Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution

Gerald Schönknecht[1)]*, *Andreas P. M. Weber*[2)3)] *and Martin J. Lercher*[3)4)]

In contrast to vertical gene transfer from parent to offspring, horizontal (or lateral) gene transfer moves genetic information between different species. Bacteria and archaea often adapt through horizontal gene transfer. Recent analyses indicate that eukaryotic genomes, too, have acquired numerous genes via horizontal transfer from prokaryotes and other lineages. Based on this we raise the hypothesis that horizontally acquired genes may have contributed more to adaptive evolution of eukaryotes than previously assumed. Current candidate sets of horizontally acquired eukaryotic genes may just be the tip of an iceberg. We have recently shown that adaptation of the thermoacidophilic red alga *Galdieria sulphuraria* to its hot, acid, toxic-metal laden, volcanic environment was facilitated by the acquisition of numerous genes from extremophile bacteria and archaea. Other recently published examples of horizontal acquisitions involved in adaptation include ice-binding proteins in marine algae, enzymes for carotenoid biosynthesis in aphids, and genes involved in fungal metabolism.

Keywords:
adaptation; evolution; genome; horizontal gene transfer; phylogeny

Introduction

Novel traits in eukaryotes are thought to mainly evolve via gene duplication followed by mutations that functionally modify the duplicate (neofunctionalization) [1, 2]. This route to novelty is a slow, incremental process, often requiring the accumulation of mutations over many generations. In contrast, a very important evolutionary mechanism in bacteria and archaea is horizontal gene transfer (HGT, also termed lateral gene transfer), i.e. the genomic integration of genetic material originating from another species [3–5]. HGT was originally discovered via the observation of rapid emergence of drug resistance in *Shigella* strains [6], and the spread of antibiotic resistance remains a paradigm for bacterial HGT. By effectively combining the information from numerous genomes that coexist in overlapping habitats, HGT makes prokaryotic evolution a massively parallel process. HGT facilitates the immediate acquisition of sequences already optimized for functions previously unknown to the receiving genome, and thus allows for rapid and drastic phenotypic changes. In each HGT event, at most a few genes are transmitted between species [4]. HGT thus favors genes without complex interactions, frequently copying operational genes rather than genes involved in the processing of genetic information [7, 8].

Genes can be transferred horizontally between organisms across the different domains of life – between archaea and bacteria, but also from archaea and bacteria into eukaryotes. Evidence for eukaryotic functional innovation via HGT from free-living bacteria or archaea has only started to accumulate recently. Yet it has long been known that eukaryotic nuclear genomes harbor massive amounts of genes originating from organelles. Mitochondria descended from an alphaproteobacterial endosymbiont at the origin of eukaryotes [9], while the plastids of photosynthetic eukaryotes

DOI 10.1002/bies.201300095

[1)] Department of Botany, Oklahoma State University, Stillwater, OK, USA
[2)] Institute of Plant Biochemistry, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
[3)] Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
[4)] Institute for Computer Science, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

***Corresponding author:**
Gerald Schönknecht
E-mail: gerald.schoenknecht@okstate.edu

Abbreviation:
HGT, horizontal gene transfer.

descended from a later cyanobacterial endosymbiont. The evolution from endosymbionts into organelles was accompanied by massive transfers of genes to the nucleus of the host cell, a process termed endosymbiotic gene transfer [10]. Further bursts of endosymbiotic gene transfer followed the secondary and higher order endosymbioses that gave rise to the diversity of algae [11]. While endosymbiosis is quite frequent – examples being alga living in corals, and nitrogen-fixing bacteria in root nodules of legumes – the endosymbiotic origin of cellular organelles is an extremely rare event. Only the endosymbiotic origin of organelles was accompanied by "massive" endosymbiotic gene transfer, while more generally, genome reduction in endosymbiotic bacteria does not seem to be accompanied by large-scale gene transfer to the eukaryotic host [12]. Endosymbiotic gene transfer facilitated the coordinated unidirectional move of hundreds or even thousands of genes from the evolving organelle to the host genome. Because of this, endosymbiotic organelle formation was capable of transferring large, complex functional systems, such as oxygenic photosynthesis, across domains of life, thereby facilitating major evolutionary transitions [13]. The tremendous impact of endosymbiotic gene transfer on the evolution of eukaryotes is well established. But how important was the recurrent horizontal acquisition of single or a few genes in eukaryotic adaptation?

Early reports of HGT in eukaryotes date back to more than 20 years ago, but many of those early candidates were not confirmed by later work [14, 15]. Similarly, the claim that more than 100 genes in the human genome were acquired horizontally from bacteria [16] was quickly corrected [17, 18]. Many of these early false positives were caused by a bias in available genome databases, which were dominated by bacteria and contained few eukaryotic model organisms. Meanwhile, numerous eukaryotic genomes have been sequenced; the wider taxa coverage facilitates largely unbiased analyses, resulting in a rapidly increasing number of well supported cases of HGT into eukaryotes [19–21]. We recently reported the genome of the thermoacidophilic red alga *Galdieria sulphuraria* and found evidence that adaptation to its hot, toxic metal-rich, acidic environment was facilitated by HGT from various bacteria and archaea [22]. Here, *G. sulphuraria* will be used as a reference, since a relative large number of 337 HGT candidates, a thorough phylogenetic analysis including statistical tests, and obvious adaptive benefits of many HGT candidates make this organism a good model to study HGT in eukaryotes.

The "gold standard" for establishing that a certain gene has been acquired by horizontal transfer from a different clade is phylogenetic incongruence, where an evolutionary tree for a specific protein family clearly differs from the established organismal phylogeny [18] (where the reliability of this approach depends on sequence alignment quality [23]). For example, a tree of kynurenine formamidases (Fig. 1A) shows the protein sequence from *G. sulphuraria* embedded within bacterial sequences. The most parsimonious explanation for this is HGT from a bacterium. Other approaches to identify HGT candidates and possible pitfalls are described in Boxes 1 and 2.

How many HGT candidates were detected in eukaryotic genomes?

Table 1 compares phylogenetic screens of complete eukaryotic genomes where each HGT candidate is supported by an evolutionary tree. The percentage of genes originating from HGT ranges from 0.035% to 9.6%, spanning more than two orders of magnitude. This breadth probably reflects different evolutionary

Box 1

How are HGT candidates identified?

Higher BLAST scores for distant than for closer relatives, or the patchy distribution of a protein family, can be starting points to identify HGT candidates (see Table 1), but must be confirmed by a thorough analysis of the evolutionary history of each HGT candidate. Such analyses reconstruct an evolutionary tree for a given gene family and look for statistically supported incongruence between the tree and the established species phylogeny. However, phylogenetic analyses by evolutionary trees are also not fail-safe. Even in the absence of HGT, evolutionary trees based on a single gene or protein family often give rise to incongruences [24], e.g. because of statistical noise in the evolutionary substitution process, model mis-specification, or long-branch attraction artifacts. In some cases, initial indications of HGT disappeared when phylogenetic analyses were repeated with larger sample sizes, such as for glycosyl hydrolase family 9 (GHF9) cellulases [25].

Genes that have been acquired by HGT can display a base composition that differs from the rest of the genome, reflecting the base composition of the donor genome. As an alternative to phylogenetic approaches, parametric methods thus use deviations in GC content, oligonucleotide frequencies, or codon usage to identify HGT candidates [26, 27]. While deviations in base composition are hard to detect for short DNA fragments such as single genes, parametric methods can outperform phylogenetic methods in detecting the recent transfer of large DNA fragments. Parametric methods are more sensitive for the detection of recent than of ancient HGTs, because the alien base composition of horizontally acquired genes ameliorates over time [3]. The large heterogeneity of GC content within many eukaryotic genomes limits the application of this method to the identification of horizontal gene acquisition in eukaryotes.

Each approach to identify HGT candidates in eukaryotic genomes has its drawbacks, and none is fail save. Combining several orthogonal approaches – such as phylogenetic and parametric analyses – therefore seems to be the best strategy to prevent erroneous conclusions.

Box 2

Possible pitfalls in the identification of HGT candidates

Why is it so difficult to reliably identify genes that originate from HGT, and how can the occurrence of false positives be prevented, or at least minimized? A rather trivial source of erroneously called HGT events are DNA contaminations resulting from bacteria and other microorganisms that live in close association with an organism from which DNA is isolated for sequencing. Contamination artifacts may also arise in the lab, either from research staff [28] or from experimental protocols involving other species; e.g. the inference of HGT from salmonids to schistosomes has been shown to result from DNA contamination with salmon DNA used as carrier material [29].

While bioinformatics approaches can filter out contaminations after sequencing, this might instead result in the removal of true positives. However, a range of analyses can be carried out to control for contamination with foreign DNA. HGT candidates in genomes of eukaryotes as different as unicellular algae [22, 30] and nematodes [31] usually originate from a wide variety of different clades of Archaea, Bacteria, or non-related Eukaryota. Multiple, non-homologous HGT candidates originating from the same clade of Archaea or Bacteria can point to a single source of contamination, warranting further examination. All HGT candidates should be mapped onto the assembled genome to test for clustering of HGT candidates (a task more difficult with recent sequencing methods that produce short reads); clustering would be expected when sequencing stretches of contaminating DNA, although in principle it might also arise as a result of co-transfer of several genes. In *G. sulphuraria* [22] and other eukaryotes [30, 31], the vast majority of HGT candidates is located on large scaffolds, where they are flanked by genes not identified as HGT candidates. This indicates that these HGT candidates are indeed part of the *eukaryotic* genome and are not the result of contaminating DNA. Eukaryotic genes that were acquired horizontally from bacteria or archaea can obtain introns after the acquisition [22, 31, 32]; the presence of introns in HGT candidates (often at a lower frequency) also rules out bacterial contamination artifacts.

time scales covered, different criteria applied in the phylogenetic screens, and different rates of horizontal gene acquisition in different lineages. The evolutionary time scale is expected to have a particularly strong effect. In *G. sulphuraria*, only proteins in Cyanidiophyceae (*G. sulphuraria* and *Cyanidioschyzon merolae*) that had no homologs (i.e. members of the same protein family) in other eukaryotes were analyzed further [22]. Since Cyanidiophyceae diverged within the red algae (Rhodophyta) more than a billion years ago [33, 34], *G. sulphuraria* may have accumulated its 337 HGT candidates over more than a billion years. In contrast, the 43 HGT candidates detected the silkworm (*Bombyx mori*) genome were probably all acquired after speciation of Bombycoidea [35], less than 100 million years ago. A phylogenomic analysis of the diatom *Phaeodactylum tricornutum* (see legend Table 1) nicely demonstrates the accumulation of HGT candidates over time. Of 222 HGT candidates detected in *P. tricornutum*, 122 have homologs in the diatom *Thalassiosira pseudonana*, and 28 have homologs in the oomycete *Phytophthora* spp. [30]. This means that at least 13% of HGT candidates in *P. tricornutum* are ancient and were acquired before the divergence of photosynthetic stramenopiles (including diatoms, brown algae, etc.) and oomycota, almost one billion years ago [34], and at least 55% were acquired before the diatoms *P. tricornutum* and *T. pseudonana* diverged around 195 million years ago [36]. This accumulation of HGT candidates over time is one possible explanation for different percentages of HGT candidates reported for nematode genomes (Table 1). For *Caenorhabditis elegans* (1.8%), nematode-specific HGT candidates were selected by removing proteins with BLAST hits in other metazoan clades from further consideration [37]. For *Meloidogyne incognita* (0.26%), root-knot nematode-specific HGT candidates were selected by removing proteins with BLAST hits in other metazoa or in four other nematode genomes from further consideration [31]. For *Bursaphelenchus xylophilus* (0.13%), superfamily-specific HGT candidates were selected by removing proteins with BLAST hits in any nematode outside the superfamily of Aphelenchoidoidea (or in other metazoa) [38]. As expected, limiting screens to more recent timescales results in smaller numbers of HGT candidates.

The stringency of a phylogenetic screen also has a large effect on the number of HGT candidates detected. High specificity, i.e. a low number of false positives, inevitably results in low sensitivity, i.e. a larger number of false negatives. In a recent, detailed, HGT screen of the red alga *Porphyridium cruentum*, 86 genes fulfilled very stringent criteria, being supported by trees with >40 terminal taxa and bootstrap values >90%, while 266 candidates were supported by trees with >4 terminal taxa and bootstrap values >50% [39]. Our highly stringent screen for HGT candidates in the genome of *G. sulphuraria* missed six out of 18 HGT candidates that had been identified during detailed, manual phylogenetic analyses of some protein families, such as the periplasmic metal binding protein shown in Fig. 1B (Gasu_41950). Disparate methods (see legend Table 1), identifying different subsets of HGT candidates, may explain the surprisingly small overlap seen in cases where the same genome has been screened twice. For example, a first HGT screen (limited to proteins that had no BLAST hits in eukaryotes) of the amoeba *Dictyostelium discoideum* identified 29 candidate genes resulting from 18 transfer events from bacteria [40]. A later screen searching for protein families with patchy distribution identified 50 candidate genes from 26 transfer events of non-amoebozoan origin [41], but recovered only one of the 18 HGT events from

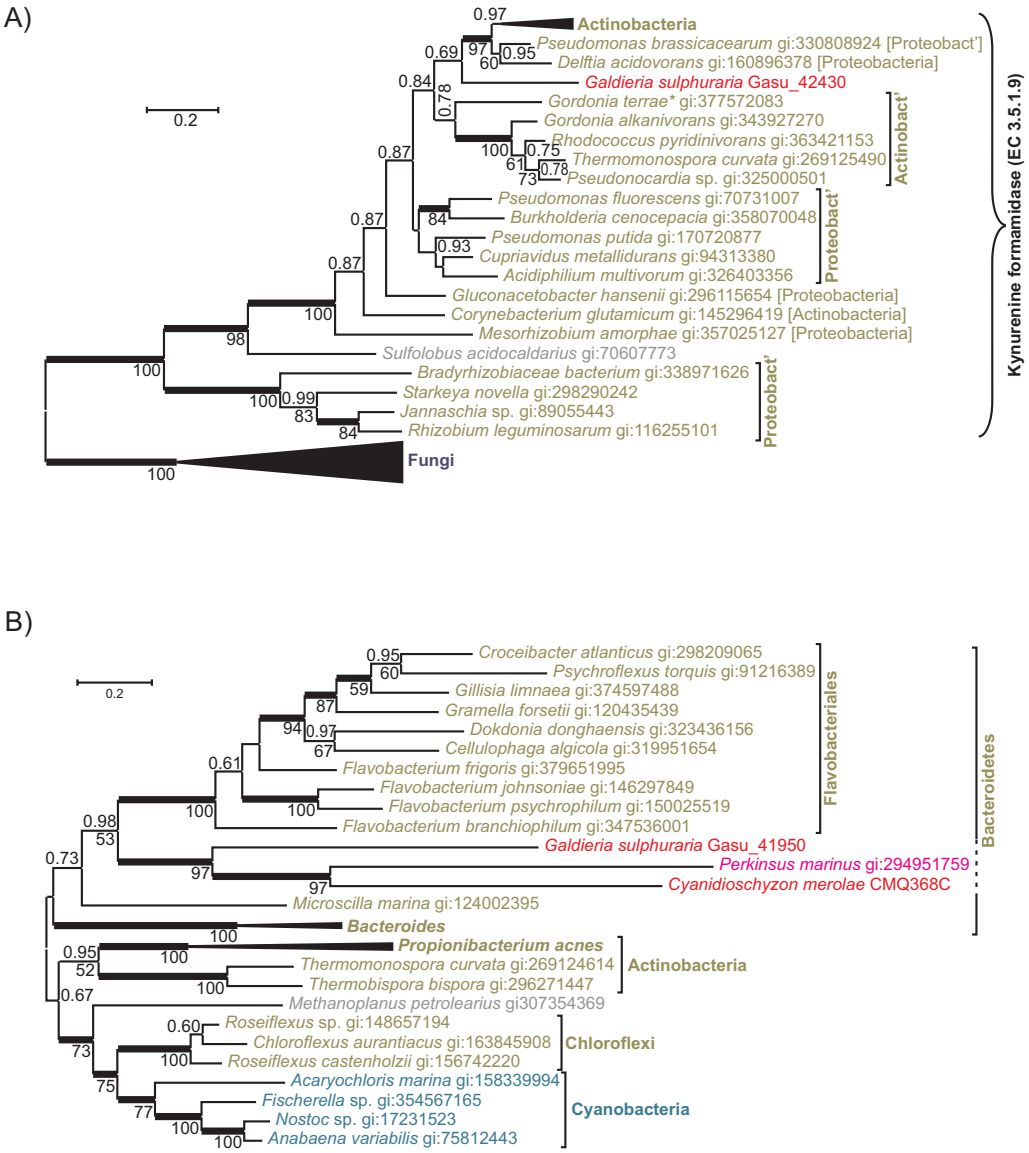

Figure 1. Phylogenetic trees showing horizontal gene transfer (HGT) from bacteria into eukaryotic genomes. **A:** Phylogenetic tree of a family of kynurenine formamidases (EC 3.5.1.9) and homologous cyclases. The kynurenine formamidase in *G. sulphuraria* has high similarity to bacterial kynurenine formamidases but no similarity to the non-orthologous kynurenine formamidases in eukaryotes, which are not cyclases but α/β hydrolase fold enzymes with an esterase/lipase domain. A HGT from Proteobacteria into the archaeon *Sulfolobus acidocaldarius* is seen as well. **Gordonia terrae* has been renamed into *Rhodococcus ruber*. The unrooted Bayesian [92] tree calculated with a WAG + I + G model of protein evolution shows posterior probabilities above the branches and PhyML [93] percent bootstrap support (using LG + I + G) below the branches. Curly bracket indicates annotation according to NCBI's Conserved Domain Database [94, 95]. **B:** Phylogenetic tree of periplasmic metal binding proteins, showing an ancient HGT. The common ancestor of *G. sulphuraria* and *C. merolae*, which diverged around 900 million years ago [22], acquired a gene encoding a periplasmic metal binding protein from a member of Bacteroidetes. *Perkinsus marinus*, even though non-photosynthetic, belongs to a lineage that acquired a plastid via secondary endosymbiosis from a red alga [11], resulting in endosymbiotic gene transfer. A single sequence from an archaeon, *Methanoplanus petrolearius*, indicates another HGT from Bacteria to Archaea. The unrooted Bayesian [92] tree calculated with a WAG + I + G model of protein evolution shows posterior probabilities above the branches and PhyML [93] percent bootstrap support (using LG + I + G + F) below the branches. Thick branches indicate 1.0 posterior probability. For clarity some sub-branches have been collapsed (elongated triangles). Color coding of major phylogenetic groups: Bacteria, Cyanobacteria, Archaea, Rhodophyta, Alveolata, Fungi. Bars represent 0.2 changes per site.

the first screen! Similarly, a screen of the *Cryptosporidium parvum* genome identified 24 HGT candidates originating from prokaryotes [42]. A later screen limited to genes encoding metabolic enzymes identified 12 HGT candidates, five of which were identical to HGT candidates from the earlier screen.

Knowledge of the frequencies of HGT in different eukaryotic lineages would be of great interest. Some studies that used identical methodologies to analyze several species indicated large differences in gene acquisition frequencies across species [37, 39]. The majority of HGT screens applied different methods and criteria to pre-select possible HGT candidates before evolutionary trees were constructed (see legend

Table 1. Eukaryotic genomes that were screened for HGT candidates

Species	Group	#HGT	From	# Genes	%	Refs.
Naegleria gruberi	Percolozoa (Excavata)	45	Prokaryotes	15,727	0.29%	[96]
Trichomonas vaginalis	Metamonada (Excavata)	152	Prokaryotes	60,000	0.25%	[97]
Cryptosporidium parvum	Apicomplexa (Aleveolata)	24	Prokaryotes	5,519	0.43%	[42]
Emiliania huxleyi	Haptophyte	77	Bacteria and Viruses	30,569	0.25%	[98]
Phaeodactylum tricornutum (diatom)	Bacillariophyta	222	Prokaryotes	10,402	2.1%	[30]
Galdieria sulphuraria	Rhodophyta	337	Prokaryotes	6,623	5.1%	[22]
Cyanidioschyzon merolae	Rhodophyta	51	Prokaryotes	5,331	1.0%	[39]
Porphyridium purpureum	Rhodophyta	144	Prokaryotes	8,355	1.7%	[39]
Bathycoccus prasinos	Chlorophyta	108	Non-Viridiplanta	7,847	1.4%	[99]
Physcomitrella patens (moss)	Streptophyta	128	Prokaryotes, Fungi, Viruses	35,938	0.36%	[91]
60 fungal genomes	Fungi		Prokaryotes		0.12%	[32]
Hydra magnipapillata	Cnidaria (Metazoa)	71	Bacteria	20,000	0.36%	[100]
Adineta ricciae (bdelloid rotifer)	Rotifera (Metazoa)	2,771	Non-metazoan	28,922	9.6%	[37]
Caenorhabditis elegans (nematode)	Nematoda (Metazoa)	198	Non-metazoan	11,168	1.8%	[37]
Meloidogyne incognita (nematode)	Nematoda (Metazoa)	52	Non-metazoan	20,359	0.26%	[31]
Bursaphelenchus xylophilus (nematode)	Nematoda (Metazoa)	24	Bacteria and Fungi	18,074	0.13%	[38]
Bombyx mori (silkworm)	Arthropoda (Metazoa)	43	Bacteria and Plant	14,623	0.29%	[35]
Acyrthosiphon pisum (aphid)	Arthropoda (Metazoa)	12	Bacteria	33,816	0.035%	[12]
Ciona intestinalis (sea squirt)	Chordata (Metazoa)	92	"Alga" and Cyanobacteria	14,002	0.66%	[90]
Monosiga brevicollis	Choanoflagellate	103	"Alga" and Cyanobacteria	9,200	1.12%	[101]
Dictyostelium discoideum[1]	Amoebozoa	29	Bacteria	12,500	0.23%	[40]
Dictyostelium discoideum[2]	Amoebozoa	50	Non-amoebozoa	13,522	0.37%	[41]
Entamoeba histolytica	Amoebozoa	96	Prokaryotes	9,938	0.97%	[102]

For each species (left), the table lists: the systematic group to which it belongs; the number of detected genes resulting from HGT, #HGT; the groups that were considered as "donors" for the HGT; the number of annotated genes in the genome of the species; and the percentage of genes resulting from HGT. Only studies screening completely sequenced genomes and supporting each reported HGT candidate by an evolutionary tree were included. The studies used a variety of different phylogenetic methods, alone or in combination. Some studies used a 1-step approach, applying a phylogenomic analysis to construct an evolutionary tree for each annotated protein; other studies employed a 2-step algorithm by first applying a pre-selection step (in most cases BLAST-based) to identify possible HGT candidates, which were then tested further by constructing evolutionary trees. 1-Step: For the genomes of *P.t.*, *C.m.*, *P.p.*, and *B.p.*, phylogenomic analyses were used, and resulting evolutionary trees were screened for topologies indicative of HGT [18]. Of the 587 HGT candidates reported for the *P.ct*. genome, only 222 with a bootstrap support $\geq$75% and not originating from cyanobacteria are included here. For *C.m.* and *P.p.*, #HGT given here have bootstrap support $\geq$70% and $N \geq 20$. For the *B.p.* genome, #HGT were reported for bootstrap support $\geq$80%. 2-Step: In all other genome screens, a pre-selection was performed to reduce the number of evolutionary trees to be constructed. The methods used for this pre-selection vary widely. Phylogenomic analyses (using PyPhy) were combined with best BLAST hits in prokaryotes (*T.v.*, *C.p.*, and *E.hi.*). Phylogenomic analyses (using PhyloGenie) were used to identify protein families with patchy distribution (*D.d.*[2]), or were combined with special software (Darkhorse) to identify HGT candidates (*M.b.*). Special software (AlienG) to identify HGT candidates was used (*P.p.* and *C.i.*). BLAST was used to pre-select HGT candidates that had no BLAST hits in eukaryotes (*N.g.* and *D.d.*[1]), or no BLAST hits in a defined eukaryotic clade were allowed (*B.x.*); or the best BLAST hit had to be in prokaryotes (*E.hu.* and *G.s.*), a certain distribution of BLAST hits in different clades was required ($<$10 BLAST hits in 60 fungal genomes, $>$30 BLAST hits in prokaryotes, no BLAST hits in other eukaryotes, or $\geq$50% of 10 best BLAST hits in non-metazoans for *M.i.*). BLAST scores inside the clade to be screened and outside the clade were compared to calculate an "alien index" ($>$30 for *H.m.*), or calculate a "HGT index" ($>$30 for *A.r.* and *C.e.*). Other studies used a combination of different BLAST searches (*B.m.* and *A.p.*). As a result of these different methods, some pre-selections yielded large numbers of possible HGT candidates, which were significantly reduced by the following analyses based on evolutionary trees (*N.g.*, *T.v.*, *C.p.*, *E.hu.*, *G.s.*, *P.p. M.i.*, *B.x.*, *C.i.*, *M.b.*, and *E.hi.*), while for other, very stringent pre-selections $>$90% of possible HGT candidates were confirmed by evolutionary trees (60 fungal genomes, *A.r.*, *C.e.*, *A.p.*). The latter screens seem more likely to produce a larger number of false negatives. While details (minimum length of scaffold and/or HGT candidate, minimum BLAST score, minimum number of species and/or clades in evolutionary trees, etc.) vary, each reported HGT candidate was supported by an evolutionary tree. Evolutionary trees were evaluated in different ways. Statistical tests (RELL test for *G.s.* and SH-test for *E.hu.* and *B.m.*) were performed. Cutoff branch support values were reported (*N.g.*, *T.v.*, *E.hu.*, *P.p.*, 60 fungal genomes, *B.m.*, *C.i.*, and *M.b.*). Strong branch support (*C.p.*, *H.m.*, *A.r.*, *C.e.*, and *B.x.*), tree topologies indicative of HGT (*M.i.*, *D.d.*[1], and *E.hi.*), or stringent manual curation were mentioned (*C.p.*, *E.hi.*, *G.s.*, *H.m.*, and *B.x.*). Some studies present a tree for each HGT candidate (*P.t.*, *C.e.*, *A.p.*, *C.i.*, *D.d.*[2], and *E.hi.*). For *D.d.*[2] only HGT candidates from 26 out of 49 trees are included here. In screens of genomes from photosynthetic eukaryotes where HGT candidates from cyanobacteria were included (*E.hi.* and *P.t.*) these are removed for the #HGT given here. The relatively high #HGT in the bdelloid rotifer *Adineta ricciae* is supported by studies with *Adineta vaga* [61, 62], which calculate an "alien index" (based on BLAST scores) but do not provide evolutionary trees.

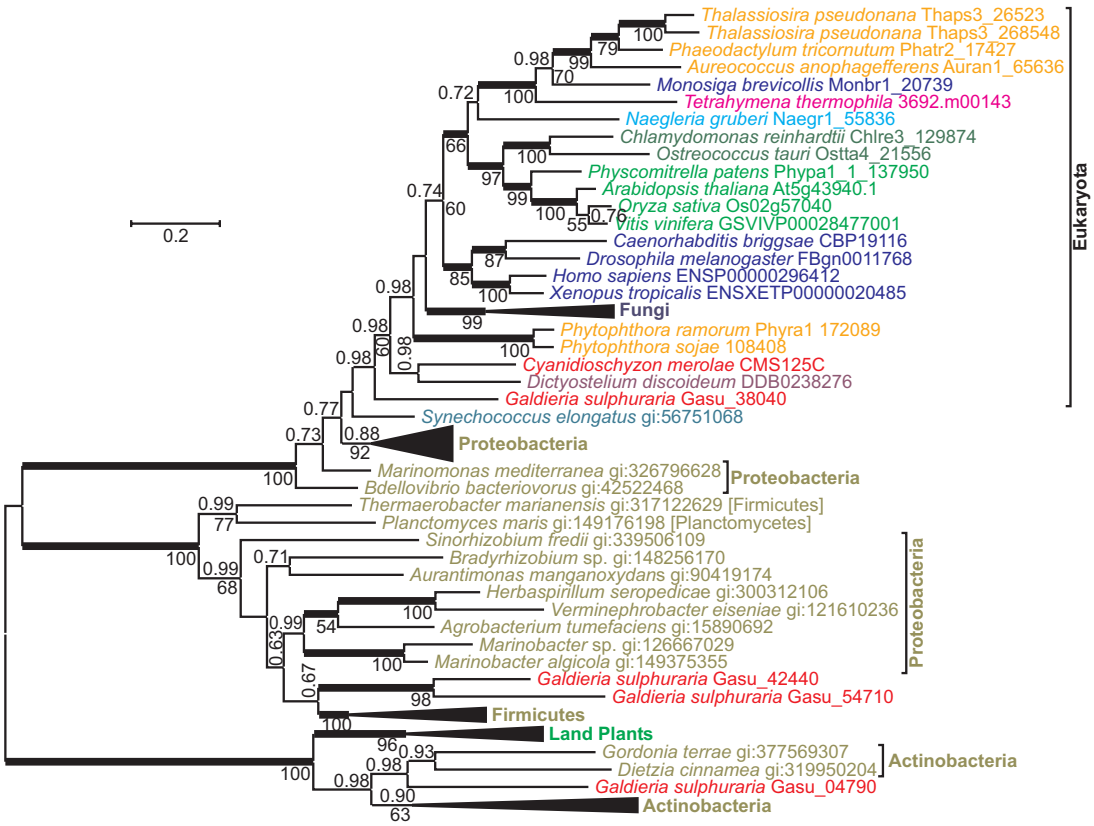

Figure 2. Phylogenetic tree of alcohol dehydrogenases (ADH, EC 1.1.1.1) indicating that *G. sulphuraria* acquired pseudoparalogs via HGT**.** The genome of *G. sulphuraria* encodes one alcohol dehydrogenase (EC 1.1.1.1) that is of eukaryotic origin (Gasu_38040) and has a homolog in *C. merolae* (CMS125C). In addition, there is one alcohol dehydrogenase gene that was acquired by HGT from Actinobacteria (Gasu_04790), and another two that probably result from an HGT from Firmicutes (Gasu_42440 & Gasu_54710). The unrooted Bayesian [92] tree calculated with a WAG+I+G model of protein evolution shows posterior probabilities above the branches and PhyML [93] percent bootstrap support (using LG+I+G) below the branches. Thick branches indicate 1.0 posterior probability. For clarity some subbranches have been collapsed (elongated triangles). Color coding of major phylogenetic groups: Bacteria, Cyanobacteria, Streptophyta, Chlorophyta, Rhodophyta, Stramenopiles, Alveolata, Excavata, Amoebozoa, Fungi, Animals & Choanoflagellates. Bar represents 0.2 changes per site.

Table 1), most likely resulting in the identification of small subsets of HGT candidates. Comparing phylogenomic analyses that avoided such filtering by constructing an evolutionary tree for each protein encoded in a genome (*P.t.*, *C.m.*, *P.p*, and *B.p*. in Table 1), we find that reported numbers of HGT candidates vary less than suggested by the direct comparisons [37, 39], ranging from 1.0% to 2.1%. However, such unbiased phylogenomic screens are currently only available for a small number of algae, and it is not clear if the reported range of HGT candidate numbers is also typical for other lineages.

Detected HGT candidates – Just the tip of the iceberg?

Only a few percent, or even less, of genes in eukaryotic genomes were identified as HGT candidates (Table 1). Does this mean that >90% of genes can be assumed to be of vertical descent? At their root, eukaryotes may be considered a merger between an archaeal host and a bacterial endosymbiont, and the ensuing massive endosymbiotic gene transfer may have affected a majority of present-day eukaryotic genes (see above). Another avalanche of non-vertical inheritance followed the acquisition of the plastidal ancestor: for *Arabidopsis*, about 18% of nuclear genes have been reported to result from endosymbiotic gene transfer from plastids [43]. But even beyond these bursts of endosymbiotic gene transfers, there is reason to believe that vertical inheritance may not be as dominant in eukaryotic gene evolution as often assumed. For most proteins encoded by eukaryotic genomes, phylogenomic analyses do not permit reliable conclusions about descent [44]. Evolutionary trees with >70% bootstrap support can be obtained for less than 20% of all proteins [39]. For example, in the red algae *Porphyridium purpureum* and *C. merolae*, where 1.7% and 1% of all genes were identified as HGT candidates, respectively (Table 1), these percentages correspond to 15.2% and 8.9% of all proteins with evolutionary trees with >70% bootstrap support [39]. In phylogenomic analyses, roughly the same percentage of trees provides clear support for either vertical descent or HGT [39].

For these reasons, the numbers of HGT candidates listed in Table 1 likely only give a lower bound. Moreover, most HGT screens are limited to the evolutionary time frame following the

divergence of a particular clade from their last common ancestor. Our analysis of the *G. sulphuraria* genome was limited to protein sequences that occur in *G. sulphuraria* and *C. merolae*, but no other eukaryote. Therefore, HGT events that occurred before the red and green algae diverged were ignored; such HGT events in the common ancestor of red (Rhodophyta) and green (Viridiplantae) plants have been reported by other studies [44, 45]. Only phylogenomic analyses that attempt to reconstruct the evolutionary history of each protein encoded in a genome stand a chance of recovering, in addition, those ancient HGT events. However, most HGT screens of eukaryotic genomes first used BLAST scores (or similar approaches; see legend to Table 1) to narrow down the number of proteins to be analyzed by evolutionary trees.

Most screens for HGT candidates in eukaryotic genomes have been limited to transfers from prokaryotes (Table 1). Because of the huge phylogenetic distance, it is usually easier to identify HGTs from bacteria or archaea into eukaryotic genomes (Figs 1 and 2), compared with HGTs between different eukaryotic lineages. We obtained numerous evolutionary trees where protein sequences from *G. sulphuraria* and fungi or metazoa form a monophyletic group to the exclusion of all other photosynthetic eukaryotes, in clear violation of established organismal systematics. When these incongruences were analyzed further, in most cases it was impossible to decide whether conservation from a common ancestor and multiple gene loss in other clades or HGT was the more likely explanation. Interestingly, in fungi, where >240 complete genomes are available, a recent, detailed reconstruction of the evolutionary history of the high-affinity fructose:H^+ transporter (TC 2.A.1.1.33) shows a complex pattern of gains and losses with a minimum of ten intra-kingdom HGT events [46]. The authors speculate that intra-kingdom HGT may be more common in eukaryotes than assumed, and was missed so far because of a limited number of eukaryotic genomes. Other than in fungi, homologs of high-affinity fructose:H^+ transporters have so far only been detected in *G. sulphuraria*.

It is also important to emphasize again that when testing genes for HGT, vertical inheritance is the null hypothesis. But most single gene/protein evolutionary analyses do not provide strong evidence for vertical inheritance either. Many phylogenetic analyses are inconclusive, and parametric methods can only detect HGT events when donor genomes show significant differences in composition, and when existing differences have not yet deteriorated. If the burden of proof were shifted and HGT be taken as null hypothesis, only a few percent of genes in eukaryotic genomes would probably pass a test for vertical inheritance; these are those few "core genes" that have been used to reconstruct the early branching of the major eukaryotic lineages [47]. When interpreting the percentage of HGT candidates in eukaryotic genomes (Table 1), one has to keep in mind that for the majority of all genes or proteins a reliable phylogeny cannot currently be obtained. In cases where a reliable phylogeny can be obtained, the percentage of genes identified as HGT candidates is roughly the same as the percentage of genes following the pattern expected from vertical descent.

What are the mechanisms for HGT in eukaryotes?

In Bacteria, the three mechanisms for horizontal transfer of DNA are (i) transformation – direct uptake of free exogenous DNA; (ii) transduction – virus-mediated DNA transfer by phages; and (iii) conjugation – plasmid-mediated DNA transfer, requiring cell-to-cell contact. Compared to hundreds of gene transfers from prokaryotes into eukaryotes, few transfer events in the opposite direction, from eukaryotes into a prokaryotic genome, have been documented [48–50].

In eukaryotes, DNA transfer from intracellular symbionts or parasites and DNA transfer from phagotrophic food vesicles into the nuclear genome has been postulated. Transformation has been observed for yeast [51], and agrobacterium-mediated transformation is possible in fungi and plants. In contrast to competent bacteria, no DNA uptake system has been reported for eukaryotes.

Transduction by viral vectors has recently been postulated as a mechanism for HGT also in eukaryotes. In the last years, giant viruses – with genomes up to 2.5 Mb, encoding more than 1,000 genes – have been discovered [52, 53], and there are indications that these giant viruses may be involved in transferring genes between different cellular organisms [54–56]. HGT between endosymbiotic *Wolbachia* bacteria and their insect host has been suggested to be mediated by a phage vector [57]. HGT from double-stranded RNA viruses into eukaryotes seems widespread [58], and phylogenetic analyses indicate that horizontal transmissions of viruses between different host eukaryotic clades is possible [59]. While nothing is currently known about viruses that might infect *G. sulphuraria*, seven annotated proteins had best BLAST hits in proteins encoded by dsDNA viruses. After manual inspection, two of these proteins were included in the list of HGT candidates. Gasu_65150 is annotated as 3′–5′ exonuclease domain of family-B DNA polymerases, and the protein family tree (Fig. 3) indicates that Gasu_65150 is not monophyletic with other eukaryotic 3′–5′ exonuclease domains. Instead, *G. sulphuraria* might have acquired this gene via HGT from a dsDNA viral genome. In this context, it is interesting to note that dsDNA viruses usually have to enter the nucleus of their host before they can replicate [60], suggesting a pathway of DNA transfer to the nucleus. In bdelloid rotifers [61, 62] and nematodes [31], genes acquired by HGT are found in regions that are significantly enriched in transposable elements. Transposable elements might function in "mobilizing" genetic material by moving or copying it between nuclear chromosomes and mobile vectors, such as viruses or parasites. Transposable elements move horizontally in prokaryotes and eukaryotes as well as between prokaryotes and eukaryotes [63, 64].

Conjugation between *E. coli* and yeast [65] has been observed; yet it is unclear how important conjugation is for HGT between prokaryotes and eukaryotes. Direct cell-to-cell contact has been shown to be conducive to DNA transfer, and in plants natural grafting has been suggested as a path for HGT between sexually incompatible species [66]. Similarly, there are well documented cases of HGT from plant

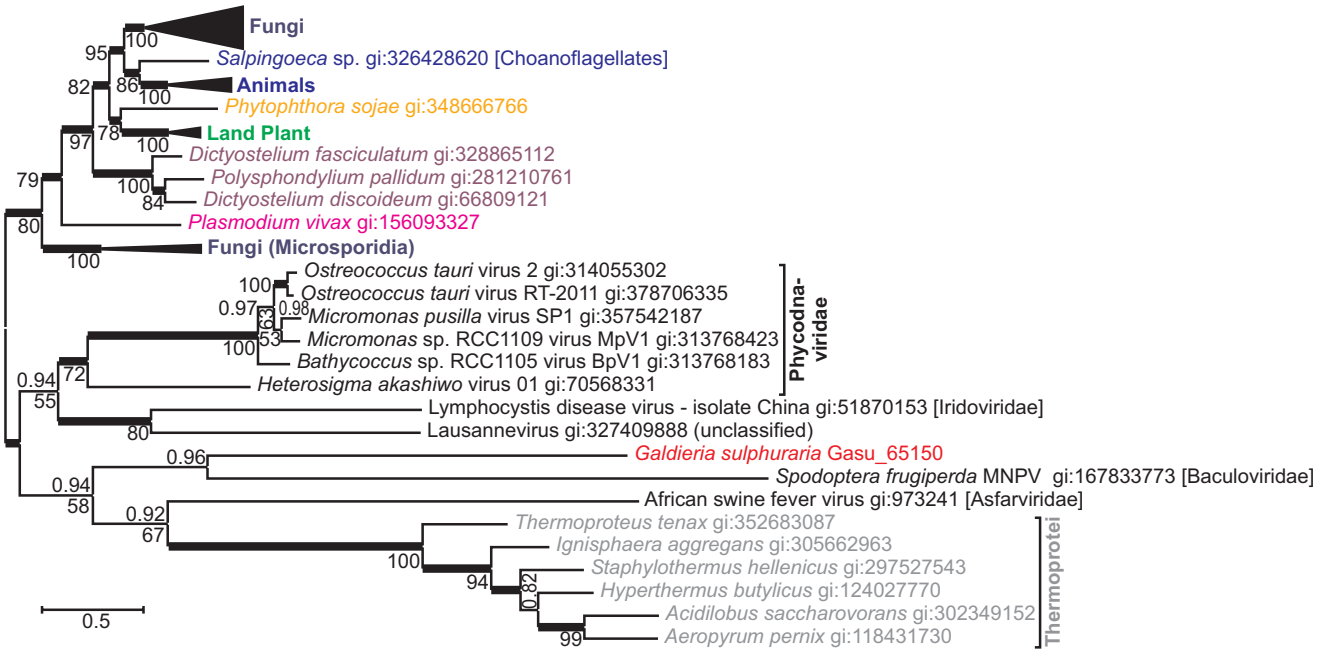

Figure 3. Phylogenetic tree of 3′-5′ exonuclease domains of family-B DNA polymerases indicating HGT from dsDNA viruses into the *G. sulphuraria* genome**.** The unrooted Bayesian [92] tree calculated with a WAG+I+G model of protein evolution shows posterior probabilities above the branches and PhyML [93] percent bootstrap support (using LG+I+G+F) below the branches. Thick branches indicate 1.0 posterior probability. For clarity some sub-branches have been collapsed (elongated triangles). Color coding of major phylogenetic groups: Viruses, Archaea, Streptophyta, Rhodophyta, Stramenopiles, Alveolata, Amoebozoa, Fungi, Animals & Choanoflagellates. Bar represents 0.5 changes per site.

hosts to parasitic plants [67], and of widespread HGT from intracellular bacteria to multicellular eukaryotes [68]. On the other hand, microsporidian parasites have assembled an entire metabolic pathway via multiple independent HGTs that do not originate from the metazoan host [69, 70].

Early on, DNA transfer from prey, especially in phagotrophic protists, was postulated as a mechanism for HGT [71]. However, the high percentage of HGT candidates observed in cell wall enclosed photoautotrophic algae (Table 1) can hardly be explained by this mechanism.

The high number of HGT candidates in bdelloid rotifers has been speculated to result from DNA uptake and integration facilitated by membrane disruption and DNA fragmentation and repair during desiccation [61, 62]. Similar mechanisms might work in nematodes that form highly draught resistant Dauer larvae [72], and in *G. sulphuraria* living on the surface on soil and rocks. Yet there is currently no experimental evidence indicating whether HGT in eukaryotes is promoted by massive stress and injury.

All the mechanisms discussed above emphasize the mechanistic basis of horizontal DNA transmission from one individual organism to another. A larger number of individuals (i.e. a larger population size) means that the population as a whole goes through more of such attempts to integrate foreign DNA. Thus, larger populations are expected to show higher rates of adaptation by HGT. This may in part explain why cell wall-encased algae living in the open ocean can harbor larger numbers of HGT candidates than phagotrophic protists or parasitic protists that live in close association with their host, which individually should have a much higher chance to acquire foreign DNA [73, 74]. It is still surprising that neither lifestyle nor absence or presence of a cell wall seem to correlate with inferred HGT rates. Similarly, because of the separation of metazoan reproductive cells from the environment by somatic cells, it is usually assumed that horizontal gene acquisition rates are lower in metazoans than in many other eukaryotes. However, existing data do not seem to support this notion, as some of the highest percentages of HGT candidates were detected in genomes of nematodes [31].

Evolution requires not only the generation of variability (horizontal acquisition of a new sequence), but also its spread through a population until fixation, a process that is more likely for variants that are more beneficial. The surprising species distribution of eukaryotic gene acquisition may thus be dominated by the functional utility of transferrable genes: HGT might not be limited as much by a foreign gene *getting* into a new genome but by the foreign gene *fitting* into the new genome and providing an adaptive advantage, emphasizing the role of fixation. This conclusion is in line with findings in bacteria. HGT patterns of bacterial metabolic genes appear to be driven by adaptation to changing environments [3]. Ecological similarity among bacteria appears to be more important to explain the distribution of recently transferred genes than close phylogenetic relationship or even close geographic proximity [75]. Mirroring the situation in prokaryotes [3, 76], all systematic screens for HGTs in eukaryotes (Table 1) show that HGT candidates are enriched in enzymes [77]. It seems highly unlikely that these genes have a higher probability of being transmitted. This bias towards enzymes

and the very low proportion of informational genes among HGT candidates emphasizes the importance of fixation of acquired genes.

The mechanisms by which foreign DNA molecules enter eukaryotic nuclei remain enigmatic, although recent work indicates that viruses might play an important role. Close physical contact of eukaryotes with symbionts, parasites, or prey might promote HGT; however, the occurrence of HGT may be limited less by the frequency of attempts to integrate foreign DNA than by the probability of fixation.

Which genes are transferred?

In bacteria, metabolic enzymes and transport proteins have been shown to be especially good candidates for HGT [3, 76], likely both because these proteins act at the interface between the organism and its environment and because of the less complex interactions with other genes (the "complexity hypothesis" [7, 8]). Similarly, in all eukaryotic genomes examined (Table 1), a large fraction of HGT candidates encodes metabolic enzymes [77]. A statistical analysis of HGT candidates from *G. sulphuraria* shows a disproportionally high fraction of genes encoding metabolic enzymes and transport proteins [22], comparable to bacteria.

Comprehensive work in Bacteria has shown that ecological similarity is most important to explain the distribution of recently transferred genes [75]. One would therefore expect that species that are closely related to putative gene donors, i.e. those species that are only distantly related to the examined eukaryote but show up close to HGT candidates in protein family trees, live in a similar environment as the organism under study. Indeed, for the soil-dwelling plant-parasitic nematode *Meloidogyne incognita*, many potential donors for HGT are soil bacteria [31], and for the thermoacidophilic red alga *G. sulphuraria* we found a significant enrichment in thermophilic and acidothermophilic bacteria among potential HGT donors [22].

Early on, HGT in bacteria was classified into three categories: (i) genes that replace existing genes encoding the same function (xenologous replacement); (ii) genes that result in a diversification of existing protein families, i.e. add pseudoparalogs to existing genes; and (iii) genes that provide new functions [78, 79]. For genes providing new functions, the potential adaptive value is easy to see. The same holds for pseudoparalogs that might contribute to functional diversification within a protein family (with the added benefit that this neo- or sub-functionalization may be instantaneous in the case of HGT). In contrast, the potential adaptive advantage for HGT to replace an existing gene is less obvious: selection should typically favor retention of the original gene, which is already integrated into the functional and regulatory networks of the cell. Accordingly, in *G. sulphuraria*, only two out of 75 HGT events seem to have replaced an existing gene [22]. One is the eukaryotic gene for kynurenine formamidase being substituted by a bacterial one (Fig. 1A), the other is the gene encoding nitrate reductase.

Most HGT events detected in *G. sulphuraria* thus seem either to add pseudoparalogs to existing genes or, more often, add new functions. The acquisition of a bacterial periplasmic metal binding protein (TroA/SBP A-1), as shown in Fig. 1B, is an example of the acquisition of a new function. In Bacteria, this protein is often encoded as part of an operon for metal transporting ABC transporters, and functions as an initial receptor that binds metal ions with high affinity [80, 81]. The homologous protein from *G. sulphuraria* is also likely involved in the selective uptake of metal cations. Figure 2 displays two potential HGTs adding pseudoparalogs (in the sense of a pan-genomic gene duplication [82]) to a family of alcohol dehydrogenases in *G. sulphuraria*. It seems likely that the different alcohol dehydrogenases in *G. sulphuraria* have different substrate specificities, as is known to occur in other species [83]. In eukaryotes, gene families with several paralogs usually result from gene duplications followed by neo- or sub-functionalization [1, 2]. In contrast, in Bacteria, HGT is the main driving force for protein family expansion [5]. Similar to what has been established for Bacteria, in *G. sulphuraria* expansion of protein families is not only driven by gene duplication but also by HGT. These two mechanisms often act in concert: according to our analyses, 75 horizontal gene acquisitions of *G. sulphuraria* gave rise to 339 genes due to multiple rounds of gene duplication. Twenty of the HGT candidates were subsequently duplicated, and out of the 20 largest protein families, five contain horizontally acquired genes [22]. The same pattern has been observed for HGT candidates in other eukaryotic genomes (Table 1): gene acquisition, which adds novel functional potential to the genome, is often followed by multiple rounds of gene duplication [31].

Thus, as in bacteria, HGT candidates in eukaryotic genomes often encode metabolic enzymes or transport proteins, which were acquired from ecologically similar organisms. In most cases, the acquired gene seems to provide the organism with a novel functionality; the benefit from this functional gain may be expanded through several rounds of gene duplication.

What is the evolutionary impact of HGT?

Our analyses of *G. sulphuraria* show that most HGT candidates added new functionality, likely providing adaptive advantages. Most of the astounding properties of *G. sulphuraria* can, at least in part, be explained by genes acquired via HGT. The extreme heat tolerance is likey related to two large families of Archaeal ATPases found in hyperthermophilic archaea and bacteria; salt tolerance is aided by a bacterial enzyme producing compatible solutes as well as by sodium pumps of bacterial origin; mercury resistance is probably conferred by a bacterial mercury reductase; a bacterial arsenic pump contributes to arsenic resistance; numerous metabolite transporters and enzymes not detected in other eukaryotes contribute to the enormous metabolic flexibility of *G. sulphuraria* [22]. It is fair to say that *G. sulphuraria* is what it is at least in part because of multiple HGTs from archaea and bacteria. Without these HGT candidates, *G. sulphuraria* would be unlikely to thrive in the hot, salt and toxic-metal laden, volcanic environment where few other organisms survive.

Recent publications indicate that HGT provided major adaptive advantages

in a variety of eukaryotes. Diatoms and other algae colonizing sea ice contain ice-binding proteins not detected in other diatoms. The genes encoding these ice-binding proteins were probably obtained via HGT from bacteria [84]. Some pea aphids can synthesize bright red carotenoids, which animals usually cannot produce. The aphid genes that encode the enzymes for carotenoid biosynthesis were probably acquired from fungi [85]. Red aphids producing the carotenoid, compared with green ones not producing the carotenoid, while more likely to be preyed on, have lower parasitism rates [86].

In plant-parasitic nematodes, several genes involved in parasitic life style, such as carbohydrate and protein degrading or modifying enzymes, probably originated from soil bacteria or fungi via HGT [31]. It has been postulated that the origin of plant parasitism in different nematode phyla was driven by HGT, and that HGT may indeed be a prerequisite for successful plant parasitism in nematodes [87]. In fungi, the rate of HGT seems to be relatively low, less than 1% (Table 1). Yet there is evidence that HGT has promoted niche specification, disease emergence, and the acquisition of new metabolic pathways in fungi [88]. The 43 genes acquired by the silkworm (*Bombyx mori*) via HGT account for three per mille of the genome, but have been suggested to provide adaptive advantages by enhancing disease resistance, nutrient and energy metabolism, and toxin degradation [35]. Finally, the evolution of C_4 photosynthesis in some grass species might have been promoted by plant-plant HGT of key C_4 genes [89]. Even though the number of genes acquired via HGT may be low in a given genome, there is a rapidly increasing number of examples where HGT does seem to contribute to major eukaryotic adaptations.

HGT in eukaryotes may even have contributed to major evolutionary transitions. Genes of algal origin in the tunicate *Ciona intestinalis* related to molecular transport and signaling were speculated to have contributed to the origin of multi-cellularity in animals [90]. An HGT screen in the moss *Physcomitrella patens* identified genes originating from viruses, prokaryotes, or fungi that were suggested to have played important roles in plant colonization of land [91].

Conclusions and outlook

In addition to endosymbiotic gene transfer from mitochondria and chloroplasts, eukaryotic genomes have acquired numerous genes via HGT, often from bacteria or archaea. Comparable to HGT among bacteria, these genes originate from different donor species and have been acquired over long evolutionary periods one-by-one or a few genes at a time. The number of HGTs into eukaryotic genomes might be higher than previously thought. As in bacteria, there is a preference for the acquisition of operational genes, such as those encoding enzymes or membrane transport proteins. Recent analyses provide increasing evidence that horizontally acquired genes play an important role in the adaptive evolution of eukaryotes.

To compare HGT frequencies across different eukaryotic lineages, large-scale phylogenomic screens of multiple eukaryotic genomes using identical criteria are required. Based on such data, detailed functional characterizations of established HGT candidates will provide a more systematic insight into the adaptive advantages provided by eukaryotic gene acquisitions. A better understanding of HGT in eukaryotes could eventually contribute to a better design of bioengineered eukaryotic organisms.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (EXC 1028 to A.P.M.W. & M.J.L.; CRC TR1 to A.P.M.W.; and CRC 680 to M.J.L.).

The authors have declared no conflict of interest.

References

1. **Taylor JS**, **Raes J.** 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* **38**: 615–43.
2. **Innan H**, **Kondrashov F.** 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**: 97–108.
3. **Pal C**, **Papp B**, **Lercher MJ.** 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37**: 1372–5.
4. **Popa O**, **Hazkani-Covo E**, **Landan G**, **Martin W**, et al**.** 2011. Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res* **21**: 599–609.
5. **Treangen TJ**, **Rocha EPC.** 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**: e1001284.
6. **Akiba T**, **Koyama K**, **Ishiki Y**, **Kimura S**, et al**.** 1960. On the mechanism of the development of multiple-drug-resistant clones of *Shigella*. *Jpn J Microbiol* **4**: 219–27.
7. **Jain R**, **Rivera MC**, **Lake JA.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96**: 3801–6.
8. **Cohen O**, **Gophna U**, **Pupko T.** 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* **28**: 1481–9.
9. **Lane N**, **Martin W.** 2010. The energetics of genome complexity. *Nature* **467**: 929–34.
10. **Timmis JN**, **Ayliffe MA**, **Huang CY**, **Martin W.** 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5**: 123–35.
11. **Keeling PJ.** 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Bio* **64**: 583–607.
12. **Nikoh N**, **McCutcheon JP**, **Kudo T**, **Miyagishima SY**, et al**.** 2010. Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet* **6**: e1000827.
13. **Maynard Smith J**, **Szathmáry E.** 1995. *The Major Transitions in Evolution*. Oxford, NY: W.H. Freeman Spektrum.
14. **Kidwell MG.** 1993. Lateral transfer in natural populations of eukaryotes. *Annu Rev Genet* **27**: 235–56.
15. **Syvanen M.** 1994. Horizontal gene transfer: evidence and possible consequences. *Annu Rev Genet* **28**: 237–61.
16. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
17. **Salzberg SL**, **White O**, **Peterson J**, **Eisen JA.** 2001. Microbial genes in the human genome: lateral transfer or gene loss? *Science* **292**: 1903–6.
18. **Stanhope MJ**, **Lupas A**, **Italia MJ**, **Koretke KK**, et al**.** 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* **411**: 940–4.
19. **Boto L.** 2010. Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **277**: 819–27.
20. **Keeling PJ**, **Palmer JD.** 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**: 605–18.
21. **Syvanen M.** 2012. Evolutionary implications of horizontal gene transfer. *Annu Rev Genet* **46**: 341–58.
22. **Schönknecht G**, **Chen W-H**, **Ternes CM**, **Barbier GG**, et al**.** 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**: 1207–10.

23. **Roettger M**, **Martin W**, **Dagan T.** 2009. A machine-learning approach reveals that alignment properties alone can accurately predict inference of lateral gene transfer from discordant phylogenies. *Mol Biol Evol* **26**: 1931–9.
24. **Salichos L**, **Rokas A.** 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**: 327–31.
25. **Davison A**, **Blaxter M.** 2005. Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* **22**: 1273–84.
26. **Becq J**, **Churlaud C**, **Deschavanne P.** 2010. A benchmark of parametric methods for horizontal transfers detection. *PLoS One* **5**: e9989.
27. **Azad RK**, **Lawrence JG.** 2011. Towards more robust methods of alien gene detection. *Nucleic Acids Res* **39**: e56.
28. **Longo MS**, **O'Neill MJ**, **O'Neill RJ.** 2011. Abundant human DNA contamination identified in non-primate genome databases. *PLoS One* **6**: e16410.
29. **Grunau C**, **Boissier J.** 2010. No evidence for lateral gene transfer between salmonids and schistosomes. *Nat Genet* **42**: 918–9.
30. **Bowler C**, **Allen AE**, **Badger JH**, **Grimwood J**, et al**.** 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–44.
31. **Paganini J**, **Campan-Fournier A**, **Da Rocha M**, **Gouret P**, et al**.** 2012. Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLoS One* **7**: e50875.
32. **Marcet-Houben M**, **Gabaldón T.** 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet* **26**: 5–8.
33. **Yoon HS**, **Zuccarello GC**, **Bhattacharya D.** 2010. Red Algae in the Genomic Age. In Seckbach J, Chapman DJ, eds; *Evolutionary History and Taxonomy of Red Algae*. Netherlands: Springer. p. 25–42.
34. **Bhattacharya D**, **Yoon HS**, **Hedges SB**, **Hackett JD.** 2009. Eukaryotes (Eukaryota). In Hedges SB, Kumar S, eds, *The Timetree of Life*. NY: Oxford University Press. p. 116–20.
35. **Zhu B**, **Lou M-M**, **Xie G-L**, **Zhang G-Q**, et al**.** 2011. Horizontal gene transfer in silkworm *Bombyx mori*. *BMC Genomics* **12**: 248.
36. **Medlin LK.** 2009. Diatoms (Bacillariophyta). In Hedges SB, Kumar S, eds, *The Timetree of Life*. NY: Oxford University Press. p. 127–30.
37. **Boschetti C**, **Carr A**, **Crisp A**, **Eyres I**, et al**.** 2012. Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genet* **8**: e1003035.
38. **Kikuchi T**, **Cotton JA**, **Dalzell JJ**, **Hasegawa K**, et al**.** 2011. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog* **7**: e1002219.
39. **Bhattacharya D**, **Price DC**, **Chan CX**, **Qiu H**, et al**.** 2013. Genome of the red alga *Porphyridium purpureum*. *Nat Commun* **4**: 1941.
40. **Eichinger L**, **Pachebat JA**, **Glockner G**, **Rajandream MA**, et al**.** 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**: 43–57.
41. **Andersson JO.** 2011. Evolution of patchily distributed proteins shared between eukaryotes and prokaryotes: *Dictyostelium* as a case study. *J Mol Microbiol Biotechnol* **20**: 83–95.
42. **Huang J**, **Mullapudi N**, **Lancto C**, **Scott M**, et al**.** 2004. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol* **5**: R88.
43. **Martin W**, **Rujan T**, **Richly E**, **Hansen A**, et al**.** 2002. 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**: 12246–51.
44. **Price DC**, **Chan CX**, **Yoon HS**, **Yang EC**, et al**.** 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**: 843–7.
45. **Huang J**, **Gogarten JP.** 2008. Concerted gene recruitment in early plant evolution. *Genome Biol* **9**: R109.
46. **Coelho MA**, **Gonçalves C**, **Sampaio JP**, **Gonçalves P.** 2013. Extensive intra-kingdom horizontal gene transfer converging on a fungal fructose transporter gene. *PLoS Genet* **9**: e1003587.
47. **Burki F**, **Okamoto N**, **Pombert J-F**, **Keeling PJ.** 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* **279**: 2246–54.
48. **Duplouy A**, **Iturbe-Ormaetxe I**, **Beatson S**, **Szubert J**, et al**.** 2013. Draft genome sequence of the male-killing *Wolbachia* strain wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* **14**: 20.
49. **Woolfit M**, **Iturbe-Ormaetxe I**, **McGraw EA**, **O'Neill SL.** 2009. An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis*. *Mol Biol Evol* **26**: 367–74.
50. **Rinke C**, **Schwientek P**, **Sczyrba A**, **Ivanova NN**, et al**.** 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–7.
51. **Nevoigt E**, **Fassbender A**, **Stahl U.** 2000. Cells of the yeast *Saccharomyces cerevisiae* are transformable by DNA under non-artificial conditions. *Yeast* **16**: 1107–10.
52. **Scola BL**, **Audic S**, **Robert C**, **Jungang L**, et al**.** 2003. A giant virus in amoebae. *Science* **299**: 2033.
53. **Philippe N**, **Legendre M**, **Doutre G**, **Couté Y**, et al**.** 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* **341**: 281–6.
54. **Routh A**, **Domitrovic T**, **Johnson JE.** 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci USA* **109**: 1907–12.
55. **Yoon HS**, **Price DC**, **Stepanauskas R**, **Rajah VD**, et al**.** 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**: 714–7.
56. **Sharon I**, **Alperovitch A**, **Rohwer F**, **Haynes M**, et al**.** 2009. Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461**: 258–62.
57. **Klasson L**, **Kambris Z**, **Cook P**, **Walker T**, et al**.** 2009. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genomics* **10**: 33.
58. **Liu H**, **Fu Y**, **Jiang D**, **Li G**, et al**.** 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J Virol* **84**: 11876–87.
59. **Liu H**, **Fu Y**, **Xie J**, **Cheng J**, et al**.** 2012. Discovery of novel dsRNA viral sequences by *in silico* cloning and implications for viral diversity, host range and evolution. *PLoS One* **7**: e42147.
60. **Dimmock NJ**, **Easton AJ**, **Leppard KN.** 2007. *Introduction to Modern Virology*. Malden, MA: Blackwell Publishing.
61. **Gladyshev EA**, **Meselson M**, **Arkhipova IR.** 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320**: 1210–3.
62. **Flot J-F**, **Hespeels B**, **Li X**, **Noel B**, et al**.** 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* **500**: 453–7.
63. **Walsh AM**, **Kortschak RD**, **Gardner MG**, **Bertozzi T**, et al**.** 2013. Widespread horizontal transfer of retrotransposons. *Proc Natl Acad Sci USA* **110**: 1012–6.
64. **Gilbert C**, **Cordaux R.** 2013. Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. *Genome Biol Evol* **5**: 822–32.
65. **Heinemann JA**, **Sprague GF.** 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**: 205–9.
66. **Stegemann S**, **Keuthe M**, **Greiner S**, **Bock R.** 2012. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl Acad Sci USA* **109**: 2434–8.
67. **Xi Z**, **Bradley R**, **Wurdack K**, **Wong KM**, et al**.** 2012. Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* **13**: 227.
68. **Dunning Hotopp JC**, **Clark ME**, **Oliveira DC**, **Foster JM**, et al**.** 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**: 1753–6.
69. **Pombert J-F**, **Selman M**, **Burki F**, **Bardell FT**, et al**.** 2012. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. *Proc Natl Acad Sci USA* **109**: 12638–43.
70. **Corradi N**, **Selman M.** 2013. Latest progress in microsporidian genome research. *J Eukaryot Microbiol* **60**: 309–12.
71. **Doolittle WE.** 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* **14**: 307–11.
72. **Rybarczyk-Mydlowska K**, **Maboreke HR**, **van Megen H**, **van den Elsen S**, et al**.** 2012. Rather than by direct acquisition via lateral gene transfer, GHF5 cellulases were passed on from early Pratylenchidae to root-knot and cyst nematodes. *BMC Evol Biol* **12**: 221.
73. **Andersson JO.** 2005. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* **62**: 1182–97.
74. **Keeling PJ.** 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin Genet Dev* **19**: 613–9.
75. **Smillie CS**, **Smith MB**, **Friedman J**, **Cordero OX**, et al**.** 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**: 241–4.
76. **Kreimer A**, **Borenstein E**, **Gophna U**, **Ruppin E.** 2008. The evolution of modularity in bacterial metabolic networks. *Proc Natl Acad Sci USA* **105**: 6976–81.
77. **Whitaker J**, **McConkey G**, **Westhead D.** 2009. The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. *Genome Biol* **10**: R36.

78. **Koonin EV**, **Makarova KS**, **Aravind L.** 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709–42.
79. **Makarova KS**, **Wolf YI**, **Mekhedov SL**, **Mirkin BG**, et al**.** 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* **33**: 4626–38.
80. **Berntsson RPA**, **Smits SHJ**, **Schmitt L**, **Slotboom D-J**, et al**.** 2010. A structural classification of substrate-binding proteins. *FEBS Lett* **584**: 2606–17.
81. **Zheng B**, **Zhang Q**, **Gao J**, **Han H**, et al**.** 2011. Insight into the interaction of metal ions with TroA from *Streptococcus suis*. *PLoS One* **6**: e19510.
82. **Grassi L**, **Caselle M**, **Lercher MJ**, **Lagomarsino MC.** 2012. Horizontal gene transfers as metagenomic gene duplications. *Mol BioSyst* **8**: 790–5.
83. **Sealy-Lewis HM**, **Fairhurst V.** 1995. Substrate specificity of nine NAD^+-dependent alcohol dehydrogenases in *Aspergillus nidulans*. *Microbiol* **141**: 2295–300.
84. **Raymond JA**, **Kim HJ.** 2012. Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS One* **7**: e35968.
85. **Moran NA**, **Jarvik T.** 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**: 624–7.
86. **Losey JE**, **Harmon J**, **Ballantyne F**, **Brown C.** 1997. A polymorphism maintained by opposite patterns of parasitism and predation. *Nature* **388**: 269–72.
87. **Haegeman A**, **Jones JT**, **Danchin EGJ.** 2011. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol Plant Microbe Interact* **24**: 879–87.
88. **Fitzpatrick DA.** 2012. Horizontal gene transfer in fungi. *FEMS Microbiol Lett* **329**: 1–8.
89. **Christin P-A**, **Edwards Erika J**, **Besnard G**, **Boxall Susanna F**, et al**.** 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr Biol* **22**: 445–9.
90. **Ni T**, **Yue J**, **Sun G**, **Zou Y**, et al**.** 2012. Ancient gene transfer from algae to animals: mechanisms and evolutionary significance. *BMC Evol Biol* **12**: 83.
91. **Yue J**, **Hu X**, **Sun H**, **Yang Y**, et al**.** 2012. Widespread impact of horizontal gene transfer on plant colonization of land. *Nat Commun* **3**: 1152.
92. **Altekar G**, **Dwarkadas S**, **Huelsenbeck JP**, **Ronquist F.** 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**: 407–15.
93. **Guindon S**, **Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
94. **Marchler-Bauer A**, **Anderson JB**, **Derbyshire MK**, **Weese-Scott C**, et al**.** 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**: D237–40.
95. **Marchler-Bauer A**, **Lu SN**, **Anderson JB**, **Chitsaz F**, et al**.** 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**: D225–9.
96. **Fritz-Laylin LK**, **Prochnik SE**, **Ginger ML**, **Dacks JB**, et al**.** 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**: 631–42.
97. **Carlton JM**, **Hirt RP**, **Silva JC**, **Delcher AL**, et al**.** 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* **315**: 207–12.
98. **Read BA**, **Kegel J**, **Klute MJ**, **Kuo A**, et al**.** 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**: 209–13.
99. **Moreau H**, **Verhelst B**, **Couloux A**, **Derelle E**, et al**.** 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol* **13**: R74.
100. **Chapman JA**, **Kirkness EF**, **Simakov O**, **Hampson SE**, et al**.** 2010. The dynamic genome of Hydra. *Nature* **464**: 592–6.
101. **Sun GL**, **Yang ZF**, **Ishwar A**, **Huang JL.** 2010. Algal genes in the closest relatives of animals. *Mol Biol Evol* **27**: 2879–89.
102. **Loftus B**, **Anderson I**, **Davies R**, **Alsmark UC**, et al**.** 2005. The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**: 865–8.