Prokaryotic Evolution in Light of Gene Transfer

J. Peter Gogarten, W. Ford Doolittle,† and Jeffrey G. Lawrence‡*

*Department of Molecular and Cell Biology, University of Connecticut; †Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada; and ‡Department of Biological Sciences, University of Pittsburgh

Accumulating prokaryotic gene and genome sequences reveal that the exchange of genetic information through both homology-dependent recombination and horizontal (lateral) gene transfer (HGT) is far more important, in quantity and quality, than hitherto imagined. The traditional view, that prokaryotic evolution can be understood primarily in terms of clonal divergence and periodic selection, must be augmented to embrace gene exchange as a creative force, itself responsible for much of the pattern of similarities and differences we see between prokaryotic microbes. Rather than replacing periodic selection on genetic diversity, gene loss, and other chromosomal alterations as important players in adaptive evolution, gene exchange acts in concert with these processes to provide a rich explanatory paradigm—some of whose implications we explore here. In particular, we discuss (1) the role of recombination and HGT in giving phenotypic ''coherence'' to prokaryotic taxa at all levels of inclusiveness, (2) the implications of these processes for the reconstruction and meaning of ''phylogeny,'' and (3) new views of prokaryotic adaptation and diversification based on gene acquisition and exchange.

As prokaryotes, Bacteria and Archaea propagate themselves primarily by binary fission. Cell fusion and recombination are not necessary steps in their reproduction, unlike in the reproduction of complex eukaryotes. As a result, early models for understanding adaptation, evolution, and speciation in these organisms often focused on clonality and periodic selection (Levin 1981). According to such models, all individuals within a species resemble each other because they descend from a single ancestor that bested its siblings by virtue of some beneficial mutation (or sequence of mutations)—fixing not only the favored mutation but the entire genome in which it first occurred. (Microbiologists vigorously debate the *applicability* of species concepts developed by animal and plant biologists, as if the concepts themselves were clear [Ward 1998; Cohan 2001; Lawrence 2001, 2002]. In fact even for organisms with regular and obligatory recombination and obvious barriers to intertaxon mating, the notion of species is onerous [Wilson 1999]. Here, we use ''species'' to designate assemblages of related organisms to which microbiologists have attached specific names, rather than natural kinds). Thus, earlier thinking, as summarized by Levin and Bergstrom (2000), was that ''adaptive evolution will proceed by the sequential accumulation of favorable mutations, rather than by recombinational generation of gene combinations; in this respect bacterial evolution will be similar to that depicted in the top portion of Muller's famous diagram of evolution in asexual and sexual populations.''

Decay of Clonality: Role of Homologous Recombination

Bacterial population geneticists have known for some time that prokaryotic genomes *do* sometimes recombine (Guttman and Dykhuizen 1994), but early estimates suggested that rates of recombination were sufficiently low to be ignored when considering periodic selection events (Cohan 1994*a*, 1994*b*). Expanding multilocus sequence typing (MLST) surveys now show that it is often homologous recombination—not the stepwise accumulation of mutations after separation of lineages—that accounts for the lion's share of sequence differences between isolates. Feil et al. (2001), in a study of conserved loci in bacterial pathogens, conclude for lineages within a species that ''over the long term, the impact of relatively frequent recombination is to obliterate the phylogenetic signal in gene trees such that the relationships between major lineages of many bacterial species should be depicted as a network rather than a tree.'' The genomes of individuals within such a species would thus resemble each other because of frequent exchange of genes and parts of genes via a common gene pool and not (except indirectly) because they share a common ancestor. The degree to which homologous recombination abrogates clonal history depends, of course, on which group is examined.

Gratifyingly, the bacteria studied by Feil and collaborators (including *Neiserria menigitidis*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, and *Staphylococcus aureus*) thus conform to the first of two operational criteria proposed by Dykhuizen and Green (1991) for the recognition of bacterial species. Starting from the well-known ''biological species concept'' (Mayr 1942, 1963), these authors observed that, because of recombination, ''phylogenies of different genes from individuals of the same species should be significantly different, whereas the phylogeny of genes from individuals of different species should not be significantly different'' (fig. 1). That is, frequent recombination should result in conflicting molecular phylogenies for genes in conspecific organisms. In contrast, molecular phylogenies for different genes in different species should be congruent if interspecies recombination is infrequent. However, recombination across species boundaries—howsoever defined—does occur much more frequently than envisioned by Dykhuizen and Green, producing incongruent phylogenies between species as well as within them (Smith et al. 1999).

Indeed it is not clear that any evolved barriers to intergroup exchange (other than those effective against

Key words: lateral gene transfer, horizontal gene transfer, bacterial speciation, recombination, niche.

Address for correspondence and reprints: Jeffrey G. Lawrence, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260. E-mail: jlawrenc@pitt.edu.

Mol. Biol. Evol. 19(12):2226–2238. 2002
 ISSN: 0737-4038

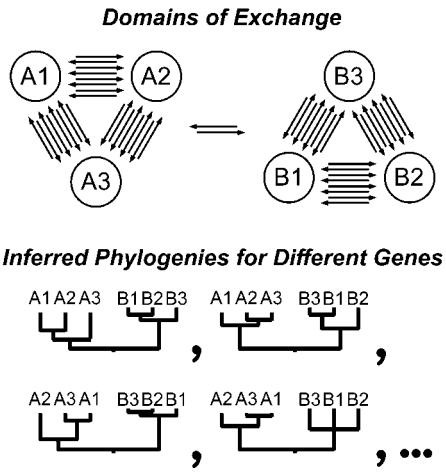

FIG. 1.—Gene transfer will obliterate patterns of vertical descent within groups that exchange genes at high frequency, producing discordant relationship among genes with different ancestries within the same cells.

lethal viruses and parasitic genetic elements, such as restriction-modification systems [Arber 1979]) *should* exist in prokaryotes. For animals that must recombine to reproduce, selection disfavors interspecific matings which, almost by definition, produce unfit or no progeny. In contrast, even the most promiscuous prokaryotes experience recombination much less frequently than they reproduce, and the exchange involves only a tiny fraction of their genomes in any one event. There seems very little selective advantage in preventing such rare interspecific exchange. Furthermore, many of the agents of exchange (bacteriophages, transmissible and conjugative plasmids) are themselves best viewed as selfish elements. For them, interspecific transfer is selectively advantageous and might even be required for long-term persistence (Goddard and Burt 1999).

Homologous recombination is, to be sure, strongly constrained by degree of sequence difference and the nature of the machinery involved (Vulic et al. 1997). Many careful studies show, not unexpectedly, that the ease with which genes recombine declines dramatically as their sequences diverge (Zawadzki, Roberts, and Cohan 1995; Majewski and Cohan 1998, 1999; Wolf et al. 2001). This constraint might be taken as a barrier to interspecific exchange and could be used as an upper limit in the delineation of a microbial species. Yet the mismatch correction system (the principal obstacle to heterologous exchange) provides at best a leaky and imprecise barrier. Some genes (and some parts of all genes) are more conserved in sequence than others and could potentially be exchanged among broader groups of organisms. Ironically, the barrier for highly conserved ribosomal RNA genes should be among the leakiest, whereas genes with rapidly changing sequences (such as those under diversifying selection) should observe tighter limits. Furthermore, recombinational barriers imposed by the mismatch repair system are abrogated in cognate mutants. There is an appealing theory that such mutants play a key role in adaptation via recombination, and the mismatch repair genes themselves show a complex history of recombination attributed to loss and recovery (through horizontal transfer) (Vulic, Lenski, and Radman 1999; Denamur et al. 2000). Evolution in these contexts (Oliver et al. 2000; Denamur et al. 2002) illustrates how both mutational and recombinatorial processes can play important roles in adaptive evolution. Homologous recombination of large fragments of DNA may also be impeded by barriers imposed by restriction endonucleases. Yet these barriers, which change at a very rapid pace, are not correlated with the degree of overall sequence divergence (Wilson and Murray 1991); and they do not preclude DNA transfer but only limit the size of transferred fragments (McKane and Milkman 1995).

Of course natural selection will act as the arbiter of success for all recombinant cells. That is, the evolutionary importance of recombinatorial events will depend on the probability that the products of gene exchange offer selective advantages. If recombination has introduced maladaptive changes, eliminated niche-specific information, or disrupted coadapted alleles, then recombinant progeny will be counterselected (however see below). Therefore, ecological differentiation may impose a selective constraint on facile genetic exchange even in the absence of any mechanistic barriers imposed by the mismatch correction system.

Decay of Clonality: Role of Horizontal Gene Transfer

Horizontal, or lateral, gene transfer (HGT) is different, both in mechanism and in impact. Barriers to homologous (legitimate) recombination do not preclude its occurrence—even between very distantly related organisms—because numerous illegitimate means are available for integrating foreign DNA into the genome (Ochman, Lawrence, and Groisman 2000). HGT can occur between even very distantly related organisms, e.g., between bacteria and plants or fungi (Heinemann and Sprague 1989; Garcia-Vallve, Romeu, and Palau 2000). The impact of such horizontal transfer is that molecular phylogenies calculated for different molecules from the same set of species, while often agreeing in broad outline (e.g., Ludwig et al. 1998), are only rarely completely congruent (Gogarten et al. 1992; Gogarten 1995). A decade ago, evolutionary biologists were hesitant to invoke HGT as an explanation for these discrepancies. Then a few cases in which a simple bifurcating tree was clearly an insufficient evolutionary metaphor were recognized (Gogarten 1995). Now, complete genome sequences offer an abundance of evidence for HGT (table 1) and highlight its confounding effects in reconstructing the history of organismal evolution (Koonin et al. 1997; Doolittle 1999*b*; Nelson et al. 1999; Pennisi 1999; Koonin, Aravind, and Kondrashov 2000; Boucher, Nesbo, and Doolittle 2001; Nesbo, Boucher, and Doolittle 2001; Zhaxybayeva and Gogarten 2002).

Detecting HGT

Methods for collecting evidence of potential gene transfer events generally fall into two classes. Phylo-

Table 1
Examples for Phylogenetic Incongruities Likely Resulting from HGT

Protein or RNA	Phylogenetic Incongruities	References
Information transfer		
Ribosomal RNA (*rrn*)	(1) *Thermomonospora* contains *rrn* operon donated from *Thermobispora* (2) *Haloarcula* contains *rrn* operon from likely halobacterial donor	(Mylvaganam and Dennis, 1992; Yap, Zhang, and Wang, 1999)
Ribosomal protein L32 (RpmF)	*Lactococcus lactis* groups within Proteobacteria	(Makarova, Ponomarev, and Koonin, 2001)
Ribosomal protein L33 (RpmG)	(1) *Deinococcus* groups with *Aquifex* instead of *Thermus* (2) *Mycobacterium leprae* groups separate from *Mycobacterium tuberculosis*	(Makarova, Ponomarev, and Koonin, 2001)
Ribosomal protein S14 (RpsN)	(1) Mycoplamas are separate from other low-GC gram-positive bacteria (2) *Deinococcus* is separated from *Thermus* and groups with some low-GC gram-positive bacteria	(Brochier, Philippe, and Moreira, 2000)
Ribosomal protein S18 (RpsR)	Three Mycoplasmatales species group with ϵ-proteobacteria	(Makarova, Ponomarev, and Koonin, 2001)
Elongation factor Tu (TufB)	Streptococcaceae group with Enterococcaceae	(Ke et al., 2000)
Lysyl-tRNA synthase	*Borrelia* groups with Archaea	(Ibba et al., 1997)
Phenylalanyl-tRNA synthase	Spirochettes group with Archaea	(Woese et al., 2000)
Prolyl-tRNA synthase	(1) *Deinococcus*, *Mycoplasma* and *Borrelia* group with the Archaea (2) *Borrelia* does not group with the spirochete *Treponema*, which remains within the Bacterial clade	(Gogarten, Murphey, and Olendzenski, 1999; Woese et al., 2000)
Seryl-tRNA synthase	The archaean *Haloarcula* groups with Bacteria	(Doolittle and Handy, 1998; Woese et al., 2000)
Mismatch repair (MutL/S)	*Methanosarcina mazei* groups with Bacteria	(Deppenmeier et al., 2002)
DNA polymerase IV	*Methanosarcina mazei* groups with Bacteria	(Deppenmeier et al., 2002)
Significant metabolism		
3-Hydroxy-3-methylglutaryl coenzyme A reductase	The archaeoglobales and *Thermoplasma* (Archaea) group with pseudomonads (Bacteria)	(Boucher et al., 2001)
APS reductase	Syntrophobacteraceae and Nitrospinaceae group with gram-positive bacteria	(Friedrich 2002)
A/F–ATPase	(1) *Deinococcus* and *Borrelia* group with the Archaea (2) The crenarchaeote *Desulfurococcus mobilis* groups with Euryarchaeota	(Shibui et al., 1997; Olendzenski et al., 2000; Senejani, Hilario, and Gogarten 2001)
Catalase-peroxidase	γ-Proteobacteria group with Archaea	(Faguy and Doolittle, 2000)
Cytochrome *c* biogenesis	Transfer between *Deinococcus*, Proteobacteria and Archaea (polarity of transfer unclear)	(Kranz and Goldman, 1998)
Dissimilatory sulfite reductase (DsrAB)	*Desulfitobacterium* (low-GC gram-positive) and *Thermodesulfobacterium* both group with δ-proteobacteria	(Klein et al., 2001)
Glucosyl hydrolases	Fungi group with Bacteria	(Garcia-Vallve, Romeu, and Palau, 2000)
Glutamate synthase	Multiple transfers within and between the Bacteria and the Archaea	(Nesbo et al., 2001)
Glutamine synthetase (GSI)	Low-GC gram-positive bacteria group with Euryarchaeotes	(Pesole et al., 1995)
Glutamine synthetase (GSII)	Multiple transfers within the rhizobia	(Turner and Young, 2000)
Glyceraldehyde-3-phosphate dehydrogenase (GapA)	*Escherichia coli* and other enteric bacteria group within the eukaryotes	(Doolittle et al., 1990)
GroES/GroEL chaperone	*Methanosarcina mazei* groups with Bacteria	(Deppenmeier et al., 2002)
HSP70 (DnaK)	Archaeal lineages are dispersed among Bacteria	(Gupta and Golding, 1993; Gogarten, Hilario, and Olendzenski, 1996; Roger and Brown, 1996)
Multiple enzymes for catalysis and cofactor synthesis for H_4MPT-mediated methyl-group transfer	*Methylobacterium extorquens*, an aerobic methyltrophic proteobacterium, clusters tightly with Archaeal methanogens; parsimony argues strongly that the lack of methanogenic apparatus from other lineages reflects recent transfer rather than large numbers of independent losses	(Chistoserdova et al., 1998)

Table 1
Continued.

Protein or RNA	Phylogenetic Incongruities	References
Myo-inositol-1-phosphate synthase ...	Multiple transfers within and between the Bacteria and the Archaea	(Nesbo et al., 2001)
Phospglucose isomerase (PGI)	*Escherichia coli* and *Haemophilus* group with eukaryotes	(Katz, 1996)
Proline metabolism (3 proteins)	*Methanosarcina mazei* groups with Bacteria	(Deppenmeier et al., 2002)
Ribulose-bisphosphate carboxylase (RuBisCo)	β-Proteobacterium *Nitrosomonas europaea* groups with γ-proteobacteria	(Utaker et al., 2002)

genetic methods look for atypical distributions of genes across organisms and may include the identification of (1) genes with highly restricted distributions, present in isolated taxa but absent from closely related species, (2) genes with an unduly high level of similarity to genes found in otherwise unrelated taxa, and (3) genes whose phylogenetic relationships are not congruent with the relationships inferred from other genes in their respective genomes (Doolittle 1999*a*, 1999*b*, 2000; Olendzenski et al. 2000; Lawrence 2001). Phylogeny-independent methods seek to identify genes that appear aberrant in their current genomic context, likely reflecting long-term evolution in genomes with different mutational biases. These methods examine nucleotide and dinucleotide frequencies (Karlin and Burge 1995; Lawrence and Ochman 1997, 1998), codon usage bias (Karlin, Mrazek, and Campbell 1998; Moszer, Rocha, and Danchin 1999; Karlin and Mrazek 2000), or patterns extracted by Markov chain analyses (Hayes and Borodovsky 1998).

While each of these methods has been used to infer that substantial portions of different genomes have arisen by HGT, they examine different properties of the genomes, identify different subsets of genes, and therefore are appropriate for testing different sorts of hypotheses (Ragan 2001*b*; Lawrence and Ochman 2002). Lawrence and Ochman (1997) proposed that at least 15% of the *Escherichia coli* genome is atypical and may have arisen by recent gene transfer, while Nelson et al. (1999) concluded that nearly 25% of *Thermotoga maritima* genes are most closely related to Archaeal genes and bespeak a history of gene transfers between these lineages. These estimates may be low: methods detecting atypical sequences fail to identify ancient transfer events, while phylogenetic methods rely upon robust sampling of potential donor lineages. While different genes may be transferred with different propensities (Jain, Rivera, and Lake 1999; Makarova et al. 1999; Graham et al. 2000; Zhaxybayeva and Gogarten 2002), no gene appears immune to HGT. Genes encoding core metabolic functions (Doolittle et al. 1990; Olendzenski et al. 2000), conserved biosynthetic pathways (Kranz and Goldman 1998; Boucher et al. 2001), components of the transcription and translation machinery (Ibba et al. 1997; Wolf et al. 1999; Brochier, Philippe, and Moreira 2000; Woese et al. 2000), and even ribosomal RNA (Yap, Zhang, and Wang 1999) have been subject to HGT.

Organismal Phylogeny: Remnant of Vertical Inheritance or Barometer of HGT?

HGT leads to genomes whose constituent genes have different evolutionary histories. Can one retain the concept of a single organismal lineage in the face of apparently frequent HGT, or is this concept fatally flawed? This is a nontrivial issue; various ad hoc, gene-based operational definitions of organismal relationships have been proposed (Fitz-Gibbon and House 1999; Snel, Bork, and Huynen 1999; Tekaia, Lazcano, and Dujon 1999; Brown et al. 2001) and may have utility, but use of the term phylogeny in this context may be inappropriate. Gene-content phylogenies (see below) might be more properly viewed as taxonomies or phenetic classifications, while the equation of organismal phylogeny with the genealogy of only a small fraction of any organism's genes is, at least, a radical departure from traditional practice. We have each discussed these philosophical issues elsewhere and here concern ourselves only with quantitative and qualitative effects of HGT on genome history.

Even here there is controversy; present viewpoints (Doolittle 1999*a*, 1999*b*, 2000; Woese 2000) form a continuum between two extremes. In the conservative view, most transfers take place between closely related organisms, and the transfer rate between divergent organisms is low. In this case, most molecular phylogenies will agree in their overall topology. Lineages might be ''fuzzy lines'' (Woese 2000), but the larger evolutionary pattern would be reflected in the majority consensus (Martin 1999). Genes transferred between more divergent species would be easily detected as conflicts with this consensus. (If reality is close to this model, dominated by vertical inheritance, then interdomain and interphylum HGT events promise to provide an excellent means of correlating evolutionary events in the different parts of the tree of life. For example, the origin of the cyanobacteria must predate the acquisition of chloroplast by early eukaryotes.)

The radical construction of HGT envisions high rates of gene transfer even between divergent organisms. If partners for transfer were randomly chosen among different taxa, then no congruent topologies should emerge from different molecular phylogenies. However, if partners were chosen nonrandomly, then patterns deduced from molecular phylogenies will reflect propen-

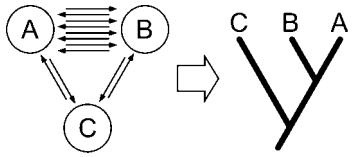

FIG. 2.—Gene transfer can create patterns of similarity and difference that mimic patterns produced by vertical descent. If taxa A and B successfully exchange genetic information (by homologous recombination or HGT) more frequently with each other than with taxon C, they will come to resemble each other more closely than they do C, both in gene content and gene sequence. Treelike patterns based on gene content or sequence will reflect these different frequencies, not some underlying organismal phylogeny. Frequency of successful exchange between taxa will depend on (1) propinquity, (2) metabolic compatibility, (3) adaptations to their abiotic environment, (4) gene expression systems, and (5) gene transfer mechanisms. As these factors change, patterns of relationship (apparent organismal phylogeny) based on gene content or sequence will also change. Deep branching (as of taxon C) may reflect genetic isolation, not early divergence. Conversely, ''distantly related'' taxa that begin to exchange genes very frequently will ultimately resemble ''sister'' taxa.

sities for gene transfer rather than vertical inheritance (fig. 2). Ironically, such preferential gene exchange could create many of the very same patterns of similarity and difference we usually attribute to vertical inheritance. Under this model, α-proteobacteria are more similar to other α-proteobacteria in gene content (or gene sequence) because they exchange genes more frequently with other α-proteobacteria than with β-, γ-, or δ-proteobacteria, cyanobacteria are self-similar because they most frequently exchange genes with each other, and so forth. Here, recognized taxonomic categories would be created exclusively through likelihood of HGT. Any taxon that began exchanging genes with α-proteobacteria would eventually be recognized as an α-proteobacterium. Similarly, lineages that adapt to an ecological niche that decreases HGT will become isolated from their surrounding lineages and will be recovered as deeply branching clades in most molecular phylogenies.

Under this scenario, one might consider the possibility that molecular phylogenies place extremely thermophilic bacteria as the oldest bacterial lineages because they live in an environment where most of the available genes are from Archaea and where they can participate less in HGT with other bacteria. Biochemical and physiological changes can also lead to genetic isolation and thus alter an organism's apparent position in trees based on gene content or sequence. For instance, perhaps the novel transcriptional apparatus of the Archaea could have made it less likely for them to incorporate genes from organisms using bacterial transcription machinery. The evolution of a bacteriophage-type RNA polymerase functioning in mitochondria provides an example to show that drastic replacements in the transcription machinery can occur (Schinkel and Tabak 1989; Cermakian et al. 1997; Rousvoal et al. 1998).

A Matter of Scale

While the occurrence of HGT is not doubted, there is apparent controversy in assessing its impact in microbial evolution, with opinions ranging from serious concerns about its confounding effects on phylogenetics (Doolittle 1999*b*) to critical reviews which downplay any major significance (Kurland 2000). The source of much of the disagreement lies in the scale at which one is assessing a group of organisms for the effects of HGT. If one chooses a group of closely related bacteria (e.g., the enterobacteria) and examines phylogenies of genes shared among them, many different genes may re-create the same phylogeny of species (even though recombination can destroy congruence of gene phylogenies within species) Similarly, estimates of HGT based on atypical gene content imply that a minority (albeit a significant minority) of genes arrived into these genomes recently by HGT (Ochman, Lawrence, and Groisman 2000; Perna et al. 2001).

Yet such results are not inconsistent with HGT having a dominant impact on the evolution of prokaryotic genomes in the long term. Transfers occurring prior to the diversification of a group such as the enterobacteria can only be detected in larger phylogenetic reconstructions (e.g., Woese et al. 2000). Similarly, surveys which examine phylogenetic incongruity as well as atypical gene sequences as an index of HGT within a genome invariably discover a larger proportion of genes that have been subject to transfer (Ragan 2001*a*; Lawrence and Ochman 2002) because methods identifying atypical sequences are limited to detecting only recent transfers. HGT confounds evolutionary relationships most strongly on broad timescales, whereas vertical inheritance—propagating mutational changes, gene rearrangements, and other intragenomic alterations—and gene exchange by homologous recombination dominate over the short term. Moreover, HGT likely affects different lineages in different fashions, perhaps illustrated most dramatically by the minimal contribution of HGT in the evolution of intracellular parasites undergoing genome reduction (Andersson and Andersson 1999; Wernegreen et al. 2000). Consideration of scale and source can serve as effective arbiters when reconciling data collected from diverse systems.

Tests and Predictions

Implicitly, Dykhuizen and Green (1991) proposed that homologous recombination provided taxonomic coherence among groups of strains. Frequent gene exchange by homologous recombination results in strains within a species that resemble each other more than they resemble strains outside the species (fig. 1). Similarly, HGT could provide phylogenetic coherence at higher taxonomic levels. In both cases, genes within the groups should show incongruent phylogenies, although the groups themselves remain monophyletic for most genes.

This framework allows the analysis of HGT to extend beyond a collection of anecdotal evidence, enabling quantitative assessment of where the truth lies (somewhere between the extremes of the scenarios described above, no doubt). This could be established by careful and robust measurement of HGT frequencies both within and between taxonomic groupings of increasing levels of inclusiveness. If HGT has been instrumental in

shaping microbial taxonomy, then one would predict that within-group transfers would outnumber between-group transfers, whereas a random donor model would predict greater numbers of between-group exchanges (due to the larger numbers of taxa outside any one group).

An obvious caveat to this approach is that the accuracy with which phylogenies can be reconstructed, and by which HGTs can be detected, depends on the degree of divergence of the organisms and molecules under study. Phylogenetic reconstruction relies both on the occurrence of substitution events that generate informative patterns and on this information not being eroded by multiple substitutions. Potentially incongruent phylogenetic relationships found for different genes might not result from HGT at all but may be due to inadequate phylogenetic signals. Only a small subset of HGTs can be detected with confidence; the majority of transfers, especially those that occurred long ago or between closely related species, will likely escape detection. To compensate for differences in signal-to-noise ratio when comparing within-group with between-group rates of HGT, it will be important to test findings using parametric bootstrap and other quantitative approaches that incorporate vertical inheritance as well as graded HGT frequencies.

Impact of HGT on Gene Content ''Trees'' and rRNA Phylogenies

Several groups have inferred organismal phylogeny using so-called gene-content trees (Fitz-Gibbon and House 1999; Snel, Bork, and Huynen 1999; Tekaia, Lazcano, and Dujon 1999). This approach uses the mere presence of a gene as a character, and initial dendrograms produced this way do show significant congruence with established 16S rRNA phylogenies, reproducing the three-domain partition and the association of the genomes from members of the same phylum. Although more recent analyses conclude that HGT has played a significant role in determining gene content (Snel, Bork, and Huynen 2002), these results contrast with most resolved phylogenies of individual protein-coding genes, which show dramatic conflicts to both the 16S rRNA and genome content trees (see table 1 for a few notable examples). For example, some Bacteria group among the Archaea in ATP synthase phylogenies (Olendzenski et al. 2000), and phylogenies of elongation factor Tu group the Streptococcaceae with Enterococcaceae (Ke et al. 2000). These cases of well-resolved phylogenetic incongruity offer strong support for HGT. Yet other cases, such as RNA polymerase phylogenies placing *Aquifex pyrophilus* among gram-negative bacteria, and defining mycoplasmas as the deepest branch among the Bacteria (Klenk et al. 1999), entail primarily the rearrangement of bacterial phyla with respect to their placement in rRNA phylogenies and remind us that other factors—such as long-branch attraction and evolutionary rate heterogeneity—contribute to phylogenetic disparity and must be considered when interpreting these data.

While the overall correspondence between gene-content trees based on whole genome sequences and 16S rRNA phylogenies would seem to argue that HGT has played a limited role in shaping the evolution of microbial lineages, we offer two observations—in addition to the reevaluation of gene-content trees themselves (Snel, Bork, and Huynen 2002)—that suggest that this conclusion might not be warranted. First, the correspondence between gene-content trees and the 16S rRNA phylogenies often seems less impressive when additional genomes are included in gene-content trees. For example, using BLAST searches to identify shared genes and different distance measures and algorithms for tree construction, the Euryarcheote *Halobacterium* sp. was found to group at the bottom of the Archaeal domain, branching off before the split between the Crenarcheota and other Euryarcheota (Olendzenski, Zhaxybayeva, and Gogarten 2001; Korbel et al. 2002). This finding indicates that congruence between whole genome–based trees and 16S rRNA phylogeny is less robust than previously concluded.

Second, there is another possible explanation for congruence between gene-content trees and phylogenies based on rRNA. Ironically, rRNA phylogenies might agree with gene-content analyses because rRNA genes are themselves mosaic and both phylogenies reflect large-scale gene transfer. Intragenic recombination has been observed in numerous genes, and gene-conversion events tend to make copies of duplicated genes more similar to one another (Gogarten and Olendzenski 1999). The segments involved in intragenic recombination usually are less than a few hundred nucleotides in length (Sweetser et al. 1994; Betran et al. 1997; Yang and Waldman 1997), much less than the length of typical genes. As a result, different regions within a single gene may have different evolutionary histories. Mosaic rRNA operons showing extensive recombination have been observed and have been demonstrated to function (Mylvaganam and Dennis 1992; Wang, Zhang, and Ramanan 1997). This is not surprising because functioning ribosomes can be formed from constituents produced by different organisms (Nomura, Traub, and Bechmann 1968; Bellemare, Vigne, and Jordan 1973; Wrede and Erdmann 1973). Ribosomal operons of an organism can be replaced under laboratory conditions with those from another species (Asai et al. 1999), and divergent rRNA operons can coexist in the same genome (Mylvaganam and Dennis 1992; Yap, Zhang, and Wang 1999). Following introduction of a foreign rRNA operon—e.g., the *rrnB* operon of *Thermomonospora chromagena* was likely derived from an organism related to *Thermobispora bispora*—gene conversion will eventually homogenize the disparate copies of these genes. The *T. chromagena rrnB* operon is an excellent example of an intermediate in this process, where the segmental nature of the transitional form is still easily recognizable (fig. 3). Upon aligning the rRNA sequence with that of *Thermus thermophilus* (for which a high-resolution crystal structure has been determined), regions inferred to have participated in recombination (gene conversion) plausibly lie at sequences corresponding to conserved stems, and

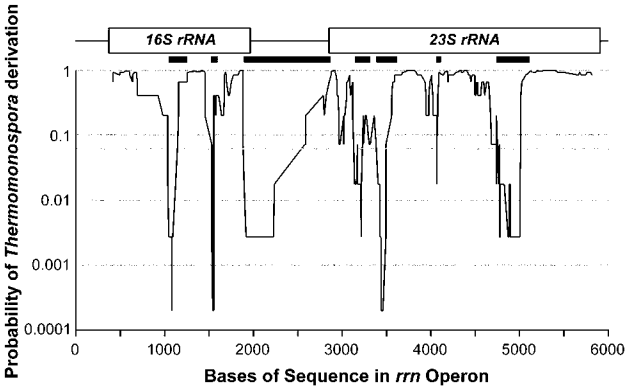

FIG. 3.—Mosaicism within the *T. chromogena rrnB* operon, which bears regions of identity to the *rrn* operons of *T. bispora*. Informative sites were identified as positions where (1) three full-length *T. chromogena rrn* operons (*rrnD, rrnE, rrnF*) bore identical bases, (2) two full-length *T. bispora rrn* operons (*rrnA, rrnC*) were identical to each other but differed from the *T. chromogena* sequences, and (3) 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, while 276 showed identity across all four *T. chromogena rrn* loci examined. A window of 10 informative sites was used to calculate directly (via the binomial distribution) the probability of the *T. chromogena rrnB* operon matching the other three *T. chromogena rrn* loci; hence, the minimal *P* value calculated (0.0002) indicated that all 10 sites within the window matched the *T. bispora rrn* loci. The dark bars denote regions likely to be of *T. bispora* origin ($P < 0.05$).

sections with discrete ancestry correspond to large structure units (e.g., loop 11 in the lower body, loops 31 and 39 in the head, and the majority of the platform; see also Wang and Zhang [2000]).

Due to strong functional constraints, rRNA genes contain continuous stretches of sequence that are conserved between divergent species. As seen in *Thermomonospora*, these sequences might facilitate recombination (Wang and Zhang 2000). Because of the redundancy of the genetic code, protein-coding DNA typically does not exhibit long stretches of nucleotide identity, precluding the frequent generation of mosaic sequences by homologous recombination. Thus the potentially mosaic nature of rRNA operons could explain the congruence of rRNA phylogenies with those inferred from gene-content comparisons. Zhaxybayeva and Gogarten (unpublished data) analyzed an alignment of 100 nearly full-length bacterial 16S rRNAs and found 17 instances of statistically significant conflicts between phylogenies reconstructed from partial versus full-length alignments. Most of these disparities persisted when different phylogenetic methods were employed or when extended data sets were analyzed that contained many additional sequences closely related to the conflicting taxa. If these data, and previous published reports (Sneath 1993; Smith et al. 1999; Ueda et al. 1999; Wang and Zhang 2000; Parker 2001), hold up to further scrutiny, a reinterpretation of the rRNA phylogeny may be necessary. Perhaps rRNA genes appear to be such useful molecules for prokaryotic taxonomy precisely because they are mosaics and reflect the mosaic character of the genome as a whole.

Evolutionary Processes Revisited

Recognition of gene transfer within and among lineages restructures microbial evolution in more ways than offering new interpretations of the *pattern* of microbial phylogeny. Here, we address four areas in which new appreciation of the role of homologous recombination and HGT might transform our understanding of *process* in prokaryotic evolution.

The Dynamic Niche

Traditional models of microbial evolution by mutational processes, combined with the measurement of environmental tolerances in laboratory environments, imparts a view of ecological niches as relatively static domains, within which organisms evolve predictably toward maximal fitness. For example, we can measure how an organism improves in fitness when grown for thousands of generations under glucose-limited conditions (Papadopoulos et al. 1999); here the environment and the evolutionary challenges seem clearly defined. However, even bacteria confined to chemostats can subvert our plans for them, inventing new niches instead of refining methods for exploiting those we had defined. For instance, strains selected for glucose utilization will spawn populations specializing in the scavenging of acetate waste products (Treves, Manning, and Adams 1998). Such adaptive changes could spawn niche-specific, ''orphan'' genes found uniquely in particular bacterial genomes.

However, even more radically inventive solutions can be expected when organisms have access to a rich variety of ready-made genes and gene complexes, as they do in real environments. HGT can fundamentally alter the character of a microbial species by introducing fully functional genes and gene clusters that can confer complex phenotypes and functions that allow effective and competitive exploitation of new niches (Lawrence 1997, 1999; Hacker and Kaper 2000). In contrast, variation introduced by point mutation will, most of the time, only adjust preexisting phenotypes. As the size of the sequence database grows, the number of orphan genes in a group at any taxonomic level decreases. This is due to our increased ability to identify distant homologues as well as the better sampling of genetic diversity which allows us to identify HGT events across large phylogenetic distances.

Because gene acquisitions can increase the metabolic repertoire of the cell, we need not view the microbial niche as a static domain, within which fitter variants constantly arise and sweep through the population. Although an organism may evolve to improve its fitness within its current niche, it is more likely that gene acquisition will allow exploitation of a related environment. In this way, the microbial niche can be considered a dynamic domain, which is redefined after each gene transfer event. This alternation of niche boundaries then imposes a different filter on the influx of foreign DNA, imparting different selective values on incoming genes. For example, an organism acquiring one pathogenicity island would begin exploring pathogenic niches that

were previously unavailable, therefore making the acquisition of subsequent pathogenicity islands far more favorable. Thus, the bacterial niche and HGT interact, each affecting the other as lineages evolve.

Implications for Fitness

Understanding evolution by HGT as a process of niche acquisition rather than refinement of niche exploitation has unexpected implications. For instance, a mesophilic heterotroph might gain access to a nearby substrate-rich but too-warm environment occupied by moderately thermophilic autotrophs, through acquisition from them of genes encoding more thermostable versions of proteins whose labilities determine its upper growth temperature. Conceivably, the newly acquired genes are very poorly adapted to the heterotrophs' other cellular machinery, so that growth rate in either environment is very slow and organisms bearing these new genes cannot compete in the original environment. They would nevertheless be the only heterotrophs at the higher temperature and could come to dominate there. Thus, frequent niche acquisition could mean that many organisms are successful because of the uniqueness of the niches they have recently discovered rather than because of fine-tuning of their cellular machinery toward the exploitation of that niche.

Lineage Diversification

Because the dynamic microbial niche is redefined after every HGT, lineage separation would occur if the populations exploring two newly derived niches were both successful. Surveys among closely related bacterial species support the hypothesis that the differences between them arose primarily by gene loss and gene acquisition, not by mutational processes. For example, all features which can discriminate between the enteric bacteria *E. coli* and *Salmonella enterica*—perhaps the best studied pair of sister species—have arisen through introduction of functions via HGT (e.g., pathogenicity in *Salmonella* or lactose utilization in *E. coli*) or by gene loss in one lineage. Both processes serve to redefine the bacterial niche.

Scope and Persistence of New Niches

The niches created by gene transfer events vary widely in their stability or novelty. Some events, like the acquisition of an antibiotic resistance gene, allow for transient exploration of a new environment, but this lineage may not persist over evolutionary time (that is, this event will likely not found a clade of antibiotic-resistant bacteria distinguished by their shared ability to be resistant to a particular antibiotic). Other events are correlated with the stable exploration of new niches, like the acquisition of the *lac* operon by *E. coli* or pathogenicity islands by *Salmonella*. Rarely, a gene transfer event may allow for the formation of radically different organisms that inhabit niches completely unreachable by organisms relying on mutational processes alone to explore environments. Examples of such lineages include the green plants (acquiring chloroplast by endosymbiosis [Bonen and Doolittle 1975]), methanotrophs (gaining the ability to synthesize critical cofactors by acquiring genes from methanogenic archaea [Chistoserdova et al. 1998]), cyanobacteria (gaining a second photosystem allowing oxygenic photosynthesis [Xiong, Inoue, and Bauer 1998]), and bacteria exploiting halorhodopsin homologues as light-driven proton pumps (Beja et al. 2001).

Shifting the Shifting Balance

A classic model for adaptation has been the Shifting Balance Theory (Wright 1932, 1982), wherein populations of organisms are found at selective peaks on an adaptive landscape. Changing an ecological niche is tantamount to relocation to a different peak on this landscape, necessitating travel through a ''valley'' of poor fitness. This process has been demonstrated experimentally in the engineering of enzymes with altered substrate specificities (Golding and Dean 1998). Adaptive changes may occur through sequential selection of mutations, and perhaps some genome-specific, orphan genes are the products of such classically Darwinian processes. But intragenic recombination can facilitate rapid exploration of this adaptive landscape because the valleys of low fitness need never be crossed (Bogarad and Deem 1999). Variant alleles with near-optimal fitnesses may be recombined to introduce multiple changes simultaneously, thereby avoiding the formation of suboptimal intermediate states.

HGT offers an expanded scope to these models, which show conclusively that recombination among existing variants offers accelerated pathways to fitness peaks. While fitness peaks may never be explored if they must be reached one gene at a time, multiple genes may be acquired in the form of bacterial operons and gene clusters (Lawrence 1997; Lawrence and Ochman 1997; Hacker and Kaper 2000; Lawrence 2001). Many examples of HGT involve the introduction of complex, multigene pathways (e.g., Jiang et al. 1995; Kranz and Goldman 1998; Perna et al. 2001).

Time Frame for Diversification

From an evolutionary perspective, lineage diversification is often viewed as an instantaneous event, a point after which genes in two groups of organisms are no longer in genetic communication. Plausible models for lineage separation invoke the initial acquisition of characters that make populations ecologically distinct (Cohan 2001; Lawrence 2002). Here, recombination between these populations at these loci would produce less fit offspring that would be counterselected (postmating reproductive isolation). Yet homologous recombination may exchange alleles between these populations at loci uninvolved in initial ecological differentiation. Neutral mutations would accumulate at loci adjacent to genes that confer ecological distinctiveness owing to the reduced levels of recombination there, ultimately leading to premating reproductive isolation mediated by mis-

match correction systems as discussed above (Lawrence 2002).

Eventually, all genes in the two ecologically distinct populations may become sufficiently different for gene exchange by homologous recombination not to be observed at any locus. If one considers this point the time of speciation, one may seriously underestimate the time of separation of genes which have been genetically isolated for longer periods of time—such as those linked to loci conferring early ecological distinctiveness—if the time from initial lineage separation (genetically isolating some genes) to final premating isolation (genetically isolating all genes) is large.

Because many metrics of molecular evolution between distinct lineages (such as rates of substitution) rely on a single divergence time for all genes in the chromosome, variation in the time of lineage separation among genes may be responsible for a substantial portion of the variance in these measures across genes. For example, genes with similar codon usage biases—reflecting similar degrees of selection on their synonymous sites—should have similar values for synonymous substitution rates, yet the correlation coefficient is just over 0.5 for genes shared between *E. coli* and *S. enterica* (Sharp et al. 1989; Sharp 1991). How much of the remaining variation is due to differences in divergence times of these genes, where an apparently large K_s for a gene, given its degree of codon usage bias, may have resulted from an earlier time of separation in the two lineages rather than a disproportionately large rate of accumulation of mutations?

The Font of Innovation

Most bacterial genome sequences reveal an abundance of paralogs which are often viewed as products of within-lineage duplication and divergence. However, many must be the result of the reuniting through HGT of orthologs that have diverged in separate lineages. Recognizing this encourages a rethinking of standard models of gene duplication and divergence. These models resemble sympatric speciation for genes, where ecological distinctiveness must arise before reproductive isolation. (In sympatric speciation events, species arise while dwelling in the same physical location. Diversifying selection allows for the propagation of distinct subpopulations, each bearing some fraction of the original genetic variation that allows ecologically distinct roles to be played. Reproductive isolation is necessary to prevent mixture of subpopulations and coalescence of the two nascent lineages into a single population. In allopatric speciation, a parental species is physically divided into two reproductively isolated populations; reproductive isolation may occur stochastically, and ecological differences may arise. If rejoined, populations will persist only if novel ecological roles have been established. Thus, the physical separation of species may allow for reproductive isolation to occur while the two lineages develop ecological distinctiveness.)

Their drawbacks can be circumvented if we adopt an ''allopatric'' approach to the evolution of novel gene function. It is clear that a gene is duplicated every time a cell divides. Yet in this case, the two gene copies are present in different cytoplasms—the equivalent of gene allopatry. In separate organisms, genes are free to evolve distinct biochemical functions. Either the functional breadth of the gene product may expand to include additional activities or some of the gene product's original functions may be lost if those functions are not critical in this organism. If genes are never reintroduced into the same cytoplasm, ecologically different roles need never be established and orthologous genes persist in separate cytoplasmic contexts. If the genes are reunited in the same cytoplasm, they must have achieved physiological distinctiveness (paralogous functions) for both to persist. Reintroduction of genes into the same genome is mediated by gene transfer, including both (1) homologous recombination with unequal crossing-over—here, a merodiploid strain (one bearing a duplication of a portion of its chromosome) is created at the initial point of DNA exchange, and (2) HGT, which is the most dramatic way of allowing gene transfer to introduce paralogous genes into the same cell.

Summary and Conclusions

Comparative analyses of gene and genome sequences indicate that exchange of genetic information within and between prokaryotic species, however defined, is far more frequent and general than previously thought. Although exchange by homologous recombination is limited by sequence divergence and should decrease markedly with ''phylogenetic distance,'' exchange by the various illegitimate recombinational processes collectively designated HGT is not so constrained. New understanding of both phenomena and their potential interplay suggests that traditional models for prokaryotic evolution based on clonality and periodic selection are inadequate to describe the process of prokaryotic evolution at the species level and that tree-like phylogenies are inadequate to represent the pattern of prokaryotic evolution at any level. Here we elaborate on this new understanding to show that a coherent model for prokaryotic evolution which invokes gene transfer as its principle explanatory force is feasible and would have many benefits for understanding diversification and adaptation. In particular, we could resolve the ''species problem'' (perhaps by dismissing it), appreciate the real differences in tempo and mode between prokaryote and ''higher'' eukaryote evolution, let unraveling of the complex histories of genes and genomes supersede the quest for one true ''organismal phylogeny,'' develop new models for diversification of prokaryotic niches and definitions of adaptedness, and, at the level of the gene, propose new scenarios for evolution of novel function.

Components of this new view as it relates to species and adaptation have already been clearly articulated, especially by Maynard Smith, Spratt, and Levin and their collaborators (Levin and Bergstrom 2000; Maynard Smith, Feil, and Smith 2000; Feil et al. 2001). Phylogenetic implications have also been explored by us and by Martin (1999) and Woese (2000), among others. Our

intent here was to show that embracing gene transfer promises a broad and radical revision of the prokaryotic evolutionary paradigm. This will come from a fusion of population genetics, molecular genetics, epidemiological and environmental genomics, microbial ecology, and molecular phylogeny, fields that have developed mostly in isolation from each other. Although we have presented the new view as if it were antithetical to traditional understandings of prokaryotic evolution, in the long run we endorse a synthesis that will acknowledge gene exchange and clonality, weblike and treelike behavior, and adaptation and the evolution of new function by many modes. We believe the most immediate task is to determine whether frequencies of within- and between-lineage gene exchange support a model like that depicted in figure 2 or whether vertical descent remains the best descriptor of the history of most genes over evolutionary time. While there are complex issues of measurement and definition to overcome, rapidly accumulating genome sequences provide no shortage of data.

Acknowledgments

We thank the Astrobiology Institute and Exobiology program of NASA (NAG5-7915), the Canadian Institutes for Health Research and the Canada Research Chair, and the David and Lucile Packard Foundation for support of research in our laboratories, Camilla Nesbø, Yan Boucher, and Olga Zhaxybayeva for helpful discussions, and the Canadian Institute for Advanced Research for supporting the preparation of this manuscript.

LITERATURE CITED

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

ARBER, W. 1979. Promotion and limitation of genetic exchange. Experientia **35**:287–293.

ASAI, T., D. ZAPOROJETS, C. SQUIRES, and C. L. SQUIRES. 1999. An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: complete exchange of rRNA genes between bacteria. Proc. Natl. Acad. Sci. USA **96**:1971–1976.

BEJA, O., E. N. SPUDICH, J. L. SPUDICH, M. LECLERC, and E. F. DELONG. 2001. Proteorhodopsin phototrophy in the ocean. Nature **411**:786–789.

BELLEMARE, G., R. VIGNE, and B. R. JORDAN. 1973. Interaction between *Escherichia coli* ribosomal proteins and 5S RNA molecules: recognition of prokaryotic 5S RNAs and rejection of eukaryotic 5S RNAs. Biochimie **55**:29–35.

BETRAN, E., J. ROZAS, A. NAVARRO, and A. BARBADILLA. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146**:89–99.

BOGARAD, L. D., and M. W. DEEM. 1999. A hierarchical approach to protein molecular evolution. Proc. Natl. Acad. Sci. USA **96**:2591–2595.

BONEN, L., and W. F. DOOLITTLE. 1975. On the prokaryotic nature of red algal chloroplasts. Proc. Natl. Acad. Sci. USA **72**:2310–2314.

BOUCHER, Y., H. HUBER, S. L'HARIDON, K. O. STETTER, and W. F. DOOLITTLE. 2001. Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. Mol. Biol. Evol. **18**:1378–1388.

BOUCHER, Y., C. L. NESBO, and W. F. DOOLITTLE. 2001. Microbial genomes: dealing with diversity. Curr. Opin. Microbiol. **4**:285–289.

BROCHIER, C., H. PHILIPPE, and D. MOREIRA. 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., C. J. DOUADY, M. J. ITALIA, W. E. MARSHALL, and M. J. STANHOPE. 2001. Universal trees based on large combined protein sequence data sets. Nature Genet. **28**: 281–285.

CERMAKIAN, N., T. M. IKEDA, P. MIRAMONTES, B. F. LANG, M. W. GRAY, and R. CEDERGREN. 1997. On the evolution of the single-subunit RNA polymerases. J. Mol. Evol. **45**:671–681.

CHISTOSERDOVA, L., J. A. VORHOLT, R. K. THAUER, and M. E. LIDSTROM. 1998. C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic archaea. Science **281**:99–102.

COHAN, F. M. 1994*a*. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am. Nat. **143**:965–986.

———. 1994*b*. Genetic exchange and evolutionary divergence in prokaryotes. Trends Ecol. Evol. **9**:175–180.

———. 2001. Bacterial species and speciation. Syst. Biol. **50**: 513–524.

DENAMUR, E., S. BONACORSI, A. GIRAUD ET AL. (11 co-authors). 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. J. Bacteriol. **184**:605–609.

DENAMUR, E., G. LECOINTRE, P. DARLU ET AL. (12 co-authors). 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103**:711–721.

DEPPENMEIER, U., A. JOHANN, T. HARTSCH ET AL. (22 co-authors). 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J. Mol. Microbiol. Biotechnol. **4**:453–461.

DOOLITTLE, R. F., D. F. FENG, K. L. ANDERSON, and M. R. ALBERRO. 1990. A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. J. Mol. Evol. **31**: 383–388.

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

DOOLITTLE, W. F. 1999*a*. Lateral genomics. Trends Cell Biol. **9**:M5–M8.

———. 1999*b*. Phylogenetic classification and the universal tree. Science **284**:2124–2129.

———. 2000. The nature of the universal ancestor and the evolution of the proteome. Curr. Opin. Struct. Biol. **10**:355–358.

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

FAGUY, D. M., and W. F. DOOLITTLE. 2000. Horizontal transfer of catalase-peroxidase genes between Archaea and pathogenic bacteria. Trends Genet. **16**:196–197.

FEIL, E. J., E. C. HOLMES, D. E. BESSEN ET AL. (12 co-authors). 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 C. H. HOUSE. 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. **27**:4218–4222.

FRIEDRICH, M. W. 2002. Phylogenetic analysis reveals multiple lateral transfers of adenosine-5′-phosphosulfate reductase genes among sulfate-reducing microorganisms. J. Bacteriol. **184**:278–289.

GARCIA-VALLVE, S., A. ROMEU, and J. PALAU. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol. Biol. Evol. **17**:352–361.

GODDARD, M. R., and A. BURT. 1999. Recurrent invasion and extinction of a selfish gene. Proc. Natl. Acad. Sci. USA **96**: 13880–13885.

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

GOGARTEN, J. P., E. HILARIO, and L. OLENDZENSKI. 1996. Gene duplications and horizontal gene transfer during early evolution. Pp. 267–292 *in* D. M. ROBERTS, P. SHARP, G. ALDERSON, and M. A. COLLINS, eds. Symposium of the society for general microbiology; evolution of microbial life. Cambridge University Press, Cambridge, United Kingdom.

GOGARTEN, J. P., R. D. MURPHEY, and L. OLENDZENSKI. 1999. Horizontal gene transfer: pitfalls and promises. Biol. Bull. **196**:359–361.

GOGARTEN, J. P., and L. OLENDZENSKI. 1999. Orthologs, paralogs and genome comparisons. Curr. Opin. Genet. Dev. **9**: 630–636.

GOGARTEN, J. P., T. STARKE, H. KIBAK, J. FISHMAN, and L. TAIZ. 1992. Evolution and isoforms of V-ATPase subunits. J. Exp. Biol. **172**:137–147.

GOLDING, G. B., and A. M. DEAN. 1998. The structural basis of molecular adaptation. Mol. Biol. Evol. **15**:355–369.

GRAHAM, D. E., R. OVERBEEK, G. J. OLSEN, and C. R. WOESE. 2000. An archaeal genomic signature. Proc. Natl. Acad. Sci. USA **97**:3304–3308.

GUPTA, R. S., and G. B. GOLDING. 1993. Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. J. Mol. Evol. **37**: 573–582.

GUTTMAN, D. S., and D. E. DYKHUIZEN. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266**:1380–1383.

HACKER, J., and J. B. KAPER. 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54**:641–679.

HAYES, W. S., and M. BORODOVSKY. 1998. How to interpret an anonymous bacterial genome: machine learning approach to gene identification. Genome Res. **8**:1154–1171.

HEINEMANN, J. A., and G. F. J. SPRAGUE. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature **340**:205–209.

IBBA, M., J. L. BONO, P. A. ROSA, and D. SOLL. 1997. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. USA **94**:14383–14388.

JAIN, R., M. C. RIVERA, and J. A. LAKE. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA **96**:3801–3806.

JIANG, W., W. W. METCALF, K. S. LEE, and B. L. WANNER. 1995. Molecular cloning, mapping, and regulation of Pho regulon genes for phosphonate breakdown by the phosphonatase pathway of *Salmonella typhimurium* LT2. J. Bacteriol. **177**:6411–6421.

KARLIN, S., and C. BURGE. 1995. Dinucleotide relative abundance extremes: a genomic signature. Trends Genet. **11**: 283–290.

KARLIN, S., and J. MRAZEK. 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. **182**: 5238–5250.

KARLIN, S., J. MRAZEK, and A. M. CAMPBELL. 1998. Codon usages in different gene classes of the *Escherichia coli* genome. Mol. Microbiol. **29**:1341–1355.

KATZ, L. A. 1996. Transkingdom transfer of the phosphoglucose isomerase gene. J. Mol. Evol. **43**:453–459.

KE, D., M. BOISSINOT, A. HULETSKY, F. J. PICARD, J. FRENETTE, M. OUELLETTE, P. H. ROY, and M. G. BERGERON. 2000. Evidence for horizontal gene transfer in evolution of elongation factor Tu in enterococci. J. Bacteriol. **182**:6913–6920.

KLEIN, M., M. W. FRIEDICH, A. J. ROGER, P. HUGENHOLTZ, S. FISHBAIN, H. ABICHT, L. L. BLACKALL, D. L. STAHL, and M. WAGNER. 2001. Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. J. Bacteriol. **183**:6028–6035.

KLENK, H. P., T. D. MEIER, P. DUROVIC, V. SCHWASS, F. LOTTSPEICH, P. P. DENNIS, and W. ZILLIG. 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., L. ARAVIND, and A. S. KONDRASHOV. 2000. The impact of comparative genomics on our understanding of evolution. Cell **101**:573–576.

KOONIN, E. V., A. R. MUSHEGIAN, M. Y. GALPERIN, and D. R. WALKER. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. **25**:619–637.

KORBEL, J. O., B. SNEL, M. A. HUYNEN, and P. BORK. 2002. SHOT: a web server for the construction of genome phylogenies. Trends Genet. **18**:158–162.

KRANZ, R. G., and B. S. GOLDMAN. 1998. Evolution and horizontal transfer of an entire biosynthetic pathway for cytochrome *c* biogenesis: *Helicobacter*, *Deinococcus*, Archaea and more. Mol. Microbiol. **27**:871–874.

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.

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

———. 2001. Catalyzing bacterial speciation: correlating lateral transfer with genetic headroom. Syst. Biol. **50**:479–496.

———. 2002. Gene transfer in bacteria: speciation without species? Theor. Popul. Biol. **61**:449–460.

LAWRENCE, J. G., and H. OCHMAN. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. **44**:383–397.

———. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95**:9413–9417.

———. 2002. Reconciling the many faces of gene transfer. Trends Microbiol. **10**:1–4.

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

LEVIN, B. R., and C. T. BERGSTROM. 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proc. Natl. Acad. Sci. USA **97**:6981–6985.

LUDWIG, W., O. STRUNK, S. KLUGBAUER, N. KLUGBAUER, M. WEIZENEGGER, J. NEUMAIER, M. BACHLEITNER, and K. H. SCHLEIFER. 1998. Bacterial phylogeny based on comparative sequence analysis. Electrophoresis **19**:554–568.

MAJEWSKI, J., and F. M. COHAN. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. Genetics **148**:13–18.

———. 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. Genetics **153**:1525–1533.

MAKAROVA, K. S., L. ARAVIND, M. Y. GALPERIN, N. V. GRISHIN, R. L. TATUSOV, Y. I. WOLF, and E. V. KOONIN. 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., V. A. PONOMAREV, and E. V. KOONIN. 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**:research0033.0031–0014.

MARTIN, W. 1999. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. Bioessays **21**:99–104.

MAYNARD SMITH, J., E. J. FEIL, and N. H. SMITH. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays **22**:1115–1122.

MAYR, E. 1942. Systematics and the origin of species. Columbia University Press, New York.

———. 1963. Animal species and evolution. Harvard University Press, Cambridge, Mass.

MCKANE, M., and R. MILKMAN. 1995. Transduction, restriction and recombination patterns in *Escherichia coli*. Genetics **139**:35–43.

MOSZER, I., E. P. ROCHA, and A. DANCHIN. 1999. Codon usage and lateral gene transfer in *Bacillus subtilis*. Curr. Opin. Microbiol. **2**:524–528.

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

NELSON, K. E., R. A. CLAYTON, S. R. GILL ET AL. (25 coauthors). 1999. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399**:323–329.

NESBO, C. L., Y. BOUCHER, and W. F. DOOLITTLE. 2001. Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. J. Mol. Evol. **53**:340–350.

NESBO, C. L., S. L'HARIDON, K. O. STETTER, and W. F. DOOLITTLE. 2001. Phylogenetic analyses of two ''archaeal'' genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. Mol. Biol. Evol. **18**:362–375.

NOMURA, M., P. TRAUB, and H. BECHMANN. 1968. Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins. Nature **219**:793–799.

OCHMAN, H., J. G. LAWRENCE, and E. GROISMAN. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299–304.

OLENDZENSKI, L., L. LIU, O. ZHAXYBAYEVA, R. MURPHEY, D. G. SHIN, and J. P. GOGARTEN. 2000. Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. J. Mol. Evol. **51**: 587–599.

OLENDZENSKI, L., O. ZHAXYBAYEVA, and J. P. GOGARTEN. 2001. What's in a tree? Does horizontal gene transfer determine microbial taxonomy? Pp 65–78 *in* J. SECKBACH, ed. Symbiosis. Mechanisms and model systems. Kluwer Academic Publishers, Dordrecht.

OLIVER, A., R. CANTON, P. CAMPO, F. BAQUERO, and J. BLAZQUEZ. 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science **288**: 1251–1254.

PAPADOPOULOS, D., D. SCHNEIDER, J. MEIER-EISS, W. ARBER, R. E. LENSKI, and M. BLOT. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA **96**:3807–3812.

PARKER, M. W. 2001. Case of localized recombination in 23S rRNA genes from divergent *Bradyrhizobium* lineages associated with neotropical legumes. Appl. Environ. Microbiol. **67**:2076–2082.

PENNISI, E. 1999. Genome data shake tree of life. Science **280**: 672–674.

PERNA, N. T., G. PLUNKETT, V. BURLAND ET AL. (28 co-authors). 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409**:529–533.

PESOLE, G., C. GISSI, C. LANAVE, and C. SACCONE. 1995. Glutamine synthetase gene evolution in bacteria. Mol. Biol. Evol. **12**:189–197.

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

———. 2001*b*. On surrogate methods for detecting lateral gene transfer. FEMS Microbiol. Lett. **201**:187–191.

ROGER, A. J., and J. R. BROWN. 1996. A chimeric origin for eukaryotes re-examined. Trends Biochem. Sci. **21**:370–372.

ROUSVOAL, S., M. OUDOT, J. FONTAINE, B. KLOAREG, and S. L. GOER. 1998. Witnessing the evolution of transcription in mitochondria: the mitochondrial genome of the primitive brown alga *Pylaiella littoralis* (L.) Kjellm. Encodes a T7-like RNA polymerase. J. Mol. Biol. **277**:1047–1057.

SCHINKEL, A. H., and H. F. TABAK. 1989. Mitochondrial RNA polymerase: dual role in transcription and replication. Trends Genet. **5**:149–154.

SENEJANI, A. G., E. HILARIO, and J. P. GOGARTEN. 2001. The intein of the *Thermoplasma* A-ATPase A subunit: structure, evolution and expression in *E. coli*. BMC Biochem. **2**:13.

SHARP, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33**:23–33.

SHARP, P. M., D. C. SHIELDS, K. H. WOLFE, and W.-H. LI. 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. Science **246**:808–810.

SHIBUI, H., T. HAMAMOTO, M. YOHDA, and Y. KAGAWA. 1997. The stabilizing residues and the functional domains in the hyperthermophilic V-ATPase of *Desulfurococcus*. Biochem. Biophys. Res. Commun. **234**:341–345.

SMITH, N. H., E. C. HOLMES, G. M. DONOVAN, G. A. CARPENTER, and B. G. SPRATT. 1999. Networks and groups within the genus *Neisseria*: analysis of *argF, recA, rho,* and 16S rRNA sequences from human *Neisseria* species. Mol. Biol. Evol. **16**:773–783.

SNEATH, P. H. 1993. Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. Int. J. Syst. Bacteriol. **43**:626–629.

SNEL, B., P. BORK, and M. HUYNEN. 1999. Genome phylogeny based on gene content. Nature Genet. **21**:108–110.

———. 2002. Genomes in flux: the evolution of archaeal and proteobacterial gene content. Genome Res. **12**:17–25.

SWEETSER, D. B., H. HOUGH, J. F. WHELDEN, M. ARBUCKLE, and J. A. NICKOLOFF. 1994. Fine-resolution mapping of spontaneous and double-strand break-induced gene conversion tracts in *Saccharomyces cerevisiae* reveals reversible mitotic conversion polarity. Mol. Cell Biol. **14**:3863–3875.

TEKAIA, F., A. LAZCANO, and B. DUJON. 1999. The genomic tree as revealed from whole proteome comparisons. Genome Res. **9**:550–557.

TREVES, D. S., S. MANNING, and J. ADAMS. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. Mol. Biol. Evol. **15**: 789–797.

TURNER, S. L., and J. P. YOUNG. 2000. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol. Biol. Evol. **17**:309–319.

UEDA, K., T. SEKI, T. KUDO, T. YOSHIDA, and M. KATAOKA. 1999. Two distinct mechanisms cause heterogeneity of 16S rRNA. J. Bacteriol. **181**:78–82.

UTAKER, J. B., K. ANDERSEN, A. AAKRA, B. MOEN, and I. F. NES. 2002. Phylogeny and functional expression of ribulose 1,5-bisphosphate carboxylase/oxygenase from the autotrophic ammonia-oxidizing bacterium *Nitrosospira* sp. isolate 40KI. J. Bacteriol. **184**:468–478.

VULIC, M., F. DIONISIO, F. TADDEI, and M. RADMAN. 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA **94**:9763–9767.

VULIC, M., R. E. LENSKI, and M. RADMAN. 1999. Mutation, recombination, and incipient speciation of bacteria in the laboratory. Proc. Natl. Acad. Sci. USA **96**:7348–7351.

WANG, Y., and Z. ZHANG. 2000. Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology **146**:2845–2854.

WANG, Y., Z. ZHANG, and N. RAMANAN. 1997. The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J. Bacteriol. **179**:3270–3276.

WARD, D. M. 1998. A natural species concept for prokaryotes. Curr. Opin. Microbiol. **1**:271–277.

WERNEGREEN, J. J., H. OCHMAN, I. B. JONES, and N. A. MORAN. 2000. Decoupling of genome size and sequence divergence in a symbiotic bacterium. J. Bacteriol. **182**:3867–3869.

WILSON, G. G., and N. E. MURRAY. 1991. Restriction and modification systems. Annu. Rev. Genet. **25**:585–627.

WILSON, R. A. 1999. Species, new interdisciplinary essays. Massachusetts Institute of Technology, Boston.

WOESE, C. R. 2000. Interpreting the universal phylogenetic tree. Proc. Natl. Acad. Sci. USA **97**:8392–8396.

WOESE, C. R., G. J. OLSEN, M. IBBA, and D. SOLL. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. **64**:202–236.

WOLF, Y. I., L. ARAVIND, N. V. GRISHIN, and E. V. KOONIN. 1999. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res. **9**:689–710.

WOLF, Y. I., I. B. ROGOZIN, A. S. KONDRASHOV, and E. V. KOONIN. 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res. **11**:356–372.

WREDE, P., and V. A. ERDMANN. 1973. Activities of *B. stearothermophilus* 50 S ribosomes reconstituted with prokaryotic and eukaryotic 5 S RNA. FEBS Lett. **33**:315–319.

WRIGHT, S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. 6th Int. Congr. Genet. **1**:356–366.

———. 1982. The shifting balance theory and macroevolution. Annu. Rev. Genet. **16**:1–19.

XIONG, J., K. INOUE, and C. E. BAUER. 1998. Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. Proc. Natl. Acad. Sci. USA **95**:14851–14856.

YANG, D., and A. S. WALDMAN. 1997. Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. Mol. Cell Biol. **17**:3614–3628.

YAP, W. H., Z. ZHANG, and Y. WANG. 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., M. S. ROBERTS, and F. M. COHAN. 1995. The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. Genetics **140**:917–932.

ZHAXYBAYEVA, O., and J. P. GOGARTEN. 2002. Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analyses. BMC Genomics **3**:4.

MARK RAGAN, reviewing editor

Accepted August 15, 2002