Mosaic bacterial chromosomes: a challenge en route to a tree of genomes

William Martin

Summary

In a recent analysis J.G. Lawrence and H. Ochman [Proc Natl Acad Sci USA 1998;95:9413–9417 (Reference 1)] surmised that about 10% of the current *E. coli* genome consists of genes that were acquired in over 200 events of lateral gene transfer, which occurred subsequent to the divergence of *E. coli* and *Salmonella* some 100 million years ago. Overall, the data suggest that no less than 18% of *E. coli*'s genes might be relatively recent foreign acquisitions, and that the average rate of acquisition may be close to about 16 kb per million years. These quantitative estimates of comparatively recent genome flux have profound impact on evolutionary genome comparisons. They tend to suggest that a search should be on to identify principles that might ultimately govern gene distribution patterns across prokaryotic genomes. *BioEssays* 1999; 21:99–104. © 1999 John Wiley & Sons, Inc.

Introduction

''The structure of genetic variation in a bacterial species'', as an experienced *E. coli* geneticist described it,[2] "is the result of recombination superimposed upon the repeated formation and spread of clones." Notably, the statement does not read "recombination between individuals of the same species"; rather, simply "recombination." For bacterial geneticists, horizontal gene transfer between distantly related species is nothing new. To introduce foreign genes into *Synechocystis* PCC6803 (a cyanobacterium), for example, one transforms a plasmid into *E. coli*, grows a culture of the *E. coli* cells, and mixes them overnight with some *Synechocystis* cells in the light. If the *E. coli* strain carries the right genes to produce sex pili, the *E. coli* cells will mate with the cyanobacteria and transfer the manipulated plasmid to them (transconjugation). If the restriction/modification systems of the mating partners are properly matched, and if there are short stretches of identical (or nearly so) sequences on plasmid and the cyanobacterial chromosome, the foreign DNA will be stably integrated into the cyanobacterial genome so that transconjugants can be selected. Natural gene transfer, overnight, as horizontal as it gets, under laboratory conditions.

But the fact that horizontal gene transfer occurs in the lab bears neither upon its occurrence nor its prevalence in nature. Although previous studies had clearly documented individual and evolutionarily recent cases of horizontal transfer in *E. coli* and *Salmonella*,[3–5] the whole-genome question of just how much genetic promiscuity has gone on in *E. coli*'s more distant past (and that of other bacteria) has been a nagging one, particularly to evolutionary biologists. Lawrence and Ochman[1] have obtained some answers that are both exciting and ominous. But they did so without the help of phylogeny-oriented comparisons of genes in many genomes and searches for unusual topologies, so it is worthwhile to briefly summarize their approach.

How to tell which genes might be alien

Foreign genes in bacterial chromosomes, if recently introduced, can betray their intruder status in two ways that are independent of sequence similarity comparisons to reference genomes: GC content and codon bias. The GC-content of bacterial genomes varies substantially across species, but within a given bacterial genome, it tends to be quite uniform. This is generally attributed to the cumulative effects of countless rounds of DNA replication and repair by the polymerases and repair enzymes specific to a given species, a process known as mutational bias.[6,7] If a gene from a

Funding agency: Deutsche Forschungsgemeinschaft.
Correspondence to: William Martin, Institut für Genetik, Technische Universität Braunschweig, Spielmannstr. 7, D-38023 Braunschweig, Federal Republic of Germany. E-mail: w.martin@tu-bs.de

donor species is introduced into a new recipient chromosome with differing GC-content, it can be detected as alien through computer analysis by virtue of its distinct GC-content. But once introduced, it begins a process called amelioration—through mutation it gradually becomes more similar in GC-content to the rest of the genome due to the mutational bias of the recipient. Over time, amelioration will go to completion and the gene will eventually become undetectable at the level of GC-content as having stemmed from a foreign genome.

Similar considerations apply to bias for synonymous codons. Different bacterial species tend to preferentially express different isoaccepting tRNAs for the same amino acid. Those genes that are highly expressed by a given genome are highly adapted at the level of codon preference to the abundantly expressed isoaccepting tRNAs of that genome; less abundantly expressed genes show a much less pronounced codon preference.[8,9] In the highly expressed *E. coli gapA* gene, for example, there are 20 codons encoding leucine, but of the six possible codons that could be used, only two are (CTG 19 times, TTA once). Different bacterial species prefer different sets of isoaccepting tRNAs, such that the *gapA* gene in *Bacillus stearothermophilus*, for example, is also highly biased at leucine codons, but has no CTG and 22 TTA codons. Thus, a gene that has been recently acquired by a bacterial genome with different preferences will initially possess a detectably anomalous codon bias, that over time will also undergo amelioration and adaptation to the recipient's codon and tRNA preferences.

Gene flux on a large scale

Using such benchmarks, Lawrence and Ochman found that 755 of *E. coli*'s 4,288 protein-coding genes deviate significantly from the overall average endogenous GC-content and codon bias of the genome[1] and—by these criteria—have apparently been acquired from outside sources through 234 independent lateral transfer events. With the help of some theoretical tools, they modeled the amelioration process over time, thus permitting an estimation of approximately how long ago each intruder was acquired. About half of the foreign sequences seem to have been acquired extremely recently in evolution, less than about 1 million years ago; the vast majority of the foreign DNA was acquired less than about 10 million years ago, and—by their criteria—only a small fraction of genes appears to have persisted in the genome from more ancient acquisitions. Taking various factors into account, Lawrence and Ochman estimate that since the divergence of *E. coli* and *Salmonella*, foreign DNA has been flowing into the *E. coli* chromosome at a rate of about 0.016 Mb (million base pairs) per million years, whereby most—but not all—of it seems to be deleted just as rapidly. Using this influx rate, one can estimate that the amount of DNA currently contained within the *E. coli* MG1655 genome (4.6 mb) flows into that genome during the course of about 300 million years. That would mean that the *E. coli* chromosome size would double during that period, were no sequences deleted in turn.

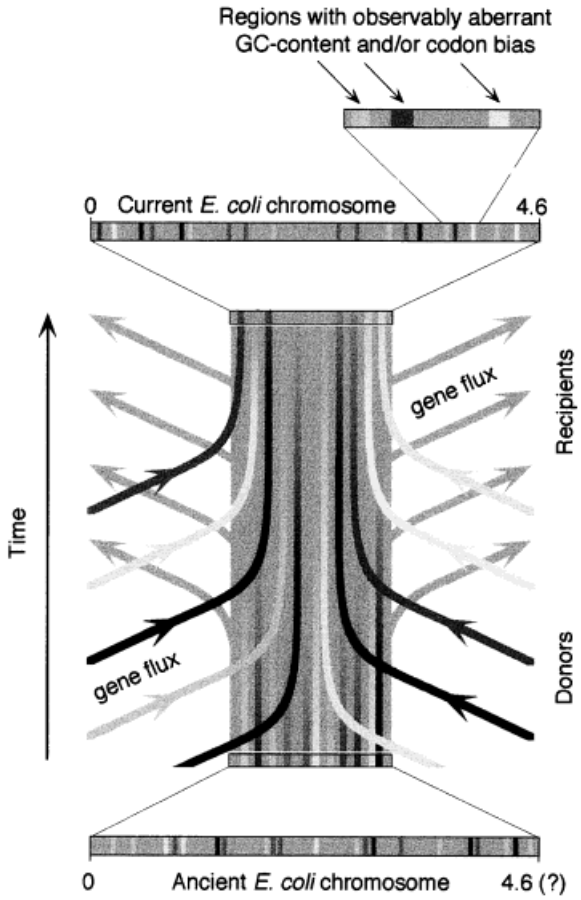

Figure 1. Schematic representation of gene flow through the *E. coli* chromosome through time. Foreign acquisitions are represented as arrows flowing into the chromosome; gene loss and donations to other bacteria are represented as arrows flowing out of the chromosome. Different gray shades symbolize GC-content and/or codon bias of donor DNA that differs from that of *E. coli*. The amelioration process is symbolized as a gradual blend of foreign acquisitions toward the average *E. coli* content and bias (intermediate gray). The figure makes no statement about mechanism(s) of acquisition.

Converting these findings into a naive and schematic picture of the *E. coli* chromosome over time might result in something like Figure 1. Foreign sequences of aberrant GC-content and/or codon bias (indicated by different gray shades in the figure) pour into the chromosome over time, become more similar to the rest of the genome in these attributes if they remain, and—in order to keep the chromosome from inflating to an unacceptable or inefficient size—sequences pour out of the chromosome simultaneously.

Certainly, some portions of the deleted DNA will simply be lost, while others will have found a suitable home in another genome that received genes from *E. coli*, rather than having donated to it. Obviously, donors and recipients involved in gene flux through the *E. coli* chromosome could in principle belong any number of distinct bacterial groups, such that the complete network of gene exchange between all participants would be very difficult to graphically represent, were it known (which it is not).

Clearly, the bold quantitative estimates put forth by Lawrence and Ochman are subject to various parameters of uncertainy that will be the subject of future critical inspection by others. For example, if the estimated divergence time for *E. coli* and *Salmonella* used (100 million years) is imprecise, the absolute influx rate estimates would require revision. Similarly, if the amelioration rate estimation is off the mark, so would be the influx rates inferred from it. Furthermore, a fraction of genes should deviate from the average GC-content of the genome by chance, and if processes other than lateral transfer influence local GC-content around the chromosome, they might mimick lateral acquisitions. But at the same time, the approach used to detect foreign sequences will pick up only such genes as were obtained from donors with GC-content and/or codon bias that differ significantly from that of *E. coli*—lateral acquisitions from genomes of similar bias will go undetected altogether. So, there is no immediate cause for crying out that the degree or rate of lateral acquisition was perspicuously overestimated, and it is possible that it may have been underestimated.

Bacterial chromosomes through time

Let us simply assume, for the purposes of the paper, 1) that the rates and quantities of acquired DNA in *E. coli* are quite real; and 2) that they may be representative for other bacterial genomes in general, and, on the basis of those assumptions, consider some of the ramifications of the findings.

From the standpoint of comparative genome analyses, evolutionarily recent horizontal transfer of the magnitude described for *E. coli* has something quite ominous about it, when it is projected—as a continuum—into the depths of geological time. If we consider a hypothetical succession of gene transfers between closely and distantly related eubacteria from the origin of this group several billion years ago to the present, it is immediately apparrent that countless new combinations of genes in eubacterial genomes will have arisen, which can confer new attributes to progeny that are 1) heritable, and 2) selectable.

This is somehow reminiscent of the combination and distribution of alleles-in-eukaryotes across distinct but potentially interbreeding (diploid) populations, for which there are many good mathematical models.[10] But mathematical models that could describe transspecific combinations and distributions of genes-in-eubacterial-genomes will necessarily possess variables that do not exist in the case of mendelian systems.

For example, probabilities of gene exchange will in some way relate to ecological specialization, because only if partners can physically meet under their specific life-supporting circumstances will they be likely to transfer genes in the first place. But as Lawrence and Ochman point out, newly acquired genes can themselves provide access to new ecological niches, opening up the door to transfer with new partners. In some new niches, many classes of preexisting genes might become expendable, leading to very poorly predictable variation in the number of loci between generations. It is also evident that the modular organization of biochemical pathways will eventually figure into these matters, but it is not evident how.

And whereas contemporary eukaryotes generally restrict themselves to acquiring and transmitting alleles within the genetic confines of their own species (meiosis ensures that), eubacteria appear to be much less choosy about which species they mate with, so that—at the extreme—new (and potentially useful) combinations of genes from a plethora of sources could, in principle, be assembled from countless, highly divergent, and independent donors within a given eubacterial chromosome. Over time, that would yield purely patchwork genomes. Is that the way that eubacterial genomes evolve? Are there no barriers to lateral transfer?

Although comparative genomics has yet to provide a general picture of how genomes evolve in toto, there clearly are barriers to interspecies transfer in bacteria, at least as far as the frequency of sexual transmission goes. Several recent papers indicate that a major barrier helping to maintain *E. coli* and *Salmonella* as distinct species may simply be sequence divergence (point mutations) in homologous DNA regions.[11,12] Since genes in the chromosomes of *E. coli* and *Salmonella* are not identical in sequence, during the process of homologous recombination, mismatches occur in the heteroduplex formed between donor and recipient DNA. These mismatches are recognized by the mismatch repair machinery[13] that identifies the intruding DNA strand in the heteroduplex as foreign so that it can be degraded, maintaining a genetically clean chromosome. The greater the sequence divergence between recipient and donor sequences, the lower the frequency of successful recombination[14]—in the lab. A similar relationship between sequence divergence and transspecific recombination frequency has also been observed for the gram-positive genus *Bacillus*[15]; but in that system, mechanisms other than mismatch repair seem to govern transspecific recombination frequencies—in the lab.

It is not known whether the intruding genes observed in *E. coli* were acquired by conjugation or by other mechanisms possibly involving phages[1] (transduction), and there are reasons to believe that the relative roles of conjugation and transduction in genetic isolation may differ between *E. coli*

and *Bacillus*.[15] But even if transfer events—irrespective of mechanism—between distantly related bacterial species in the wild are excruciatingly rare, in those cases where they confer a distinct selective advantage to recipients in a given ecological niche, they will tend to be rapidly fixed and clonally propagated, as in the case of surface antigens in *Salmonella*,[4,5] or pathogenicity islands in other bacteria.[16,17] Worse yet, regardless of how rare a given recombination event might be in terms of events per genome and generation, the hefty numbers of bacterial individuals and the inexhaustible patience of geological time would tend to weigh in favour of its having occurred at some time in the evolutionary past—a disturbing implication.

From prokaryotes to eukaryotes: Inheritance of (transient?) acquisitions

The impact of horizontal gene transfer between eubacteria extends firmly into the realm of eukaryotic genomes as well, since eukaryotes possess numerous nuclear genes that were acquired from eubacteria. Many nuclear genes entered the eukaryotic lineage through the genomes of the eubacterial antecedents of mitochondria and chloroplasts, and were simply transferred to the nucleus in the process of symbiosis.[18] This argument is straightforward for nuclear genes that, in phylogenetic analyses, branch robustly with cyanobacterial and α-proteobacterial homologues, respectively, for example the nuclear-encoded chaperonins hsp60[19] and hsp70.[20] It is even more straightforward when the nuclear genes branch with homologues that are still encoded in one or the other organelle genome.[19,21]

But there are cases in which the eubacterial donor of a nuclear gene is difficult to identify, notwithstanding phylogenetic methodological issues concerning the limited amount of information contained within any individual gene or protein.[21,22] Recent analyses have shown that there are a number of genes in the yeast genome that are clearly eubacterial, but not specifically proteobacterial in origin, at least on the basis of comparisons to the proteobacterial genomes that have been sequenced to date (n.b.: at the time of writing, no α-proteobacterial genome has been published as a reference for mitochondria). Riviera et al.,[23] for example, found that there are a few genes in the yeast nuclear genome that share more similarity with homologues found in the *Synechocystis* genome than with homologues found in the *E. coli* or *Haemophilus* genomes. Does this mean that yeast once possessed a plastid and is secondarily nonphotosynthetic? At face value, that is one straightforward interpretation of such findings. However in light of rates and amounts laterally acquired DNA observed in *E. coli*,[1] another very real and equally straightforward interpretation becomes apparent: It is quite possible that such genes that were transferred to (what became) the yeast nuclear genome from the α-proteobacterial antecedents of mitochondria, but that the same genes in the free-living relatives of those antecedents were subsequently exchanged between eubacteria in such a manner that they are now found in (at least one) cyanobacterial genomes, rather than in (currently characterized) proteobacterial genomes.[24]

Clearly, the origin of mitochondria involved a straightforward sampling process of a genome-sized aliquot of eubacterial genes by the host's genome.[24] But ancient transfer events *between eubacteria* complicate matters surrounding the biological source (hence the biological context of acquisition) of eubacterial genes in eukaryotic chromosomes, as schematically depicted in Figure 2. This is not a trivial issue, because there are many more genes of eubacterial origin in eukaryotic chromosomes than can be accounted for by traditional formulations of the endosymbiont hypothesis.[25,26] Three recent studies arrived independently at the conclusion that of those genes in the yeast genome for which a statement is currently possible, about two-thirds appear to be eubacterial in origin, whereas about one-third are archaebacterial.[23,30,31] Earlier large-scale gene-by-gene comparisons[25,32] of homologues common to eukaryotes, archaebacteria, and eubacteria—but including genes from incompletely sequenced genomes—had reached very similar conclusions.

There are currently three explanations for the existence of too many eubacterial genes in eukaryotic chromosomes that: a) differ in views on the origin of eukaryotes, b) differ in views concerning the constitution of the host's genome proir to the acquisition of mitochondria,[33] and c) generate different predictions about the relationships between genes *that are common to prokaryotes and eukaryotes*. 1) It is possible that eubacterial genes in eukaryotic genomes were acquired via lateral transfer from prokaryotes that were ingested by eukaryotes as food particles,[34] in which case different eukaryotic lineages can be predicted to have acquired and inherited genes from many distinctly different eubacterial sources. 2) It is possible that the eubacterial genes of eukaryotes descend from an ancestral pool of freely exchangeable genes that was simply sorted out into the genomes of cells that became the ancestors of archaebacteria, eubacteria, and eukaryotes—as argued by Kandler[35] and more recently by Woese[36]—in which case the diversity of genes common to these groups can be predicted be *roughly* equal in all three lineages (depending upon whether eukaryotes are viewed as being equally as old as prokaryotes,[36] or somewhat younger by virtue of their heterotrophy[35]). 3) It is possible that the eubacterial fraction of eukaryotic nuclear genes simply stems from the α-proteobacterial antecedant of mitochondria,[26,37] in which case gene diversity in eukaryotes can be predicted to be a distinct and commonly inherited subset of prokaryotic gene diversity.

Notably, many eubacterial genes in eukaryotic chromosomes encode cytosolic gene products.[18] Classical formulations of the endosymbiont hypothesis concerning the origin of

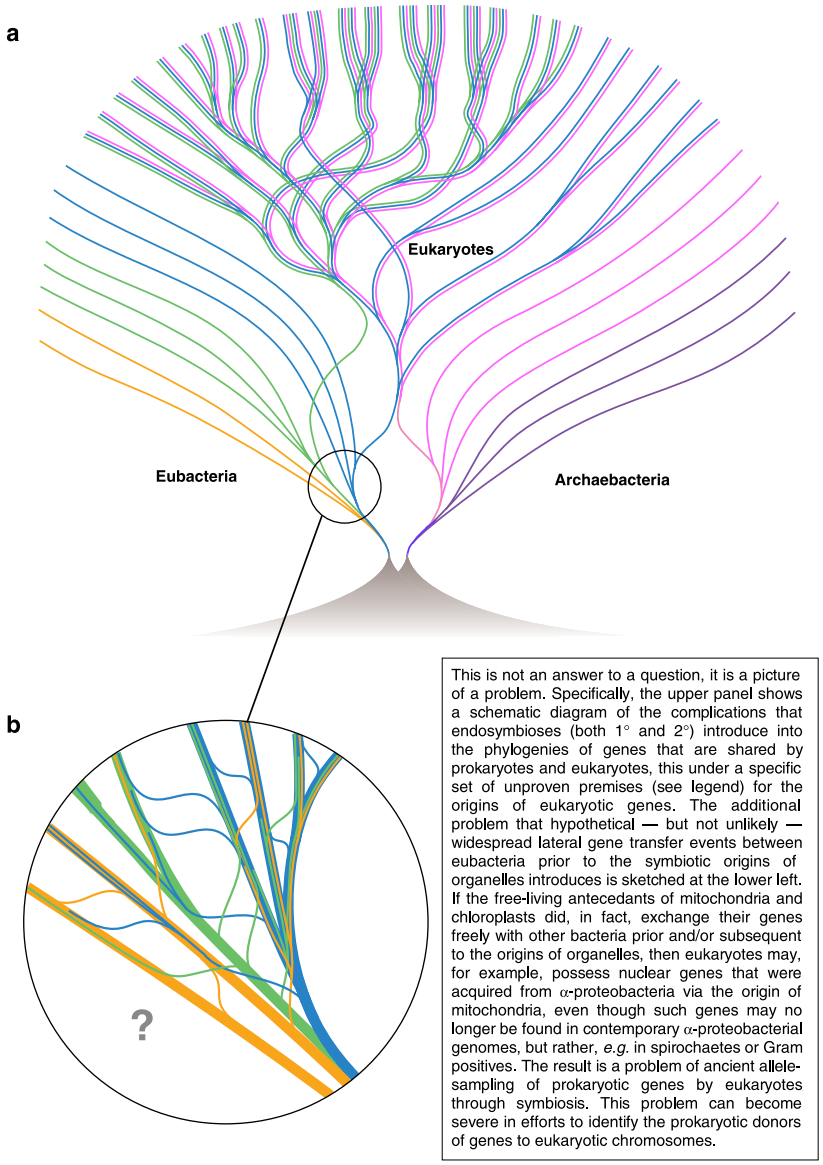

Figure 2. A tree of genomes. Each prokaryotic genome is represented as a single colored line; different colors symbolize different groups of prokaryotes. a: A working hypothesis for the origin of eukaryotic genes that is fundamentally the same as figures in refs. 25 and 26, except that this figure departs from those by incorporating two important elements from ref. 35: 1) that in the earliest stages of evolution, gene pools, rather than distinct lineages predominated[35]; and 2) that all contemporary cells ultimately descend from autotrophic ancestors.[35] Furthermore, the figure embraces the unproven but explicit premises: 1) that the host that acquired the mitochondrion was an archaebacterium (not a eukaryote);[37] and 2) that no eukaryotes ever existed that did not possess the mitochondrial symbiont.[37] Another minor difference is that this figure extends symbiotic associations (merging of genomes into the same cellular confines = merging of colored lines) to include schematic indication of several independent secondary symbioses for the acquisition of plastids during eukaryotic history.[27,28] Importantly, among eukaryotes, colored lines indicate merely that prokaryotic genomes existed at one time within the cellular confines of a given eukaryotic lineage, not that they have persisted to the present as an independently compartmented genome. For example, some eukaryotes with secondary symbionts are schematically indicated with six lines, but only have four distinctly compartmented genomes.[27,28] Similarly, eukaryotes that lack mitochondria apparently possessed such organelles in the past[20,22,25,26,29,33,37] but only have one genome: that in the nucleus. b: In the enlargement of a portion of a, lateral gene transfer *between eubacteria* prior to—and implicitly, but not shown, subsequent to—the origin of mitochondria (blue lines) and plastids (green lines) is schematically represented. Genomes are represented as heavy lines; individual gene transfer events (regardless of possible numbers of genes involved) as thin lines.

mitochondria cannot directly account for this finding,[26,31] unless corollary assumptions of widespread horizontal transfer from eubacteria to eukaryotes are added.[34] A recently formulated alternative to the endosymbiont hypothesis for the origin of mitochondria directly accounts for (moreover demands the existence of) many eubacterial gene products in the cytosol,[37] but might need a corollary assumption of subsequent lateral transfer between eubacteria—similar to that observed by Lawrence and Ochman—to account for overall patterns of similarites between nuclear genomes and *large* (*Rhodobacter*-like,[37] rather than *Rickettsia*-like[38]) α-proteobacterial genomes when these become available for analysis.

Conclusion

The rate with which the *E. coli* genome has amassed foreign DNA and the amount of foreign DNA in that genome tend to suggest that bacterial chromosomes are dynamic structures, rather than static. Just how dynamic they are, how dynamic they have been in the past, and whether all of them are

dynamic or not, are important issues. If *all* genes were being passed around through *all* bacterial genomes in a haphazard manner through time, then the immediate prediction would follow that *no* genes would even remotely emulate rRNA phylogeny. But since RecA, as one example, *does* produce trees that strongly resemble rRNA topology in head-to-head comparisons,[39] it seems that some principles must govern the distribution of genes across bacterial genomes. It is likely that there will be cases where selection is at work for useful gene combinations—as has been discussed in the context of an intruiging α-proteobacterium-methanogen transfer[40]—and that there will be cases where the principles are not obvious at all. Clearly, careful gene-by-gene phylogenetic comparisons[32,33] in addition to genome-by-genome comparisons for the general presence and absence of genes are needed. It is a substantial challenge for comparative genomics to merely describe the distribution of genes across genomes. An even greater challenge will be to uncover its governing principles.

Acknowledgments

I thank W. Ford Doolittle, C. Kurland, M.A. Ragan, and T. Martin Embley for constructive criticisms on the manuscript, U.-G. Maier, M.A. Ragan, and T. Martin Embley for making data available prior to publication, many colleagues for discussions, and the Deutsche Forschungsgemeinschaft for financial support.

References

1. Lawrence JG, Ochman, H. Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci USA 1998;95:9413–9417.
2. Milkman R. Recombinational exchange between clonal populations. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonmella.* Washington, DC: American Society for Microbiology Press; 1996. pp 2663–2684.
3. Boyd EF, Nelson K, Wang F-S, Whittam TS, Selander RK. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and *Salmonella enterica.* Proc Natl Acad Sci USA 1994;91:1280–1284.
4. Reeves P. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet 1993;9:17–22.
5. Li J, Nelson K, McWhorter AC, Whittam TS, Selander RK. Recombinational basis of serovar diversity in *Salmonella enterica.* Proc Natl Acad Sci USA 1994;91:2552–2556.
6. Sueoka N. Directional mutation pressure, mutator mutations and the dynamics of molecular evolution. J Mol Evol 1993;37:137–153.
7. Berg OG, Silva PJ. Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection. Nucleic Acids Res 1997;25:1397–1404.
8. Sharp PM, Li W-H. The codon adaptation index—a measure of synonymous codon usage bias, and its potential application. Nucleic Acids Res 1987;14:1281–1295.
9. Sharp PM, Li W-H. The rate of synonymus substitution in enterobacterial genes is inversely related to codon usage bias. Mol Biol Evol 1987;4:222–230.
10. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987. 512 p.
11. Matic I, Rayssiguier C, Radman M. Interspecies gene exchange in bacteria: the role of sos and mismatch repair systems in evolution of species. Cell 1995;80:507–515.
12. Vulic M, Dionisio F, Taddei F, Radman M. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc Natl Acad Sci USA 1997;94:9763–9767.
13. Modrich P. Mechanisms and biological effects of mismatch repair. Ann Rev Genet 1991;25:229–253.
14. Sniegowski P. Mismatch repair: Origin of species? Curr Biol 1998;8:59–61.
15. Majewski M, Cohan FM. The effect of mismatch repair and heterodulpex formation on sexual isolation in *Bacillus.* Genetics 1998;148:13–18.
16. Groisman EA, Ochman H. Pathogenicity islands: Bacterial evolution in quantum leaps. Cell 1996;87:791–794.
17. Mel SF, Mekalanos JJ. Modulation of horizontal transfer in pathogenic bacteria by in vivo signals. Cell 1996;87:795–798.
18. Martin W, Herrmann RG. Gene transfer from organelles to the nucleus: How much, what happens and why? Plant Physiol 1998;118:9–17.
19. Wastl J, Fraunholz M, Zauner S, Douglas S, Maier U-G. Ancient gene duplication and differential gene flow in plastid lineages: the GroEL/Cpn60 example. J Mol Evol; in press.
20. Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Curr Biol 1997;7:995–998.
21. Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 1998;393:162–165.
22. Keeling PJ, Doolittle WF. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. Proc Natl Acad Sci USA 1997;94:1270–1275.
23. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes Proc Natl Acad Sci USA 1998;95:6239–6244.
24. Nowitzki U, Flechner A, Kellermann J, Hasegawa M, Schnarrenberger C, Martin W. Eubacterial origin of eukaryotic nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase: Sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. Gene 1998;214:205–213.
25. Doolittle WF. Fun with genealogy. Proc Natl Acad Sci USA 1997;94:12751–12753.
26. Doolittle WF. A paradigm gets shifty. Nature 1998;392:15–16.
27. Van de Peer Y, Rensing S, Maier U-G, De Wachter R. Substitution rate calibration of small subunit RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. Proc Natl Acad Sci USA 1996;93:7744–7748.
28. McFadden GI, Gilson PR, Douglas SE, Cavalier-Smith T, Hofmann CJ, Maier U-G. Bonsai genomics: sequencing the smallest eukaryotic genomes. Trends Genet 1997;13:46–49.
29. Embley TM, Horner DA, Hirt RP. Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? Trends Ecol Evol 1997;12:437–441.
30. Ribiero S, Golding GB. The mosaic nature of the eukaryotic nucleus. Mol Biol Evol 1998;15:779–788.
31. Ragan MA, Gaasterland T. Microbial genescapes: A prokaryotic view of the yeast genome. Microbial and Comparative Genomics; 1999.
32. Brown JR, Doolittle WF. *Archaea* and the prokaryote-to-eukaryote transition. Microbiol Mol Biol Rev 1997;61:456–502.
33. Embley TM, Hirt RP. Early branching eukaryotes? Curr Opin Genet Dev; 1998;8:624–629.
34. Doolittle WF. You are what you eat: a gene transfer ratchet could account for eubacterial genes in eukaryotic genomes. Trends Genet 1998;14:307–311.
35. Kandler O. The early diversification of life. In: Bengston S, editor. Early life on earth. New York: Columbia University Press; 1994. p 152–160.
36. Woese C. The universal ancestor. Proc Natl Acad Sci. USA 1998;95:6854–6859.
37. Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. Nature 1998;392:37–41.
38. Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard M, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G. Genome structure and gene content in protists mitochondrial DNAs. Nucleic Acids Res 1998;26:865–878.
39. Eisen JA. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J Mol Evol 1995;41:1105–1123.
40. Chistoserdova L, Vorholt JA, Thauer RK, Lidstrom ME. C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. Science 1998;281:99–102.