

RESEARCH ARTICLE

Open Access



# Heme pathway evolution in kinetoplastid protists

Ugo Cenci<sup>1,2</sup>, Daniel Moog<sup>1,2</sup>, Bruce A. Curtis<sup>1,2</sup>, Goro Tanifuji<sup>3</sup>, Laura Eme<sup>1,2</sup>, Julius Lukeš<sup>4,5</sup> and John M. Archibald<sup>1,2,5\*</sup>

## Abstract

**Background:** Kinetoplastea is a diverse protist lineage composed of several of the most successful parasites on Earth, organisms whose metabolisms have coevolved with those of the organisms they infect. Parasitic kinetoplastids have emerged from free-living, non-pathogenic ancestors on multiple occasions during the evolutionary history of the group. Interestingly, in both parasitic and free-living kinetoplastids, the heme pathway—a core metabolic pathway in a wide range of organisms—is incomplete or entirely absent. Indeed, Kinetoplastea investigated thus far seem to bypass the need for heme biosynthesis by acquiring heme or intermediate metabolites directly from their environment.

**Results:** Here we report the existence of a near-complete heme biosynthetic pathway in *Perkinsela* spp., kinetoplastids that live as obligate endosymbionts inside amoebozoans belonging to the genus *Paramoeba/Neoparamoeba*. We also use phylogenetic analysis to infer the evolution of the heme pathway in Kinetoplastea.

**Conclusion:** We show that *Perkinsela* spp. is a deep-branching kinetoplastid lineage, and that lateral gene transfer has played a role in the evolution of heme biosynthesis in *Perkinsela* spp. and other Kinetoplastea. We also discuss the significance of the presence of seven of eight heme pathway genes in the *Perkinsela* genome as it relates to its endosymbiotic relationship with *Paramoeba*.

**Keywords:** Heme, Kinetoplastea, *Paramoeba pemaquidensis*, *Perkinsela*, Evolution, Endosymbiosis, Prokinetoplastina, Lateral gene transfer

## Background

Kinetoplastea is a diverse group of unicellular flagellated organisms, most of which are parasites. The best known group of kinetoplastid parasites is the Trypanosomatida, which parasitize plants (e.g., *Phytomonas* [1]), insects (e.g., *Angomonas* [2]) and humans (e.g., *Leishmania* [3] and *Trypanosoma* [4]). However, the Kinetoplastea also includes non-parasitic organisms such as free-living bodonids like *Bodo saltans* [5] and *Neobodo designis* [6]. The bodonids are comprised of Neobodonida, Eubodonida and Parabodonida, which are considered early branching Kinetoplastea [7–10] and serve as an important reference point for the evolution of parasitism within this species-rich group [11]. However, these organisms

are poorly understood and the evolutionary relationship amongst bodonids is still debated [8, 10, 12]. The Prokinetoplastina is an even deeper branching group of kinetoplastid flagellates [7–9], and is composed of organisms such as *Ichthyobodo necator*, a fish ectoparasite, and *Perkinsela* sp. [13], which is an endosymbiont of opportunistic pathogenic amoebae belonging to *Neoparamoeba/Paramoeba* spp. [14–18]. The Kinetoplastea themselves belong to the Excavata, more specifically the Euglenozoa, which includes Diplonemida (e.g., *Diplonema papillatum*) and Euglenida such as the plastid-bearing organisms *Eutreptiella gymnastica* and *Euglena gracilis*.

Parasitic Kinetoplastea have complex life cycles and have undergone extensive reductive evolution as a consequence of their parasite-host interactions. One commonly observed change is the reduction or complete loss of biochemical pathways that can be augmented or provided by their hosts [11]. This includes the lack of tetrahydrobiopterin biosynthesis required for folate and pteridine [19, 20] and, in trypanosomatids, purine auxotrophy [21]. One particularly

\* Correspondence: john.archibald@dal.ca

<sup>1</sup>Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada

<sup>2</sup>Centre for Comparative Genomics and Evolutionary Bioinformatics, Halifax, Nova Scotia, Canada

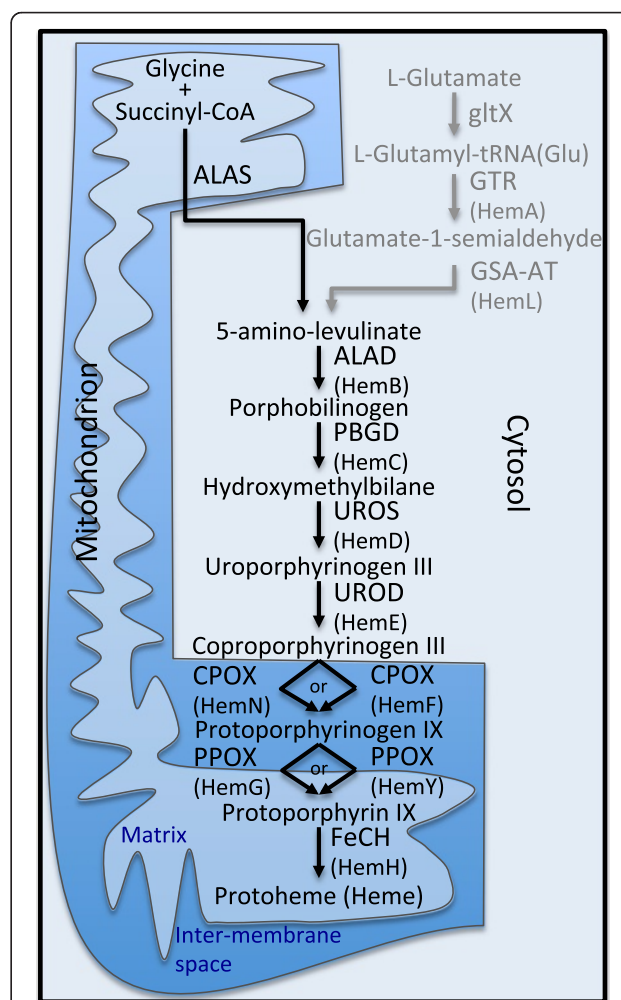
Full list of author information is available at the end of the article



striking example of such metabolic reduction in Kinetoplastea is the heme pathway, which is either incomplete or missing entirely (in some parasitic species). In the latter case, essential metabolites are acquired from their hosts [22, 23] or, in the case of *Strigomonas culicis* and *Angomonas deanei*, from bacterial endosymbionts [24]. Metabolite import could involve intermediates in the heme pathway from coproporphyrinogen III, as suggested by Kořený et al. [25], or heme itself [26, 27]. Furthermore, the plant pathogen *Phytomonas serpens* seems not to require heme at all [28]. The heme pathway is not found in *Trypanosoma* and only the last three steps of the pathway are present in other trypanosomatids such as Leishmaniinae (composed of *Leptomonas* spp., *Crithidia* spp. and *Leishmania* spp.) [29] and Strigomonadinae (composed of *Angomonas* spp. and *Strigomonas* spp.). No complete heme pathway has been described for a member of the Kinetoplastea [25].

The heme pathway is an important part of cellular metabolism. It produces the cofactor heme, which is required for key biochemical processes such as oxidative phosphorylation. In most eukaryotes, heme biosynthesis involves eight steps (Fig. 1), the first being the transformation of glycine or L-glutamate into 5-amino-levulinate by 5-aminolevulinic acid synthase (ALAS) [30]. An alternate is the synthesis of 5-amino-levulinate by the sequential action of glutamyl-tRNA synthetase (GltX), glutamyl-tRNA reductase (GTR), and glutamate-1-semialdehyde 2,1-aminomutase (GSA-AT). This second pathway is found in most bacteria and in most eukaryotes with a plastid [31, 32]. 5-amino-levulinate is then converted into porphobilinogen by porphobilinogen synthase (also known as delta-aminolevulinic acid dehydratase (ALAD)), and then into hydroxymethylbilane by hydroxymethylbilane synthase (or porphobilinogen deaminase (PBGD)). Subsequently, uroporphyrinogen-III synthase (UROS) produces uroporphyrinogen III, which is then converted into coproporphyrinogen III by uroporphyrinogen decarboxylase (UROD) [33, 34].

In the next step of the heme pathway, coproporphyrinogen III is modified to protoporphyrin IX by coproporphyrinogen III oxidase (CPOX/HemF) under aerobic conditions, and under anaerobic conditions, by oxygen-independent coproporphyrinogen III oxidase (CPOX/HemN) [35]. Subsequently, protoporphyrin IX is transformed into protoporphyrin IX by oxygen-dependent protoporphyrinogen oxidase (PPOX/HemY), or menaquinone-dependent protoporphyrinogen oxidase (PPOX/HemG), the latter enzyme possessing the ability to perform the reaction under aerobic and anaerobic conditions [36–38]. Finally, protoporphyrin IX is converted to protoheme by the action of ferrochelatase (FeCH) [39]. Interestingly, the heme biosynthetic pathway in eukaryotes involves proteins in three different



**Fig. 1** Heme pathway in eukaryotic cells. The Heme pathway in Opisthokonta, and most likely other heterotrophic eukaryotes, as described in Kořený et al. [25], is represented in black. An alternate entry to the pathway, present in bacteria and in the plastids of algae, is represented in grey. For each protein name the corresponding protein in bacteria is indicated in parenthesis. The heme pathway in eukaryotes takes place in the cytosol for the steps involving ALAD, PBGD, UROS, and UROD, while ALAS and FeCH act in the mitochondrial matrix, and the PPOX and CPOX act in the inter-membrane space. Abbreviations: ALAS: 5-aminolevulinic acid synthase, GltX: glutamyl-tRNA synthetase, GTR: glutamyl-tRNA reductase, GSA-AT: glutamate-1-semialdehyde 2,1-aminomutase, ALAD: delta-aminolevulinic acid dehydratase, PBGD: porphobilinogen deaminase, UROS: uroporphyrinogen-III synthase, UROD: uroporphyrinogen decarboxylase, CPOX/HemF: coproporphyrinogen III oxidase, CPOX/HemN: oxygen-independent coproporphyrinogen III oxidase, PPOX/HemY: oxygen-dependent protoporphyrinogen oxidase, PPOX/HemG: menaquinone-dependent protoporphyrinogen oxidase, FeCH: ferrochelatase

cellular locations. While the ALAD, PBGD, UROS, and UROD enzymes act in the cytosol, the ALAS and FeCH enzymes are usually localized to the mitochondrial matrix, and PPOX and CPOX function in the inter-membrane space of the mitochondrion [25, 40] (Fig. 1).

In this study, we analyzed the heme pathway of two species of *Perkinsella*, which are endosymbionts of the amoebozoans *Paramoeba pemaquidensis* and *P. invadens*. Using molecular phylogenetics, we show that *Perkinsella* belongs to the earliest branching kinetoplastid lineage currently known, and demonstrate that these highly reduced endosymbionts nevertheless possess a near-complete heme biosynthesis pathway, the first of its kind described for a member of the Kinetoplastea. We hypothesize that at least a subset of these enzymes represents elements of the ancestral heme pathway in the group. Finally, we discuss the importance of this pathway in early-branching kinetoplastid flagellates for understanding the adaptations that have occurred during the evolution of bodonids and parasitic trypanosomatids.

## Methods

### Heme pathway protein sequence acquisition

We searched for genes encoding heme biosynthetic enzymes in the nuclear genome of the *Perkinsella* endosymbiont living within the amoebozoan *Paramoeba pemaquidensis* CCAP1560/4. The nuclear genome (and associated transcriptome) of the host *P. pemaquidensis* was also queried, as was transcriptomic data from another species, *P. invadens*. GenBank accession numbers for the *Perkinsella* spp. and *Paramoeba* spp. data used in this study are LFNC00000000 and KU609011-KU609043. BLASTp/tBLASTn searches were carried out using a curated set of heme pathway enzymes from diverse eukaryotes as queries with an E-value cut-off 1e-05. For identification of UROS enzymes, local HMMER [41] searches (*hmmsearch*) were initially performed against the total *P. invadens* transcriptome database (6-frame translation into protein) using default settings. Profile HMMs were constructed via *hmmbuild* with Stockholm alignments ('Seed' and 'NCBI') for HEM4 (PF02602) retrieved from the Pfam website (<http://pfam.xfam.org>). Hits with an E-value  $\leq 1e-05$  were then used as queries to screen the *P. pemaquidensis* total genomic and transcriptomic assemblies via local tBLASTn. Homologs were also identified using Ghostkoala (<http://www.kegg.jp/ghostkoala/>). Sequences used in phylogenetic analyses were obtained by BLAST from the NCBI nr database, the MMETSP database of transcriptomes [42], from TritypDB [43], and from the *B. saltans* genome (Wellcome Trust Sanger Institute). To further verify their predicted roles in heme biosynthesis, all sequences analyzed in this study were annotated using InterProScan [44] and the InterPro classification [45].

### Protein localization predictions

Sequences of proteins putatively involved in heme biosynthesis were subjected to a localization prediction pipeline using the following tools: SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP-3.0/>) [46], TargetP 1.1

(<http://www.cbs.dtu.dk/services/TargetP/>) [47], PredSL (<http://aias.biol.uoa.gr/PredSL/input.html>) [48] and Predotar (<https://urgi.versailles.inra.fr/predotar/predotar.html>) [49] with standard settings for prediction of N-terminal targeting signals such as secretory signal peptides (SPs) and mitochondrial targeting peptides (mTPs). TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) [50] was used for analysis of potential transmembrane domains (TMDs). Euglenophytes harbor complex three membrane-bound plastids of green algal origin and the proteins targeted to these organelles usually possess bipartite targeting signals (BTS) consisting of a SP followed by a transit peptide (TP) [51, 52]. Plastid targeting of proteins in photosynthetic euglenids was thus predicted using TargetP, PredSL and Predotar (see above) in 'plant/chloroplast' mode after removal of the signal peptide (SP) (based on SignalP prediction). For classification of a protein into one of four categories (secretory, mitochondrial, plastidial, 'other'), the output of at least two of the searches had to be positive for a specific category (see Additional file 1: Table S1). Only those protein sequences starting with a methionine residue were classified.

### Phylogenetic analysis

Eleven proteins were selected for their potential to resolve the evolution of Kinetoplastea in general and the taxonomic position of *Perkinsella* spp. in particular (Additional file 2: Table S2; proteins used for *Perkinsella* sp. from *P. pemaquidensis* were: KNH09580, KNH04116, KNH06333, KNH09360, KNH08922, KNH06227, KNH03620, KNH06559, KNH08032, KNH06818, KNH05478). Phylogenetic trees were first constructed individually for each protein. Homologs were aligned using MUSCLE [53] and blocks were selected using BMGE [54] with the BLOSUM40 similarity matrix and a block size of four. Each individual protein tree was obtained with IQ-TREE using the ultrafast bootstrap method under the LG4X model and was checked manually before concatenation. Sequences for each organism were then concatenated (30 taxa, 5,060 sites), and a single phylogenetic tree was built using Phylobayes version 4.1 [55] under the catfix C20 + Poisson model [56]. The two chains were stopped when convergence was reached ( $\text{maxdiff} < 0.1$ ) after 230 cycles and a burn-in of 300 trees. We then mapped bootstrap values obtained from 1,000 replicates under the LG4X [57] model with IQ-TREE software [58]. Trees were visualized using Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>). Topological tree tests were performed using RAXML version 8.0.19 [59] under the PROTGAMMALG4X model. The different topologies were then compared according to the tree topology test available in IQ-TREE (RELL approximation [60], the Kishino-Hasegawa test [61], the Shimodaira-Hasegawa test [62], and expected likelihood weights [63]). The percentage of missing data

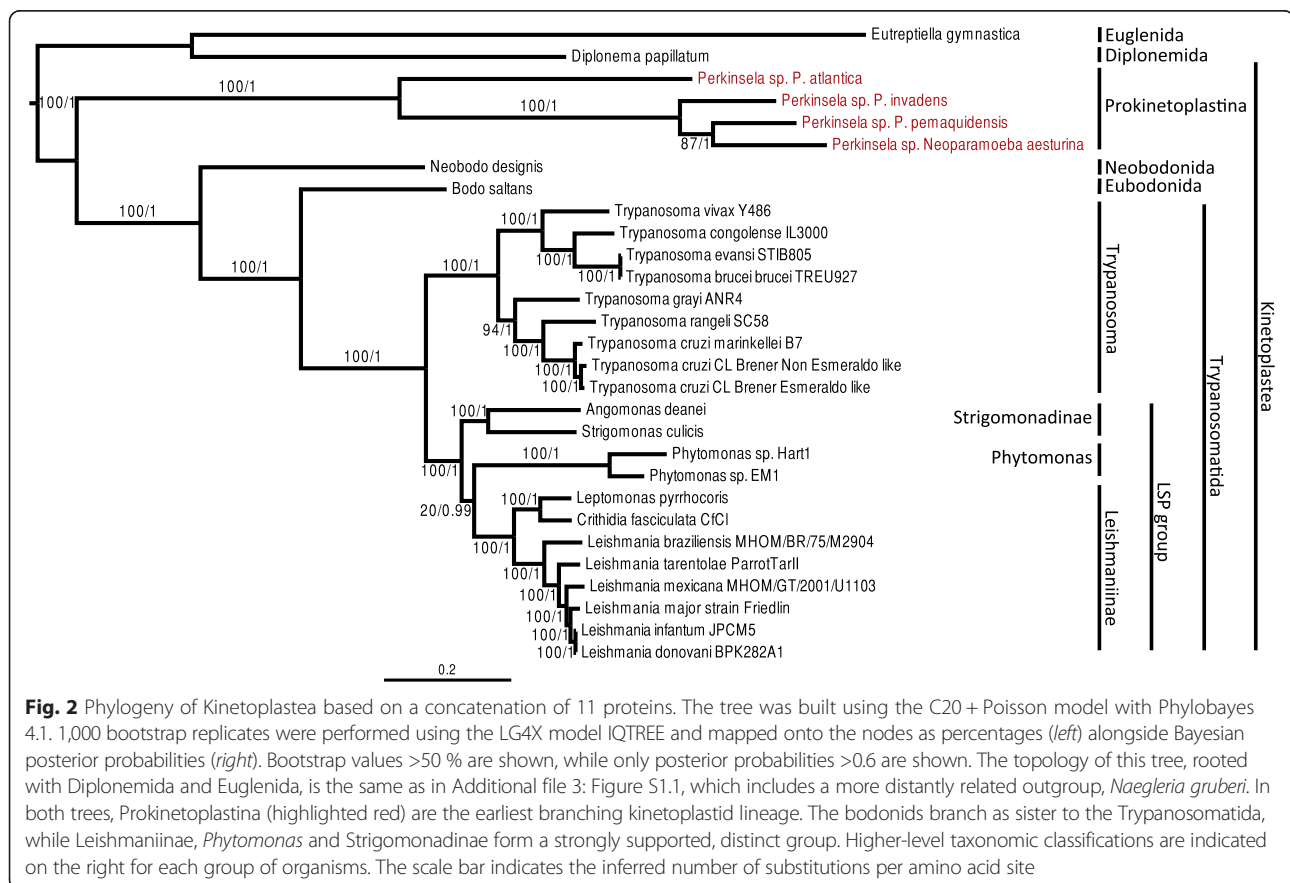
for each organism in the concatenated alignment is provided in Additional file 3: Table S1.4.

We also built phylogenetic trees for enzymes involved in the heme pathway in Kinetoplastea. Sequences were retrieved using homology searches by BLAST against sequences obtained from different sources (see above). All sequences with an E-value less than  $1e^{-5}$  were selected. We then aligned these sequences using MAFFT with the fast alignment settings [64]. Block selection was then performed using BMGE with a block size of 4 and the BLOSUM30 similarity matrix. Preliminary trees were generated using Fasttree [65] and ‘dereplication’ was applied to robustly supported monophyletic clades using TreeTrimmer [66] in order to reduce sequence redundancy. For each protein, the final set of sequences was selected manually. Proteins were re-aligned with MUSCLE, block selection was carried out using BMGE with a block size of four and the matrix BLOSUM30, and trees were generated using Phylobayes-4.1 under the catfix C20 + Poisson model with the two chains stopped when convergence was reached ( $\text{maxdiff} < 0.1$ ) after at least 200 cycles, discarding 1,000 burn-in trees. Bootstrap support values were estimated from 100 replicates using IQ-TREE under the LG4X model and mapped onto the Bayesian tree.

## Results

### *Perkinsela*: an early branching kinetoplastid lineage

Together with Euglenida and Diplonemida, Kinetoplastea belong to the Euglenozoa. We built a phylogeny of eleven concatenated proteins (Fig. 2 and Additional file 3: Figure S1.1), sampled from the Prokinetoplastina containing the *Perkinsela* spp. group, with representatives from trypanosomatids and bodonids, and rooted with the diplomemid *Diplonema papillatum*, the euglenid *Eutreptiella gymnastica* and the heterolobosean *Naegleria gruberi*. The eleven proteins were carefully selected based on their availability in public databases in the lineages of interest (i.e., *Eutreptiella gymnastica*, *Diplonema papillatum*, *Perkinsela* spp., *Neobodo designis* and *Bodo saltans*) so as to minimize missing data in our supermatrix. Our results confirm that the diversity of *Perkinsela* spp. for which genomic and/or transcriptomic sequence data are currently available represent a monophyletic assemblage, as previously described [16]. We tested the effects of a distantly related outgroup by removing *N. gruberi* to see if the Prokinetoplastina was still positioned as the deepest branch of the Kinetoplastea (Fig. 2). We then ran topology tests to assess the early branching nature of *Perkinsela* spp. within Kinetoplastea, and to determine if alternative topologies to the Trypanosomatida clade formed by *Trypanosoma*, *Phytomonas*, the



Strigomonadinae and Leishmaniinae were rejected (Additional file 3: Table S1.2). Our results suggest that the *Perkinsela* spp. group represents a monophyletic deep branching clade, positioned between the diplomonids and bodonids. While there is some uncertainty with the position of Prokinetoplastina relative to Diplonema, the Kishino-Hasegawa test rejected the possibility that the former branches deeper in the tree than the latter (Additional file 3: Table S1.2). Moreover, the group formed by *Phytomonas*, the Strigomonadinae and Leishmaniinae was found to be robust. However, the relative branching of Strigomonadinae, *Phytomonas* and Leishmaniinae is not clear.

#### A near-complete set of heme pathway enzymes in *Perkinsela* spp.

Unlike other kinetoplastid flagellates, and despite having a reduced genome due to its endosymbiotic lifestyle (Tanifuji et al., unpublished data), with the exception of one protein (UROS, see below), we found a full set of heme biosynthesis enzymes encoded in the nuclear genome of the *Perkinsela* sp. living within the amoeba *Paramoeba pemaquidensis*. We also analyzed genes for heme pathway enzymes in the transcriptomes of *Paramoeba atlantica*, *P. invadens*, and *Neoparamoeba aestuarina*. For each of these species, the sequence data are a mixture of host- and endosymbiont-derived transcripts, due to the fact that it is not possible to separate the *Perkinsela* endosymbionts from their amoeba hosts. Despite this complication, our results are consistent with the existence of an almost complete heme pathway in the *Perkinsela* sp. endosymbionts within *P. invadens* and *N. aestuarina*. In the case of the *Perkinsela* sp. of *P. atlantica*, genes for only two enzymes were identified, perhaps due to the limited number of endosymbiont-derived transcripts in the data.

Beyond *Perkinsela* spp., we inferred the presence/absence of heme biosynthesis enzymes in all the organisms present in the phylogenetic tree shown in Fig. 2, as well as for the amoebozoan hosts in which *Perkinsela* spp. reside (Table 1 and Additional file 4: Table S3). Our results are consistent with those of others [24, 25, 28] in showing the complete absence of the heme pathway in members of the genus *Trypanosoma*. The Strigomonadinae/Leishmaniinae group was found to possess only the last three steps of heme biosynthesis, while *Phytomonas* spp. appear to have only the ferrochelatase (FeCH) enzyme. However, we also found a potential uroporphyrinogen-III synthase enzyme in Leishmaniinae, which has not previously been discussed. In addition, we identified an almost complete heme pathway in the euglenid *E. gymnastica*, and, as expected for a plastid-bearing organism, the alga-associated GTR and GSA-AT enzymes [67]. We found no evidence for the existence of a heme biosynthesis pathway

in the diplomonid *Diplonema papillatum*, noting that at present only a small amount of sequence data is publicly available.

#### Subcellular localization of the heme pathway in *Perkinsela* spp.

As shown in Table 1, several *Perkinsela* spp. appear to lack only one enzyme of the heme biosynthetic pathway (UROS), making it the most complete set identified for a kinetoplastid thus far. Other kinetoplastids have only a partial pathway or have lost the capacity to synthesize heme entirely; presumably they obtain heme from their host or do not require it. Perhaps due to its incomplete nature (when present), the subcellular localization of the heme pathway in other Kinetoplastea is different from the classical heme pathway in other eukaryotes [25]. We predicted the localization of the putative heme synthesis enzymes in *Perkinsela* sp. and compared these data to what is known from other kinetoplastids. We also predicted the localization of the heme pathway in the host amoebae, and compared them to other amoebozoans (Additional file 1: Table S1). As expected, FeCH seems to be targeted to the mitochondrion of *Perkinsela* spp., as in other heterotrophic organisms and in the FeCH-containing trypanosomatids. In addition, the heme pathway in *Perkinsela* spp., as well as in their *Neoparamoeba/Paramoeba* hosts, is predicted to produce 5-amino-levulinate in the mitochondrion. However, CPOX/HemF and PPOX/HemY enzymes were not predicted to be targeted to the mitochondrion.

#### Complex phylogenetic patterns for heme biosynthesis enzymes in *Perkinsela* spp.

Given the patchy distribution of heme pathway enzymes in Kinetoplastea, we performed phylogenetic analyses for the complete set of *Perkinsela* spp. enzymes predicted to be involved in this pathway in an attempt to infer their evolutionary history. For two enzymes, UROD (Fig. 3) and ALAS (Fig. 4), the *Perkinsela* spp. homologs form a monophyletic clade with their counterparts in organisms to which they are known to be related, i.e., *Euglena* and/or *Eutreptiella*. Statistical support for this monophyletic relationship is reasonably strong in the case of UROD (bootstrap (BS) and posterior probabilities (PP) of 70 % and 0.99, respectively), but weak for ALAS. In both cases, homologues can be found in virtually all major eukaryotic groups and they seem to form a monophyletic clade (although the backbone of the trees is poorly resolved). This suggests that UROD and ALAS were likely present in the eukaryotic common ancestor and have been vertically inherited in *Perkinsela* spp. In addition, the eukaryotic ALAS homologues branch within alpha-proteobacteria, strongly suggesting a mitochondrial origin.

**Table 1** Presence/absence of enzymes involved in the biosynthesis of heme

		ALAS	gitX	GTR (HemA)	GSA-AT (HemL)	ALAD (HemB)	PBGD (HemC)	UROS (HemD)	UROD (HemE)	CPOX (HemF)	CPOX (HemN)	PPOX (HemY)	PPOX (HemG)	FeCH (HemH)
Naegleria	<i>Naegleria gruberi</i>		+		+						+			+
Euglenida	<i>Eutreptiella gymnastica</i>	+	+	+	+	+	+	+	+	+	+	+		+
Diplonemida	<i>Diplonema papillatum</i> *													
Prokinetoplastina	putative <i>Perkinsela</i> sp. from <i>Paramoeba atlantica</i>		+											
	putative <i>Perkinsela</i> sp. from <i>Neoparamoeba aestuarina</i>	+	+			+	+		+	+		+		+
	<i>Perkinsela</i> sp. from <i>P. pemaquidensis</i> CCAP1560	+	+			+	+		+	+		+		+
	<i>Perkinsela</i> sp. from <i>P. invadens</i>	+	+			+	+		+	+		+		+
Neobodonida	<i>Neobodo designis</i>		+											
Eubodonida	<i>Bodo saltans</i>		+											
Trypanosoma	<i>Trypanosoma brucei</i> TREU927		+											
	<i>Trypanosoma congolense</i> IL3000		+											
	<i>Trypanosoma cruzi</i> CL Brener Esmeraldo-like		+											
	<i>Trypanosoma cruzi</i> CL Brener Non-Esmeraldo-like		+											
	<i>Trypanosoma cruzi marinkellei</i> B7		+											
	<i>Trypanosoma evansi</i> STIB805		+											
	<i>Trypanosoma grayi</i> ANR4		+											
	<i>Trypanosoma rangeli</i> SC58		+											
	<i>Trypanosoma vivax</i> Y486		+											
Phytomonas	<i>Phytomonas</i> Hart1		+											+
	<i>Phytomonas</i> EM1		+											+
Strigomonadinae	<i>Strigomonas culicis</i>		+							+			+	+
	<i>Angomonas deanei</i>		+							+			+	+
Leishmaniinae	<i>Leptomonas pyrrocoris</i>		+							+			+	+
	<i>Crithidia fasciculata</i> CfCl		+					+		+			+	+
	<i>Leishmania braziliensis</i> MHOMBR75M2904		+					+		+			+	+
	<i>Leishmania donovani</i> BPK282A1		+					+		+			+	+
	<i>Leishmania infantum</i> JPCM5		+					+		+			+	+
	<i>Leishmania major</i> Friedlin		+					+		+			+	+
	<i>Leishmania mexicana</i> MHOMGT2001U1103		+					+		+			+	+
	<i>Leishmania tarentolae</i> ParrotTarII		+					+		+			+	+
Amoebozoa	<i>Paramoeba pemaquidensis</i> (host)	+	+			+		+	+		+	+		
	<i>Neoparamoeba aesturina</i> (host)	+	+			+	+	+	+	+	+			+
	<i>Paramoeba atlantica</i> (host)	+	+			+		+	+	+	+	+		

A '+' indicates the presence of this enzyme. A '\*' indicates uncertainty associated with incomplete genomic or transcriptomic data. Blanks indicate protein absence or not detected

In phylogenies of ALAD (Fig. 5), PBGD (Additional file 5: Figure S2.1), and PPOX/HemY (Fig. 6), *Perkinsela* spp. proteins generally branched amongst eukaryotic sequences, with bacterial sequences either being more distantly related or patchily distributed (see, e.g., Chlamydia homologs in the PPOX/HemY tree (Fig. 6)). In addition, PPOX/HemY and ALAD homologs found in *Perkinsela* spp. form a clade near the base of eukaryotes. In the case of CPOX/HemF (Fig. 7), while the *Perkinsela* spp. sequences are close to eukaryotes, the tree is complicated by the fact that the main eukaryotic clade contains sequences from Bacteroidetes. It is thus not possible to make definitive statements about the evolutionary origin of this gene in *Perkinsela* spp., although it seems likely that it represents the canonical eukaryotic CPOX. Our preliminary phylogenetic analyses of FeCH were suggestive of several lateral gene transfer (LGT) events in the history of this enzyme in Kinetoplastea (Additional file 5: Figures S2.2 and S2.4). However, the extremely large number of bacterial homologs available for this enzyme made comprehensive analyses difficult, and we suspected that the divergent nature of certain clades within the global FeCH tree were introducing long branch attraction artifacts. As described below, we thus systematically analyzed sub-sections of the FeCH tree in order to better understand the evolutionary history of this enzyme in the kinetoplastids.

#### Lateral gene transfer and the evolution of heme biosynthetic enzymes in Kinetoplastea

Kořený et al. proposed that heme enzymes in trypanosomatids have been acquired by LGT from various sources [25]. Indeed, the trypanosomatids are thought to have lost the heme pathway early in the evolution of the group [22], and while the Leishmaniinae/Strigomonadinae probably acquired genes for CPOX/HemF, PPOX/HemG and FeCH by LGT, the bodonids appear to have acquired FeCH separately, as have members of the genus *Phytomonas* [25].

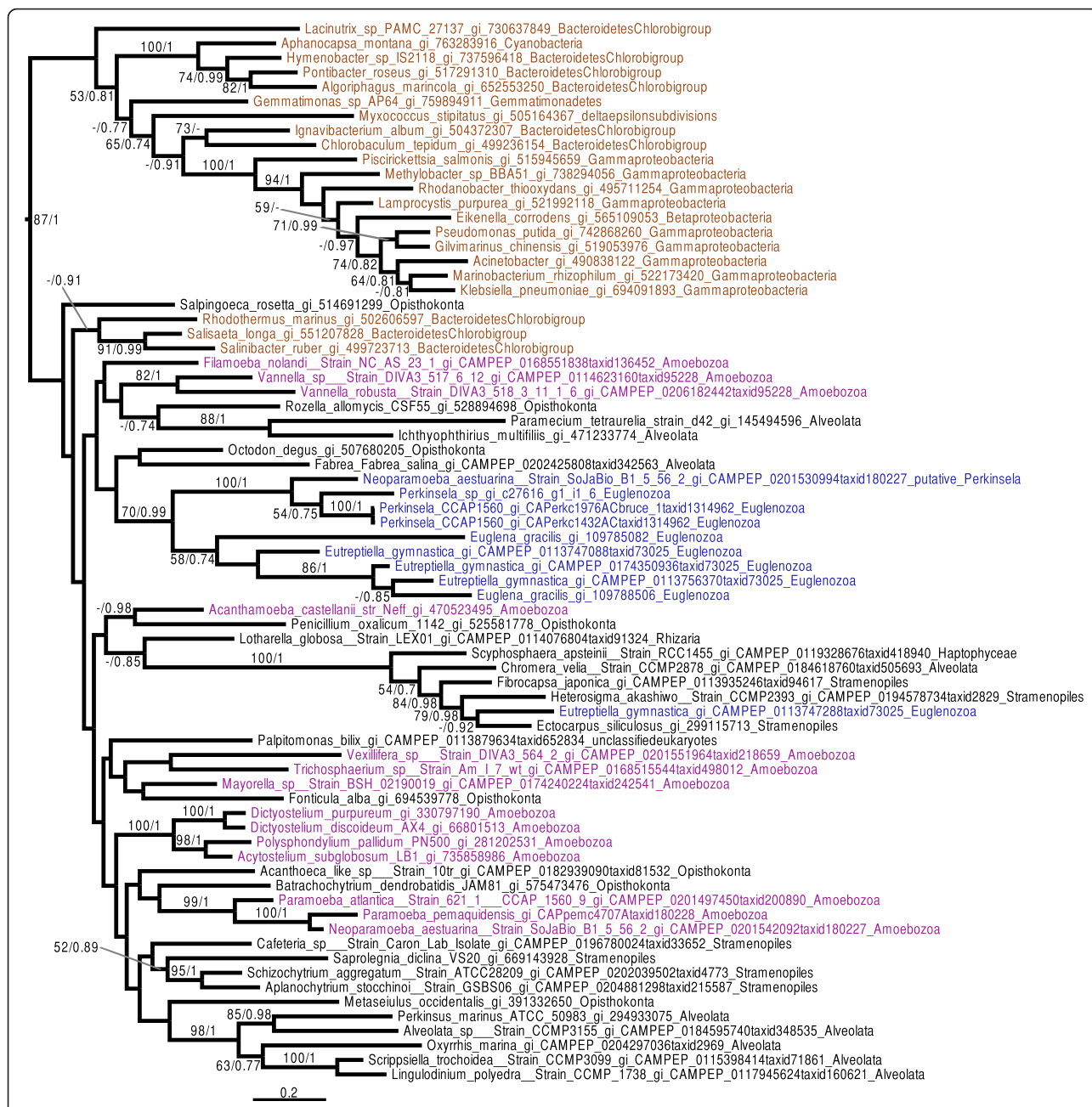
Our results provide further support for the likelihood of LGT events from Gammaproteobacteria to an ancestor of Leishmaniinae and Strigomonadinae for CPOX/HemF (Fig. 7) and PPOX/HemG (Additional file 5: Figure S2.3). In addition, the FeCH phylogeny for these organisms suggests that Leishmaniinae and Strigomonadinae acquired the gene by LGT either from Gammaproteobacteria or from Firmicutes (Additional file 5: Figure S2.5). In the case of *Parabodo caudatus*, the FeCH homolog appears to be derived by LGT from Gammaproteobacteria (Additional file 5: Figure S2.6). The phylogeny of FeCH homologs in *Phytomonas* species is complicated by their divergent nature, but they nevertheless seem to have a different origin from their counterparts in the other Kinetoplastea discussed above (Additional file 5: Figure S2.7). However, FeCH enzymes in *Phytomonas* spp. appear to share recent

common ancestry with FeCH homologs found in *Leptomonas* species.

#### UROS and FeCH enzymes in Kinetoplastea: ancestral or recent acquisitions from bacteria?

While we were unable to detect any UROS related enzyme genes in the genome of the *Perkinsela* sp. within *P. pemaquidensis*, we found two different UROS gene candidates in *Paramoeba* spp., which, due to the presence of several introns, have both been assigned to be of host origin. The first UROS enzyme (KU609032) – the one included in the phylogenetic tree (Fig. 8) – shows a high degree of primary sequence conservation, and possesses homologs in all *Paramoeba* spp. included in this study (Table 1 and Additional file 4: Table. S3). The amino acid sequence of the other UROS candidate (KU609033), which was detected via HMMER search, is highly divergent; no phylogenetic trees could be constructed with this sequence. It is thus unclear whether this second host UROS-like enzyme functions in the heme pathway at all. In addition, we detected a putative UROS enzyme in Leishmaniinae, and carried out the first investigation of its evolutionary history. In our phylogenetic analysis, the Leishmaniinae and other eukaryotes, including the *Paramoeba* spp. group (i.e., the hosts of *Perkinsela* spp.), branch somewhat close to one another (Fig. 8), with the sequences from *Paramoeba* spp. showing affiliations to cyanobacterial homologs. *Leishmania* shows affiliations with a restricted group of Bacteroidetes. In an attempt to better understand the evolution of the *Leishmania* sequences, we removed the divergent bacterial group and reconstructed the tree. The resulting phylogeny (Additional file 5: Figure S2.8) still does not reveal any obvious link between Leishmaniinae and other eukaryotes. However, since the other eukaryotic sequences (*Paramoeba* spp. and Viridiplantae) found in this part of the UROS tree are separated by different bacteria, it seems reasonable to assume that the putative UROS gene of Leishmaniinae was acquired by LGT from bacteria.

As noted above, the FeCH enzyme of Prokinetoplastina is of ambiguous origin. In phylogenetic trees *Perkinsela* spp. homologs do not branch with those of Euglenozoa or Eukaryota (Additional file 5: Figures S2.9 and S2.10), although *Perkinsela* spp. and *Euglena gracilis* homologs are both nested within a restricted Bacteroidetes clade that could suggest lateral acquisition from this bacterial group. Strictly speaking, and despite the apparent rarity of LGT eukaryotes-to-prokaryotes [68], transfer from an early euglenozoan to Bacteroidetes cannot be excluded, and it is formally possible that the FeCH found in both *Euglena gracilis* and Prokinetoplastina represents the ancestral state for Euglenozoa.



**Fig. 3** UROD phylogenetic tree. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50 % are shown, while only posterior probabilities >0.6 are shown. The tree is rooted with the proteobacterial sequences, as in Kořený and Oborník [67]. Sequences are colored according to their taxonomic affiliation: Amoebozoa are in purple, Euglenozoa are in blue, other Eukaryota are in black, and Bacteria are brown. A trio of sequences from the euglenozoan *Eutreptiella gymnastica*, as well as two *Euglena gracilis* sequences, group with Prokinetoplastina with a bootstrap value of 70 % and a posterior probability of 0.99. Another *E. gymnastica* sequence branches elsewhere in the tree. The scale bar shows the inferred number of amino acid substitutions per site

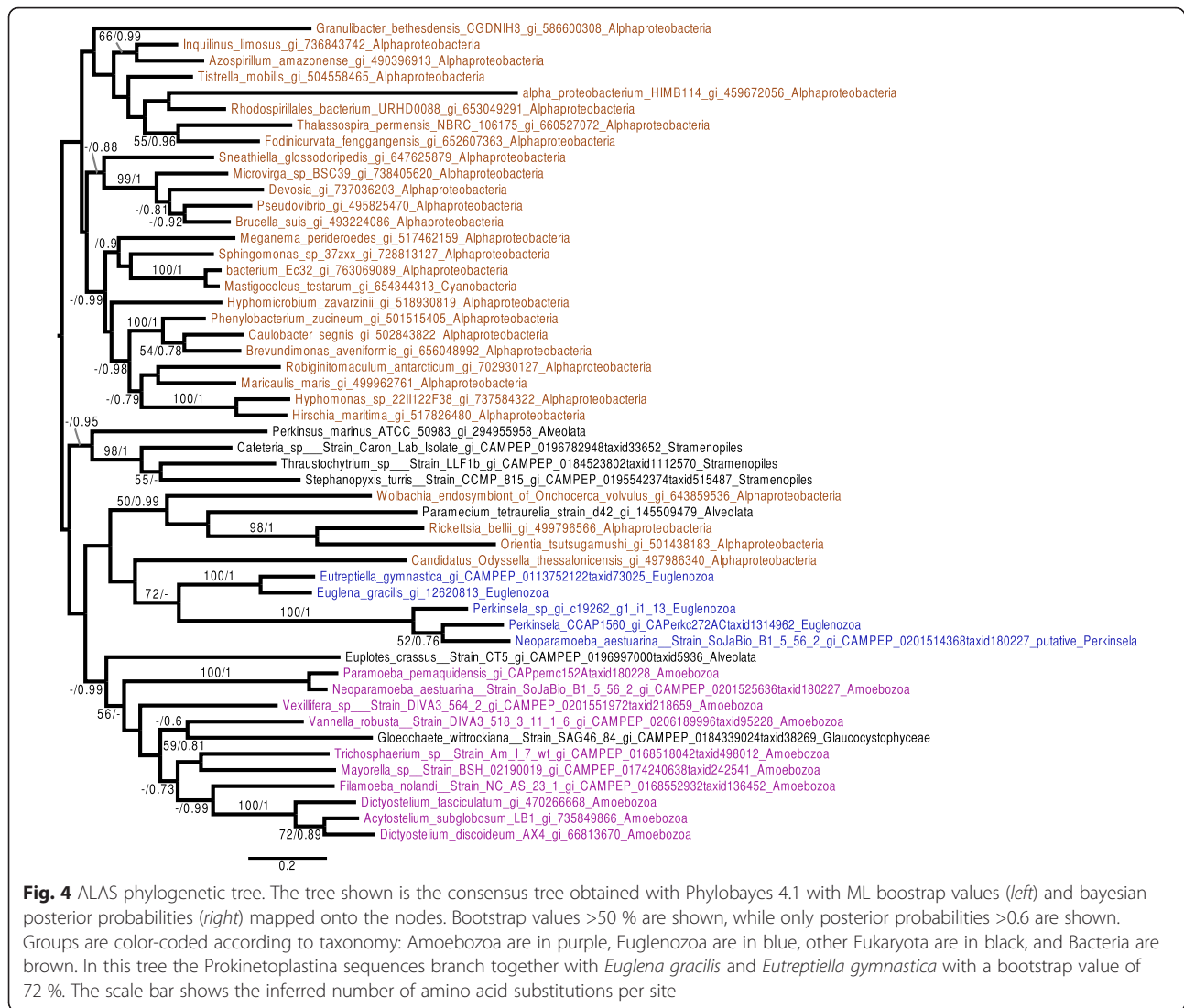
**Discussion**

**Evolution of Kinetoplastea**

A wealth of genome and transcriptome data is now available from members of the Kinetoplastea. Recent studies on the evolution of this group [10, 69], together with the new data presented here, support a robust phylogeny with the Prokinetoplastina at the root of the

Kinetoplastea. In addition, we show that the group formed by Leishmaniinae, Strigomonadinae and *Phytomonas* is robust (Fig. 2 and Additional file 3: Figure S1.1 and Table S1.2), as suggested in prior studies [1, 69]. Moreover, this group, called the LSP group in Fig. 2, is of particular interest from the perspective of heme biosynthetic enzymes in kinetoplastids, and the extent to





which these organisms have lost or acquired heme pathway genes throughout their evolutionary history. We were unable to clearly resolve the relative branching order of *Phytomonas* spp., Leishmaniinae and Strigomonadinae. Maximum likelihood trees suggested a grouping of Leishmaniinae and Strigomonadinae, while Bayesian trees grouped Leishmaniinae and *Phytomonas* spp. together. However, these topological differences could be due to the different evolutionary models employed, as suggested by Brown et al. [70]. To test if this was the case, we built phylogenetic trees in which the evolutionary models and methods (i.e., maximum likelihood and Bayesian inference) were swapped. When the maximum likelihood analysis was performed with empirical profile mixture models, and the Bayesian approach was used with the LG model, the topologies corresponding to the original models were found in both cases, suggesting that it was the evolutionary model, not

the tree building method, that was influencing the results (Additional file 3: Figure S1.3). In addition, since the empirical profile mixture models are considered to be more robust [71, 72] and our topological test with the LG4X model could not reject the close relationship between *Phytomonas* and Leishmaniinae (Additional file 3: Table S1.2), we consider a specific relationship between these two groups to the exclusion of Strigomonadinae to be most likely.

#### Heme biosynthetic pathway evolution in *Perkinsela* spp.

We have shown that the *Perkinsela* spp. enzymes ALAS and UROD share recent common ancestry with those of euglenids, to which these endosymbionts are closely related [73, 74], and also that for ALAD, PBGD, CPOX/HemE, and PPOX/HemY enzymes, the *Perkinsela* spp. sequences generally branch within a eukaryotic clade

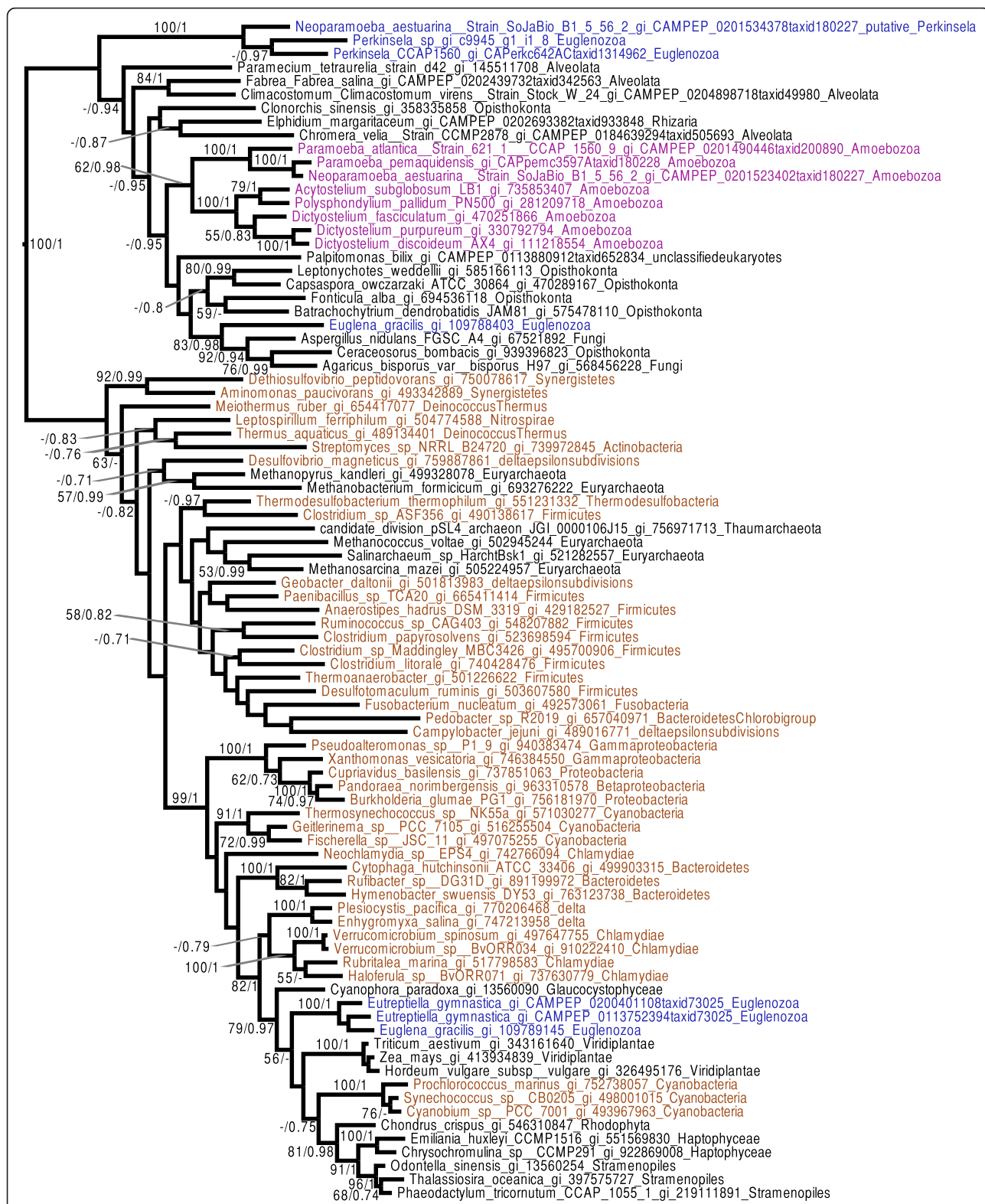


Fig. 5 (See legend on next page.)

(See figure on previous page.)

**Fig. 5** ALAD phylogenetic tree. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50 % are shown, while only posterior probabilities >0.6 are shown. The tree is rooted with the distant group composed of bacteria. Color-coding: purple = Amoebozoa, blue = Euglenozoa, other eukaryotes and Archaea = black, Bacteria = brown. The Prokinetoplastina sequences branch at the base of a clade of eukaryotic homologs from Opisthokonta, Alveolata, Rhizaria and Amoebozoa. The scale bar shows the inferred number of amino acid substitutions per site

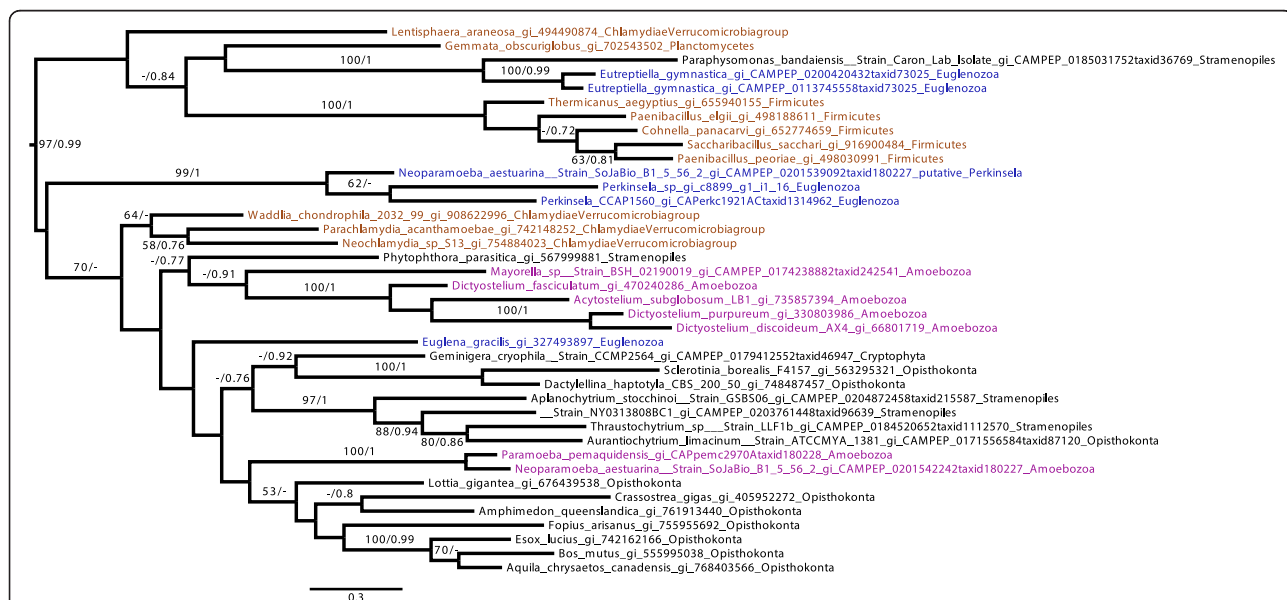
represented by most or all of the major eukaryotic lineages. This is consistent with the hypothesis that the heme pathway in *Perkinsela* spp. represents, at least in part, the ancestral pathway present in Kinetoplastea. Nevertheless, there is growing evidence supporting an important role for LGT in eukaryotes (for review see [75]), one that is particular well described in unicellular organisms [76–78]. A role for LGT in the case of *Perkinsela* spp. heme biosynthesis should thus not be dismissed.

On one hand, the ALAS and UROD enzymes of *Perkinsela* spp. were most likely inherited vertically, since they seem to share recent common ancestry with euglenozoan enzymes. On the other hand, the genes for ALAD, PBGD, CPOX/HemF, and PPOX/HemY could represent more recent acquisitions by LGT, perhaps from a eukaryotic donor. However, homologs of CPOX/HemF, PPOX/HemY and ALAD are widely distributed amongst eukaryotes and they seem to form a monophyletic clade. In addition, the position of the Prokinetoplastina at the root of eukaryotes seems to not argue in

favor of recent acquisitions by LGT, particularly in light of recent articles that place excavates at the root of the eukaryotic tree [79–81].

In the case of FeCH, the *Perkinsela* spp. gene was either acquired by LGT from Bacteroidetes or present in a common ancestor shared with Euglenozoa. Under the scenario of it being present ancestrally, the presence of Bacteroidetes homologs in this part of the tree could mean LGT from Euglenozoa to Bacteroidetes (Additional file 5: Figures S2.9 and S2.10). It seems reasonable to speculate that poor taxonomic sampling for this gene across the full breadth of the eukaryotic tree is contributing to our inability to distinguish between a lateral or vertical evolutionary history for this enzyme in Kinetoplastea.

The only enzyme of the heme biosynthetic pathway that appears to be absent from the *Perkinsela* sp. genome is UROS. However, as we have identified a potential UROS enzyme in Leishmaniinae (see Table 1 and Fig. 8), we cannot definitively reject the ancestral nature of this protein. Indeed, the fact that the Leishmaniinae homologs



**Fig. 6** PPOX/HemY phylogenetic tree. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50 % are shown, while only posterior probabilities >0.6 are shown. The tree is rooted with the distant group composed of Firmicutes, Planctomycetes and Lentisphaerae. Sequences are color-coded as follows: Amoebozoa are in purple, Euglenozoa are in blue, other eukaryotes are black, and Bacteria are brown. *Eutreptiella gymnastica*, *Euglena gracilis*, *Paraphysomonas bandaisensis* (all plastid-bearing organisms) and *Perkinsela* spp. sequences occupy distinct positions in the tree relative to other eukaryotes and bacterial homologs. The scale bar shows the inferred number of amino acid substitutions per site

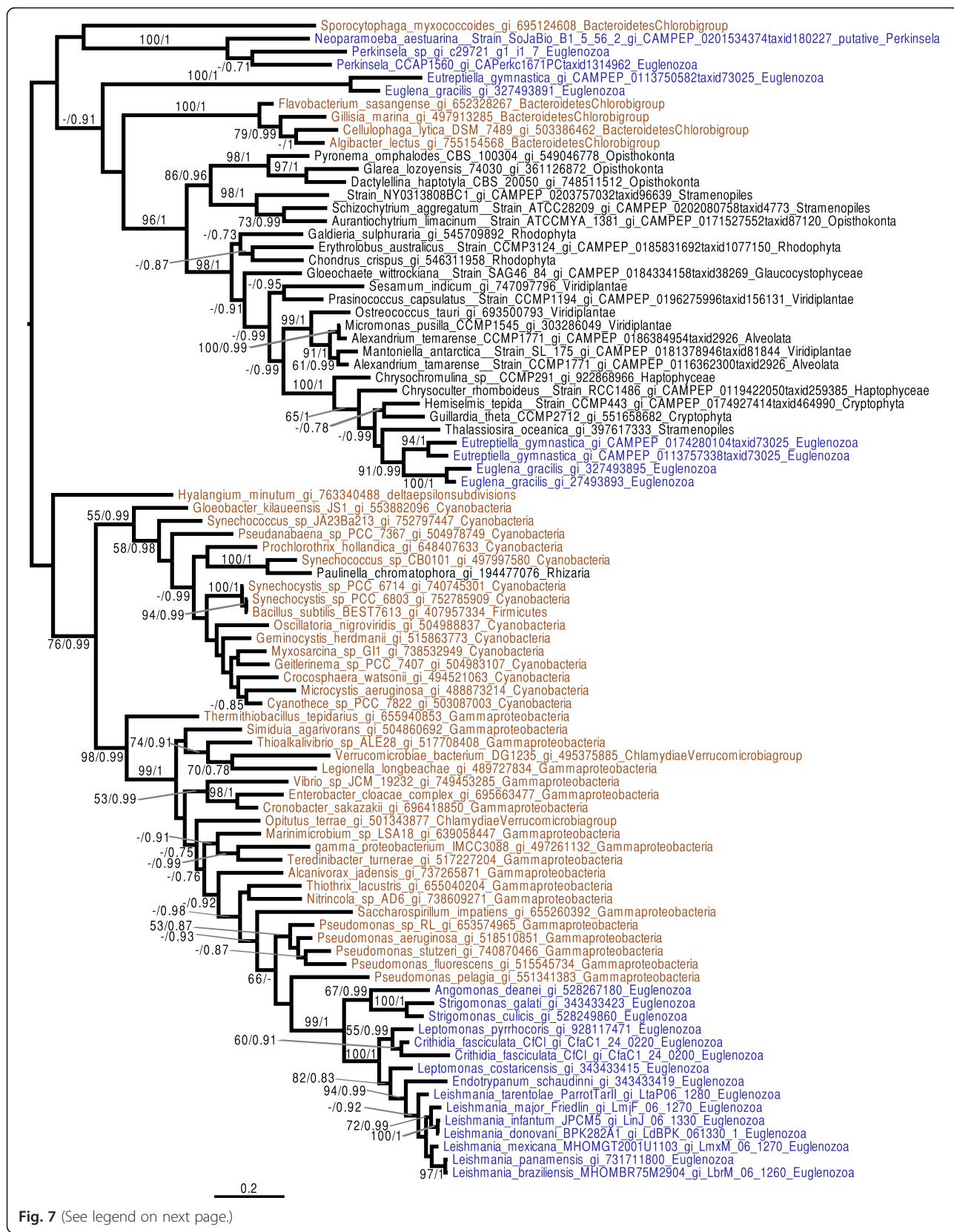


Fig. 7 (See legend on next page.)

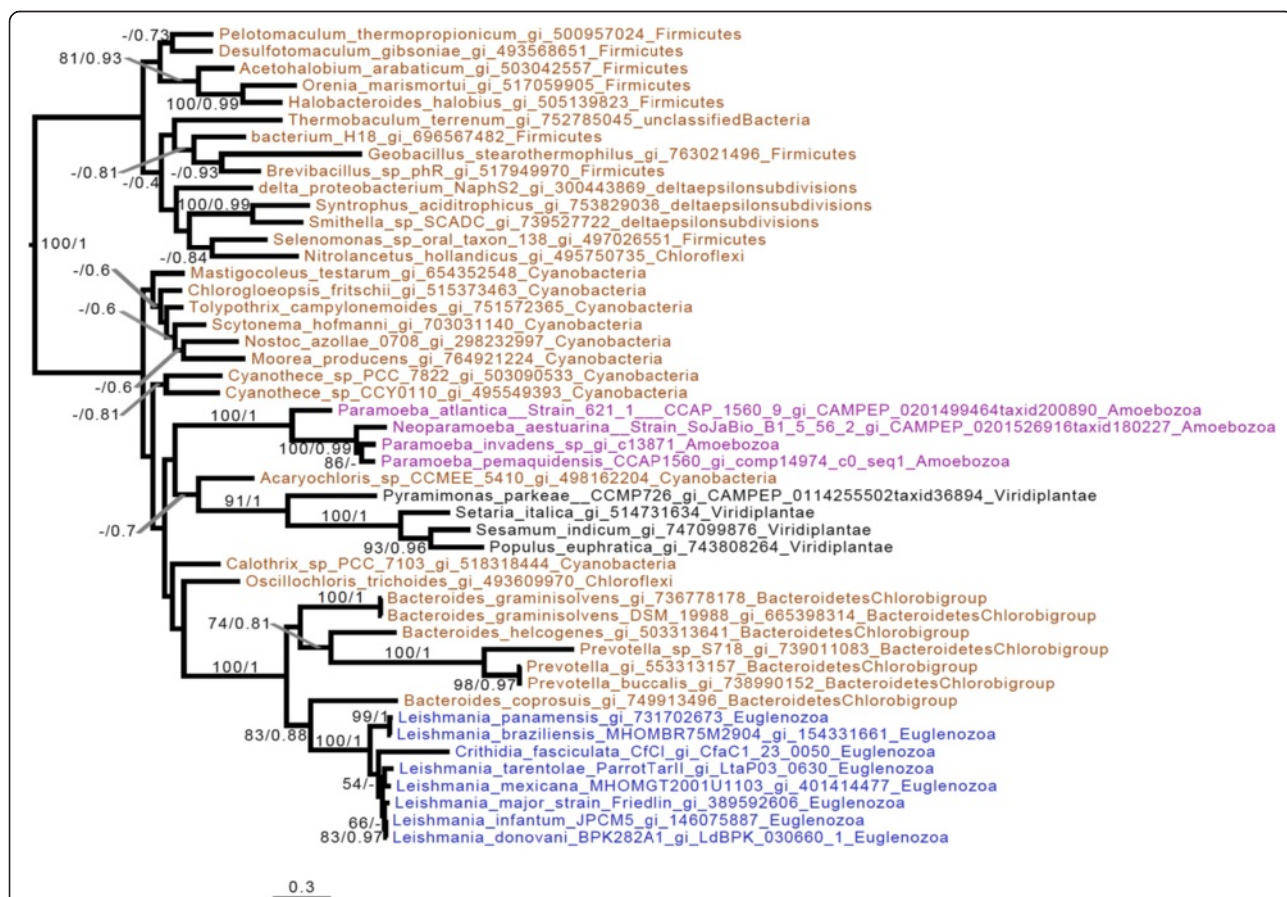
(See figure on previous page.)

**Fig. 7** CPOX/HemF phylogenetic tree. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50 % are shown, while only posterior probabilities >0.6 are shown. Groups are colored depending on their taxonomic group: Euglenozoa are in blue, other Eukaryotes are in black while Bacteria are colored brown. The *Perkinsela* spp. sequences group near the other eukaryotic sequences, albeit intermingled with bacterial sequences. The Leishmaniinae/Strigomonadinae sequences are nested within Gammaproteobacteria. The scale bar shows the inferred number of amino acid substitutions per site

branch relatively close to UROS from other eukaryotes (including the *Paramoeba* sp. host) is consistent with this possibility. Alternatively, the Leishmaniinae and/or other eukaryotes could have acquired the putative UROS gene by LGT from bacteria independently.

Interestingly, we found two potential UROS candidates in the *Paramoeba pemaquidensis* nuclear genome, which could indicate two different UROS functions in the host amoeba. However, the level of sequence conservation of one of the two putative host UROS enzymes (KU609033; not identifiable by BLASTp, only through HMMER searches) is so low that it was not possible to include it in

our phylogenetic analysis. We are thus presently only confident in the presence of one UROS enzyme in the *Paramoeba* species we investigated (KU609032 in *P. pemaquidensis*). In any case, the apparent absence of a gene encoding UROS in the *Perkinsela* sp. genome, and the presence of at least one UROS in *Paramoeba* spp., raises the possibility of metabolic interdependency between the two cells. It is conceivable that a host-encoded UROS protein and/or metabolic intermediates make their way to the kinetoplastid endosymbiont in order to compensate for the gap in the heme biosynthetic pathway of the endosymbiont. Further research



**Fig. 8** UROS phylogenetic tree. The tree shown is the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50 % are shown, while only posterior probabilities >0.6 are shown. The tree is rooted with a distant bacterial group composed of Firmicutes/Proteobacteria/Chloroflexi. Taxonomic labels are as follows: Bacteria are brown, Euglenozoa are in blue, Amoebozoa are in purple, and other eukaryotes are in black. The sequences of the kinetoplastid genus *Leishmania* branch with a group of Bacteroidetes, which belong itself to a bigger group composed of cyanobacteria, Viridiplantae, and the *Paramoeba* spp. *Perkinsela* spp. sequences are not included, as no UROS genes were detected. The scale bar shows the inferred number of amino acid substitutions per site

is needed to confirm or refute this hypothesis (see below).

#### Several independent lateral acquisitions of the heme biosynthetic pathway in Leishmaniinae/Strigomonadinae, *Phytomonas* and *Parabodo*

The discovery of a near-full heme pathway in a basal kinetoplastid allows us to assess loss, acquisition by LGT, or retention of the heme pathway in Kinetoplastea. Indeed, recent studies suggest that CPOX/HemF and PPOX/HemG were acquired by LGTs in Leishmaniinae, while FeCH appears to have been acquired three different times in kinetoplastids: in Leishmaniinae/Strigomonadinae, in *Phytomonas* and also in Bodonida [22, 25]. Here, with the inclusion of new sequences from basal-branching kinetoplastids and other Euglenozoa, LGT of CPOX/HemF and PPOX/HemG from Gammaproteobacteria to Leishmaniinae/Strigomonadinae can be inferred (Fig. 7 and Additional file 5: Figure S2.3), as well as LGT of FeCH from Firmicutes or Gammaproteobacteria to Leishmaniinae/Strigomonadinae (Additional file 5: Figure S2.5). However, the FeCH present in *Phytomonas* groups with the Leishmaniinae genus *Leptomonas*, and could be the product of a separate LGT event (Additional file 5: Figures S2.2 and S2.7). Altogether these results strongly support the hypothesis that heme biosynthetic enzymes in Leishmaniinae/Strigomonadinae, *Phytomonas/Leptomonas* and *Parabodo* were acquired by LGT.

#### Heme pathway evolution in Kinetoplastea

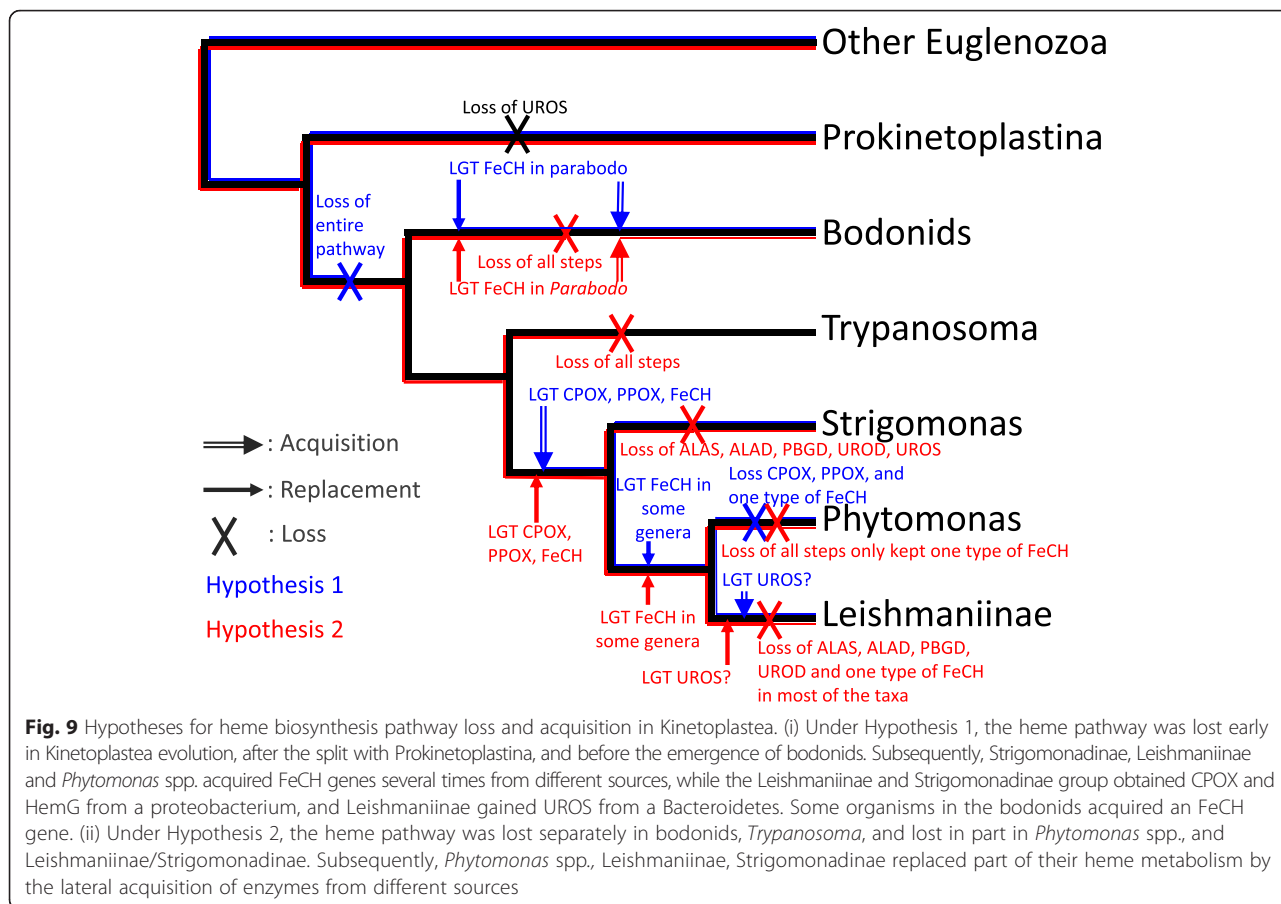
Based on our analyses, we considered three different scenarios for the evolution of the heme biosynthetic pathway in Kinetoplastea (Fig. 9). The first possibility is that the heme pathway was present ancestrally but lost early in the evolution of Kinetoplastea, after the split with Prokinetoplastina, and before the emergence of bodonids. Subsequently, Strigomonadinae, Leishmaniinae and *Phytomonas* spp. acquired FeCH several times from different sources, while the Leishmaniinae and Strigomonadinae taxa obtained CPOX/HemF and PPOX/HemG from a proteobacterium, and Leishmaniinae alone gained UROS from a Bacteroidetes. Some organisms within the bodonids gained FeCH, as proposed by Kořený et al. [22]. A second hypothesis is that the heme pathway was lost separately in bodonids, *Trypanosoma*, and lost in part in *Phytomonas* spp., and Leishmaniinae/Strigomonadinae. Subsequently, *Phytomonas* spp., Leishmaniinae, Strigomonadinae replaced part of their metabolism by LGT from different sources. Those replacements might have been related to their adaptation to a parasitic lifestyle, as is the case of the PPOX/HemG enzyme that possesses the dual ability to act under anaerobic and aerobic conditions [36]. Finally, a third hypothesis revolves around the notion that LGT has not been a significant factor in eukaryotic evolution in

general and the evolution of the heme pathway in particular. This scenario would necessitate the existence of a gene-rich organism at the root of Kinetoplastea and eukaryotes, where the presence/absence of the different components of the heme pathway in modern-day organisms is the result of differential gene loss. This latter hypothesis is in line with recent proposals for the rarity of LGT in eukaryotes [82, 83].

Given that the putative LGTs of CPOX/HemY and PPOX/HemG genes from Gammaproteobacteria to Leishmaniinae and Strigomonadinae are highly supported and no other eukaryotic homologs appear closely related (Fig. 7, Additional file 5: Figure 2.3), we consider the first two hypotheses to be more likely. Furthermore, we favor the second hypothesis over the first, for two reasons, first because shared common ancestry between the UROS enzymes of Leishmaniinae and other eukaryotes (including the *Paramoeba* spp. host and Viridiplantae) cannot be rejected outright (Fig. 8 and Additional file 5: Figure S2.8), and second, from a biochemical standpoint, gene replacement is perhaps more feasible. Indeed, it is easier to replace enzymes in an existing biochemical pathway than to acquire individual enzymes *de novo*, since there is no need for an organism to retain single enzymes from a pathway in isolation, particularly those catalyzing intermediate steps. Nevertheless, because of the exceedingly complicated phylogenies of several of the enzymes analyzed herein, it is difficult to determine which of these two scenarios is more likely. The individual phylogenies of the heme pathway proteins could in fact be explained by a combination of hypotheses one and two, e.g., an early loss of part of the heme pathway followed by differential losses of other enzymes in specific lineages.

#### Presence/absence of key enzymes in *Paramoeba pemaquidensis* and its *Perkinsela* endosymbiont suggests an intimate link between these organisms

Given that seven of the eight 'core' heme pathway enzymes exist in *Perkinsela* spp., and most of the enzymes, like those of their hosts, appear to have 'standard' subcellular localizations (Additional file 1: Table S1), the endosymbiont and host heme pathways appear to operate mainly separately. However, the absence of a UROS gene in the *Perkinsela* sp. genome and the presence of at least one UROS in the *Paramoeba* sp. genome is suggestive of metabolic 'cross-talk'. Could there be heme (intermediate) exchange between the two? Interestingly, heme transporters have recently been characterized in *Leishmania* [26, 84] and heme uptake in *Trypanosoma brucei* might be receptor-mediated [27, 85]. We searched for, but did not find, obvious homologs of the *Leishmania* and *Trypanosoma* heme transporters in the nuclear genome of *Perkinsela* sp. Two scenarios are possible: either the evolution of such heme transporters in an ancestor of



kinetoplastids took place after the divergence of Prokinetoplastina, or the transporters were lost in Prokinetoplastina. Since the Prokinetoplastina sequences currently available belong to intracellular endosymbionts, it is possible that the related transporters have been lost as a result of reductive evolution.

While the heme biosynthetic pathway provides heme as a cofactor for metabolic processes such as oxidative phosphorylation, growing evidence indicates that some organisms can thrive without it [22, 86]. Reduction and outright loss of the pathway in several Kinetoplastea has been documented, e.g., in *Leishmania* and *Trypanosoma* [25], which makes the finding of a near-complete pathway in the kinetoplastid endosymbiont *Perkinsela* sp. all the more intriguing. Given its obligate intracellular lifestyle, it would not be surprising if it were capable of taking heme, or at least intermediates of the heme biosynthetic pathway, from its host. A common theme in cellular evolution is the reduction of pathways in organisms that live (endo) symbiotically [87, 88], particularly in cases where the host can produce necessary metabolites, as is presumably the case with the amoeba hosts of *Perkinsela* spp. Although only one heme pathway enzyme appears to be missing in

*Perkinsela* spp., it is a key component and suggestive of host-endosymbiont metabolic dependency.

In addition, it is interesting to note that in the *Perkinsela* sp. endosymbiont, the reaction producing protoporphyrinogen IX is most likely catalyzed by CPOX/HemF, while the amoeba host possesses both CPOX/HemF, which acts in aerobic environments, and a CPOX/HemN homolog predicted to function anaerobically. Thus under anaerobic conditions, only the host might be capable of producing the required heme intermediates needed by both organisms. Indeed, even in the absence of specific transporters for heme or heme pathway intermediates, it is possible that non-specific mechanisms such as endocytosis might play a role in host-endosymbiont heme exchange. A better understanding of the metabolic and cell biological communication between host and endosymbiont will hopefully shed light on the nature of this unusual endosymbiotic relationship.

### Conclusions

Here we have explored the basal evolutionary position of Prokinetoplastina within Kinetoplastea, and provided the first evidence for the existence of a near-complete heme

biosynthesis pathway in a member of the Kinetoplastea. Phylogenetic analyses suggest that a major portion of the heme pathway present in Prokinetoplastina (in particular, *Perkinsela* spp.) was ancestrally present. LGT also appears to have contributed heme biosynthetic genes after the various Kinetoplastea lineages diverged from one another, perhaps related to their adaptation to a parasitic lifestyle. The presence of an almost complete heme pathway in *Perkinsela* spp. provides an opportunity for future studies aimed at shedding light on the nature of the metabolic connections between *Perkinsela* and its *Paramoeba* host.

### Availability of data and materials

All data are available through NCBI. Materials (*Paramoeba* strains) are available on request and from the CCAP. Phylogenetic data are available in Dryad: <https://datadryad.org/resource/doi:10.5061/dryad.664cp.2>.

### Additional files

**Additional file 1: Table S1.** Table of localization predictions for proteins involved in heme metabolism in Kinetoplastea. (XLSX 15 kb)

**Additional file 2: Table S2.** Table with accession numbers of proteins used for the Kinetoplastea phylogeny. (XLSX 12 kb)

**Additional file 3: Figure S1.** Supplementary figures (phylogenetic trees and tables) for the Kinetoplastea phylogeny. (PDF 361 kb)

**Additional file 4: Table S3.** Table with accession numbers of the proteins involved in the heme pathway, encoded in Kinetoplastea and the host *Paramoeba/Neoparamoeba*. (XLSX 72 kb))

**Additional file 5: Figure S2.** Supplementary phylogenetic trees of proteins involved in heme pathway, encoded in the different Kinetoplastea. (PDF 7181 kb)

### Abbreviations

ALAD: delta-aminolevulinic acid dehydratase; ALAS: 5-aminolevulinic acid synthase; BTS: bipartite targeting signal; CPOX/HemF: coproporphyrinogen III oxidase; CPOX/HemN: oxygen-independent coproporphyrinogen III oxidase; FeCh: ferrochelatase; GlT: glutamyl-tRNA synthetase; GSA-AT: glutamate-1-semialdehyde 2,1-aminomutase; GTR: glutamyl-tRNA reductase; LGT: lateral gene transfer; PBGD: porphobilinogen deaminase; PPOX/HemG: menaquinone-dependent protoporphyrinogen oxidase; PPOX/HemY: oxygen-dependent protoporphyrinogen oxidase; SP: signal peptide; TMD: transmembrane domain; UROD: uroporphyrinogen decarboxylase; UROS: uroporphyrinogen-III synthase.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

UC conceived the study, UC and DM annotated the heme pathway enzymes in *Perkinsela* sp. and *Paramoeba* sp., and UC and LE performed the phylogenetic analyses. All (UC, DM, BAC, GT, LE, JL, JMA) authors helped interpret the data and UC and JMA wrote the manuscript. All authors have read and approved the final version of the manuscript.

### Acknowledgements

This work was supported by an operating grant awarded to JMA from the Canadian Institutes of Health Research (MOP-115141). DM was supported by a grant from Dalhousie University's Centre for Comparative Genomics and Evolutionary Bioinformatics. JMA and JL acknowledge support from the Canadian Institute for Advanced Research. We thank Andrew Jackson

and Matthew Berriman for permission to analyze the *Bodo saltans* genome prior to publication.

### Author details

<sup>1</sup>Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Canada. <sup>2</sup>Centre for Comparative Genomics and Evolutionary Bioinformatics, Halifax, Nova Scotia, Canada. <sup>3</sup>Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan. <sup>4</sup>Institute of Parasitology, Biology Centre, Czech Academy of Sciences, and Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic. <sup>5</sup>Canadian Institute for Advanced Research, Toronto, Canada.

Received: 2 April 2016 Accepted: 21 April 2016

Published online: 18 May 2016

### References

- Porcel BM, Denoed F, Opperdoes F, Noel B, Madoui M-A, Hammarton TC, Field MC, Da Silva C, Couloux A, Poulain J, Katinka M, Jabbari K, Aury J-M, Campbell DA, Cintron R, Dickens NJ, Docampo R, Sturm NR, Koumandou VL, Fabre S, Flegontov P, Lukeš J, Michaeli S, Mottram JC, Szöör B, Zilberstein D, Bringaud F, Wincker P, Dollet M. The streamlined genome of *Phytomonas* spp. Relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet*. 2014;10:e1004007.
- Motta MC, Martins AC, De Souza SS, Catta-Preta CM, Silva R, Klein CC, De Almeida LGP, de Lima Cunha O, Ciapina LP, Brocchi M, et al. Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. *PLoS One*. 2013;8:e60209. doi:10.1371/journal.pone.0060209.
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 2005;309:436–42.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 2005;309:416–22.
- Jackson AP, Quail MA, Berriman M. Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa). *BMC Genomics*. 2008;9:594.
- von der Heyden S, Cavalier-Smith T. Culturing and environmental DNA sequencing uncover hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater *Neobodo designis*. *Int J Syst Evol Microbiol*. 2005;55:2605–21.
- Moreira D, López-García P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol*. 2004;54:1861–75.
- von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T. Ribosomal RNA phylogeny of bodonid and diplomonid flagellates and the evolution of Euglenozoa. *J Eukaryot Microbiol*. 2004;51:402–16.
- Simpson AGB, Stevens JR, Lukeš J. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol*. 2006;22:168–74.
- Deschamps P, Lara E, Marandé W, López-García P, Ekelund F, Moreira D. Phylogenomic analysis of Kinetoplastids supports that Trypanosomatids Arose from within Bodonids. *Mol Biol Evol*. 2011;28:53–8.
- Jackson AP. Genome evolution in trypanosomatid parasites. *Parasitology*. 2015;142:S40–56.
- Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol*. 2014;195:115–22.
- Dyková I, Fiala I, Lom J, Lukeš J. *Perkinsiella* amoebae-like endosymbionts of *Neoparamoeba* spp., relatives of the kinetoplastid *Ichthyobodo*. *Eur J Protistol*. 2003;39:37–52.
- Dyková I, Fiala I, Pecková H. *Neoparamoeba* spp. and their eukaryotic endosymbionts similar to *Perkinsiella* amoebae (Hollande, 1980): Coevolution demonstrated by SSU rRNA gene phylogenies. *Eur J Protistol*. 2008;44:269–77.
- Kudryavtsev A, Pawlowski J, Hausmann K. Description of *Paramoeba atlantica* n. sp. (Amoebozoa, Dactylopodida)—a marine amoeba from the eastern Atlantic, with emendation of the dactylopodid families. *Acta Protozool*. 2011;50:239. 10.4467/16890027AP.11.023.0023.
- Feehan C, Johnson-Mackinnon J, Scheibling R, Lauzon-Guay J, Simpson A. Validating the identity of *Paramoeba invadens*, the causative agent of recurrent mass mortality of sea urchins in Nova Scotia, Canada. *Dis Aquat Organ*. 2013;103:209–27.



17. Young ND, Dyková I, Crosbie PBB, Wolf M, Morrison RN, Bridle AR, Nowak BF. Support for the coevolution of *Neoparamoeba* and their endosymbionts, *Perkinsela* amoebae-like organisms. *Eur J Protistol.* 2014;50:509–23.
18. Tanifuji G, Kim E, Onodera NT, Gibeault R, Dlutek M, Cawthorn RJ, Fiala I, Lukeš J, Greenwood SJ, Archibald JM. Genomic Characterization of *Neoparamoeba pemaquidensis* (Amoebozoa) and Its Kinetoplastid Endosymbiont. *Eukaryot Cell.* 2011;10:1143–6.
19. Beck J, Ullman B. Nutritional requirements of wild-type and folate transport-deficient *Leishmania donovani* for pterins and folates. *Mol Biochem Parasitol.* 1990;43:221–30.
20. Ouellette M, Drummelsmith J, El Fadili A, Kündig C, Richard D, Roy G. Pterin transport and metabolism in *Leishmania* and related trypanosomatid parasites. *Int J Parasitol.* 2002;32:385–98.
21. Gutteridge W, Gaborak MA. re-examination of purine and pyrimidine synthesis in the three main forms of *Trypanosoma cruzi*. *Int J Biochem.* 1979;10:415–22.
22. Kořený L, Lukeš J, Oborník M. Evolution of the haem synthetic pathway in kinetoplastid flagellates: An essential pathway that is not essential after all? *Int J Parasitol.* 2010;40:149–56.
23. Basu S, Horáková E, Lukeš J. Iron-associated biology of *Trypanosoma brucei*. *Biochim Biophys Acta BBA Gen Subj.* 2015;1860(2):363–70.
24. Alves JMP, Voegtly L, Matveyev AV, Lara AM, da Silva FM, Serrano MG, Buck GA, Teixeira MMG, Camargo EP. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS One.* 2011;6:e23518.
25. Kořený L, Oborník M, Lukeš J. Make it, take it, or leave it: heme metabolism of parasites. *PLoS Pathog.* 2013;9:e1003088.
26. Renberg RL, Yuan X, Samuel TK, Miguel DC, Hamza I, Andrews NW, Flannery AR. The heme transport capacity of LHR1 determines the extent of Virulence in *Leishmania amazonensis*. *PLoS Negl Trop Dis.* 2015;9:e0003804.
27. Vanhollenbeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, Raes M, Moestrup SK, Pays E. A Haptoglobin-Hemoglobin receptor conveys innate immunity to *Trypanosoma brucei* in humans. *Science.* 2008;320:677–81.
28. Kořený L, Sobotka R, Kovářová J, Gnipová A, Flegontov P, Horváth A, Oborník M, Ayala FJ, Lukeš J. Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proc Natl Acad Sci.* 2012;109:3808–13.
29. Jirků M, Yurchenko VY, Lukeš J, Maslov DA. New species of insect Trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. *J Eukaryot Microbiol.* 2012;59:537–47.
30. Hunter GA, Ferreira GC. Molecular enzymology of 5-Aminolevulinic synthase, the gatekeeper of heme biosynthesis. *Biochim Biophys Acta BBA Proteins Proteomics.* 1814;2011:1467–73. doi:10.1016/j.bbapap.2010.12.015.
31. Brzezowski P, Richter AS, Grimm B. Regulation and function of tetrapyrrole biosynthesis in plants and algae. *Biochim Biophys Acta BBA Bioenerg.* 1847; 2015:968–85.
32. Panek H, O'Brian MR. A whole genome view of prokaryotic haem biosynthesis. *Microbiology.* 2002;148:2273–82.
33. Layer G, Reichelt J, Jahn D, Heinz DW. Structure and function of enzymes in heme biosynthesis. *Protein Sci.* 2010;19:1137–61.
34. Senior N, Brocklehurst K, Cooper J, Wood S, Erskine P, Shoolingin-Jordan P, Thomas P, Warren M. Comparative studies on the 5-aminolaevulinic acid dehydratases from *Pisum sativum*, *Escherichia coli* and *Saccharomyces cerevisiae*. *Biochem J.* 1996;320:401–12.
35. Kim E-J, Oh EK, Lee JK. Role of HemF and HemN in the heme biosynthesis of *V. ibrio vulnificus* under S-adenosylmethionine-limiting conditions: *V. vulnificus* HemF and HemN activities. *Mol Microbiol.* 2015;96:497–512.
36. Boynton TO, Daugherty LE, Dailey TA, Dailey HA. Identification of *Escherichia coli* HemG as a Novel, Menadiene-Dependent Flavodoxin with Protoporphyrinogen Oxidase Activity. *Biochemistry (Mosc).* 2009;48:6705–11.
37. Camadro J-M, Labbe P. Cloning and characterization of the yeast HEM14 gene coding for protoporphyrinogen oxidase, the molecular target of diphenyl ether-type herbicides. *J Biol Chem.* 1996;271:9120–8.
38. Kobayashi K, Masuda T, Tajima N, Wada H, Sato N. Molecular phylogeny and intricate evolutionary history of the three isofunctional enzymes involved in the oxidation of protoporphyrinogen IX. *Genome Biol Evol.* 2014;6:2141–55.
39. Dailey HA, Dailey TA, Wu C-K, Medlock AE, Rose JP, Wang K-F. Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters. *Cell Mol Life Sci CMLS.* 2000;57:1909–26.
40. Hamza I, Dailey HA. One ring to rule them all: Trafficking of heme and heme synthesis intermediates in the metazoans. *Biochim Biophys Acta BBA - Mol Cell Res.* 1823;2012:1617–32.
41. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 2009;23:205–11.
42. Keeling P, Burki F, Wilcox J, Allam B, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 2014;12, e1001889. doi:10.1371/journal.pbio.1001889.
43. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 2010;38(Database): D457–62. doi:10.1093/nar/gkp851.
44. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40.
45. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong S-Y, Bateman A, Punta M, Attwood TK, Sigrist CJA, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 2015;43:D213–21.
46. Dyrlov Bendtsen J, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: signalP 3.0. *J Mol Biol.* 2004;340:783–95.
47. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000;300:1005–16.
48. Petsalaki EI, Bagos PG, Litou ZL, Hamodrakas SJ. PredSL: A tool for the N-terminal sequence-based prediction protein subcellular localization. *Genomics Proteomics Bioinformatics.* 2006;4:48–55.
49. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *PROTEOMICS.* 2004;4:1581–90.
50. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
51. Bolte K, Bullmann L, Hempel F, Bozarth A, Zauner S, Maier U-G. Protein targeting into secondary plastids. *J Eukaryot Microbiol.* 2009;56:9–15.
52. Patron NJ, Waller RF. Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays.* 2007;29:1048–58.
53. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
54. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol.* 2010;10:210. doi:10.1186/1471-2148-10-210.
55. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 2009; 25:2286–8.
56. Le S, Gascuel O, Lartillot N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics.* 2008;24:2317–23.
57. Le SQ, Dang CC, Gascuel O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol Biol Evol.* 2012;29:2921–36.
58. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
59. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
60. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol.* 1990;31:151–60.
61. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol.* 1989;29:170–9.
62. Shimodaira H, Hasegawa M. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 1999;16:1114.
63. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc R Soc B Biol Sci.* 2002;269:137–42. doi:10.1098/rspb.2001.1862.
64. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
65. Price MN, Dehal PS, Arkin AP, et al. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5:e9490.

66. Maruyama S, Eveleigh RJ, Archibald JM. Treertrimmer: a method for phylogenetic dataset size reduction. *BMC Res Notes*. 2013;6:145.
67. Kořený L, Oborník M. Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biol Evol*. 2011;3:359–64.
68. Alsmark C, Foster PG, Sicheritz-Ponten T, Nakjang S, Embley TM, Hirt RP. Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol*. 2013;14:R19.
69. Flegontov P, Votýpka J, Skalický T, Logacheva MD, Penin AA, Tanifuji G, Onodera NT, Kondrashov AS, Volf P, Archibald JM, Lukeš J. *Paratrypanosoma* Is a novel early-branching Trypanosomatid. *Curr Biol*. 2013;23:1787–93. doi:10.1098/rspb.2013.1755.
70. Brown MW, Sharpe SC, Silberman JD, Heiss AA, Lang BF, Simpson AGB, Roger AJ. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc R Soc B Biol Sci*. 2013;280:20131755.
71. Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol*. 2007;7 Suppl 1:S4.
72. Wang H-C, Susko E, Roger AJ. An amino acid substitution-selection model adjusts residue fitness to improve phylogenetic estimation. *Mol Biol Evol*. 2014;31:779–92.
73. Gile GH, Faktorová D, Castlejohn CA, Burger G, Lang BF, Farmer MA, Lukeš J, Keeling PJ. Distribution and phylogeny of EFL and EF-1 $\alpha$  in euglenozoa suggest ancestral Co-occurrence followed by differential loss. *PLoS One*. 2009;4:e5162.
74. Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, le Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawłowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW. The revised classification of eukaryotes. *J Eukaryot Microbiol*. 2012;59:429–514. doi:10.1111/j.1550-7408.2012.00644.x.
75. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. 2015;16:472–82.
76. Gawryluk RMR, Eme L, Roger AJ. Gene fusion, fission, lateral transfer, and loss: Not-so-rare events in the evolution of eukaryotic ATP citrate lyase. *Mol Phylogenet Evol*. 2015;91:12–6.
77. He D, Fu C-J, Baldauf SL. Multiple Origins of Eukaryotic *cox15* Suggest Horizontal Gene Transfer from Bacteria to Jakobid Mitochondrial DNA. *Mol Biol Evol* 2015:msv201. doi: 10.1093/molbev/msv201.
78. Corradi N. Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Annu Rev Microbiol*. 2015;69:167–83.
79. He D, Fiz-Palacios O, Fu C-J, Fehling J, Tsai C-C, Baldauf SL. An alternative root for the eukaryote tree of life. *Curr Biol*. 2014;24:465–70.
80. Lasek-Nesselquist E, Gogarten JP. The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol Phylogenet Evol*. 2013;69:17–38.
81. Cavalier-Smith T. The neomuran revolution and phagotrophic origin of eukaryotes and cilia in the light of intracellular coevolution and a revised tree of life. *Cold Spring Harb Perspect Biol*. 2014;6:a016006.
82. Ku C, Nelson-Sathi S, Roettger M, Garg S, Hazkani-Covo E, Martin WF. Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *Proc Natl Acad Sci*. 2015;112:10139–46.
83. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature*. 2015;524:427–32.
84. Huynh C, Yuan X, Miguel DC, Renberg RL, Protchenko O, Philpott CC, Hamza I, Andrews NW. Heme Uptake by *Leishmania amazonensis* Is mediated by the transmembrane protein LHR1. *PLoS Pathog*. 2012;8:e1002795.
85. Stødkilde K, Torvund-Jensen M, Moestrup SK, Andersen CBF. Structural basis for trypanosomal haem acquisition and susceptibility to the host innate immune system. *Nat Commun*. 2014;5:5487.
86. Rao AU, Carta LK, Lesuisse E, Hamza I. Lack of heme synthesis in a free-living eukaryote. *Proc Natl Acad Sci U S A*. 2005;102:4270–5.
87. Cavalier-Smith T. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann Bot*. 2005;95:147–75.
88. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2011;10:13–26.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

