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 (Guttmann and Dykhuizen 1994), but early estimates suggested that rates of recombination were sufficiently low to be ignored when considering periodic selection events (Cohan 1994a, 1994b). 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 Neisseria meningitidis, 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
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 homologous recombination of large fragments of DNA) imposes 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-Vallvé, 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 1999b; 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

<table>
<thead>
<tr>
<th>Protein or RNA</th>
<th>Phylogenetic Incongruities</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information transfer</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Ribosomal RNA (rrn) | 1. *Thermomonospora* contains *rrn* operon donated from *Thermobispora*  
2. *Halococcus* 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. Mycoplasmas 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 *Mycoplasma* species groups with *e*-proteobacteria | (Makarova, Ponomarev, and Koonin, 2001) |
| Elongation factor Tu (TufB) | Streptococccaceae group with Enterococccaceae | (Ke et al., 2000) |
| Lysyl-tRNA synthase | *Borrelia* groups with *Archaea* | (Ibba et al., 1997) |
| Phenylalanyl-tRNA synthase | Spirochetes group with *Archaea*  
1. *Deinococcus*, *Mycoplasmata* 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 archaeal *Halococcus* 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 | Synthrophobacteraceae and *Nitrosopinaeae* 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) | *Desulfobacterium* (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 *Euryarchaeota* | (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 with the *Eukaryotes* | (Doolittle et al., 1990) |
| GroES/GroEL chaperone | *Methanosarcina mazei* groups with *Bacteria* | (Deppenmeier et al., 2002) |
| HSP70 (DnaK) | Archael 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 H4MPT-mediated methyl-group transfer | *Methylobacterium extorquens*, an aerobic methyl-trophic proteobacterium, clusters tightly with *Archaea* 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) |
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 1999a, 1999b, 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, Mrázek, and Campbell 1998; Moszer, Rocha, and Danchin 1999; Karlin and Mrázek 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 2001b; 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; Tekai, 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 organisomal 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 1999a, 1999b, 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 intra-phylogeny 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.

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 1999b) 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 2001a; 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 tRNA Phylogenies

Several groups have inferred organismal phylogeny using so-called gene-content trees (Fitz-Gibbon and House 1999; Snel, Bork, and Huynen 1999; Tekia, Lacan, 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 Streptococci with Enterococci (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 *Aquific 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 (Myl vaganam and Dennis 1992; Wang, Zhang, and Raman 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 (Myl vaganam and Dennis 1992; Yap, Zhang, and Wang 1999). Following introduction of a foreign rRNA operon—e.g., the *rrnB* operon of *Thermomonospora chromagensa*—was likely derived from an organism related to *Thermobispora bispora*—gene conversion will eventually homogenize the disparate copies of these genes. The *T. chromagensa* *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
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 (Trevés, 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 remaining variation is due to differences in divergence (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 $K$, 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


