

HORIZONTAL GENE TRANSFER IN PROKARYOTES: Quantification and Classification¹

Eugene V. Koonin,¹ Kira S. Makarova,^{1,2} and L. Aravind¹

¹National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, and ²Department of Pathology, F.E. Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799; e-mail: koonin@ncbi.nlm.nih.gov, makarova@ncbi.nlm.nih.gov, aravind@ncbi.nlm.nih.gov

Key Words phylogenetic analysis, gene acquisition, gene displacement, adaptation, orthologs

■ **Abstract** Comparative analysis of bacterial, archaeal, and eukaryotic genomes indicates that a significant fraction of the genes in the prokaryotic genomes have been subject to horizontal transfer. In some cases, the amount and source of horizontal gene transfer can be linked to an organism's lifestyle. For example, bacterial hyperthermophiles seem to have exchanged genes with archaea to a greater extent than other bacteria, whereas transfer of certain classes of eukaryotic genes is most common in parasitic and symbiotic bacteria. Horizontal transfer events can be classified into distinct categories of acquisition of new genes, acquisition of paralogs of existing genes, and xenologous gene displacement whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog). Each of these types of horizontal gene transfer is common among prokaryotes, but their relative contributions differ in different lineages. The fixation and long-term persistence of horizontally transferred genes suggests that they confer a selective advantage on the recipient organism. In most cases, the nature of this advantage remains unclear, but detailed examination of several cases of acquisition of eukaryotic genes by bacteria seems to reveal the evolutionary forces involved. Examples include isoleucyl-tRNA synthetases whose acquisition from eukaryotes by several bacteria is linked to antibiotic resistance, ATP/ADP translocases acquired by intracellular parasitic bacteria, *Chlamydia* and *Rickettsia*, apparently from plants, and proteases that may be implicated in chlamydial pathogenesis.

CONTENTS

INTRODUCTION	710
CRITERIA FOR DETECTING HORIZONTALLY TRANSFERRED GENES	713

¹The U.S. Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

Unexpected Ranking of Sequence Similarity Among Homologs	713
Unexpected Phylogenetic Tree Topology	713
Unusual Phyletic Patterns	714
Conservation of Gene Order Between Distant	
Taxa: Horizontal Transfer of Operons	715
Anomalous Nucleotide Composition	717
Establishing the Direction of Horizontal Gene Transfer	717
CLASSIFICATION AND QUANTIFICATION	
OF HORIZONTAL GENE TRANSFER EVENTS	718
HORIZONTAL GENE TRANSFER FROM EUKARYOTES	
TO BACTERIA AND ARCHAEA	723
Aminoacyl-tRNA Synthetases	729
Miscellaneous Eukaryotic Genes Acquired by Prokaryotes	733
CONCLUSIONS	735

INTRODUCTION

Horizontal (lateral) gene transfer, the transfer of genes between different species, is an evolutionary phenomenon whose extent and even very existence have been the subject of a longtime debate that tends to become particularly vigorous when cases of horizontal gene transfer that involve eukaryotes are considered (98, 103). This is understandable because (a) horizontal gene transfer seems to challenge the traditional, tree-based view of the evolution of life and the core neo-Darwinist belief in the central role of reproductive isolation between species in evolution (21–23, 53, 74, 89, 90), concepts that, at least initially, have been developed in studies on the evolution of sexually reproducing eukaryotes, and (b) like many evolutionary phenomena, horizontal transfer is hard to prove unambiguously. Until the advent of the genome sequencing era, although striking anecdotal examples of horizontal gene transfer have been described (97) and some prescient speculation on the potential major evolutionary impact of such events has been published (102), the prevailing opinion seemed to be that this phenomenon was rare enough to be insignificant for our general understanding of evolution. The only instance where the importance of horizontal gene transfer had been clearly recognized was the apparent flow of genes from the genomes of endosymbiotic organelles—mitochondria in all eukaryotes and chloroplasts in plants—to the eukaryotic nuclear genome (37, 38, 58, 75).

However, sequenced-based genomics has quickly shown that these “illegitimate” evolutionary events are too common to be dismissed as inconsequential (22, 84). The first strong indication probably came from the multifactorial analysis of codon frequencies in portions of the *Escherichia coli* genome that indicated significant deviation from the general pattern of codon usage in about 15% of this bacterium’s genes (76). Because some of the genes in this group showed clear relationships with bacteriophage genes, the hypothesis has been proposed that all these genes were alien to *E. coli* and have been acquired horizontally from various sources. These types of observations seemed to strongly

suggest substantial and relatively recent horizontal gene flow. On a different evolutionary scale, the possibility of horizontal transfer being a major factor in evolution has been suggested by the difficulty of constructing congruent phylogenetic trees for different sets of orthologous genes from a wide range of organisms. For example, some archaeal genes showed a clear affinity to their eukaryotic counterparts, whereas others equally strongly clustered with bacterial homologs (12, 33, 40).

The availability of multiple prokaryotic genomes for comparative analysis ushered in the new age of "lateral genomics" (21). Dramatic differences in gene repertoires even among bacteria that belong to the same evolutionary lineage, such as *E. coli* and *Haemophilus influenzae* (106), indicated that genome evolution could not be reasonably described in terms of vertical descent alone. It is clear that much of the difference was attributable to differential gene loss, particularly in parasites, but horizontal gene transfer is the other major evolutionary factor that could help explain the emerging complex picture of prokaryotic genomes. The archaeal genomes presented a particularly striking "genomescape" strongly suggestive of massive horizontal gene transfer. In agreement with the earlier indications from phylogenetic studies, but now on the whole-genome scale, it has become clear that archaeal proteins split into those genes that were most similar to their bacterial homologs and that looked "eukaryotic" (24, 56, 70). Some exceptions notwithstanding, the bacterial and eukaryotic proteins in archaea were neatly divided along functional lines, with those involved in information processing (translation, transcription, and replication) showing the eukaryotic affinity, and metabolic enzymes, structural components, and a variety of uncharacterized proteins that appeared to be "bacterial." Because the informational components generally appear to be less subject to horizontal gene transfer (however, some important exceptions are discussed below) and in accord with the standard model of early evolution whereby eukaryotes share a common ancestor with archaea, these observations have been tentatively explained by massive gene exchange between archaea and bacteria (56). This view has been further supported when the genomes of two hyperthermophilic bacteria, *Aquifex aeolicus* and *Thermotoga maritima*, were sequenced. Each of these genomes contained a significantly greater fraction of archaeal genes than any of the other bacterial genomes, establishing a plausible connection between the similarity in the lifestyles of evolutionarily distant organisms and the apparent rate of horizontal gene exchange between them (4, 83). Also, these findings emphasized the issue of the adaptive versus opportunistic nature of horizontal gene transfer: Do the genes that might have been acquired from archaea directly enable these bacteria to thrive in hyperthermal conditions, or have they acquired more archaeal genes simply because they have been more exposed to contacts with archaea because of their thermophily? Another case of nonrandomness in the apparent horizontal gene transfer has been observed in the genome of the cyanobacterium *Synechocystis* sp., which encodes a variety of proteins associated with different forms of signaling that have been thought of as eukaryotic (50, 91).

The finding that the contributions of horizontal gene transfer and lineage-specific gene loss to the gene repertoire of prokaryotes was comparable to that of vertical descent amounted to a major shift in our understanding of evolution. Indeed, it became apparent that, in many cases, phylogenetic trees for different genes were incongruent not because of artifacts inherent in tree-construction methods but because of genuine differences in the evolutionary histories of these genes brought about by horizontal transfer. Thus, a true tree of life, a species tree, could not be constructed, not because of the complexity of the problem and erosion of the phylogenetic signal from ancient divergence events, but perhaps in principle (22, 23). The best one could hope for was a consensus tree that would reflect the history of a gene core conserved in all or the great majority of species and not subject to horizontal gene transfer. But the very existence of such a stable core and more so its actual delineation remain questionable.

In retrospect, the magnitude of apparent horizontal gene transfer in prokaryotes perhaps should not have come as a complete surprise. Indeed, the ability of microbes to absorb DNA from the environment and to integrate it into the genome had been dramatically demonstrated in the Avery-McLeod-McCarthy experiment of 1943 that proved the role of DNA as the genetic material (7). Subsequently, high transformability has been demonstrated for a variety of microbial species (69). Moreover, bacteriophages and plasmids well known to cross-species barriers provide additional, potentially highly effective vehicles for horizontal gene transfer (42, 77, 101). Given the fact that microbes typically coexist in tightly knit communities such as microbial mats and the microflora of animal intestines (49, 79, 93, 104, 111), it appears that opportunities should abound for DNA transfer by various means between diverse prokaryotes and potentially even between eukaryotes and prokaryotes, although in the latter case, the extra complication of getting rid of introns resident in eukaryotic genes is involved.

Despite its obvious growth in prominence with the progress of sequence-based comparative genomics, the issue of horizontal gene transfer as a major evolutionary force remains highly controversial (5, 57, 67). One reason seems to be that the major change in the general picture of evolution seems inevitable if the notion of lateral genomics is vindicated and prompts understandable and perhaps epistemologically justified (the evidence must be fully convincing to justify a paradigm shift) caution in many quarters. The other problem is that, whereas the general significance of horizontal transfer seems to ensue from genome comparisons, the validity of this evolutionary scenario in many individual cases can be questioned owing to uncertainties in phylogenetic tree topologies, unequal evolution rates in different lineages, and other complications.

In this review, we consider the criteria used to ascertain horizontal gene transfer, present conservative quantitative estimates of the amount of likely horizontal gene transfer in the completely sequenced prokaryotic genomes, propose a classification of the distinct types of horizontal transfers, and discuss examples of apparent acquisition of eukaryotic genes by bacteria and archaea.

CRITERIA FOR DETECTING HORIZONTALLY TRANSFERRED GENES

All criteria for identifying probable horizontal gene transfer, or more precisely acquisition of foreign genes by a particular genome, inevitably rely on some unusual feature(s) of subsets of genes that distinguishes them from the bulk of genes in the genome. Traditional tests for horizontal gene transfer involve phylogenetic tree analysis and inherit all the pitfalls typical of these methods. However, the accumulation of multiple genome sequences provides for new, perhaps simpler criteria. When considering these criteria, one has to keep in mind that direct proofs for horizontal gene transfer may be unavailable for the simple reason that there is no record of these evolutionary events other than what could be deciphered by comparison of extant genomes. Therefore all indications for horizontal transfer necessarily remain probabilistic, and the point of using different criteria is maximizing the likelihood of these events being identified correctly.

Unexpected Ranking of Sequence Similarity Among Homologs

The suspicion of horizontal gene transfer usually emerges when a gene sequence (or rather a protein sequence because database searches are typically performed at the protein level) from a particular organism shows the strongest similarity to a homolog from a distant taxon. For example, when all protein sequences encoded in a bacterial genome are compared with the entire protein database and the detected (probable) homologs (or hits in the jargon of computational biology) are classified according to their taxonomic origin, a certain fraction of proteins shows the greatest similarity to eukaryotic homologs rather than to those from other bacteria. The size of this fraction depends, evidently, on the genome and also on the cutoff (usually expressed in terms of alignment score or expect value) used to define “more similar” [the taxonomic breakdown of the best hits for all proteins encoded in completely sequenced prokaryotic genomes is available through the genome division of the Entrez retrieval system (108) at <http://www.ncbi.nlm.nih.gov:80/PMGifs/Genomes/org.html>; then see the “Distribution of BLAST Protein Homologs by Taxa” for individual genomes; see also details below]. These genes make a list of candidates for possible horizontal gene exchange between the given bacterium (or, more precisely, the evolutionary lineage it represents) and eukaryotes. The strength of the claim depends on the cutoff used, but generally the evidence from sequence comparisons should be considered preliminary. To make the case for horizontal transfer convincing, phylogenetic analysis is required.

Unexpected Phylogenetic Tree Topology

Analysis of phylogenetic tree topologies is traditionally the principal means to decipher evolutionary scenarios, including horizontal transfer events (103). Indeed,

if, for example, in a well-supported tree, a bacterial protein groups with its eukaryotic homologs to the exclusion of homologs from other bacteria and, best of all, shows a reliable affinity with a particular eukaryotic lineage, the conclusion that horizontal gene transfer is at play seems inevitable. Moreover, in a convincing case like this, even the most likely direction of transfer, from eukaryotes to bacteria, seems clear. It is unfortunate, however, that phylogenetic analysis does not offer such clear-cut solutions in all suspected cases of horizontal gene transfer, not necessarily even in a majority of these cases. It is common knowledge that phylogenetic methods are prone to a variety of artifacts, perhaps the most notorious being long-branch attraction (78). This phenomenon is particularly relevant for the analysis of probable horizontal gene transfer because these events may be accompanied by accelerated evolution, hence long branches in phylogenetic trees. Tree topology is a good indicator of the probable course of evolution only in cases when the critical nodes are strongly supported statistically, by bootstrap analysis or other methods (11, 26). However, many gene (protein) families seem to have undergone “star evolution,” with short internal branches. In such cases, the actual tree topology remains uncertain and phylogenetic analysis becomes useless for verifying the candidate–horizontal transfer events. On a more practical note, phylogenetic analysis is time and labor consuming, critically depends on correct sequence alignments, and is hard to automate without compromising the quality.

Unusual Phyletic Patterns

With many complete genome sequences available, new and relatively simple, but potentially powerful, approaches to evolutionary analysis become feasible. With the systematic delineation of families of orthologs (direct evolutionary counterparts related by vertical descent), the notion of a phyletic (phylogenetic) pattern has been introduced (29, 30, 105). In the most straightforward formulation, a phyletic pattern is simply the pattern of species present or missing in the given cluster of orthologs. The striking observation made during the construction of the collection of clusters of orthologous groups (COGs) of proteins is the extreme diversity of the observed phyletic patterns (105, 107), most of which include only a small number of genomes (Figure 1). This distribution of COGs by the number of represented species immediately suggests major roles of lineage-specific gene loss and horizontal gene transfer in evolution. Certain types of phyletic patterns, however, appear to signal horizontal transfer in a more specific fashion (Table 1). For example, when a robustly defined set of orthologs shows the presence of a typical “archaeal-eukaryotic” protein in a single bacterial lineage, the odds for the horizontal transfer explanation seem to be high. The B-family DNA polymerase (*E. coli* DNA polymerase II) is a straightforward example of such obvious horizontal gene transfer. The DNA polymerase is a highly conserved protein, and in particular, the similarity between the gamma-proteobacterial Pol II and its archaeal and eukaryotic homologs is highly significant, which seems to render practically irrelevant the typical objections that other bacteria might in fact encode orthologs that have diverged beyond recognition. The presence of the Pol II gene in several

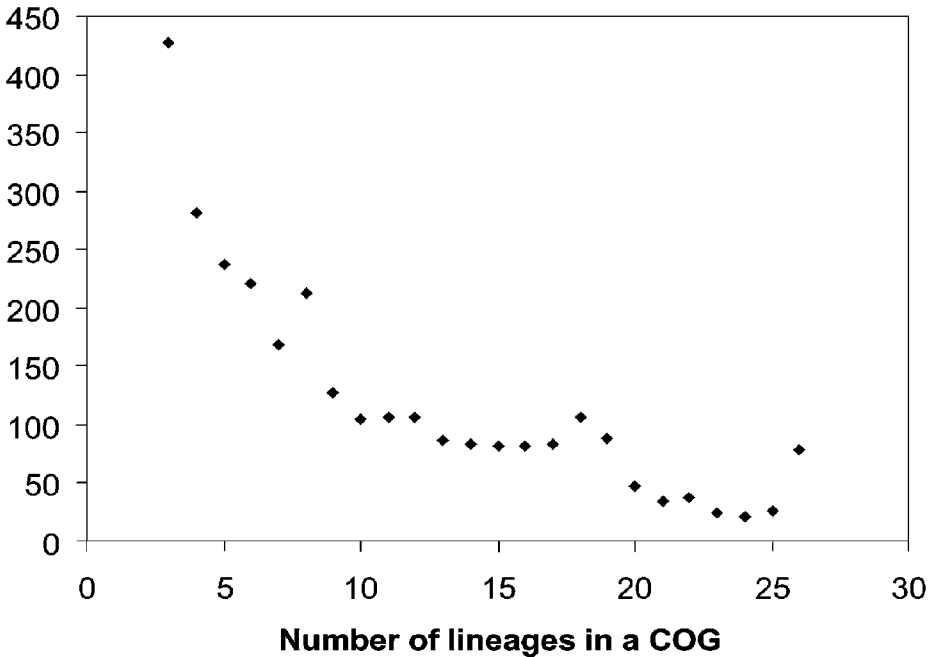


Figure 1 Distribution of the clusters of orthologous groups of proteins (COGs) by the number of represented species. Each COG includes predicted orthologs from at least three genomes that belong to 26 distinct lineages (107).

sequenced gamma-proteobacterial genomes seems to rule out any possibility of artifacts such as contamination of a bacterial genome with eukaryotic sequences. In this particular case, there even seems to be an indication as to possible vehicle of gene transfer because family B polymerases are encoded by numerous bacteriophages and animal viruses (9, 51).

Conservation of Gene Order Between Distant Taxa: Horizontal Transfer of Operons

The evolution of bacterial and archaeal genomes involved extensive gene shuffling, and there is little conservation of gene order between distantly related genomes (18, 47, 81). It has been determined that the presence of three or more genes in the same order in distant genomes is extremely unlikely unless these genes form an operon (118). The same analysis implies that each operon typically emerges only once during evolution and is maintained by selection ever after (59, 60). Therefore, when a (predicted) operon is present in only a few distantly related genomes, horizontal gene transfer seems to be the most likely scenario. If such cases can be confirmed by phylogenetic tree analysis for multiple genes comprising the operon, they figure among the strongest indications of horizontal transfer. Horizontal mobility of operons that encode restriction-modification systems is probably the most

TABLE 1 Examples of probable horizontal gene transfers identified using phyletic patterns in COGs^a

COG	Function	Occurrence in complete bacterial genomes	Occurrence in archaeal and eukaryotic genomes
From archaea or eukaryotes to bacteria			
COG0417	DNA polymerase, B family	Ec, Pa, Vc	All (also viruses and bacteriophages)
COG0430	RNA phosphate cyclase	Ec, Pa, Aa	All
COG0467	KaiC-like ATPases of RecA-superfamily implicated in signal transduction	Ssp, Aa, Tm	All
COG0615	Predicted cytidylyltransferases	Aa, Bs	All
COG1257	Hydroxymethylglutaryl-CoA reductase	Vc, Bb	All
COG1577	Mevalonate kinase	Bb	All
COG2519	Predicted SAM-dependent methyltransferase involved in tRNA ^{Met} maturation	Aa, Mtu	All
From bacteria to archaea (and eukaryotes?)			
COG0847	DNA polymerase III epsilon subunit/domain (3'-5' exonuclease)	All except Bb, mycoplasmas	Af, Sc, Ce, Dm
COG0566	rRNA methylase	All	Af, Sc, Ce, Dm
COG0188	DNA gyrase (topoisomerase II) A subunit	All	Af, Sc, Ce, Dm
COG0187	DNA gyrase (topoisomerase II) B subunit	All	Af, Sc, Ce, Dm
COG0138	AICAR transformylase/IMP cyclohydrolase PurH (purine biosynthesis)	All except Hp, Rp, spirochetes, chlamydia, mycoplasmas	Af, Sc, Ce, Dm
COG0807	GTP cyclohydrolase II (riboflavin biosynthesis)	All except Rp, spirochetes, mycoplasmas	Af, Sc

^aSpecies name abbreviations in this and subsequent tables: Aa, *Aquifex aeolicus*; Af, *Archaeoglobus fulgidus*; Ap, *Aeropyrum pernix*; Bb, *Borrelia burgdorferi*; Bh, *Bacillus halodurans*; Bs, *Bacillus subtilis*; Ce, *Caenorhabditis elegans*; Cj, *Campylobacter jejuni*; Cp, *Chlamydia pneumoniae*; Ct, *Chlamydia trachomatis*; Dm, *Drosophila melanogaster*; Dr, *Deinococcus radiodurans*; Ec, *Escherichia coli*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Mth, *Methanobacterium thermoautotrophicum*; Mtu, *Mycobacterium tuberculosis*; Nm, *Neisseria meningitidis*; Pa, *Pseudomonas aeruginosa*; Rp, *Rickettsia prowazekii*; Sc, *Saccharomyces cerevisiae*; Ssp, *Synechocystis* sp.; Tm, *Thermotoga maritima*; Tp, *Treponema pallidum*; Vc, *Vibrio cholerae*; Uu, *Ureaplasma urealyticum*; Xf, *Xylella fastidiosa*.

TABLE 2 Examples of horizontally transferred operons

Operon	Genes	Phyletic distribution
Nitrate reductase	NarGHJI	Ec, Mtu, Bs, Pa, Ap
Urease	UreABCFG	Mtu, Bs, Bh, Hi, Pa, Hp, Dr, Uu
Formate hydrogen lyase	HyfBCDEFG	Ec, Mtu
Na ⁺ -transporting NADH:ubiquinone oxidoreductase	NqrABCEF	Nm, Vc, Pa, Tm, Cp, Ct
Archaeal/vacuolar-type H ⁺ -ATPase	NtpABCDEFGFIK	All archaea, Dr, Tp, Bb, Ct, Cp
Na ⁺ /H ⁺ -antiporter	MnhABCEFG	Dr, Bs, Bh, Vc, Tm

compelling example of horizontal mobility of operons (52, 82), but for many normal operons, dissemination by horizontal transfer also appears extremely probable; some examples are given in Table 2. The archaeal-type H⁺-ATPase operon is a well-characterized example of such apparent horizontal dissemination of an operon among bacteria, with displacement of the classical bacterial ATPase operon (43, 85).

Anomalous Nucleotide Composition

Anomalous nucleotide composition is widely used but is applicable only to recent horizontal transfers. This approach is based on the “genome hypothesis,” according to which codon usage and GC content are distinct signatures of each genome (35, 36). Thus, genes whose nucleotide or codon composition are significantly different from the mean for a given genome are considered as probable horizontal acquisitions although the likely source of these alien genes generally cannot be identified (31, 61, 76, 80, 84). A significant fraction of prokaryotic genomes, up to 15%–20% of the genes, belongs to this class of recent horizontal acquisitions (31, 61, 76). Many of the horizontally transferred genes revealed by these criteria are prophages, transposons, and other genetic elements for which such evolutionary mobility is not unexpected. This type of horizontal gene transfer and approaches used for its identification have been recently discussed in some detail (84) and are not specifically considered here.

Establishing the Direction of Horizontal Gene Transfer

Difficult as it might be to prove beyond reasonable doubt that horizontal gene transfer has occurred during the evolution of a particular gene family, it is even harder to unequivocally determine which organism is the donor and which one is the recipient in each case. To begin with, the available collection of genome sequences is but a tiny sampling of the genome universe (86), because of which we cannot reasonably hope to identify the true source of any gene present in a given genome. At best, it might be possible to propose a credible hypothesis as to the

donor lineage. The logic used to formulate such hypotheses is based primarily on the “out of Africa” principle (15)—which assumes that if horizontal transfer has indeed occurred, the taxon with the most diverse representation of the given family is the most likely source. The examples in Table 1 were selected to illustrate this approach; these gene families are either widely (usually universally) represented in archaea and eukaryotes but are found in only one or a few bacterial species, or vice versa, which strongly suggests the transfer polarity. However, in the rather common cases of a limited representation of a given family in two distant taxa, horizontal transfer per se may be (almost) indisputable, but the direction cannot be established with any confidence.

CLASSIFICATION AND QUANTIFICATION OF HORIZONTAL GENE TRANSFER EVENTS

Horizontal gene transfer events can be classified into at least three distinct categories with respect to the relationships between the horizontally acquired gene and homologous genes (if any) preexisting in the recipient lineage: (*a*) acquisition of a new gene missing in other members of a given clade, (*b*) acquisition of a paralog of the given gene with a distinct evolutionary ancestry, and (*c*) acquisition of a phylogenetically distant ortholog followed by xenologous gene displacement—that is, elimination of the ancestral gene [xenology has been defined as homology of genes incongruent with the species tree and so implies horizontal gene transfer (32, 88)].

In terms of the actual evolutionary scenarios, the first two classes of events in some cases may reflect nonorthologous gene displacement, that is, acquisition of an unrelated (or distantly related) gene with the same function as an essential ancestral gene typical of the given clade, with subsequent elimination of the latter (54, 55).

We sought to quantitatively assess the amount of horizontal gene transfer in bacterial and archaeal genomes and to classify the transfer events, at least tentatively, into the above categories. With all the caveats discussed above, taxonomic classification of database hits is the only practicable method to identify candidate horizontal gene transfer events on a genome scale. Therefore we applied this approach with conservative cutoffs and identified proteins that are significantly more similar to homologs from other taxa than to those from the taxon to which the given species belongs (hereinafter paradoxical best hits and reference taxa). Protein sets from 31 complete prokaryotic genomes (9 archaeal and 22 bacterial) available at the time of this analysis were extracted from the genome division of the Entrez retrieval system (108) and used as queries to search the nonredundant (NR) protein sequence database at the National Center for Biotechnology Information (National Institutes of Health, Bethesda, MD) with the gapped BLASTP program (1). From the results of these searches, three sets of paradoxical best hits were identified using the Tax_Collector program of the SEALS package

(110). The first set was designed to include candidate horizontal transfers between phylogenetically most distant organisms. Specifically, for nine archaeal species, all proteins were detected whose best hits to bacterial or eukaryotic proteins expect (E)-values were significantly lower than the E-value of the best hit to an archaeal protein (see Table 3). Similarly, for 22 bacterial proteomes, the paradoxical best hits to archaeal and eukaryotic proteins were collected. A separate subset of the paradoxical best hits was formed by proteins with a significant hit ($E < 0.001$) detected only outside the reference taxon.

The second group of paradoxical best hits was to include candidate gene exchange events between major bacterial lineages. With the current state of genome sequencing, this type of analysis is best applicable to small genomes of parasitic bacteria when at least one larger genome sequence of a related species is available. Thus, for *H. influenzae* and *Rickettsia prowazekii*, hits to Proteobacteria (two large proteobacterial genomes, those of *E. coli* and *Pseudomonas aeruginosa*, have been sequenced) were compared with hits to all other bacteria, and paradoxical hits were collected. To isolate likely horizontal transfer events between distantly related bacteria, the paradoxical hits to archaea and eukaryotes included in the first set were subtracted. Similarly, for *Mycoplasma genitalium*, *M. pneumoniae*, and *Ureaplasma urealyticum*, the reference taxon was Firmicutes (gram-positive bacteria; available large genomes—*Bacillus subtilis* and *Mycobacterium tuberculosis*). In the case of the spirochetes *Treponema pallidum* and *Borrelia burgdorferi*, the reference taxon was *Spirochaetales* (no large genome available for this lineage).

The third set was to include paradoxical best hits owing to probable recent horizontal transfers. For this purpose, two closely related species pairs, namely *Chlamydomytila pneumoniae/Chlamydia trachomatis* and *Mycoplasma genitalium/Mycoplasma pneumoniae* were compared. The criteria used to register a paradoxical best hit were the same for all three sets.

Because paradoxical best hits, even with a conservative threshold used in this analysis, provide only a first-approximation estimate of horizontal gene transfer events, a detailed analysis of all candidates was performed for four selected genomes, the bacteria *H. influenzae*, *Vibrio cholerae*, and *A. aeolicus*, and the archaeon *Methanobacterium thermoautotrophicum*. All paradoxical best hits were examined case by case, which involved establishing the phyletic distribution of the corresponding protein family and constructing phylogenetic trees. Multiple protein sequence alignments were constructed with the ClustalW program (109), checked for the conservation of salient sequence motifs, and used for constructing phylogenetic trees, with the neighbor-joining method (94) as implemented in the NEIGHBOR program of the PHYLIP program package (28).

Table 3 shows the tally of candidate interdomain horizontal transfers for all prokaryotic genomes. In most free-living bacteria, the interdomain transfers seemed to involve ~3% of the genes. This fraction was significantly lower in parasitic bacteria, with the exception of *Chlamydia* and *Rickettsia*. In contrast, archaea had a greater fraction of candidate gene transfers of this type, typically between 4% and 8%. It should be emphasized that the protocol used to obtain these numbers

TABLE 3 Candidate horizontal transfers between bacteria, archaea, and eukaryotes: a quantitative assessment^a

Species	Reference taxon	Paralog acquisition or displacement (number and % in genome)	Acquisition of new genes (number and % in genome)
<i>Aeropyrum pernix</i>	Archaea	34 (1.8)	47 (2.5)
<i>Archaeoglobus fulgidus</i>	Archaea	103 (4.3)	100 (4.2)
<i>Methanobacterium thermoautotrophicum</i>	Archaea	100 (5.3)	61 (3.3)
<i>Methanococcus jannaschii</i>	Archaea	43 (2.5)	39 (2.3)
<i>Pyrococcus horikoshii</i>	Archaea	55 (2.7)	39 (1.9)
<i>Pyrococcus abyssi</i>	Archaea	72 (4.1)	39 (2.2)
<i>Thermoplasma acidophilum</i>	Archaea	112 (7.8)	54 (3.7)
<i>Halobacterium</i> sp.	Archaea	204 (8.4)	174 (7.2)
<i>Aquifex aeolicus</i>	Bacteria	87 (5.7)	45 (3.0)
<i>Thermotoga maritima</i>	Bacteria	207 (11.1)	53 (2.9)
<i>Deinococcus radiodurans</i>	Bacteria	47 (1.5)	45 (1.5)
<i>Bacillus subtilis</i> ^b	Bacteria	71 (1.7)	28 (0.7)
<i>Bacillus halodurans</i> ^b	Bacteria	79 (1.9)	40 (1.0)
<i>Mycobacterium tuberculosis</i>	Bacteria	50 (1.3)	62 (1.7)
<i>Escherichia coli</i>	Bacteria	26 (0.6)	13 (0.3)
<i>Haemophilus influenzae</i>	Bacteria	3 (0.2)	3 (0.2)
<i>Rickettsia prowazekii</i>	Bacteria	23 (2.8)	7 (0.8)
<i>Pseudomonas aeruginosa</i>	Bacteria	66 (1.2)	39 (0.7)
<i>Neisseria meningitidis</i>	Bacteria	6 (0.3)	5 (0.2)
<i>Vibrio cholerae</i>	Bacteria	12 (0.3)	16 (0.4)
<i>Xylella fastidiosa</i>	Bacteria	22 (0.8)	8 (0.3)
<i>Buchnera</i> sp.	Bacteria	0 (0.0)	0 (0.0)
<i>Treponema pallidum</i>	Bacteria	10 (1.0)	4 (0.4)
<i>Borrelia burgdorferi</i>	Bacteria	3 (0.4)	6 (0.7)
<i>Synechocystis PCC6803</i>	Bacteria	219 (6.9)	115 (3.6)
<i>Chlamydomonada pneumoniae</i> ^c	Bacteria	23 (2.2)	9 (0.9)
<i>Mycoplasma pneumoniae</i>	Bacteria	0 (0.0)	1 (0.1)
<i>Ureaplasma urealyticum</i>	Bacteria	1 (0.2)	1 (0.2)
<i>Helicobacter pylori</i>	Bacteria	5 (0.3)	3 (0.2)
<i>Campylobacter jejuni</i>	Bacteria	5 (0.3)	4 (0.2)

^aAll protein sequences from each genome were compared with the NR database using the BLASTP program [expect (E)-value cut off 0.001, no filtering for low complexity] and the results were searched for paradoxical best hits, i.e., those that either had a hit to a homolog from a nonreference taxon with an E-value 10 orders of magnitude lower (more significant) than that of the best hit to a homolog from the reference taxon, or had statistically significant hits to homologs from nonreference taxa only. All automatically detected paradoxical best hits were manually checked to eliminate possible false positives.

^bHits to the other *Bacillus* species were disregarded.

^cHits to other Chlamydial species were disregarded.

will detect primarily relatively recent horizontal transfer events because ancient ones (e.g., those that could have occurred prior to the divergence of the analyzed archaeal species) would have been obscured by interarchaeal hits. For the same reason, the fraction of transfers recorded for each of the gamma-proteobacterial species is likely to be an underestimate given that this lineage is represented by several genomes, including two large ones, *E. coli* and *P. aeruginosa*. Given these limitations in detection of paradoxical best hits and the conservative cut-off values used, the level of likely gene exchange between different domains of life seems to be quite substantial.

Several organisms clearly stood out in terms of the number of genes that probably have been horizontally acquired from a different domain of life. In two of these, *Synechocystis* sp. [a cyanobacterium, the progenitor of chloroplasts (38)] and *R. prowazekii* [an alpha-Proteobacterium, the group of bacteria to which the progenitor of the mitochondria is thought to belong (2)], the reported high numbers probably reflect the most obvious interdomain horizontal gene transfers, namely those between chloroplasts and mitochondria and eukaryotic nuclear genomes. As noticed previously, hyperthermophilic bacteria, *A. aeolicus* and especially *Thermotoga maritima*, are significantly enriched in genes apparently horizontally transferred from archaea (4, 83, 119). Conversely, the archaea *Thermoplasma acidophilum* and especially *Halobacterium* sp. appear to possess a much greater number of acquired bacterial genes than other archaeal species, perhaps because these organisms that are moderate thermophiles share their habitats with multiple bacterial species. Along the same line, the difference in the number of acquired genes between two close bacterial species, *Bacillus subtilis* and *B. halodurans*, is notable. The excess of archaeal genes in the latter may be plausibly explained by coexistence with halophilic archaea.

In this genome-wide analysis, apparent acquisition of new genes (cases when a given protein simply has no detectable homologs in the reference taxon) could be automatically distinguished from paralog acquisition/xenologous displacement, but differentiating between the latter was not possible without additional detailed study. Notably, the number of probable events of paralog acquisition/xenologous displacement was generally comparable to that of the acquisition of new genes (Table 3).

The data in Table 4 provide a more complete estimate of probable horizontal gene transfer events by using the respective bacterial taxa as the reference taxa for each bacterial genome and accordingly including gene exchange between major lineages. These estimates are expected to be particularly reliable for small genomes because, with the exception of the spirochetes, larger genomes of related bacteria are available, making it unlikely that the paradoxical best hits are due to an insufficient representation of the given taxon in the sequence database. The estimates of horizontal transfer rate obtained by this approach widely differ, from the modest 1.6% of the genes for *M. genitalium* to the striking 32.6% in *T. pallidum*. The spirochete data could be an overestimate owing to differential gene loss in the two parasitic spirochetes with different lifestyles (100), but in general, a substantial amount of relatively recent horizontal gene transfer seems to be supported by

TABLE 4 Candidate horizontal gene transfers between major bacterial lineages: a quantitative assessment^a

Species	Reference taxon	Paralog acquisition or displacement (number and % in genome)	Acquisition of a new gene (number and % in genome)
<i>Mycoplasma genitalium</i>	Firmicutes	6 (1.2)	2 (0.4)
<i>Mycoplasma pneumoniae</i>	Firmicutes	9 (0.9)	8 (1.2)
<i>Bacillus subtilis</i>	Firmicutes	685 (16.7)	383 (9.3)
<i>Bacillus halodurans</i>	Firmicutes	772 (19.0)	400 (9.8)
<i>Treponema pallidum</i>	Spirochaetales	132 (12.8)	204 (19.8)
<i>Borrelia burgdorferi</i>	Spirochaetales	109 (12.8)	141 (16.6)
<i>Haemophilus influenzae</i>	Proteobacteria	32 (1.9)	21 (1.2)
<i>Rickettsia prowazekii</i>	Proteobacteria	49 (5.9)	32 (3.8)
<i>Escherichia coli</i>	Proteobacteria	223 (5.2)	102 (2.4)
<i>Pseudomonas aeruginosa</i>	Proteobacteria	448 (8.1)	275 (5.0)
<i>Neisseria meningitidis</i>	Proteobacteria	55 (2.7)	34 (1.7)
<i>Vibrio cholerae</i>	Proteobacteria	130 (3.4)	85 (2.2)
<i>Xylella fastidiosa</i>	Proteobacteria	88 (3.2)	83 (3.0)
<i>Buchnera</i> sp.	Proteobacteria	0 (0.0)	0 (0.0)
<i>Mycoplasma genitalium</i>	Mycoplasma	0 (0.0)	0 (0.0)
<i>Chlamydia pneumoniae</i>	Chlamydiales	4 (0.4)	25 (2.4)

^aThe schema for detection of candidate horizontal transfers was the same as in Table 3.

the data. We also estimated the number of more recent horizontal transfer events by collecting the paradoxical best hits for two pairs of closely related bacterial species—*M. genitalium*/*M. pneumoniae* and *C. trachomatis*/*C. pneumoniae*. Predictably, this analysis revealed a smaller number of candidate horizontal transfers, with none at all seen in *M. genitalium* (Table 4).

In an attempt to distinguish between xenologous gene displacement and acquisition of a paralog of a preexisting gene, we performed a phylogenetic analysis of the candidate horizontally transferred genes for three bacterial and one archaeal genomes. In many cases, the phylogenetic tree topology was too complex to make the distinction, but multiple clear-cut cases of both types were identified in each of the analyzed genomes (Table 5 and data not shown). The relative contributions of xenologous gene displacement and paralog acquisition appeared to differ significantly for the compared genomes, with the former phenomenon being prevalent in the hyperthermophilic bacterium *A. aeolicus* and the archaeon *Methanobacterium thermoautotrophicum* and the latter in the parasitic bacterium *V. cholerae* (Table 5).

TABLE 5 Classification of candidate horizontal gene transfer events in selected genomes^a

Species	Reference taxon	Xenologous gene displacement	Acquisition of paralog	Uncertain
<i>Haemophilus influenzae</i>	Proteobacteria	6	6	20
<i>Vibrio cholerae</i>	Proteobacteria	5	40	85
<i>Aquifex aeolicus</i>	Bacteria	31	8	48
<i>Methanobacterium thermoautotrophicum</i>	Archaea	19	17	64

^aParadoxical best hits representing the "paralog acquisition or displacement" category (Table 4, third column) were examined case by case, which involved establishing the phylogenetic distribution of the corresponding protein family and constructing phylogenetic trees. Trees were generated from multiple alignments, constructed using the ClustalW program, with the neighbor-joining method as implemented in the NEIGHBOR program of the PHYLIP program package. If a protein sequence showed a clear phylogenetic affinity (with an at least 60% bootstrap support) to a nonreference taxon, whereas another member of the same protein family from the same species (a paralog of the protein in question) belonged to the reference-taxon branch, a case of probable acquisition of a paralog was recorded. When the given species was represented by a single ortholog in a family and that sequence displayed a clear affinity to a nonreference taxon, this was considered a case of xenologous gene displacement. Other complicated cases and those that were not supported by bootstrap analysis were classified as "uncertain."

The trees in Figure 2 illustrate acquisition of paralogs and xenologous gene displacement. The 3-isopropylmalate dehydrogenase tree (Figure 2A) exemplifies the apparent massive gene exchange between archaea and bacteria, particularly thermophilic ones. Xenologous displacement of the original bacterial gene by the archaeal counterpart seems to have independently occurred in *Aquifex*, *Thermotoga*, and *Deinococcus*. In addition, *Thermotoga* apparently acquired a paralog from a distinct archaeal source. The 6-phosphogluconate dehydrogenase tree (Figure 2B) shows probable xenologous gene displacement within two pairs of relatively close bacterial species, namely *H. influenzae*/*E. coli* and *T. pallidum*/*B. burgdorferi*. In the former case, the evolutionary scenario seems clear because the *H. influenzae* protein belongs to a tight cluster of proteobacterial homologs, whereas the *E. coli* protein unexpectedly falls within the gram-positive lineage. Therefore the displacement event can be confidently mapped to the *E. coli* genome. In contrast, it is hardly possible to determine which of the spirochetes has undergone displacement because no orthologous sequences from other species from the same taxon are currently available.

HORIZONTAL GENE TRANSFER FROM EUKARYOTES TO BACTERIA AND ARCHAEA

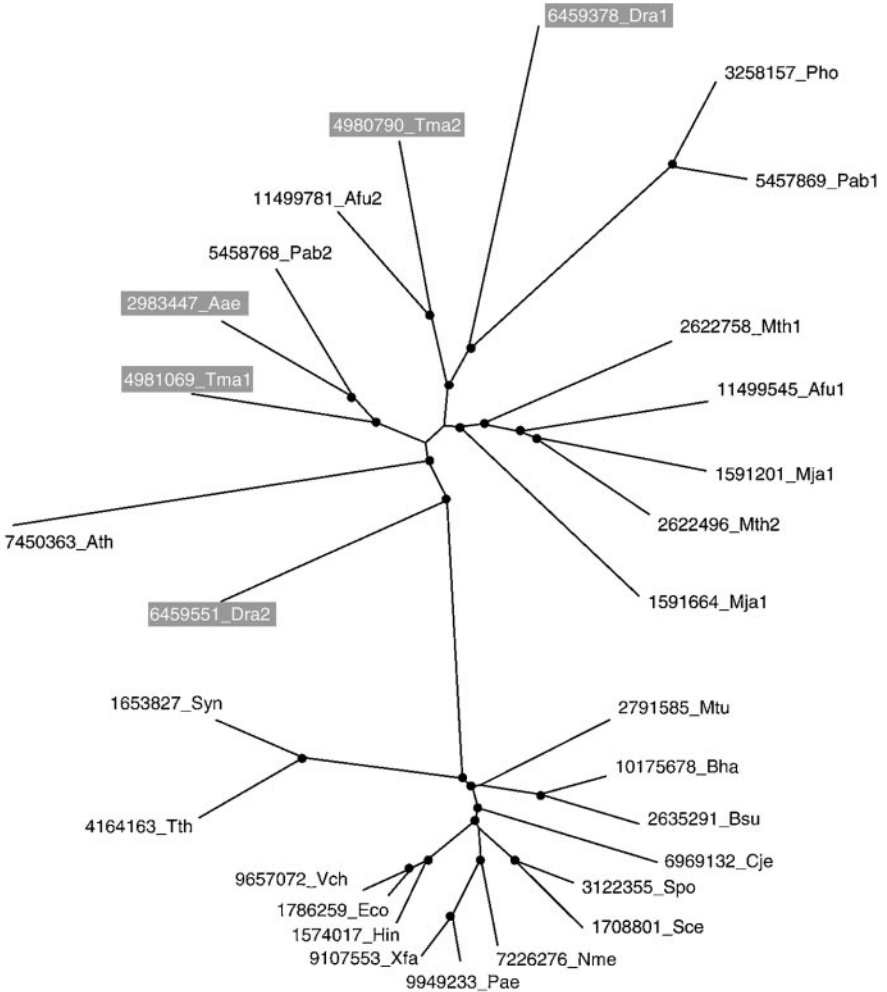
Acquisition of eukaryotic genes by bacteria is potentially of particular interest because of the possible role of such horizontally transferred genes in bacterial pathogenicity (39, 84). The amount of such horizontal transfer evaluated using the

paradoxical best hit approach seems to be relatively small, typically on the order of 1% of the genes in a given prokaryotic genome (Table 6; the high number of “plant” genes in *Synechocystis* is an artifact due to the relationship between Cyanobacteria and chloroplasts). There seems to be a modest but consistent excess of acquired eukaryotic genes in at least some parasites, such as *M. tuberculosis*, *P. aeruginosa*, *Xylella fastidiosa*, and *C. pneumoniae* (Table 1). However, for other parasitic bacteria, such as spirochetes, only a small number of probable acquired eukaryotic genes are detected with this approach. The obvious Cyanobacterium-chloroplast relationship notwithstanding, *Synechocystis* possesses an unusual excess of eukaryotic genes, including those that are otherwise animal specific, such as proteins that share conserved domains with cadherins and other animal extracellular receptors (91). This could suggest that, similar to the extant cyanobacterial symbionts of poriferans, the ancestors of *Synechocystis* have passed through a symbiotic phase (87). It is interesting that some apparent acquisition of eukaryotic genes is seen in each of the archaeal genomes, with the greatest number detected in *Halobacterium* sp. (the archaeal species that also appears to have acquired the greatest number of bacterial genes; see Table 3 and discussion above). A more detailed phyletic breakdown of the eukaryotic acquisitions shows some limited correlations with the parasite-host affinities. For example, there is an apparent excess of animal genes in *P. aeruginosa*, and an excess of “plant” genes in the plant pathogen *X. fastidiosa* (Table 6). A paradoxical situation was observed in *Chlamydia* that seem to have acquired a greater number of genes from plants than from animals. As discussed previously (14), it seems possible that Chlamydiae and their close relatives had a long history of parasitic or symbiotic relationships with eukaryotes and at some stages of their evolution could have been parasites of plants or their relatives.

It should be emphasized that, for the transfer of eukaryotic genes to prokaryotes, the numbers produced in this fashion are likely to be underestimates, perhaps significant, because some of the proteins encoded by the transferred genes may not show highly significant similarity to their eukaryotic ancestors and therefore may be easily missed. This is the case for most signaling proteins discussed below. Apparent acquisition of eukaryotic genes is particularly characteristic of certain functional classes of bacterial genes. Here we discuss two such groups of prokaryotic genes and some notable sporadic cases of apparent horizontal transfer.

Signal Transduction Systems Eukaryotes possess incomparably more versatile signal transduction systems than most bacteria and archaea (19, 45, 62), although in some bacteria, such as Cyanobacteria, Myxobacteria, and Actinomycetes, these systems also show remarkable complexity (25, 44). A detailed survey of the phyletic distribution of eukaryotic signaling domains has shown, rather unexpectedly, their frequent presence (sometimes in highly divergent forms) in prokaryotes (91). Based upon their prevalence among bacteria and archaea, these domains have been classified into those that probably have been inherited from the last universal common ancestor, and those that have evolved in eukaryotes and have been subsequently horizontally transferred to bacteria (and less frequently to archaea). A sampling of the signal transduction domains that show evolutionary mobility

and their typical protein contexts in eukaryotes and prokaryotes is presented in Table 7. In cases when a domain is present and involved in an essential function in all eukaryotes and, in contrast, is found in only one or two bacterial lineages, the case for horizontal transfer appears compelling. The SWIB domain, which is present in subunits of the SWI/SNF chromatin-associated proteins in all eukaryotes (14), was found in only one bacterial lineage, *Chlamydia* (99). In this case, the



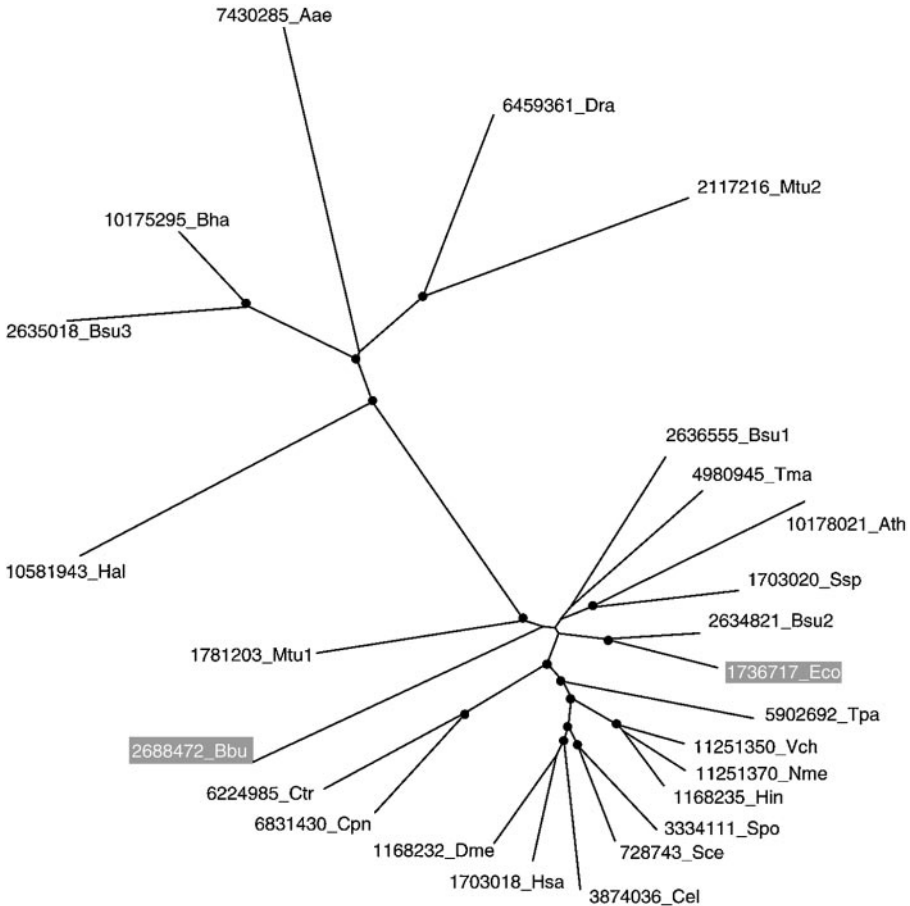
0.1

(4)

Figure 2A

possibility exists that SWIB domain, one copy of which is fused to the chlamydial topoisomerase I, participates in chromatin condensation, a distinguishing feature of this group of intracellular parasites (8). Similarly, the SET domain, a signature eukaryotic histone methylase (92), has been detected in *Chlamydia* (99) and *X. fastidiosa* (95). The role of the SET methylase in these bacteria remains unclear and may point to regulatory mechanisms so far unsuspected in prokaryotes.

Some of the eukaryotic signaling domains acquired by bacteria probably perform functions that, at least mechanistically, are similar to their functions in



0.1

(B)

Figure 2B

eukaryotic systems. This probable functional conservation is exemplified by the phosphoserine–peptide-binding FHA domain (64) whose partners in signal transduction, protein kinases, and phosphatases are represented in both eukaryotes and prokaryotes, indicating a role for this domain in similar phosphorylation-based signaling pathways (63).

Other eukaryotic signaling domains have probably been exapted (34) for completely different functions in bacteria. The examples of such exaptation are the predicted cysteine proteases of two distinct families detected in *Chlamydia* (71, 99; Table 7). In eukaryotes the adenovirus-type protease family is known (65, 66), and the other family has been predicted (71), to participate in the ubiquitin system of controlled protein degradation. However, the ubiquitin pathway does not exist in bacteria, which rules out functional conservation for these proteins. It is most likely that in *Chlamydia*, the proteases contribute to the pathogen host cell interaction as indicated, in particular, by the predicted membrane localization of the adenovirus-family proteases in *C. trachomatis* (99).

Some of the eukaryotic signal transduction domains detected in prokaryotes illustrate the interesting phenomenon of interkingdom domain fusion whereby the acquired eukaryotic domains joined with preexisting bacterial domains within the same protein (117). The aforementioned SWIB domain in *Chlamydia* is a clear-cut case of fusion of a signaling domain acquired from eukaryotes with a typical bacterial enzyme. The presence of two versions of this domain in the bacterial genome,

←

Figure 2 Phylogenetic trees illustrating xenologous gene displacement and acquisition of paralogous gene. (A) 3-Isopropylmalate dehydrogenase. (B) 6-Phosphogluconate dehydrogenase. The rootless neighbor-joining trees (94) were generated using the NEIGHBOR program of the PHYLIP package (28) from multiple alignments that were constructed using the CLUSTALW program (109). The black circles show the nodes with at least 60% bootstrap support (1000 replications). Proteins whose position in the tree is indicative of horizontal gene transfer (see text) are boxed in gray. Each protein is designated by its gene identification number in the nonredundant protein sequence database followed by the species name abbreviation: Aae, *Aquifex aeolicus*; Afu, *Archaeoglobus fulgidus*; Bbu, *Borrelia burgdorferi*; Bsu, *Bacillus subtilis*; Bha, *Bacillus halodurans*; Cje, *Campylobacter jejuni*; Cel, *Caenorhabditis elegans*; Cpn, *Chlamydomophila pneumoniae*; Ctr, *Chlamydia trachomatis*; Dme, *Drosophila melanogaster*; Dra, *Deinococcus radiodurans*; Eco, *Escherichia coli*; Hal, *Halobacterium* sp. NRC-1; Hin, *Haemophilus influenzae*; Has, *Homo sapiens*; Mja, *Methanococcus jannaschii*; Mth, *Methanobacterium thermoautotrophicum*; Mtu, *Mycobacterium tuberculosis*; Nme, *Neisseria meningitidis*; Pab, *Pyrococcus abyssi*; Pho, *Pyrococcus horikoshii*; Pse, *Pseudomonas aeruginosa*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Ssp, *Synechocystis* sp.; Tma, *Thermotoga maritima*; Tpa, *Treponema pallidum*; Tth, *Thermus thermophilus*; Vch, *Vibrio cholerae*; Xfa, *Xylella fastidiosa*.

TABLE 6 Apparent phylogenetic affinities of eukaryotic best hits in bacteria and archaea

Species	Metazoa	Plants	Fungi	Other
<i>Aeropyrum pernix</i>	1	2	0	0
<i>Archaeoglobus fulgidus</i>	5	6	2	2
<i>Methanobacterium thermoautotrophicum</i>	5	3	2	1
<i>Methanococcus jannaschii</i>	2	2	0	0
<i>Pyrococcus horikoshii</i>	3	3	2	1
<i>Pyrococcus abyssi</i>	5	1	2	1
<i>Thermoplasma acidophilum</i>	7	2	2	2
<i>Halobacterium</i> sp.	9	10	5	4
<i>Aquifex aeolicus</i>	0	2	1	0
<i>Thermotoga maritima</i>	2	4	1	1
<i>Deinococcus radiodurans</i>	11	11	8	1
<i>Bacillus subtilis</i>	10	10	6	3
<i>Bacillus halodurans</i>	9	5	3	2
<i>Mycobacterium tuberculosis</i>	16	5	0	2
<i>Escherichia coli</i>	7	2	2	1
<i>Haemophilus influenzae</i>	0	0	0	1
<i>Rickettsia prowazekii</i>	15	6	3	3
<i>Pseudomonas aeruginosa</i>	26	14	16	7
<i>Neisseria meningitidis</i>	1	2	2	0
<i>Vibrio cholerae</i>	4	5	3	0
<i>Xylella fastidiosa</i>	5	10	4	2
<i>Treponema pallidum</i>	3	3	0	2
<i>Borrelia burgdorferi</i>	1	1	1	1
<i>Synechocystis PCC6803</i>	22	167	4	69
<i>Chlamydia pneumoniae</i>	9	12	2	0
<i>Helicobacter pylori</i>	8	2	0	0
<i>Campylobacter jejuni</i>	1	0	1	1
<i>Mycoplasma pneumoniae</i>	0	0	1	0
<i>Ureaplasma urealyticum</i>	0	1	0	0

one that stands alone and one fused to topoisomerase, suggests the probable two-step evolutionary scenario for the origin of such fusions, which includes transfer of a eukaryotic gene followed by recombination with the respective bacterial gene. For reasons that remain unclear, fusion of eukaryotic domains with bacterial ones is particularly common in *Actinomycetes* (117).

Aminoacyl-tRNA Synthetases

Generally, genes coding for the components of the translation machinery appear to belong to the conserved core of the genome that is less prone to horizontal transfer than are other categories of genes (48). An interesting exception is ribosomal protein S14 for which several probable horizontal transfer events have been revealed by phylogenetic analysis (10). However, aminoacyl-tRNA synthetases (aaRSs) are essential components of the translation machinery whose evolution involves horizontal gene transfer as a common trend that could reflect the relative functional autonomy of these enzymes as opposed, for example, to ribosomal proteins that function as subunits of a tight complex (20, 114, 115). Phylogenetic analysis of the aaRS of all 20 specificities indicated that horizontal transfers have probably occurred in almost each case (115). Several of these events involve transfer of an eukaryotic aaRS to bacteria, typically with replacement of the original bacterial gene (Table 8). Acquisition of eukaryotic aaRS genes is most prominent in two groups of parasitic bacteria, Spirochetes and Chlamydia. That parasites have acquired more aaRS genes than free-living bacteria may not be unexpected, but the reasons why this number is unusually high in spirochetes remain unclear. One plausible possibility is that spirochetes have been identified as parasites or symbionts in a wide variety of eukaryotes (73) and might have had a longer history of such relationships than other groups of parasitic and symbiotic bacteria for which complete genome sequences are currently available. However, spirochetes do not seem to show a general preponderance of horizontal gene transfer from eukaryotes compared with other bacteria (Table 6), which suggests that extensive acquisition of aaRS by these organisms could involve some specific, not yet understood features of their biology.

A most unexpected case of probable horizontal transfer from eukaryotes is the TrpRS from the archaeal genus *Pyrococcus*, which includes free-living, hyperthermophilic microbes (115). Given this lifestyle, the presence of a eukaryotic aaRS in the *Pyrococci* (and so far not in any other hyperthermophilic archaea) is almost too unusual to believe, but the result of phylogenetic analysis in this case has been unequivocal. Thus, it seems most likely that *Pyrococcus* indeed has acquired a eukaryotic TrpRS gene from a thermophilic eukaryote such as a polychaete annelid (96).

In all cases, with one notable exception discussed below, the apparent horizontal transfer of eukaryotic aaRS genes into bacteria involves xenologous gene displacement, i.e., the corresponding ancestral bacterial (archaeal) aaRS is never

TABLE 7 Examples of eukaryotic-bacterial transfer of genes coding for proteins and domains involved in signal transduction^a

Domain	Occurrence in eukaryotes	Occurrence in prokaryotes	Prevalent protein context		Functions in eukaryotes
			Eukaryotes	Prokaryotes	
WD40	All	Ssp, Hi, Dr, Tm, Streptomyces, Cenarchium symbiosum (archaeon)	Multiple repeats in various regulatory proteins including G-protein subunits	Stand-alone, multiple repeats, Ser/Thr protein kinases	Various nuclear and cytoplasmic regulatory protein-protein interactions
Ankyrin	All	Ec, Ssp, Bb, Tp, Nm, Rp, Dr, Streptomyces	Multiple repeats in various cytoplasmic and regulatory proteins	Stand-alone, multiple repeats	Various forms of signal transduction including transcription regulation, cell cycle control and PCD
TIR	Animals, plants	Bs, Ssp, Streptomyces Rhizobium	Toll/Interleukin Receptors, PCD adaptors	Stand-alone, AP-ATPases, WD40	Programmed cell death, interleukin signaling
EF-hand	All	Ssp, Streptomyces	Calmodulins, calcineurin phosphatase regulatory subunits, other Ca-binding proteins	Predicted Ca-binding proteins; transaldolase	Various forms of Ca-dependent regulation
FHA (Forkhead associated)	All	Ct, Cp, Mtu, Ssp, Xi, Streptomyces, Myxococcus	Protein Ser/Thr kinases and phosphatase; various nuclear regulatory proteins	Stand-alone, adenylate cyclase, histidine kinase	Phosphoserine binding, protein kinase-mediated nuclear signaling

SET	All	Ct, Cp, Xf	Various multidomain chromatin proteins	Stand-alone	Histone methyltransferase
SWIB	All	Ct, Cp	SWI/SNF complex subunits	Stand-alone, Topoisomerase type I	Chromatin remodeling, transcription regulation
Sec7	All	Rp	Guanine-nucleotide exchange factors; protein transport system components	Stand-alone (protein transport?)	Protein transport, GTPase regulation
Kelch	All	Ec, Hi, Af	Multiple repeats, fusions with POZ and other signaling domains; actin-binding proteins, transcription regulators	Stand-alone, multiple repeats	Transcription regulation; cytoskeleton assembly
Adenovirus-type protease	All; DNA-viruses	Ct, Ec	Stand-alone; ubiquitin-like protein hydrolases	Stand-alone; membrane-associated (Ct)	Regulation of ubiquitin-like-protein-dependent protein degradation
OTU-family protease	All; RNA-and DNA-viruses	Cp	Stand-alone or fused to ubiquitin hydrolase	Large protein with nonglobular domains	Unknown; possible role in ubiquitin-mediated protein degradation
START domains	Animals and plants	Pa	Stand-alone or fused to homeodomains and GTPase regulatory domains	Stand-alone	Lipid binding

^aAn update of the results presented in Reference 91.

TABLE 8 Horizontal transfer of eukaryotic aaRS genes into different bacterial lineages^a

Bacterial group	Horizontally transferred aaRS genes (amino acid specificity)	Comment
Spirochetes	Pro (<i>Borrelia</i> only), Ile, Met, Arg, His, Asn Ser, Glu	Apparent acquisition of the mitochondrial gene
Chlamydia	Ile, Met, Arg, Asn?? Glu	Apparent acquisition of the mitochondrial gene
Bacillus	Asn	
Mycobacteria	Ile, Asn??	
Mycoplasma	Pro, Asn	
γ -Proteobacteria	Asn, Gln	
Helicobacter	His	
Deinococcus	Gln, Asn	
Cyanobacteria	Arg, Asn	
Pyrococcus (archaeon)	Trp	

^aAn update of the results presented in Reference 115.

present along with the eukaryotic one. GlnRS that apparently first emerged in eukaryotes through a duplication of the GluRS gene, and subsequently has been horizontally acquired by gamma-Proteobacteria, is a striking case of nonorthologous gene displacement. In most other bacteria, glutamine incorporation into protein is mediated by a completely different mechanism, namely transamidation, whereby glutamine is formed from Glu-tRNA^{Gln} in a reaction catalyzed by the specific transamidation complex GatABC (16, 46). Thus, in this case, horizontal gene transfer accompanied by nonorthologous gene displacement has resulted in a switch to a completely different pathway of an essential biochemical process. In several Proteobacteria, for which only partial genome sequences are available, and in *Deinococcus radiodurans*, GlnRS and the GatABC complex coexist (41), indicating that in gamma-Proteobacteria the loss of the transamidation system has been a relatively late event compared with the acquisition of the eukaryotic GlnRS gene. The *Deinococcus* system has been studied in detail, and in this organism, the transamidation mechanism only functions to produce Asn-tRNA^{Asn} from mischarged Asp-tRNA^{Asn}, whereas the formation of Gln-tRNA^{Gln} is mediated exclusively by GlnRS (17). Notably, the *Deinococcus* GlnRS contains a fused domain homologous to the C-terminal domain of the tRNA-recognizing GatB protein, which is thought to increase the specificity toward tRNA^{Gln} (72). This case illustrates the intricate functional complexity that may result from gene

acquisition by horizontal gene transfer accompanied by interkingdom domain fusion.

It has been proposed that the topology of some of the aaRS trees, particularly that for IleRS, could be accounted for by postulating just one horizontal gene transfer from eukaryotes, with subsequent dissemination among bacteria (13). This is a plausible hypothesis that is compatible with the reliable clustering of all bacterial species that are suspected to have acquired the respective eukaryotic gene in the IleRS and the HisRS trees. However, the topologies of the trees for MetRS, ArgRS, and Asp-AsnRS are not compatible with the single-transfer scenario, which instead suggests multiple cases of acquisition of the respective eukaryotic genes by different bacterial lineages (115).

The gene for the eukaryotic-type IleRS disseminates through bacterial populations on plasmids, conferring resistance to the antibiotic mupirocin (13). This is one of the rare cases when not only the vehicle of horizontal transfer appears clear, but the nature of the selective pressure that results in the fixation of the transferred gene in the bacterial population seems obvious.

Miscellaneous Eukaryotic Genes Acquired by Prokaryotes

Numerous genes that appear to have been horizontally transferred from eukaryotes to bacteria or archaea cannot be conveniently classified into just one or a few functional categories; examples of such diverse genes are given in Table 9. It should be admitted that, on most occasions, the selective advantage that could be conferred on the prokaryote by the acquired eukaryotic gene cannot be easily gleaned from comparative sequence analysis. Nevertheless, the few exceptions when this is feasible allow biologically interesting inferences. Perhaps the most clear-cut example is the chloroplast-type ATP/ADP translocase detected in *Chlamydia*, *Rickettsia*, and the plant pathogen *X. fastidiosa*. The advantage of having this enzyme for the intracellular parasites *Chlamydia* and *Rickettsia* is obvious because it allows them to scavenge ATP from the host, thus becoming, at least in part, “energy parasites” (113). However, the discovery of the ATP/ADP translocase in *X. fastidiosa* (M.Y. Galperin, V. Anantharaman, L. Aravind & E.V. Koonin, unpublished data) is unexpected and might indicate that such use of the energetic facilities of the host is, after all, not limited to bacteria that grow inside host cells. The acquisition of the ATP/ADP translocase by bacteria from plants is beyond reasonable doubt. Furthermore, it seems that this has occurred on at least two independent occasions. Unexpectedly, phylogenetic analysis of bacterial and plant translocases indicated that *Chlamydia* and *Rickettsia* probably have exchanged these genes (116). The most plausible hypothesis has it that the gene was first acquired from plants by ancestors of *Chlamydia* that might have been plant parasites (see above) and was subsequently passed to *Rickettsia*.

Another case when the adaptive value of the acquired gene seems clear is the sodium/phosphate cotransporter that was detected in *V. cholerae* but not in any other

TABLE 9 Examples of eukaryotic-prokaryotic transfer of functionally diverse genes

Gene function	Representative (gene name species)	Occurrence in prokaryotes	Occurrence in eukaryotes	Apparent source	Type of horizontal transfer
Hydrolase, possibly RNase	AF2335_Af	Af (archaeon) only	Animals, Leishmania	Animal?	Acquisition of paralog
Heme-binding protein	MTH115_Mth	Mth only	Plants, animals	Plant?	Acquisition of a new function
Glutamate-cysteine ligase	XF1428_Xf	Xf, Zymomonas, Bradyrhizobium	Plants	Plant	Xenologous gene displacement
Fructose-bisphosphate aldolase	XF0826_Xf	Xf, Cyanobacteria	Plants, animals	Plant	Nonorthologous gene displacement (of the typical bacterial FBA)
ATP/ADP translocase	XF1738_Xf	Xf, Chlamydia, Rickettsia	Plants	Plant	Acquisition of a new function
Sulfotransferase	BH3370_Bh	Bh	Animals	Animal	Acquisition of a new function
Gamma-D-glutamyl-L-diamino acid endopeptidase I	ENP1_Bs	Bacillus	Animals	Animal	Acquisition of a new function
General stress protein	GsiB_Bs	Bs only	Plants	Plant	Acquisition of a new function
Superfamily I helicase	Cj0945c_Cj	Cj only	Fungi, animals, plants	Eukaryotic	Anquisition of paralog
Guanylate cyclase	Rv1625c_Mtu	Mtu only	Animals, slime mold	Animal	Acquisition of paralog
Purple acid phosphatase	Rv2577_Mtu	Mtu only	Plants, fungi, animals	Plant?	Acquisition of paralog
α/β hydrolase (possible cutinase or related esterase)	Rv1984, Rv3451, Rv2301, Rv1758 (Mtu)	Mycobacteria; multiple paralogs	Fungi	Fungal	Acquisition of a new gene
C-5 sterol desaturase	Slr0224_Ssp	Synechocystis, Mtu, Vibrio	Fungi, animals, plants	Eukaryotic	Acquisition of a new gene
Carnitine O-palmitoyltransferase	MPN114_Mp	Mp only	Fungi, animals	Eukaryotic	Acquisition of a new gene
Arylsulfatase	B1498_Ec	Ec only	Animals	Animal	Acquisition of paralog
Cation transport system component	ChaC_Ec	Ec only	Fungi, animals, plants	Eukaryotic	Acquisition of a new gene
Thiamine pyrophosphokinase	NMB2041_Nm	Nm only	Fungi, animals	Fungal?	Nonorthologous gene displacement?
Phospholipase A2	VC0178_Vc	Vc only	Plants, animals	Plant	Acquisition of paralog
Sodium/phosphate cotransporter	VC0676_Vc	Vc only	Animals	Animal	Acquisition of a new gene
Topoisomerase IB	DR0690_Dr	Dr only	All eukaryotes	Eukaryotic	Acquisition of a new gene
RNA binding protein Ro	DR1262	Dr, Streptomyces	Animals	Animal	Acquisition of a new gene

bacterium; this transporter probably facilitates the survival of the bacterium in the host gastrointestinal tract and could be directly relevant for pathogenesis. A variety of enzymes acquired by bacteria from eukaryotes appear to have been exapted for various functions including interactions with their eukaryotic hosts. One probable example of this is a hemoglobinase-like protease that is encoded by the genome of *Pseudomonas* (PA4016) and that functions as a virulence factor that degrades host proteins. The fukutin-like enzymes found in certain pathogenic bacteria, such as *Haemophilus* and *Streptococcus*, represent another case of horizontally transferred eukaryotic proteins apparently adapted by these bacteria for the modification of their own surface molecules (3). In some cases, such as an α/β -fold hydrolase [either a lipolytic cutinase or a related esterase (68)] that has probably been acquired by *Mycobacterium* from fungi, the horizontally transferred gene has apparently undergone functional diversification on entry into the bacterial lineage through a series of lineage-specific gene duplications.

On a more general note, the acquisition and utilization of eukaryotic genes by prokaryotes shows the remarkable functional plasticity of many cellular systems that is manifested in the compatibility of components evolved in phylogenetically distant organisms. A striking example is topoisomerase IB, an enzyme that is ubiquitous in eukaryotes but that had not been seen in the prokaryotic world until the genome of the extreme radioresistant bacterium *D. radiodurans* had been sequenced (112). Despite the major differences in the protein composition and mechanisms between the bacterial and eukaryotic repair systems (6, 27), this eukaryotic topoisomerase contributes to the ultraviolet resistance of the bacterium and hence apparently does have a function in repair (72).

CONCLUSIONS

Rough estimates based upon the analysis taxon-specific best hits indicate a high level of horizontal gene transfer for most bacterial and archaeal genomes. For selected genomes analyzed in detail, this is largely confirmed by phylogenetic analysis. Probable horizontal transfer events could be classified into the distinct categories of acquisition of new genes and acquisition of paralogs, sometimes followed by nonorthologous gene, and acquisition of phylogenetically distant orthologs followed by xenologous gene displacement. The rates of these different types of horizontal gene transfer events seem to be generally comparable although their relative contributions differ among bacterial and archaeal lineages.

Acquisition of eukaryotic genes by bacterial genomes, particularly those of parasites, and symbionts, and, to a lesser extent, by archaeal genomes, is one of the important directions of horizontal gene flow. Apparent horizontal gene transfer was detected in various functional classes of genes, although it is particularly characteristic of certain categories, such as aminoacyl-tRNA synthetases and different signal transduction systems.

From an evolutionary-theoretical perspective, horizontal transfer, particularly when it occurs between eukaryotes and bacteria, is a testimony to the remarkable

unity of molecular-biological mechanisms in all types of cells that result in the compatibility of eukaryotic and bacterial proteins that have evolved in their distinct milieux for billions of years. Although coadaptation of proteins in the course of evolution might impede horizontal transfer of certain types of genes (e.g., those coding for ribosomal proteins of RNA polymerase subunits), components of many functional systems appear to be fully compatible. One could consider this a moot point given the successful expression of eukaryotic proteins in bacteria routinely exploited in the laboratory. This parallel is not really valid, however, because in order to be fixed in the bacterial population and retained in the long term, a eukaryotic gene must confer selective advantage on the recipient bacterium. This is particularly pertinent for xenologous gene displacement because in these cases, the transferred, heterologous version of a gene should immediately become superior, from the standpoint of selection, to the original version typical of the recipient species. In one case, that of eukaryotic isoleucyl-tRNA synthetase displacing the original gene in some bacteria, this has been convincingly explained by acquisition of antibiotic resistance. It seems likely that these observations have general implications for xenologous gene displacement. In some cases of acquisition of new genes, the nature of the selective advantage also appears clear, such as for the ATP/ADP translocases acquired by intracellular parasitic bacteria, *Chlamydia* and *Rickettsia*. In most instances, however, comparative genomics can only point to the genes that have probably entered the given genome by horizontal transfer. Understanding the biological significance of horizontal gene transfer will require direct experimental studies with these genes.

ACKNOWLEDGMENTS

We wish to offer our apologies and express our appreciation to all researchers whose important work, particularly on individual cases of probable horizontal gene transfer, is not cited here because of inevitable space restrictions. We thank A.S. Kondrashov for useful discussions, M.Y. Galperin and V. Anantharaman for sharing unpublished observations, and Y.I. Wolf for help with some of the computational analyses. Kira Makarova is supported by grants DE-FG02-98ER62583, DE-FG02-97ER62492, and DE-FG07-97ER20293 from the U.S. Department of Energy.

Visit the Annual Reviews home page at www.AnualReviews.org

LITERATURE CITED

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–402
2. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133–40

3. Aravind L, Koonin EV. 1999. The fukutin protein family—predicted enzymes modifying cell-surface molecules. *Curr. Biol.* 9:R836–37
4. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV. 1998. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14:442–44
5. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV. 1999. Reply. Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* 15:299–300
6. Aravind L, Walker DR, Koonin EV. 1999. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* 27:1223–42
7. Avery OT, MacLeod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* 149:297–326
8. Barry CED, Brickman TJ, Hackstadt T. 1993. Hc1-mediated effects on DNA structure: a potential regulator of chlamydial development. *Mol. Microbiol.* 9:273–83
9. Braithwaite DK, Ito J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21:787–802
10. Brochier C, Philippe H, Moreira D. 2000. The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet.* 16:529–33
11. Brown JK. 1994. Bootstrap hypothesis tests for evolutionary trees and other dendrograms. *Proc. Natl. Acad. Sci. USA* 91:12293–97
12. Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* 61:456–502
13. Brown JR, Zhang J, Hodgson JE. 1998. A bacterial antibiotic resistance gene with eukaryotic origins. *Curr. Biol.* 8:R365–67
14. Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, et al. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87:1249–60
15. Cavalli-Sforza LL. 1998. The DNA revolution in population genetics. *Trends Genet.* 14:60–65
16. Curnow AW, Hong K, Yuan R, Kim S, Martins O, et al. 1997. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl. Acad. Sci. USA* 94:11819–26
17. Curnow AW, Tumbula DL, Pelaschier JT, Min B, Soll D. 1998. Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl. Acad. Sci. USA* 95:12838–43
18. Dandekar T, Snel B, Huynen M, Bork P. 1998. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23:324–28
19. Davie JR, Spencer VA. 2000. Signal transduction pathways and the modification of chromatin structure. *Prog. Nucleic Acid Res. Mol. Biol.* 65:299–340
20. Doolittle RF, Handy J. 1998. Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Curr. Opin. Genet. Dev.* 8:630–36
21. Doolittle WF. 1999. Lateral genomics. *Trends Cell Biol.* 9:M5–8
22. Doolittle WF. 1999. Phylogenetic classification and the universal tree. *Science* 284:2124–29
23. Doolittle WF. 2000. Uprooting the tree of life. *Sci. Am.* 282:90–95
24. Doolittle WF, Logsdon JM Jr. 1998. Archaeal genomics: do archaea have a mixed heritage? *Curr. Biol.* 8:R209–11
25. Dunny GM, Leonard BA. 1997. Cell-cell communication in gram-positive bacteria. *Annu. Rev. Microbiol.* 51:527–64
26. Efron B, Halloran E, Holmes S. 1996.

- Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:13429–34
27. Eisen JA, Hanawalt PC. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* 435:171–213
28. Felsenstein J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* 266:418–27
29. Gaasterland T, Ragan MA. 1998. Microbial genescapes: phyletic and functional patterns of ORF distribution among prokaryotes. *Microb. Comp. Genomics* 3:199–217
30. Galperin MY, Koonin EV. 2000. Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotechnol.* 18:609–13
31. Garcia-Vallve S, Romeu A, Palau J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10:1719–25
32. Gogarten JP. 1994. Which is the most conserved group of proteins? Homology-orthology, paralogy, xenology, and the fusion of independent lineages. *J. Mol. Evol.* 39:541–43
33. Golding GB, Gupta RS. 1995. Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* 12:1–6
34. Gould SJ. 1997. The exaptive excellence of spandrels as a term and prototype. *Proc. Natl. Acad. Sci. USA* 94:10750–55
35. Grantham R, Gautier C, Gouy M. 1980. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* 8:1893–912
36. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. 1980. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 8:R49–62
37. Gray MW. 1992. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* 141:233–357
38. Gray MW. 1999. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* 9:678–87
39. Groisman EA, Ochman H. 1997. How *Salmonella* became a pathogen. *Trends Microbiol.* 5:343–49
40. Gupta RS, Golding GB. 1996. The origin of the eukaryotic cell. *Trends Biochem. Sci.* 21:166–71
41. Handy J, Doolittle RF. 1999. An attempt to pinpoint the phylogenetic introduction of glutamyl-tRNA synthetase among bacteria. *J. Mol. Evol.* 49:709–15
42. Hartl DL, Dykhuizen DE, Berg DE. 1984. Accessory DNAs in the bacterial gene pool: playground for coevolution. *Ciba Found. Symp.* 102:233–45
43. Hilario E, Gogarten JP. 1993. Horizontal transfer of ATPase genes—the tree of life becomes a net of life. *Biosystems* 31:111–19
44. Hopwood DA, Chater KF, Bibb MJ. 1995. Genetics of antibiotic production in *Streptomyces coelicolor* A3(2), a model streptomycete. *Biotechnology* 28:65–102
45. Hunter T. 2000. Signaling—2000 and beyond. *Cell* 100:113–27
46. Ibbá M, Curnow AW, Soll D. 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. *Trends Biochem. Sci.* 22:39–42
47. Itoh T, Takemoto K, Mori H, Gojobori T. 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.* 16:332–46
48. Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96:3801–6
49. Jeanthon C. 2000. Molecular ecology of hydrothermal vent microbial communities. *Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* 77:117–33
50. Kaneko T, Tabata S. 1997. Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol.* 38:1171–76

51. Knopf CW. 1998. Evolution of viral DNA-dependent DNA polymerases. *Virus Genes* 16:47–58
52. Kobayashi I, Nobusato A, Kobayashi-Takahashi N, Uchiyama I. 1999. Shaping the genome—restriction-modification systems as mobile genetic elements. *Curr. Opin. Genet. Dev.* 9:649–56
53. Koonin EV, Aravind L, Kondrashov AS. 2000. The impact of comparative genomics on our understanding of evolution. *Cell* 101:573–76
54. Koonin EV, Mushegian AR. 1996. Complete genome sequences of cellular life forms: glimpses of theoretical evolutionary genomics. *Curr. Opin. Genet. Dev.* 6:757–62
55. Koonin EV, Mushegian AR, Bork P. 1996. Non-orthologous gene displacement. *Trends Genet.* 12:334–36
56. Koonin EV, Mushegian AR, Galperin MY, Walker DR. 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.* 25:619–37
57. Kyrpides NC, Olsen GJ. 1999. Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* 15:298–99
58. Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33:351–97
59. Lawrence JG. 1997. Selfish operons and speciation by gene transfer. *Trends Microbiol.* 5:355–59
60. Lawrence JG. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.* 9:642–48
61. Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44: 383–97
62. Lengeler KB, Davidson RC, D'Souza C, Harashima T, Shen WC, et al. 2000. Signal transduction cascades regulating fungal development and virulence. *Microbiol. Mol. Biol. Rev.* 64:746–85
63. Leonard CJ, Aravind L, Koonin EV. 1998. Novel families of putative protein kinases in bacteria and archaea: evolution of the “eukaryotic” protein kinase superfamily. *Genome Res.* 8:1038–47
64. Li J, Lee GI, Van Doren SR, Walker JC. 2000. The FHA domain mediates phosphoprotein interactions. *J. Cell Sci.* 113:4143–49
65. Li SJ, Hochstrasser M. 1999. A new protease required for cell-cycle progression in yeast. *Nature* 398:246–51
66. Li SJ, Hochstrasser M. 2000. The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol. Cell. Biol.* 20:2367–77
67. Logsdon JM, Faguy DM. 1999. Thermotoga heats up lateral gene transfer. *Curr. Biol.* 9:R747–51
68. Longhi S, Cambillau C. 1999. Structure-activity of cutinase, a small lipolytic enzyme. *Biochim. Biophys. Acta* 1441:185–96
69. Lorenz MG, Wackernagel W. 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* 58:563–602
70. Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, et al. 1999. Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* 9:608–28
71. Makarova KS, Aravind L, Koonin EV. 2000. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* 25:50–52
72. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, et al. 2001. The genome of the extremely radioresistant bacterium *Deinococcus radiodurans*: comparative genomics. *Microbiol. Mol. Biol. Rev.* 65:44–79

73. Margulis L, Chase D, To LP. 1979. Possible evolutionary significance of spirochaetes. *Proc. R. Soc. London Ser. B* 204: 189–98
74. Martin W. 1999. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *BioEssays* 21:99–104
75. Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 118:9–17
76. Medigue C, Rouxel T, Vigier P, Henaut A, Danchin A. 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* 222:851–56
77. Moreira D. 2000. Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. *Mol. Microbiol.* 35:1–5
78. Moreira D, Philippe H. 2000. Molecular phylogeny: pitfalls and progress. *Int. Microbiol.* 3:9–16
79. Moyer CL, Dobbs FC, Karl DM. 1995. Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl. Environ. Microbiol.* 61:1555–62
80. Mrazek J, Karlin S. 1999. Detecting alien genes in bacterial genomes. *Ann. NY Acad. Sci.* 870:314–29
81. Mushegian AR, Koonin EV. 1996. Gene order is not conserved in bacterial evolution. *Trends Genet.* 12:289–90
82. Naito T, Kusano K, Kobayashi I. 1995. Selfish behavior of restriction-modification systems. *Science* 267:897–99
83. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323–29
84. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304
85. Olendzenski L, Liu L, Zhaxybayeva O, Murphey R, Shin DG, Gogarten JP. 2000. Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. *J. Mol. Evol.* 51:587–99
86. Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–40
87. Paerl HW. 1996. Microscale physiological and ecological studies of aquatic cyanobacteria: macroscale implications. *Microsc. Res. Technol.* 33:47–72
88. Patterson C. 1988. Homology in classical and molecular biology. *Mol. Biol. Evol.* 5:603–25
89. Pennisi E. 1998. Genome data shake tree of life. *Science* 280:672–74
90. Pennisi E. 1999. Is it time to uproot the tree of life? *Science* 284:1305–7
91. Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV. 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.* 289:729–45
92. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 406:593–99
93. Risatti JB, Capman WC, Stahl DA. 1994. Community structure of a microbial mat: the phylogenetic dimension. *Proc. Natl. Acad. Sci. USA* 91:10173–77
94. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–25
95. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28:231–34
96. Sicot FX, Mesnage M, Masselot M, Exposito JY, Garrone R, et al. 2000. Molecular adaptation to an extreme environment: origin of the thermal stability of

- the pompeii worm collagen. *J. Mol. Biol.* 302:811–20
97. Smith MW, Feng DF, Doolittle RF. 1992. Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* 17:489–93
 98. Sprague GF Jr. 1991. Genetic exchange between kingdoms. *Curr. Opin. Genet. Dev.* 1:530–33
 99. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, et al. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* 282:754–59
 100. Subramanian G, Koonin EV, Aravind L. 2000. Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. *Infect. Immun.* 68:1633–48
 101. Sundstrom L. 1998. The potential of integrons and connected programmed rearrangements for mediating horizontal gene transfer. *APMIS Suppl.* 84:37–42
 102. Syvanen M. 1985. Cross-species gene transfer; implications for a new theory of evolution. *J. Theor. Biol.* 112:333–43
 103. Syvanen M. 1994. Horizontal gene transfer: evidence and possible consequences. *Annu. Rev. Genet.* 28:237–61
 104. Tannock GW. 2000. The intestinal microflora: potentially fertile ground for microbial physiologists. *Adv. Microb. Physiol.* 42:25–46
 105. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–37
 106. Tatusov RL, Mushegian AR, Bork P, Brown NP, Hayes WS, et al. 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* 6:279–91
 107. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29:22–28
 108. Tatusova TA, Karsch-Mizrachi I, Ostell JA. 1999. Complete genomes in WWW Entrez: data representation and analysis. *Bioinformatics* 15:536–43
 109. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–80
 110. Walker DR, Koonin EV. 1997. SEALS: a system for easy analysis of lots of sequences. *Intell. Syst. Mol. Biol.* 5:333–39
 111. Ward DM, Ferris MJ, Nold SC, Bateson MM. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* 62:1353–70
 112. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, et al. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* 286:1571–77
 113. Winkler HH, Neuhaus HE. 1999. Non-mitochondrial ATP transport. *Trends Biochem. Sci.* 24:64–68
 114. Woese CR, Olsen GJ, Ibba M, Soll D. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* 64:202–36
 115. Wolf YI, Aravind L, Grishin NV, Koonin EV. 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
 116. Wolf YI, Aravind L, Koonin EV. 1999. Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* 15:173–75
 117. Wolf YI, Kondrashov AS, Koonin EV. 2000. Interkingdom gene fusions. *Genome Biol.* 1:research0013.1–0013.15
 118. Wolf YI, Rogozin IB, Kondrashov AS,

Koonin EV. 2001. Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. *Genome Res.* 11:356–72

119. Worning P, Jensen LJ, Nelson KE, Brunak S, Ussery DW. 2000. Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*. *Nucleic Acids Res.* 28:706–9