

Research article

Open Access

The iron-sulfur cluster assembly genes *iscS* and *iscU* of *Entamoeba histolytica* were acquired by horizontal gene transfer

Mark van der Giezen*, Siân Cox and Jorge Tovar

Address: School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

Email: Mark van der Giezen* - mark.vandergiezen@rhul.ac.uk; Siân Cox - s.s.e.cox@rhul.ac.uk; Jorge Tovar - j.tovar@rhul.ac.uk

* Corresponding author

Published: 20 February 2004

Received: 15 December 2003

BMC Evolutionary Biology 2004, 4:7

Accepted: 20 February 2004

This article is available from: <http://www.biomedcentral.com/1471-2148/4/7>

© 2004 van der Giezen et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Iron-sulfur (FeS) proteins are present in all living organisms and play important roles in electron transport and metalloenzyme catalysis. The maturation of FeS proteins in eukaryotes is an essential function of mitochondria, but little is known about this process in amitochondriate eukaryotes. Here we report on the identification and analysis of two genes encoding critical FeS cluster (Isc) biosynthetic proteins from the amitochondriate human pathogen *Entamoeba histolytica*.

Results: *E. histolytica* IscU and IscS were found to contain all features considered essential for their biological activity, including amino acid residues involved in substrate and/or co-factor binding. The IscU protein differs significantly from other eukaryotic homologs and resembles the long type isoforms encountered in some bacteria. Phylogenetic analyses of *E. histolytica* IscS and IscU showed a close relationship with homologs from *Helicobacter pylori* and *Campylobacter jejuni*, to the exclusion of mitochondrial isoforms.

Conclusions: The bacterial-type FeS cluster assembly genes of *E. histolytica* suggest their lateral acquisition from epsilon proteobacteria. This is a clear example of horizontal gene transfer (HGT) from eubacteria to unicellular eukaryotic organisms, a phenomenon known to contribute significantly to the evolution of eukaryotic genomes.

Background

After malaria, *Entamoeba histolytica* is the second leading cause of death due to parasitic disease in humans [1]. *E. histolytica* has been cited as infecting one tenth of the world population, although it is now known that these infections are caused by two very similar species, *E. histolytica* and *E. dispar*. The former is the cause of all invasive disease, with an estimated 50,000 to 100,000 fatalities each year [2]. This human parasite was traditionally considered a classic example of a primitive eukaryote due to its apparent lack of 'typical' eukaryotic cell structures such as mitochondria, peroxisomes, Golgi apparatus and endoplasmic reticulum [3]. The lack of morphologically iden-

tifiable mitochondria led to the suggestion that its ancestors predate the endosymbiotic acquisition of this organelle [4], despite the observation that *Entamoeba* branches after well established mitochondrial groups in ribosomal RNA phylogenies [5].

The "primitively amitochondrial" view was overturned by the discovery of genes encoding mitochondrial proteins (e.g., chaperonin 60 (Cpn60), mitochondrial-type Hsp70 (mtHsp70), pyridine nucleotide transhydrogenase (PNT)), and by the demonstration that mitochondrial remnant organelles (mitosomes) housing chaperonin Cpn60 have been retained in this organism [6-9]. Several

lines of evidence support the mitochondrial ancestry of mitosomes: i) Cpn60 and mtHsp70 cluster with mitochondrial homologs to the exclusion of prokaryotic sequences in phylogenetic reconstructions; ii) Cpn60, mtHsp70 and PNT contain amino terminal regions rich in hydroxylated and positively charged amino acids, reminiscent of mitochondrial/hydrogenosomal targeting presequences; iii) Deletion of amino acids 2–15 from the putative targeting presequence of Cpn60 leads to an accumulation of the truncated protein in the cytosol, a phenotype that can be reversed by the addition of a functional mitochondrial targeting signal from *Trypanosoma cruzi* Hsp70 to the truncated protein [8].

Since the discovery of mitosomes in *E. histolytica*, mitochondrial remnant organelles have also been identified in the microsporidian *Trachipleistophora hominis* [10], the apicomplexan *Cryptosporidium parvum* [11] and, most recently, in the diplomonad *Giardia intestinalis* [12]. *Giardia* mitosomes have been shown to function in FeS cluster biosynthesis and FeS protein maturation [12], essential mitochondrial functions of eukaryotic organisms [13]. FeS proteins are involved in energy metabolism, DNA repair, transcriptional regulation, and biosynthesis of nucleotides and amino acids [14]. The identification of genes encoding putative Isc proteins in the genomes of all amitochondrial protists sequenced so far [15–19] suggests that this mitochondrial function might have been retained in all amitochondrial protists and may be a general functional feature of all mitochondrion-derived organelles [12,16,20–22].

Here we report on the cloning, structural characterization and phylogenetic analysis of *E. histolytica* genes encoding Isc proteins. Both *E. histolytica* IscU and IscS homologs were found to contain all the structural features required for their biological activity, including substrate and co-factor binding sites, suggesting a fully operational FeS cluster biosynthetic pathway in *E. histolytica*. Phylogenetic analyses show that both Isc proteins have a different evolutionary history to that of mitochondrial homologs, indicating their lateral acquisition from bacteria. Moreover, the observation that both proteins seem to have been acquired from the same bacterial taxon might suggest a single transfer event of a small bacterial Isc operon.

Results and Discussion

Identification and primary sequence analyses of *E. histolytica* genes encoding the FeS assembly proteins IscS and IscU

BLAST searches of preliminary data generated by the *E. histolytica* genome-sequencing project revealed clones with extensive sequence similarity to the *G. intestinalis* *iscS* gene. PCR amplification of *E. histolytica* genomic DNA using primers based on these putative *E. histolytica* *EhiscS*

sequences and on a putative *EhiscU* sequence (accession number: AY040613) generated products of the expected size. DNA sequencing confirmed the identity of the amplified clones. The 5' untranslated regions of *EhiscU* and *EhiscS* contain distinct putative promoter elements reported to be typical for *E. histolytica* [23]. All three conserved regions are present in the first 40 bases upstream the initiation codon of *iscU* and *iscS* (Fig. 1), suggesting both genes are functional, although the GAAC-element is less well conserved in the *iscU* promoter region. The *E. histolytica* IscU protein is 348 amino acids in length and has a predicted molecular mass of 38.9 kDa and a predicted isoelectric point of 5.71. Its large size indicates it is a long-form IscU, similar to the one described for *Azotobacter vinelandii* [24], and not a short form as found in other eukaryotes (Fig. 2). For IscS these values are 390 amino acids, 42.8 kDa, and 5.92, respectively. The GC values for the *iscU* (*iscS*) genes are 33 % (32 %) for the coding region, 29 % (25 %) for the 5' untranslated region and 29 % (18 %) for the 3' untranslated region (250 bp each). These values are in agreement with GC values reported for other *E. histolytica* genes based on 75,615 codons analyzed [25]. Codon usage is also similar to *E. histolytica* codon usage and no introns are present in either of these two genes.

Both *E. histolytica* IscU and IscS contain structural motifs typical of FeS assembly proteins. Pfam (PF01106, PF01592), PRODOM (PD002830), and InterProScan (IPR001075, IPR002871) motifs characteristic of IscU and NifU proteins are present in the *E. histolytica* homolog (Fig. 2A) [26–28]. *E. histolytica* IscS contains Pfam (PF00266), PROSITE (PS00595), and InterProScan (IPR000192) motifs that are normally associated with aminotransferase class V proteins, a subfamily of the aminotransferase proteins. IscS is one of the eight members of the class V subfamily (Fig. 2A). As indicated above, IscU has an extension at the carboxy-terminus relative to most IscU homologs. This extension is also present on the *A. vinelandii* *NifU* gene whose amino-terminal part is homologous to that of IscU. In addition, this C-terminal extension is similar to a completely different gene from *Saccharomyces cerevisiae*, *Nfu1* (NifU-like in Fig. 2B). Since *Nfu1*- and *Isu*-like sequences are part of the same gene in *Azotobacter*, *Campylobacter*, *Entamoeba* and *Helicobacter* it could be inferred that both proteins interact with each other when found on two separate genes. Such informative fusion proteins (or Rosetta Stone sequences) indicate an interaction between protein pairs [29]. The existence of long IscU isoforms would therefore suggest that the *Nfu1* and *Isu1/2* proteins do interact in yeast as postulated by Garland et al. [30].

Both proteins align along their whole length to homologous proteins from other organisms (Fig. 3). Residues

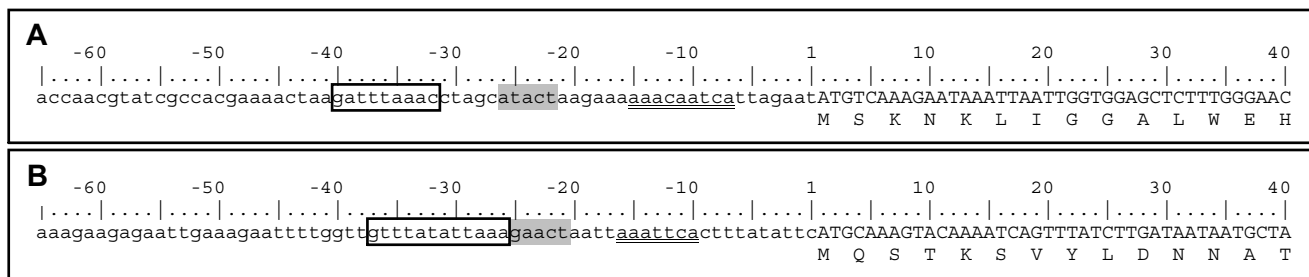


Figure 1
Overview of the 5'-flanking region of *E. histolytica* *iscU* (A) and *iscS* (B) encoding genes. The three typical upstream regulatory elements are depicted as by Purdy et al [23]; the putative initiator element, double underlined; the 'GAAC'-element, grey box; and the putative TATA element, boxed.

implicated in function are conserved in both IscU and IscS proteins. The three cysteine residues that are conserved in *Escherichia coli* IscU which provide a scaffold for the assembly of iron-sulfur clusters [14] are conserved in the *E. histolytica* protein (Fig. 3A). In addition, in *E. coli* one of these IscU cysteines interacts with a conserved cysteine from IscS which is also present in the *E. histolytica* IscS (Figs. 3A and 3B). Most residues considered to be important for IscS function are also present on the *E. histolytica* protein (Fig. 3B). To test whether the *E. histolytica* IscS protein assumes a normal three-dimensional conformation, this protein was modeled on the solved NifS protein structure from *Thermotoga maritima*. The overall topology of both proteins is quite similar and the force field energy of the computed *E. histolytica* IscS model is -13,800 kJ/mol, indicating an energetically plausible model [31]. The putative active site architecture of *E. histolytica* IscS and the solved active site of *T. maritima* NifS show similar structures (Fig. 4). The ring of the cofactor vitamin B6 (or pyridoxal-5'-phosphate; PLP) is sandwiched between EhHis106/TmHis99 and EhThr184/TmVal179 and further fixed by residues EhAsp182/TmAsp177 and EhGln185/TmGln180. The phosphate-group is anchored by six hydrogen bonds from EhThr76/TmThr71, EhHis207/TmHis202, EhThr198/TmSer200, and EhThr243/TmThr238 [32]. The presence of all residues considered to be important for IscU and IscS activity on the *E. histolytica* proteins suggest that these proteins are indeed involved in FeS cluster assembly.

No N-terminal or C-terminal organelle targeting domains could be unambiguously identified in *E. histolytica* IscS/U proteins using subcellular localization and targeting prediction software (e.g., PSORT II [33], MitoProt [34], NNPSL [35]). The C-terminal signature motif which is considered to be characteristic of proteobacterial and eukaryotic IscS proteins [20] is not present in homologs

from *E. histolytica*, *Campylobacter* or *Azotobacter* (Fig. 3B). Because these organisms all possess the long-type IscU isoforms, it is possible that the extended IscU protein might negate the need for the C-terminal signature residues on the interacting IscS protein. However, functional studies using deletion mutants are needed to confirm this hypothesis.

Phylogenetic analyses of the *E. histolytica* FeS cluster assembly proteins

Bayesian and maximum likelihood (ML) phylogenetic analyses of *E. histolytica* IscU and IscS protein sequences revealed that the *Entamoeba* Isc proteins form a well supported clade with *Helicobacter pylori* and *Campylobacter jejuni* – two bacteria encountered in the human digestive tract – to the exclusion of all other prokaryotic and eukaryotic homologs (Fig. 5). All three independent Bayesian analyses converged on the same tree with similar posterior probabilities. For IscU, the ML tree had a slightly better likelihood than the Bayesian tree, while for IscS both trees had similar likelihoods. The overall topologies of IscS and IscU phylogenetic trees are very similar to each other and major taxonomic clades like plants, animals, and fungi are well conserved. The position of the microsporidium *Encephalitozoon cuniculi* in the IscU tree is poorly resolved as indicated by the very low support for this node at the base of the metazoa, contrary to its well-documented association with fungi [36].

The position of *Rickettsia prowazekii* IscS basal to the eukaryotes suggests that eukaryotic IscS proteins originated from the mitochondrial endosymbiont, since this bacterium is considered to be a close relative to the mitochondrial ancestor. Indeed, the mitochondrial ancestry of *E. cuniculi*, *T. vaginalis* and *G. intestinalis* IscS proteins is strongly supported by their clustering with mitochondrial homologs [15,20,37]. For IscU, the base of the eukaryotic

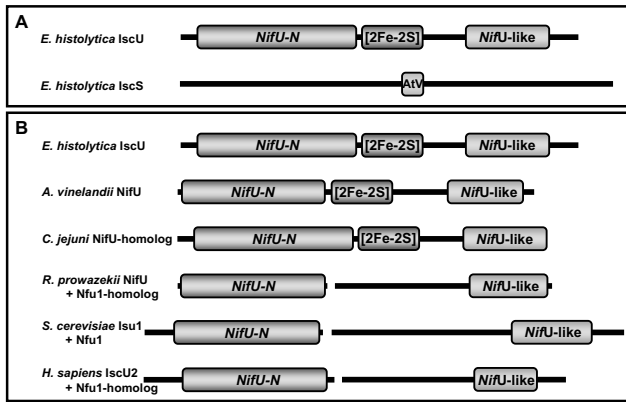


Figure 2
Primary sequence features of the *E. histolytica* IscU and IscS proteins. A. Schematic drawing of the *E. histolytica* IscU and IscS protein sequences indicating the positions and sizes of Pfam [26] signature motifs PF01592 (NifU-N), PF04324 ([2Fe-2S]), and PF01106 (NifU-like) on the putative IscU protein and Pfam motif PF00266 (aminotransferase class V – AtV) on the putative IscS protein. The presence of these domains on a protein is used by the various databases to classify a protein and to infer its function. B. Comparison of the *E. histolytica* IscU protein depicted as above with homologous proteins from *Azotobacter vinelandii* (NifU), *Campylobacter jejuni* (NifU), *Rickettsia prowazekii* (NifU and RP667), *Saccharomyces cerevisiae* (Icu1 and Nfu1), and *Homo sapiens* (IscU2 and HIRIP5).

clade is not well resolved. Animals and plants cluster together with a proteobacterial sister clade containing the α -proteobacterium *R. prowazekii*, while fungi, *G. intestinalis*, and the alveolates are basal to this clade. However, the well-supported clustering of *E. histolytica* Isc proteins with homologs from the bacteria *H. pylori* and *C. jejuni*, to the exclusion of all other eukaryotes, suggests that *E. histolytica* acquired its *isc* genes laterally from ϵ -proteobacteria (Fig. 5). This suggestion is further supported by the fact that *Campylobacter*, *Helicobacter* and *E. histolytica* all possess long form IscU proteins to the exclusion of the short isoforms found in eukaryotic organisms and in many bacterial taxa (see orange branches in Fig. 5).

Mitochondrial-type IscS/U proteins have been identified in several amitochondrial eukaryotes including *Giardia*, *Encephalitozoon*, *Trichomonas* and *Cryptosporidium*, and there is significant direct and indirect evidence that these proteins are targeted into their highly derived mitochondrion-related organelles [12,15,16,20]. Thus, *E. histolytica* appears to be unique amongst eukaryotic organisms that contain mitochondrion-related organelles in harbouring bacterial-type IscS/U proteins. That no mitochondrial-

type IscS/U proteins have thus far been identified in *E. histolytica* would suggest that its original mitochondrial-type *iscS/U* genes were replaced during the course of evolution by the more recently acquired bacterial homologs. However, since the *E. histolytica* genome has not yet been fully sequenced, the possibility that mitochondrial type *iscS/U* genes might have escaped detection cannot be formally excluded.

Since both *E. histolytica* Isc proteins form a strongly supported clade with homologs from gut bacteria, we investigated whether other intestinal inhabitants would form part of this clade. The genomes of 23 bacterial and 2 eukaryotic inhabitants of the human gut were screened using *E. histolytica* IscU and IscS as query sequences, but no additional homologs were identified. Only a fraction of the estimated 400–500 bacteria species living in the human intestine [38] have been sequenced and therefore we may not have been able to identify any other members of this clade due to sampling limitations. Nevertheless, the most parsimonious explanation for the clustering of *E. histolytica* Isc proteins with those of bacteria is that *E. histolytica*, or its ancestors, acquired its *iscS/U* genes by horizontal gene transfer (HGT), a well-documented contributor to prokaryotic and eukaryotic genome evolution. In higher eukaryotes the most obvious example of HGT is the relocation of genes from endosymbiosis-derived organelles to the cell nucleus, which might be regarded as a special case of HGT. However, over the past few years evidence has accumulated of the frequent incorporation of genes into the genomes of microbial eukaryotes by HGT [39-47]. The transfer of bacterial genes into eukaryotes might occur in several possible ways. One hypothesis is the 'you are what you eat' gene transfer ratchet of HGT which suggests that when a genome is continuously bombarded with DNA, some of these genes might eventually replace the host's own genes [48]. Since both *Helicobacter* and *Campylobacter* occupy the same ecological niche as *E. histolytica*, an avid consumer of gut bacteria, HGT via this mechanism seems plausible. Establishing unequivocally the timing of HGT will be important to test this hypothesis.

Analysis of the organization of Isc/Nif loci on the genomes of several bacteria revealed the presence of a small Isc operon consisting exclusively of IscU and IscS in *H. pylori* and *C. jejuni*, whilst the well-studied *E. coli* and *A. vinelandii* *isc* operons contained several other genes involved in FeS cluster assembly (see Fig. 6). This observation provides a mechanistic explanation for the presence of two interacting proteins with similar ancestry in the genome of *E. histolytica*. It is possible that *E. histolytica* might have incorporated the entire *isc* operon from *Helicobacter/Campylobacter*, or from their ancestors, into its genome in a single transfer event. Once freed from the

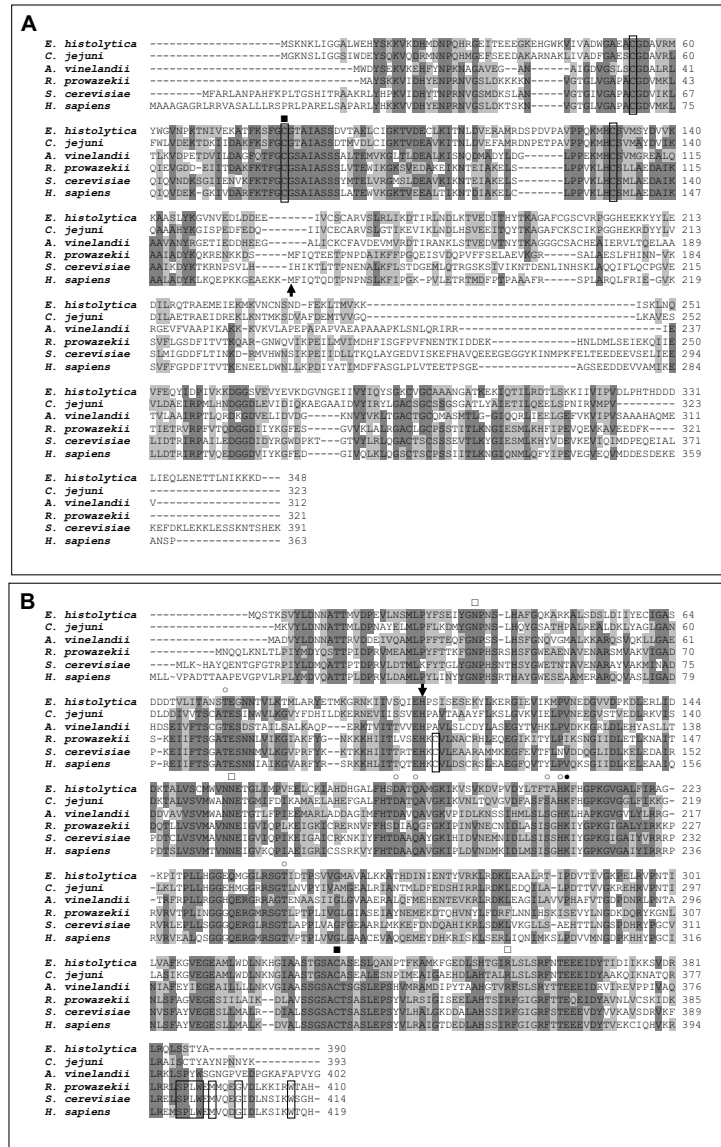


Figure 3
Alignment of the putative *E. histolytica* IscU and IscS with homologs from *C. jejuni*, *A. vinelandii*, *R. prowazekii*, *S. cerevisiae*, and *H. sapiens*. A. Alignment of the *E. histolytica* long-form IscU with similar isoforms from *C. jejuni* and *A. vinelandii*. In addition, short-form IscU homologs from *R. prowazekii*, *S. cerevisiae*, and *H. sapiens* are aligned concatenated with their NfuI homologs (arrow indicates start of NfuI homologs) which resemble the C-terminal extension found on the long-form IscU. The conserved cysteine residues which provide a scaffold for the IscS-directed sequential assembly of labile FeS-clusters [14] are boxed. The cysteine residue that forms a disulfide bridge with a conserved cysteine residue on IscS (see B) is indicated by a closed square (■). The yeast NfuI mitochondrial transit peptide has been deleted. B. Alignment of the *E. histolytica* IscS with homologs from the above mentioned organisms. Important residues for function are as described by Tachezy, Sánchez and Müller [20]; the conserved lysine involved in co-factor binding (pyridoxal-5'-phosphate, PLP) is indicated by a closed circle (●), other residues involved in PLP interaction are indicated by open circles (○), the cysteine residue that forms a disulfide bridge with a cysteine residue on IscU (see A) is indicated by a closed box (■), residues involved in substrate binding (L-cysteine) are indicated by open squares (□), the conserved histidine involved in substrate deprotonation is indicated by an arrow. Typical eukaryotic/eubacterial conserved cysteine and C-terminal residues are boxed. Note that organisms that contain a long-form IscU (see Fig. 2B and 3A) do not have these conserved residues suggesting that the C-terminal IscU extension might take over the role of these residues. Part of the mitochondrial transit peptides from the yeast and human IscS homologs have been omitted (-) for reasons of clarity. Amino acids were shaded according to similarity/identity scores: dark grey indicates fully conserved residues while light grey indicates similar residues according to the PAM250 matrix [59].

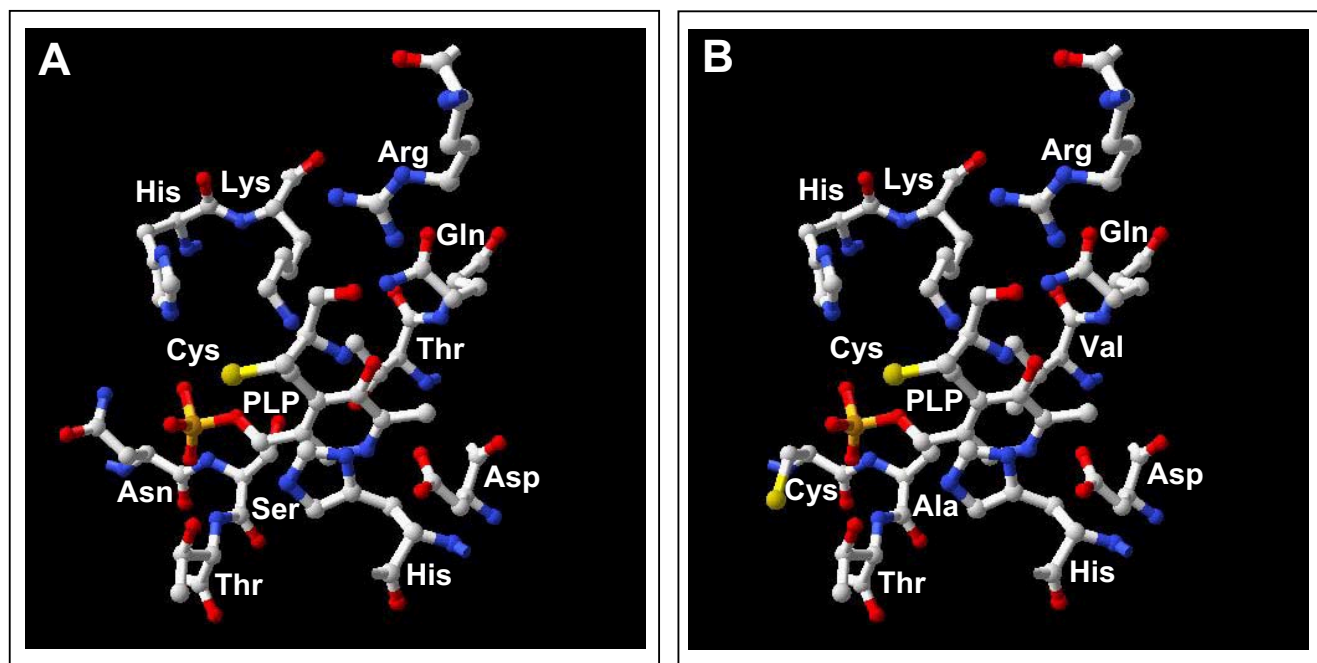


Figure 4
Model of the active site of *E. histolytica* IscS and *Thermotoga maritima* NifS. The *E. histolytica* IscS putative three-dimensional structure (A) was deduced using the conceptually translated *iscS* sequence. The previously solved crystal structure of *T. maritima* NifS (B) [PDB accession number: 1EG5, [32]] was used as a template. The *E. histolytica* IscS sequence was aligned to the *T. maritima* NifS sequence using DeepView v3.7 [<http://www.expasy.org/spdbv/>, [60]] and manually improved based on an independent Clustal W alignment [54]. Only residues involved in co-factor (PLP) and substrate (Cys) binding were selected in order to show the active site.

constraints of operon-type prokaryotic gene expression, the *iscS/U* genes might have become separated in the *E. histolytica* genome during the course of evolution.

Conclusions

E. histolytica or its ancestors appear to have acquired their *iscS/U* genes by HGT from ϵ -proteobacteria. The apparent absence of mitochondrial-type IscS/U proteins in an organism with mitochondrion-bearing ancestors such as *E. histolytica* suggests that its original mitochondrial *iscS/U* genes might have been replaced with the more recently acquired bacterial homologs. This finding, like several other recently reported cases of prokaryote to eukaryote gene transfers [39-47], highlights the important role played by HGT in protozoan genome evolution. Since no recent HGT events from prokaryotes to humans have been detected in the human genome [49], HGT from bacteria to protozoan parasites might have important implications for public health. Targeting enzymes or metabolic pathways of bacterial origin in human pathogens should have more severe consequences for the parasite than for its host, making these proteins promising targets for chemotherapy.

Methods

Organism and DNA isolation

E. histolytica HM-1:IMSS clone 9 was maintained axenically by subculture in YI-S medium with 15% adult bovine serum as described [50]. *Entamoeba* genomic DNA was isolated using cetyltrimethylammonium bromide (CTAB) according to Clark [51].

Cloning and sequencing of the *E. histolytica* *iscS* and *iscU* genes

Standard recombinant DNA techniques were used as described elsewhere [52]. PCR was performed on isolated *E. histolytica* genomic DNA. Primers were designed using Primer3 [53]. The *EhiscU* gene was amplified using primers based on a NifU-like *E. histolytica* sequence (accession number AY040613). The primers were Eh_IscU_936F, 5'-CCA ACG TAT CGC CAC GAA AA-3' and Eh_IscU_2270R, 5'-GCA AAA CAA AGT ATG GCA GAA GCA-3' for forward and reverse primers, respectively. The *EhiscS* gene was identified on the *E. histolytica* genome by BLAST searches of preliminary data generated by the *Entamoeba* genome sequencing project [17] using *G. intestinalis* *GiiscS* (accession number AAK39427) as the query sequence. Putative

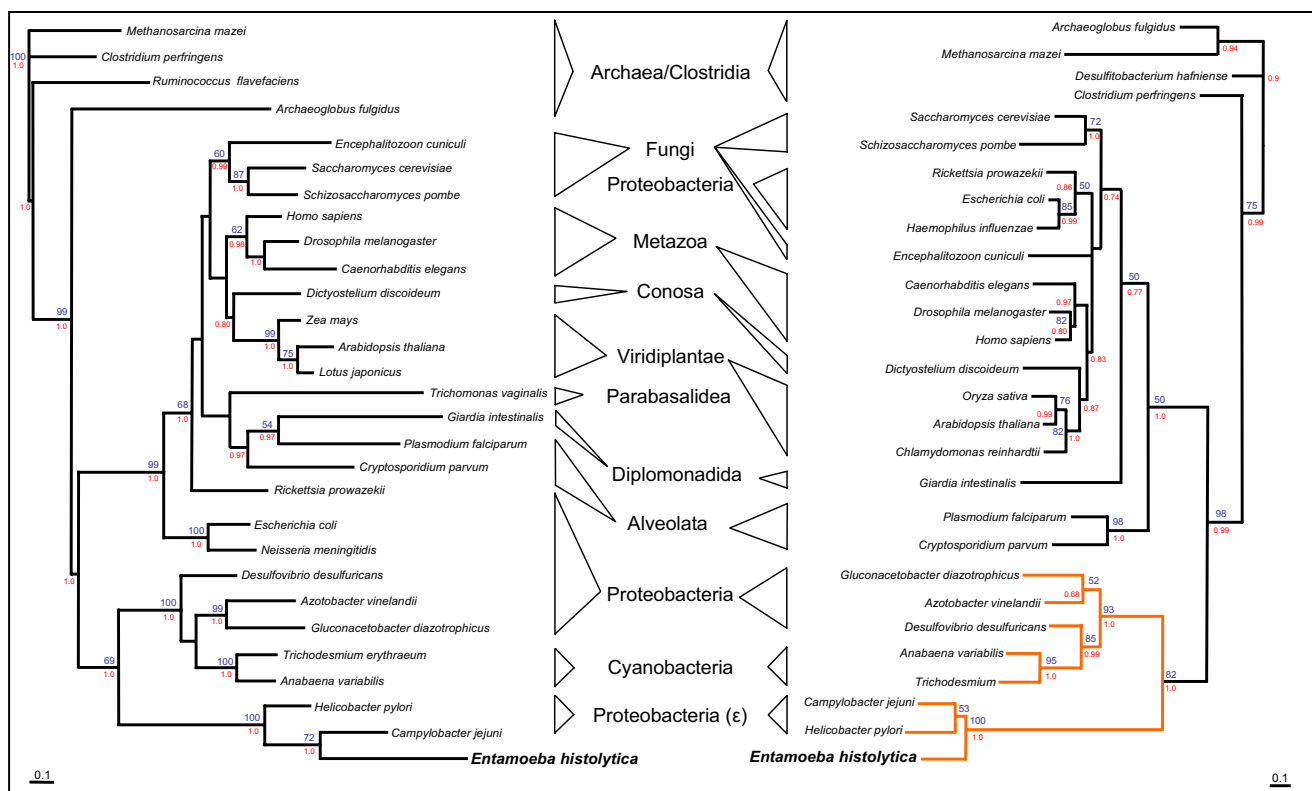


Figure 5
Phylogenetic analysis of *E. histolytica* IscS and IscU protein sequences using a similar taxonomic sampling.
 Depicted are unrooted maximum likelihood phylogenetic trees of 29 IscS (left) and 28 IscU (right) protein sequences. The *E. histolytica* sequences are recovered as part of a well supported monophyletic group comprising the gut bacteria *H. pylori* and *C. jejuni*. The orange branches represent those sequences containing the long IscU isoform. Numbers in red represent posterior probabilities as determined by MrBayes [55] where a value of 1.0 represents maximum support (only values above 0.75 are shown). Values in blue represent bootstrap values as determined using PHYML [57], only bootstrap values above 50% are shown.

EhiscS gene sequences (1000 bases up- and downstream of the ORF) were used for primer design. The *EhiscS* coding region was amplified using primers Eh_IscS_681F, 5'-CAA GTG CGA ATA CCC AAT TTG AA-3' and Eh_IscS_2515R, 5'-GGC TGA AGC CAT GAC ACC TC-3' (forward and reverse primers, respectively). The resulting PCR fragments were all cloned into pGEM-T-Easy (Promega) and sequenced to confirm their identity. The new *E. histolytica* IscS sequence has been deposited in Genbank (accession number AY277946).

Phylogenetic analyses

The conceptually translated *E. histolytica* IscS and IscU amino acid sequences were aligned using Clustal W [54] to reference sequences from Genbank. The alignments were manually refined and only unambiguously aligned regions without gaps were used for phylogenetic analysis,

leaving data sets of 28 taxa with 116 amino acid positions (IscU) and a similar taxon set consisting of 29 taxa with 326 amino acid positions (IscS). Likelihood searches were performed in a Bayesian framework under the JTT-f substitution model accommodating site rate variation (fraction of invariable sites plus four variable gamma rates) using the program MrBayes [55]. All analyses started with randomly generated trees and ran for 200,000 generations, with sampling at intervals of 100 generations that produced 2,000 trees. To ensure that the analyses were not trapped on local optima, the data set was run three times independently, each run beginning with a different starting tree. The log-likelihood values of the 2,000 trees in each analysis were plotted against the generation time (not shown). Although the likelihood model stabilized very rapidly, only the last 1,500 trees in each of the three independent analyses were used to estimate separate 50%

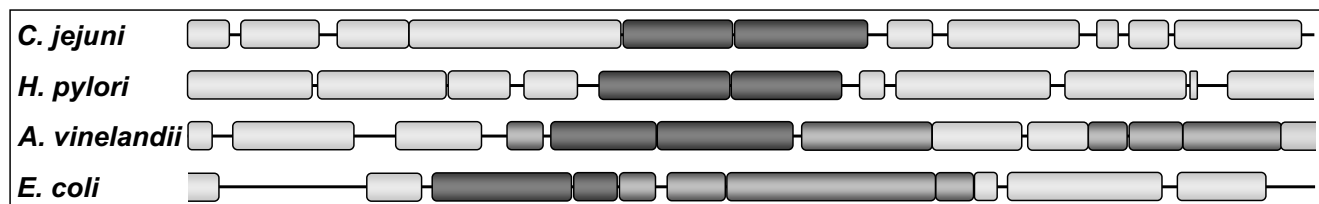


Figure 6
Schematic representation of lsc/Nif operons from different bacteria. Shown is an area of about 10 kb around the lscU/S or NifU/S genes from *C. jejuni*, *H. pylori*, *A. vinelandii*, and *E. coli*. lsc/Nif genes are indicated by a dark grey box and other genes that are part of the lsc-operon are indicated by a lighter shade of grey. Genes that are not part of the lsc/Nif operon are of yet a lighter shade. Boxes are drawn proportionally with regards to length of the ORF.

majority rule consensus trees for these. The frequency of any particular clade, among the individual trees contributing to the consensus tree, represents the posterior probability of that clade [55]. For the maximum likelihood analyses, protein data sets were resampled 100 times using SEQBOOT from PHYLIP [56]. These resampled datasets were analysed using PHYML [57] with alpha and invariant sites parameters optimized on the Bayesian tree in TREE-PUZZLE 5.0 [58] with a mixed four-category discrete gamma plus invariable sites model of rate heterogeneity. The JTT substitution model was used in the protein analyses. Majority rule consensus trees were obtained from the resulting 100 trees using CONSENSE (PHYLIP).

Authors' contributions

MvdG designed and coordinated the molecular genetic studies, carried out the homology modelling, phylogenetic analyses and drafted the manuscript. SC carried out the molecular genetic studies and participated in the sequence alignment. JT edited the manuscript and participated in the design and coordination of the study. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank Drs Salvador Carranza (University of Barcelona, Spain) and David S. Horner (University of Milan, Italy) for advice and help with the phylogenetic analyses and Dr. Hermie Harmsen (University of Groningen, The Netherlands) for compiling a list of gut bacteria currently being sequenced. Preliminary sequence data for *E. histolytica* is deposited regularly into the GSS division of GenBank. The Sequencing effort is part of the International *E. histolytica* Genome Sequencing Project and is supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. SC was supported by an undergraduate research bursary from the Nuffield Foundation (URB/00970/G). This work was supported by a grant from the BBSRC (I11/C13820) to JT.

References

- Stanley SL: **Amoebiasis.** *Lancet* 2003, **361**:1025-1034.
- WHO: *The world health report Geneva: World Health Organization; 1995.*
- Meza I: ***Entamoeba histolytica*: phylogenetic considerations.** *Arch Med Res* 1992, **23**:1-5.
- Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T: **Early branchings in the evolution of eukaryotes: ancient divergence of entamoeba that lacks mitochondria revealed by protein sequence data.** *J Mol Evol* 1993, **36**:380-388.
- Sogin ML: **Early evolution and the origin of eukaryotes.** *Curr Opin Genet Dev* 1991, **1**:457-463.
- Bakatselou C, Kidgell C, Clark CG: **A mitochondrial-type hsp70 gene of *Entamoeba histolytica*.** *Mol Biochem Parasitol* 2000, **110**:177-182.
- Clark CG, Roger AJ: **Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*.** *Proc Natl Acad Sci USA* 1995, **92**:6518-6521.
- Tovar J, Fischer A, Clark CG: **The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*.** *Mol Microbiol* 1999, **32**:1013-1021.
- Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J: **Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*.** *Mol Cell Biol* 1999, **19**:2198-2205.
- Williams BAP, Hirt RP, Lucocq JM, Embley TM: **A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*.** *Nature* 2002, **418**:865-869.
- Riordan CE, Ault J, Langreth SG, Keithly JS: ***Cryptosporidium parvum* Cpn60 targets a relict organelle.** *Curr Genet* 2003, **44**:138-147.
- Tovar J, León-Avila G, Sánchez L, Sutak R, Tachezy J, van der Giezen M, Hernández M, Müller M, Lucocq JM: **Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation.** *Nature* 2003, **426**:172-176.
- Lill R, Kispal G: **Maturation of cellular Fe-S proteins: an essential function of mitochondria.** *Trends Biochem Sci* 2000, **25**:352-356.
- Kato S, Mihara H, Kurihara T, Takahashi Y, Tokumoto U, Yoshimura T, Esaki N: **Cys-328 of lscS and Cys-63 of lscU are the sites of disulfide bridge formation in a covalently bound lscS/lscU complex: implications for the mechanism of iron-sulfur cluster assembly.** *Proc Natl Acad Sci USA* 2002, **99**:5948-5952.
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissbach J, Vivares CP: **Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*.** *Nature* 2001, **414**:450-453.
- LaGier MJ, Tachezy J, Stejskal F, Kutisova K, Keithly JS: **Mitochondrial-type iron-sulfur cluster biosynthesis genes (lscS and lscU) in the apicomplexan *Cryptosporidium parvum*.** *Microbiology* 2003, **149**:3519-3530.
- Mann BJ: ***Entamoeba histolytica* Genome Project: an update.** *Trends Parasitol* 2002, **18**:147-148.
- Bankier AT, Spriggs HF, Fartmann B, Konfortov BA, Madera M, Vogel C, Teichmann SA, Ivens A, Dear PH: **Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*.** *Genome Res* 2003, **13**:1787-1799.
- McArthur AG, Morrison HG, Nixon JE, Passamaneck NQ, Kim U, Hinkle G, Crocker MK, Holder ME, Farr R, Reich CI, Olsen GE, Aley

- SB, Adam RD, Gillin FD, Sogin ML: **The *Giardia* genome project database.** *FEMS Microbiol Lett* 2000, **189**:271-273.
20. Tachezy J, Sánchez LB, Müller M: **Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS.** *Mol Biol Evol* 2001, **18**:1919-1928.
21. Embley TM, van der Giezen M, Horner DS, Dyal PL, Foster P: **Mitochondria and hydrogenosomes are two forms of the same fundamental organelle.** *Phil Trans R Soc Lond* 2003, **358**:191-204.
22. van der Giezen M, Tovar J: **Hydrogenosomes, mitosomes and mitochondria; variations on a theme?** In: *Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics* Edited by: Horner DS, Hirt RP. CRC Press; 2004.
23. Purdy JE, Pho LT, Mann BJ, Petri WA Jr: **Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin.** *Mol Biochem Parasitol* 1996, **78**:91-103.
24. Zheng L, Cash VL, Flint DH, Dean DR: **Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA-fox gene cluster from *Azotobacter vinelandii*.** *J Biol Chem* 1998, **273**:13264-13272.
25. Nakamura Y, Gojibori T, Ikemura T: **Codon usage tabulated from international DNA sequence databases: status for the year 2000.** *Nucl Acids Res* 2000, **28**:292.
26. Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL: **The Pfam Protein Families Database.** *Nucl Acids Res* 2002, **30**:276-280.
27. Servant F, Bru C, Carrere S, Courcelle E, Gouzy J, Peyruc D, Kahn D: **ProDom: automated clustering of homologous domains.** *Brief Bioinform* 2002, **3**:246-251.
28. Zdobnov EM, Apweiler R: **InterProScan – an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847-848.
29. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting protein function and protein-protein interactions from genome sequences.** *Science* 1999, **285**:751-753.
30. Garland SA, Hoff K, Vickery LE, Culotta VC: **Saccharomyces cerevisiae ISU1 and ISU2: members of a well-conserved gene family for iron-sulfur cluster assembly.** *J Mol Biol* 1999, **294**:897-907.
31. Manzetti S: *Structural genomics and molecular visualization – part 2; Basic homology modelling of proteins.* Proinformatic.com 2001.
32. Kaiser JT, Clausen T, Bourenkow GP, Bartunik HD, Steinbacher S, Huber R: **Crystal structure of a NifS-like protein from *Thermotoga maritima*: implications for iron sulphur cluster assembly.** *J Mol Biol* 2000, **297**:451-464.
33. Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization.** *Trends Biochem Sci* 1999, **24**:34-36.
34. Claros MG, Vincens P: **Computational method to predict mitochondrially imported proteins and their targeting sequences.** *Eur J Biochem* 1996, **241**:779-786.
35. Reinhardt A, Hubbard T: **Using neural networks for prediction of the subcellular location of proteins.** *Nucleic Acids Res* 1998, **26**:2230-2236.
36. Van de Peer Y, Ben Ali A, Meyer A: **Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi.** *Gene* 2000, **246**:1-8.
37. Emelyanov VV: **Phylogenetic affinity of a *Giardia lamblia* cysteine desulfurase conforms to canonical pattern of mitochondrial ancestry.** *FEMS Microbiol Lett* 2003, **226**:257-266.
38. Berg RD: **The indigenous gastrointestinal microflora.** *Trends Microbiol* 1996, **4**:430-435.
39. Andersson JO, Roger AJ: **Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions, and prokaryote-to-eukaryote lateral gene transfers.** *Eukaryot Cell* 2002, **1**:304-310.
40. Andersson JO, Sjogren AM, Davis LA, Embley TM, Roger AJ: **Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes.** *Curr Biol* 2003, **13**:94-104.
41. Field J, Rosenthal B, Samuelson J: **Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*.** *Mol Microbiol* 2000, **38**:446-455.
42. Rosenthal B, Mai Z, Caplivski D, Ghosh S, de la Vega H, Graf T, Samuelson J: **Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*.** *J Bacteriol* 1997, **179**:3736-3745.
43. Boucher Y, Doolittle WF: **The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways.** *Mol Microbiol* 2000, **37**:703-716.
44. Sanchez L, Horner D, Moore D, Henze K, Embley T, Müller M: **Fructose-1,6-bisphosphate aldolases in amitochondriate protists constitute a single protein subfamily with eubacterial relationships.** *Gene* 2002, **295**:51.
45. Henze K, Horner DS, Suguri S, Moore DV, Sánchez LB, Müller M, Embley TM: **Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*.** *Gene* 2001, **281**:123-131.
46. de Koning AP, Brinkman FS, Jones SJ, Keeling PJ: **Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*.** *Mol Biol Evol* 2000, **17**:1769-1773.
47. Suguri S, Henze K, Sánchez LB, Moore DV, Müller M: **Archaeobacterial relationships of the phosphoenolpyruvate carboxykinase gene reveal mosaicism of *Giardia intestinalis* core metabolism.** *J Eukaryot Microbiol* 2001, **48**:493-497.
48. Doolittle WF: **You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.** *Trends Genet* 1998, **14**:307-311.
49. Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR: **Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates.** *Nature* 2001, **411**:940-944.
50. Clark CG, Diamond LS: **Methods for cultivation of luminal parasitic protists of clinical importance.** *Clin Microbiol Rev* 2002, **15**:329-341.
51. Clark CG: **DNA purification from polysaccharide-rich cells.** In: *Protocols in Protozoology Volume 1.* Edited by: Lee JJ, Soldo AT. Lawrence, Kansas: Allen Press; 1992:D3.1-D3.2.
52. Sambrook J, Fritsch E, Maniatis T: *Molecular cloning, a laboratory manual* New York, USA: Cold Spring Harbor Laboratory Press; 1989.
53. Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers.** In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* Edited by: Krawetz S, Misener S. Totowa, NJ: Humana Press; 2000:365-386.
54. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22**:4673-4680.
55. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
56. Felsenstein J: *PHYMLIP (Phylogeny Inference Package)* Felsenstein, Department of Genetics, University of Washington, Seattle; 1993.
57. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
58. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
59. Schwartz RM, Dayhoff MO: **Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts.** *Science* 1978, **199**:395-403.
60. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.** *Electrophoresis* 1997, **18**:2714-2723.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

