

# Horizontal gene transfer in eukaryotic evolution

Patrick J. Keeling\* and Jeffrey D. Palmer†

**Abstract** | Horizontal gene transfer (HGT; also known as lateral gene transfer) has had an important role in eukaryotic genome evolution, but its importance is often overshadowed by the greater prevalence and our more advanced understanding of gene transfer in prokaryotes. Recurrent endosymbioses and the generally poor sampling of most nuclear genes from diverse lineages have also complicated the search for transferred genes. Nevertheless, the number of well-supported cases of transfer from both prokaryotes and eukaryotes, many with significant functional implications, is now expanding rapidly. Major recent trends include the important role of HGT in adaptation to certain specialized niches and the highly variable impact of HGT in different lineages.

Horizontal gene transfer (HGT), also known as lateral gene transfer, refers to the movement of genetic information across normal mating barriers, between more or less distantly related organisms, and thus stands in distinction to the standard vertical transmission of genes from parent to offspring.

Extant eukaryotes arose by HGT on the grandest scale possible — the endosymbiosis and subsequent genetic integration of entire organisms that gave rise to the mitochondrion and the plastid. However, the importance of subsequent, ongoing HGT in eukaryotes has been under-appreciated since the outset of the genomics era for a variety of reasons. First, the early recognition of the importance of HGT in bacterial evolution, and continuing debate about exactly how profound this has been (see REFS 1–6 for examples), have overshadowed the role of HGT in eukaryotes. Second, most of the earliest and even the most current evidence for HGT in eukaryotes comes from protists, whereas most biologists study animals (or at least plants or fungi) and therefore often overlook, or dismiss as insignificantly aberrant, findings from such obscure creatures. Lastly, the erroneous report of substantial HGT in the human genome<sup>7–9</sup> has probably chilled the field, tarnished more credible claims for eukaryotic HGT, and possibly inhibited scientists from looking for HGT in the many subsequently sequenced animal genomes, and perhaps even in fungal and plant genomes.

Nevertheless, HGT has altered our concepts of the genome, the species and the tree (or web) of life in the bacterial world, and now cases of HGT in eukaryotes are emerging at an increasing rate and account for

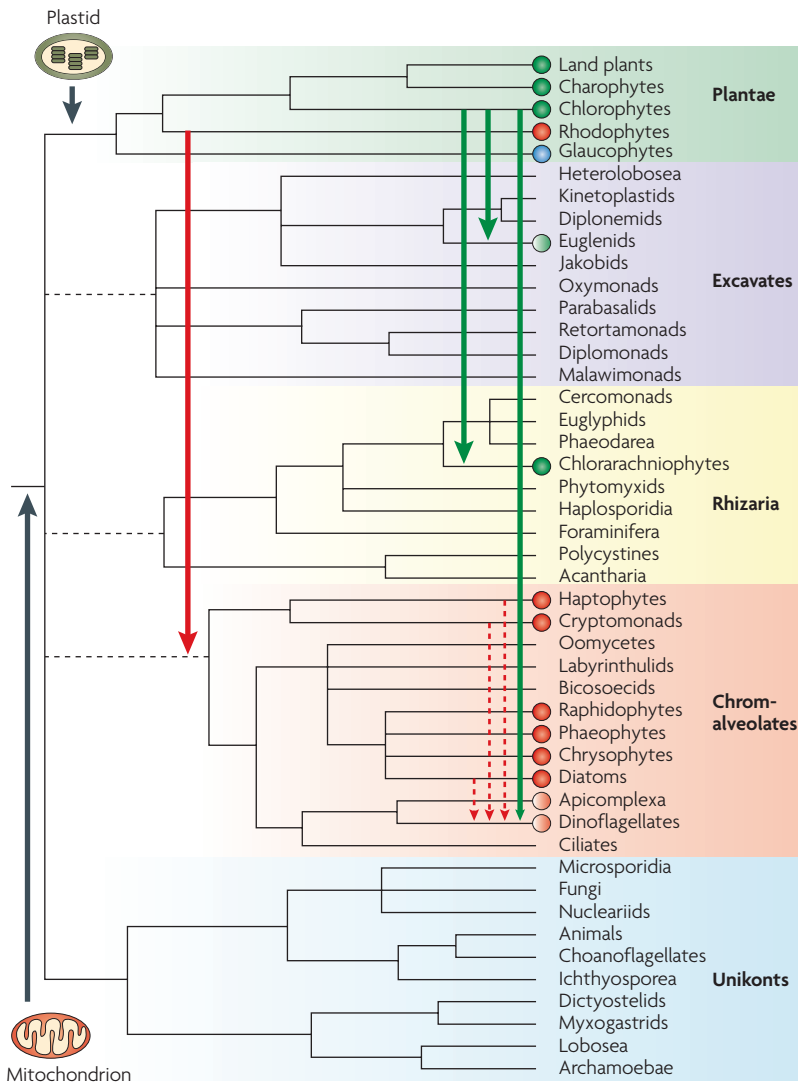
many adaptively important traits. Indeed, the number of cases of HGT involving eukaryotes is now so great that it no longer makes sense to attempt to comprehensively compile individual cases, as has been done in past reviews<sup>10–12</sup>. Instead, we will use selected examples to highlight emerging trends and questions.

## Detecting and evaluating cases of HGT

Several approaches can lead to the suspicion that HGT has occurred, but so-called ‘surrogate’ methods<sup>13</sup> to identify potential cases of HGT are insufficient unless followed up by formal phylogenetic analysis. The widely publicized but overstated claim in 2001 that the draft human genome contained about one hundred genes acquired from bacteria at the onset of or during vertebrate evolution<sup>7</sup> illustrates this problem: top BLAST hits were used to identify human genes, the closest ‘relatives’ of which were bacterial homologues, but subsequent phylogenetic analysis often negated this relationship<sup>9</sup>. HGT was also inferred (along with the alternative possibility of massive gene loss) for cases in which human genes had homologues in bacteria but not in the then four other ‘completely’ sequenced eukaryotic genomes<sup>7</sup>. These cases were undermined by finding closer homologues in incompletely sequenced eukaryotic genomes, supporting gene loss<sup>8,14</sup>.

The gold standard for identifying HGT with confidence is phylogenetic incongruence — this occurs if there is strong conflict between the phylogenies of the gene and of the organism. Even then, several potentially confounding factors must still be taken into account. Although our understanding of eukaryotic phylogeny

\*Department of Botany,  
University of British  
Columbia, Vancouver, British  
Columbia V6T 1Z, Canada.  
†Department of Biology,  
Indiana University,  
Bloomington,  
Indiana 47405, USA.  
Correspondence to P.J.K.  
e-mail:  
[pkeeling@interchange.ubc.ca](mailto:pkeeling@interchange.ubc.ca)  
doi:10.1038/nrg2386  
Published online 1 July 2008



**Figure 1 | The major bifurcations and reticulations in eukaryotic evolution.** A simplified tree of eukaryotes is shown, consisting of five putative 'supergroups'. Some relationships are highly uncertain, and these are shown as polytomies, whereas other relationships are supported by some evidence but need further testing, and these are shown as dashed lines (summarized in REF. 158). In particular, there are recent and important indications that Rhizaria fall within Chromalveolates<sup>103,159</sup>, which has interesting implications for plastid evolution if proven to be true. Major endosymbiotic events that led to the origin and spread of mitochondria and plastids are shown as vertical arrows. Black solid arrows indicate the primary endosymbioses at the origin of mitochondria and plastids. Red and green arrows indicate the movement of red or green plastids, respectively, by secondary (solid arrows) and tertiary (dashed arrows) endosymbioses. Lineages in which a plastid organelle is presently known to exist are indicated by coloured circles beside the name (red or green for red algae or green algae and all the plastids derived from them, and blue for glaucophytes to emphasize their independence from red and green algae). Lineages with a substantial mix of photosynthetic and non-photosynthetic species are partially filled. The dinoflagellate plastid is shown as red for simplicity, because only a single small lineage contains the secondary green plastid: the majority of dinoflagellate plastids are derived from red algae. Accurate knowledge of the tree of eukaryotes is essential for studying horizontal gene transfer (HGT) because, as described in the main text, clear-cut and well-supported deviation from this organismal tree in a given gene tree provides the best evidence for HGT. Similarly, our knowledge of endosymbiotic events allows us to recognize where to expect to find the genes that are derived from such events. Cases like the tertiary endosymbionts in dinoflagellates offer complex situations that are useful models to study the effects of symbiosis on gene movement.

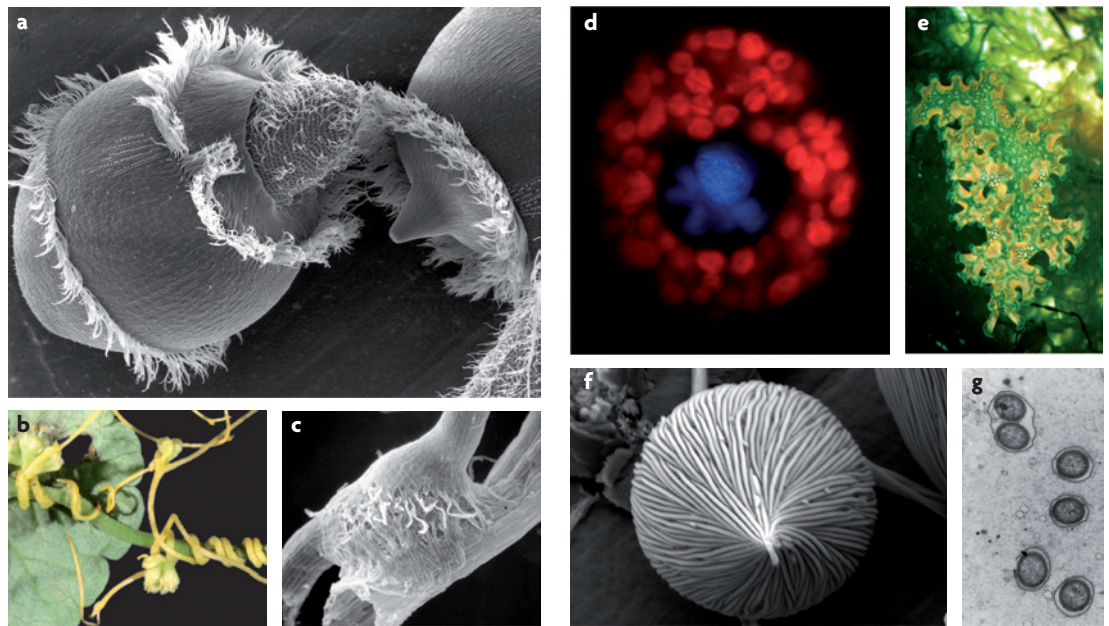
is rapidly improving (FIG. 1), to confidently detect phylogenetic incongruence requires a level of taxonomic sampling for the gene in question that is often lacking at present, and therefore some HGTs will be overlooked. In some cases, HGT is difficult to distinguish from ancestral gene duplication and differential gene loss<sup>15-17</sup>. In other cases, unusual trees are simply misleading because they are artefacts of phylogenetic reconstruction owing to an unusual mode or rate of evolution (which can yield strongly supported but incorrect trees). Nonetheless, there are a large and growing number of cases at the other end of the spectrum, in which well-supported and well-sampled phylogenies demonstrate a glaring conflict with organismal relationships that are also well-supported. This suggests that some proportion of the more poorly supported trees is probably the result of HGT.

**Endosymbiosis and massive HGT**

A profound difference between prokaryotes and eukaryotes is the importance of endosymbiosis. The origin of the cytoskeleton and endocytosis enabled eukaryotes to readily engulf and feed on other cells<sup>18</sup>, which are occasionally retained as endosymbionts (FIG. 1). Numerous well-established endosymbiotic partnerships have been described between a variety of eukaryotic hosts and prokaryotic or eukaryotic endosymbionts<sup>19-22</sup>. The level of host-symbiont integration ranges from transient affairs to permanent and obligatory marriages, each of which provides an opportunity for genes to move to a new genome (FIG. 2).

*Primary symbiosis and endosymbiotic gene transfer.* The two prominent endosymbioses in eukaryotic evolution are, of course, the origin of mitochondria and plastids from an alpha-proteobacterium and a cyanobacterium, respectively. All of the examined organellar genomes encode only a small fraction of the organelle's proteins: from 3 (or perhaps 2 in the case of *Oxyrrhis marina*<sup>23</sup>) to 67 for mitochondria and from 15 to 209 for plastids. The majority of organelle proteins are encoded by nuclear genes, so massive endosymbiotic gene transfer (EGT) has been inferred to have taken place in the evolution of both organelles<sup>24,25</sup>. In the case of plastids, some of these genes were transferred repeatedly, initially from organelle to nucleus (discussed in this section) and, in secondary and tertiary plastids, from endosymbiont nucleus to host nucleus (FIG. 1; BOX 1). Many organelle-derived genes have also been hypothesized to now supply proteins to other cellular compartments, so enzymes of some biochemical pathways are derived from different sources, targeted to a compartment other than that from which they originated, or both<sup>26-28</sup>. The scale of such a contribution is not entirely clear, but some analyses suggest it is significant<sup>24,25,29</sup>.

The similarities in residual gene content among plastids, and among mitochondria, suggest that most organellar genes were transferred in early, and perhaps rapid, migrations. The subsequent tempo of EGT has been highly punctuated, with bursts of transfer interspersed with, or terminating in, long periods of stasis<sup>30</sup>.



**Figure 2 | You are what you eat, what you live on, what lives on you, and what lives in you.** Several behaviours and life-styles can enhance horizontal gene transfer (HGT), some of which are shown here. The predatory ciliate *Didinium* engulfs and digests another ciliate, *Paramecium*, as shown by SEM (a) (photo courtesy of W. Foissner, University of Salzburg). Phagotrophy, especially in microbial eukaryotes in which the germ and soma cells are the same, could greatly enhance the access of an organism to foreign DNA. The stem of the parasitic plant *Cuscuta* (dodder) entwines the stem of its host plant, *Glechoma* (b) (photo courtesy of K. Robertson, Illinois Natural History Survey). The parasite *Cuscuta* forms intimate connections with the vascular system of its host, *Glechoma*, which are called haustoria (c) (photo courtesy of H. Albrecht and J. Yoder, University of California, Davis). Several cases of mitochondrial HGT between parasitic plants and their hosts have been described (BOX 3). Plastid movement by successive rounds of endosymbiosis has affected several groups of eukaryotes. *Kryptoperidinium* is a consortium between a dinoflagellate and a diatom (d). Nuclei of both partners (blue) are adjacent to one another (the dinoflagellate nucleus is round with bright spots, the diatom nucleus is multi-lobed and less bright) and the diatom plastids are red (photo by P.K.). The sea slug *Elysia clarki* is bright green (e) because it retains photosynthetically active plastids from its food algae for months after the food has otherwise been digested (photo courtesy of S. Pierce, Gulfbase at Texas A&M University–Corpus Christi). Permanent and long-term endosymbiotic associations lead to large-scale sequence migrations, but the depth of this impact has not been wholly investigated in most systems, including in *Kryptoperidinium* and *Elysia clarki*. Some interspecies associations that are highly specific and long term are probably less integrated at the genetic level but nonetheless provide opportunities for transfer. The surface of a devescovineid flagellate is completely covered with a uniform layer of bacteria (f), whereas the flagellate *Barbulonympha* also takes up bacteria into its cytoplasm (g) (both photos by K. Carpenter, University of British Columbia and P.K.).

#### Polytomy

In a phylogeny, when a branching pattern cannot be resolved, the branches in question can be collapsed to show the absence of a hypothesis for the relationships among the lineages that they represent.

#### Endocytosis

A general eukaryotic cellular process using the cytoskeleton and endomembrane system to take up material from the environment.

#### Ectosymbiont

An organism living in a symbiotic association with another, specifically by attachment to the surface of its host.

#### Endosymbiont

An organism living in a symbiotic association with another, specifically by living inside a host cell.

Many genes also have a patchy distribution across extant organellar genomes, implying repeated loss or repeated EGT<sup>24,30</sup>. This pattern is most dramatically evident in plant plastids<sup>31</sup> and even more so in mitochondria: the study of 280 diverse angiosperms has revealed a massively parallel, highly punctuated and gene-specific pattern of evolutionarily recent EGT for 16 of the 40 mitochondrial-protein genes that were examined<sup>32,33</sup>. By contrast, animals exhibit profound stasis in mitochondrial gene content: not one case of functional EGT has been reported across the 600 million years of metazoan evolution, despite sequencing of thousands of animal mitochondrial genomes (BOX 2).

Many nuclear genomes contain many organellar-derived sequences, some as large as entire organellar genomes, whereas comparative and experimental studies indicate high rates of introduction and elimination of these sequences<sup>34–38</sup>. These findings establish that nuclear genomes are adept at taking up foreign DNA, which is

also an important precondition for HGT. However, the frequent transfer of mitochondrial genes in angiosperms seems to be largely an RNA-mediated process, as most functionally transferred genes are distinguishable from their mitochondrial progenitors by the lack of introns and/or RNA editing sites<sup>30,33</sup> (see REF. 39 for an alternative view). Many recently transferred genes acquired their targeting sequences — and probably their upstream promoter and regulatory sequences — from long-standing nuclear genes for proteins that are targeted to the same compartment<sup>30,33,40,41</sup>. This emphasizes the fluidity of nuclear genomes and the potential of fortuitous recombination events to create properly expressed and targeted genes.

**Gene transfer from bacterial endosymbionts.** Unlike mitochondria and plastids, we know relatively little about many of the other intimate associations with ectosymbionts and endosymbionts (FIG. 2). Two recent findings



## Box 1 | Secondary and tertiary endosymbiosis: Matryoshka dolls of cells and genomes

Although the great majority of evidence suggests that plastids arose only once in eukaryotic evolution (for discussion and review see REFS 110–113), they subsequently spread to many other diverse eukaryotic lineages by secondary eukaryotic–eukaryotic endosymbioses<sup>110,114</sup> (FIG. 1). In most algal lineages that trace back to secondary symbioses, all of the hundreds of plastid-derived genes from the endosymbiont nucleus have been transferred to the host nucleus and the endosymbiont nucleus has disappeared, but in two cases it persists as a highly reduced relic called a nucleomorph<sup>110,114</sup>. In the most extreme cases known, certain dinoflagellates have engaged in yet another round of tertiary endosymbiosis by taking up another alga with a different secondary plastid, resulting in a complex genomic heritage (FIG. 1). Thus, many of the genes that are acquired by primary cyanobacterial endosymbiosis have been transferred over and over again during eukaryotic evolution, initially from plastid to nucleus (and often many times separately<sup>24,31</sup>), repeatedly and *en masse* from nucleus to nucleus during each secondary symbiosis, and yet again from nucleus to nucleus in each tertiary symbiosis.

An important complication in all tertiary endosymbioses is that the host is also from a photosynthetic lineage (that is, dinoflagellates), so it either possessed a plastid during the symbiosis, or once possessed one. The potential complexities in gene transfer and genomic chimerism afforded by such redundancy are illustrated by a study of endosymbiotic gene transfer in the dinoflagellate *Karlorodinium*. The tertiary plastid in this alga is present owing to the replacement of the original dinoflagellate plastid (of secondary origin) with a new one (also of secondary origin) derived from a haptophyte alga (FIG. 1). Nuclear genes for plastid-targeted proteins in this organism would be expected to be derived from haptophytes, because that is where the plastid came from, and indeed the first such proteins to be described were exactly so<sup>115</sup>. Yet, analysis of a large number of plastid-targeted proteins subsequently showed that both haptophyte and dinoflagellate proteins were retained, and both are seemingly targeted to the new plastid, resulting in a highly chimeric plastid proteome<sup>75</sup>.

Another complexity is that many lineages of protists are hypothesized to have once contained a plastid<sup>105,116,117</sup> (FIG. 1), but do not apparently contain one now<sup>118–120</sup>. Strong support for this with respect to the oomycetes comes from the discovery of 855 genes of putatively cyanobacterial or red-algal origin in the nucleus of *Phytophthora*, 295 of which are shared with the diatom *Thalassiosira*<sup>120</sup>. However, the ciliate genomes that have been sequenced seem to lack any plastid-derived genes<sup>119,121</sup>, and so either these two species have totally lost all vestige of a photosynthetic ancestry, or some plastid-derived genes were overlooked, or the Chromalveolate hypothesis<sup>105,116,117</sup> (FIG. 1) is wrong. The first two possibilities are not unlikely given the complexities of whole-genome phylogenetic analyses and the near-total loss of plastid-derived genes (only seven seem to remain) in the aplastidic apicomplexan *Cryptosporidium parvum*<sup>122</sup>, which clearly once had a plastid<sup>123</sup>. By contrast, the presence of ‘plastid genes’ in trypanosomatid parasites was used to argue for a photosynthetic ancestry of this group<sup>124</sup>, but subsequent analyses have undermined these claims<sup>76,125</sup>.

in such systems suggest that symbiosis-mediated gene transfers will be found in many more genomes. First, large tracts of *Wolbachia* DNA have moved, often quite recently, from these common intracellular bacterial endosymbionts to the nuclear genome of their insect and nematode hosts<sup>42,43</sup>. In the most extreme case so far, an entire copy of the *Wolbachia* genome was found in the genome of a fruitfly, and 2% of the transferred genes were shown to be transcribed<sup>42,43</sup>. Whether these transfers are functional, much less adaptive, is unclear (except for two large, relatively old transfers, the genes of which are all degenerate), but it is clear that the intimate association of the two genomes has greatly enhanced the chances for gene transfer. An important cautionary lesson from these studies<sup>42,43</sup> is that many cases of recent bacterial-to-nuclear HGT might be overlooked if bacterial sequences are routinely excluded from nuclear sequence assemblies from the many lineages containing such symbionts. The second unexpected finding is the discovery of two unrelated bacterial endosymbionts, the genomes of which are comparably reduced in size (160 kb and 245 kb) and protein-coding gene number (182 and 228) to the genomes of plastids. The smaller of the two endosymbiont genomes<sup>44,45</sup> seems to lack genes that are thought to be required for DNA replication, transcription and translation, and is therefore likely to import host proteins, quite possibly from transferred bacterial genes; the larger genome<sup>46</sup> might also lack some essential genes.

### Prokaryote–eukaryote transfers

**Variable amounts of prokaryote–eukaryote HGT.** Excluding genes of obviously endosymbiotic origin, the number of nuclear genes acquired from bacterial sources varies enormously between species, with estimates ranging from zero to hundreds. Relatively few multicellular genomes have been systematically searched for evidence of bacterial HGT: none of the six sequenced genomes of land plants and only one of the dozens of sequenced animal genomes have been analysed. This exception, notoriously, is our own genome, for which there are no credible cases of bacterial HGT<sup>7–9,14</sup>. Relatively few of the sequenced fungal genomes have been reported to have been systematically searched, and those that have are thought to contain only a few bacterial genes (but see below for a major caveat).

An important limitation in most genome-wide studies is the use of search strategies with initial steps that inherently bias towards the detection of relatively recent transfers from bacteria. These include studies that begin either by a BLAST search to identify all genes in a given genome that have best hits matching bacterial genes (followed by phylogenetic analysis of all candidate cases of HGT) or by identifying genes that are present only in the organism under study (relative to all eukaryotes or to a set of related ones) and then asking whether they are of bacterial origin. The identification of genes that are only present in the organism in question has strongly shaped the perception that bacterial transfer to fungal genomes

**Box 2 | Protein import: the major constraint on endosymbiotic gene transfer**

A crucial component of the transformation of the mitochondrial and plastid endosymbionts into cellular organelles was the establishment of a complex multi-subunit import system to efficiently target proteins that are encoded in the nucleus back to the organelle<sup>126</sup>, as this allowed the large-scale movement of genes from organellar genomes to the nucleus. In the case of the plastid, however, whether this import system is the outcome of, rather than the reason for, organellar establishment is the subject of vigorous debate<sup>111,127,128</sup>. So too is the corollary issue of whether organelles should be distinguished from their progenitor endosymbionts by genetic integration with the host cell (and thus endosymbiotic gene transfer, EGT) or simply by metabolic integration. Targeting to secondary and tertiary plastids (BOX 1) is not so well characterized, but is known to have extra layers of complexity, reflecting the need to traverse three to four plastid-bounding membranes (rather than the two in primary plastids) of divergent origin and properties<sup>129</sup>.

Ease of 'importability' — which includes the full range of issues associated with efficient protein targeting to the correct organelle, import and proper assembly — is probably the major factor determining whether and how readily a given organellar gene is functionally transferred to the nucleus<sup>30,130–132</sup>; for an alternative view, see REF. 133. The core set of organellar-encoded proteins — the proteins retained by the many highly and independently reduced mitochondrial genomes<sup>72,134</sup> and by the uniquely reduced plastid genome of stereotypical dinoflagellates<sup>135</sup> — are virtually always hydrophobic proteins, which were long thought and increasingly shown to be the most difficult to properly import and assemble<sup>131</sup>. For example, comparative analysis<sup>136,137</sup> has shown that exceptional cases of the successful functional transfer of these most EGT-refractory genes are invariably associated with marked reductions in hydrophobicity. In one such case, an elegant experimental study<sup>132</sup> identified specific hydrophobicity-reducing amino-acid changes as necessary and sufficient to make the protein importable. The hydrophobicity–importability hypothesis is also supported by studies showing that mitochondrial genes encoding hydrophobic proteins are essentially never transferred in plants despite remarkably frequent transfer of genes for non-hydrophobic proteins<sup>32,33</sup>. Import constraints, together with the evolution of divergent mitochondrial genetic codes in the most reduced mitochondrial genomes, seem to have locked many lineages (animals most prominently) into a small, virtually irreducible set of 'core' mitochondrial genes.

is quite rare. For instance, the genomes of six diverse hemiascomycetes (yeasts or yeast-like fungi) have been inferred to contain very few (0–10) genes of bacterial origin, but all genes that were present in even one other yeast under consideration were rejected as candidates for potential bacterial HGT<sup>47,48</sup>. Less restrictive and more comprehensive searches for bacterial HGT in fungi would probably yield more cases, especially given mounting evidence from single-gene studies (see below) that certain fungi are actively acquiring genes by HGT.

Despite such analytical limitations, most of the protist genomes that have been examined contain a significant number of genes of probable bacterial origin (Supplementary information S1 (table)). On a percentage basis, HGT (from bacteria at least) contributes less to protistan genomes than it does to bacterial genomes, although rumen-dwelling ciliates, which are estimated to have acquired 4% of their genes from bacteria<sup>49</sup>, approach levels of HGT that are commonly found in bacteria. The number of bacterial genes in any particular nuclear genome is likely to be a complex function of multiple factors affecting the likelihood of both the acquisition of bacterial genes and their fixation and persistence. One commonly noted factor affecting incorporation of bacterial genes is opportunity, or exposure to bacterial DNA. Many protists are phagotrophs, they subsist by eating bacteria and sometimes other eukaryotes (FIG. 2). This

has led to the 'you are what you eat' theory of HGT<sup>50</sup>. Other protists, although not phagotrophs now, might have been so for long periods in their past and/or might live in environments where they are frequently exposed to bacterial DNA (for example, parasites, rumen-dwellers and so on). But not all protists are exposed in this way, and this probably correlates with a lack of bacterial genes. For example, the non-phagotrophic green alga *Chlamydomonas reinhardtii* so far seems to lack bacterial genes<sup>51,52</sup>, although its recently sequenced genome has not yet been examined in this regard.

Another factor that is commonly mentioned<sup>53,54</sup> is that eukaryotes with a highly segregated germ line (that is, animals) will tend to be most sheltered from heritably meaningful exposure to foreign DNA (and might, for example, have a role in the contrast between animal and plant mitochondria mentioned below). This is probably a relatively strong deterrent to HGT in animals, although certainly not an absolute one. Several strong cases for HGT into animal genomes are now known<sup>55,56</sup>, and we suspect that when more animal genomes are searched systematically for HGT other cases will emerge. Each of these factors, and others, all probably affect the frequency of HGT differently in different lineages, but it would be too simple to identify one factor as having a key role across eukaryotes. More probably, the frequency of HGT is a complex sum of various factors in any extant lineage. It is important to remember that the interplay of different factors will change over time as the organism evolves, but the legacy of previous tendencies will be preserved in the genome: an organism with traits that seem to discourage HGT today might have evolved from ancestors with traits that encouraged it, and it might still harbour many genes acquired by that ancestor.

Although we can safely generalize that overall levels of bacterial genes in nuclear genomes vary substantially among eukaryotes (Supplementary information S1 (table)), too few 'whole' genomes have been examined to allow much insight into patterns, much less rates, of both horizontal transfer and loss of transferred genes. One exception is the three recently sequenced trypanosomid genomes, which were found to contain in aggregate 47 putative genes of bacterial origin<sup>57</sup>. Only 20 of these genes were common to all three trypanosomes, whereas 6 genes were found in only two of the three, and 21 were unique to a single genome, indicating considerable variability in gain and/or loss of bacterial genes within the estimated 250 million year divergence<sup>58</sup> of the three species. A contrasting comparison of *Spiroplasma salmonicida* and *Giardia lamblia*, which represent the earliest known split within diplomonads, reveals considerable (72%) sharing of the 68 bacterial genes found in *S. salmonicida*, whereas at least a dozen of these transfers seem to have occurred even more anciently, in the common ancestor of diplomonads and parabasalians<sup>59,60</sup>.

Finally, a noteworthy but poorly understood case of potentially massive HGT comes from the green alga *Ostreococcus tauri*. HGT seems not to be a prevalent feature of the genome as a whole, but chromosome 19 differs from the rest of the genome with respect to GC ratio, gene density and the proportion of mobile

**Rumen**

A fermentative compartment of the digestive system in many cellulose-digesting vertebrates, the contents of which are rich in anaerobic protists and prokaryotes.

**Ciliate**

A lineage of protists (for example, *Tetrahymena* and *Paramecium*), predominantly predators, defined by the presence of dimorphic nuclei and large numbers of short flagella (cilia) on the surface. They are members of the Chromalveolates.

**Trypanosomid**

A lineage of protist flagellates (for example, *Trypanosoma*, the sleeping-sickness agent), predominantly made up of parasites, and home to many unusual characteristics of genome structure (for example, RNA editing). They are members of the Excavates.

elements<sup>61</sup>. Moreover, genes on chromosome 19 are significantly less likely to share a phylogenetic relationship with other green algae, and many are weakly related to bacterial homologues. It was concluded that the entire chromosome might be derived from some exogenous source, although whether that source was a bacterium remains to be shown.

#### *Adaptive functions of bacterial genes in eukaryotes.*

In addition to the variation among eukaryotes in the number of genes acquired from bacteria, there is considerable variation in the tendency for certain genes or kinds of genes to be transferred. For example, some organisms that have become ecologically specialized are rich in horizontally transferred genes, and the genes that allow such adaptations seem to be among the most commonly acquired. Anaerobic parasites that rely on fermentation tend to contain many prokaryotic genes that are related to fermentation or to other aspects of anaerobic metabolism, and in some cases the same gene has been acquired from different bacteria in distantly related anaerobic eukaryotes<sup>60,62,63</sup>. This suggests that the different lineages adapted to the anaerobic environment by borrowing genes from bacteria *ad hoc*, a conclusion that has been reinforced by the complete genomes of *G. lamblia*, *Trichomonas vaginalis*, and particularly *Entamoeba histolytica*, in which many of the 96 putatively bacterial genes are related to its niche of anaerobic metabolism<sup>64–66</sup>. The genomes of these highly specialized parasites might be the tip of the anaerobic iceberg: as mentioned earlier, anaerobic ciliates that live in the animal rumen contain the highest reported proportion of bacterial genes in a nuclear genome, and again many of these genes are involved in metabolic processes that are unique to their environment<sup>49</sup>. Adaptation to parasitism might also favour the acquisition of new genes by HGT. Numerous instances of HGT have been reported in parasites, despite the often reduced complexity of their proteome<sup>12</sup>.

The ecological advantages that HGT might confer during a period of adaptation, such as becoming anaerobic or parasitic, are obvious (and this is not restricted to bacterial genes), but this is not to imply that HGT is only associated with extreme adaptations. The *Dictyostelium discoideum* genome, for example, was reported to have 18 potential cases of bacterial genes, many of them related to life in the soil (for example, dipeptidases for degrading bacterial cell walls, a siderophore for the transport of ferric iron, and soil toxin-resistance genes)<sup>67</sup>.

**Why are prokaryote–eukaryote transfers relatively common?** Most described cases of HGT in eukaryotes involve bacterial genes. To a significant degree, this reflects two interrelated experimental biases. First, bacterial transfers are generally much easier to detect than within-eukaryote transfers, this is because even with relatively poor sampling they stand out in phylogenetic analyses owing to the large evolutionary distance between bacteria and eukaryotes. Second, and consequently, most studies that have systematically examined whole nuclear genomes for evidence of HGT have focused on genes of potential bacterial origin.

This is not to say, however, that the pattern does not reflect a real biological bias. Some forces might favour transfer between eukaryotes; for example, a virus cross-infecting close relatives could increase the flow of genes between them, and a eukaryotic gene product generally might be more likely to interact appropriately with existing proteins and therefore be retained in another eukaryotic genome, particularly if it is replacing an existing gene that encodes a protein that is part of a co-adapted pathway or complex. On the other hand, several factors such as gene and genome organization might favour the transfer of bacterial genes to eukaryotic genomes. All eukaryotes have introns and are therefore able to splice them in principle, but in practice introns from one species might not splice accurately or efficiently in a foreign genome if their characteristics differ. Prokaryotic genomes contain many operons with clustered functionally related genes, whereas operons are absent from most nuclear genomes; thus, bacteria offer an unmatched potential for acquiring an entire pathway by a single, relatively small transfer event. In many cases, functionally related genes seem to have been derived from different bacteria (see REF. 66 for an example), so this might not be an important factor favouring the acquisition of bacterial genes, although multiple gene transfer is rarely well documented. One clear example is of two adjacent cyanobacterial genes that fused when they transferred to the ancestor of dinoflagellates, leading to a novel subcellular localization of the second protein<sup>68</sup> — other cases have also been documented (see REF. 69 for an example).

Although characteristics like these might have some effect on the frequencies of transfer, the largest biological effect is probably simple opportunity. Bacterial populations are immense in comparison with eukaryotic populations, so the pool of bacterial genes that can be acquired in the environment is proportionately larger. Similarly, many eukaryotes now and throughout their evolution have eaten bacteria, whereas eukaryotrophs are comparatively rare, so overall the availability of prokaryotic genes for transfer into a eukaryotic genome is probably much greater than for genes from other eukaryotes.

**HGT from prokaryotes to organelles.** The above section deals exclusively with the nuclear genome. Despite ideal conditions for detection (far more organellar genomes have been sequenced than nuclear genomes, and transfers should be easy to spot because organelle genes are generally strongly united phylogenetically), phylogenomic and other studies have revealed just three cases of eubacteria-to-plastid HGT. Two are ancient homologous replacements of pre-existing plastid genes (in one case, an operon), both marking major clades of algae<sup>70</sup>. The third is a more recent acquisition of a gene of novel function<sup>71</sup>. Setting aside ambiguous cases involving genes for phage-type DNA or RNA polymerase, which were probably acquired from mitochondrial plasmids of uncertain origin, there is no good evidence for the direct introduction of bacterial genes into mitochondrial genomes<sup>72</sup>.

#### Diplomonad

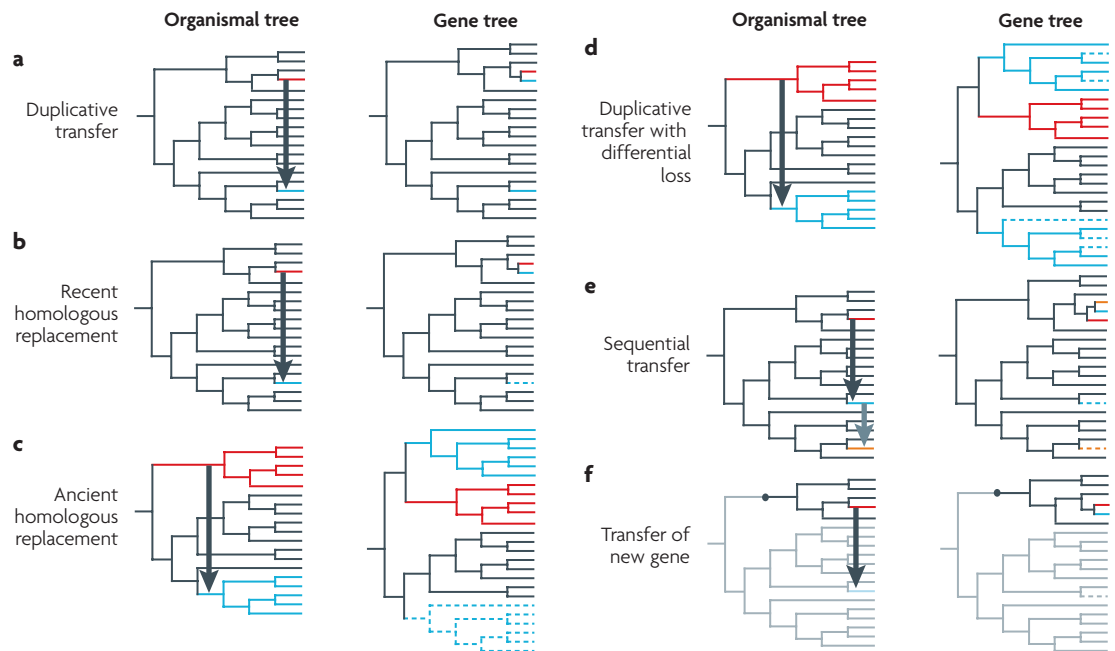
A lineage of anaerobic or microaerophilic protist flagellates (for example, *Giardia lamblia*), predominantly parasitic and often studied because of their reduced metabolism and mitochondria. They are members of the Excavates.

#### Parabasalium

A lineage of anaerobic or microaerophilic protist flagellates (for example, *Trichomonas*), predominantly parasitic and often studied owing to their reduced metabolism and their hydrogenosome, a hydrogen-producing mitochondrial relict. They are members of the Excavates.

#### Dinoflagellate

A lineage of protist flagellates (for example, *Alexandrium*, a red tide alga) with photosynthetic, heterotrophic and parasitic representatives, which are known for many unusual modifications to genome structure — they are members of the Chromalveolates.



**Figure 3 | Different kinds of transfer and their effects on gene phylogeny.** The organismal trees each represent a hypothetical phylogeny of organisms for which a different type of horizontal gene transfer (HGT) event has taken place from a red branch to a blue branch. The effects of these events are shown in the gene trees, which represent the phylogeny of the transferred gene, with the dashed branches indicating lineages that have lost a particular gene. **a,b** | These trees show the relatively simple cases of duplication or replacement, the interpretation of which is straightforward given adequate sampling. **b,c** | These trees show the difference between recent and ancient events. **d** | These trees show the effects of a duplicative transfer followed by differential loss of one gene or the other in the lineage, a case in which incomplete sampling would greatly distort the interpretation of the events. **e** | These trees show two sequential transfers and how this can lead to a complex distribution of a closely related gene among distantly related organisms. **f** | These trees show a gene that exists in only a subset of organisms (its origin is indicated by the dot, and the grey branches lack the gene altogether), the transfer of which leads to a patchy distribution. If the origin of the gene is by HGT from an even more distantly related group (for example, a prokaryotic gene transferred to a eukaryote), then this is a special case of sequential transfer.

### Eukaryote–eukaryote transfers

**Inferring eukaryote–eukaryote HGT and its functional implications.** Eukaryote–eukaryote transfer of nuclear genes is underestimated for a number of reasons: the sampling of most eukaryotic genes is only now approaching the level needed to see such events; there is extensive, often confounding, gene duplication within many nuclear genomes; and there has not, to our knowledge, been any systematic search of completely sequenced genomes for within-eukaryote HGT (this is not true for EST projects, in which several large-scale searches for HGT have been carried out, see REFS 51,52,73,77,78 for examples). Given all this, it is noteworthy how many gene phylogenies have led to the conclusion that genes are in fact transferred between eukaryotes<sup>10,11,26,27,51,59,60,62,74–83</sup>. Many of these acquired genes replaced an existing homologue rather than introducing a new function, but this might largely reflect the way these transfers were detected rather than any real bias. Indeed, cases in which novel functions have been acquired are particularly well described in fungi, and include changes to mating that can affect population structure<sup>81,84</sup>, uptake and synthesis of small molecules<sup>48,82,85</sup>, or the transfer of virulence factors<sup>80</sup>.

This last case is notable for its recentness: an 11 kb region containing a toxin gene is thought to have been transferred to a previously avirulent fungal species only about 70 years ago. Many of these are transfers between two fungi (some of which were closely related strains<sup>86</sup>), but fungal genes have been transferred to other eukaryotic groups with important effects: in one particularly interesting case, 11 genes from filamentous fungi were found in oomycetes, and because many of these have important functions in osmotrophy this led to the conclusion that HGT played a part in the convergent adaptation to plant pathogenesis in the two groups<sup>79</sup>.

Of emerging importance is an alternative means for identifying transfers between two eukaryotes, which takes advantage of rare events that allow HGT between eukaryotes to be tracked more easily (FIG. 3). For example, if a eukaryotic lineage acquired a prokaryotic gene by HGT and this gene was subsequently transferred to other eukaryotes, the result would be two or more distantly related eukaryotes possessing closely related prokaryotic genes (FIG. 3e,f). The prokaryotic origin of the gene (in addition to being another case of HGT) is a 'tag' that allows the subsequent transfer of the gene between eukaryotes to be detected, even without ideal sampling

#### Oomycete

A lineage of protist parasites (for example, *Phytophthora*, the potato late-blight agent) that are responsible for numerous plant diseases, and were once mistakenly thought to be fungi but are really heterokonts. They are members of the Chromalveolates.

#### Osmotrophy

Feeding by absorption of nutrients directly from the environment (which can include a host organism in the case of parasites).



or an unusually well-resolved phylogeny. A number of such cases have been described and interpreted in this way, on the basis of several distinctive tags<sup>17,59,62,74,87–89</sup>. In some cases, genes that previously seemed to be the result of relatively simple HGT with replacement or duplication (FIG. 3a,b) have, with greater sampling, been revealed to be more consistent with multiple losses or multiple transfers between eukaryotes<sup>88</sup> (FIG. 3d,e,f). The most complex case is elongation factor-like protein (EFL), a GTPase hypothesized to have been transferred between eukaryotes, replacing an ancient paralogue, translation elongation factor EF-1 $\alpha$ , as it moves<sup>17,87</sup>. This and other genes with similarly complex distributions suggest that transfers between eukaryotes might be more common than we currently realize; however, finer-scale sampling of EFL and EF-1 $\alpha$  in green algae conversely suggests that differential gene loss probably also has a role in its distribution, at least at this scale<sup>17</sup>. Whether this applies to larger scales is not certain, but as the number of distinct groups with EFL continues to expand, the case for eukaryote–eukaryote HGT alone explaining its overall distribution loses strength. Ultimately, the distribution of EFL is probably due to HGT and other forces, emphasizing the need for more sampling of even relatively well-sampled genes.

**Plastid-targeted proteins.** Without the benefit of such special circumstances, only a few proteins have been adequately sampled and have a sufficiently clear evolutionary history to reveal major trends in HGT across a diversity of eukaryotes. One such class of genes are the nuclear genes for proteins that are targeted to the plastid. These genes are relatively well sampled, and in nearly all algal groups we have a clear expectation that they should be related to either red or green algal homologues, depending on the origin of the plastid in question (FIG. 1; BOX 1). However, in dinoflagellates, heterokonts, haptophytes and chlorarachniophytes, EST projects have revealed that a sizable minority of these genes are related to the wrong algal lineage, probably as a result of HGT between algae (other genes are derived from non-cyanobacterial prokaryotic lineages and are also interpreted to be the result of HGT)<sup>51,73,75,77,78,89</sup>. By contrast, the same genes from the green alga *C. reinhardtii* do not show signs of HGT<sup>51</sup>, once again revealing the variability in HGT to different lineages. It is noteworthy that in some of these algae<sup>51</sup> the proportion of genes inferred to be derived by HGT is greater than the highest reported proportion of bacterial genes in eukaryotic genomes (Supplementary information S1 (table)), so if plastid-targeted proteins are representative of the rest of the genome, transfers of genes between eukaryotes might be abundant in some lineages.

**Organelle–organelle HGT.** Although thousands of animal mitochondrial genomes have been sequenced, only a single study has claimed that HGT has occurred in animal mitochondrial DNA. Nevertheless, the claim is extraordinary in that multiple recent transfers of the mitochondrial cytochrome b (*cytb*) gene are inferred among three sympatric species of beetles in Mexico<sup>90</sup>.

Like the mitochondrial DNA of animals (but with the possible exception of these beetles), plastid DNA of land plants is devoid of HGT, whereas plant mitochondrial DNA is unexpectedly active in HGT. Potential reasons for this disparity, as well as important findings from the study of HGT in plant mitochondria, are discussed in BOX 3.

### Eukaryote–prokaryote transfers

**Providing new functions to bacteria.** Few cases of eukaryotic-to-prokaryotic HGT have been reported, the most interesting being several apparently recent transfers of genes that are otherwise found only in eukaryotes. In two cases the proteins have central roles in the cytoskeleton, and therefore they must impart some novel function to their new hosts. Alpha- and beta-tubulins are encoded in the bacterium *Prostheco bacter* by an operon that includes another eukaryotic gene — for kinesin light chain<sup>91</sup>. Both tubulins seem to have structural differences compared with canonical tubulins<sup>92</sup>, but they do form protofilaments with similar properties and seem to polymerize by some cooperative assembly mechanism<sup>92,93</sup>. Similarly, genes for actin and the functionally associated profilin have been found adjacent to each other in the genome of a single strain of the cyanobacterium *Microcystis aeruginosa*<sup>94</sup>. As with the *Prostheco bacter* tubulins, the actin has been localized and found to form a shell within the cell wall, suggesting again that the protein has taken on some structural role in the bacterium<sup>94</sup>. Another interesting case is fructose bisphosphate aldolase (FBA), an enzyme that exists as two non-homologous analogues, one common to bacteria and one common to eukaryotes. Several isolates of the closely related cyanobacteria *Prochlorococcus* and *Synechococcus* have been found to possess a eukaryotic FBA, which is located adjacent to the bacterial analogue that is ancestral to cyanobacteria<sup>95</sup>. The transferred gene is clearly derived from the plastid-targeted FBA of red algae, suggesting it might have taken on a role in carbon fixation in these ecologically important cyanobacteria<sup>95</sup>.

### Why are eukaryotic genes rare in prokaryotic genomes?

Given how many prokaryotic genomes have been sequenced and how much prokaryote-to-prokaryote HGT has been found, the paucity of eukaryote-to-prokaryote transfers begs explanation, all the more so because transfers from a distantly related source such as eukaryotes should be easy to detect. In particular, no large-scale eukaryotic contribution has been documented in any prokaryotic genome. One potentially interesting case was made for eukaryotic genes related to pathogenicity in *Mycobacterium tuberculosis*<sup>96</sup>, but this was subsequently undermined by further analyses<sup>97</sup>. Introns are major barriers when present, but they are absent from many genes in many eukaryotes. Opportunity might be a major factor; as described above, population sizes of bacteria are far greater and therefore offer a larger pool of potential donors, and the overall amount and intimacy of direct interaction between cells is probably much greater within the bacterial

#### Heterokont

A lineage of protist (for example, oomycete parasites and kelps) with photosynthetic, heterotrophic and parasitic representatives, all of which are united by the possession of uniquely dimorphic flagella. They are members of the Chromalveolates.

#### Haptophyte

A lineage of photosynthetic protist (for example, *Emiliania*), predominantly marine, some of which form massive marine blooms, and many of which make distinctive calcium carbonate scales that have contributed significantly to limestone deposits. They are members of the Chromalveolates.

#### Chlorarachniophyte

A lineage of photosynthetic protist (for example, *Bigelowiella*) with amoeboid and flagellate life stages, best known for their retention of a relict nucleus of their green algal plastid endosymbiont, known as a nucleomorph. They are members of the Rhizaria.



world than between bacteria and eukaryotic cells. Gene transfer from eukaryotes to bacteria should be disfavoured by the processes of conjugation and transduction, whereas transformation should in general be neutral with respect to donor DNA. Finally, eukaryotes might simply not have much to offer in the way of genes that are potentially useful in bacterial evolution.

**HGT and eukaryotic phylogeny**

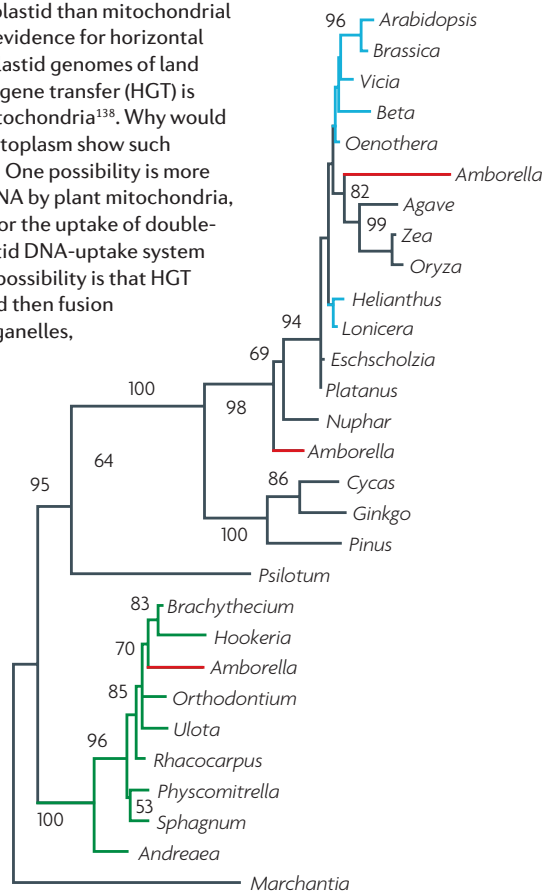
Although HGT should now be recognized as an important force in the evolution of many groups of eukaryotes and their genes, including some genes previously thought to be ‘immune’ to HGT<sup>83,87</sup>, there is no reason to think that it is so prevalent as to undermine efforts to reconstruct a dichotomously branching

**Box 3 | Horizontal gene transfer in the plant cytoplasm**

Despite far greater sampling of plastid than mitochondrial genes and genomes, there is no evidence for horizontal transfer of eukaryotic genes in plastid genomes of land plants<sup>70</sup>, even though horizontal gene transfer (HGT) is surprisingly common in plant mitochondria<sup>138</sup>. Why would two genomes within the same cytoplasm show such strikingly different rates of HGT? One possibility is more efficient uptake of exogenous DNA by plant mitochondria, which possess an active system for the uptake of double-stranded DNA<sup>139</sup>, whereas a plastid DNA-uptake system has not been reported. Another possibility is that HGT is mediated by direct contact and then fusion between donor and recipient organelles, in which case the far greater propensity of mitochondria than plastids to fuse with one another<sup>140,141</sup> could explain the rate differences. According to the fusion hypothesis, plant mitochondrial genomes should preferentially take up other mitochondrial sequences by HGT, and indeed this is the case so far, with the donors limited to other land plants<sup>138</sup>. However, these two patterns could be a sampling artefact of the PCR approaches used, and so sequencing HGT-rich plant mitochondrial genomes is crucial to distinguishing between the fusion and uptake hypotheses. Differences in streamlining pressures might also contribute to the striking disparity in organellar HGT rates, as the more compact genomes of plastids are probably less likely to incorporate foreign DNA and more likely to lose it.

Several other findings have emerged from the study of HGT in plant mitochondria:

- Most transfers are recent events, being restricted to a single plant genus or even a limited subset of species within a genus<sup>142–144</sup>.
- Most transferred genes published so far seem to be non-functional and to coexist with a native, functional homologue.
- A few intriguing cases have been described in which a foreign gene has recombined with the native copy, creating a chimeric, probably functional gene<sup>145,146</sup>, and many more cases of ‘chimeric HGT’ await publication (J.D.P., unpublished observations).
- Mitochondrial HGT sometimes occurs on a massive scale: the mitochondrial genome of the basal angiosperm *Amborella* possesses proportionately more foreign DNA than even the most HGT-rich bacterial genomes<sup>147</sup> (J.D.P., unpublished observations). As an example, the figure shows a native copy of *nad5* in *Amborella* (middle *Amborella* branch), and foreign copies derived from angiosperms (upper *Amborella* branch) and mosses (lower *Amborella* branch). The numbers on the branches represent bootstrap support, a statistical measure of confidence in that relationship.
- Although there are many ways in which DNA could theoretically move between plants (by any number of biological vectors, by plant-to-plant contact, by illegitimate pollination or by transformation, for example), five studies provide evidence from phylogeny and sometimes biogeography or life history, that a major route involves direct contact between donor and recipient plants. In all these cases, a parasitic flowering plant is involved as either the transfer donor<sup>142,144</sup> or as recipient from its obligate host plant<sup>146,148,149</sup> (FIG. 2).



An angiosperm



Amborella



A moss

Box 4 | **Transposable elements: indicators of overlooked horizontal gene transfer?**

There is a wealth of evidence that transposable elements (TEs) move horizontally in eukaryotes, often at considerable frequency (see REFS 150–152 for examples). Here we focus exclusively on the predictive use of TEs as sensitive indicators of the potential for horizontal gene transfer (HGT). This reflects both the mobility property of TEs, which enhances their rate of incorporation into a foreign genome relative to regular genes, and the fact that they tend, for reasons that vary according to type of element, to be highly persistent even though they are not under selection in their new genome. We predict that in groups of eukaryotes in which frequent horizontal transfer of TEs has been found, HGT of non-mobile elements — genes in particular — will also be discovered, albeit at lower frequency. For example, 5 years after a homing group I intron was shown to be frequently transferred between mitochondrial genomes of angiosperms<sup>153</sup>, HGT of plant mitochondrial housekeeping genes was discovered<sup>143,145</sup> and is now known to be fairly common (BOX 3).

Fruitflies and yeasts stand out because, despite ample evidence for frequent TE transfer within each group<sup>150</sup> as well as the availability of many complete genome sequences in each group, within-group HGT has not, to our knowledge, been reported. This apparent lack of HGT could be real, which would raise interesting biological questions. For example, are TEs super-mobile within these groups, or are rates of introduction, fixation and/or persistence of foreign genes somehow retarded? Alternatively, perhaps nobody has looked thoroughly for HGT — especially within-group HGT — in flies and yeasts, and some cases of HGT might be waiting to be found. In this regard, note that it took a full 9 years after the *Saccharomyces cerevisiae* genome was sequenced for it to be searched for even a limited subset of potential bacterial-to-yeast HGTs, yet 10 such transfers were found<sup>48</sup>. To our knowledge, other than the human genome, none of the many sequenced animal genomes, including the dozen sequenced fruitfly genomes, have been methodically searched for bacterial HGT, much less for eukaryote–eukaryote HGT. We are unaware of any systematic pan-fungal searches for eukaryote–eukaryote HGT. Few of the many sequenced yeast and other fungal genomes have been thoroughly examined for bacterial HGT<sup>48,154,155</sup>, and even fewer have been searched for fungal–fungal HGT<sup>86</sup>.

If the evidence from TEs for especially frequent phylogenetically local horizontal transfer extends to the genome more generally, as it does in angiosperm mitochondria (see above), then HGT could be particularly difficult to fix and detect in organisms such as *S. cerevisiae*, in which foreign DNA integrates predominately by homologous recombination. Relatively distantly related foreign genes are likely to lead to abortive recombination, or, if recombination is successful, to deleterious chimeric genes, whereas closely related genes will tend to produce subtly chimeric genes that are difficult to detect. Detecting HGT-generated chimeric genes, even between distantly related eukaryotes, often requires dense taxonomic sampling and/or luck in the form of fortuitous signature sequences<sup>74,145,146</sup> (J.D.P., unpublished observations).

We have used flies and yeasts as examples, but we wish to emphasize that a growing number of studies report the occasional to frequent horizontal spread of TEs (especially group I introns<sup>151,156,157</sup>) in a diversity of eukaryotes (especially fungi<sup>151</sup>), and thus these elements might serve as widely useful indicators. As above, a disproportionate share of these TE transfers seem to occur between relatively closely related eukaryotes, raising intriguing questions as to why, and even more so if, these patterns do indeed extend to genic HGT.

**Monophyly**

In phylogeny, a common ancestor and all its descendants are monophyletic (for example, animals), as opposed to a collection of organisms that does not include their common ancestor, which are polyphyletic (for example, flying animals). Monophyletic is sometimes subdivided into holophyletic (the most recent common ancestor and all things that evolved from it, for example, animals) and paraphyletic (the most recent common ancestor, but not all the things derived from it, for example, reptiles — from which birds evolved).

**Chromalveolates**

A hypothetical 'supergroup' of protists, including apicomplexa, dinoflagellates, ciliates, heterokonts, haptophytes and cryptomonads, all of which are hypothesized to have diverged from an ancient common ancestor that has acquired a plastid by secondary endosymbiosis with a red alga.

tree of eukaryotic phylogeny, much less call for the replacement of the tree metaphor with a 'web of life' metaphor, as some have controversially suggested for prokaryotes<sup>1</sup>. The history of some genes is probably so encumbered with multiple transfers and differential losses that these transfers are unlikely to ever be useful as phylogenetic markers<sup>11,16</sup>, but, in a few cases at least, HGT can be viewed as a positive for reconstructing eukaryotic phylogeny, because in principle each transfer has the potential to serve as a valuable phylogenetic marker<sup>98</sup>. In these cases, a transferred gene is found in many or all members of several diverse lineages, supporting their common ancestry. Generally, the shared transfer is noticed because it supports an existing hypothesis<sup>16,27,59,60,68,70,99–101</sup>, and at other times it provides the first strong evidence for a novel major group — for example, cryptophytes plus haptophytes<sup>70</sup>, the monophyly of which was swiftly confirmed by multigene phylogenetic analyses<sup>102,103</sup>. In two cases, transfers have been proposed to even pre-date one of the eukaryotic supergroups (FIG. 1): animals and fungi possess a haloarchaeal tyrosyl-tRNA synthetase<sup>104</sup>, and photosynthetic chromalveolates possess the same bacterial fructose bisphosphate aldolase<sup>105</sup>.

**How does age affect HGT?**

The relatively ancient transfers discussed above constitute a small minority of the current catalogue of prokaryote–eukaryote transfers, the bulk of which are restricted to a single eukaryote or to a small group of relatively closely related eukaryotes. This imbalance is striking, and there are several factors that could contribute to such a pattern. First, the more ancient a transfer, the harder it is to detect because a narrow distribution and a close relationship to donor are characteristics that make evidence for transfer most compelling, but they both deteriorate with age (for example, compare FIG. 3b,c with FIG. 3a,d). Second, as described above, the approaches used to detect bacterial-to-eukaryote HGT in sequenced genomes have a sampling bias that favours detection of relatively recent transfers. Third, our perception of what constitutes ancient versus recent transfers is necessarily tied to our understanding of divergence times within the eukaryotic radiation, and here there is great uncertainty. Most importantly, if the major groups of eukaryotes arose during a relatively compressed period, as suggested by some recent molecular-clock analyses<sup>58,106</sup>, then this might account for at least part of the pattern. Fourth, many transfers might be relatively transient.

If transferred genes are likely to be lost over time, then a snap-shot of foreign genes in a genome at any one time will be dominated by relatively recent transfers.

The fate of foreign genes over the long term has not been investigated in eukaryotes, but an intriguing analysis of genes transferred into one bacterial genus showed that most transfers were indeed transient<sup>107,108</sup>. This is consistent with the kinds of genes that are most often found to have been transferred: these include operational genes — those involved in the metabolism of small molecules that might confer a short-term advantage tied to a particular habitat. This correlation has been interpreted differently before: the complexity hypothesis<sup>109</sup> argues that the likelihood of the ‘transfer’ of a gene (more properly, the likelihood of its fixation and the replacement of its native homologue) is inversely correlated to the complexity of the interactions between its product and other cellular components. This is probably true, but the same conditions can be argued to make the eventual loss of operational proteins easier, and so the genes that are most likely to be transferred are probably those that are most likely to be later discarded. To the extent that this fourth factor is operating in eukaryotes, HGT might have a more significant role in the short term relative to the long haul of evolutionary time. A major challenge for future HGT studies in eukaryotes is to minimize the confounding effects of the first three factors in order to ultimately measure rates of both HGT and the loss of genes acquired by HGT.

### Conclusions and emerging questions

The burgeoning database of eukaryotic genomic sequences has transformed our appreciation for the role of HGT in eukaryotic genomes, but there are still

many outstanding questions: why is there such an apparently strong differential impact of HGT in different lineages and in different genomes within a cell? Why aren't eukaryotic genes transferred to prokaryotes at a higher frequency? Why are ancient transfers apparently less common than more recent ones? In general, we also know little about the mechanisms of HGT, or the behavioural and ecological factors that enhance or discourage it. There are many lineages that need a closer look, or even a first look (BOX 4), for HGT, whereas some lineages (for example, the fungi) are emerging as excellent systems to look at long- and short-term effects of HGT. Looking back at the history of our knowledge of HGT in bacteria, it seems likely that some aspects of HGT in eukaryotes will get increasingly complex and difficult to explain as our sampling grows (for example, the emergence of more complex patterns of distribution with greater sampling). However, looking forward from the many transfers that are already known to be a source of new functions, it is clear that HGT should feature prominently in future studies on the comparative genomics of eukaryotes.

### Note added in proof

A very recent paper<sup>160</sup> reports the first evidence in animals, bdelloid rotifers in particular, for extensive horizontal acquisition of genes from diverse bacterial and eukaryotic sources. The many foreign genes are strongly concentrated, along with diverse transposable elements, in telomeric regions of bdelloid chromosomes, and their uptake might be facilitated by the repeated cycles of desiccation-induced membrane disruption and DNA breakage and repair that occur as part of the bdelloid life style.

- Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129 (1999).
- Eisen, J. A. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* **10**, 606–611 (2000).
- Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742 (2001).
- Kurland, C. G. What tangled web: barriers to rampant horizontal gene transfer. *Bioessays* **27**, 741–747 (2005).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Lawrence, J. G. Gene transfer in bacteria: speciation without species? *Theor. Popul. Biol.* **61**, 449–460 (2002).
- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Salzberg, S. L., White, O., Peterson, J. & Eisen, J. A. Microbial genes in the human genome: lateral transfer or gene loss? *Science* **292**, 1903–1906 (2001).
- Stanhope, M. J. *et al.* Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* **411**, 940–944 (2001).
- Andersson, J. O. in *Genomics and Evolution of Microbial Eukaryotes* (eds Katz, L. A. & Bhattacharya, D.) 109–122 (Oxford Univ. Press, Oxford, 2006).
- Andersson, J. O. Lateral gene transfer in eukaryotes. *Cell. Mol. Life Sci.* **62**, 1182–1197 (2005).
- Richards, T. A., Hirt, R. P., Williams, B. A. & Embley, T. M. Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17–32 (2003).
- Ragan, M. A., Harlow, T. J. & Beiko, R. G. Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends Microbiol.* **14**, 4–8 (2006).
- Roelofs, J. & Van Haastert, P. J. Genes lost during evolution. *Nature* **411**, 1013–1014 (2001).
- Roger, A. J. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146–S163 (1999).
- Andersson, J. O., Hirt, R. P., Foster, P. G. & Roger, A. J. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *BMC Evol. Biol.* **6**, 27 (2006).
- Even among just four gene families, a large number of both bacteria–eukaryote and eukaryote–eukaryote transfers are shown to have accumulated sequentially in the genomes of several diverse lineages of protists.**
- Noble, G. P., Rogers, M. B. & Keeling, P. J. Complex distribution of EFL and EF-1 $\alpha$  proteins in the green algal lineage. *BMC Evol. Biol.* **7**, 82 (2007).
- The overall distribution of the EFL protein suggests several eukaryote–eukaryote HGT events (see reference 87), but the distribution of EFL within green algae is so complex at the fine scale that both transfer and lineage sorting probably contributed to its evolution.**
- Stanier, R. Y. Some aspects of the biology of cells and their possible evolutionary significance. *Symp. Soc. Gen. Microbiol.* **20**, 1–38 (1970).
- Hoffmeister, M. & Martin, W. Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. *Environ. Microbiol.* **5**, 641–649 (2003).
- de Souza, W. & Motta, M. C. Endosymbiosis in protozoa of the Trypanosomatidae family. *FEMS Microbiol. Lett.* **173**, 1–8 (1999).
- Hackstein, J. H. & Vogels, G. D. Endosymbiotic interactions in an anaerobic protozoa. *Antonie Van Leeuwenhoek* **71**, 151–158 (1997).
- Wernegren, J. J. Endosymbiosis: lessons in conflict resolution. *PLoS Biol.* **2**, e68 (2004).
- Slamovits, C. H., Saldarriaga, J. F., Larocque, A. & Keeling, P. J. The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *J. Mol. Biol.* **372**, 356–368 (2007).
- Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
- Kurland, C. G. & Andersson, S. G. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820 (2000).
- Obornik, M. & Green, B. R. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol. Biol. Evol.* **22**, 2343–2353 (2005).
- Richards, T. A. *et al.* Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* **5**, 1517–1531 (2006).
- Patron, N. J. & Keeling, P. J. Common evolutionary origin of starch biosynthetic enzymes in green and red algae. *J. Phycol.* **41**, 1131–1141 (2005).
- Esser, C. *et al.* A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
- Adams, K. L. & Palmer, J. D. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29**, 380–395 (2003).



31. Millen, R. S. *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**, 645–658 (2001).
32. Adams, K. L., Qiu, Y. L., Stoutemyer, M. & Palmer, J. D. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl Acad. Sci. USA* **99**, 9905–9912 (2002). **A large-scale study showing that rates of functional transfer of mitochondrial genes to the nucleus in angiosperms vary enormously, often exceeding rates of even synonymous nucleotide substitutions, and that genes for hydrophobic proteins are far more recalcitrant to functional transfer than are those encoding non-hydrophobic proteins.**
33. Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354–357 (2000).
34. Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**, 72–76 (2003).
35. Thorsness, P. E. & Weber, E. R. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int. Rev. Cytol.* **165**, 207–234 (1996).
36. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5**, 123–135 (2004).
37. Leister, D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* **21**, 655–663 (2005).
38. Matsuo, M., Ito, Y., Yamauchi, R. & Obokata, J. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast–nuclear DNA flux. *Plant Cell* **17**, 665–675 (2005). **Elegant bioinformatic analyses reveal that, in rice at least, plastid DNA sequences are transferred to the nucleus at even higher rates than previously observed but are then lost and fragmented at surprisingly high rates too. Another unexpected finding is that much of this chloroplast–nuclear flux occurs in pericentromeric regions of chromosomes.**
39. Henze, K. & Martin, W. How do mitochondrial genes get into the nucleus? *Trends Genet.* **17**, 383–387 (2001).
40. Choi, C., Liu, Z. & Adams, K. L. Evolutionary transfers of mitochondrial genes to the nucleus in the *Populus* lineage and coexpression of nuclear and mitochondrial *Sdh4* genes. *New Phytol.* **172**, 429–439 (2006).
41. van Dooren, G. C., Su, V., D’Ombra, M. C. & McFadden, G. I. Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J. Biol. Chem.* **277**, 23612–23619 (2002).
42. Hotopp, J. C. *et al.* Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753–1756 (2007). **Many animals harbour bacterial endosymbionts, and this paper shows that many genes, or even whole copies of these endosymbiont genomes, have been transferred to the nuclear genome of the animal host. This paper is also a cautionary tale for the automated removal of ‘contaminant’ bacterial sequence from genome projects of eukaryotes that harbour bacterial symbionts.**
43. Nikoh, N. *et al.* *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res.* **18**, 272–280 (2008).
44. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**, 267 (2006).
45. Tamames, J. *et al.* The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol. Biol.* **7**, 181 (2007).
46. McCutcheon, J. P. & Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl Acad. Sci. USA* **104**, 19392–19397 (2007).
47. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
48. Hall, C., Brachat, S. & Dietrich, F. S. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**, 1102–1115 (2005).
49. Ricard, G. *et al.* Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrate-rich environment. *BMC Genomics* **7**, 22 (2006).
- The rumen gut is an environment that is rich in bacteria–protist interactions, and this study shows that rumen ciliates are taking up bacterial genes at a high frequency.**
50. Doolittle, W. F. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**, 307–311 (1998).
51. Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K. & Keeling, P. J. Lateral gene transfer and the evolution of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa. *J. Mol. Evol.* **63**, 801–814 (2006). **One of the first studies to describe widespread HGT to a nuclear genome, in this case showing that plastid-targeted genes are derived from many different kinds of algae and bacteria.**
52. Watkins, R. F. & Gray, M. W. The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa. *J. Mol. Evol.* **63**, 801–814 (2006).
53. de Koning, A. P., Brinkman, F. S., Jones, S. J. & Keeling, P. J. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**, 1769–1773 (2000).
54. Andersson, J. O., Doolittle, W. F. & Nesbo, C. L. Genomics. Are there bugs in our genome? *Science* **292**, 1848–1850 (2001).
55. Kondrashov, F. A., Koonin, E. V., Morgunov, I. G., Finogenova, T. V. & Kondrashova, M. N. Evolution of glyoxylate cycle enzymes in metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol. Direct* **1**, 31 (2006).
56. Nakashima, K., Yamada, L., Satou, Y., Azuma, J. & Satoh, N. The evolutionary origin of animal cellulose synthase. *Dev. Genes Evol.* **214**, 81–88 (2004).
57. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422 (2005).
58. Douzery, E. J., Snell, E. A., Baptiste, E., Delsuc, F. & Philippe, H. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl Acad. Sci. USA* **101**, 15386–15391 (2004).
59. Andersson, J. O., Sarchfield, S. W. & Roger, A. J. Gene transfers from nanoarchaeota to an ancestor of diplomonads and parabasalids. *Mol. Biol. Evol.* **22**, 85–90 (2005).
60. Andersson, J. O. *et al.* A genomic survey of the fish parasite *Spirionucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution. *BMC Genomics* **8**, 51 (2007). **Dozens of putative HGTs were identified in the genome of one diplomonad, with most of these events pre-dating the divergence of the two major lineages of diplomonads and some apparently pre-dating the divergence of diplomonads and parabasalids. Most transferred genes encode metabolic proteins and were acquired from bacteria, but a significant minority seem to be eukaryote-to-eukaryote transfers.**
61. Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl Acad. Sci. USA* **103**, 11647–11652 (2006).
62. Andersson, J. O., Sjogren, A. M., Davis, L. A., Embley, T. M. & Roger, A. J. Phylogenetic analyses of diplomonad genomes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* **13**, 94–104 (2003).
63. Slamovits, C. H. & Keeling, P. J. Pyruvate-phosphate dikinase of oxymonads and parabasalids and the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes. *Eukaryot. Cell* **5**, 148–154 (2006).
64. Carlton, J. M. *et al.* Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* **315**, 207–212 (2007).
65. Morrison, H. G. *et al.* Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**, 1921–1926 (2007).
66. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868 (2005). **Genome sequencing of anaerobic eukaryotic parasites (also see references 64 and 65) has confirmed the conclusion, which was based on earlier, single-gene studies, that HGT has had a major role in the *ad hoc* assembly of the unusual metabolic pathways of these organisms.**
67. Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57 (2005).
68. Waller, R. F., Slamovits, C. H. & Keeling, P. J. Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein. *Mol. Biol. Evol.* **23**, 1437–1443 (2006).
69. Andersson, J. O. & Roger, A. J. Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions, and prokaryote-to-eukaryote lateral gene transfers. *Eukaryot. Cell* **1**, 304–310 (2002).
70. Rice, D. W. & Palmer, J. D. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* **4**, 31 (2006). **In contrast to plant mitochondria, this large-scale analysis of HGT in plastid genomes revealed very little transfer. One case that was identified, however, is an important, rare character that serves to phylogenetically unite two major groups of algae.**
71. Khan, H. *et al.* Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP 1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol. Biol. Evol.* **24**, 1832–1842 (2007).
72. Bevan, R. B. & Lang, B. F. in *Topics in Current Genetics* Vol. 8. (eds Kohler, C. & Bauer, M. F.) 1–35 (Springer, Berlin, 2004).
73. Bachvaroff, T. R., Concepcion, G. T., Rogers, C. R., Herman, E. M. & Delwiche, C. F. Dinoflagellate expressed sequence tags data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* **155**, 65–78 (2004).
74. Keeling, P. J. & Palmer, J. D. Lateral transfer at the gene and subgenomic levels in the evolution of eukaryotic enolase. *Proc. Natl Acad. Sci. USA* **98**, 10745–10750 (2001).
75. Patron, N. J., Waller, R. F. & Keeling, P. J. A tertiary plastid uses genes from two endosymbionts. *J. Mol. Biol.* **357**, 1373–1382 (2006). **Some dinoflagellates have a complex history of replacing their plastids with new ones from other lineages. Here it is shown that genes derived from both plastids can be retained in the nuclear genome and their products also targeted to the new plastid, making its proteome chimeric.**
76. Rogers, M. B. & Keeling, P. J. Lateral gene transfer and re-compartmentalisation of Calvin cycle enzymes in plants and algae. *J. Mol. Evol.* **58**, 367–375 (2003).
77. Hackett, J. D. *et al.* Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr. Biol.* **14**, 213–218 (2004).
78. Waller, R. F., Patron, N. J. & Keeling, P. J. Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. *Int. J. Syst. Evol. Microbiol.* **56**, 1439–1347 (2006).
79. Richards, T. A., Dacks, J. B., Jenkinson, J. M., Thornton, C. R. & Talbot, N. J. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr. Biol.* **16**, 1857–1864 (2006). **Fungal and oomycete plant pathogens are distantly related but have many mechanistic similarities, and here it is shown that several convergent characteristics of their pathogenicity and feeding might be due to transfer of genes between them.**
80. Friesen, T. L. *et al.* Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genet.* **38**, 953–956 (2006). **Virulence-related factors have been observed to transfer in several systems, including fungi, but in this case a virulence factor has moved from one fungus to another so recently that it can be dated to around 1941, when it led to the emergence of a new disease.**
81. Inderbitzin, P., Harkness, J., Turgeon, B. G. & Berbee, M. L. Lateral transfer of mating system in *Stemphylium*. *Proc. Natl Acad. Sci. USA* **102**, 11390–11395 (2005).
82. Slot, J. C. & Hibbett, D. S. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS ONE* **2**, e1097 (2007).



83. Simpson, A. G., Perley, T. A. & Lara, E. Lateral transfer of the gene for a widely used marker, alpha-tubulin, indicated by a multi-protein study of the phylogenetic position of *Andalucia* (Excavata). *Mol. Phylogenet. Evol.* **47**, 366–377 (2008).
84. Paoletti, M., Buck, K. W. & Brasier, C. M. Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novulmi*. *Mol. Ecol.* **15**, 249–262 (2006).
85. Hall, C. & Dietrich, F. S. The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* **177**, 2293–2307 (2007).
86. Kavanaugh, L. A., Fraser, J. A. & Dietrich, F. S. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Mol. Biol. Evol.* **23**, 1879–1890 (2006).
87. Keeling, P. J. & Inagaki, Y. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc. Natl Acad. Sci. USA* **101**, 15380–15385 (2004).
88. Rogers, M. B. *et al.* A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. *BMC Evol. Biol.* **7**, 89 (2007). **There are many examples of relatively simple bacteria–eukaryote HGT events, but in this paper broader sampling of three cases shows that complex, multiple transfer events might often seem to be simpler than they really are, and that transfer between eukaryotes might be more common than we think.**
89. Nosenko, T. & Bhattacharya, D. Horizontal gene transfer in chromalveolates. *BMC Evol. Biol.* **7**, 173 (2007).
90. Alvarez, N. *et al.* Phylogeographic support for horizontal gene transfer involving sympatric bruchid species. *Biol. Direct* **1**, 21 (2006).
91. Jenkins, C. *et al.* Genes for the cytoskeletal protein tubulin in the bacterial genus *Prostheco bacter*. *Proc. Natl Acad. Sci. USA* **99**, 17049–17054 (2002). **Tubulins are hallmark proteins of the eukaryotic cytoskeleton, so it was surprising to find that this bacterial genome had acquired both alpha- and beta-tubulin genes by HGT, and that the proteins seem to retain some structural function.**
92. Schlieper, D., Oliva, M. A., Andreu, J. M. & Lowe, J. Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc. Natl Acad. Sci. USA* **102**, 9170–9175 (2005).
93. Sontag, C. A., Staley, J. T. & Erickson, H. P. *In vitro* assembly and GTP hydrolysis by bacterial tubulins BtubA and BtubB. *J. Cell Biol.* **169**, 233–238 (2005).
94. Guljamov, A. *et al.* Horizontal gene transfer of two cytoskeletal elements from a eukaryote to a cyanobacterium. *Curr. Biol.* **17**, R757–R759 (2007). **Similar to reference 91, the cytoskeletal protein actin was thought to be unique to eukaryotes, but this cyanobacterium was found to have taken up genes for actin and actin accessory proteins, which are now used in a structural capacity.**
95. Rogers, M. B., Patron, N. J. & Keeling, P. J. Horizontal transfer of a eukaryotic plastid-targeted protein gene to cyanobacteria. *BMC Biol.* **5**, 26 (2007). **Many eukaryotic genes are derived from the plastid endosymbiont, but here a eukaryotic gene that took over plastid function has also been transferred back to cyanobacteria, where it integrated into the genome adjacent to its bacterial analogue.**
96. Gamielidien, J., Ptiitsyn, A. & Hide, W. Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. *Trends Genet.* **18**, 5–8 (2002).
97. Kinsella, R. J. & McInerney, J. O. Eukaryotic genes in *Mycobacterium tuberculosis*? Possible alternative explanations. *Trends Genet.* **19**, 687–689 (2003).
98. Huang, J. & Gogarten, J. P. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* **22**, 361–366 (2006).
99. Stechmann, A., Baumgartner, M., Silberman, J. D. & Roger, A. J. The glycolytic pathway of *Trimastix pyriformis* is an evolutionary mosaic. *BMC Evol. Biol.* **6**, 101 (2006).
100. Henze, K. *et al.* Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*. *Gene* **281**, 123–131 (2001).
101. Ledent, V. & Vervoort, M. Comparative genomics of the class 4 histone deacetylase family indicates a complex evolutionary history. *BMC Biol.* **4**, 24 (2006).
102. Patron, N. J., Inagaki, Y. & Keeling, P. J. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr. Biol.* **17**, 887–891 (2007).
103. Hackett, J. D. *et al.* Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol. Biol. Evol.* **24**, 1702–1713 (2007).
104. Huang, J., Xu, Y. & Gogarten, J. P. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol. Biol. Evol.* **22**, 2142–2146 (2005).
105. Patron, N. J., Rogers, M. B. & Keeling, P. J. Gene replacement of fructose-1,6-bisphosphate aldolase (FBA) supports a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell* **3**, 1169–1175 (2004).
106. Berney, C. & Pawlowski, J. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. Biol. Sci.* **273**, 1867–1872 (2006).
107. Hao, W. & Golding, G. B. Patterns of bacterial gene movement. *Mol. Biol. Evol.* **21**, 1294–1307 (2004).
108. Hao, W. & Golding, G. B. The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res.* **16**, 636–643 (2006).
109. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806 (1999).
110. Keeling, P. J. The diversity and evolutionary history of plastids and their hosts. *Am. J. Bot.* **91**, 1481–1493 (2004).
111. Reyes-Prieto, A., Weber, A. P. & Bhattacharya, D. The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**, 147–168 (2007).
112. Huang, J. & Gogarten, J. P. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* **8**, R99 (2007).
113. Stiller, J. W. Plastid endosymbiosis, genome evolution and the origin of green plants. *Trends Plant Sci.* **12**, 391–396 (2007).
114. Archibald, J. M. Jumping genes and shrinking genomes — probing the evolution of eukaryotic photosynthesis with genomics. *IUBMB Life* **57**, 539–547 (2005).
115. Ishida, K. & Green, B. R. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc. Natl Acad. Sci. USA* **99**, 9294–9299 (2002).
116. Cavalier-Smith, T. In *Organelles, Genomes and Eukaryotic Evolution* (ed. Horner, D.) 71–103 (Taylor and Francis, London, 2004).
117. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18**, 418–426 (2001).
118. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
119. Eisen, J. A. *et al.* Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* **4**, e286 (2006).
120. Tyler, B. M. *et al.* *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–1266 (2006).
121. Aury, J. M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178 (2006).
122. Huang, J. *et al.* Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol.* **5**, R88 (2004).
123. Moore, R. B. *et al.* A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959–963 (2008).
124. Hannaert, V. *et al.* Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl Acad. Sci. USA* **100**, 1067–1071 (2003).
125. El-Sayed, N. M. *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409–415 (2005).
126. Pfanner, N. & Geissler, A. Versatility of the mitochondrial protein import machinery. *Nature Rev. Mol. Cell Biol.* **2**, 339–349 (2001).
127. Bhattacharya, D., Archibald, J. M., Weber, A. P. & Reyes-Prieto, A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1239–1246 (2007).
128. Theissen, U. & Martin, W. The difference between organelles and endosymbionts. *Curr. Biol.* **16**, R1016–1017; author reply R1017–R1018 (2006).
129. McFadden, G. I. Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* **37**, 951–959 (2001).
130. de Grey, A. D. Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity? *Bioessays* **27**, 436–446 (2005).
131. Daley, D. O. & Whelan, J. Why genes persist in organelle genomes. *Genome Biol.* **6**, 110 (2005). **This is a balanced and insightful review of the many hypotheses to account for the persistence of certain genes in organelle genomes, with the importability hypothesis being the best-supported hypothesis.**
132. Daley, D. O., Clifton, R. & Whelan, J. Intracellular gene transfer: reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **99**, 10510–10515 (2002).
133. Allen, J. F., Puthiyaveetil, S., Strom, J. & Allen, C. A. Energy transduction anchors genes in organelles. *Bioessays* **27**, 426–435 (2005).
134. Burger, G., Gray, M. W. & Lang, B. F. Mitochondrial genomes: anything goes. *Trends Genet.* **19**, 709–716 (2003).
135. Green, B. R. The chloroplast genome of dinoflagellates — a reduced instruction set? *Protist* **155**, 23–31 (2004).
136. Funes, S. *et al.* The typically mitochondrial DNA-encoded ATP6 subunit of the F1F0-ATPase is encoded by a nuclear gene in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **277**, 6051–6058 (2002).
137. Perez-Martinez, X. *et al.* Subunit II of cytochrome c oxidase in *Chlamydomonas* algae is a heterodimer encoded by two independent nuclear genes. *J. Biol. Chem.* **276**, 11302–11309 (2001).
138. Richardson, A. O. & Palmer, J. D. Horizontal gene transfer in plants. *J. Exp. Bot.* **58**, 1–9 (2007).
139. Koulintchenko, M., Konstantinov, Y. & Dietrich, A. Plant mitochondria actively import DNA via the permeability transition pore complex. *EMBO J.* **22**, 1245–1254 (2003).
140. Arimura, S., Yamamoto, J., Aida, G. P., Nakazono, M. & Tsutsumi, N. Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proc. Natl Acad. Sci. USA* **101**, 7805–7808 (2004).
141. Sheahan, M. B., McCurdy, D. W. & Rose, R. J. Mitochondria as a connected population: ensuring continuity of the mitochondrial genome during plant cell dedifferentiation through massive mitochondrial fusion. *Plant J.* **44**, 744–755 (2005).
142. Mower, J. P., Stefanovic, S., Young, G. J. & Palmer, J. D. Plant genetics: gene transfer from parasitic to host plants. *Nature* **432**, 165–166 (2004).
143. Won, H. & Renner, S. S. Horizontal gene transfer from flowering plants to *Gnetum*. *Proc. Natl Acad. Sci. USA* **100**, 10824–10829 (2003).
144. Davis, C. C., Anderson, W. R. & Wurdack, K. J. Gene transfer from a parasitic flowering plant to a fern. *Proc. Biol. Sci.* **272**, 2237–2242 (2005).
145. Berghthorsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**, 197–201 (2003).
146. Barkman, T. J. *et al.* Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol. Biol.* **7**, 248 (2007).
147. Berghthorsson, U., Richardson, A. O., Young, G. J., Goertzen, L. R. & Palmer, J. D. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc. Natl Acad. Sci. USA* **101**, 17747–17752 (2004). **A remarkable case in which a flowering plant mitochondrial genome possesses as many foreign genes, acquired from a diverse range of land plant mitochondrial donors, as it does native genes.**

148. Nickrent, D. L., Blarer, A., Qiu, Y. L., Vidal-Russell, R. & Anderson, F. E. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. *BMC Evol. Biol.* **4**, 40 (2004).
149. Davis, C. C. & Wurdack, K. J. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305**, 676–678 (2004).
150. Sanchez-Gracia, A., Maside, X. & Charlesworth, B. High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet.* **21**, 200–203 (2005).
151. Haugen, P., Simon, D. M. & Bhattacharya, D. The natural history of group I introns. *Trends Genet.* **21**, 111–119 (2005).
152. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38**, 1–35 (2004).
153. Cho, Y., Qiu, Y. L., Kuhlman, P. & Palmer, J. D. Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl Acad. Sci. USA* **95**, 14244–14249 (1998).
154. Wei, W. *et al.* Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl Acad. Sci. USA* **104**, 12825–12830 (2007).
155. Woolfit, M., Rozpedowska, E., Piskur, J. & Wolfe, K. H. Genome survey sequencing of the wine spoilage yeast *Dekkera (Brettanomyces) bruxellensis*. *Eukaryot. Cell* **6**, 721–733 (2007).
156. Fukami, H., Chen, C. A., Chiou, C. Y. & Knowlton, N. Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox 1* gene of scleractinian corals. *J. Mol. Evol.* **64**, 591–600 (2007).
157. Goddard, M. R., Leigh, J., Roger, A. J. & Pemberton, A. J. Invasion and persistence of a selfish gene in the Cnidaria. *PLoS ONE* **1**, e3 (2006).
158. Keeling, P. J. *et al.* The tree of eukaryotes. *Trends Ecol. Evol.* **20**, 670–676 (2005).
159. Burki, F. *et al.* Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* **2**, e790 (2007).
160. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320**, 1210–1213 (2008).

## Acknowledgements

The authors wish to thank numerous colleagues for supplying unpublished data and several of the photos in Figure 2 (see the figure for photo credits). P.J.K.'s research on HGT is supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Tula Foundation, and he is a Fellow of the Canadian Institute for Advanced Research. J.D.P.'s research on HGT is supported by a grant from the National Institutes of Health and by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc.

## FURTHER INFORMATION

The Keeling laboratory: <http://www.botany.ubc.ca/keeling>  
 Jeffrey D. Palmer's homepage: <http://www.bio.indiana.edu/facultyresearch/faculty/Palmer.html>

## SUPPLEMENTARY INFORMATION

See online article: [S1](#) (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF