Lateral gene transfers and the evolution of eukaryotes: theories and data

Lateral gene transfers and the evolution of eukaryotes: theories and data

Laura A. Katz

Vertical transmission of heritable material, a cornerstone of the Darwinian theory of evolution, is inadequate to describe the evolution of eukaryotes, particularly microbial eukaryotes. This is because eukaryotic cells and eukaryotic genomes are chimeric, having evolved through a combination of vertical (parent to offspring) and lateral (trans-species) transmission. Observations on widespread chimerism in eukaryotes have led to new and revised hypothesis for the origin and diversification of eukaryotes that provide specific predictions on the tempo (early vs continuous transfers) and mode (nature of donor and recipient lineages) of lateral gene transfers (LGTs). Analyses of available data indicate that LGTs in eukaryotes largely fall into two categories: (1) LGTs from organelles to the nucleus, only a few of which appear to have occurred at the time of the origin of eukaryotes, and (2) anomalous LGTs involving diverse donor and recipient lineages. Further testing of hypotheses on the origin and diversification of eukaryotes will require complete genome sequences from a number of diverse eukaryotes and prokaryotes combined with sequences of targeted genes from a broad phylogenetic sample.

Keywords: horizontal gene transfer, origin of eukaryotes, gene genealogies, chimerism

INTRODUCTION

Although long a part of models of prokaryotic evolution, the impact of lateral gene transfer (LGT) on eukaryotes has only recently been appreciated. Evidence first from analyses of multiple genes, and now genomes, indicates that eukaryotic genomes are chimeric with respect to archaea and bacteria. Data on this chimerism, defined here as having descended through a combination of vertical and lateral transmission, combined with recent evidence on the timing of the acquisition of mitochondria and increasing numbers of examples of LGTs, invoked revision of hypotheses on eukaryotic origins (Table 1). Each of these hypotheses generates testable predictions on the nature of donor and host lineages and/or the timing of LGTs in eukaryotes. The aim of this paper is to synthesize current data on the contributions of LGTs to the chimeric eukaryotic genome and to evaluate the support for hypotheses of eukaryotic evolution.

Identifying LGTs

There are numerous ways to detect LGTs, including (1) the discovery of discordant gene genealogies, (2) aberrant patterns of codon/compositional bias across genomes, and (3) unusual patterns of the presence/absence of genes within genomes. The first method provides the most direct evidence of LGT as potential donor and recipient lineages can be identified by comparing discordant topologies. Genealogies in which relationships among sequences contradict species relationships inferred from analyses of morphology and/or other molecular markers indicate that LGTs have occurred. Perhaps the most striking of such transfers are trans-domain transfers in which a small clade of sequences from one domain fall within sequences from another domain: e.g. when a limited
Table 1. Predictions from a subset of existing hypotheses for the origin of eukaryotes

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Reference</th>
<th>Donor</th>
<th>Recipient</th>
<th>Timing*</th>
<th>Data†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion</td>
<td>Zillig (1991); Zillig et al. (1989)</td>
<td>An archaeon and a bacterium</td>
<td>Archaeon</td>
<td>Early+</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>e.g. Lang et al. (1999); Roger (1999)</td>
<td>z-Proteobacterium</td>
<td>Early+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contemporaneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen</td>
<td>Martin &amp; Müller (1998)</td>
<td>z-Proteobacterium</td>
<td>Methanogenic archaeon</td>
<td>Early+</td>
<td>?</td>
</tr>
<tr>
<td>Syntrophy</td>
<td>Moreira &amp; Lopez-Garcia (1998)</td>
<td>δ- and z-Proteobacteria</td>
<td>Euryarchaeota</td>
<td>Early+</td>
<td>?</td>
</tr>
<tr>
<td>SET‡</td>
<td>Margulis (1996); Margulis et al. (2000)</td>
<td>z-Proteobacterium, spirochaete</td>
<td>Archaeon</td>
<td>Early+</td>
<td>?</td>
</tr>
<tr>
<td>Genetic annealing</td>
<td>Woese (1998)</td>
<td>Predominantly early lineages</td>
<td>Early only</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>You are what you eat</td>
<td>Doolittle (1998)</td>
<td>Many</td>
<td>Many</td>
<td>Continuous</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* ‘Timing’ refers to the tempo of LGTs. Some theories focus on LGTs at the time of the origin of eukaryotes, but do not preclude later transfers (early+).
† ‘Data’ refers to the support of a substantial number of gene genealogies for a particular model.
‡ Serial endosymbiosis theory.

number of eukaryotic sequences fall within a well-supported clade of prokaryotic sequences to the exclusion of other eukaryotic sequences.

Analyses of genome characteristics such as codon usage, compositional bias and the presence/absence of genes also have been successfully used to identify recent LGT events, particularly among prokaryotes (Eisen, 2000; Garcia-Vallvé et al., 2000a; Lawrence, 1999; Lawrence & Ochman, 1998). However, over time, codon/compositional signatures of LGTs decay due to the mode of the host genome evolution. Moreover, it is difficult to discern the nature of the donor lineage from this type of data.

The caveats

There are numerous potential pitfalls in attempts to identify LGT events. Particularly worrisome for understanding LGTs in eukaryotes is taxon sampling, as data on the nature of eukaryotic genes and genomes continue to be dominated by a relatively limited phylogenetic sample of eukaryotes – plant and animal sequences are particularly abundant while potentially early diverging eukaryotic sequences are still relatively underrepresented. Taxon sampling is known to have a considerable impact on the topology of genealogies through analyses of simulated genealogies and combined datasets (Graybeal, 1998; Poe, 1998a, b; Swoford et al., 1996; Yoder & Irwin, 1999). Comparisons of complete genomes from phylogenetically diverse eukaryotes, combined with broadly sampled potential donor lineages among prokaryotes, will be essential for clarifying the nature and extent of LGTs.

Equally troubling is the fact that LGTs among prokaryotes will obscure our interpretation of the origin of eukaryotic genes; LGTs that occur among prokaryotes after a gene is transferred from one of these prokaryotic (or organelle) lineages into a eukaryotic genome will prevent us from clearly interpreting the donor lineages for that gene (e.g. Lange et al., 2000; Martin, 1999). For example, if a mitochondrial gene was transferred to the nucleus around the time of the origin of eukaryotes, and subsequently a z-proteobacterial homologue was replaced by a LGT from a different type of bacterium, then it will be difficult to reconstruct the ancestry of the gene in the eukaryotic genome. Finally, the undetected presence of paralogues and unequal rates of evolution between lineages also poses challenges in using genealogies to interpret the history of LGTs.

EMERGING DATA ON THE EVOLUTION OF EUKARYOTES

The past decade has seen significant revision in our understanding of the evolution of eukaryotic genomes and the timing of the acquisition of mitochondria.

The nature of eukaryotic genomes

Eukaryotic genomes are chimeric in that they contain genes from multiple donor lineages. In contrast to earlier notions that eukaryotes evolved through simple descent with modification from a single prokaryotic lineage, analyses of individual genes and complete genomes indicate that some eukaryotic genes are of bacterial origin and others are of archaeal origin (Brown & Doolittle, 1997; Brown et al., 1994; Doolittle, 1999; Golding & Gupta, 1995; Gupta & Golding, 1996; Lange et al., 2000; Ribeiro & Golding, 1998; Roger & Brown, 1996). Genealogical analyses of many genes, particularly analyses of genes involved in the informational systems of eukaryotes (DNA transcription, translation, repair), tend to unite eukaryotes with archaea (Best & Olsen, 2001; Brown & Doolittle,
Timing of the acquisition of mitochondria

A second development that has radically revised theories for the evolution of eukaryotes is the demise of the ‘archezoa’, defined as primitively amitochondriate eukaryotes (Cavalier-Smith, 1983). The ‘archezoa’ were seen as early lineages in a two-step process in the evolution of eukaryotes: first there was the evolution of the nucleus and cytoskeleton in the ‘archezoa’ and only later, the acquisition of mitochondria. However, the taxa included in the ‘archezoa’ have changed considerably over time as the putative early position of some archezoans has been rejected (Clark & Roger, 1995; Germot et al., 1997; Hirt et al., 1997; Horner et al., 1996; Keeling, 1998; Patterson, 1999; Roger et al., 1998). For example, although initial analysis of the unusually small SSU-rRNA of microsporidians suggested that this lineage was an early diverging eukaryote, subsequent analyses of protein-encoding genes indicate that these organisms are instead highly modified fungi (Hirt, 1999)! Further, the definition of this group has been questioned by the discovery of genes of possible mitochondrial origin in the putative archezoans Entamoeba, parabasalids, microsporidia and diplomonads (Bui et al., 1996; Germot et al., 1996; Hashimoto et al., 1998; Roger, 1999; Roger & Brown, 1996; Roger et al., 1998). Together, these data suggest the possibility that the acquisition of mitochondria occurred at the time of, or soon after, the origin of eukaryotes (e.g. Gray, 1999; Katz, 1998; Lang et al., 1999; Roger, 1999).

HYPOTHESES AND PREDICTIONS

Data on the chimeric nature of eukaryotic genomes and the timing of the acquisition of mitochondria have inspired numerous hypotheses, and revisions to previous hypotheses, on the origin of eukaryotes (Table 1). For example, the simple ‘fusion’ theory of Zillig (Zillig, 1991; Zillig et al., 1989) explains the apparent chimerism in the biochemistry of eukaryotes as being the result of fusion between an archaeon and a bacterium. Hence, if this theory is true and more recent LGTs have been relatively rare, genes within the eukaryotic nucleus should trace back to one archaeal lineage and one bacterial lineage.

The presence of genes of potentially mitochondrial origin in the nucleus of some ‘archezoa’ suggests that the acquisition of mitochondria occurred at the time of the origin of eukaryotes (Lang et al., 1999; Roger, 1999) and the chimerism within the eukaryotic genome was at least in part due to the transfer of genes from the mitochondria to the nucleus (‘contemporaneous’ hypothesis, Table 1). This theory suggests that a subset of genes in the eukaryotic nucleus, including early diverging eukaryotes, should trace back to the lineage that gave rise to mitochondria.

In its most recent version, the serial endosymbiosis theory (SET) argues that three lineages contributed to the origin of eukaryotes (Margulis, 1996; Margulis et al., 2000). Under this theory, symbioses between two bacteria, a methanogenic archaeon, and a spirochaete-like bacterium gave rise to eukaryotes with a cytoskeleton and nucleus, specifically a karyomastigont system. Later, mitochondria were derived through the acquisition of a z-proteobacterial symbiote, with the symbiosis driven by respiration (Margulis, 1996; Margulis et al., 2000). This theory predicts three specific donor lineages to eukaryotic genomes: an archaeon, a spirochaete and a z-proteobacterium.

Combining observations on the genetic makeup of eukaryotic nuclei, the biochemistry of hydrogenosomes (hydrogen-producing organelles found in some potentially early diverging lineages), and the nature of extant syntrophic relationships (symbiosis based on metabolic interactions), Martin & Müller (1998) proposed ‘the hydrogen hypothesis’ for the origin of eukaryotes. Under this hypothesis, in which eukaryotes arose through a symbiosis between hydrogen producing z-proteobacterial symbiote and a methaneproducing archaeon, we expect that at least a subset of genes in the eukaryotic nucleus should trace back to these two lineages. Similarly, the ‘syntrophy hypothesis’ of Moreira & Lopez-Garcia (1998) argues that eukaryotes arose through a symbiotic relationship between a methanogenic euryarchaeota and a δ-proteobacterium such as a sulfate-reducing myxobacterium. Under this theory, the acquisition of mitochondria occurred as an independent symbiosis, indicating three possible lineages as the source of the chimerism of eukaryotic nuclei: methanogenic archaea, δ-proteobacterium and z-proteobacterium.

Two additional hypotheses that account for the chimeric nature of the eukaryotic nucleus do not rely on a simple set of donor and recipient lineages and instead emphasize specifics on the tempo of LGTs. Under ‘genetic annealing’, LGTs are predicted to have had the greatest impact early in the divergences of the domains of life when lineages in the three domains of life are imagined to be ‘leaky’ (Woese, 1998). Later, when the integrity of the lineages increased, the amount of LGT decreased. Finally, there is the ‘you are what you eat’ theory of Doolittle (1998), in which the mosaic nature of eukaryotic genomes is the result of a continuous transfer of genes into the nuclei of microbial eukaryotes from organelles and ingested prey items. Under this theory, eukaryotic genes trace back to multiple donor lineages.

LGTs in eukaryotes

As described above, different hypotheses on eukaryotic origins vary in their predictions on the tempo and mode of LGTs (Table 1). Two of these hypotheses...
Several hypotheses for the origin of eukaryotes, including the hydrogen hypothesis, the syntrophy hypothesis, serial endosymbiosis theory (SET) and the ‘contemporaneous’ hypothesis, invoke a transfer of genes from the symbiont that gave rise to mitochondria (and, in some models, hydrogenosomes) to the nucleus. While there are numerous genes that function in mitochondria of plants, animals and fungi that are of z-proteobacterial origin (e.g. Kurland & Andersson, 2000), the current phylogenetic sampling of many of these genes is limited and it is impossible to test whether these genes are the result of early or late LGTs. Moreover, there is evidence of LGTs from diverse bacterial lineages into the mitochondrial proteome of animals and fungi (e.g. Kurland & Andersson, 2000; Wolf & Koonin, 2001). In only a few cases, including those described below, do genealogies place sequences from a broad sampling of eukaryotes as sister to z-proteobacteria. Such a broad distribution is essential if we are to explain the widespread chimerism observed in eukaryotes as being the result of early transfers from the symbiont that gave rise to mitochondria.

Perhaps the best supported instance is for chaperonin 60 (cpn60), where genealogical analyses of diverse eukaryotic CPN60, including sequences from potentially early diverging amitochondriate eukaryotes Giardia, Spironucleus, Entamoeba and Trichomonas, provide strong support for this clade falling within z-proteobacterial CPN60 sequences and sister to a clade containing the Rickettsia sequence (Roger et al., 1999; Horner & Embley, 2001). Genealogical analyses of heat-shock protein 70 (HSP70), which is also encoded in the genome of some amitochondriate eukaryotes, also place eukaryotic sequences within a clade of z-proteobacterial sequences but with limited bootstrap support for relationships within this clade (Morrison et al., 2001). Similarly, analyses of triosephosphate isomerase sequences from 22 diverse eukaryotes, including several microbial eukaryotes, 18 bacteria and one archaeon, also place the eukaryotic sequences as sister to a clade containing extant relatives of mitochondria [e.g. Rhizobium etli; Keeling & Doolittle (1997) – although unpublished data described in Kurland & Andersson (2000) challenge this topology]. Such topologies are consistent with the idea that eukaryotes are chimeric in part because of the impact of LGT from mitochondria to nucleus soon after the origin of eukaryotes.

**LGTs I: from organelles to the nucleus**

The first type of LGT involves transfers of genes from organelles into the nucleus. Clearly, the acquisition of an organelle through endosymbiosis provides a new potential source of material to be laterally transferred and it may be that following the acquisition of mitochondria, there was a burst of transfer of genes from organelle to nucleus. Further, we know that genes can be transferred from mitochondria to nucleus through analyses of both comparative and experimental data (e.g. Adams et al., 1999; Gray, 1999; Henze & Martin, 2001; Martin & Herrmann, 1989; Thorsness & Fox, 1990).

**Mitochondria.** Several hypotheses for the origin of eukaryotes, including the hydrogen hypothesis, the syntrophy hypothesis, serial endosymbiosis theory (SET) and the ‘contemporaneous’ hypothesis, invoke a transfer of genes from the symbiont that gave rise to mitochondria (and, in some models, hydrogenosomes) to the nucleus. While there are numerous genes that function in mitochondria of plants, animals and fungi that are of z-proteobacterial origin (e.g. Kurland & Andersson, 2000), the current phylogenetic sampling of many of these genes is limited and it is impossible to test whether these genes are the result of early or late LGTs. Moreover, there is evidence of LGTs from
Table 2. Sample of potential anomalous LGTs in eukaryotes

<table>
<thead>
<tr>
<th>Gene</th>
<th>‘Host’</th>
<th>Suspected donor</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMG-CoA class 2 reductase</td>
<td>Giardia</td>
<td>Bacteria</td>
<td>Boucher &amp; Doolittle (2000)</td>
</tr>
<tr>
<td>Isoprenoid biosynthesis (DXP)</td>
<td>Plastid-containing eukaryotes</td>
<td>Potentially diverse bacteria</td>
<td>Lange et al. (2000)</td>
</tr>
<tr>
<td>Iron hydrogenase</td>
<td>Nyctotherus</td>
<td>Desulfovibrio (δ-proteobacterium)</td>
<td>Horner et al. (2000)</td>
</tr>
<tr>
<td>Class II fumarase</td>
<td>Trichomonads</td>
<td>Bacteria</td>
<td>Gerbod et al. (2001)</td>
</tr>
<tr>
<td>Many possible</td>
<td>Fungi</td>
<td>Diverse bacteria</td>
<td>Rosewich &amp; Kistler (2000)</td>
</tr>
<tr>
<td>Glycosyl hydrolases</td>
<td>Chytrids</td>
<td>Rumen bacteria</td>
<td>Garcia-Vallvé et al. (2000b)</td>
</tr>
<tr>
<td>Catalases</td>
<td>Aspergillus</td>
<td>Gram +ve bacteria, proteobacteria</td>
<td>de Koning et al. (2000)</td>
</tr>
<tr>
<td>N-Acetylneuraminate lyase</td>
<td>Trichomonads</td>
<td>Pasteurellaeae (γ-proteobacteria)</td>
<td>Salzberg et al. (2001)</td>
</tr>
<tr>
<td>Many possible (~ 40)</td>
<td>Human (vertebrate) genomes</td>
<td>Potentially diverse bacteria</td>
<td>Andersson et al. (2001)</td>
</tr>
<tr>
<td>N-Acetylneuraminate lyase</td>
<td>Human (vertebrate) genomes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Many analyses include very limited taxonomic sampling and estimates of genealogical relationships are likely to change as additional taxa are added. See text for further details.

Genealogical analyses that include a broad sample of potential donor and recipient lineages. Support for the anomalous transfers described below and in Table 2 varies depending on the sampling and methodologies employed in the studies. It is likely that reanalyses of individual genealogies with additional sequences and new methodologies will likely alter some, but not all, of the anomalous transfers discussed.

Isoprenoid biosynthesis genes. Eukaryotes possess two major pathways for the synthesis of isoprenoids: the melonavate (MVA) pathway, also common in archaea, and the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway, which is characteristic of many bacteria and photosynthetic eukaryotes (Boucher & Doolittle, 2000; Lange et al., 2000). Genealogical analyses of one MVA pathway gene, HMG-CoA reductase (HMGR), indicate that eukaryotic sequences fall in two distinct clades: HMGR sequences from most eukaryotes, including several microbial eukaryotes, are sister to a predominantly archaeal clade while the HMGR sequence from Giardia lamblia falls within a predominantly bacterial clade (Boucher & Doolittle, 2000). This polyphyletic distribution of eukaryotic HMGR sequences is consistent with an anomalous transfer from bacteria into the ancestor of the diplomonad Giardia lamblia (Boucher & Doolittle, 2000).

The phylogenetic distribution of the DXP pathway suggests that photosynthetic eukaryotes acquired genes for the DXP pathway by transfer from plastid genomes. Yet, of the nine genes in these pathways for which genealogies were constructed, only one genealogy, for 1-deoxyxylulose-5-phosphate reductoisomerase (DXR), shows strong support for relationships between homologues in higher plants and cyanobacteria – consistent with a plastid origin – but even here, the sequence in the plastid-containing eukaryote Plasmodiunm clusters with a sequence from hyperthermophilic bacterium Aquifex (Lange et al., 2000). Another genealogy, for cytosolic acetyl-CoA acyltransferase (AACT) unites a subset of eukaryotic sequences with homologues from an α-proteobacterium – consistent with a mitochondrial origin of this enzyme (Lange et al., 2000). Relationships among taxa in the other seven genealogies are even more complex, indicating multiple LGTs. For example, the 1-deoxyxylulose-5-phosphate synthase (DXPS) gene of higher plants and the flagellate Chlamydomonas cluster with a sequences from Rhodobacter capsulatus while the sequence from the apicomplexan Plasmodium is sister to a clade containing enteric bacterial sequences, albeit without strong bootstrap support (Lange et al., 2000). The authors interpret the discordance among gene genealogies as evidence of LGT among prokaryotes that occurred after the acquisition of plastids (Lange et al., 2000). However, it is also possible that multiple independent LGTs from prokaryotes to eukaryotes explain part of the discordance, particularly for the DXP pathway genes from photosynthetic eukaryotes that appear polyphyletic.

Metabolism in anaerobic eukaryotes. Metabolism within hydrogenosomes, hydrogen-producing organelles found in chytrids, ciliates and parabasalids, requires biochemical pathways and enzymes that are uncommon in eukaryotes. The genealogies of two such enzymes, ferredoxin oxidoreductase (PFO) and iron hydrogenases, show contrasting patterns. PFO genes sampled from diverse anaerobic eukaryotes form a monophyletic group relative to prokaryotic sequences, but fail to show a strong relationship to any particular prokaryotic lineage (Horner et al., 1999). These data suggest a common and ancient origin of this enzyme in eukaryotes, but do not resolve the potential ‘donor’ lineage of this enzyme. In contrast, genealogical analyses of iron hydrogenase indicates that there are two separate origins of this gene in eukaryotes (Horner et al., 2000): sequences from Entamoeba, the diplomonad Spironucleus and the parabasalid Trichomonas form a monophyletic group in some analyses (albeit with poor bootstrap support), while the hydrogenase from the ciliate Nyctotherus appears to be an independent acquisition from bacteria.
and weakly clusters with δ-proteobacterium Desulfovibrio. Although support for specific ‘donor’ lineages is lacking in these analyses, Horner et al. (2000) used parametric bootstrap and hierarchical ratio tests to demonstrate that the eukaryotic Fe-hydrogenases do not form a monophyletic group (Horner et al., 2000). These analyses are consistent with the idea that at least one clade of eukaryotic iron hydrogenases is derived from an anomalous LGT event.

**Class II fumarases.** Genealogical analyses of class II fumarases indicate that sequences from several trichomonads have an independent origin relative to other eukaryotic class II fumarases (Gerbod et al., 2001). While sequences from plants, animals and fungi cluster with α-proteobacterial sequences, class II fumarase from trichomonads fall within different clades of bacteria (Gerbod et al., 2001). The nature of the donor lineage cannot be resolved from these genealogies as there is low bootstrap support for many nodes. Moreover, as with many studies, the limited sampling of other eukaryotic sequences, particularly microbial eukaryotes, makes it difficult to interpret the meaning of the reported topologies.

**LGTs in fungi.** The role of LGTs in fungi has been reviewed by Rosewich & Kistler (2000), where numerous potential trans-domain LGTs are described. In some cases, e.g. xylanases, endoglucanase B and β-glucanase, LGTs are inferred from data on the presence/absence of genes and levels of sequence divergence rather than genealogies, and further work is needed to confirm these putative transfers. Genealogical analyses of glycosyl hydrolases reveal two clades of chytrid sequences that cluster within bacterial glycosylases and distinct from the clade containing all other eukaryotic sequences (Garcia-Vallvé et al., 2000b). Another example of discordant gene genealogies involving fungi is that of analyses of diverse sequences of large-subunit catalases, which unite some fungal sequences with bacterial sequences from Gram-positive bacteria and proteobacteria (Klotz et al., 1997; Rosewich & Kistler, 2000). In their review, Rosewich and Kistler (2000) argue that the life history of many fungi, including their often close association with other organisms and saprophytic mode of feeding, make fungi particularly susceptible to LGTs. Additional data are required to test this correlation.

**Parasitism/defence genes.** One striking example of an anomalous LGT in eukaryotes is the transfer of N-acetylneuraminic lyase from a γ-proteobacterium in the Pasteurellaceae (parasites of vertebrate epithelia) to the parasite Trichomonas vaginalis (de Koning et al., 2000). Genealogical analysis of the N-acetylneuraminic lyase gene from Trichomonas put this sequence within a clade of bacterial sequences, sister to sequences from Pasteurellaceae (de Koning et al., 2000). These data suggest a relatively recent transfer of this enzyme may have been involved in the adaptation of Trichomonas to its parasitic lifestyle (de Koning et al., 2000).

**LGTs and the human genome.** Initial reports of sequence of the human genome suggested that between 113 and 223 genes in the human genome had been transferred from bacteria to a vertebrate ancestor (Lander et al., 2001). Reanalysis of these genes demonstrate that a number of these putative transfers are misidentified due to differential gene loss, taxon sampling and unequal evolutionary rates among lineages (Andersson et al., 2001; Salzberg et al., 2001; Stanhope et al., 2001). Excluding these factors, comparisons of similarity scores from BLAST searches, plus a limited number of genealogical analyses, reduced the number of potential LGTs to ~ 40 genes (Salzberg et al., 2001). However, even this number should be interpreted with caution as well-explored genealogical analyses, including sequences from broadly sampled lineages, are necessary to confirm cases of LGTs. For example, only one of seven genealogies constructed from candidate genes identified by Salzberg et al. (2001) is consistent with a possible bacterium to vertebrate (or vertebrate to bacterium) transfer; the vertebrate copy of N-acetylneuraminic lyase clusters in a clade with Vibrio cholerae and Yersinia pestis (Andersson et al., 2001).

**LGTs in ecological settings.** Many micro-organisms contain mechanisms that allow them to exchange genes. Such mediated transfers are best known between prokaryotes, where conjugation, transduction and transformation may all allow the uptake of exogenous DNA. Several analogous mediated transfer systems do exist, allowing the transfer of DNA from prokaryotes to eukaryotes. In fact, Agrobacterium is routinely used to transform plants (e.g. Chateau et al., 2000; De Block, 1994). LGTs have also been observed between plants and fungi and plants and micro-organisms in ‘common garden’ experiments (Bertolli & Simonet, 1999; Hoffmann et al., 1994; Paget et al., 1998). Finally, yeast have been shown to be transformed by plasmid DNA under ‘non-artificial’ starvation conditions (Nevoigt et al., 2000). These examples of contemporary LGTs are anomalous and indicate possible mechanisms for LGTs that have occurred over evolutionary time scales.

**CONCLUSIONS**

Hypotheses for the origin of eukaryotes that invoke symbioses between limited numbers of lineages are not supported by current data on LGTs. For example, the fusion model, contemporaneous hypothesis, hydrogen hypothesis, syntrophy hypothesis and SET (Table 1) indicate two or three prokaryotic lineages that contributed to the chimerism of eukaryotic genomes. Significant to testing these hypotheses, current data do not provide any strong signal for particular donor lineages involved in the origin of eukaryotic cells. Moreover, only a few sequences, including CPN60, HSP70 and TPI, provide varying degrees of support for early transfer of genes from mitochondria to the nucleus. It is unclear whether the lack of support for symbiosis-based hypotheses is due to inadequate sampling of
genes and taxa, the obscuring effects of more recent LGTs or the inappropriateness of the proposed hypotheses.

The one hypothesis that is not consistent with current data on LGTs is ‘genetic annealing’ (Woese, 1998), as anomalous transfers are not restricted to early diverging lineages (Table 2). Instead, the data are consistent with the continuous tempo of the ‘you are what you eat’ hypothesis (Doolittle, 1998) as the LGTs described above include both potentially early diverging eukaryotes (e.g. Trichomonas and Giardia) and more recently derived lineages (e.g. vertebrates and chytrids).

Although currently we cannot distinguish among hypotheses on eukaryotic evolution from comparisons of multiple gene genealogies, the growing numbers of anomalous LGTs necessitate revisions of ideas on eukaryotic evolution. These discordant genealogies suggest that diversification of living organisms has been ‘web’- or ‘net’-like (e.g. Doolittle, 1999; Hilario & Gogarten, 1993; Katz, 1998, 1999; Martin, 1999). Hence, we must abandon, or at least substantially redefine, the notion of ‘The Tree of Life’ in favour of more realistic models of diversification of eukaryotes, as we can no longer assume that vertical transmission has been the rule throughout evolution.

ACKNOWLEDGEMENTS

I am grateful to Bill Martin for discussion of an earlier version of the manuscript as well as for the comments of two anonymous reviewers. This work is supported by an NSF Small Grant in Exploratory Research (NSF #0079325).

REFERENCES


provides further evidence for secondary loss of mitochondria among Chaperonin 60 phylogeny. Current Genetics culture with transgenic higher-plants.


