Plant-Type Trehalose Synthetic Pathway in *Cryptosporidium* and Some Other Apicomplexans

Yonglan Yu[1,2], Haili Zhang[2], Guan Zhu[2,3]*

1 College of Veterinary Medicine, China Agricultural University, Beijing, China, 2 Department of Veterinary Pathobiology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, Texas, United States of America, 3 Faculty of Genetics Program, Texas A&M University, College Station, Texas, United States of America

Abstract

Background: The trehalose synthetic pathway is present in bacteria, fungi, plants and invertebrate animals, but is absent in vertebrates. This disaccharide mainly functions as a stress protectant against desiccation, heat, cold and oxidation. Genes involved in trehalose synthesis have been observed in apicomplexan parasites, but little was known about these enzymes. Study on trehalose synthesis in apicomplexans would not only shed new light into the evolution of this pathway, but also provide data for exploring this pathway as novel drug target.

Methodology/Principal Findings: We have observed the presence of the trehalose synthetic pathway in *Cryptosporidium* and other apicomplexans and alveolates. Two key enzymes (trehalose 6-phosphate synthase [T6PS; EC 2.4.1.15] and trehalose phosphatase [TPase; EC 3.1.3.12] are present as Class II bifunctional proteins (T6PS-TPase) in the majority of apicomplexans with the exception of *Plasmodium* species. The enzyme for synthesizing the precursor (UDP-glucose) is homologous to dual-substrate UDP-galactose/glucose pyrophosphorylases (UGGPases), rather than the "classic" UDP-glucose pyrophosphorylase (UGPase). Phylogenetic reconstructions indicate that both T6PS-TPases and UGGPases in apicomplexans and other alveolates are evolutionarily affiliated with stramenopiles and plants. The expression level of *T6PS-TPase* in *C. parvum* is highly elevated in the late intracellular developmental stage prior to or during the production of oocysts, implying that trehalose may be important in oocysts as a protectant against environmental stresses. Finally, trehalose has been detected in *C. parvum* oocysts, thus confirming the trehalose synthetic activity in this parasite.

Conclusions/Significance: A trehalose synthetic pathway is described in the majority of apicomplexan parasites including *Cryptosporidium* and the presence of trehalose was confirmed in the *C. parvum* oocyst. Key enzymes in the pathway (i.e., T6PS-TPase and UGGPase) are plant-type and absent in humans and animals, and may potentially serve as novel drug targets in the apicomplexans.

Introduction

Trehalose (α-D-glucopyranosyl-1,1-α-D-glucopyranoside) is a disaccharide consisting of two units of glucose linked by an α,α-1, 1-glycosidic linkage. It is present in a wide range of organisms, including prokaryotes, fungi, plants, and invertebrate animals [1–4]. Trehalose may serve as an energy source, but its major function is known as a protectant against various stresses including desiccation/dehydration, heat, cold and oxidation [5,6]. Additionally, trehalose may be an integral component of cell wall glycolipids in mycobacteria and corynebacteria [7], and the pathway and intermediary metabolites may also play regulatory roles in signaling, sugar metabolism or stress-responses [1–5,8–13].

Bacteria, fungi, plants and invertebrates can synthesize trehalose from UDP-glucose by trehalose 6-phosphate synthase (**T6PS or TPS**, EC 2.4.1.15) and trehalose phosphatase (**TPase or TP**, EC 3.1.3.12) (**Figure 1**). Vertebrates including humans and other

mammals are incapable of synthesizing, but able to metabolize this disaccharide by possessing a trehalase (EC 3.2.1.28). T6PS and TPase can be discrete enzymes or fused together as bifunctional proteins, which, in plants, are referred to as Class I and Class II enzymes, respectively [14,15].

The Phylum Apicomplexa is comprised of unicellular parasites including many important pathogens in humans and animals, such as *Plasmodium*, *Babesia*, *Theileria*, *Toxoplasma*, *Eimeria* and *Cryptosporidium* [16–18]. This group of protists are evolutionarily close to dinoflagellates and ciliates, and these three phyla form a super group termed alveolates (Alveolata) [18,19]. The capacity for trehalose synthesis in apicomplexans was first recognized when a gene encoding a T6PS domain was identified in the *Cryptosporidium* genome [20,21]. Trehalose synthetic genes were also observed in the *Theileria* and *Babesia* genomes [22–24]. Trehalose may act as a stress protectant in these parasites, particularly for the oocyst stage in the natural environment (e.g., *Cryptosporidium*) or the sexual
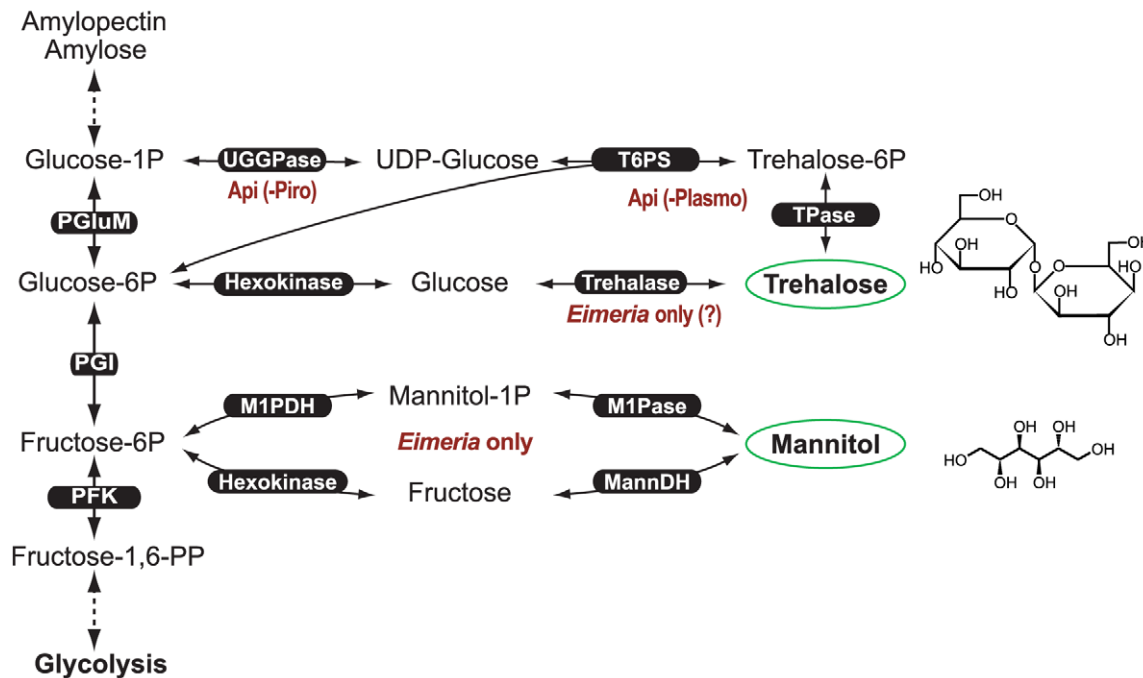
Figure 1. Presence of trehalose synthetic pathway in the apicomplexans as determined from available genome sequences. Within the pathway, UDP-glucose/galatose pyrophosphorylase (UGGPase) is present in almost all apicomplexans with the exception for piroplasmids (e.g., *Theileria* and *Babesia*) [marked as Api(-Piro)], while trehalose-6P synthase and trehalose phosphatase (TPase) are fused as a bifunctional protein that is present in all apicomplexans with the exception for *Plasmodium* species [marked as Api(-Plasmo)]. Trehalase may only be present in the intestinal coccidia (*Eimeria*). For comparison, the mannitol cycle present only in the intestinal coccidia (represented by *Eimeria tenella*) is also illustrated.
doi:10.1371/journal.pone.0012593.g001

developmental stage in vectors (e.g., *Theileria* and *Babesia*). Therefore, it can be reasonably assumed that trehalose may play an important role in protecting apicomplexans from various stresses in their complex life cycle. However, virtually nothing is known on the molecular and biochemical features of this important pathway in apicomplexans, and related genes are also not well annotated in the genome databases.

In the present study we annotated all available apicomplexan genomes and performed phylogenetic reconstructions to delineate the evolutionary history of the key enzymes in the trehalose synthetic pathway. Trehalose has been detected in *C. parvum* oocysts, thus confirming that this pathway is active in this apicomplexan. We show that the synthesis of trehalose in the apicomplexans is mediated by a single bifunctional T6PS-TPase that is evolutionarily affiliated with plant Class II enzymes. Considering that the anti-stress mannitol cycle has been proven to be a drug target in the coccidian *Eimeria* [25,26], we speculate that the unique, plant-type trehalose synthesis may also be explored as a novel drug target in the apicomplexans.

Results and Discussion

Genomic evidence of trehalose synthetic pathway in the majority of apicomplexan lineages except for the *Plasmodium*

Trehalose is synthesized by two reactions, in which T6PS first converts UDP-glucose to trehalose 6-phosphate and this product is in turn converted to trehalose by TPase (**Figure 1**). The production of UDP-glucose from glucose 1-phosphate can be mediated by a classic UDP-glucose pyrophosphorylase (UGPase, EC 2.7.7.9; also termed UTP-glucose-1-phosphate uridylyltransferase) [27,28], or by a UDP-galactose/glucose pyrophosphorylase

(UGGPase, EC 2.7.7.64) that is dual-functional and capable of synthesizing both UDP-glucose and UDP-galactose [29]. Genome analysis indicates the presence of UGGPase among alveolates including the majority of apicomplexan species, with the exception of *Theileria* and *Babesia*; the dinoflagellate *Perkinsus marinus*; and ciliates such as *Paramecium* and *Tetrahymena* (**Table 1**). All aveolates, however, lack the classic UGPase, whereas humans and other mammals possess both UGPase and UAP, but lack UGGPase. Among apicomplexans, the lack of UGPase and UGGPase in *Theileria* and *Babesia* suggests that these two piroplasmids may rely on host cells to supply UDP-glucose, or there might be an unknown pathway to synthesize this nucleotide sugar.

UGGPase belongs to a glycosyltransferase family A (GTA) group that also includes UDP-*N*-acetylglucosamine pyrophosphorylase (**UAP**) and other glycosyltransferases [29]. This grouping might explain why some UGGPase genes, such as those from *C. parvum* (GenBank No. XP_628360; locus_tag, cdg7_1830), *T. gondii* (e.g., XP_002370608 and EEB03468), *Perkinsus* (EER18291) and some other eukaryotes are annotated as **UAP** family proteins. Although UAP (EC 2.7.7.23) is evolutionarily related to UGGPases (see phylogenetic data below for more detail), it catalyzes the formation of UDP-*N*-acetyl-α-D-glucosamine from *N*-acetyl-α-D-glucosamine 1-phosphate and UTP [30–33]. On the other hand, *Cryptosporidium* and many other alveolates indeed possess authentic UAP orthologs in their genomes (e.g., XP_625683 in *C. parvum*, XP_665528 in *C. hominis*, and XP_002140971 in *C. muris*).

UGGPases in *Cryptosporidium* and other apicomplexans consist of more than 650 amino acids (aa), and contain all putative active site motifs conserved among the GTA proteins (**Figure 2**). Among the 5 motifs, the fourth one in the apicomplexans is highly distinct from other GTA proteins in their amino acid

Table 1. Evidence of trehalose synthetic pathways in *Cryptosporidium* and some other apicomplexans and alveolates as shown by the presence of genes encoding UDP-galactose/glucose pyrophosphorylase (UGGPase), Class II trehalose-6P synthase-trehalose phosphatase (T6PS-TPase) in their genomes.

Group	Genus	Trehalose metabolism				Mannitol cycle		
		UGPase	UGGPase	T6PS-TPase	Trehalase	M1PDH	M1Pase	MannDH
	(No. species with available genomes)	2.7.7.9	2.7.7.64	2.4.1.15 3.1.3.12	3.2.1.28	1.1.1.17	3.1.3.22	1.1.1.255
Apicomplexan	*Cryptosporidium (3)*	-	+	+	-	-	-	-
	Plasmodium (7)	-	+	-	-	-	-	-
	Theileria (2)	-	-	+	-	-	-	-
	Babesia (1)	-	-	+	-	-	-	-
	Toxoplasma (1)	-	+	+	-	-	-	-
	Eimeria (1)	-	+	+	+	+	+	+
Dinoflagellate	*Perkinsus (1)*	-	+	+	+	-	-	-
Ciliate	*Paramecium (1)*	-	+	+	+	-	-	-
	Tetrahymera (2)	-	+	+	-	-	-	-
Mammal	*Homo (1)*	+	-	-	+	-	-	-

Notes: Genes involved in the anti-stress related mannitol cycle were also minded for comparison. Plus and minus symbols indicate the presence and absence of specified genes in various apicomplexans, a dinoflagellate and two ciliates at genus level. Humans are listed to represent mammalian hosts for apicomplexan parasites.
doi:10.1371/journal.pone.0012593.t001

compositions, from which we predict that this motif is probably important in defining the substrate preferences of this superfamily of enzymes.

Two enzymes (i.e., T6PS and TPase) are involved in converting UDG-glucose to trehalose (**Figure 1**), in which T6PS is a member of the glucosyltransferase family-B (GTB) proteins [34,35], while TPase belongs to HAD superfamily type IIB enzymes [36]. Nearly all alveolates for which genome data are available possess a putative bifunctional protein containing both T6PS and TPase domains (**Table 1 and Figure 3**). The only exception is *Plasmodium* that lacks either bifunctional or discrete trehalose synthetic enzymes, although their genomes encode UGGPases. Due to the limitation of genome data, it is yet unclear whether other haemosporida also lack trehalose synthesis. Genes encoding T6PS-TPases are intronless in *Cryptosporidium*, but contain introns in other apicomplexans. Conceptually translated apicomplexan T6PS-TPase proteins are typically large, comprised of more than 1400 amino acids. A number of motifs conserved in both T6PS and TPase domains could be identified from the N- and C-termini of apicomplexan T6PS-TPases (**Figure 3**), which further support their identities.

As indicated by the phylogenetic data below, apicomplexan T6PS-TPases are more closely related to the plant Class II TPS enzymes for which the functional roles in plants are still under investigation [14,37–39]. There was a speculation that some class II proteins from *Arabidopsis thaliana* might be simply TPases (Leyman et al. 2001) [40], and it was also reported that Class II *AtTPS* genes were incapable of complementing the T6PS-deficient *Saccharomyces cerevisiae* (Tps1 mutant) [14]. On the other hand, another study demonstrated that at least one Class II gene (i.e., *AtTPS6*) involved in the regulation of cell shape and plant architecture was able to rescue the yeast Tsp1 mutant phenotype, suggesting that "*AtTPS6* gene may be unique among class II *AtTPS* genes in affecting the cell shape of leaf pavement cells" [37]. Our comprehensive genomic analysis has revealed that apicomplexans, dinoflagellates and ciliates lack discrete T6PS and TPase genes, thus they may only rely on the Class II-like genes to make the trehalose that they possess (as described below).

Most organisms including animals, plants, fungi and bacteria can reutilize trehalose by converting it back to glucose by trehalase [41–47]. Among the aveolates, trehalase is present in at least some dinoflagellates and ciliates, but absent in the majority of apicomplexans with a possible exception of *Eimeria* (**Table 1**). Within the partially sequenced *E. tenella* genome, we were able to identify a contig that encodes fragmented protein sequences with significant homology to invertebrate trehalase (i.e., dev_EIMER_ contig_00020795), for which its authenticity remains to be experimentally determined. The lack of a trehalase is indicative that the majority of apicomplexans may not recycle trehalose back into its energy metabolism, or they may simply reuse the bifunctional T6PS-TPases in the reverse direction when needed.

Plant-affinity of apicomplexan trehalose synthetic genes

The availability of trehalose synthetic gene sequences in a large number of diverse taxonomic groups permits a good and unbiased sampling of taxa based on sequence similarity for phylogenetic reconstructions. In our initial **BI** trees inferred from a large dataset, T6PS-TPases were mainly clustered by major taxonomic groups, in which apicomplexans were clustered together with dinoflagellates, red algae, diatoms and then Class II enzymes from green plants and algae (**Figure S1**). The "plant-affinity" of apicomplexan enzymes was further confirmed by **BI** and **ML** analyses of a second dataset with fewer taxa, but a greater number of alignable positions after excluding the more distant prokaryotic and invertebrate sequences (**Figure 4**). Trees inferred from both datasets were robust as the majority of nodes were strongly supported by the PP and BP values in BI and ML analyses. As shown in the second tree (**Figure 4**), the fungal clade (outgroup) is split into two subgroups: one contained regulatory Tsl1/Tps3 subunits, while the other contained Tps2 subunits of the trehalose synthesis complex. Plants, green and red algae, diatoms and dinoflagellates generally possess both Class I and Class II proteins, thus forming distinguished clades. Apicomplexan enzymes joined with *P. marinus* (dinoflagellate), diatoms (stramenopiles) and then red algae to form a monophyletic group as a sister to the Class II enzymes from green plants and algae.
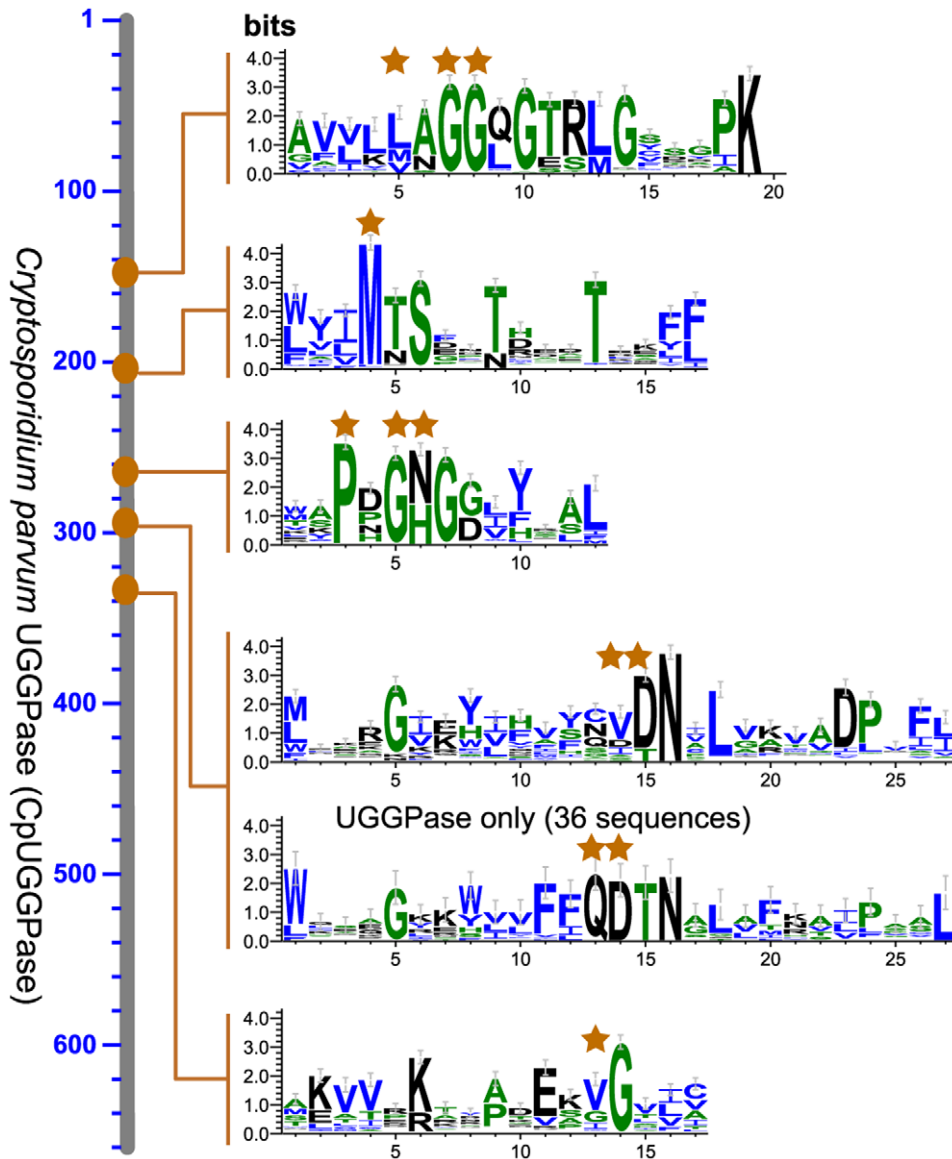
Figure 2. Structure of apicomplexan UDP-glucose/galatose pyrophosphorylase (UGGPase) as exemplified by *Cryptosporidium parvum* protein (CpUGGPase). Sequence logos represent conserved motifs and domains as determined from 224 sequences of glucosyltransferase family-A (GTA) proteins including UGGPases and UDP-*N*-acetylglucosamine pyrophosphorylases (UAPs). In the fourth domain, UGGPases displayed a very unique sequence pattern that differs significantly from other GTA proteins. Stars indicate amino acids important at the active sites.
doi:10.1371/journal.pone.0012593.g002

These observations indicate that all Class II enzymes likely originated from a common ancestor, and those in alveolates including apicomplexans are more closely related to the stramenopiles (heterokonts) and red algae than to the green algae and plants. Additionally, the presence of multiple Class II isoforms in many organisms that were placed together within single species, or intermixed with taxonomically closely related species (as exemplified in *A. thaliana* and *S. cerevisiae*) suggests that the trehalose synthetic gene family has undergone expansion by a number of gene duplication events at different taxonomic levels. However, such a gene expansion did not (or at least less commonly) occur in the apicomplexans, as in many cases, only single-copy genes are present in this lineage (**Figure 4**).

UGGPases are a newly identified group of enzymes capable of making UDP-glucose and UDP-galactose [29], in which UDP-glucose is a precursor for making trehalose. As two closely related members of GTA family proteins, UGGPases share a number of sequence features with UAP enzymes. UGGPases appear to be mainly present in plants, stramenopiles, kinetoplasts and alveolates, whereas UAPs are present in all major prokaryotic and eukaryotic taxonomic groups. Phylogenetic analysis using a large dataset of more than 224 sequences containing both UGGPases and UAPs was performed using BI method, but the runs were not well converged even after more than 10^6 generations. However, although the poorly converged BI tree was mostly polytomic and unable to resolve the phylogenetic relationships among the majority of sequences, it confidently grouped all UGGPases into a single clade (**Figure 5A**). Additionally, we have also employed a less time-consuming quartet-puzzling analysis using TreePuzzle v5.2 program (*WAG +*
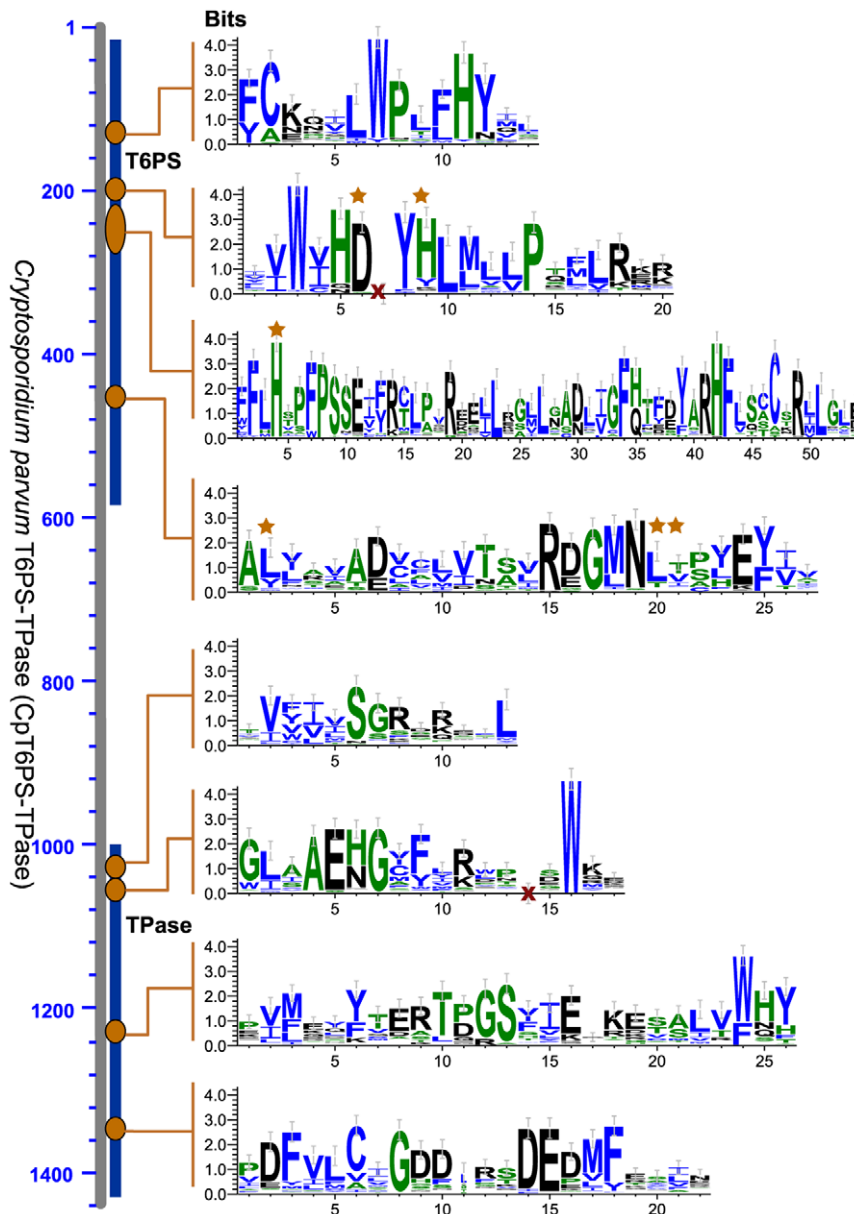
Figure 3. Structure of putative apicomplexan Class II, bifunctional trehalose-6P synthase–trehalose phosphatase (T6PS-TPase) as exemplified by *Cryptosporidium parvum* protein (CpT6PS-TPase). Sequence logos corresponding to the T6PS and TPase domains represent conserved motifs determined from 140 orthologs. Stars indicate amino acids important at the active sites.
doi:10.1371/journal.pone.0012593.g003

$F_{inv} + \Gamma_{(8)}$). The puzzle tree produced after 10,000 puzzling steps was better resolved than the BI tree, and again supported the monophyly of UGGPases (**Figure 5B**).

We have also reconstructed trees only from the UGGPase sequences, in which green plants and algae are grouped together, followed by two stramenopiles (**Figure 5C**). However, the remaining sequences formed a polytomic node, in which apicomplexans were only resolved at genus levels. Although sequences within ciliates, stramenopiles and kinetoplastids were better resolved, relationships between these major taxonomic groups were unresolved at phyla levels. These observations indicates that either these sequences do not contain sufficient informative positions for phylogeny and/or there is an insufficient number of available sequences to fully resolve the phylogenetic relationship among UGGPases. It is noticed that plant-type

UGGPase genes are also present in kinetoplastids and the chlamydia group bacteria (**Figure 5C**). However, it is not unusual since the "plant" relationship in some enzymes and metabolic pathways in kinetoplastids and chlamydia are well documented [48–52]. Collectively, both sequence similarity and phylogenetic data clearly support the notion that apicomplexan UGGPases are of plant affinity and are highly divergent from the closest enzymes in their host animals and humans.

The plant affinity of apicomplexan trehalose synthetic enzymes have previously been implied by the primary genome sequencing projects mainly based on the top hits in BLAST searches [20,24], and more recently by a large scale metaTIGER phylogenetic tree database project mainly based on a high-throughput PhyML analysis [53,54]. Our findings are congruent with these earlier observations, but much more robust by employing detailed and

BI/ML

● =1/100
○ =1/95-99

Theileria annulata XP_954185
Theileria parva XP_766116
Babesia bovis XP_001609560
Cryptosporidium muris XP_002142543
Cryptosporidium parvum XP_627376
Toxoplasma gondii XP_002371192
Perkinsus marinus EER18482
Perkinsus marinus EER16132

Apicomplexans
Dinoflagellate

0.96/61
1/89
1/100

0.97/72

1/93

Thalassiosira pseudonana XP_002286273
Phaeodactylum tricornutum XP_002183251
Thalassiosira pseudonana XP_002287922
Phaeodactylum tricornutum XP_002180425
Thalassiosira pseudonana XP_002288483
Porphyra haitanensis ABG75728
Cyanidioschyzon merolae CMI293C*

Stramenopiles
Red Algae

1/93

Arabidopsis thaliana NP_173799 (TPS9)
Arabidopsis thaliana NP_177186 (TPS8)
Arabidopsis thaliana NP_176221 (TPS10)
Ricinus communis XP_002521023
Ricinus communis XP_002511428
Oryza sativa NP_001061805
Oryza sativa EAY87814
Oryza sativa NP_001061925
Arabidopsis thaliana NP_179460 (TPS11)
Ricinus communis XP_002533178
Oryza sativa NP_001063104
Oryza sativa NP_001049428
Arabidopsis thaliana NP_974105 (TPS6)
Arabidopsis thaliana NP_567538 (TPS5)
Oryza sativa NP_001056050
Oryza sativa NP_001044248
Oryza sativa NP_001044141
Ricinus communis XP_002522255
Arabidopsis thaliana NP_172129
Physcomitrella patens XP_001765959
Physcomitrella patens XP_001772649
Physcomitrella patens XP_001766932
Physcomitrella patens XP_001766049
Ostreococcus lucimarinus XP_001420835
Ostreococcus tauri CAL57334
Micromonas sp. XP_002504033
Micromonas pusilla EEH60166
Dunaliella viridis ACH87586

Class II

Green Plants
Geen Algae

0.99/82
1/93
0.51/52
1/93
1/84
0.71/67
1/94
1/92
0.58/51
0.5/x
1/79
1

Arabidopsis thaliana NP_173142 (TPS2)
Arabidopsis thaliana NP_194485 (TPS4)
Arabidopsis thaliana NP_177979 (TPS1)
Ricinus communis XP_002531237
Ricinus communis XP_002522296
Oryza sativa EAY98715
Selaginella lepidophylla AAD00829
Physcomitrella patens XP_001769595
Physcomitrella patens XP_001778505
Micromonas pusilla EEH59517
Micromonas sp. XP_002507506
Ostreococcus_tauri CAL50152

Class I

0.72/x
0.73/x
0.74/x
0.95/84
1/89

Perkinsus marinus EER11216
Perkinsus marinus EER12881

Dinoflagellate

Cyanidioschyzon merolae CMO053C*
Cyanidioschyzon merolae CMP219C*
Cyanidioschyzon merolae CMS178C*

Red Algae

0.94/69
0.84/61
0.98/67

Saccharomyces cerevisiae NP_013608 Tsl1
Candida glabrata XP_446899
Saccharomyces cerevisiae NP_013988 Tps3
Zygosaccharomyces rouxii XP_002498572
Kluyveromyces lactis XP_454457
Ashbya gossypii NP_985133
Lachancea thermotolerans XP_002552421
Debaryomyces hansenii XP_460651
Candida albicans XP_717151
Pichia angusta AAF73139
Yarrowia lipolytica XP_504619
Schizosaccharomyces pombe NP_594728

Tsl1/Tps3-like

0.55
0.67
1/97
0.88
0.68
1/100
1/99
1/100
1/91
1/100

Penicillium chrysogenum XP_002563154
Penicillium chrysogenum XP_002567539
Aspergillus fumigatus XP_749114
Penicillium chrysogenum XP_002562460
Aspergillus fumigatus XP_749534
Neurospora crassa XP_964949

Fungi

0.53/56
0.56/x
1/78

Candida glabrata XP_446592
Saccharomyces cerevisiae CAA50025 Tps2-like
Vanderwaltozyma polyspora XP_001646298
Zygosaccharomyces rouxii XP_002498622
Kluyveromyces lactis XP_454407
Lachancea thermotolerans XP_002556207
Ashbya gossypii NP_986798
Candida albicans XP_721536
Debaryomyces hansenii XP_462075
Yarrowia lipolytica XP_502823
Aspergillus fumigatus XP_755036
Aspergillus clavatus XP_001270589
Aspergillus niger XP_001395056
Neurospora crassa XP_956349
Schizosaccharomyces pombe NP_594430

Tps2-like

1/91
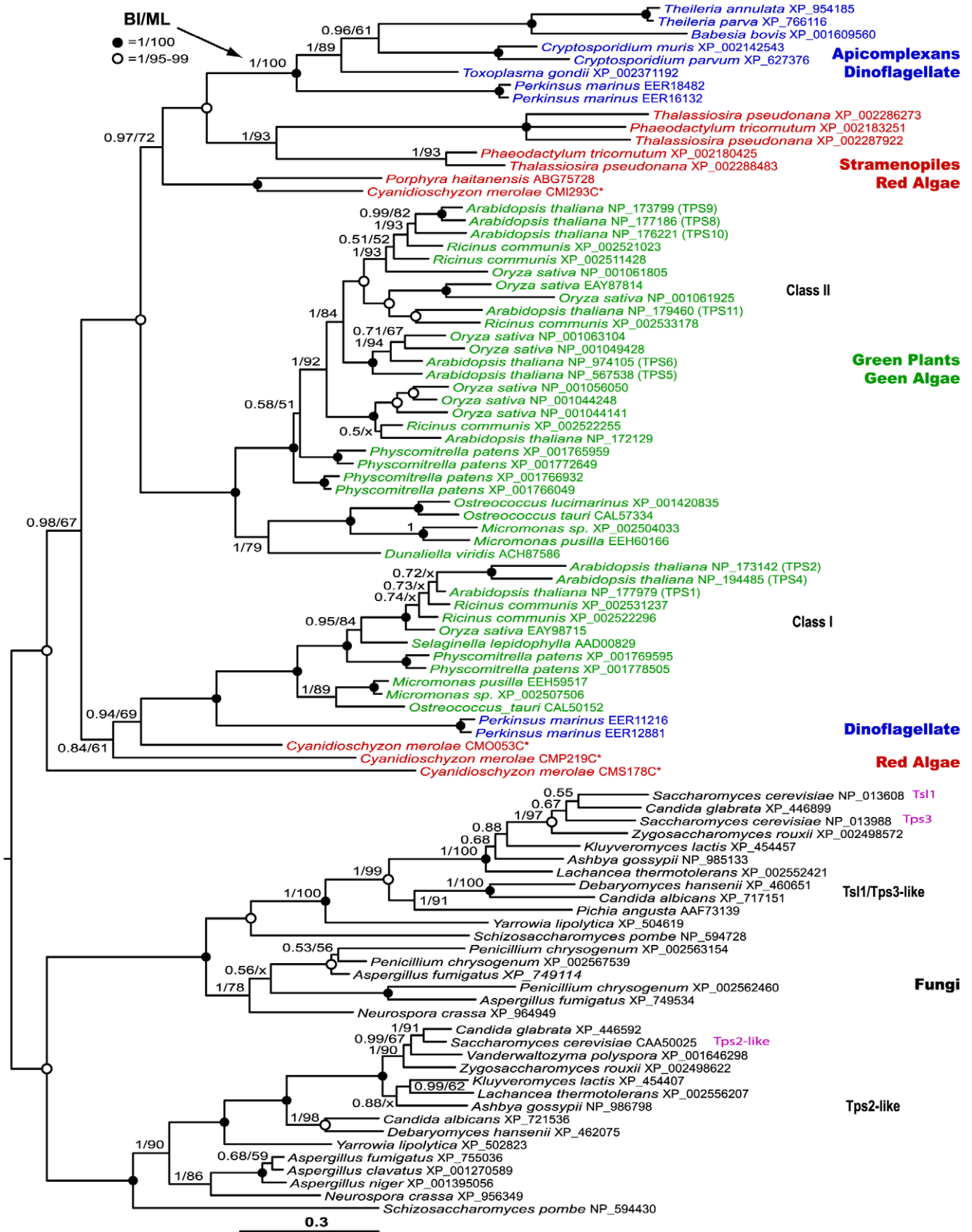0.99/67
1/90
0.99/62
0.88/x
1/98
1/90
0.68/59
1/86

0.3

Figure 4. Phylogenetic tree inferred from 93 Class II trehalose-6P synthase-trehalose phosphatase (T6PS-TPase) protein sequences by Bayesian inference (BI) and maximum likelihood (ML) methods. The BI tree is shown here and arbitrarily rooted using fungal sequences as an outgroup. Numbers at nodes are posterior probability (PP) and bootstrap (BP) supporting values determined by BI and ML analyses, respectively. Solid circles indicate nodes with 100% supports by both PP and BP values, while open circles indicates nodes with 100% PP and 95%-99% BP supports.
doi:10.1371/journal.pone.0012593.g004

comprehensive phylogenetic models with BP/PP supporting values and conserved domain analysis.

T6PS-TPase gene expression in *Cryptosporidium* is elevated in the late intracellular developmental stages, corresponding to the development of environmental oocysts

We hypothesize that trehalose in apicomplexans is an important stress protectant, such as conferring resistance to desiccation or freezing, and secondly, that trehalose functions in the external oocyst stage that is subject to stresses of environmental exposure. Our real-time qRT-PCR analysis with parasite 18S rRNA levels as internal controls indicates that *CpT6PS-TPase* gene is detectable in all life cycle stages, but highly elevated in the late developmental stage. Specifically, at 72 hr post-infection, the relative level of *CpT6PS-TPase* transcripts was ~3.5-fold higher than the overall mean level, or ~34-fold higher than the level at 6 h post-infection) (**Figure 6**). The 72 hr post-infection time is correlated with the sexual development and the production of oocysts in *C. parvum*, indicating that this parasite is likely producing a significantly higher level of trehalose for oocysts that would be eventually released into external environment. Similar expression patterns were also observed for genes encoding *Cryptosporidium* oocyst wall proteins (COWPs) [55,56], and more importantly, also for genes encoding anti-stress related mannitol cycle enzymes that were also found highly expressed in the oocyst production stage in *Eimeria* [57]. One small surprise is the low level expression of *CpT6PS-TPase* gene in the oocysts (**Figure 6**). However, this is not unexpected since trehalose, as a protectant, likely needs to be already synthesized before oocysts are mature and released into natural environment, similar to the oocyst wall proteins.

Trehalose is present in the *C. parvum* oocysts

Using a trahalase/hexokinase/gluconate-6-phosphate dehydrogenase-coupled spectrophotometric assay, we have determined that trehalose is present in the *C. parvum* oocysts with a concentration at 0.199 ± 0.014 μg/10^7 oocysts. This roughly equals to 3.5×10^7 trehalose molecules per oocyst. While we are still refining the assay to detect the trehalose contents in the parasite intracellular life cycle stages, which is complicated by the presence of host cells, the presence of trehalose in the oocysts ultimately confirms that the trehalose synthetic pathway is active in the apicomplexan *C. parvum*. As mentioned earlier, the expression of *CpT6P-TPase* gene was the lowest in the free oocyst stage, but elevated in the later intracellular developmental stages, which implies that trehalose might mainly be synthesized during the formation of oocysts before they were defecated into the external environment.

In summary, we have performed a thorough genomic analysis and reconstructed the trehalose synthetic pathway in the apicomplexans. This pathway is active in *C. parvum* as a gene encoding the key enzyme T6PS-TPase is expressed during the parasite life cycle and trehalose can be detected in the parasite oocysts. Our sequence analysis and phylogenetic recontructions indicate that the two genes involved in the apicomplexan trehalose synthesis (i.e., UGGPase and T6PS-TPase) are plant-type genes. The closest orthologs of these enzymes are those from other

alveolates and stramenopiles, suggesting the these genes were probably acquired from the plant-lineage by an ancestral alveolate, retained in some apicomplexan lineages. On the other hand, humans and other mammals lack the Class II bifunctional T6PS-TPase and UGGPase, suggesting that, similar to the enzymes involved in the mannitol cycle in *Eimeria*, these two enzymes (particularly the plant-type T6PS-TPase) may be explored as novel drug targets in *Cryptosporidium* or other apicomplexans. Finally, although in silico analysis and phylogenetic reconstructions can provide some important clues on the functions and evolutionary affinity of proteins under investigation, the true functions and biological roles of enzymes in the apicomplexan trehalose synthetic pathway can only be derived from further biochemical and biological studies.

Materials and Methods

Reconstruction of the trehalose synthetic pathway from the genomes

All publicly available apicomplexan genome sequences (completely or nearly completely sequenced) were searched for genes encoding enzymes involved in or connected to the trehalose metabolism. These include 15 species from the genera *Cryptosporidium*, *Toxoplasma*, *Eimeria*, *Plasmodium*, *Theileria* and *Babesia*. Most sequences genome sequences are available at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/), except for the *Eimeria tenella* genome that is available at the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/Projects/E_tenella/). Databases at the EuPathDB (http://eupathdb.org/) containing raw and annotated genomes of most apicomplexans (except for those of *Eimeria*, *Theileria* and *Babesia* species) were also searched to ensure quality [58]. In addition to the apicomplexans, searches were also extended to include dinoflagellates and ciliates to gain a more complete picture of the pathway among alveolates.

Targeted enzymes include T6PS, TPase, trehalase, UDP-glucose pyrophosphorylase (**UGPase**) and UDP-galactose/glucose pyrophosphorylase (**UGGPase**). For comparison, enzymes associated with mannitol cycle (i.e., mannitol 1-phosphate dehydrogenase [M1PDH], mannitol-1-phosphatase [M1Pase] and mannitol dehydrogenase [MannDH]) were also searched.

Annotated protein sequences of these enzymes from apicomplexans and other major taxonomic groups (i.e., plants, bacteria, fungi and animals) were used as queries to repeatedly search the nucleotide or translated protein sequences in these databases using TBLASTN and BLASTP algorithms [59], respectively. Hits were then used as queries to search the non-redundant protein sequences and conserved domain databases at NCBI to validate their identities and conserved motifs. Additionally, the mappings and annotation of corresponding apicomplexan genes at the Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) were also inspected for annotated and missing enzymes [60]. Annotated apicomplexan sequences within the trehalose metabolic pathway at KEGG were retrieved and analyzed for their identities.

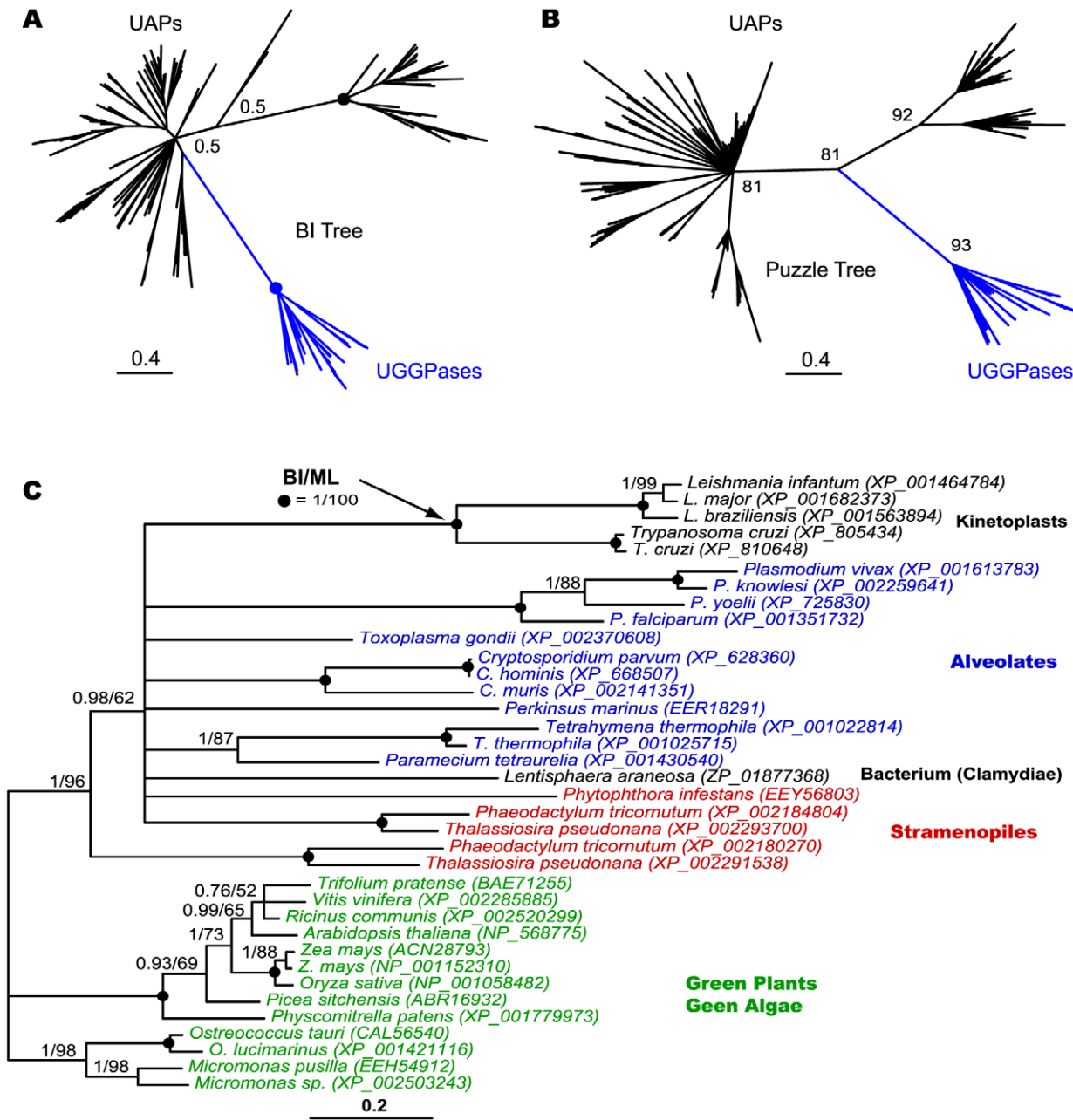Retrieved sequences were aligned using MUSCLE program (version 3.6) [61,62], and conserved domains and motifs were

Figure 5. Phylogenetic relationship of apicomplexan UDP-glucose/galatose pyrophosphorylase (UGGPase) among glucosyltransferase family A (GTA) proteins including UDP-*N*-acetylglucosamine pyrophosphorylases (UAPs). A) Tree derived from 224 GTA taxa by Bayesian inference (BI) with posterior probability values indicated at major nodes. **B**) Maximum likelihood (ML) tree derived from 224 GTA taxa by quartet puzzling method with quartet puzzling support values indicated at major nodes. **C**) Trees inferred from 93 taxa containing only UGGPase sequences by BI and ML methods using MrBayes and TreeFinder programs, respectively. Numbers at nodes are posterior probability (PP) and bootstrap (BP) supporting values determined by BI and ML analyses. Solid circles indicate nodes with 100% supports by both PP and BP values.
doi:10.1371/journal.pone.0012593.g005

identified and visualized as sequence logos by bits with the height of symbols within the stack indicates the relative frequency of each amino of the positions using WebLogo 3 (http://weblogo.threeplusone.com/) [63].

Phylogenetic reconstructions

Phylogenetic reconstructions were performed for the three key enzymes - the bifunctional T6PS-TPase and UGGPase to elucidate their evolutionary histories. To build datasets for phylogenetic analysis, various apicomplexan protein sequences

were used as queries to search orthologs from the NBCI non-redundant protein databases. Orthologs were also searched and retrieved from a red alga *Cyanidioschyzon merolae* Genome Project (http://merolae.biol.s.u-tokyo.ac.jp/) [64,65]. In our initial analysis, nearly all retrieved sequences, up to a few hundreds taxa, were aligned using a MUSCL program (version 3.6). After visual inspection of the alignments with a MacVector program (version 11.0.4), short and nearly identical sequences were removed from the datasets. Neighbor-joining (NJ) trees were first constructed with MacVector program from positions containing no gaps with
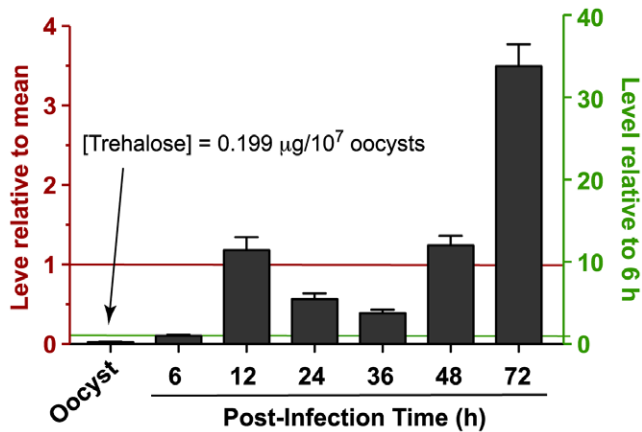
Figure 6. Relative levels of *Cryptosporidium parvum* trehalose-6P synthase-trehalose phosphatase (*CpT6PS-TPase*) gene expressed as determined by real-time quantitative RT-PCR in mature oocysts and intracellular developmental stages in HCT-8 host cells for various post-infection times. The levels of *CpT6PS-TPase* transcripts were first normalized with those parasite 18S rRNA and then displayed as relative to the overall mean (left-side Y-axis) and to the level at 6 h post-infection (right-side Y-axis). The detected trehalose concentration in the parasite oocysts is also indicated.
doi:10.1371/journal.pone.0012593.g006

Poisson-corrected protein distances. These NJ trees were used to guide unbiased selections of sequences that represented all major taxonomic groups (e.g., prokaryotes, fungi, plants, animals, stramenopiles and aveolates) that possessed corresponding enzyme orthologs. In the final datasets, ambiguous and gap-containing positions were excluded from subsequent phylogenetic reconstructions by Bayesian inference (**BI**) and maximum likelihood (**ML**) methods.

The final T6PS-TPase dataset contained 93 taxa and 447 aa positions derived from both T6PS and TPase domains. For UGGPase, we first retrieved orthologs within the GTA superfamily including UAP sequences from the public databases, and built a dataset to contain 224 taxa and 200 amino acid (aa) positions. We also built a smaller dataset containing only UGGPase orthologs with 33 taxa and 228 aa positions. Two datasets were similarly built for T6PS-TPase protein sequences, in which the large dataset contained 140 taxa with 354 aa sampled from all major taxonomic groups. After phylogenetic analysis with BI method, a second dataset was built to include only groups that were more closely related to the apicomplexans (93 taxa), which gave us a larger number of alignable positions (i.e., 472 aa).

Phylogenetic trees were then inferred from the UGGPase and T6PS-TPase datasets by BI and ML methods (and a quartet-puzzling ML method was also applied to the large UGGPase dataset). BI analysis was performed with a parallel version of MrBayes program (version 3.1.2; http://mrbayes.csit.fsu.edu/) [66], in which at least 10^6 generation of searches were performed with two independent runs, each containing 4 chains running simultaneously. The current trees were saved every 1,000 generations, and the posterior probability (**PP**) values were calculated from the saved BI trees obtained after the runs converged (typically after the first 25% trees were discarded). ML bootstrapping analysis was conducted from 100 replicated sequences with a TreeFinder program (version October 2008; http://www.treefinder.de/) [67], in which consensus ML trees and bootstrap proportion (**BP**) supporting values were calculated from the 100 ML trees by a majority ruling law. Both BI and ML analyses used a WAG amino acid substitution model. Among-site

rate heterogeneity considered the fraction of invariance (F_{inv}) and a discrete 8-rate gamma distribution (i.e., $WAG + F_{inv} + \Gamma_{(8)}$). Trees were visualized using a FigTree program (version 1.2.3 or 1.3.1; http://tree.bio.ed.ac.uk/software/figtree/) and annotated with an Adobe Illustrator program (version CS4; http://www.adobe.com/).

Expression profile of T6PS-TP gene in *C. parvum*

To verify that trehalose synthesis is associated with stresses in the apicomplexans, we analyzed the expression pattern of T6PS-TP gene from *C. parvum (CpT6PS-TPase)* by a SYBR-green-based one-step real-time quantitative RT-PCR (qRT-PCR) method as previously described (e.g., [68–70]). Total RNA was isolated from various life cycle stages of *C. parvum* (IOWA-1 strain) using an RNeasy isolation kit (Qiagen). These include the mature oocysts (an environmental stage), intracellular developmental stages obtained by infecting human HCT-8 cells for various times (i.e., from 6 to 72 hr post-infection) that cover the parasites development from first and second generations of merogony, to gametogenesis and oocyst production.

The expression levels of *CpT6PS-TPase* gene were determined with an iCycler iQ Real-Time PCR Detection System (Bio-Rad) using primer pairs specific to *CpT6PS-TPase* (i.e., CpT6P-3473, 5′-AGG CAA GCT TTG ACT TGG ATT-3′ and CpT6P-3587R, 5′ TGC TTT TGC TTC TGT TGG AGT-3′) and *C. parvum* 18S rRNA (Cp18S-1011F, 5′-TTG TTC CTT ACT CCT TCA GCA C-3′ and Cp18S-1185R, 5′-TCC TTC CTA TGT CTG GAC CTG-3′). Reactions containing 20 ng of total RNA and 0.2 μM of specified primers were first incubated at 48°C for 30 min to synthesize cDNA, heated at 95°C for 15 min to inactivate reverse transcriptase, and then subjected to 40 thermal cycles (95°C 20 sec, 50°C 30 sec and 72°C 30 sec) of PCR amplification. Because RNA isolated from intracellular parasites contained a large portion of host cell RNA, the levels of parasite 18S rRNA were first used as baselines to normalize those of *CpT6PS-TPase* transcripts in all samples by calculating the ΔC_T (i.e., $C_{T[CpT6P-TPase]} - C_{T[Cp18S]}$) for calculating the relative levels of *CpT6PS-TPase* transcripts with $2^{-\Delta CT}$. The relative level of *CpT6PS-TPase* transcripts in each developmental stage was plotted in relative to the overall mean level of all samples (i.e., ratio between individual normalized level and the overall mean) and to the level of the earliest intracellular sample (i.e., 6 h post-infection).

Detection of trehalose in *C. parvum*

Fresh *C. parvum* oocysts (IOWA-1 strain) were purchased from Bunch Grass Farm (Deary, Idaho, USA) and stored at 4°C until use. Before the assay, oocysts were treated with 10% Clorox in ice for 10 min, washed for 5–8 times with water by centrifugation, and purified by a Percoll gradient centrifugation protocol as described elsewhere. To detect trehalose, 2×10^8 oocysts were suspended in 250 μl PBS (pH 7.2) and subject to 5 times of freeze/thaw cycles. The disrupted oocysts were treated at 65°C for 20 min to inactivate enzymes, and then centrifuged at 10,000×g to remove insoluble materials. A spectrophotometry-based trehalose detection kit purchased from Megazyme International Ireland Limited (catalog # K-TREH) was used in this study. The kit detected trehalose by a three-enzyme coupled assay, in which trehalose was first hydrolyzed to D-glucose that was then phosphorylated to glucose-6-phosphate (G-6-P). Finally, G-6-P was oxidized to gluconate-6-phosphate by G-6-P dehydrogenase (G6PDH) using $NADP^+$ as an electron receiver, in which the formation of NADPH could be monitored by measuring the increase of absorbance at 340 nm (OD_{340}).

The assay was performed according to the manufacturer's protocol, but the reaction volumes were reduced to 63 μl using a 384-well UV transmissible microplate to save parasite materials. The OD_{340} values were measured with a Multiskan Spectrum Microplate Spectrophotometer (Thermo Scientific). For each sample, the difference in the OD_{340} values between reactions with and without trehalase were used to calculate trehalose concentration against a standard curve generated with serial concentrations of trehalose under the same experimental conditions. In this assay, we have noticed that heat-inactivation of oocyst extracts was critical as active enzymes in the non-inactivated samples could rapidly consume NADPH to the background level.

Supporting Information

Figure S1 Unrooted tree inferred from T6PS-TPase protein sequences (93 taxa, 444 amino acid positions) by Bayesian inference (BI) method using the same amino acid substitution model and the consideration of rate heterogeneity as described in the Methods section. Solid circles indicate select major nodes that were 100% supported by posterior probability (PP) values. Only representative genus names are labeled to indicate taxonomic affiliations of major clusters.
Found at: doi:10.1371/journal.pone.0012593.s001 (0.39 MB JPG)

Author Contributions

Conceived and designed the experiments: YY HZ GZ. Performed the experiments: YY HZ GZ. Analyzed the data: GZ. Wrote the paper: GZ.

References

1. Paul MJ, Primavesi LF, Jhurreea D, Zhang Y (2008) Trehalose metabolism and signaling. Annu Rev Plant Biol 59: 417–441.
2. Gancedo C, Flores CL (2004) The importance of a functional trehalose biosynthetic pathway for the life of yeasts and fungi. FEMS Yeast Res 4: 351–359.
3. Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. Glycobiology 13: 17R–27R.
4. Wingler A (2002) The function of trehalose biosynthesis in plants. Phytochemistry 60: 437–440.
5. Chen Q, Haddad GG (2004) Role of trehalose phosphate synthase and trehalose during hypoxia: from flies to mammals. J Exp Biol 207: 3125–3129.
6. De Silva-Udawatta MN, Cannon JF (2001) Roles of trehalose phosphate synthase in yeast glycogen metabolism and sporulation. Mol Microbiol 40: 1345–1356.
7. Takayama K, Wang C, Besra GS (2005) Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin Microbiol Rev 18: 81–101.
8. Hanson J, Smeekens S (2009) Sugar perception and signaling–an update. Curr Opin Plant Biol 12: 562–567.
9. Paul MJ (2008) Trehalose 6-phosphate: a signal of sucrose status. Biochem J 412: e1–2.
10. Paul M (2007) Trehalose 6-phosphate. Curr Opin Plant Biol 10: 303–309.
11. Halford NG, Paul MJ (2003) Carbon metabolite sensing and signalling. Plant Biotechnol J 1: 381–398.
12. Eastmond PJ, Li Y, Graham IA (2003) Is trehalose-6-phosphate a regulator of sugar metabolism in plants? J Exp Bot 54: 533–537.
13. Eastmond PJ, Graham IA (2003) Trehalose metabolism: a regulatory role for trehalose-6-phosphate? Curr Opin Plant Biol 6: 231–235.
14. Ramon M, De Smet I, Vandesteene L, Naudts M, Leyman B, et al. (2009) Extensive expression regulation and lack of heterologous enzymatic activity of the Class II trehalose metabolism proteins from *Arabidopsis thaliana*. Plant Cell Environ 32: 1015–1032.
15. Li P, Ma S, Bohnert HJ (2008) Coexpression characteristics of trehalose-6-phosphate phosphatase subfamily genes reveal different functions in a network context. Physiol Plant 133: 544–556.
16. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52: 399–451.
17. Finlay BJ (2004) Protist taxonomy: an ecological perspective. Philos Trans R Soc Lond B Biol Sci 359: 599–610.
18. Morrison DA (2009) Evolution of the Apicomplexa: where are we now? Trends Parasitol 25: 375–382.
19. Cavalier-Smith T (1991) Cell diversification in heterotrophic flagellates. In: Patterson DJ, Larsen J, eds. The Biology of Free-living Heterotrophic Flagellates. Oxford: Oxford University Press. pp 113–131.
20. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, et al. (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science 304: 441–445.
21. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, et al. (2004) The genome of *Cryptosporidium hominis*. Nature 431: 1107–1112.
22. Brayton KA, Lau AO, Herndon DR, Hannick L, Kappmeyer LS, et al. (2007) Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog 3: 1401–1413.
23. Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, et al. (2005) Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. Science 309: 134–137.
24. Pain A, Renauld H, Berriman M, Murphy L, Yeats CA, et al. (2005) Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. Science 309: 131–133.
25. Allocco JJ, Nare B, Myers RW, Feiglin M, Schmatz DM, et al. (2001) Nitrophenide (Megasul) blocks *Eimeria tenella* development by inhibiting the mannitol cycle enzyme mannitol-1-phosphate dehydrogenase. J Parasitol 87: 1441–1448.
26. Schmatz DM (1997) The mannitol cycle in *Eimeria*. Parasitology 114(Suppl): S81–89.
27. Kleczkowski LA, Geisler M, Ciereszko I, Johansson H (2004) UDP-glucose pyrophosphorylase. An old protein with new tricks. Plant Physiol 134: 912–918.
28. Meng M, Geisler M, Johansson H, Harholt J, Scheller HV, et al. (2009) UDP-glucose pyrophosphorylase is not rate limiting, but is essential in *Arabidopsis*. Plant Cell Physiol 50: 998–1011.
29. Dai N, Petreikov M, Portnoy V, Katzir N, Pharr DM, et al. (2006) Cloning and expression analysis of a UDP-galactose/glucose pyrophosphorylase from melon fruit provides evidence for the major metabolic pathway of galactose metabolism in raffinose oligosaccharide metabolizing plants. Plant Physiol 142: 294–304.
30. Bulik DA, van Ophem P, Manning JM, Shen Z, Newburg DS, et al. (2000) UDP-N-acetylglucosamine pyrophosphorylase, a key enzyme in encysting *Giardia*, is allosterically regulated. J Biol Chem 275: 14722–14728.
31. Mio T, Yabe T, Arisawa M, Yamada-Okabe H (1998) The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression, and catalytic mechanism. J Biol Chem 273: 14392–14397.
32. Mok MT, Edwards MR (2005) Kinetic and physical characterization of the inducible UDP-N-acetylglucosamine pyrophosphorylase from *Giardia intestinalis*. J Biol Chem 280: 39363–39372.
33. Peneff C, Ferrari P, Charrier V, Taburet Y, Monnier C, et al. (2001) Crystal structures of two human pyrophosphorylase isoforms in complexes with UDPGlc(Gal)NAc: role of the alternatively spliced insert in the enzyme oligomeric assembly and active site architecture. EMBO J 20: 6191–6202.
34. Lariviere L, Sommer N, Morera S (2005) Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. J Mol Biol 352: 139–150.
35. Gibson RP, Turkenburg JP, Charnock SJ, Lloyd R, Davies GJ (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem Biol 9: 1337–1346.
36. Lu Z, Dunaway-Mariano D, Allen KN (2005) HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. Biochemistry 44: 8684–8696.
37. Chary SN, Hicks GR, Choi YG, Carter D, Raikhel NV (2008) Trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in *Arabidopsis*. Plant Physiol 146: 97–107.
38. Wu W, Pang Y, Shen GA, Lu J, Lin J, et al. (2006) Molecular cloning, characterization and expression of a novel trehalose-6-phosphate synthase homologue from *Ginkgo biloba*. J Biochem Mol Biol 39: 158–166.
39. Harthill JE, Meek SE, Morrice N, Peggie MW, Borch J, et al. (2006) Phosphorylation and 14-3-3 binding of *Arabidopsis* trehalose-phosphate synthase 5 in response to 2-deoxyglucose. Plant J 47: 211–223.
40. Leyman B, Van Dijck P, Thevelein JM (2001) An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. Trends Plant Sci 6: 510–513.
41. Frison M, Parrou JL, Guillaumot D, Masquelier D, Francois J, et al. (2007) The *Arabidopsis thaliana* trehalase is a plasma membrane-bound enzyme with extracellular activity. FEBS Lett 581: 4010–4016.

42. Carroll JD, Pastuszak I, Edavana VK, Pan YT, Elbein AD (2007) A novel trehalase from *Mycobacterium smegmatis* - purification, properties, requirements. FEBS J 274: 1701–1714.

43. Parrou JL, Jules M, Beltran G, Francois J (2005) Acid trehalase in yeasts and filamentous fungi: localization, regulation and physiological function. FEMS Yeast Res 5: 503–511.

44. Dmitryjuk M, Zoltowska K (2003) Purification and characterization of acid trehalase from muscle of *Ascaris suum* (Nematoda). Comp Biochem Physiol B Biochem Mol Biol 136: 61–69.

45. Oesterreicher TJ, Markesich DC, Henning SJ (2001) Cloning, characterization and mapping of the mouse trehalase (Treh) gene. Gene 270: 211–220.

46. Muller J, Aeschbacher RA, Wingler A, Boller T, Wiemken A (2001) Trehalose and trehalase in *Arabidopsis*. Plant Physiol 125: 1086–1093.

47. Uhland K, Mondigler M, Spiess C, Prinz W, Ehrmann M (2000) Determinants of translocation and folding of TreF, a trehalase of *Escherichia coli*. J Biol Chem 275: 23439–23445.

48. Moustafa A, Reyes-Prieto A, Bhattacharya D (2008) Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One 3: e2205.

49. Brinkman FS, Blanchard JL, Cherkasov A, Av-Gay Y, Brunham RC, et al. (2002) Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. Genome Res 12: 1159–1167.

50. Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55: 709–742.

51. Lee SH, Stephens JL, Englund PT (2007) A fatty-acid synthesis mechanism specialized for parasitism. Nat Rev Microbiol 5: 287–297.

52. Lee SH, Stephens JL, Paul KS, Englund PT (2006) Fatty acid synthesis by elongases in trypanosomes. Cell 126: 691–699.

53. Whitaker JW, McConkey GA, Westhead DR (2009) Prediction of horizontal gene transfers in eukaryotes: approaches and challenges. Biochem Soc Trans 37: 792–795.

54. Whitaker JW, Letunic I, McConkey GA, Westhead DR (2009) metaTIGER: a metabolic evolution resource. Nucleic Acids Res 37: D531–538.

55. Templeton TJ, Lancto CA, Vigdorovich V, Liu C, London NR, et al. (2004) The *Cryptosporidium* oocyst wall protein is a member of a multigene family and has a homolog in *Toxoplasma*. Infect Immun 72: 980–987.

56. Abrahamsen MS, Schroeder AA (1999) Characterization of intracellular *Cryptosporidium parvum* gene expression. Mol Biochem Parasitol 104: 141–146.

57. Allocco JJ, Profous-Juchelka H, Myers RW, Nare B, Schmatz DM (1999) Biosynthesis and catabolism of mannitol is developmentally regulated in the protozoan parasite *Eimeria tenella*. J Parasitol 85: 167–173.

58. Aurrecoechea C, Brestelli J, Brunk BP, Fischer S, Gajria B, et al. (2010) EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res 38: D415–419.

59. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

60. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36: D480–484.

61. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

62. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

63. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.

64. Maruyama S, Matsuzaki M, Kuroiwa H, Miyagishima SY, Tanaka K, et al. (2008) Centromere structures highlighted by the 100%-complete *Cyanidioschyzon merolae* genome. Plant Signal Behav 3: 140–141.

65. Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 428: 653–657.

66. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

67. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4: 18.

68. Cai X, Woods KM, Upton SJ, Zhu G (2005) Application of quantitative real-time reverse transcription-PCR in assessing drug efficacy against the intracellular pathogen *Cryptosporidium parvum* in vitro. Antimicrob Agents Chemother 49: 4437–4442.

69. Rider SD, Jr., Zhu G (2008) Differential expression of the two distinct replication protein A subunits from *Cryptosporidium parvum*. J Cell Biochem 104: 2207–2216.

70. Zeng B, Cai X, Zhu G (2006) Functional characterization of a fatty acyl-CoA-binding protein (ACBP) from the apicomplexan *Cryptosporidium parvum*. Microbiology 152: 2355–2363.