How big is the iceberg of which organellar genes in nuclear genomes are but the tip?

W. F. Doolittle^{1,2*}, Y. Boucher², C. L. Nesbø^{1,2}, C. J. Douady², J. O. Andersson²
and A. J. Roger^{1,2}

¹Genome Atlantic, and ²Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, NS B3H 1X5, Canada

As more and more complete bacterial and archaeal genome sequences become available, the role of lateral gene transfer (LGT) in shaping them becomes more and more clear. Over the long term, it may be the dominant force, affecting most genes in most prokaryotes. We review the history of LGT, suggesting reasons why its prevalence and impact were so long dismissed. We discuss various methods purporting to measure the extent of LGT, and evidence for and against the notion that there is a core of never-exchanged genes shared by all genomes, from which we can deduce the 'true' organismal tree. We also consider evidence for, and implications of, LGT between prokaryotes and phagocytic eukaryotes.

Keywords: lateral gene transfer; prokaryotes; eukaryotes; genome evolution

1. PATTERN AND PROCESS

The neo-Darwinian synthesis explains both evolutionary pattern and evolutionary process. Surely, it is what Darwin would have come up with himself, if he had lived another hundred years after publishing *The Origin of Species*. Contemporary species resemble each other phenotypically, to greater or lesser extents, because of their descent (with modification) from common ancestors fewer or more generations removed—that is the *pattern*. The *process* is fixation—by natural selection operating within species—of favourable alleles. This microevolutionary mechanism is the primary cause of modification during descent. Macroevolution (speciation and the appearance of evolutionary trends) is its by-product.

Darwin, of course, did not know about genes, and the formulators of the neo-Darwinian synthesis did not anticipate the phenomenal growth of molecular phylogenetics and comparative genomics. But because of the success of these new disciplines, we have now come to rely on *gene phylogenies* to recreate evolution's pattern, and to see *genomes* as surrogates for organisms, as well as chronicles of the process of evolution: we have reduced evolution to (phylo)genetics and genomics. It is, in particular, this most highly reductive, genome-based neo-neo-Darwinism—especially as it is applied to prokaryotes—that LGT deconstructs, and *could* destroy.

Woese (2000) muses that, 'vertically generated and horizontally acquired variation could be viewed as the yin and the yang of the evolutionary process'. Neo-Darwinism acknowledged both at the population and species level, but only the yin of vertical descent beyond that. There is

no unique tree relating all humans in the current population, because no two human genomes (except twins') derive from a single ancestral genome. But there will be a unique tree relating genomes of humans, mice, birds and fishes, because any exchange between these lineages ceased shortly after separation of their respective ancestral gene pools. Surely, this *is* the correct model for such 'higher' eukaryotic taxa.

However, for prokaryotes (which are Woese's main concern) the yang of 'horizontal acquisition' is also a major force, perhaps even the dominant one. Furthermore, its impact on the evolution of simpler unicellular eukaryotes, without sequestered germ lines and often intimately exposed to foreign DNA, is still unknown. In this review, we address this yang, and its meaning for evolutionary pattern and process, in prokaryotes and such eukaryotes.

2. THE EARLY (BUT IGNORED) THREAT FROM LATERAL GENE TRANSFER

Amazingly, many of the issues that still interest and perplex molecular evolutionists were first articulated by Zuckerkandl and Pauling in the mid-1960s. Their chapter in Bryson & Vogel's (1965) *Evolving Genes and Proteins*, for instance, explains how molecular sequences can be used to reconstruct organismal phylogeny, argues forcefully that they are better for this than traditional phenotypic characters, introduces the concept of the neutrality of molecular sequence evolution and draws the distinction between orthology and paralogy (though without those words). Zuckerkandl and Pauling also set forth the basic challenge that molecular phylogeny needed to meet, and spelled out the enormous intellectual payoff that success at doing this would provide. They wrote: 'it will be determined to what extent the phylogenetic tree, as derived from molecular data in complete independence

* Author for correspondence (ford@is.dal.ca).

One contribution of 21 to a Discussion Meeting Issue 'Chloroplasts and mitochondria: functional genomics and evolution'.

from the results of organismal biology, coincides with the phylogenetic tree constructed on the basis of organismal biology. If the two phylogenetic trees are mostly in agreement with respect to the topology of branching, the best available single proof of the reality of macro-evolution would be furnished' (Zuckerlandl & Pauling 1965).

What Zuckerlandl and Pauling did not consider was the possibility that different molecules (genes) might yield different trees—not only different than that constructed from 'organismal biology' but from each other—and *still* all be telling the truth. However, Japanese microbiologists had already uncovered genetic processes that inevitably must produce just such conflicts. In the late 1950s, several investigators observed that resistance to multiple antibiotics could be transferred simultaneously from *Shigella* to *Escherichia coli*, and that such between-species transfer was probably responsible for the increase in drug resistance among dysentery-causing shigellae in Japan, after World War II. In an influential 1963 English language review, Watanabe elaborated on this form of 'infectious heredity' and the evidence that the underlying genetic determinants (RTFs) were in fact episomes, or plasmids, similar to the F factors first identified through their (rather more harmless) roles in the 'sexuality' of *E. coli* (Watanabe 1963).

As the facts of plasmid biology (and the transduction abilities of phages) became more widely known, a few bold theorists addressed the implications of between-species transfer for our broader understanding of evolution and systematics. In a prescient 1966 *Nature* article, E. S. Anderson reviewed F and RTF biology and evidence for broad host ranges among RTFs, and noted that often 'the same transfer factor will transfer different characters and the same character may be carried by different transfer factors'. In spelling out implications for our understanding of bacterial evolution, he focused on how LGT could drive adaptation, so that 'the evolutionary time scale might have been telescoped into a shorter time span than that envisaged purely in terms of the selection of mutants with survival advantages'. Furthermore, he suggested that, if plasmid-mediated LGT proved to occur outside the Enterobacteriaceae (which no one then knew it to do), 'some rethinking may be called for in relation to evolution in other fields of biology' (Anderson 1966).

Since that time, the Quebecois microbiologist Sorin Sonea and his colleagues have run the farthest with this particular ball. In a series of articles and books beginning in the early 1970s (Sonea 1971; Sonea & Panniset 1976; Sonea & Mathieu 2000, 2001) they have argued that, because of LGT mediated by plasmids and phages, we should think of *all* prokaryotes as but a single species. Recently, for instance, Sonea & Mathieu (2000) suggest that, somewhere in Life's first billion years, 'lateral gene transfer mechanisms appeared and were progressively improved, furthering the development of diversity. The prokaryotes' constructive evolution resulted in the formation of a worldwide web of genetic information, and a global bacterial superbiosystem (superorganism).' Similar, if more conservatively expressed, ideas were mooted by Reanney, in the mid-1970s (Reanney 1974), and of course are consistent with Margulis' influential writings on the role of genetic chimerism (cellular symbiosis) in evolution

in general over the last three-and-a-half decades (Margulis 1996).

Somehow, though, molecular phylogeneticists were not deterred in their attempts to impress Zuckerlandl and Pauling's paradigm on the microbial world. It is interesting to speculate why. Of importance, we think, is the fact that the discipline's pioneers were molecular biologists and biochemists, *not* geneticists. Moreover, the phenomena uncovered by the Japanese infectious disease microbiologists seemed to involve only highly specialized genes determining host-parasite interactions, not the conserved core of genes required for basic metabolic and information transfer processes. Indeed, the arms race between humans and their pathogens might be especially complex and intense, and thus have quite recently engendered or encouraged new and faster modes of gene exchange. (Evidence for this prejudice is the fact that only lately have we started to think of pathogenicity islands as members of a more general class of 'genomic islands', which in non-pathogens can bear determinants for symbiosis, saprophytic activities or adaptation to pollutants, for instance (Hentschel & Hacker 2001).)

3. THE EXCEPTION THAT PROVED THE RULE: THE SERIAL ENDOSYMBIOSIS HYPOTHESIS

By contrast, molecular evolutionists enthusiastically embraced the even more radical type of LGT involved in the endosymbiotic origin of the organellar genes and genomes of eukaryotes. Indeed, it could be argued that it was in finding compelling molecular evidence for *these* transfers that the infant discipline of molecular phylogenetics cut its first teeth.

Two years after Zuckerlandl and Pauling explicated the virtues of molecular sequence approaches, Lynn Sagan had revived and revitalized the notion that was later to be known as the SEH (see also Taylor 1979). According to this theory, mitochondria and plastids are (respectively) the degenerate descendants of once free-living respiring bacteria and cyanobacteria that had entered into endosymbiotic relationships with heterotrophic, anaerobic proto-eukaryotic host cells (Sagan 1967). In 1970, Sagan (now Margulis) fleshed out this scenario more fully in her *Origin of Eukaryotic Cells* (Margulis 1970). This book introduced many molecular biologists to the problems of early cell evolution and was crucial in encouraging the initial application of molecular methods to 'deep phylogeny' (by which we mean attempts to establish branching order for Life's three domains, and primary divisions within them). At the same time, Stanier elaborated on the probable properties of the proto-eukaryote (Stanier 1970). He reasoned that the capacity to harbour cellular endosymbionts was an essential character of such proto-eukaryotic hosts, and that this derived from the *capacity to perform endocytosis* (phagocytosis). Thus, before mitochondria and plastids came on the scene, cells with complex internal membrane systems and cytoskeletal elements—in essence eukaryotes without organelles—had already appeared. Cavalier-Smith later coined the term *archezoa* for such cells (Cavalier-Smith 1987), and this is still a good name for primitively amitochondriate eukaryotes, even though it now seems possible that none currently exist (Roger 1999).

We could sequence proteins before we could sequence RNA or DNA, and the early days of 'deep' molecular phylogeny saw a heavy focus on bacterial ferredoxins and cytochromes. Substantial sequence data accumulated from the experimental efforts of Kamen and Ambler, for instance, and were organized by Margaret Dayhoff into what we would now call a database. Schwartz and Dayhoff summarized much of this work in a benchmark article in *Science*, in 1978 (Schwartz & Dayhoff 1978). On the evidence of cytochromes and ferredoxins, chloroplasts did show a particularly compelling relationship to cyanobacteria ('blue-green algae'), as everyone expected—whereas mitochondria looked to be derived proteobacteria, as some people (for instance, John & Whatley 1975) would have predicted. Of course these proteins were known to be encoded (usually) by nuclear, not organellar genes, so such tests of the SEH were actually a test of the more complex notion that organelles derive from prokaryotes, and that nuclear genes whose product function in organelles derive from organellar DNA. Also, because, according to the SEH, they had no homologues encoded in the original proto-nuclear genome itself, ferredoxins and cytochromes could not tell us where the bulk of the eukaryotic cell (its nuclear-cytoplasmic component) came from. To graft the eukaryotic nuclear lineage to the prokaryotic tree, Schwartz and Dayhoff used 5S rRNAs, for which many sequences were already then available.

Unfortunately, 5S rRNA holds little deep phylogenetic information (and Schwartz and Dayhoff eventually linked the eukaryotic nucleus to Gram-positive bacteria!). Compelling proof that the nucleus was neither of cyanobacterial or proteobacterial in origin and thus that both organelles did have endosymbiotic (non-nuclear) origins, required sequence information from SSU rRNAs. These are encoded by homologous genes on all prokaryotic, mitochondrial, chloroplast and nuclear genomes—one of the reasons that Carl Woese, by the mid-1970s (Fox *et al.* 1977) had settled on SSU rRNA as the ideal 'universal molecular chronometer'. Within the decade, SSU rRNA data (at first partial, in the form of 'oligonucleotide catalogues', and then as complete rRNA or rDNA sequences) provided incontrovertible proof of the SEH, as applied to mitochondria and plastids (Gray & Doolittle 1982; Yang *et al.* 1985).

4. THE HEGEMONY OF SSU rRNA

Proof of the SEH was one feather in SSU rRNA's cap. Another was the ease, once sequencing became routine, of clustering bacterial species into higher-order taxa that, in general, correspond to divisions in Bergey's manual. Here, there has been just the right combination of major discovery (the assignment of many previously problematic taxa to the Archaea, for instance) and confirmation of expectation (such as the monophyly of the cyanobacteria) to engender both interest and confidence. Furthermore, SSU sequence characterization makes it possible to survey the phylogenetic diversity of natural environments without isolating, cultivating or even *seeing* a single organism. Such 'phylotype' analysis, as pioneered by Norman Pace and his students (Pace 1997), has not only revolutionized microbial ecology, it has uncovered still further major pro-

karyotic 'phyla', of whose existence the authors of Bergey's manuals had no clue.

Ironically, SSU rRNA has underperformed in one of the most important tasks set for it—determining branching order among bacterial phyla. It was their own failure in this endeavour that caused Stanier and van Niel to abandon phylogenetic pretensions in favour of more limited, practical, systematic practices, and for which Woese has, in several instances, castigated them (Olsen *et al.* 1994). However, most SSU rRNA-based universal trees *also* show bacterial phyla in an unresolved polychotomy (a 'star phylogeny'). Investigators who retain strong faith in this marker suggest this is a real result—that most modern bacterial groups appeared almost simultaneously, in some Archaean adaptive radiation (Pace 1997). Others see in this star phylogeny only poor resolution, and evidence of the inability of this (or any other single) gene to probe Life's earliest evolution (Forterre & Philippe 1999; Roger *et al.* 1999; Philippe *et al.* 2000).

For a time, SSU rRNA-based and cytochrome *c*-based phylogenetic work proceeded in parallel. During the early 1980s, a minor rebellion surfaced from the latter quarter. Ambler, Meyer, Kamen and collaborators (Ambler *et al.* 1979a,b) noted that trees for purple non-sulphur photosynthetic bacteria (we would now call them proteobacteria) made with different proteins (cytochrome *c*-551 and *c'*, high potential iron-sulphur proteins) had different topologies. Furthermore, these molecular trees disagreed with traditional classifications based on cell physiology, biochemistry and morphology. Although acknowledging several possible causes, they concluded that 'the simplest (if most difficult to prove) explanation for all the anomalies described is that the proteins concerned have reached some of the organisms by processes of interspecific gene transfer' (Ambler *et al.* 1979a).

In a strong rejoinder, Woese *et al.* (1980) dismissed this interpretation (although Woese and Fox had always endorsed LGT as a major force in *early* cell evolution (Woese 1987)). They showed congruence between SSU rRNA and one of the cytochromes, and argued that disagreements with other proteins only reflected the inadequate phylogenetic information content of the latter. Similarly, in their view, disagreement with traditional classification simply confirmed Stanier and van Niel's earlier fears for the taxonomic uselessness of cell morphology and physiology, and thus the superiority of SSU rRNA (for a similar view see Dickerson 1980). Dismissing transfer, however, obliged Woese and colleagues to explain the patchy distribution of photosynthetic functions among SSU-based groups as evidence of multiple losses of photosynthesis in lineages derived from photosynthetically competent ancestors. Surely, such loss *has* happened, although we also now have compelling evidence that genes for proteins that are intimately involved in the formation of the bacterial photosynthetic apparatus have been transferred widely and often among bacteria (Igarashi *et al.* 2001; Schutz *et al.* 2000; Xiong & Bauer 2002).

The doubts of the cytochrome sequencers were not assuaged by the arguments of Woese or Dickerson. In fact, Meyer *et al.* (1986) soon added mutational saturation to LGT as a confounding factor in tree construction, arguing that sequence differences cease to reflect evolutionary distances beyond the within-phylum scale. Thus, it could be

noise, as well as LGT, that cause different proteins to give different trees. Furthermore, they claimed, 'similar restrictions apply to the use of rRNA sequence data'. It is interesting to note that their fear that 'only God can make a tree' is with us again. Many authors now maintain that mutational saturation, long branch attraction and rate variations across sites and across trees make it impossible, or at least very problematical, to establish deep branching order (Forterre & Philippe 1999; Roger *et al.* 1999; Philippe *et al.* 2000).

5. THE DARK AGES OF LATERAL GENE TRANSFER

The 1980s produced many molecular phylogenies for many genes and many organisms, as gene sequencing became a standard tool—and as evolutionary speculations became one way to lend interest to otherwise routine results. At the same time, new methods of phylogenetic analysis appeared and found enthusiastic use, even though statistical methods for determining just how likely a best tree was to be a true tree were poorly developed. It is not surprising that the literature saw exciting claims for LGT that were in fact poorly supported and unlikely, in retrospect, to be true. In 1992, Russell F. Doolittle's group set a more cautious tone (Smith *et al.* 1992). Although endorsing half a dozen cases of between-domain transfer, they issued a general '*caveat comparer*', suggesting that credible cases for transfer should include support from multiple methods, a rooted tree with good taxon sampling, and some reasonable biological scenario about how the transfer might have occurred. Michael Syvanen, long a booster of LGT, similarly cautioned restraint in a 1994 review. 'Because so many examples have been offered and then retracted, and because there remains considerable scepticism as to the occurrence of horizontal transfer' one should be obliged to 'consider the rigor of the various criteria used for making this judgement' (Syvanen 1994).

Such conservative attitudes are much rarer in these post-genomic days. Nevertheless, Eisen wrote very recently: 'However, interesting these assertions are, I recommend caution before jumping on the HGT bandwagon. In most cases, the proposals for HGT involve only one type of evidence, such as nucleotide composition bias or unusual distribution patterns. As each of these can be caused by factors other than HGT, it will be important to see if they are supported by other lines of evidence' (Eisen 2000). Eisen's position seems to be that unless LGT is clearly proven, its absence (vertical descent) should be assumed—that vertical descent should retain the status of the null hypothesis. This would be justified if LGT were always rare, but both we, and he, know that there are genes, situations and timescales for which it is not. We suggest that for individual gene trees, poor support should leave us agnostic, while strong support for LGT events ought now to be accepted without 'other lines of evidence'. Some instances of 'unusual distribution patterns' provide overwhelmingly good evidence for transfer, if rules of parsimony are to be given any weight.

6. RENAISSANCE

When Russell Doolittle issued his '*caveat comparer*', the databases were a tiny fraction of their current size, and complete genome sequences were still the stuff of dreams. Five years later, Brown & Doolittle (1997) re-examined the issue of congruence between trees, at the deepest level. By that time, the SSU rRNA-based tripartite view of Life's history (that Bacteria, Archaea and Eukarya comprise all its domains) was widely accepted, as was the notion (based on phylogenies of some paralogous protein coding gene pairs) that Archaea and Eukarya are sister taxa (Iwabe *et al.* 1989; Gogarten *et al.* 1989; Baldauf *et al.* 1996). Although we could only assume midpoint rootings for the trees compared, and there were, as yet, few genes for which there was good representation from each domain, it was clear that trees *were* different—that extensive between-domain transfer *had* occurred. The literature now holds scores or hundreds of well-supported trees supporting lateral transfers within and between the bacterial and archaeal domains.

A sampling of such anecdotal cases is provided in table 1. Our own particular favourite is the gene for HMGC_{CoA} reductase. This catalyses the key step in the synthesis of mevalonic acid, an essential intermediate in the synthesis (for instance) of sterols in eukaryotes, quinones and menaquinones in bacteria, and the characteristic isoprenoid ether lipids of archaea. Trees prepared (by whatever method) from the two dozen or so prokaryotic and eukaryotic sequences available in 1999 showed at least four unquestionable cases of transfer: from bacteria into the archaean *Archaeoglobus fulgidis*, from archaea into a *Streptomyces* species, from archaea into *Vibrio cholerae* and from bacteria into the diplomonad eukaryote *Giardia lamblia* (Boucher & Doolittle 2000). Subsequent augmentation of the databases with further archaeal sequences showed that the *Archaeoglobus* transfer probably occurred at the base of the Archaeoglobales and that an independent transfer from bacteria happened early in the history of the Thermoplasmatales, or between these two archaeal groups after a bacterial transfer into one of them (Boucher *et al.* 2001).

Because, at least for *Archaeoglobus* and *Giardia*, the product of this gene is essential for survival, HMGC_{CoA} reductase transfers provide evidence for what we call orthologous replacement (figure 1c), the substitution of one version of a gene by another, probably through an intermediate stage in which both the resident (original) and incoming (replacement) copies are simultaneously present. The literature now holds many exciting cases of this previously unexpected evolutionary process. Aminoacyl-tRNA synthetases are especially prone to exchange and few, if any, show the 'Standard Tree' (monophyletic Bacteria, Archaea and Eukarya, with the last two sisters). Even components of the translational machinery which are more tightly integrated into complex structures, for instance ribosomal proteins, provide several clear cut examples (Brochier *et al.* 2000; Makarova *et al.* 2001).

7. THERMOTOGA AND METHANOSARCINA AS EXEMPLARS

Complete genome sequences of course provide hundreds or thousands of *potential* instances of LGT, and it

Table 1. Representative cases of lateral gene transfer.

gene	functional category	type of LGT ^a	direction of LGT ^b	detection method	reference
(a) prokaryotes					
aminoacyl-tRNA synthetases	translation, ribosomal structure and biogenesis	GA, OR or HR	multiple within and between all domains	PD	Woese <i>et al.</i> (2000)
ribosomal protein RpS14	translation, ribosomal structure and biogenesis	GA, OR	multiple within Eubacteria	PD, GC	Wolf <i>et al.</i> (1999) Brochier <i>et al.</i> (2000)
ribosomal proteins RpL31, RpL33, RpS14, RpS18, RpL32, RpL28	translation, ribosomal structure and biogenesis	GA, OR	multiple within Eubacteria	PD, GC	Makarova <i>et al.</i> (2001)
elongation factor Tu (tufB)	translation, ribosomal structure and biogenesis	GA	Streptococcaceae to Enterococcaceae	PD, SM	Ke <i>et al.</i> (2000)
rRNA operons	translation, ribosomal structure and biogenesis	GA or OR	<i>Thermobispora bispore</i> to <i>Thermomonospora chromogena</i>	PD	Yap <i>et al.</i> (1999)
recombination/repair system (mutS, mutL, mutH, mutU)	DNA replication, recombination and repair	HR	multiple between <i>E. coli</i> strains	PD	Denamur <i>et al.</i> (2000)
Walker-type chromosome partitioning ATPase (parA)	cell division and chromosome partitioning	GA, OR or HR	multiple between Archaea and Eubacteria	PD	Gerdes <i>et al.</i> (2000)
nodulation genes (nodA, nodB)	signal transduction mechanisms	OR	multiple inter- and intra-specific within Rhizobiaceae	PD	Wernegreen & Riley (1999)
two-component signal transduction systems	signal transduction mechanisms	GA, OR	multiple from Eubacteria to Archaea and Eukarya	PD, DP	Koretke <i>et al.</i> (2000)
lysine biosynthesis gene cluster (lys20, hacA, hacB, rimK, argC, argB)	amino acid transport and metabolism	GA, OR or HR	between <i>Pyrococcus horikoshii</i> and <i>Thermus thermophilus</i>	PD, DP	Nishida <i>et al.</i> (1999)
histidine biosynthesis genes (HisD, HisG)	amino acid transport and metabolism	GA or OR	multiple between Eubacteria and Archaea	PD	Bond & Francklyn (2000)
glutamine synthetase (GSI)	amino acid transport and metabolism	OR	Euryarchaea to low G+C Gram positives	PD	Pesole <i>et al.</i> (1995)
glutamine synthetase (GSII)	amino acid transport and metabolism	OR	multiple inter- and intra-generic within Rhizobiaceae	PD	Turner & Young (2000)
glutamate synthase	amino acid transport and metabolism	OR	multiple within and between Archaea and Eubacteria	PD	Nesbø <i>et al.</i> (2001a)
myo-inositol 1P synthase	carbohydrate transport and metabolism	GA, OR	multiple within and between Archaea and Eubacteria	PD	Nesbø <i>et al.</i> (2001a)
phosphoglucose isomerase	carbohydrate transport and metabolism	GA or OR	Eukarya to <i>Escherichia</i> and <i>Haemophilus</i>	PD	Katz (1996)
glyceraldehyde-3-phosphate dehydrogenase (Gap1, Gap2)	carbohydrate transport and metabolism	GA or OR	β -proteobacteria to CFB Eubacteria	PD	Figge <i>et al.</i> (1999)

(Continued.)

Table 1. (Continued.)

gene	functional category	type of LGT ^a	direction of LGT ^b	detection method	reference
ABC transport system (maltose, trehalose)	carbohydrate transport and metabolism	GA	between <i>Pyrococcus furiosus</i> and <i>Thermococcus litoralis</i>	DP, SM	Diruggiero <i>et al.</i> (2000)
rubisco (form I, rbcL)	carbohydrate transport and metabolism	GA or OR	multiple within Eubacteria and from Eubacteria to Eukarya	PD	Palmer (1995)
HMG-CoA reductase	lipid metabolism	OR	multiple between all domains	PD	Boucher & Doolittle (2000)
MEP cytidyltransferase	lipid metabolism	GA	Eubacteria to <i>Pyrococcus horikoshii</i>	PD, DP	Lange <i>et al.</i> (2000)
A/V ATPases (A and B subunits)	energy production and conversion	GA	multiple from Archaea to Eubacteria	PD, DP	Olendzenski <i>et al.</i> (2000)
photosynthesis gene cluster (crtEF-bchCXYZ-puf and bchFNBHLM-lhaA-puhA)	energy production and conversion	GA or OR	α -proteobacteria to δ -proteobacteria	PD	Igarashi <i>et al.</i> (2001)
HSP70 (dnaK)	post-translational modification, protein turnover, chaperones	GA	multiple from Eubacteria to Archaea	PD	Grihaldo <i>et al.</i> (1999)
catalase-peroxidase	inorganic ion transport and metabolism	GA or OR	Archaea to δ -proteobacteria	PD	Faguy & Doolittle (2000)
Nitrogenase (nifH)	inorganic ion transport and metabolism	GA or OR	α -proteobacteria to β -proteobacteria	PD	Hurek <i>et al.</i> (1997)
(b) eukaryotes					
glutamate synthase (small subunit)	amino acid transport and metabolism	OR	low G+C Gram pos. to Eukaryotes	PD	Andersson & Roger (2002)
glyceraldehyde-3-phosphate dehydrogenase	carbohydrate transport and metabolism	OR	Eubacteria to kinetoplastids and Eubacteria to diplomonads	PD	Qian & Keeling (2001)
phosphoenolpyruvate carboxykinase	carbohydrate transport and metabolism	GA or OR	Archaea to <i>Giardia</i>	PD	Suguri <i>et al.</i> (2001)
N-acetylneuraminate lyase	carbohydrate transport and metabolism	GA	Eubacteria to <i>Trichomonas</i>	PD, DP	de Koning <i>et al.</i> (2000)
5-monophosphate-dehydrogenase	nucleotide metabolism	OR	proteobacteria to <i>Cryptosporidium</i>	PD	Striepen <i>et al.</i> (2002)
sulphide dehydrogenase	sulphur metabolism	GA	anaerobic prokaryote to diplomonads	PD, DP	Andersson & Roger (2002)
malic enzyme	energy production and conversion	OR	Archaea to <i>Entamoeba</i>	PD, DP	Field <i>et al.</i> (2000)
NADH oxidase	energy production and conversion	GA	Eubacteria to <i>Giardia</i> and Eubacteria to <i>Entamoeba</i>	PD, DP	Nixon <i>et al.</i> (2002)

^a Type of LGT: GA, gene acquisition; HR, homologous recombination; OR, orthologous replacement.^b Detection method: PD, phylogenetic discordance; SM, overall sequence or motif similarity; GC, genomic context; DP, distribution pattern.

Figure 1. Four possible outcomes of LGT. (a) Novel acquisition: a gene conferring novel advantage is fixed for the first time in a lineage. (b) Loss and regain: an unused gene deleted and replaced by a homologue when again selected for. (c) Orthologous replacement: a homologue of an essential gene added, original subsequently lost. (d) Non-orthologous displacement: as orthologous replacement except that new gene(s) are analogues.

has become standard for publications reporting the completion of a prokaryotic genome project to present estimates of the fraction of genes that might be of 'foreign' origin. Two papers that illustrate the genre especially well (and in fact feature 'lateral gene transfer' in the their titles) are Nelson *et al.* (1999) and Deppenmeier *et al.* (2002), describing the genomes of *Thermotoga maritima* and *Methanosarcina mazei*, respectively. In the *Thermotoga* case, proper phylogenetic analyses of a limited number of genes and BLAST-based comparisons of the remainder led to the conclusion that 24% of this thermophilic bacterium's ORFs were derived by lateral transfer from archaea (and are not found in other bacteria). Two groups of investigators (Kyrpides & Olsen 1999; Logsdon & Faguy 1999) soon complained that, if *Thermotoga* were indeed the deepest bacterial branch (as SSU rRNA trees show), then the 'archaeal' genes it contains might as easily

be retained from the last common bacterial/archaeal ancestor and lost from the main bacterial trunk after *Thermotoga* diverged. In fact, such an explanation would be more parsimonious than the interpretation that they were transferred in, if we view transfer as generally less likely than loss. Worning *et al.* (2000) have, however, bolstered the claim for lateral transfer with evidence for structural heterogeneity of the DNA, whereas we (Nesbø *et al.* 2001a) have shown, in more rigorous phylogenetic analyses, that several *Thermotoga* genes are indeed the product of transfer, from specific archaeal groups, to lineages within the Thermotogales. It seems noteworthy that *Aquifex aeolicus*, the other fully sequenced bacterial hyperthermophile, also boasts many 'archaeal' genes. Because most sequenced archaeal genomes are also hyperthermophilic, perhaps the shared sequences are, more correctly, 'hyperthermophilic' genes.

Methanosarcina mazei and *M. acetivorans* have the largest genomes of any archaea that have so far enjoyed total genomic scrutiny (Deppenmeier *et al.* 2002; Galagan *et al.* 2002). They also have more 'bacterial genes' (ORFs with the closest homologue in a bacterial genome) than any other archaea—1043 (30%) and 945 (21%) of their respective totals of 3371 and 4524. (Note that numbers for the two genomes, differently annotated and analysed, are probably not strictly comparable.) For *M. mazei*, half of these bacterial ORFs have no significant match in archaea at all and *must* be transfers, if parsimony applies. The general message here is not simply that larger genomes have more room for foreign genes, although of course they do. By the same BLAST-based analysis, the next largest archaeal genome (that of *Sulfolobus solfataricus*, at 2977 genes) has one of the lowest bacterial gene contents (12.0%). Both *Methanosarcina* species are much more broadly talented, metabolically, than the first sequenced, smaller genomed *Methanobacterium thermoautotrophicum* (1918 genes) and *Methanococcus jannaschii* (1738), and some of their 'bacterial genes' are clearly responsible for this, although several complex pathways seem to mix-and-match bacterial and archaeal components. It is definitely *not* the case (with *M. mazei* anyway) that the bacterial contribution is limited to clusters or islands of functionally related or contiguous genes: the genome is a *fine-scale mosaic*.

8. GLOBAL MEASURES

These record holders aside, most completed prokaryotic genome sequences (of which there should be approximately 100 by the time this issue is published) show some proportion of 'foreign' genes, by one measure or another. In addition to scanning each genome for specific LGT events, it ought to be possible to use this enormous resource to get some global measure of the overall extent of LGT for any, and all, genomes. At least three general approaches have been applied.

In 1991, Danchin and collaborators (Medigue *et al.* 1991) noted that the *E. coli* genome comprised, on the basis of codon usage, three classes of genes, one of which they inferred to be the products of LGT. Ochman *et al.* (2000; see also Garcia-Vallve *et al.* 2000) have developed this compositional heterogeneity approach further, targeting genes of unusual base composition or codon usage characteristics (compared with most of the genome in which they reside), as probable transfers. This is a one-genome-at-a-time approach and, using it, they have recently reported foreign gene contents as low as zero (for *Mycoplasma genitalium* and *Rickettsia prowazekii*) and as high as 16.6% (for *Synechococcus* PCC6803). The technique is not without its detractors (Koski *et al.* 2001; Ragan 2001a,b; Hooper & Berg 2002), and in any case has a 'look back time' of the order of a few hundred million years (Lawrence & Ochman 1997).

Koonin and his collaborators, and many other groups after them, have used BLAST tools to measure the extents of LGT, defining as transferred those genes that have more similarity to homologues in distant rather than in closely related genomes—with varying criteria for accepting and differentiating matches (Koonin *et al.* 2001). This second, BLAST-based approach for measuring

global extents is by no means a substitute for thorough phylogenetic analysis: the method will have both false positives and false negatives, each sensitive to the parameters set for accepting or rejecting 'hits'. It is common to express distrust of such analyses (for instance Eisen 2000). However, genomes do differ substantially in size, so many genes must truly be present in some genomes and absent from others. When a genome contains a gene with *no* significant hit in any of several other 'related' genomes, but strong similarity to genes in a distant taxon, it seems perverse not to accept LGT as the best explanation. Thus we might quarrel about some genes among the 1043 in *M. mazei* that Deppenmeier *et al.* (2002) claim as transfers simply because the exponent of the e-value of the best bacterial hit exceeded that of the best archaeal hit by 14. However, in fact 544 of these 1043 had *no* significant archaeal match whatever. In any single such case, the doubter might still go on to claim (unparsimoniously) that the gene was present in the last common bacterial/archaeal ancestor and lost independently at least three and maybe five times (depending on what we take to be the tree relating sequenced archaea). However, this line of reasoning when applied consistently to all cases of patchy gene distribution assigns more and more genes to ancient common ancestral genomes, with every new genome sequenced. In the end, it requires a last universal common ancestor with an enormous range of metabolic capabilities—essentially any gene that is now seen in at least one contemporary archaean, and one contemporary bacterium must find a home in that ancient cell. Such a 'genome of Eden' hypothesis seems, to us, unappealing.

In a recent review, Koonin's group have summarized estimates of transferred genes within many prokaryotic genomes, using their BLAST method, which they call 'unexpected ranking of sequence similarity among homologues' (Koonin *et al.* 2001). Considering only exchanges *between* domains (bacterial transfers into archaea or the reverse), they report a minimum of 0% (*Buchnera* sp.) and a maximum of 15.6 (*Halobacterium* NRC-1). (*Thermotoga maritima* weighs in at 14%.) Perhaps such numbers are not at first impressive, and the average fraction of archaeal genes in the bacterial genomes was only 3%, whereas values for bacterial genes in archaea ranged from 4 to 8%. But, as Koonin *et al.* (2001) note, their method would exclude early inter-domain transfers, because they accept only genes whose inter-domain hits are all significantly less than a between-domain hit. Furthermore, between-domain transfers are the least likely to be fixed: domain-specific peculiarities in transcriptional and translational signals make the process more onerous than within-domain exchange. Applying the same method to detect exchange between major lineages *within* Bacteria, these workers do indeed get higher numbers (again 0% for *Buchnera*, but 32.6% for *Treponema pallidum* and 28.8% for *Bacillus halodurans*). Somewhat more than half of these apparent exchanges are what Koonin calls 'paralogous acquisition' or 'displacement' events (the latter equivalent to our orthologous replacement). For such transfers, within-phylum exchange (that is, among cyanobacteria or among α -proteobacteria) should be *even more* frequent. Each lineage has (is sometimes defined by) characteristic biochemistry: a cyanobacterium might benefit by replacing a component of its light-harvesting apparatus by a dif-

ferently tuned homologue from a sister cyanobacterium, whereas a spirochaete would simply have no use for such a gene. If we add together within-species (or genus) recombination and within-lineage, within-domain and between-domain transfers, it may be that very few genes now found in the same genome will have the same evolutionary history, all the way back to the last universal common ancestor.

A new BLAST-based method for identifying LGTs that seems independent of any phylogenetic assumptions has been described by Charlebois, Ragan and collaborators (Clarke *et al.* 2002). GenBank is sorted by species, and ORFs from a chosen genome are classified as 'phylogenetically discordant' or not depending on whether their pattern of BLASTP matches are significantly different from the median pattern displayed by all the ORFs of that genome. Interestingly, this method detects apparent LGTs in the genomes of intracellular parasites, where Koonin's group finds few or none.

All analysts agree that the very best evidence for any individual LGT event is a robust phylogenetic tree, with good taxon sampling and strong bootstrap values, showing that on the basis of gene W, taxon X should be placed *within* clade Y, when otherwise it belongs *within* clade Z. (As a philosophical *caveat* we note that 'otherwise it belongs' is a problematical phrase. All we can really say without further unproven assumptions is that genes robustly disagree, probably because of LGT, since we have no logically uncontested way of designating some gene or genes as unique markers of organismal phylogeny.) Good trees are hard to make: most experts distrust automatic alignment algorithms and much human judgement still seems to be involved in choices of taxa and evolutionary models in any analysis (see any issue of *Molecular Biology and Evolution*). Thus, no one has yet performed phylogenetic analyses of maximum rigour for all, or even most of, the genes in any sequenced prokaryotic genome. There *have* been a few brave attempts to set up automated all-the-trees-for-all-the-genes analyses, perhaps the most successful being that of Sicheritz-Pontén & Andersson (2001). Their program makes neighbour joining or maximum-parsimony trees for all genes (automatically aligned by CLUSTALW), and categorizes them according to whether the sister gene is of the same or different group (at several different taxonomic ranks): if the sister taxon comprises multiple species, it considers whether these all belong to the same group. By this definition, the best cases for inter-domain transfer are genes that have a sister clade comprising only genes of another domain. The numbers are impressively high, given the reasons discussed above to expect inter-domain transfers to be the rarest—17% for *Thermotoga maritima* and 19% for *Pyrococcus abyssi*, for instance.

9. THE NOTION OF A CORE

Different methods for detecting and quantifying gene transfer into genomes detect different genes and yield different global estimates (Ragan 2001a,b; Koski *et al.* 2001). This is disheartening, but no excuse for dismissing (as Kurland (2000) does) very many well-supported instances of transfer, nor for ignoring the strength of the general argument advanced above concerning genes that are sim-

ply not present in a genome's closest relatives, but well known from distant taxa. An enormous amount of data of this latter sort has begun to appear as genome sequencers start to look at multiple strains within a given bacterial 'species', and the case for massive LGT over relatively short evolutionary time-scales (however long strains of a species have been diverging) is incontrovertible (see next section). Reliable global estimates for the fraction of any genome's genes that have different histories from each other (and how many different histories they might in total show) may, however, be a long time coming.

Thus, it might be fruitful to turn the problem on its head. Perhaps instead of asking 'How many genes in this or any genome come from different places, and how many places?' we should ask 'How many and which genes shared by *all* genomes in a given group, or indeed by all prokaryotes, have the same evolutionary history and produce the same phylogenetic tree?' In other words, for any group, is there a stable core of genes that do faithfully record population bifurcations and speciation events, back to the group's last common ancestor?

At the species level, for some bacteria, frequent interstrain and between-species homology-dependent recombination means that the answer to this question is demonstrably *no!* Feil and collaborators, after reviewing the evidence from multi-locus sequence-typing studies of several pathogenic bacteria, concluded that 'over the long term, the impact of relatively frequent recombination is to obliterate the phylogenetic signal in gene trees, such that the relationships between major lineages of many bacterial species should be depicted as a network rather than a tree' (Feil *et al.* 2001).

For some other species, and for all prokaryotes at some intermediate taxonomic rank, where homology-dependent recombination should be greatly reduced but LGT has had insufficient time to replace any genomes' entire contents, the answer must surely be *yes*. Some core of genes will truly give the same tree. Included among them would probably be those genes thought to be poor candidates for transfer. Woese has, since his initial choice of SSU rRNA as 'universal molecular chronometer', maintained that this gene is virtually immobile, because its product is so vital and interacts with so many other, coevolved, RNAs and proteins (Woese 1987). Jain *et al.* (1999) named this notion the *complexity hypothesis* and argued (as had Woese) that a similar immobility should afflict genes for other components of the transcriptional and translational machinery. In fact, we do now know of transfer events affecting aminoacyl tRNA synthetase, ribosomal protein, elongation factor and even rRNA genes (Ke *et al.* 2000; Woese *et al.* 2000; Makarova *et al.* 2001; Parker 2001). In general, however, we would expect such genes and perhaps many others to define a core for many taxa at *some* moderate level of inclusiveness.

Different taxa would not necessarily have the same stable core. Indeed, if *all* phyla or divisions within Bacteria (or Archaea) had the same core, or even if just some part of all taxon-specific cores within either domain comprised the same genes, then we could call that shared part the domain-specific stable core. Furthermore, if there were an overlap between these bacterial and archaeal cores, we would have a universal stable core. Add to this possibility the assumption that such universally shared genes show

the same tree because they have never been transferred (rather than because they have always been transferred together), and one derives what we think is the prevailing current belief about the relationship between gene and universal organismal phylogenies. Namely, *the phylogeny of genes in some universal stable core (almost certainly including rRNA genes) is the phylogeny of all organisms.*

Are these appealing notions testable or true? Are there such stable domain or universal cores? Three kinds of test have been applied. First, attempts have been made to individually construct and then compare universal trees for all genes shared by all (or a representative subset of) sequenced genomes, or by genes widely thought to be part of such a shared set, such as ribosomal proteins (Teichman & Mitchison 1999; Hansmann & Martin 2000; Matte-Taille et al. 2002; Brochier et al. 2002). Individually, such genes have too little phylogenetic signal. All reproduce the bacterial/archaeal dichotomy but few, if any, resolve major groups within the domains. The trees are not congruent, but they are also not robust: it could be noise that makes them disagree.

Because of the noise problem, Philippe and coworkers (Brochier et al. 2002; Matte-Taille et al. 2002) have developed a second method. They concatenate (string together) the sequences of a large fraction of bacterial or archaeal ribosomal and translation-related proteins, make domain-specific trees, and then use principal component analysis to determine whether any genes within the set disagree significantly with the signal of the concatenated sequence as a whole. Indeed, some do disagree, and are eliminated as cases of LGT, but 44 (of 57) bacterial and 45 (of 53) archaeal proteins do, by this test, seem to be singing the same phylogenetic song. Brown et al. (2001), applying a similar concatenation approach to link both bacteria and archaea into a single, universal tree, also obtained what they considered a robust result, but had to reduce the final set to 14 genes (less than 1% of any of the organisms' gene complement!).

The authors (Nesbø et al. 2001b) and Zhaxybayeva & Gogarten (2002) have tackled the goal of testing whether genes found in apparently orthologous copies in all genomes of some apparently monophyletic assemblages have the same tree, using yet a third approach. We simplify the analyses greatly by choosing genomes four at a time. Because there are only three possible unrooted topologies for a four-taxon tree, we can ask of each alignable set of four apparent orthologues which tree it prefers, and tabulate the results. If all shared orthologues have the same history (in other words, if there have been no orthologous replacement events), one of three topologies should be overwhelmingly preferred. In such analyses to date, whenever four roughly equally distant taxa have been chosen, this seems not to be the case. All three possible topologies are well represented. One might still be the true topology of the core, but there is no obvious way to identify it, and it is clear that there has been much orthologous replacement among genes shared by any set of genomes examined.

Thus, no attempt to prove that there is any substantial stable core of genes by which universal organismal phylogeny might be reliably tracked has succeeded, unless we are willing to accept just a few per cent of a genome (such as the translational proteins analysed by Philippe and

colleagues) as that core (Brochier et al. 2002). In this case, because the genes thus identified are often tightly clustered, the possibility of co-transfer cannot be dismissed, at least not without due consideration.

10. CLOSELY WATCHED GENOMES

All doubt about the physiological, taxonomic and phylogenetic importance of LGT over the short evolutionary term should vanish in contemplation of the genomes of *Escherichia coli* O157:H7 and K12. The former, a nasty pathogen, bears some 1387 genes not found in the latter, a benign laboratory strain, which nevertheless itself boasts 528 genes missing from the former (Perna et al. 2001). These two genomes differ in 26 or 12% of their genes (depending on which way you cast the comparison). There is probably more variation in genetic information between these two strains of what is considered a single species than within, say, all the mammals, maybe even all the vertebrates! Such between-strain variation is now known to be typical for pathogenic bacteria, and establishing and quantifying within-species genomic diversity, by complete or limited-pass sequencing, micro-array analysis and subtractive hybridization has become a growth industry (for a summary, see table 2). Our own survival in an increasingly aggressive and versatile microbial environment may depend on such knowledge. At the same time, good evidence that extensive genomic variability is not restricted to pathogens has begun to accumulate. From a study of large environmental genomic fragments bearing the SSU rRNA signature of *Cenarchaeum symbiosum*, Beja et al. (2002) conclude that 'considerable functional diversity may exist within single populations of coexisting microbial strains, even those with identical 16S sequences'. In our own suppressive subtractive hybridization study with very closely related strains of *T. maritima* hybridized as 'testers' against the sequenced *T. maritima* (the 'driver'), we find between-genome variations in gene content as extensive as known for many pathogens (Nesbø et al. 2002). Genes that are present in some *Thermotoga* strains determine a great range of biochemical activities, and seem to come from a variety of bacterial and archaeal sources. We also find, among clones isolated by subtractive hybridization, distant homologues of genes in the sequenced genomes: orthologous replacement may well be a common process in *Thermotoga*'s environment.

11. FOUR KINDS OF GENE EXCHANGE

Abstractly considered, LGT events (by which we mean exchanges successfully fixed in recipient species, not the various physical mechanisms by which DNA might get into prokaryotic cells) can be of four types (figure 1), each with its own possibilities and limitations. Novel acquisition (of a brand new function) can have dramatic effects. Examples might be: (i) the founding of a clade of marine proteobacteria using light to pump protons thanks to acquisition of a halorhodopsin gene from an archaean (Beja et al. 2001); (ii) the establishment of methanotrophy in bacteria though co-optation of archaeal genes of methanogenesis (Graham et al. 2000); and (possibly) (iii) the conversion from anaerobic to aerobic lifestyle in

Table 2. Some examples of differences in genome content among closely related strains. (For micro-array hybridization studies the percentage refers to the proportion of genes missing relative to the reference genome. For subtraction studies the percentage refers to the number of genes present in the tester species not found in the reference genome. % VG refers to percentage variable genes.)

species	% VG	comments	reference
<i>(a)</i> by comparison of genome sequences			
<i>Chlamydia trachomatis</i>	< 1	two genomes, orthologues show 90% DNA identity	Read <i>et al.</i> (2000)
<i>Chlamydia pneumoniae</i>	< 1	two genomes, orthologues show 99.9% DNA identity	Read <i>et al.</i> (2000)
<i>Escherichia coli</i>	12–26	two genomes, orthologues show 98.4% DNA identity	Perna <i>et al.</i> (2001); Hayashi <i>et al.</i> (2001)
<i>Helicobacter pylori</i>	6–7	two genomes, orthologues show 92.6% DNA identity	Alm & Trust (1999)
<i>Neisseria meningitidis</i>	9–11	two genomes	Lan & Reeves (2000); Tettelin <i>et al.</i> (2000)
<i>Salmonella enterica</i>	11–13	<i>S. enterica</i> serovar <i>typhimurium</i> contains 89% of the ORFs found in serovar <i>typhi</i> and 87% of the ORFs in serovar <i>paratyphi</i> A, shared orthologues show 98% DNA identity for both strains	McClelland <i>et al.</i> (2001)
<i>Salmonella enterica</i>	10–12	five serovars— <i>dublin</i> , <i>enteritidis</i> , <i>paratyphi</i> , <i>typhimurium</i> , <i>typhi</i> , shared orthologues show 99.5–97.6% DNA identity	Edwards <i>et al.</i> (2002)
<i>Staphylococcus aureus</i>	7	two strains from the clone type II-A (N315 and Mu50), shared orthologues show greater than 99% DNA identity, only Mu50 specific genes are reported	Kuroda <i>et al.</i> (2001)
group A <i>Streptococcus</i>	6–9	two serovars (M1 and M18), shared orthologues show DNA identity in the range 80–100%	Smoot <i>et al.</i> (2002)
<i>(b)</i> by microarray hybridization			
<i>Campylobacter jejuni</i>	0.5–7	11 strains, total 21% variable genes in reference genome	Dorrell <i>et al.</i> (2001)
<i>Escherichia coli</i>	1–10	5 strains	Ochman & Jones (2000)
<i>Helicobacter pylori</i>	6–18	15 strains, total 22% variable genes in reference genome	Salama <i>et al.</i> (2000)
<i>Helicobacter pylori</i>	0.3–1.5	13 clones from the same patient as that from whom J99 was isolated but sampled 6 years later, total 2.3% variable genes in reference genome	Israel <i>et al.</i> (2001)
<i>Mycobacterium tuberculosis</i>	0–0.7	16 strains of which 13 were isolated from patients in San Francisco, total 1.7% variable genes in reference genome	Kato-Maeda <i>et al.</i> (2001)
<i>Mycobacterium tuberculosis</i> complex	1.6–2.4	13 BCG strains and 8 virulent <i>M. bovis</i> strains compared with a <i>M. tuberculosis</i> H37Rv array, total 3.4% variable genes in reference genome, strains within the <i>M. tuberculosis</i> complex show greater than 99.9% DNA identity for shared orthologues	Behr <i>et al.</i> (1999)
<i>Salmonella enterica</i>	8–17	three strains (serovars)	McClelland <i>et al.</i> (2001)
<i>Staphylococcus aureus</i>	0.8–12	36 strains, total 22% variable genes in reference genome	Fitzgerald <i>et al.</i> (2001)
<i>Streptococcus pneumoniae</i>	2.9–4.4	two strains, total 4.7% variable genes in reference genome	Tettelin <i>et al.</i> (2001)
<i>Streptococcus pneumoniae</i>	2.7–11	20 strains, all strains except one differ by 8–11%, total 24% variable genes in reference genome	Hakenbeck <i>et al.</i> (2001)
group A <i>Streptococcus</i>	0–3	36 strains of serotype M18 collected between 1931 and 1999, total 12% variable genes in reference genome	Smoot <i>et al.</i> (2002)
<i>Vibrio cholerae</i>	0–1	nine pathogenic strains, total 4% variable genes in reference genome	Dziejman <i>et al.</i> (2002)
<i>(c)</i> by genomic subtraction			
<i>Escherichia coli</i>	8	an <i>E. coli</i> K12 strain was used as driver against an avian pathogenic strain, the subtracted DNA was hybridized to a cosmid library	Brown & Curtiss (1996)
<i>Listeria monocytogenes</i>	5–6	a prototype virulent strain genome was subtracted from a prototype epidemic strain	Herd & Kocks (2001)
<i>Salmonella enterica</i>	2–20	genomic DNA from LT2 was subtracted against four different strains and hybridized to cosmid library	Lan & Reeves (1996)
<i>Salmonella enterica</i> serovar Typhimurium	3	genomic subtraction using serovar Typhi as driver, serovar Typhimurium estimated to contain 140 kb not found in Typhi	Emmerth <i>et al.</i> (1999)
<i>Thermotoga</i> spp. RQ2	20	20% using <i>T. maritima</i> MSB8 as driver, orthologues show ca. 96% DNA identity	Nesbø <i>et al.</i> (2002)

haloarchaea, after obtaining genes of respiration from bacteria (Kennedy *et al.* 2001). A pattern of loss and regain would be expected for genes of the sort addressed by Lawrence and Roth's 'selfish operon' theory (Lawrence & Roth 1996). That is, genes that have become dispensable (genes for sugar uptake in the absence of the sugar), will suffer mutational erosion and loss, but might later be regained through selection of transferred orthologues or analogues (unrelated genes of the same function), when the need returns (sugar again available). One could argue that individual prokaryotic lineages and prokaryotes collectively could not avoid loss of many sometimes optional biochemical pathways were it not for the operation of such a process. Without it, the prokaryotic world might consist very largely of extreme specialists. Orthologous replacement we envision as a two-step process, sometimes driven by a transient advantage of having both the foreign and resident orthologues present simultaneously. Resistance to natural antimicrobials directed against the product of the resident gene (with the incoming version being more resistant) would be one obvious driving mechanism. Adjustment to marginal physical regimes (temperature or salinity) might be another. In situations where the recipient is constantly exposed to DNA from the donor, chance alone might be enough to drive replacement. Several of the results described in the previous sections indicate that this process is much more common than we had appreciated. Analogous replacement would be similarly motivated and have similar biological consequences, but recombination between incoming and resident genes would not occur.

12. WHAT ABOUT EUKARYOTES?

The concern expressed by one of us three years ago, that LGT renders any gene-based universal Tree of Life problematical *as phylogeny*, is still a valid one (Doolittle 1999; see also Martin 1999). Any new understanding of the meaning of the SSU rRNA tree, as it depicts prokaryotes, will need to be much more complex than original versions (compare Woese 2000 with Woese 1987, 1999). However, nobody seriously thinks that all the hundreds of SSU rRNA trees relating birds and bees and bamboos are at risk because of LGT. Recent claims for bacterial transfer, specifically into primates, seem most unlikely to be true (Andersson *et al.* 2001), and the routes by which foreign genes might make it into the germ lines of complex multicellular eukaryotes are surely narrow and tortuous.

Nevertheless, all currently popular models for the origin of the eukaryotic genome would have it be genetically chimeric in its very origins and fundamental nature. Organellar genomes are after all integral parts of eukaryotic genomes, and specifically nuclear genes that derived from organelles or their immediate ancestors are numerous. Others in this volume will address these organelle-derived genes, and chimeric models (in addition to usual versions of the SEH) which describe nuclear gene complements in terms of a few founding partners (see Martin *et al.* 2001). We shall limit ourselves to considering genes that eukaryotic nuclear genomes might have acquired one, or a few at a time, by processes similar, *in their genetic outcomes*, to those described in figure 1. Some eukaryotes might, because of features of their biology and access to

foreign DNA, be as frequently the beneficiaries of trans-specific (or trans-domain) LGT as many prokaryotes, despite the lack of known evolved mechanisms encouraging interspecific exchange, such as conjugation or transformation.

Although organelle-to-nucleus transfers, as classically conceived, are thus specifically *not* examples of such trans-specific LGT, and specifically *not* our concern here, models developed to rationalize these transfers *do* provide a springboard for further discussion. Recently, Berg and Kurland have reviewed experimental and theoretical studies of gene transfer between mitochondrial and nuclear genomic compartments (Berg & Kurland 2000), whereas Selosse *et al.* (2001) survey explanations for the apparent unidirectionality of the process. The experimental system of Thorsness and collaborators (using yeast) allowed estimates of rates of transfer of intact genes in the range of one per 10^5 generations from mitochondrion to nucleus, and five orders of magnitude lower than that for transfer from nucleus to organelle (Thorsness & Fox 1993; Thorsness & Weber 1996; Campbell & Thorsness 1998; Shafer *et al.* 1999). The primary mechanism, they argue, is autophagy of mitochondria by cellular lysosomes. Although most genes should locate in inappropriate places in the nuclear DNA, Baker & Schatz (1987) demonstrated some time ago that at least a few per cent of such transferred genes might expect to find themselves downstream of suitable transit sequences and/or promoters. Of course, adequate expression of transferred genes in the nucleus would be a necessary precursor to subsequent loss through mutation (and drift or selection) of the parent copy from mitochondrial DNA, for any essential gene. One should also not forget that the relocation of organellar genes to the nucleus is a two-step process, and that each step must account for itself through normal population genetic mechanisms. It is surely inappropriate to see some sort of long-term drive towards cellular integration as a selective force.

Berg & Kurland (2000) model the organelle-to-nucleus transfer process mathematically. After exploring several parameters, they conclude that in the 'realistic case of finite populations, gene transfer favouring the nucleus and/or higher mutation rates in the mitochondrion will drive mitochondrial genes to the nucleus'. Selosse *et al.* (2001) come to the same conclusion, about relative rates, while also invoking selection for small organellar genome size as a significant force. Both argue that measurement of rates of transfer in either direction will be crucial in testing such models. In a more purely intuitive proposal we (Doolittle 1998) emphasized that, even if rates of acquisition by nuclear genomes of organellar genes are immeasurably small, as long as they are non-zero and higher than rates of transfer in the reverse direction, a genetic ratchet mechanism will ensure that all genes that *can* be relocated to the nucleus eventually *will* be. A transferred organellar gene (derived from lysed organelles) will be redundant to its organellar homologue (in the unlysed siblings of the donor organelle). Either could be lost, and thus one will be. If the nuclear copy disappears, there always be 'more where that came from': the transfer can be 'tried' again. However, loss of the organellar copy should be permanent. Measurements of real rates (per cell division) of organelle lysis, DNA integration and

promoter/targeting sequence are all desirable in themselves, but in the context of the enormous number of cell divisions since the origin of mitochondriate eukaryotes, almost irrelevant.

We (Doolittle 1998) also suggested, together with Maynard-Smith & Szathmari (1995), that versions of the SEH that describe its early stages in terms of mutual benefit are metaphorically misleading. The latter authors describe the relationship, more frankly, as 'slavery'. They suggest that 'host cells may have kept protomitochondria as humans keep pigs'. We think it is indeed instructive to describe the protomitochondrial period in terms of prokaryote husbandry. Suppose ingested bacteria could continue to produce cell mass heterotrophically or autotrophically, at the expense of chemical or physical energy sources not usable by the ingesting protist, but readily available in its environment. (Cyanobacteria and/or ingested plastids persevering in photosynthetic carbon fixation in the vacuoles of transparent hosts are examples (Margulis 1981).) Allowing prey to grow and even divide before lysis and catabolism of their cell contents multiplies the nutritional value of any single meal, and provides much added value to the predator, when prey are scarce. Means for transporting metabolites across membranes or walls of predators or prey need not appear until later (if ever). Our view of successive stages in the development of the 'organelle-host' relationship is outlined in figure 2. At each stage, phagocytosis or autophagy of food, endosymbiont or organellar particles should lead, however rarely, to the incorporation of prokaryotic genes into nuclear DNA. At each stage, something like the genetic ratchet described for organelle-to-nucleus transfer should operate.

Of course, mitochondria and plastids have each arisen only once, and prokaryote husbandry relationships that have not 'progressed' into extreme genetic modification may be far more common. Fenchel & Finlay (1995) note that modern anaerobic protists (very like what we might imagine archezoa to be) are, as a group, 'remarkable for the high proportion of species which live in symbiotic association with prokaryotes. They are also unique in the types of symbiotic partners they have acquired: endosymbiotic purple non-sulphur photosynthetic bacteria and methanogens; and ectosymbiotic sulphate-reducing bacteria.' The specific α -proteobacterial and cyanobacterial slaves that gave rise to modern mitochondria and plastids are but two of the possible prokaryotic foods, friends (endosymbiont) or foes (intracellular parasites) that might download genes into the eukaryotic nucleus. Several experimental models in which pathogenic or innocuous bacteria yield up their genes to engulfing mammalian phagotrophic cells are available, and could provide estimates of rates (e.g. Hense *et al.* 2001). Most studies involve intracellular pathogens that seek to evade digestion, and analyses aimed directly at measuring gene transfer from food bacteria into protists are still much needed. However, even a very low rate of leakage of DNA from lysosomes (as in the yeast mitochondria studies of Thorsness and his colleagues) should be effective in the evolutionarily long run.

Transferred genes will be fixed in nuclear genomes as *novel acquisitions* (figure 1) if their expression there offers advantage, and/or if at least one copy per cell is essential, and the organelle or permanent symbiont has lost its ver-

sion. More interestingly, we think, a similar genetic ratchet might effect the *orthologous replacement* of nuclear genes by homologues or analogues from the favourite food source, habitual symbionts or recurring pathogens, as illustrated in figure 3. The prediction is that, if a protist is habitually exposed to such a source of DNA, any of its endogenous genes that can (without serious detriment) be replaced by prokaryote-origin genes eventually *will* be: one crank of the ratchet (figure 3) is enough.

The proof of this pudding will of course be in the eating. *Have* protists that eat or live intimately with prokaryotes indeed incorporated genes from bacteria (or archaea) other than the 'protomitochondrion' into their nuclear genomes? Half a dozen instances of well-supported prokaryote-to-eukaryote transfer are known in the literature, and are included in table 1. All involve protists as probable recipients. With Martin Embley and his team, two of us (J.O.A. and A.J.R.) have used the available *Giardia lamblia* genomic information (McArthur *et al.* 2000) and an ongoing *Spironucleus barkhanus* sequence survey project to identify genes of likely bacterial origin in the nuclear genomes of these phagotrophic, amitochondriate diplomonad protists. Fifteen apparent cases of LGT (12 represented in available data from both protists) were examined thoroughly. Most or all of these 15 genes were introduced into the diplomonads by LGT before the divergence of *Giardia* and *Spironucleus*. Half are involved in anaerobic metabolism, suggesting that the common ancestor of *Giardia* and *Spironucleus* used genes from anaerobic protists to adapt to an oxygen-free life. Each of Life's three domains (Bacteria, Archaea and other eukaryotes) is represented among potential donors of these genes, most of which have clearly been involved in other transfers, including other transfers into eukaryotes (Andersson *et al.* 2002).

Fifteen genes is, to be sure, no large fraction of any genome, but it will not be easy to progress from anecdotal studies such as these to global estimates of the importance of LGT in eukaryote evolution, for several reasons. First, the last common ancestor of all characterized eukaryotes was probably already a 'sophisticated' cell, with most of the intracellular ultrastructural and macromolecular complexity of modern eukaryotic cells (Dacks & Doolittle 2001). An enormous amount of structural and functional reorganization has been wrought since its divergence from the presumed archaeal ancestor of the first 'archezoan', and transfers from bacteria (which would then be symple-siomorphies for extant eukaryotes) probably played a significant part. Second, without a firmer understanding of protist phylogeny (Roger 1999), we will not be able to venture any reasonable guesses about which 'bacterial' genes found in more than one protist group are symple-siomorphic (and thus not LGT events, at least since the last eukaryotic common ancestor) and which are synapomorphies for the clades they unite (and thus unequivocally LGTs). The 15 events characterized by Andersson & Roger (2002) and Andersson *et al.* (2002) were chosen because they had *no* eukaryotic BLAST hits (outside diplomonads) approaching their bacterial hits. Third, any transfers from poorly characterized phyla of bacteria or archaea will not produce trees with robust support for LGT. Such uncultivated 'phyla', known only from environmental SSU rRNA sequences, may outnumber cultivated ones (Pace 1997). Unless the transferred gene

Figure 2. An unsentimental view of gene transfer during the evolution of organelles.

is itself available from the uncultivated taxon, such transfers will show only vague bacterial affinity, insufficient for any well-supported LGT argument (Eisen 2000).

As with establishing transfer within prokaryotes, there may be a problem with the null hypothesis. The standard view is that eukaryotes (archezoa) are sisters to Archaea (Woese 1999, 2000), thus all else being equal, their genes *should* look at least vaguely 'archaeal'. Any eukaryotic nuclear gene that is no more clearly archaeal than bacterial would seem a candidate for transfer, even if not embedded within a well-sampled bacterial clade. It might be too bold to claim such genes as LGTs, but it is surely most unreasonable to take them as clear instances of 'vertical descent' as evidence favouring the null hypothesis of archaeal ancestry. This is exactly what Canback *et al.* (2002) have done in a recent analysis of genes for eight glycolytic enzymes common to eukaryotes and *Bartonella*

henselae. These show no robust affinity to any genes from known bacteria, but (in most cases) have no archaeal homologues whatever. Instead of Canback *et al.*'s verdict of 'innocent' on the charge of LGT, 'not proven' would have been appropriate (and some guilty genes are probably getting off!). As long as standards of proof for LGT and 'vertical descent' remain different, it will be difficult to obtain fair measures of the extent of either. And until there is a broader and more complete sampling of protist genomes, it will be impossible to tell which genes known so far only from the frequently sampled animals, plants and fungi are in fact part of the ancestral eukaryotic heritage.

13. HOW BIG IS THE ICEBERG?

Five years ago, most of us only considered LGT as a nuisance in phylogenetic reconstruction. We assumed that

Figure 3. The 'you are what you eat' genetic ratchet for orthologous replacement (after Doolittle 1998).

most genes in most genomes would eventually prove, with appropriately realistic models of sequence evolution and adequate taxon sampling, to have the same evolutionary history. This is actually quite an onerous notion, properly considered. *Even one* event of interspecies LGT in all of the enormous number of gene replications along the line stretching back from a gene in a modern sequenced genome to its ancestor in the genome of the last common ancestral cell, which most of us think lived more than 3.5 billion years ago, falsifies this assumption, for that gene. Perhaps we could relax a bit, demanding only that such interspecific LGT events, when they occurred, spanned only short 'phylogenetic distances' (within 'genera' or at worst within 'phyla' of bacteria, for instance). They might thus remain inconsequential; our methods may never be precise enough to detect them as differences between gene trees for different genes. Nevertheless, given the extensive data supporting transfer of many genes among many genomes reviewed above, this assumption seems unlikely to be correct, even in the relaxed form, for most genes in most genomes. The debate now must be over whether it is true in its precise form for *any* genes in *any* genomes. The community, and we ourselves, are divided on the issue.

If the iceberg of LGT does sink the ship of genome-based phylogeny, we have many lifeboats (each gene is one). Our understanding of gene and genome evolution will be richer, not poorer. Only simple metaphors (the 'Tree of Life') and dreams of hierarchical orderliness passed down to us from Linnaeus are at risk. In unravelling the complex web of gene phylogenies, we have much left to do!

Work in the laboratories of W.F.D. and A.J.R. is supported by the Canadian Institutes for Health Research (CIHR), the Natural Sciences and Engineering Research Council and Genome Atlantic. As individuals, the authors are supported by the Canada Research Chair Program (W.F.D.), CIHR (Y.B. and C.L.N.), the Killam Foundation (C.J.D.), the Wenner-Gren Foundation (J.O.A.) and the Canadian Institute for Advanced Research (A.J.R.).

REFERENCES

- Alm, R. A. & Trust, T. J. 1999 Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* **77**, 834–846.
- Ambler, R. P., Daniel, M., Hermoso, J., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. 1979a Cytochrome *c2* sequence variation among recognized species of purple nonsulfur photosynthetic bacteria. *Nature* **278**, 659–660.
- Ambler, R. P., Meyer, T. E. & Kamen, M. D. 1979b Anomalies in amino acid sequences of small cytochromes *c* and cytochromes *c'* from two species of purple photosynthetic bacteria. *Nature* **278**, 661–662.
- Anderson, E. S. 1966 Possible importance of transfer factors in bacterial evolution. *Nature* **209**, 637–638.
- Andersson, J. O. & Roger, A. J. 2002 Evolutionary analysis of the small subunit of glutamate synthase: gene order conservation, gene fusion, and prokaryote-to-eukaryote lateral gene transfers. *Euk. Cell* **1**, 304–310.
- Andersson, J. O., Doolittle, W. F. & Nesbo, C. L. 2001 Genomics: are there bugs in our genome? *Science* **292**, 1848–1850.
- Andersson, J. O., Horner, D. S., Dyal, P. L., Sjögren, Å. M., Davis, L., Embley, T. M. & Roger, A. J. 2002 Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* (In the press.)

- Baker, A. & Schatz, G. 1987 Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc. Natl Acad. Sci. USA* **84**, 3117–3121.
- Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. 1996 The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93**, 7749–7754.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. 1999 Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523.
- Beja, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. 2001 Proteorhodopsin phototrophy in the open ocean. *Nature* **411**, 786–789.
- Beja, O., Koonin, E. V., Aravind, L., Taylor, L. T., Seitz, H., Stein, J. L., Bensen, D. C., Feldman, R. A., Swanson, R. V. & DeLong, E. F. 2002 Comparative genomic analysis of archaeal genotypic variants in a single population and in two oceanic provinces. *Appl. Environ. Microbiol.* **68**, 335–345.
- Berg, O. G. & Kurland, C. G. 2000 Why mitochondrial genes are most often found in nuclei. *Mol. Biol. Evol.* **17**, 951–961.
- Bond, J. P. & Francklyn, C. 2000 Proteobacterial histidine-biosynthetic pathways are paraphyletic. *J. Mol. Evol.* **50**, 339–347.
- Boucher, Y. & Doolittle, W. F. 2000 The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.* **37**, 703–716.
- Boucher, Y., Huber, H., L'Haridon, S., Stetter, K. O. & Doolittle, W. F. 2001 Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. *Mol. Biol. Evol.* **18**, 1378–1388.
- Brochier, C., Philippe, H. & Moreira, D. 2000 The evolutionary history of ribosomal protein RpS14. *Trends Genet.* **16**, 529–533.
- Brochier, C., Baptiste, E., Moreira, D. & Philippe, H. 2002 Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* **18**, 1–5.
- Brown, J. R. & Doolittle, W. F. 1997 Archaea and the prokaryote to eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502.
- Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. 2001 Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28**, 281–285.
- Brown, P. K. & Curtiss III, R. 1996 Unique chromosomal regions associated with virulence of an avian pathogenic *Escherichia coli* strain. *Proc. Natl Acad. Sci. USA* **93**, 11 149–11 154.
- Bryson, V. & Vogel, H. J. 1965 *Evolving genes and proteins*. New York: Academic.
- Campbell, C. L. & Thorsness, P. E. 1998 Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal compartments. *J. Cell Sci.* **111**, 2455–2464.
- Canback, B., Andersson, S. G. E. & Kurland, C. G. 2002 The global phylogeny of glycolytic enzymes. *Proc. Natl Acad. Sci. USA* **99**, 6097–6102.
- Cavalier-Smith, T. 1987 Eukaryotes with no mitochondria. *Nature* **326**, 332–333.
- Clarke, G. D. P., Beiko, R. G., Ragan, M. A. & Charlebois, R. L. 2002 Inferring genome trees using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLAST scores. *J. Bacteriol.* **184**, 2072–2080.
- Dacks, J. B. & Doolittle, W. F. 2001 Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* **107**, 419–425.
- de Koning, A. P., Brinkman, F. S., Jones, S. J. & Keeling, P. J. 2000 Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**, 1769–1773.
- Denamur, E. (and 11 others) 2000 Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**, 711–721.
- Deppenemeier, U. (and 21 others) 2002 The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between Bacteria and Archaea. *J. Mol. Microbiol. Biotechnol.* **4**, 453–461.
- Dickerson, R. E. 1980 Evolution and gene transfer in purple photosynthetic bacteria. *Nature* **283**, 210–212.
- Diruggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W. & Weiss, R. B. 2000 Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol. Microbiol.* **38**, 684–693.
- 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.
- Doolittle, W. F. 1999 Phylogenetic classification and the universal tree. *Science* **282**, 2124–2129.
- Dorrell, N. (and 11 others) 2001 Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11**, 1706–1715.
- Dziejman, M., Balon, E., Boyd, D., Fraser, C. M., Heidelberg, J. F. & Mekalanos, J. J. 2002 Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl Acad. Sci. USA* **99**, 1556–1561.
- Edwards, R. A., Olsen, G. J. & Maloy, S. R. 2002 Comparative genomics of closely related salmonellae. *Trends Microbiol.* **10**, 94–99.
- Eisen, J. A. 2000 Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Gen. Dev.* **10**, 606–611.
- Emmerth, M., Goebel, W., Miller, S. I. & Hueck, C. J. 1999 Genomic subtraction identifies *Salmonella typhimurium* prophages, F-related plasmid sequences, and a novel fimbrial operon, stf, which are absent in *Salmonella typhi*. *J. Bacteriol.* **181**, 5652–5661.
- Faguy, D. M. & Doolittle, W. F. 2000 Horizontal transfer of catalase-peroxidase genes between archaea and pathogenic bacteria. *Trends Genet.* **16**, 196–197.
- Feil, E. J. (and 11 others) 2001 Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl Acad. Sci. USA* **98**, 182–187.
- Fenchel, T. & Finlay, B. J. (eds) 1995 *Ecology and evolution in anoxic worlds*. Oxford University Press.
- Field, J., Rosenthal, B. & Samuelson, J. 2000 Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.* **38**, 446–455.
- Figge, R. M., Schubert, M., Brinkmann, H. & Cerff, R. 1999 Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and inter-kingdom gene transfer. *Mol. Biol. Evol.* **16**, 429–440.
- Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. 2001 Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc. Natl Acad. Sci. USA* **98**, 8821–8826.
- Forterre, P. & Philippe, H. 1999 Where is the root of the universal tree of life? *BioEssays* **21**, 871–879.
- Fox, G. E., Peckman, K. J. & Woese, C. R. 1977 Comparative cataloging of 16S ribosomal ribonucleic acid: molecular

- approach to prokaryotic systematics. *Int. J. Syst. Bacteriol.* **27**, 44–57.
- Galagan, J. E. (and 54 others) 2002 The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**, 532–542.
- Garcia-Vallve, S., Romeau, A. & Palau, J. 2000 Horizontal transfer in bacterial and archaeal complete genomes. *Genome Res.* **10**, 1719–1725.
- Gerdes, K., Moller-Jensen, J. & Bugge Jensen, R. 2000 Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* **37**, 455–466.
- Gogarten, J. P. (and 11 others) 1989 Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665.
- Graham, D. E., Overbeek, R., Olsen, G. J. & Woese, C. R. 2000 An archaeal genomic signature. *Proc. Natl Acad. Sci. USA* **97**, 3304–3308.
- Gray, M. W. & Doolittle, W. F. 1982 Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46**, 1–42.
- Gribaldo, S., Lumia, V., Creti, R., de Macario, E. C., Sanangelantoni, A. & Cammarano, P. 1999 Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J. Bacteriol.* **18**, 434–443.
- Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W. & de Saizieu, A. 2001 Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect. Immun.* **69**, 2477–2486.
- Hansmann, S. & Martin, W. 2000 Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. *Int. J. Syst. Evol. Microbiol.* **50**, 1655–1663.
- Hayashi, T. (and 21 others) 2001 Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **8**, 11–22.
- Hense, M., Domann, E., Krusch, S., Wachholz, P., Dittmar, K. E. J., Rohde, M., Wehland, J., Chakraborty, T. & Weiss, S. 2001 Eukaryotic expression plasmid transfer from the intracellular bacterium *Listeria monocytogenes* to host cells. *Cell. Microbiol.* **3**, 599–609.
- Hentschel, U. & Hacker, J. 2001 Pathogenicity islands: the tip of the iceberg. *Microbes Infect.* **3**, 545–548.
- Herd, M. & Kocks, C. 2001 Gene fragments distinguishing an epidemic-associated strain from a virulent prototype strain of *Listeria monocytogenes* belong to a distinct functional subset of genes and partially cross-hybridize with other *Listeria* species. *Infect. Immun.* **69**, 3972–3979.
- Hooper, S. D. & Berg, O. G. 2002 Detection of genes with atypical nucleotide sequence in microbial genomes. *J. Mol. Evol.* **54**, 365–375.
- Hurek, T., Egner, T. & Reinhold-Hurek, B. 1997 Divergence in nitrogenases of *Azoarcus* spp., Proteobacteria of the beta subclass. *J. Bacteriol.* **179**, 4172–4178.
- Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K. & Nagashima, K. V. P. 2001 Horizontal transfer of the photosynthetic gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.* **52**, 333–341.
- Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S. & Peek Jr, R. M. 2001 *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl Acad. Sci. USA* **98**, 14 625–14 630.
- Iwabe, N., Kuma, K. I., Hasegawa, M., Osawa, S. & Miyata, T. 1989 Evolutionary relationships of archaeobacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359.
- Jain, R. C., Rivera, M. C. & Lake, J. A. 1999 Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806.
- John, P. & Whatley, F. R. 1975 *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254**, 495–498.
- Kato-Maeda, M., Rhee, J. T., Gingeras, T. R., Salamon, H., Drenkow, J., Smittipat, N. & Small, P. M. 2001 Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res.* **11**, 547–554.
- Katz, L. A. 1996 Transkingdom transfer of the phosphoglucose isomerase gene. *J. Mol. Evol.* **43**, 453–459.
- Ke, D., Boissinot, M., Huletsky, A., Picard, F. J., Frenette, J., Ouellette, M., Roy, P. H. & Bergeron, M. G. 2000 Evidence for horizontal gene transfer in evolution factor Tu in enterococci. *J. Bacteriol.* **182**, 6913–6920.
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L. & Das-Sarma, S. 2001 Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* **11**, 1641–1650.
- Koonin, E. V., Makarova, K. S. & Aravind, L. 2001 Horizontal gene transfer in prokaryotes: quantification and classification. *A. Rev. Microbiol.* **55**, 709–742.
- Koretke, K. K., Lupas, A. N., Warren, P. V., Rosenberg, M. & Brown, J. R. 2000 Evolution of two-component signal transduction. *Mol. Biol. Evol.* **17**, 1956–1970.
- Koski, L. B., Morton, R. A. & Golding, G. B. 2001 Codon biases and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* **18**, 404–412.
- Kurland, C. G. 2000 Something for everyone: horizontal gene transfer in evolution. *EMBO Reports* **1**, 92–95.
- Kuroda, M. (and 36 others) 2001 Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 12 225–12 240.
- Kyrpides, N. C. & Olsen, G. J. 1999 Archaeal and bacterial thermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* **15**, 298–299.
- Lan, R. & Reeves, P. R. 1996 Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* **13**, 47–55.
- Lan, R. & Reeves, P. R. 2000 Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol.* **8**, 396–401.
- Lange, B. M., Rujan, T., Martin, W. & Croteau, R. 2000 Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl Acad. Sci. USA* **97**, 13 172–13 177.
- Lawrence, J. G. & Roth, J. R. 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**, 1843–1860.
- Lawrence, J. & Ochman, H. 1997 Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**, 383–397.
- Logsdon, J. M. & Faguy, D. M. 1999 *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* **9**, R747–R751.
- McArthur, A. G. (and 14 others) 2000 The *Giardia* genome project. *FEMS Microbiol. Lett.* **189**, 271–273.
- McClelland, M. (and 25 others) 2001 Complete genome sequence of *Salmonella enterica* serovar *Thyphimurium* LT2. *Science* **413**, 852–856.
- Makarova, K. S., Ponomarev, V. A. & Koonin, E. V. 2001 Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol.* **2**, Research 0033.1–0033.14.
- Margulis, L. 1970 *Origin of eukaryotic cells*. New Haven, CT: Yale University Press.
- Margulis, L. 1981 *Symbiosis in cell evolution*. San Francisco, CA: Freeman.

- Margulis, L. 1996 Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl Acad. Sci. USA* **93**, 1071–1076.
- Martin, W. 1999 Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *BioEssays* **21**, 99–104.
- Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. 2001 An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol. Chem.* **382**, 1521–1539.
- Matte-Tailliez, O., Brochier, C., Forterre, P. & Philippe, H. 2002 Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* **19**, 631–639.
- Maynard-Smith, J. & Szathmari, E. 1995 *The major transitions in evolution*. New York: Freeman.
- Medigue, C., Rouxel, T., Vigier, P., Hénaut, A. & Danchin, A. 1991 Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* **222**, 851–856.
- Meyer, T. E., Cusanovich, M. A. & Kamen, M. D. 1986 Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proc. Natl Acad. Sci. USA* **83**, 217–220.
- Nelson, K. E. (and 28 others) 1999 Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nesbø, C. L., L'Haridon, S., Stetter, K. O. & Doolittle, W. F. 2001a Phylogenetic analyses of two 'archaeal' genes in *Thermotoga maritima* reveal multiple transfers between Archaea and Bacteria. *Mol. Biol. Evol.* **18**, 362–375.
- Nesbø, C., Boucher, Y. & Doolittle, W. F. 2001b Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. *J. Mol. Evol.* **53**, 340–350.
- Nesbø, C. L., Nelson, K. E. & Doolittle, W. F. 2002 Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. *J. Bacteriol.* **184**, 4475–4488.
- Nishida, H., Nishiyama, M., Kobashi, N., Kosuge, T., Hoshimo, T. & Yamane, H. 1999 A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Res.* **9**, 1175–1183.
- Nixon, J. E. J., Wang, A., Morrison, H. G., McArthur, A. G., Sogin, M. L., Loftus, B. J. & Samuelson, J. 2002 Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. *Euk. Cell* **1**, 181–190.
- Ochman, H. & Jones, I. B. 2000 Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J.* **19**, 6637–6643.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. 2000 Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Olendzenski, L., Liu, L., Zhaxybayeva, O., Murphey, R., Shin, D. G. & Gogarten, J. P. 2000 Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. *J. Mol. Evol.* **51**, 587–599.
- Olsen, G. J., Woese, C. R. & Overbeek, R. 1994 The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* **176**, 1–6.
- Pace, N. R. 1997 A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.
- Palmer, J. D. 1995 Rubisco rules fall; gene transfer triumphs. *BioEssays* **17**, 1005–1008.
- Parker, M. A. 2001 Case of localized recombination in 23S rRNA genes from divergent *Bradyrhizobium* lineages associated with neotropical legumes. *Appl. Environ. Microbiol.* **67**, 2076–2082.
- Perna, N. T. (and 29 others) 2001 Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533.
- Pesole, G., Gissi, C., Lanave, C. & Saccone, C. 1995 Glutamine synthetase gene evolution in bacteria. *Mol. Biol. Evol.* **12**, 189–197.
- Philippe, H., Germot, A. & Moreira, D. 2000 The new phylogeny of eukaryotes. *Curr. Opin. Gen. Dev.* **10**, 596–601.
- Qian, Q. & Keeling, P. K. 2001 Diplonemid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer. *Protist* **152**, 193–201.
- Ragan, M. A. 2001a Detection of lateral gene transfer among microbial genomes. *Curr. Opin. Gen. Dev.* **11**, 620–626.
- Ragan, M. A. 2001b On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol. Lett.* **201**, 187–191.
- Read, T. D. (and 24 others) 2000 Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**, 1397–1406.
- Reaney, D. C. 1974 Viruses and evolution. *Int. Rev. Cytol.* **37**, 212–252.
- Roger, A. J. 1999 Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146–S163.
- Roger, A. J., Sandblom, O., Doolittle, W. F. & Philippe, H. 1999 An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. *Mol. Biol. Evol.* **16**, 218–233.
- Sagan, L. 1967 On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225–274.
- Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. & Falkow, S. 2000 A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl Acad. Sci. USA* **97**, 14 668–14 773.
- Schutz, M. (and 11 others) 2000 Early evolution of cytochrome bc complexes. *J. Mol. Biol.* **300**, 663–675.
- Schwartz, R. M. & Dayhoff, M. O. 1978 Origins of prokaryotes, mitochondria and chloroplasts. *Science* **199**, 395–403.
- Selosse, M.-A., Albert, B. & Godelle, B. 2001 Reducing the genome size of organelles favours transfer to the nucleus. *Trends Ecol. Evol.* **16**, 135–142.
- Shafer, K. S., Henekamp, T., White, K. H. & Thorsness, P. E. 1999 Mechanisms of mitochondrial DNA escape to the nucleus in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **36**, 183–194.
- Sicheritz-Pontén, T. & Andersson, S. G. E. 2001 A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* **29**, 545–552.
- Smith, M. W., Feng, D. F. & Doolittle, R. F. 1992 Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem. Sci.* **17**, 489–493.
- Smoot, J. C. (and 17 others) 2002 Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proc. Natl Acad. Sci. USA* **99**, 4668–4673.
- Sonea, S. 1971 A tentative unifying view of bacteria. *Rev. Can. Biol.* **30**, 3239–3244.
- Sonea, S. & Mathieu, L. G. 2000 *Prokaryotology: a coherent view*. Montréal: Les Presses de l'Université de Montréal.
- Sonea, S. & Mathieu, L. G. 2001 Evolution of the genomic systems of prokaryotes and its momentous consequences. *Int. Microbiol.* **4**, 67–71.
- Sonea, S. & Panniset, M. 1976 Manifesto for a new bacteriology. *Rev. Can. Biol.* **35**, 103–167. [In French.]
- Stanier, R. Y. 1970 Some aspects of the biology of cells and their possible evolutionary significance. *Symp. Soc. Gen. Microbiol.* **20**, 1–38.
- Striepen, B., White, M. W., Li, C., Guerini, M. N., Malik, S. B., Logsdon Jr., J. M., Liu, C. & Abrahamsen, M. S. 2002 Genetic complementation in apicomplexan parasites. *Proc. Natl Acad. Sci. USA* **99**, 6304–6309.

- Suguri, S., Henze, K., Sanchez, L. B., Moore, D. V. & Muller, M. 2001 Archaeobacterial relationships of the phosphoenolpyruvate carboxykinase gene reveal mosaicism of *Giardia intestinalis* core metabolism. *J. Euk. Microbiol.* **48**, 493–497.
- Syvanen, M. 1994 Horizontal gene transfer: evidence and possible consequences. *A. Rev. Genet.* **28**, 237–261.
- Taylor, F. J. 1979 Symbiontism revisited: a discussion of the evolutionary impact of intracellular symbioses. *Proc. R. Soc. Lond. B* **11**, 267–286.
- Teichman, S. A. & Mitchison, G. 1999 Is there a phylogenetic signal in prokaryote proteins? *J. Mol. Evol.* **49**, 98–107.
- Tettelin, H. (and 41 others) 2000 Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815.
- Tettelin, H. (and 38 others) 2001 Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506.
- Thorsness, P. E. & Weber, E. R. 1996 Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int. Rev. Cytol.* **165**, 207–234.
- Thorsness, P. E. & Fox, T. D. 1993 Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **134**, 21–28.
- Turner, S. L. & Young, J. P. 2000 The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol. Biol. Evol.* **17**, 309–319.
- Watanabe, T. 1963 Infective heredity of multiple drug resistance in bacteria. *Bacteriol. Rev.* **27**, 87–115.
- Wernegreen, J. J. & Riley, M. A. 1999 Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* **16**, 98–113.
- Woese, C. R. 1987 Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.
- Woese, C. R. 1999 The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859.
- Woese, C. R. 2000 Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396.
- Woese, C. R., Gibson, J. & Fox, G. E. 1980 Do genealogical patterns in purple photosynthetic bacteria reflect interspecific gene transfer? *Nature* **283**, 212–214.
- Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. 2000 Aminoacyl-tRNA synthetases, the genetic code and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236.
- Wolf, Y. I., Aravind, L., Grishin, N. V. & Koonin, E. V. 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.
- Worning, P., Jensen, L. J., Nelson, K. E., Brunak, S. & Ussery, D. W. 2000 Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*. *Nucleic Acids Res.* **28**, 706–709.
- Xiong, J. & Bauer, C. E. 2002 Complex evolution of photosynthesis. *A. Rev. Plant Physiol. Plant Mol. Biol.* **53**, 503–521.
- Yang, D., Oyaizu, H., Olsen, G. J. & Woese, C. R. 1985 Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–4447.
- Yap, W. H., Zhang, Z. & Wang, Y. 1999 Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* **181**, 5201–5209.
- Zhaxybayeva, O. & Gogarten, J. P. 2002 Bootstrap, Bayesian probability and maximum likelihood mapping: exploring new tools for comparative genome analysis. *Genomics* **3**, 4.
- Zuckermandl, E. & Pauling, L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (ed. V. Bryson & H. J. Vogel), pp. 97–166. New York: Academic Press.

Discussion

S. G. E. Andersson (*Department of Molecular Evolution, Uppsala University, Uppsala, Sweden*). If I may just comment a little on your criticism of the glycolysis phylogeny, I wanted to point out that that study had a very specific aim, namely, to test the hypothesis that the glycolytic system came in with the endosymbiotic event from the α -proteobacteria, so the aim there was to see if we could see a relationship specifically with the α -proteobacteria. With the enzyme that is currently available and the genomes currently sampled, we cannot see that relationship.

W. F. Doolittle. I do understand that, and I think that I did say that in the slide. I think that some of the discussion in your paper on the phylogeny of glycolysis implies that you are dubious of lateral transfer rather more largely and I think that was not justified by those data. Yes, indeed, probably most of those genes you looked at, at least from that data, were not always very strongly bootstrapped. They do not show themselves to be α -proteobacterial, but they do, I would say, all show themselves to be bacterial in the sense that they do not show themselves not to be archaeal.

S. G. E. Andersson. No, I do not think we intended to mean that they were archaeal. We indicated it was possible they were present already in the common ancestor.

W. F. Doolittle. That is fine with me. That is consistent with my last slide, which I think there have been a lot of independent transfers in, so yes, I might be criticizing you for something you were not actually trying to say. But William Martin probably has a different take.

W. Martin (*Institute of Botany III, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany*). Yes. Among different protists, such bacterial genes often seem to come from different bacteria. The bacterial contribution was not a single massive chimeric event found in eukaryotes. Now, how do you arrive at that conclusion? You just spent 30 minutes telling us that genes are passed around through prokaryotic genomes through time.

W. F. Doolittle. OK. I know where you are going with this.

W. Martin. If you know where I am going, why do you write that? That is not true; in no way does this conclusion follow from your premises. It is not inferable by logical means. The conclusion does not follow if we embrace lateral transfer, because then, under the premise of lateral transfer, we should not expect any single bacterium to contain exactly the complement of genes. Would you please explain what that is about?

W. F. Doolittle. Well, I have no defense against that charge. What Bill is saying is that you cannot say that a gene went from this taxon to that taxon, because who knows what genes were present in the ancestors of those taxa at the inferred time of transfer? I agree with that, since I embrace lateral transfer pretty thoroughly. But to embrace it that thoroughly is to lose the ability to name the groups between which ancient transfers occurred.

W. Martin. You embrace lateral transfer except as regards the purple non-sulphur bacterium that gave rise to the mitochondrion. That statement implies that the proteobacterial symbiont that gave rise to the mitochondrion is immune to lateral transfer.

W. F. Doolittle. OK. That is right, but you have got to be careful because you are making the hydrogen hypoth-

esis basically say that these genes all came from proteobacteria, but you will never be able to prove it because proteobacteria are no longer what they were.

W. Martin. No, the same predictions, problems, apply to every model. The hydrogen hypothesis is not more immune to that than any other model.

W. F. Doolittle. Absolutely. But by that reasoning we could actually doubt the proteobacterial origin of mitochondrial genomes because modern proteobacteria were not the ones that were around at that time. I think the same argument would apply to the mitochondrial genome itself, and to the chloroplast genome for that matter, but that is indeed stretching it.

W. Martin. No, you are stretching it.

J. A. Raven (*School of Life Sciences, University of Dundee, Dundee, UK*). I think a lot of this argument about horizontality versus verticality would have been forestalled if the originator of the original tree of life, Haeckel, had had a better knowledge of trees, because they show a lot of self-grafting, particularly in the roots. If you start from a real tree you would have cross-links as well as simple branching, so better natural history originally would, perhaps, not have channelled our thinking purely in terms of verticality without horizontality.

W. F. Doolittle. It does come partly from the desire to conflate phylogeny with taxonomy, because people want a clean hierarchical classification scheme, which is that kind of tree.

J. A. Raven. Yes, I am afraid it is something that we are perhaps hard wired for.

W. F. Doolittle. Yes, I think we are.

N. Lane (*Department of Medicine, Royal Free and University College Medical School, London, UK*). I am just curious that if lateral transfer is and has been so easy, why it is that, say, oxygenic photosynthesis never made the leap

from bacteria to the Archaea, for example, or even between cyanobacteria and other types of bacteria?

W. F. Doolittle. Well, it is not that easy. We live in some a world somewhere between a state in which no transfer occurs at all, and a state in which transfer is so frequent that most of us actually were carrots in our previous lives, and we are in the middle somewhere or we would not be having any of this discussion. The most stunning bacterial genomics paper in the last several years was, in my view, Ed de Long's demonstration of proteorhodopsin in proteobacteria; based on 16S studies in the ocean, a robust and highly diversified clade of proteobacteria makes its living at least in part from photosynthesis from halorhodopsin that it has got from halobacteria, and has adapted to different light regimes. And that is almost the same thing. It is not oxygenic photosynthesis, but the more complicated the system, the more genes it involves, and the harder it will be to transfer.

F. R. Whatley (*Department of Plant Sciences, University of Oxford, Oxford, UK*). I would like just to comment on William Martin's altercation. In my greater youth I used to believe that the intermediate between a bacterium and a mitochondrion was in fact a 'Hopeful Monster', and that would, in fact, go very much against your suggestion that we have a lot of lateral transfer.

GLOSSARY

HGT: horizontal gene transfer, synonymous with LGT

HMGCoA: 3-hydroxy-3-methylglutaryl coenzyme A

LGT: lateral gene transfer

LSU: large subunit

ORF: open reading frame

RTF: resistance transfer factor

SEH: serial endosymbiosis hypothesis

SSU: small subunit