Phylogenomic Study Indicates Widespread Lateral Gene Transfer in *Entamoeba* and Suggests a Past Intimate Relationship with Parabasalids

Jessica R. Grant  
*Smith College*

Laura A. Katz  
*Smith College*, lkatz@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/bio_facpubs

Part of the Biology Commons

**Recommended Citation**

Grant, Jessica R. and Katz, Laura A., "Phylogenomic Study Indicates Widespread Lateral Gene Transfer in *Entamoeba* and Suggests a Past Intimate Relationship with Parabasalids" (2014). Biological Sciences: Faculty Publications, Smith College, Northampton, MA.  
https://scholarworks.smith.edu/bio_facpubs/103

This Article has been accepted for inclusion in Biological Sciences: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu
Phylogenomic Study Indicates Widespread Lateral Gene Transfer in Entamoeba and Suggests a Past Intimate Relationship with Parabasalids

Jessica R. Grant and Laura A. Katz

1Department of Biological Sciences, Smith College, Northampton, MA
2Program in Organismic and Evolutionary Biology, University of Massachusetts

*Corresponding author: E-mail: lkatz@smith.edu.

Accepted: August 14, 2014

Abstract

Lateral gene transfer (LGT) has impacted the evolutionary history of eukaryotes, though to a lesser extent than in bacteria and archaea. Detecting LGT and distinguishing it from single gene tree artifacts is difficult, particularly when considering very ancient events (i.e., over hundreds of millions of years). Here, we use two independent lines of evidence—a taxon-rich phylogenetic approach and an assessment of the patterns of gene presence/absence—to evaluate the extent of LGT in the parasitic amoebozoan genus Entamoeba.

Previous work has suggested that a number of genes in the genome of Entamoeba spp. were acquired by LGT. Our approach, using an automated phylogenomic pipeline to build taxon-rich gene trees, suggests that LGT is more extensive than previously thought. Our analyses reveal that genes have frequently entered the Entamoeba genome via nonvertical events, including at least 116 genes acquired directly from bacteria or archaea, plus an additional 22 genes in which Entamoeba plus one other eukaryote are nested among bacteria and/or archaea. These genes may make good candidates for novel therapeutics, as drugs targeting these genes are less likely to impact the human host. Although we recognize the challenges of inferring intradomain transfers given systematic errors in gene trees, we find 109 genes supporting LGT from a eukaryote to Entamoeba spp., and 178 genes unique to Entamoeba spp. and one other eukaryotic taxon (i.e., presence/absence data). Inspection of these intradomain LGTs provide evidence of a common sister relationship between genes of Entamoeba (Amoebozoa) and parabasalids (Excavata). We speculate that this indicates a past close relationship (e.g., symbiosis) between ancestors of these extant lineages.

Key words: microbial eukaryotes, parasites, Trichomonas vaginalis, horizontal gene transfer, LGT, HGT.

Introduction

Entamoeba histolytica is a human parasite that has a significant impact on health worldwide (Stanley 2003). Although initial phylogenetic analyses placed Entamoeba as an early diverging eukaryote, more recent studies based on greater numbers of genes and more sophisticated methods have shown that Entamoeba is a highly derived member of the Amoebozoa (see Embley 2006), one of the major clades of eukaryotes (Lühe 1913; Cavalier-Smith 1998). Previous work suggests that some genes in Entamoeba are of bacterial or archaeal origin (Yang et al. 1994; Rosenthal et al. 1997; Ali, Shigeta, et al. 2004), and the original annotation of the E. histolytica genome revealed examples of lateral gene transfer (LGT; Loftus and Hall 2005; Loftus et al. 2005; Clark et al. 2007). The original estimate of 96 genes involved in LGT was lowered to 68 when the genome was reassessed in 2007 (Clark et al. 2007) and again in 2010 (Lorenzi et al. 2010). Most of these 68 genes appear to be transfers from bacteria, but others do not have a closely related bacterial donor (at least not one with available sequence data for comparison) and may have been transferred from another eukaryote (Loftus et al. 2005). The genomes of the human parasite E. histolytica, and the closely related E. dispar and E. invadens, have been sequenced and there are expressed sequence tag (EST) data for E. nuttalli and Mastigamoeba balamuthi, a free-living relative. We used these data in a taxon-rich phylogenomic pipeline (Grant and Katz 2014) to assess the impact of LGT on the Entamoeba genome.
Despite barriers to integrating foreign DNA into a genome (Andersson 2005; Baltrus 2013), it is apparent that LGT has impacted the evolution of eukaryotes as well as bacteria and archaea (Katz 2002; Andersson et al. 2003; Keeling and Palmer 2008). Many of the genes affected by LGT are involved in metabolism (Nixon et al. 2002; Ali, Hashimoto, et al. 2004; Ali, Shigeta, et al. 2004; Anderson and Loftus 2005), and it seems likely that LGT has influenced the independent evolution of an anaerobic lifestyle in many eukaryotic lineages (Ginger et al. 2010; Hug et al. 2010).

Supported discordance among gene trees provides evidence for LGT, as a gene with a history of LGT will cluster with the donor lineage in phylogenetic trees long after it has been transferred to another genome. Yet single gene trees are notoriously error prone and errors in the inference of LGT can be made when donor lineages are not represented on the tree (Beiko and Ragan 2009). Several inferences of LGTs, based on phylogenetic relationships, have been falsified by trees with improved taxon sampling and more sophisticated methods (Richards et al. 2003; Andersson 2005). Hence, a taxon-rich approach is needed to assess cases of LGT. Further, there is greater power in detecting LGT between distinctly related species (i.e., interdomain events.) For example, a gene of bacterial ancestry is quite distinct and often easy to recognize after it has been transferred to a eukaryotic genome.

The pattern of presence or absence of a gene can also be a strong indicator of LGT, especially for genes found only among members of distantly related lineages, though the impact of gene loss cannot be discounted (Zmasek and Godzik 2011; Wolf and Koonin 2013). For example, a gene found only among diverse bacteria and one clade of eukaryotes could be explained by assuming the gene was present in the last common ancestor of eukaryotes and that it was lost in all lineages except one. However, LGT from a bacterium to the ancestor of the clade that shares the gene is a more parsimonious explanation (Ragan 2001). Thus, searching for unusual patterns of taxa represented in gene alignments can be used to detect LGT (Lake and Rivera 2004; Cohen and Pupko 2010; Cohen et al. 2011; Le et al. 2012).

Here, we take two approaches to investigate the impact of LGT on the Entamoeba genome: 1) analyze taxon-rich phylogenetic gene trees and assess evidence for LGT in Entamoeba from both bacterial and archaeal lineages, and to a lesser extent from other eukaryotes; and 2) catalog examples of patterns of gene presence/absence in Entamoeba plus bacterial/ archaea as further evidence of potential LGTs. Both approaches suggest a greater number of LGT events in the genome of Entamoeba than previously documented. To our surprise, both also provide evidence of a relationship between ancestors of Entamoeba and parabasalids such as Trichomonas vaginalis, another human parasite that is phylogenetically distant from Entamoeba on the eukaryotic tree of life.

Materials and Methods
Initial Pipeline
The starting point for these analyses is a set of orthologous groups from OrthoMCL (2003), a database of clustered orthologous groups that includes taxa from 105 whole genomes including E. histolytica, E. dispar, and E. invadens. We chose the 6,107 genes in OrthoMCL that contained E. histolytica to seed our phylogenomic pipeline (Grant and Katz 2014). Another species of Entamoeba, E. nuttalli, and a free-living relative of Entamoeba, Mastigamoeba balmamuthi, were included in the data added by the pipeline along with 237 eukaryotes, 485 bacteria, and 59 archaea (supplementary table S1, Supplementary Material online). The output of this pipeline includes, for each orthologous group, a robust single-gene alignment and a most likely tree built in RAxML 7.2.8 (Stamatakis 2006; Stamatakis et al. 2008) with model setting PROTGAMMALG.

Of the 6,107 starting genes, 4,000 were not recovered at the end of the pipeline because either they had fewer than two taxa, because no characters remained after masking positions with more than 50% missing data in the alignments, or because the Entamoeba were removed (180 genes). Removal of Entamoebae spp. can occur when the original cluster of sequences in OrthoMCL is too divergent to satisfy the stringent criteria of our phylogenomic pipeline (Grant and Katz 2014). An additional 3,664 genes were uninformative because the group included only Entamoeba and no other taxa. Finally, we chose to discard those genes that had only one Entamoeba sequence (57 genes), or Entamoeba plus only one other sequence (99 genes) to eliminate potential cases of contamination, though we understand cases of recent LGT may have been missed here.

Inferences from Pipeline
Single-gene alignments for the remaining 2,107 genes from the pipeline output were analyzed in FastTree (Price et al. 2009, 2010), as a first assessment. One thousand bootstrap (BS) replicates were built under the WAG model. A consensus tree was built from these replicates and nodes of <70% BS support were collapsed using custom python scripts and implementing the tree walking methods in p4 (Foster 2004). Trees collapsed to nodes with >70% BS were examined by script and by eye to determine the supported relationships between the Entamoeba and other taxa on the tree. A total of 993 trees were not considered in our assessment of inheritance, as their BS consensus did not provide support for relationships between Entamoeba spp. and any other taxon. This left 1,114 genes to be assessed for evidence of LGT or vertical inheritance.

For the genes that suggested LGT but included other eukaryotes (Entamoeba spp. nested in bacteria and/or archaea in a gene that contains other eukaryotes [81 genes] or Entamoeba spp. sister to nonamoeboida eukaryotes [520
genes), we further refined our inference of LGT with the approximately unbiased (AU) test, as implemented in Consel (Shimodaira and Hasegawa 2001), testing the monophyly of Amoebozoa. In these categories, 63 and 109 genes, respectively, rejected the monophyly of Amoebozoa and were retained.

These 678 (1,114 minus the genes that did not pass the AU test) genes were initially categorized based on the topology of nodes with >70% BS support in FastTree as follows: Vertical inheritance: 253 genes; *Entamoeba* spp. in a tree with only bacteria and/or archaea: 53 genes; *Entamoeba* spp. nested in bacteria and/or archaea in a gene that contains other eukaryotes: 63 genes; *Entamoeba* spp. plus one other eukaryote nested in bacteria and/or archaea or as the only eukaryotes in a gene with bacteria and/or archaea: 22 genes; *Entamoeba* spp. sister to nonamoebozoan eukaryotes: 109 genes; *Entamoeba* spp. plus one other major clade of eukaryotes: 178 genes (fig. 1).

**Sister Relationship—Two Approaches**

We addressed the issue of sister taxa two ways: First, we investigated the genes from our initial phylogenetic inferences with *Entamoeba* spp. in a relationship with non-amoebozoan eukaryotes—the 22 genes with *Entamoeba* spp. plus non-amoebozoan eukaryotes nested within bacteria and/or archaea and 109 genes with *Entamoeba* spp. in a gene with other Amoebozoa but with the monophyly of Amoebozoa rejected by the AU test (fig. 1C and D). FastTree has been compared favorably to RAxML (Liu et al. 2011) but we bootstrapped a number of alignments using RAxML and found that the BS values from FastTree were inflated as compared with the values from RAxML. To be more sure of our sister relationship inferences, we rebootstrapped the 109 trees with RAxML version 7.2.8 using rapid bootstrapping with model PROTGAMMALG and determining the proper number of independent bootstrap replicates with bootstrapping criteria autoMRE (Stamatakis et al. 2005, 2008; Stamatakis 2006). After bootstrapping, we retained only those 22 genes that had BS support >80% in RAxML to a sister clade that contained only one clade of eukaryotes (e.g., trees with sister relationships to a plant and a fungus were rejected even if BS support was high.)

Secondly, in order to investigate sister relationships of *Entamoeba* spp. independent of our phylogenetic pipeline, we analyzed all orthologous groups from OrthoMCL made up of only *Entamoeba* spp. and one other eukaryotic species. To align genes and assess the robustness of the OrthoMCL groupings, fasta files downloaded from the OrthoMCL database were passed through Guidance (Liu et al. 2011), a program that builds and bootstraps multisequence alignments and scores both taxa and characters. For our phylogenomic pipeline, we used Guidance with relaxed score cutoffs (Penn et al. 2010) because the default parameters are too stringent for phylogenomic analyses given the diversity seen with our broad taxon sampling. For this presence/absence analysis, however, we wanted to have greater confidence in our call of orthology, so we used the more conservative default parameters. Most of the groups (225 of 372), were not recovered after Guidance because there were too few sequences in the alignment—either in the original OrthoMCL group (81 orthologous groups) or after removal of poorly aligned sequences (85 orthologous groups) or because Guidance removed all of the sequences from one of the two taxa (59 orthologous groups). Relationships that were retained after this screen are reported.

---

**Fig. 1.—** Number of trees supporting LGT in *Entamoeba* ranked from strongest to weakest support. Cartoon trees exemplifying the patterns consistent with LGT. (A) Putative interdomain LGT: *Entamoeba* species in a tree that otherwise includes only bacteria and/or archaea. (B) Putative interdomain LGT: *Entamoeba* species nested within bacteria or archaea in trees with other eukaryotic taxa; monophyletic amoebozoan rejected by AU test, or no other amoebozoa in gene. (C) Putative interdomain LGT followed by intradomain LGT: *Entamoeba* species with a eukaryotic sister taxon nested within bacteria and/or archaea. Relationship with other eukaryote supported with >80% bootstrap support. (D) Putative intradomain LGT: *Entamoeba* species in a eukaryotic clade that is distinct from other amoebozoan taxa; monophyletic Amoebozoan rejected by AU test.
Functional Comparisons

To assess the function of the genes categorized as having either vertical or lateral descent, we used BLAST2GO (Conesa et al. 2005) with default parameters to assign gene ontologies to the E. histolytica sequences. Entamoeba histolytica sequences from the 253 genes with strong support for vertical inheritance and the 116 genes with strong support for interdomain LGT (table 1 and supplementary table S2, Supplementary Material online) were used. We assessed the differences in Level 2 Biological Processes, as assessed by Blast2GO, in these two groups of genes.

Results

To investigate the impact of LGT on the parasitic genus Entamoeba phylogenetically, we used aligned sequences and gene trees produced from our phylogenomic pipeline (Grant and Katz 2014). After preliminary screening, we retained 1,114 genes that met our initial criteria of sufficient taxon sampling and BS support for a relationship for Entamoebae spp. (see Materials and Methods). For those genes present in other amoebozoan taxa, we used the AU test (Shimodaira and Hasegawa 2001) to evaluate the alternative hypothesis that Entamoeba spp. inherited the gene vertically and the tree topology is spurious. We removed 436 genes for which the AU test did not reject the monophyly of Amoebozoa, leaving 678 genes in our analyses.

From our sample of 678 genes, 253 have patterns consistent with vertical inheritance from amoebozoan ancestors: 197 trees with topological evidence (i.e., Entamoeba spp. in a clade with other amoebozoan taxa) and 56 trees with presence/absence support (i.e., alignments containing only Entamoeba spp. plus other Amoebozoa; table 1 and supplementary table S2, Supplementary Material online).

We also looked for evidence of interdomain LGT to explain the presence of genes in Entamoeba spp. We identified 116 genes consistent with a single interdomain LGT event giving rise to the genes in Entamoeba spp. (table 1): 53 with gene presence/absence evidence of LGT (i.e., genes with only Entamoeba spp. and bacteria and/or archaea; fig. 1A and table 1) and 63 trees with phylogenetic evidence of LGT (i.e., Entamoeba spp. nested within bacterial or archaeal clades with >80% BS support; fig. 1B and table 1).

A smaller number of genes suggest a history of interdomain transfer followed by intradomain transfer. In 22 gene trees, Entamoeba is found sister to a single nonamoebozoan eukaryotic taxon nested within clades of bacteria and/or archaea—a topology consistent with LGT from bacteria or archaea into one of the eukaryotes followed by a second LGT event into the eukaryotic sister (fig. 1C and table 1).

We also looked at patterns of intradomain transfer, though we recognize that eukaryote-to-eukaryote LGT is more difficult to assess than interdomain LGT. Phylogenetic evidence suggests 287 genes from our pipeline analysis have support for intradomain LGT. In 109 genes, tree topologies show a sister relationship with nonamoebozoan taxa, and the monophyly of Amoebozoa (vertical descent) is rejected by the AU test (fig. 1D and table 1), evidence for putative intradomain LGT. In addition, 178 trees contained only Entamoeba spp. and taxa from one other nonamoebozoan clade, a pattern of gene presence/absence suggesting possible gene sharing between diverse eukaryotes (table 1 and supplementary table S2, Supplementary Material online). In these cases, gene loss in all other taxa is another possible explanation.

Gene Function

The distribution of gene function is different in vertically inherited genes compared with genes putatively impacted by LGT in the E. histolytica genome. Using Blast2Go (Conesa
et al. 2005), we assigned functional categories to genes from our phylogenomic pipeline that had phylogenetic signatures of interdomain LGT (116 genes; table 1) and those with compelling phylogenetic evidence of vertical descent (197 genes; table 1). The genes identified as candidate interdomain LGT genes are more likely to be involved in metabolic processes than those identified as vertically inherited genes, while vertically inherited genes are more evenly distributed among processes (fig. 2).

**Sister Relationships across All Trees**

The topologies of the single gene trees from our pipeline show a striking relationship between *Entamoeba* and the parabasalid taxa *T. vaginalis*, *Histomonas meleagris*, and *Pentatrichomonas hominis*. To investigate further, we took a dual approach to assessing sister relationships to *Entamoeba* spp.—one relying on the output of our phylogenomic pipeline and another independent of our pipeline, relying only on estimates of homology in OrthoMCL (Li et al. 2003; Chen et al. 2006; see Materials and Methods).

We examined sister relationships in two types of trees from our phylogenomic pipeline: the 22 trees with *Entamoeba* spp. sister to a single eukaryotic taxon, which were nested within bacteria or archaea (interdomain LGT followed by intradomain LGT; fig. 1C and table 1) and the 109 trees with Entamoeba and a nonamoebzan sister in trees where the monophyly of Amoebozoa is rejected by the AU test (eukaryote-to-eukaryote LGT; fig. 1D and table 1). To be conservative in our estimation of sister relationships, we bootstrapped the 109 trees in RAxML, and kept only the 22 trees with a BS support of >80% for a sister relationship between *Entamoeba* spp. and one eukaryotic taxon. This approach identified 44 gene trees with a sister relationship that can be identified with confidence and among these, the most common sister taxon was *T. vaginalis* (22 trees; fig. 3A and supplementary data S2 and table S3, Supplementary Material online). Other taxa had supported sister relationships with *Entamoeba* in many fewer trees including kinetoplastids (four trees), *Giardia* spp. (three trees), apicomplexa (three trees), mixed Excavata (three trees), and microsporidia (two trees; supplementary table S3, Supplementary Material online). These rarer occurrences may be due to aberrant LGT events, convergence, or may appear from biases in methods.

We also examined sister relationships among genes present only in *Entamoeba* spp and one other eukaryote based on clusters of orthologs determined by OrthoMCL as this approach is independent of the parameters of our pipeline. Here, we found the same association between *Entamoeba* spp. and *T. vaginalis*. Of the 147 genes that passed the stringent requirements (see Materials and Methods), the largest number (42 groups) is consistent with vertical inheritance as they include only *Entamoeba* spp. and *Dictyostelium discoideum*, the only other amoebozan taxon represented in OrthoMCL. The second most common association was between *Entamoeba* spp. and *T. vaginalis*, which was found in
29 genes (fig. 3B). The remaining genes contained *Entamoeba* spp. plus taxa found in only a few groups (e.g., at most nine groups for *Arabidopsis thaliana*), with many taxa being represented in only one orthologous group (fig. 3B and supplementary table S4, Supplementary Material online).

**Discussion**

**Phylogenetic Trees and Patterns of Gene Presence/Absence Suggest Widespread LGT in the Genome of *Entamoeba***

Our dual approach of assessing taxon-rich tree topologies and gene presence/absence patterns reveals the impact of LGT in the genomes of *Entamoeba* spp. As interpretation of past LGT events is challenging given problems inherent in analyzing evolutionary history of single genes over long periods of time, we rank our findings roughly in order of greatest to least confidence. We identified 116 candidate interdomain LGT events, 22 putative instances of interdomain LGT followed by intradomain LGT, plus 287 possible intradomain LGT events (table 1 and fig. 1). We recognize that our attempts to be conservative may have eliminated a number of “true” LGT events and that additional examples may be identified as taxon sampling improves.

The impact of LGT on *Entamoeba* spp. has been previously recognized (Yang et al. 1994; Rosenthal et al. 1997; Loftus et al. 2005; Clark et al. 2007) though our approach yields a longer list of candidate genes. Comparisons across studies are challenging because of concurrent changes in methodologies for genome assembly and LGT detection; nevertheless, we find the 68 genes identified in Clark 2007 (supplementary table S2, Supplementary Material online) plus more. The differences emerge, in part, because we use a dual approach of assessing tree topologies and identifying cases where genes are only found in *Entamoeba* spp. and bacteria and/or archaea. The genes we retain include nine candidate LGTs originally identified by Loftus et al. (2005) that were removed from consideration by Clark et al. (2007) as increased taxon sampling had revealed eukaryotic sisters. We retained genes if *Entamoeba* spp. plus one other eukaryotic group was nested among bacteria and/or archaea (i.e., interdomain followed by intradomain LGTs) and found that six of the nine genes rejected by Clark et al. (2007) showed *Entamoeba* spp. sister to *T. vaginalis* (see below).

Compared with vertically inherited genes, the genes identified by our approach as candidate LGTs are more likely to be involved in metabolism (fig. 2), a trend noted in several recent studies of LGT into other microbial eukaryotes (Embley 2006; Ginger 2006; Tsaousis et al. 2012; Imanian and Keeling 2014). These genes may make effective targets for drug discovery efforts, as drugs targeting genes with bacterial origins are less likely to have an impact on their human host (Umejiego et al. 2008; Keeling 2009; Alsmark et al. 2013).

**Phylogenetic Trees and Orthology Estimates Support a Specific Ancestral Relationship with Parabasalids***

Both phylogenetic trees and presence/absence data show a specific relationship between genes found in *Entamoeba* spp. and those in *T. vaginalis* and sometimes other parabasalids. (Although the entire genome has been sequenced from *T. vaginalis*, there are only limited EST data from *P. hominis* and *H. meleagridis* (supplementary table S1, Supplementary Material online) making assessment of the relationship with parabasalids as a whole more difficult.) We used two independent approaches here, first assessing the sister relationships in phylogenetic trees and second analyzing patterns of gene presence/absence in all orthologous groups from OrthoMCL. Both analyses show *T. vaginalis* is more highly represented than any other nonamoeboid taxon (fig. 3 and supplementary table S4, Supplementary Material online).

Inspection of individual trees from our phylogenomic pipeline yields some intriguing patterns. For example, phylogenetic analyses of a lipase-containing protein, OG5_129115 (abbreviation refers to orthologous group as determined in OrthoMCL; Li et al. 2003) place the *T. vaginalis* sequence sister to *E. dispar* and *E. invadens* and nested among other amoeboid taxon (fig. 4). Analyses of OG5_127586 (a hypothetical conserved protein) show three *Entamoeba* paralogs, each sister to *T. vaginalis* paralogs, suggesting acquisition of a recently expanded gene family (fig. 5). In this case, one of these clades also contains *Blastocystis hominis*, another
FIG. 4.—Exemplar maximum-likelihood tree of lipase-containing protein (OrthoMCL cluster OGS_129115) showing putative LGT from Entamoeba to Trichomonas vaginalis. Maximum-likelihood tree with monophyletic Amoebozoa interrupted by a T. vaginalis sequence sister to E. dispar and E. invadens suggests LGT from an ancestor of Entamoeba to T. vaginalis. Arrow points to clades of interest. Sequences are labeled with their major clade: Op, Opisthokont; Pl, Archaeplastida; Sr, SAR; Ex, Excavata; Am, Amoebozoa. Branches within the clades of interest are labeled with bootstrap support; branches outside the clades of interest with bootstrap support >70% are bold. Scale bar represents number of changes.
FIG. 5.—Maximum-likelihood tree of hypothetical conserved protein (OrthoMCL cluster OG5_127586) showing putative LGT of multiple paralogs. Maximum-likelihood tree with multiple sister groups of *Entamoeba* and *Trichomonas vaginalis* suggests a relationship between *Entamoeba* and *T. vaginalis* more recently than the gene duplication event that created the paralogs. Sequences are labeled by their major clade (see fig. 4 legend for abbreviations). Branches within the clades of interest are labeled with bootstrap support; branches outside the clades of interest with bootstrap support >70% are bold. Scale bar represents number of changes.
distantly related mucosal parasite. This topology suggests the possibility of parallel gene transfers into organisms within environmental niches, as has previously been hypothesized (Ricard et al. 2006; Alsmark et al. 2013; Clarke et al. 2013).

Given our taxon sampling, there is no clear directionality in the observed LGT between the ancestors of Entamoeba and Trichomonas: Of the 22 trees analyzed, two are consistent with transfer from Entamoeba to Trichomonas (i.e., T. vaginalis is nested among Amoebozoa), three appear to be the opposite, and in the remaining 17 trees there is poor support at deeper nodes (supplementary tables S2 and S3, Supplementary Material online). We anticipate that directionality can be determined with additional sampling of whole genomes from diverse amoebozoan and parabasalid lineages.

An association between E. histolytica and T. vaginalis has been mentioned in several studies, as organisms that have both been impacted by LGT and some putative gene sharing has been noted (Stanley 2005; Keeling and Palmer 2008; Alsmark et al. 2009, 2013). Borrowing genes, particularly metabolic genes (fig. 2), may be one way to adapt to a new, extremely different, environment. Entamoeba emerges as a parasite from a mostly nonparasitic clade (Amoebozoa) and the transition to parasitism may have occurred in a similar environment (i.e., epithelial cells) as the evolution of parasitism in parabasalids. Thus, it is possible that genes shared between these taxa are due to borrowing genes from a pool of donor lineages available in the niche environment that they share (Alsmark et al. 2013; Clarke et al. 2013). However, the sister relationship between these two taxa in our analyses suggests more than a shared tendency to independently pick up genes from their environment. The frequent sister relationship, along with the patterns of shared paralogs (e.g., fig. 5), suggests a past relationship between the ancestors of the two taxa. While this is speculative, endosymbiosis among eukaryotes does occur, for example, Excavata symbionts are found within the macronucleus of some ciliates (e.g., Fokin et al. 2008, 2014; Gomaa et al. 2014). Moreover, Tanifuji et al. (2011) document an example of an amoebozoan, Neoparamoeba pemaquidensis, that hosts a kinetoplast (an Excavata, like the parabasalids) endosymbiont—a relationship similar to the one we suggest here.

Caveats

Single gene trees are prone to error; any one gene tree that can be explained by LGT might also be explained by misidentified orthologs, gene loss, limited taxon sampling or any of a number of other causes (see Kurland et al. 2003; Martin 2005). Although some of the discordance we see is no doubt due to this sort of error, the large number of discordant genes and the independent evidence of bias in both gene function and sister taxon relationships point to something beyond error in our analyses. Nevertheless, it is important to consider alternative explanations. One possible explanation is convergent evolution: Eukaryotic parasites living in a similar niche experience similar selective pressures, including host defenses and a microaerophilic environment. However, although convergent evolution is often seen in protein functional domains (Bork and Doolittle 1992; Gandbhir et al. 1995; Tomii et al. 2012), convergence at the sequence level across the length of a gene is unlikely (Doolittle 1994; Oren 1995; Gogarten and Olendzenski 1999). Another possibility is that we still have inadequate taxon sampling to depict vertical donor lineages accurately, though this is less likely to impact cases of interdomain transfers.

Although these caveats mean that we may be mistaken in our interpretation of individual gene trees, we believe that our conservative approach—relying on strong phylogenetic support for taxon-rich gene trees (i.e., high BS support and AU tests) and identifying genes present in only Entamoeba spp. plus bacteria/archaea (i.e., presence/absence data)—may well have led to an underestimate of the number of interdomain LGTs. Moreover, our hypothesis of past relationships between the ancestors of Entamoeba spp. and parabasalids is testable as additional taxa are sampled from close relatives.

Supplementary Material

Supplementary data and tables S1–S4 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org).

Acknowledgments

This work was supported by a U.S. National Institutes of Health award 1R15GM097722-01 and U.S. National Science Foundation award DEB-1208741. The authors thank Rob Dorit (Smith College) and J. Gordon Burleigh (Univ. Florida) for helpful conversations.

Literature Cited

Ali V, Hashimoto T, Shigeta Y, Nozaki T. 2004. Molecular and biochemical characterization of D-phosphoglycerate dehydrogenase from Entamoeba histolytica. A unique enteric protozoan parasite that possesses both phosphorylated and nonphosphorylated serine metabolic pathways. Eur J Biochem. 271:2670–2681.

Ali V, Shigeta Y, Tokumoto U, Takahashi Y, Nozaki T. 2004. An intestinal parasitic protist, Entamoeba histolytica, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. J Biol Chem. 279:16863–16874.

Alsmark C, et al. 2013. Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biol. 14:R19.

Alsmark UC, et al. 2009. Horizontal gene transfer in eukaryotic parasites: a case study of Entamoeba histolytica and Trichomonas vaginalis. Methods Mol Biol. 532:489–500.

Anderson IJ, Loftus BJ. 2005. Entamoeba histolytica: observations on metabolism based on the genome sequence. Exp Parasitol. 110:173–177.

Anderson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62:1182–1197.

Anderson JO, Soguren AM, Davis LAM, Embley TM, Roger AL. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol. 13:94–104.
Stanley SL. 2003. Amoebiasis. Lancet 361:1025–1034.
Stanley SL Jr. 2005. The Entamoeba histolytica genome: something old, something new, something borrowed and sex too? Trends Parasitol. 21:451–453.
Tanifuji G, et al. 2011. Genomic characterization of Neoparamoeba pemaiquidensis (amoebozoa) and its kinetoplastid endosymbiont. Eukaryot Cell. 10:1143–1146.
Tomii K, Sawada Y, Honda S. 2012. Convergent evolution in structural elements of proteins investigated using cross profile analysis. BMC Bioinformatics 13:11.
Tsaoasis AD, et al. 2012. Evolution of Fe/S cluster biogenesis in the anaerobic parasite Blastocystis. Proc Natl Acad Sci U S A. 109:10426–10431.
Umejiego NN, et al. 2008. Targeting a prokaryotic protein in a eukaryotic pathogen: identification of lead compounds against cryptosporidiosis (vol 15, pg 70, 2008). Chem Biol. 15:200–200.
Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. Bioessays 35:829–837.
Yang W, Li E, Kairong T, Stanley SL Jr. 1994. Entamoeba histolytica has an alcohol dehydrogenase homologous to the multifunctional adhE gene product of Escherichia coli. Mol Biochem Parasitol. 64:253–260.
Zmasek CM, Godzik A. 2011. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. Genome Biol. 12:13.

Associate editor: Bill Martin