## Concluding remarks

I apologize to my virologist and mycologist colleagues for the emphasis on bacteria. There are, for reasons of space, many aspects of microbiology that I have left out. For example, I have omitted any discussion of microbial metabolism, which still has surprises such as the discovery of novel anabolic and catabolic pathways that might have important applications in bioremediation and biotransformation. As the majority of bacterial species remain beyond our grasp scientifically, there are undoubtedly many additional aspects of microbial metabolism and physiology to be revealed. In one sense, the present understanding of microbiology is like physics or astronomy in the mid-20th century; microbiologists are still trying to find out what they have to work with! If tens of thousands of genome sequences are to be made available to expand the basic framework of microbiological science, the new microbiology must embrace aspects of 'big science'. However, funding of research in microbiology is at abysmally low levels, and it is imperative that this situation be redressed, as befits the importance of the subject. For now, the ubiquitous and indispensable microbes remain misunderstood and as yet widely unexploited; they never make anything useless, it is our job to find uses for their metabolic richness. And, for all of you who enjoy good cheese, wine, yoghurt, bread, beer, quorn, soy sauce, coffee, chocolate, etc., try to imagine what life would be like without microbes!

References

1 Amann, R.I. *et al.* (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169
2 Shimkets, L.J. (1998) in *Bacterial Genomics* (Bruijn, F.J. *et al.*, eds), pp. 5–11, Chapman & Hall
3 Pallen, M.J. (1999) Microbial genomes. *Mol. Microbiol.* 32, 907–912
4 Hacker, J. *et al.* (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* 23, 1089–1097
5 Nelson, K.E. *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329
6 Woese, C.R. (1994) There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol. Rev.* 58, 1–9
7 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2128
8 Williams, P. and Stewart, G.S.A.B. (1994) in *Bacterial Biofilms and Their Control in Medicine and Industry* (Nichols, W.W. *et al.*, eds), pp. 9–13, Bioline
9 Fuqua, C. and Greenberg, E.P. (1998) Self perception in bacteria: quorum sensing with acylated homoserine lactones. *Curr. Opin. Microbiol.* 1, 183–189
10 Kell, D.B. *et al.* (1995) Pheromones, social behaviour and the functions of secondary metabolism in bacteria. *Trends Ecol. Evol.* 10, 126–129
11 Dunny, G.M. and Winans, S.C. (1999) *Cell–Cell Signaling in Bacteria*, ASM Press
12 Golden, S.S. *et al.* (1998) The cyanobacterial circadian system: a clock apart. *Curr. Opin. Microbiol.* 1, 669–673
13 Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740
14 Whitman, W.B. *et al.* (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578–6583
15 Krause, R.M. (1998) *Emerging Infections*, Academic Press
16 Relman, D.A. (1999) The search for unrecognized pathogens. *Science* 284, 1308–1310
17 Cossart, P. *et al.* (1996) Cellular microbiology emerging. *Science* 271, 315–316
18 Costerton, J.W. *et al.* (1999) Bacterial biofilms: a common cause of persistent infections. *Science* 284, 1318–1322
19 Henderson, D.A. (1999) The looming threat of bioterrorism. *Science* 283, 1279–1282

Acknowledgements

I thank the National Science and Engineering Council of Canada and the Canadian Bacterial Diseases Network for support. Rosario Bauzon and Dorothy Davies worked on the preparation of the manuscript, and Rachel Fernandez provided helpful comments.

# Lateral genomics

*W. Ford Doolittle*

More than 20 complete prokaryotic genome sequences are now publicly available, each by itself an unparalleled resource for understanding organismal biology. Collectively, these data are even more powerful: they could force a dramatic reworking of the framework in which we understand biological evolution. It is possible that a single universal phylogenetic tree is not the best way to depict relationships between all living and extinct species. Instead a web- or net-like pattern, reflecting the importance of horizontal or lateral gene transfer between lineages of organisms, might provide a more appropriate visual metaphor. Here, I ask whether this way of thinking is really justified, and explore its implications.

I depict in Fig. 1a a highly-simplified version of the 'universal tree of life', as generally accepted[1–3]. The identity and composition of the three major groups of organisms (domains) is based on comparisons of the sequences of small subunit ribosomal RNAs (SSU rRNAs). The position of the root (see Glossary) of the tree has been determined independently, using the sequences of certain key proteins[4,5]. Many other molecular data sets agree with many of the individual groupings established by rRNA, and several support rRNA-derived branching orders within and between domains. This rRNA tree is surely the most important single guide we will ever have to understanding genealogical relationships between organisms.

However, several molecular-sequence[6–12] data sets disagree, root and/or branch, with the rRNA tree. Frequently, such disagreement reveals the inadequacy of phylogenetic methods: the algorithms used to build trees are based on oversimple assumptions about within-molecule or between-lineage variation in rates of sequence change and are insensitive to mutational saturation. Sometimes, however, disagreements clearly mean that different genes have different evolutionary histories – that lateral gene transfer has occurred between lineages (Fig. 1b).

There are three kinds of evidence for this. First, there are now dozens of well-supported anecdotal cases, based on data for individual protein-coding genes, in which groups established as coherent by rRNA are broken up to produce trees differing substantially from the rRNA tree or each other (Fig. 1b). Notable examples would be genes for Hsp70, $H^+$-ATPase subunits, glutamine synthetase, glutamate dehydrogenase, carbamoylphosphate

W. Ford Doolittle
ford@is.dal.ca

Canadian Institute for Advanced Research, Dept of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7.



TCB, Vol. 9, N° 12 (0962-8924) PII: S0962-8924(99)01664-5
TIBS, Vol. 24, N° 12 (0968-0004) PII: S0968-0004(99)01471-1
TIG, Vol. 15, N° 12 (0168-9525) PII: S0168-9525(99)01877-6

## Glossary

**Last common ancestor**
A single organism (genome) from which all contemporary organisms (genomes) in a particular group are descended.

**Lateral gene transfer**
Transfer of a gene from one genome to another at some point in the evolutionary process: an outcome, not a specific genetic mechanism.

**Paralogy, paralogue**
When gene X is duplicated so a genome has two copies, X and X′, these two genes are paralogues of each other. If both paralogues are retained during subsequent evolution and speciation, their evolutionary trees should be the same.

**Phylogeny**
The process that generates phyla, or the science that studies it. The word is now commonly used to describe a pattern (a phylogenetic tree) of genealogical relationships between species, understood as a succession of bifurcations. Species phylogenies are often based on sequence comparisons of a single set of homologous genes – on the assumption that such gene phylogenies track the evolutionary history of the organisms.

**Root**
The earliest time point, or last common ancestor, in a particular gene tree. Deciding which point on a gene tree corresponds to the root requires additional information, usually a more distant outgroup gene for comparison – such as myoglobin would be for a tree of the haemoglobins.

synthetases, HMGCoA reductases and many aminoacyl tRNA synthetases[6,12,13]. Second, some within-genome analyses show, by virtue of G + C content, codon usage and genetic organization, the relatively recent importation of substantial numbers of foreign genes. For example, Lawrence and Ochman[14] have concluded that 18% of the genes in the *Escherichia coli* K12 genome were introduced by lateral gene transfer in the past 100 million years. Finally, several between-genome comparisons show that all genomes contain some genes that are more similar to homologues in distant genomes than to homologues in closer relatives, or indeed that are not found at all in genomes of closer relatives. Nelson and co-workers concluded by such reasoning that 24% of the genome of the bacterial hyperthermophile *Thermotoga maritima* has been obtained from archaeal hyperthermophiles[15].

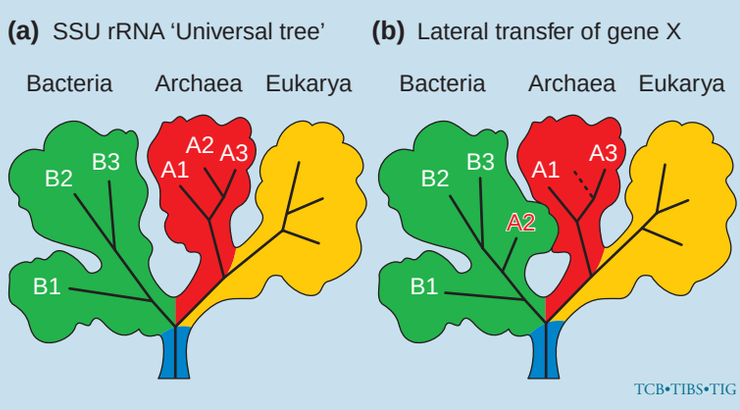


**FIGURE 1.** What lateral gene transfer looks like. (a) A simplified schematic version of a small subunit ribosomal RNA (SSU rRNA)-based universal tree, with three bacterial (B1–B3) and three archaeal (A1–A3) taxa. In fact, thousands of species can now be placed on such a tree[2], showing detailed structure within domains. (b) Pattern observed for gene X if there has been a transfer of gene X from an ancestor of B2 and B3 to an ancestor of A2, and loss by A2 of the resident homologue of X. Trees that disagree with the SSU rRNA tree in such a fashion have been described for many genes[6–12]. Many more instances of lateral gene transfer must have occurred within domains. Incongruent trees like this can also, although much less parsimoniously, be explained by invoking paralogy and differential loss. To make sense of this simple case by paralogy, we would have to assume that (i) the last common ancestor in (b) had two paralogous copies of gene X (produced by an earlier duplication), (ii) bacterial species B1 retained only paralogue 1, while B2 and B3 retained only paralogue 2, (iii) archaeal species A1 and A3, and the eukaryotes, retained paralogue 1 and (independently) lost paralogue 2, while (iv) archaeal species A2 retained paralogue 2, but lost paralogue 1.

## What we might lose

If lateral gene transfer can affect all genes, and has affected some substantial fraction of genes over the past 3.8 billion years (since the origin of life), then much of what molecular phylogeneticists have hoped to accomplish is at risk, especially in the area of prokaryote evolution. These researchers can establish genealogical relationships only through analyses of genes that organisms share by virtue of descent from common ancestors. Yet even strains of a single prokaryotic 'species' can differ by up to 20% of their chromosomal DNA[16]. Seldom, if ever, will two species of the same genus have received all their genes from a single common ancestor, and the further up the taxonomic hierarchy we go (however we choose to define higher taxonomic ranks such as classes, phyla or kingdoms), the worse the situation becomes. Several authors have now claimed that many if not most of the genes for metabolic functions in archaeal genomes are recent imports from bacteria [8,11,13,17]. Eukaryotes also might harbour as many (or more) genes of bacterial origin as they do archaeal-origin genes (in spite of the implications of Fig. 1a)[7,8,10]. This means that we can have no gene-sequence-based universal phylogenetic tree, and no coherent universal systematic scheme, unless we choose to ignore many of the data.

This is not an altogether new problem: we have known for three decades that prokaryotic genomes differ in size (by fourfold among the cyanobacteria, for instance[18]). We have known for even longer (since the discovery of antibiotic-resistance transfer factors) that important genetic determinants can be transferred between unrelated bacteria[19]. But this knowledge had not deterred us from constructing universal prokaryotic phylogenies, probably because of seldom-examined beliefs such as that:

- genome size differences would prove mostly to be due to lineage-specific gene duplications or losses, not transfer;
- transferred genes are few and mostly special in some way (determining environmental interactions such as antibiotic resistance, pathogenicity and inter-prokaryote warfare); and
- there would prove to be a core genome, comprising most genes that are fundamental to the biology of most cells, genes that cannot be or have not been transferred.

Comparative genomics refutes the first two beliefs and calls into question the third.

Similarly, efforts to deduce the genetic make up of the last common ancestor of all extant life now appear misguided. There is no guarantee that a gene currently represented in some Bacteria, Archaea and Eukarya was present in their common ancestor – it could have arisen more recently in one domain and spread to the others. Such a scenario seems more appealing for dispensable genes (antibiotic resistance, biosynthetic capabilities not useful in many environments) than for genes encoding indispensable cellular functions, but replacements of such essential genes have occurred nevertheless. For example, genes for aminoacyl-tRNA synthetases have been transferred extensively within and between domains[12]. All contemporary homologues of a given gene, by definition, trace their ancestry to a single progenitor gene of that type (ignoring intragenic recombination). Nevertheless, there is no reason to believe that all (or even most) of the progenitors of all contemporary genes of different types, even those found in all three domains, ever resided in a single ancestral genome. As

Carl Woese writes, 'the universal ancestor is not a discrete entity. It is, rather, a diverse community of cells that survives and evolves as a biological unit.'[20]

Extensive lateral gene transfer also inevitably means that 'phylogenetic taxonomies' based on individual molecules have less predictive power than we had hoped. An unknown or uncultured organism is more likely to have physiological properties like those of sister taxa than of distant cousins on the rRNA tree[2], but it does not have to. Genes for radically different capacities might have been integrated recently into its genome. Thus, determination of rRNA 'phylotypes'[2] can only be the first step in characterization of novel environments.

## Reality checks

These are serious blows to molecular phylogenetics, and lateral gene transfer might owe as much of its current popularity to these iconoclastic consequences as to the weight of evidence. Many of the data are indeed 'anecdotal' – isolated cases of incongruent trees that alone might have other explanations and seem but a drop in the genomic ocean[6,8]. Many of the exhaustive comparisons of two or more genomes base their shocking conclusions on sequence similarity scores alone, not trees[15,17,21]. Even trees, unless properly rooted[7,8], can be misleading. The observation that, in general, archaeal genes encoding metabolic functions are most similar to their bacterial homologues, while the majority of archaeal genes for replication, transcription and translation look eukaryotic[11], could be explained by assuming that genes for metabolic functions evolve (change sequence) especially fast in eukaryotes, while replication, transcription and translation genes evolve especially fast in bacteria.

Differential loss of paralogues can in principle also be used to explain patterns such as that in Fig. 1b. However, invoking paralogy in place of lateral gene transfer can seriously violate rules of parsimony. Even the very simple case presented in Fig. 1b requires six independent losses (instead of one transfer and one loss), as spelled out in detail in the legend for Fig. 1b. Furthermore, each time we accept such a deep paralogy, we must add one more paralogue to the gene complement of the last common ancestral genome, which must then have been very much larger than the genome of any known contemporary prokaryote. This is so not only for paralogues but for any unique (orthologous) genes that are found in representatives of more than one domain – surely already an enormous number. This logic, and the apparent rigour of many of the anecdotal and some of the genome-versus-genome comparisons, persuade me that there is a real fire under the smoke of current editorials and articles with catchy titles. It remains possible that there is a core of untransferable or never-yet-transferred genes, but the burden of proof has surely shifted.

## What's to be gained?

How would science benefit by whole-hearted adoption of the notion that prokaryotic evolution should be viewed as net-like, not tree-like? What might we gain by focusing on how genes themselves have evolved and the role that transfer of genes has played in generating the pattern of diversity we see in the microbial world? Note that there still will be groups of related organisms. As long as species split into new species frequently with respect to the rate of loss or gain (by transfer) of new genes by their genomes, we can talk sensibly about their relationships, and we can make useful taxonomies. Sometimes, collections of many genes might remain together for very long periods – because they are co-adapted, because they together contribute to a successful phenotype or because there is little possibility for lateral gene transfer. We might then employ them to define useful taxa of very high rank, such as Bacteria and Archaea. Nevertheless, the net or web metaphor should remind us that all prokaryotic taxa are in essence imprecisely bounded and ephemeral. We might thus realistically look at all prokaryotes as one 'global superorganism' (as Sonea[22] and Reanney[23] already suggested several decades ago) divided into subpopulations – within and between which genes are exchanged at different frequencies.

This super-species model is at one end of a spectrum defined by the frequency and generality (in terms of genes that can be transferred) of lateral gene transfer. The traditional tree model is still at the other, and there is no single experimental observation that could ever decide between them. Thus, the worth and survival of the new view of evolution will be judged by its ability to produce interesting new problems, theories and approaches. As Bill Martin[10] wrote recently: '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.' Fortunately, there already are a few good ideas or conceptual approaches in circulation, and more might easily come to mind. Let's consider five.

### The complexity hypothesis

Jain *et al.*[11] articulate the common belief that genes for RNAs or proteins that interact to form complexes with many other cellular macromolecules will be less subject to transfer because their products will less easily function in a foreign cytoplasm. There certainly should be a relationship between exchangeability and interactiveness, but perhaps not such a simple one. Proteins or RNAs that interact with just one or a few cellular partners might (through coevolution with those partners) acquire idiosyncratic structures incompatible with integration into the homologous complex in cells of a distant species. Proteins or RNAs that must interact with many other macromolecules might on the other hand be able to change very little in structure and thus better retain the ability to function in a foreign setting. In any case, genetic linkage of genes for interacting molecules could easily trump such coevolutionary barriers to exchange. Dandekar *et al.*[24] have in fact observed that genes whose products interact physically (even when they do not catalyse steps in a single pathway) are likely to exhibit genetic linkage conserved across bacteria and archaea, as if they were frequent travelling companions.

### Selfish operon theory

Lawrence and Roth's theory[25] elegantly explains why genes whose products do catalyse steps in a single pathway are clustered in operons (even when these products do not interact physically). They suggest that, 'from a gene's perspective, horizontal transfer provides a way to escape evolutionary loss [in environments where their function is not required] by allowing colonization of organisms lacking the encoded functions. Since organisms bearing clustered genes are more likely to act as successful donors, clustered genes would spread among bacterial genomes.'

### Taking the genes-eye view

The selfish DNA theory[26,27], now 20 years old, argued that plasmids and transposable elements should be viewed as genetic parasites, whose sometimes beneficial effects on the long-term evolvability of prokaryotic hosts are coincidental. This view has gained general acceptance as part of a broader hierarchical theory. Potential targets for selection can be defined at all levels of biological organization (genes, organisms, populations, species), but the relative effectiveness of this force in forging adaptations at these different levels is still vigorously debated[28,29]. The phenomenology of lateral gene transfer provides much new grist for this mill. Consider VPI, the 40-kbp pathogenicity island of

*Vibrio cholerae*, which bears many genes affecting interactions of the bacterium with its host and the host's susceptibility to the toxin-encoding cholera phage CTXΦ. Karaolis *et al.* have shown recently that VPI is itself a prophage, so this system exhibits an amazing mix of parasitic and symbiotic interactions in need of theoretical elaboration[30]. Or consider plasmid- and transposon-borne integrons and the gene cassettes (encoding antibiotic resistance) they are able to recruit from as-yet-mysterious sources[31] – a highly sophisticated system that might only make sense if we model prokaryotic genes as semi-autonomous agents within a global superorganism. The new view of prokaryotic evolution will stimulate both theoretical and experimental work on systems of gene exchange. We need to look for, and try to understand, those features of structure and function of genes that bear on their survival and spread as independent agents within the global superorganism. So far, our focus has been mostly on those features of genes that affect the fitnesses of the organisms in which we happen to find them.

### Driving forces for acquiring new genes

Similarly, we must ask more rigorously what recipients gain from acquiring new genes. New genes could be acquired and fixed by selection because:

- they confer on the recipient cell a novel biosynthetic or degradative capacity for which previously it had no gene;
- they confer resistance to an antibiotic or other toxic agent that inactivates the resident copy of the gene; or
- they encode a protein whose kinetic properties or physical characteristics (heat or oxygen sensitivity, for instance) are better adapted than those of the resident gene product to a new organismal niche.

Newly transferred genes might also be fixed neutrally: any cell that by chance has integrated a foreign gene that adequately performs the same function as a resident gene might by chance lose the latter. Both events might be rare, but, for cells constantly exposed to foreign DNA, such an outcome would be inevitable[32].

### Evolutionary novelty

New genes from far away should impart new tempo and new modes in prokaryotic evolution. Laterally transferred genes, because they can confer radically new and complex phenotypes, might often result in adaptive radiations and the formation of new subpopulations (bacterial clades or even 'phyla') – perhaps in fact more often than can mutation and selection operating on already resident genes. Lawrence and Ochman could be right when they suggest that, unlike eukaryotic speciation, bacterial speciation might be 'driven by a high rate of horizontal transfer, which introduces novel genes, confers beneficial phenotypic capabilities, and permits the rapid exploitation of competitive environments'[14].

## A new synthesis

Phylogeneticists, genomicists, molecular biologists and population geneticists have different perspectives on prokaryote evolution, and their respective literatures on lateral gene transfer seem disconnected. Come the millennium, we could hope for more disciplinary interpenetration and a more sophisticated understanding of how life's history is sometimes like a tree and sometimes like a net. An evolutionary model in which novel genes transferred between populations play a major role in adaptation is radically different from one in which adaptation is achieved by selective allele replacement within populations. Its implications for phylogeny, whether that word is interpreted to mean genealogy or the process by which major groups are formed, are also radically different. Nevertheless, both modes of adaptation drive the evolution of prokaryotes, at the same time.

Acknowledgements

I thank the Canadian Institute for Advanced Research for a Fellowship, the Medical Research Council of Canada for research support, and Dave Faguy, John Logsdon, and Andrew Roger for stimulating discussions about lateral gene transfer.

References

1 Woese, C.R. *et al.* (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* 87, 4576–4579
2 Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740
3 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2128
4 Gogarten, P.J. *et al.* (1989) Evolution of the vacuolar $H^+$-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6661–6665
5 Iwabe, N. *et al.* (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 9355–9359
6 Gogarten, P.J. *et al.* (1996) Gene duplications and horizontal gene transfers during early evolution. In *Evolution of Microbial Life* (Roberts, D. *et al.*, eds), pp. 267–292, Cambridge University Press
7 Feng, D.F. *et al.* (1997) Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13028–13033
8 Brown, J.R. and Doolittle, W.F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* 61, 456–502
9 Gupta, R.S. (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1435–1491
10 Martin, W. (1999) Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *BioEssays* 21, 99–104
11 Jain, R. *et al.* (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3801–3806
12 Doolittle, R.F. and Handy, J. (1998) Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Curr. Opin. Genet. Dev.* 8, 630–636
13 Doolittle, W.F. and Logsdon, J.M., Jr (1998) Archaeal genomics: do archaea have a mixed heritage? *Curr. Biol.* 8, R209–R211
14 Lawrence, J.G. and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9413–9417
15 Nelson, K.E. *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329
16 Lan, R. and Reeves, P.R. (1996) Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* 13, 47–55
17 Koonin, E.V. *et al.* (1997) Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* 25, 619–637
18 Herdman, M. *et al.* (1979) Genome size of cyanobacteria. *J. Gen. Microbiol.* 111, 73–85
19 Falkow, S. (1975) *Infectious Multiple Drug Resistance*, Pion
20 Woese, C.R. (1998) The universal ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6854–6859
21 Aravind, L. *et al.* (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444
22 Sonea, S. and Paniset, M. (1976) Towards a new bacteriology [in French]. *Rev. Can. Biol.* 35, 103–167
23 Reanney, D.C. (1978) Coupled evolution: adaptive interactions among the genomes of plasmids, viruses, and cells. *Int. Rev. Cytol.* (Suppl.) 8, 1–68
24 Dandekar, T. *et al.* (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23, 324–328
25 Lawrence, J.G. and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143, 1843–1860
26 Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603
27 Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA. *Nature* 288, 645–646
28 Williams, G.C. (1992) *Natural Selection: Domains, Levels and Challenges*, Oxford University Press
29 Gould, S.J. (1998) Gulliver's further travels: the necessity and difficulty of a hierarchical theory of selection. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* 353, 307–314
30 Karaolis, D.K. *et al.* (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399, 375–379
31 Hall, R.M. and Collis, C.M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* 15, 593–600
32 Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* 14, 307–311