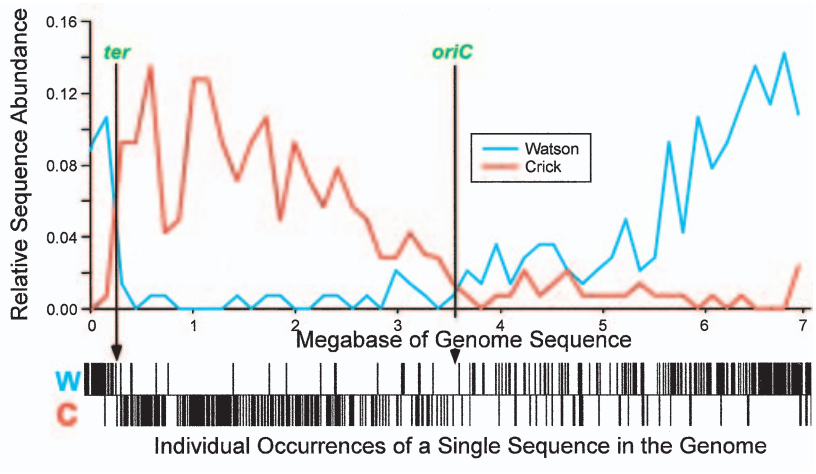

Fig. 2. Occurrences of a skewed, asymmetrically distributed sequence in the genome of the α-proteobacterium *Mesorhizobium loti*. The lower panel depicts each sequence on either the Watson (top) or Crick (bottom) strand as a hash mark. The abundance of this sequence on each strand is tabulated in the graph above; the origin and terminus of replication can be inferred from analyses of GC skew as well as the distribution of the octomer depicted here (H. Hendrickson and J. G. Lawrence, unpublished results). The origin was distinguished from the terminus by both the position of the *dnaA* gene (which is typically origin-proximal) and the orientation of *rrn* operons (typically transcribed away from the origin of replication). Comparable accumulation of octomeric sequences does not occur elsewhere in the genome, implying that this is not the result of chance. More importantly, analyses of di- and trinucleotide frequencies show that they are uniform across the genome, rejecting the hypothesis that the accumulation of these sequences is the result of significant alteration in mutation bias during replication.

great abundance on both strands in distantly related taxa, such as *Streptomyces coelicolor* (H. Hendrickson and J. G. Lawrence, unpublished results). If a DNA fragment were to be introduced into *Mesorhizobium* (or one of its relatives) from a donor taxon in which this sequence was abundant on both strands, the presence of the DNA would incur a selective detriment that could potentially offset any benefits provided by the newly acquired gene products. This barrier to gene exchange, unlike those discussed above, has only come to light with recent genomic analyses, which have both furthered our understanding of bacterial genome structure and shown us the depth of our ignorance.

The distribution of such sequences, as well as the other factors detailed above (Fig. 3A), would limit HGT frequency in a clade-specific fashion. That is, rates of HGT would be relatively high among closely related taxa, but would decrease in efficiency with phylogenetic distance by the accumulation of these numerous problematic differences (e.g. lack of proper ribosome binding sites, lack of proper promoter sequences, an excess of functionally biased sequences on the 'improper' DNA strand, etc.). If this is true, then HGT would have a profoundly different impact on phylogenetic reconstruction than the genetic panmixia that had previously been envisaged by many. Here, bacterial clades would be self-reinforcing, as most of the HGT would be occurring among more closely related taxa (Fig. 3B). As a result, one would detect fewer long-range transfers of highly conserved genes (Table 1), and many gene phylogenies would be congruent with that inferred from the rRNA sequences when examined at large phylogenetic scales. Hence, phylogenies based on gene content (Fitz-Gibbon and House, 1999; Snel *et al.*, 1999; Tekaia *et al.*, 1999) may reflect the propensity for HGT among more closely related lineages as much as the retention of their ancestral genes (Gogarten *et al.*, 2002).

How does one reach robust conclusions on the presence or absence of HGT?

The availability of multiple complete genome sequences has created the opportunity for unprecedented sophistication in phylogenetic analyses, wherein dendrograms are no longer derived from selected, and fortuitously available, DNA sequences, Rather, the entire body of information contained in the genomes can be brought to bear. Although this has solved some problems (such as poor taxon sampling or the necessity of using single gene sequences), it has created problems of its own as new methodologies have been developed to analyse genome sequences *en masse*. For example, does the creation of 'supergene' trees (Brown *et al.*, 2001) amplify weak phylogenetic signals at the expense of masking the signals of gene transfer? Moreover, the dynamics of gene loss and growth of paralogous gene families can obfuscate the identification of horizontally acquired genes and the inference of genome evolution (Jordan *et al.*, 2001; Snel *et al.*, 2002; Kunin and Ouzounis, 2003; Mirkin *et al.*, 2003), and some inferences are open to misinterpretation regarding the role of HGT (see below). Yet these works clearly show that the balance between gene loss and gene acquisition – by both lateral gene transfer and the expansion of pre-existing gene families – will also vary among lineages, making an overall assessment of the impact of gene transfer alone in genome evolution only one part of a complex process that we are only beginning to understand.

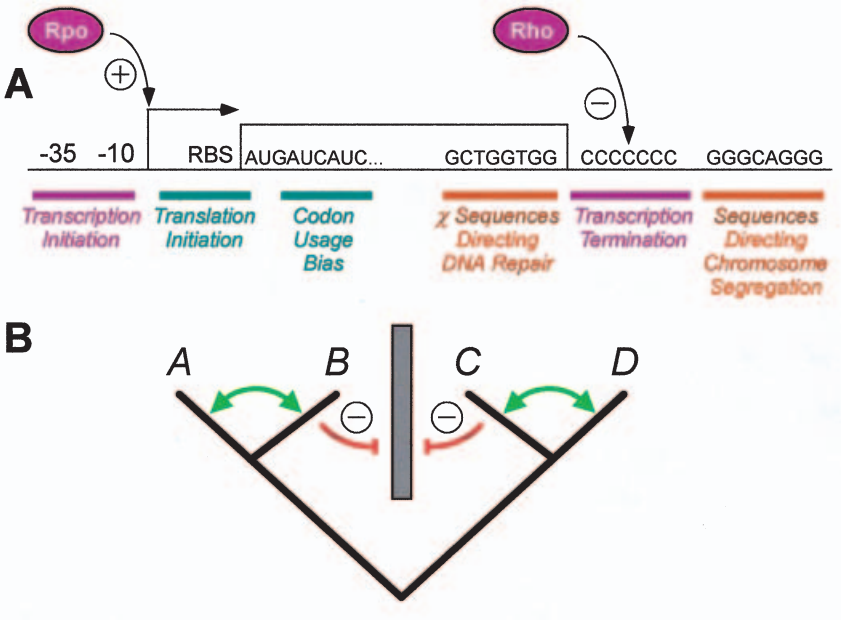

Fig. 3. Limitations on HGT among taxa.
A. Variable sequence features that can differ between taxa and decrease the likelihood of successful gene transfer, including those involved in transcription (magenta), translation (blue) and replication (orange).
B. A model in which the sequence features noted in (A) allow for more frequent transfer (green arrows) among more closely related organisms, but act as a barrier (albeit not an impervious one) to transfer between distantly related taxa (aborted red arrows).

There lies an even more pressing issue beneath the questions regarding the impact of gene-specific or taxon-specific variation in HGT on bacterial evolution: in many cases, it is difficult to ascertain with any degree of certainty whether HGT has or has not played a role in the evolutionary history of a gene. This lack of confidence stems from many sources, both those trivial to explain or correct and those that are more profoundly difficult to address (e.g. see Koonin *et al*., 2001). For example, genes probably affected by HGT have been identified by numerous methods in bacterial genomes, but these lists of 'alien' genes do not agree with each other (Ragan, 2001b). In this case, many of the discrepancies can be attributed either to statistical artifacts in the methods used or to the different classes of genes that each method was designed to detect (Lawrence and Ochman, 2002). In addition, parametric methods detecting atypical genes (presumably having evolved in a genome with different mutational biases) can lead to incorrect assignment of short open reading frames (ORFs) as being atypical (and potentially newly acquired) because of lack of data, and may be unable to identify genes recently transferred from taxa with similar mutational biases. These methods will ultimately fail to detect genes that were introduced long ago, as the mutational proclivities of their current host will ameliorate any atypical sequence features over time (Lawrence and Ochman, 1997; 1998).

Similarly, phylogenetic methods can be confounded by (i) the amplification of gene families in certain genomes, which interferes with the proper identification of orthologous genes; (ii) convergent evolution resulting from parallel phenotypic shifts (for example in the %GC content of the genome or in the thermal growth regime leading to predictable protein modifications); or (iii) phylogenetic artifacts such as variation in the rates of evolution between lineages or long branch length attraction (Stiller and Hall, 1999; Ragan 2001a,b; Simmons *et al*., 2002). Ultimately, phylogenetic methods will also fail, in this case when evolutionary changes have become so numerous as to overwhelm a useful phylogenetic signal, making inferences regarding HGT a challenge in navigating the vagaries of phylogenic reconstruction methodology, which can always be called into question.

What we find even more disturbing is the failure of most investigators examining HGT to reach a consensus as to what null hypothesis should be tested. That is, regardless of approach, how one phrases a scientific question can bias the conclusions. In the 'pregenomic era', it was assumed that genes were inherited vertically during cell division. Naturally, one tested the idea that a gene had been subject to HGT by stating vertical inheritance as the null hypothesis to be disproved by the weight of the data. If one could not disprove the null hypothesis, one then concluded that the gene was not subject to HGT. Yet one could just as easily begin with a null hypothesis whereby the genes being analysed had been subject to HGT and collect data to refute this hypothesis. Here, one would conclude that the gene was not subject to HGT only if one refuted the null hypothesis, rather than having this conclusion be the default condition upon failure to disprove an alternative null hypothesis. In most phylogenetic analyses, the first scenario is the *de facto* approach; yet, in many cases, the data are of insufficient quality – for the reasons outlined above – to make robust conclusions regardless of which null hypothesis is taken. That is, if neither null hypothesis can be rejected, robust conclusions cannot be made, and uncertainty must remain. This caveat is also applicable to the identification of putatively transferred genes by parametric approaches: the failure to identify a gene as atypical does not rule out the possibility that HGT has played a role in its evolution in this taxon. Is it fair to assume that genes have been inherited vertically and require evidence that HGT has played a role, rather than the converse?

How does one integrate HGT into the continuum of genetic exchange to arrive at meaningful microbiological concepts?

Exchange of DNA among bacterial taxa can occur between very closely related strains, where it is often termed 'recombination', and integration of DNA is mediated by homologous recombination (Guttman, 1997; Feil *et al*., 2001). Therefore, transfer of DNA between closely related taxa will be unlikely to result in a recombinant bearing two alleles of the same locus; rather, an orthologous replacement would occur. As sequence divergence increases, homologous recombination is precluded by the mismatch correction system (Zawadzki *et al*., 1995; Majewski and Cohan, 1999), and only an illegitimate or site-specific recombination event can introduce the DNA into the genome. If the sequences are closely related, one copy will be retained and the other lost by deletion, as the genes would probably not encode proteins that conferred sufficiently distinct functions to allow selection for retention of both copies. The probability of gene retention probably increases as sequence divergence between donor and recipient lineages increases, as more time would elapse for functional differences to arise (Fig. 4). However, more distantly related taxa would experience the barriers to HGT discussed above (see also Fig. 3A), thus reducing the probability of successful transfer. As a result, one can consider a 'zone of paralogy' where it is most likely that sequences introduced by HGT could be retained. This 'zone of paralogy' would also act to reinforce HGT clade identities initially established by common ancestry.

The 'zone of paralogy' also offers a cogent mechanism for the growth of gene families observed in many taxa

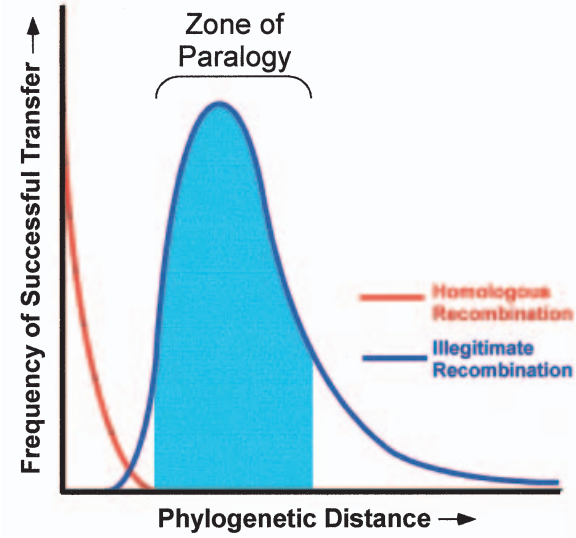

Fig. 4. The interplay between HGT mediated by homologous (red line) and illegitimate (blue line) recombination. Among closely related taxa, incoming DNA is likely to be integrated by homologous recombination, resulting in allelic replacement. More divergent sequences cannot recombine by this route, resulting in genomes with homologous genes; however, more distantly related homologues are more likely to be retained as paralogues, as they are more likely to confer separate functions (axes are depicted using arbitrary units). However, the factors shown in Fig. 3 decrease the overall frequency of HGT as taxa become more distantly related. The interplay between these effects results in a 'zone of paralogy', depicted in cyan, in which sequences are most likely to be retained.

(Jordan *et al*., 2001; Snel *et al*., 2002). The expansion of gene families by duplication and divergence of single genes within a single genome is an old idea, yet fraught with difficulty. Foremost among the difficulties is the problem of maintaining selection on both copies, thus preventing loss of the duplicated gene, until each gene develops functionally distinct roles. While clever schemes have been devised to circumvent these problems (e.g. see Stoltzfus, 1999; Lynch *et al*., 2001), differential function may arise while genes reside in different cytoplasms and experience different selective constraints. HGT would then reunite previous orthologues in the same genome, where they would appear as paralogues; this process alleviates the need for a period of co-existence of multiple copies of the same gene without selection for differential function (Lawrence, 2001; Gogarten *et al*., 2002). Therefore, one must consider carefully the mechanisms by which 'gene genesis' (Snel *et al*., 2002) occurs. Is HGT also playing a role here? Moreover, different rates of evolution among genes changes the taxonomic scope of organisms available for gene exchange by homologous recombination, and makes the 'zone of paralogy' vary in a gene-specific manner.

The rates of DNA exchange by homologous and illegitimate recombination are also intimately associated via the manner by which novel alleles are distributed in a population. If recombination among strains in a population is rare, then novel alleles arising by HGT are more likely to be lost by genetic drift than those able to be transmitted by homologous recombination. Therefore, increases in the rate of homologous recombination within populations serve not only to decrease the threshold of an effectively neutral mutation (increasing the likelihood of HGT; Lawrence, 2001) but also to disseminate newly acquired genes and prevent their stochastic loss (Berg and Kurland, 2002). Yet the introduction of novel alleles by HGT will also allow for niche-specific adaptation, which will eventually lead to bacterial 'speciation' (Cohan, 2001; Lawrence, 2002). Certain recombination events – those that disrupt such niche-specific loci – will produce less fit offspring, leading to reproductive isolation at chromosomal loci surrounding genes introduced by HGT (Lawrence, 2002).

One can view the interplay of gene exchange by these mechanisms as effectively blurring the lines between microbial taxa, making it difficult to delineate microbial 'species' or groupings at higher taxonomic levels. It is difficult to apply the biological species concept, as have Dykhuizen and Green (1991), to groups of strains that are reproductively isolated at some loci and not others. Similarly, the variable domains of exchange among taxa at different levels of inclusiveness, as well as the variable rates of exchange among different genes, makes higher ordered taxonomic classification difficult to quantify as well. As discussed previously (Gogarten *et al*., 2002), if higher ordered taxonomy is dictated by both the presence of ancestral genes (as is the case in eukaryotes) as well as biased HGT within taxonomic groups, then bacterial taxonomy reflects both history (the patterns of speciation events) as well as ongoing processes (HGT). Hence, the conclusions of Zuckerkandl and Pauling (1965), that genes are documents of evolutionary history, becomes far more complex as we integrate patterns of gene exchange – and lineage-specific gene loss – with histories of vertical inheritance.

Conclusions

Woese (2002) postulated that HGT was rampant early in microbial evolution, but plays a smaller role now, after passage through the 'Darwinian threshold'. Although the arguments that the role played by HGT differs now from the roles played in ancient lineages are compelling, it is still clear that HGT can be a potent process in microbial diversification. The questions remain as to how its impact can be quantified in lineages and genes of interest, and how these data can be integrated into a holistic understanding of how gene exchange mediates evolutionary change.

Answers are likely to come from multiple sources, including the accumulation of additional data that will allow for more conclusive identification of orthologues among distantly related taxa, the development of more robust methods for phylogenetic inference that can be used on large data sets, integration of methods used to detect atypical genes and methods used to detect genes with aberrant phylogenetic histories, and the continued integration of the numerous evolutionary forces acting on genome evolution. More importantly, these advances must be accompanied by a holistic change in mindset among microbiologists. Critical, thoughtful evaluation and interpretation of all available data can assist in making inferences and conclusions that help to clarify, rather than confound, these complex biological issues. Only in this way can horizontal gene transfer be discussed as a topic with a firm foundation in fact, rather than as a collection of anecdotes and seemingly arcane analyses.

Acknowledgements

This work was supported by grant MCB-0217278 from the National Science Foundation.

References

Andersson, J.O., and Andersson, S.G. (1999a) Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* **9:** 664–671.

Andersson, J.O., and Andersson, S.G. (1999b) Genome degradation is an ongoing process in *Rickettsia*. *Mol Biol Evol* **16:** 1178–1191.

Barinaga, M. (1996) A shared strategy for virulence. *Science* **272:** 1261–1263.

Baumler, A.J., Hargis, B.M., and Tsolis, R.M. (2000) Tracing the origins of *Salmonella* outbreaks. *Science* **287:** 50–52.

Berg, O.G., and Kurland, C.G. (2002) Evolution of microbial genomes: sequence acquisition and loss. *Mol Biol Evol* **19:** 2265–2276.

Brochier, C., Philippe, H., and Moreira, D. (2000) The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet* **16:** 529–533.

Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E., and Stanhope, M.J. (2001) Universal trees based on large combined protein sequence data sets. *Nature Genet* **28:** 281–285.

Capiaux, H., Cornet, F., Corre, J., Guijo, M., Perals, K., Rebollo, J.E., and Louarn, J. (2001) Polarization of the *Escherichia coli* chromosome. A view from the terminus. *Biochimie* **83:** 161–170.

Cohan, F.M. (2001) Bacterial species and speciation. *Syst Biol* **50:** 513–524.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., *et al.* (2001) Massive gene decay in the leprosy bacillus. *Nature* **409:** 1007–1011.

Daubin, V., Moran, N.A., and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* **301:** 829–832.

Davies, J. (1996) Origins and evolution of antibiotic resistance. *Microbiologia* **12:** 9–16.

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284:** 2124–2129.

Doolittle, R.F., and Handy, J. (1998) Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Curr Opin Genet Dev* **8:** 630–636.

Doolittle, R.F., Feng, D.F., Anderson, K.L., and Alberro, M.R. (1990) A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. *J Mol Evol* **31:** 383–388.

Doolittle, W.F., Boucher, Y., Nesbo, C.L., Douady, C.J., Andersson, J.O., and Roger, A.J. (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Phil Trans R Soc London B Biol Sci* **358:** 39–57.

Dykhuizen, D.E., and Green, L. (1991) Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* **173:** 7257–7268.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P., Enright, M.C., *et al.* (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* **98:** 182–187.

Fitz-Gibbon, S.T., and House, C.H. (1999) Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res* **27:** 4218–4222.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Garcia-Vallve, S., Romeu, A., and Palau, J. (2000a) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res* **10:** 1719–1725.

Garcia-Vallve, S., Romeu, A., and Palau, J. (2000b) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* **17:** 352–361.

Gogarten, J.P. (1995) The early evolution of cellular life. *Trends Ecol Evol* **10:** 147–151.

Gogarten, J.P., Starke, T., Kibak, H., Fishman, J., and Taiz, L. (1992) Evolution and isoforms of V-ATPase subunits. *J Exp Biol* **172:** 137–147.

Gogarten, J.P., Murphey, R.D., and Olendzenski, L. (1999) Horizontal gene transfer: pitfalls and promises. *Biol Bull* **196:** 359–361.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19:** 2226–2238.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* **147:** 1079–1085.

Gordon, D.M., Bauer, S., and Johnson, J.R. (2002) The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology* **148:** 1513–1522.

Guttman, D.S. (1997) Recombination and clonality in natural populations of *Escherichia coli*. *Trends Ecol Evol* **12:** 16–22.

Hall, R.M. (1997) Mobile gene cassettes and integrons: moving antibiotic resistance genes in gram-negative bacteria. *Ciba Found Symp* **207:** 192–202.

Hayes, W.S., and Borodovsky, M. (1998) How to interpret an

anonymous bacterial genome: machine learning approach to gene identification. *Genome Res* **8:** 1154–1171.

Ibba, M., Bono, J.L., Rosa, P.A., and Soll, D. (1997) Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc Natl Acad Sci USA* **94:** 14383–14388.

Jain, R., Rivera, M.C., and Lake, J.A. (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96:** 3801–3806.

Jordan, I.K., Makarova, K.S., Spouge, J.L., Wolf, Y.I., and Koonin, E.V. (2001) Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res* **11:** 555–565.

Karlin, S. (1998) Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr Opin Microbiol* **1:** 598–610.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Klenk, H.P., Meier, T.D., Durovic, P., Schwass, V., Lottspeich, F., Dennis, P.P., and Zillig, W. (1999) RNA polymerase of *Aquifex pyrophilus*: implications for the evolution of the bacterial *rpoBC* operon and extremely thermophilic bacteria. *J Mol Evol* **48:** 528–541.

Koonin, E.V., Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55:** 709–742.

Kunin, V., and Ouzounis, C.A. (2003) The balance of driving forces during genome evolution in prokaryotes. *Genome Res* **13:** 1589–1594.

Kurland, C.G. (2000) Something for everyone. *EMBO Rep* **1:** 92–95.

Lawrence, J.G. (1997) Selfish operons and speciation by gene transfer. *Trends Microbiol* **5:** 355–359.

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

Lawrence, J.G. (2001) Catalyzing bacterial speciation: correlating lateral transfer with genetic headroom. *Syst Biol* **50:** 479–496.

Lawrence, J.G. (2002) Gene transfer in bacteria: speciation without species? *Theor Pop Biol* **61:** 449–460.

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

Lawrence, J.G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95:** 9413–9417.

Lawrence, J.G., and Ochman, H. (2002) Reconciling the many faces of gene transfer. *Trends Microbiol* **10:** 1–4.

Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics* **143:** 1843–1860.

Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol* **184:** 4891–4905.

Levin, B. (1981) Periodic selection, infectious gene exchange, and the genetic structure of *E. coli* populations. *Genetics* **99:** 1–23.

Lobry, J.R. (1996) Asymetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13:** 660–665.

Lobry, J.R., and Louarn, J.M. (2003) Polarisation of prokaryotic chromosomes. *Curr Opin Microbiol* **6:** 101–108.

Lobry, J.R., and Sueoka, N. (2002) Asymmetric directional mutation pressures in bacteria. *Genome Biol* **3:** RESEARCH0058.0051–0058.0014.

Logsdon, J.M., and Fuguy, D.M. (1999) *Thermotoga* heats up lateral gene transfer. *Curr Biol* **9:** R747–R751.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., *et al.* (1998) Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19:** 554–568.

Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* **159:** 1789–1804.

Majewski, J., and Cohan, F.M. (1999) DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* **153:** 1525–1533.

Makarova, K.S., Aravind, L., Galperin, M.Y., Grishin, N.V., Tatusov, R.L., Wolf, Y.I., and Koonin, E.V. (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res* **9:** 608–628.

Makarova, K.S., Ponomarev, V.A., and Koonin, E.V. (2001) Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol* **2:** 1–14.

Médigue, C., Rouxel, T., Vigier, P., Hénaut, A., and Danchin, A. (1991) Evidence of horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* **222:** 851–856.

Mirkin, B.G., Fenner, T.I., Galperin, M.Y., and Koonin, E.V. (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol Biol* **3:** 2.

Moran, N.A., and Wernegreen, J.J. (2000) Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* **15:** 321–326.

Mylvaganam, S., and Dennis, P.P. (1992) Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*. *Genetics* **130:** 399–410.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399:** 323–329.

Nesbø, C.L., Boucher, Y., and Doolittle, W.F. (2001) Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. *J Mol Evol* **53:** 340–350.

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

Okada, S., and Gordon, D.M. (2001) Host and geographical factors influence the thermal niche of enteric bacteria isolated from native Australian mammals. *Mol Ecol* **10:** 2499–2513.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113:** 171–182.

Rabsch, W., Andrews, H., Kingsley, R.A., Prager, R., Tschape, H., Adams, L.G., and Baumler, A.J. (2002) *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infect Immun* **70:** 2249–2255.

Ragan, M.A. (2001a) Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev* **11:** 620–626.

Ragan, M.A. (2001b) On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol Lett* **201:** 187–191.

Serres, M.H., and Riley, M. (2000) MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb Comp Genomics* **5:** 205–222.

Simmons, M.P., Randle, C.P., Freudenstein, J.V., and Wenzel, J.W. (2002) Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. *Mol Biol Evol* **19:** 14–23.

Snel, B., Bork, P., and Huynen, M. (1999) Genome phylogeny based on gene content. *Nature Genet* **21:** 108–110.

Snel, B., Bork, P., and Huynen, M.A. (2002) Genomes in flux: the evolution of Archaeal and proteobacterial gene content. *Genome Res* **12:** 17–25.

Stiller, J.W., and Hall, B.D. (1999) Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol* **16:** 1270–1279.

Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J Mol Evol* **49:** 169–181.

Tamas, I., Klasson, L., Näslund, K., Eriksson, A.-S., Canbäck, B.J.W.J., Sandström, J.P., *et al.* (2002) Fifty million years of genomic stasis in endosymbiotic bacteria. *Science* **296:** 2376–2379.

Tekaia, F., Lazcano, A., and Dujon, B. (1999) The genomic tree as revealed from whole proteome comparisons. *Genome Res* **9:** 550–557.

Vogel, J., Normand, P., Thioulouse, J., Nesme, X., and Grundmann, G.L. (2003) Relationship between spatial and genetic distance in *Agrobacterium* spp. in 1 cubic centimeter of soil. *Appl Environ Microbiol* **69:** 1482–1487.

Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51:** 221–271.

Woese, C.R. (2002) On the evolution of cells. *Proc Natl Acad Sci USA* **99:** 8742–8747.

Woese, C.R., Olsen, G.J., Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64:** 202–236.

Worning, P., Jensen, L.J., Nelson, K.E., Brunak, S., and Ussery, D.W. (2000) Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*. *Nucleic Acids Res* **28:** 706–709.

Yap, W.H., Zhang, Z., and Wang, Y. (1999) Distinct types of rRNA operons exist in the genome of the Actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol* **181:** 5201–5209.

Zawadzki, P., Roberts, M.S., and Cohan, F.M. (1995) The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* **140:** 917–932.

Zuckerkandl, E., and Pauling, L. (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8:** 357–366.