AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Ecological and Evolutionary Science

Metabolic Reconstruction Elucidates the Lifestyle of the Last Diplomonadida Common Ancestor

Alejandro Jiménez-González,[a]* Jan O. Andersson[a]

[a]Uppsala Biomedicine Centre, Department of Cell and Molecular Biology, Molecular Evolution program, Uppsala University, Uppsala, Sweden

ABSTRACT The identification of ancestral traits is essential to understanding the evolution of any group. In the case of parasitic groups, this helps us understand the adaptation to this lifestyle and a particular host. Most diplomonads are parasites, but there are free-living members of the group nested among the host-associated diplomonads. Furthermore, most of the close relatives within Fornicata are free-living organisms. This leaves the lifestyle of the ancestor unclear. Here, we present metabolic maps of four different diplomonad species. We identified 853 metabolic reactions and 147 pathways present in at least one of the analyzed diplomonads. Our study suggests that diplomonads represent a metabolically diverse group in which differences correlate with different environments (e.g., the detoxification of arsenic). Using a parsimonious analysis, we also provide a description of the putative metabolism of the last Diplomonadida common ancestor. Our results show that the acquisition and loss of reactions have shaped metabolism since this common ancestor. There is a net loss of reaction in all branches leading to parasitic diplomonads, suggesting an ongoing reduction in the metabolic capacity. Important traits present in host-associated diplomonads (e.g., virulence factors and the synthesis of UDP-*N*-acetyl-D-galactosamine) are shared with free-living relatives. The last Diplomonadida common ancestor most likely already had acquired important enzymes for the salvage of nucleotides and had a reduced capacity to synthesize nucleotides, lipids, and amino acids *de novo*, suggesting that it was an obligate host-associated organism.

IMPORTANCE Diplomonads are a group of microbial eukaryotes found in oxygen-poor environments. There are both parasitic (e.g., *Giardia intestinalis*) and free-living (e.g., *Trepomonas*) members in the group. Diplomonads are well known for their anaerobic metabolism, which has been studied for many years. Here, we reconstructed whole metabolic networks of four extant diplomonad species as well as their ancestors, using a bioinformatics approach. We show that the metabolism within the group is under constant change throughout evolutionary time, in response to the environments that the different lineages explore. Both gene losses and gains are responsible for the adaptation processes. Interestingly, it appears that the last Diplomonadida common ancestor had a metabolism that is more similar to extant parasitic than free-living diplomonads. This suggests that the host-associated lifestyle of parasitic diplomonads, such as the human parasite *G. intestinalis*, is an old evolutionary adaptation.

KEYWORDS ancestral reconstruction, lateral gene transfer, metabolism, parasites, protists

The identification of the metabolic capacities of any species or group is an important task to understand the adaptation to the environment. For example, it could help to elucidate the interactions with other species *in vivo* and to identify the growth requirement for those organisms that fail to grow in axenic cultures (1, 2). The comparison of the metabolism of different species provides essential data to predict the

Citation Jiménez-González A, Andersson JO. 2020. Metabolic reconstruction elucidates the lifestyle of the last Diplomonadida common ancestor. mSystems 5:e00774-20. https://doi.org/10.1128/mSystems.00774-20.

Editor Holly L. Lutz, University of California, San Diego

Copyright © 2020 Jiménez-González and Andersson. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Jan O. Andersson, jan.andersson@icm.uu.se.

* Present address: Alejandro Jiménez-González, Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Vestec, Czechia.

Received 6 August 2020
Accepted 30 November 2020
Published 22 December 2020

metabolism of their common ancestor (3). The description of the ancestral state of a group helps to understand how the different lifestyles evolved within the group.

Diplomonads are an example of a group that is suitable to study lifestyle transitions because they contain both host-associated and free-living species (4). They are flagellated protists found in low-oxygen environments and classified within the group Fornicata (Metamonada) (5). *Giardia intestinalis*, the causative agent of giardiasis in humans, is the most widely known diplomonad. This organism has been used as a model to understand the evolution of parasitism and the reduction of the mitochondria (6–11). However, members of diplomonads can also colonize other mammals as well as fish, amphibians, and birds (4). Studies of the fish parasite *Spironucleus salmonicida* have deepened the understanding of parasitism in diplomonads (6, 12, 13). *Trepomonas* sp. strain PC1 has been described as a secondary free-living organism because its ancestor escaped a parasitic lifestyle thanks to the acquisition, from bacteria, of many genes associated with its free-living lifestyle (14). Recently, new genomes and transcriptomes have been published from close relatives of diplomonads (6, 15). Interestingly, the closest relatives of diplomonads among these are free-living (6, 16), raising the question of whether the last Diplomonadida common ancestor was already a parasite or if the transition to parasitism occurred multiple times within the group.

A comparison of the metabolic capacities of extant diplomonads could shed light on this question. The metabolism of various diplomonads has indeed been studied, both with experimental and bioinformatic approaches (12, 14, 17–21). Here, we present a systematic comparative reanalysis of the metabolic capacities of four diplomonad species. We show that traits associated with a host-associated lifestyle were present already in the last Diplomonadida common ancestor.

RESULTS

We manually curated the annotations of genes coding for metabolic reactions in four genomes and one transcriptome representing four diplomonad species using a number of prediction tools and databases (see Materials and Methods for details). In total, we identified 853 reactions (see Table S1 in the supplemental material) and 147 pathways present in at least one of the analyzed diplomonads. Among these, 559 reactions (66%) and 82 pathways (56%) were common to all the analyzed diplomonads, while 101 reactions (12%) and 22 pathways (15%) were unique to one diplomonad (Fig. 1). *Trepomonas* sp. strain PC1 showed the most complex metabolism, with 764 reactions and 118 pathways, while *G. muris* showed the most reduced metabolism, with 669 reactions and 95 pathways (Fig. 2).

We constructed clusters of genes based on their functional annotation. The evolutionary origin of the genes in the clusters was predicted based on the identity of the most similar sequences (see Materials and Methods for details and Table S1). The putative metabolic capacities of the last Diplomonadida common ancestor and the Giardiinae and Hexamitinae ancestors were reconstructed using this approach (Fig. 2 and Fig. S1 to S6).

Last Diplomonadida common ancestor. The last Diplomonadida common ancestor encoded 702 of the 853 reactions and 104 of the 147 pathways annotated in the extant species (Fig. 2). The predicted overall metabolism was, as expected, similar to that of extant diplomonads (Fig. S1). For example, the last Diplomonadida common ancestor appeared to have produced pyruvate via glycolysis. Pyruvate was converted to acetyl-coenzyme A (CoA) via pyruvate:ferredoxin oxidoreductase (Fig. S1). This ancestor could produce pyruvate either via pyruvate kinase or via the most efficient pyrophosphate-dependent pyruvate phosphate dikinase enzyme, similar to extant *G. intestinalis* and *Trepomonas* sp. strain PC1. The Diplomonadida ancestor could interconvert phosphoenolpyruvate and oxaloacetate via phosphoenolpyruvate carboxykinase, which allowed it to adjust its metabolism depending on the environmental conditions, similar to Giardiinae species (Fig. S1). The pentose phosphate pathway appears to be

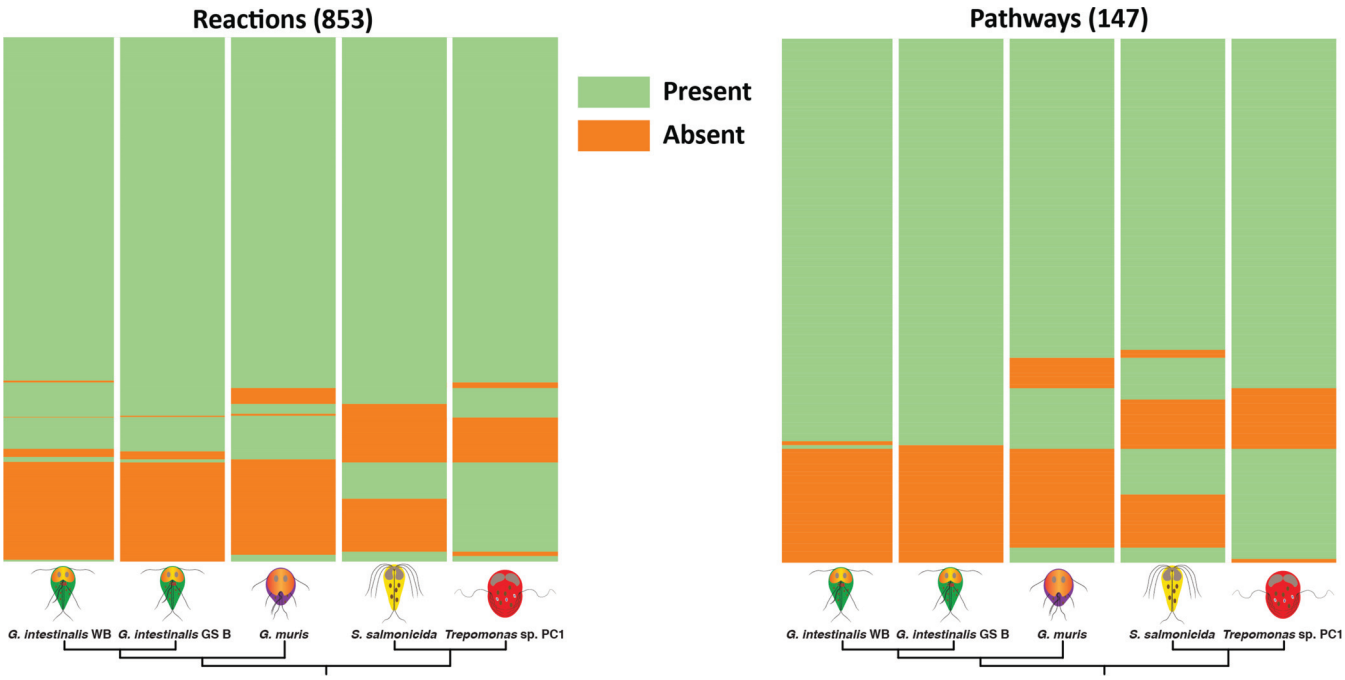

FIG 1 Heatmaps of reactions (left) and pathways (right). Reactions and pathways present and absent in a diplomonad are in green and orange, respectively.

partial and use ribulose 5-phosphate or erythrose 4-phosphate and xylulose 5-phosphate to produce glyceraldehyde 3-phosphate. This ancestor could synthesize UDP-*N*-acetyl-D-galactosamine, a compound that is essential for building the cyst wall, together with three cyst wall proteins in *Giardia* species and probably *S. salmonicida* (12, 22). The capacity for synthesizing UDP-*N*-acetyl-D-galactosamine was already present in

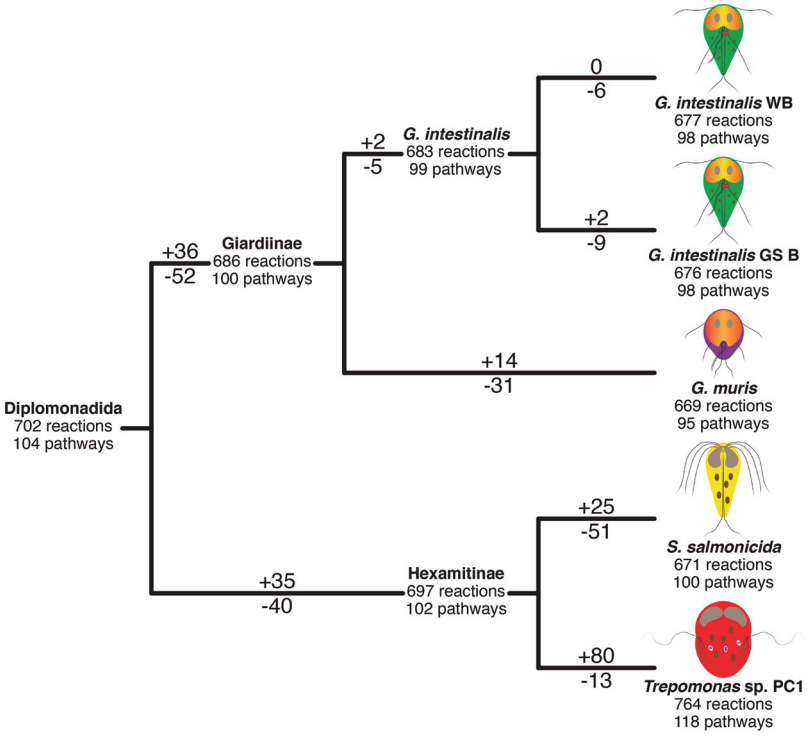

FIG 2 Total reactions and pathways identified in the studied diplomonads and the different ancestors. Positive and negative numbers indicate reactions gained and lost at that branch.

the common ancestor of diplomonads and *Kipferlia bialata*, suggesting that the Diplomonadida ancestor had the metabolic capacity to form cysts (Fig. S1).

Thirteen proteins have been suggested to be virulence factors in *G. intestinalis*. Most of these enzymes are proteases that disrupt the epithelial cells and the intestinal biofilm of the host (23). Our analysis suggested that most of these proteases were already present in the common ancestor of *K. bialata* and diplomonads (Fig. S1). Only two of the potential virulence factors were not shared with *K. bialata*: uridine phosphorylase and serine peptidase. However, both proteins were classified as vertically inherited candidates and probably were lost in *K. bialata*. These observations suggest that the candidate virulence factors evolved in a free-living ancestor of diplomonad parasites, indicating that they are not parasite-specific inventions.

Members of Metamonada that are host associated, such as *G. intestinalis*, *S. salmonicida*, *Trichomonas vaginalis*, and *Monocercomonoides exilis*, have been shown to have a metabolism that is dependent on the supply of metabolites from the environment within the host (12, 24–26). Our analyses indeed suggested that the Diplomonadida ancestor had a limited capacity to synthesize lipids, amino acids, and nucleotides *de novo* (Fig. S1), in agreement with a host-associated lifestyle.

Several transporters of lipids were present, and only pathways for lipid modification were identified, suggesting that it depended on external sources (Fig. S1). The ancestor most likely could utilize arginine, tryptophan, and serine available in the environment as sources of energy. Whereas the arginine dihydrolase pathway is widespread in Metamonada, the capacity for degradation of tryptophan and serine was acquired from bacteria in the lineage leading to the Diplomonadida ancestor.

Similarly, the Diplomonadida ancestor had several pathways for the salvage and degradation of nucleotides and nucleosides and was likely dependent on the salvage pathways to ascertain the availability of nucleotides and deoxynucleotides (Fig. 3 and Fig. S1). All nucleotide salvage pathways appear to have been acquired since the last eukaryotic ancestor. Our analyses classified all key enzymes related to salvage and degradation of purines as lateral gene transfer (LGT) candidates acquired after the divergence from *K. bialata* (Fig. 3 and Fig. S1), whereas the key enzymes for the salvage and degradation of pyrimidines shared an origin with *K. bialata*. This suggests that the Diplomonadida ancestor could salvage nucleosides and convert them into all needed nucleotides (Fig. 3). The enzyme ribonucleotide reductase synthesizes deoxynucleotides from nucleotides. All organisms, except a few parasites, encode this enzyme (27). It was previously shown that anaerobic ribonucleoside-triphosphate reductase is present in some parasitic diplomonads but absent from *G. intestinalis* and *S. salmonicida* (9, 14). Our analysis indicated that the Diplomonadida ancestor lacked the enzyme. A phylogenetic analysis suggested that a bacterial anaerobic ribonucleoside-triphosphate reductase was acquired before the split with *K. bialata* and was subsequently lost in the lineage leading to the last Diplomonadida common ancestor (Fig. 3 and Fig. S7). The enzyme was regained in the Hexamitinae ancestor, lost in several lineages, and gained a third time via LGT by *Trepomonas* sp. strain PC1 (14). Thus, it appears that the last Diplomonadida common ancestor was dependent on a source of deoxynucleotides for the synthesis of DNA, similar to *Giardia* and *S. salmonicida*.

Giardiinae ancestor, *Giardia intestinalis*, and *Giardia muris*. All members of Giardiinae are parasites, and a reduction of the metabolic capacity could be expected within the group. Fifty-two reactions were indeed lost in the lineage from the last Diplomonadida ancestor to the Giardiinae ancestor (Fig. 2 and Fig. S2A). However, 36 reactions were also gained, suggesting an evolutionary flexibility of the metabolic capacities (Fig. 2 and Fig. S2B). The nucleotide metabolism was affected by the loss of the last step for the degradation of adenosine, guanosine, and inosine-5-phosphate. However, uracil phosphoribosyltransferase was gained in the lineage leading to this ancestor, independent of the gain event in *S. salmonicida* (Fig. 3). This acquisition expanded the possibilities for nucleotide salvage. The energy metabolism was also affected because the ability to degrade galactose and triacylglycerols was lost (Fig. S2A). Our analyses suggested

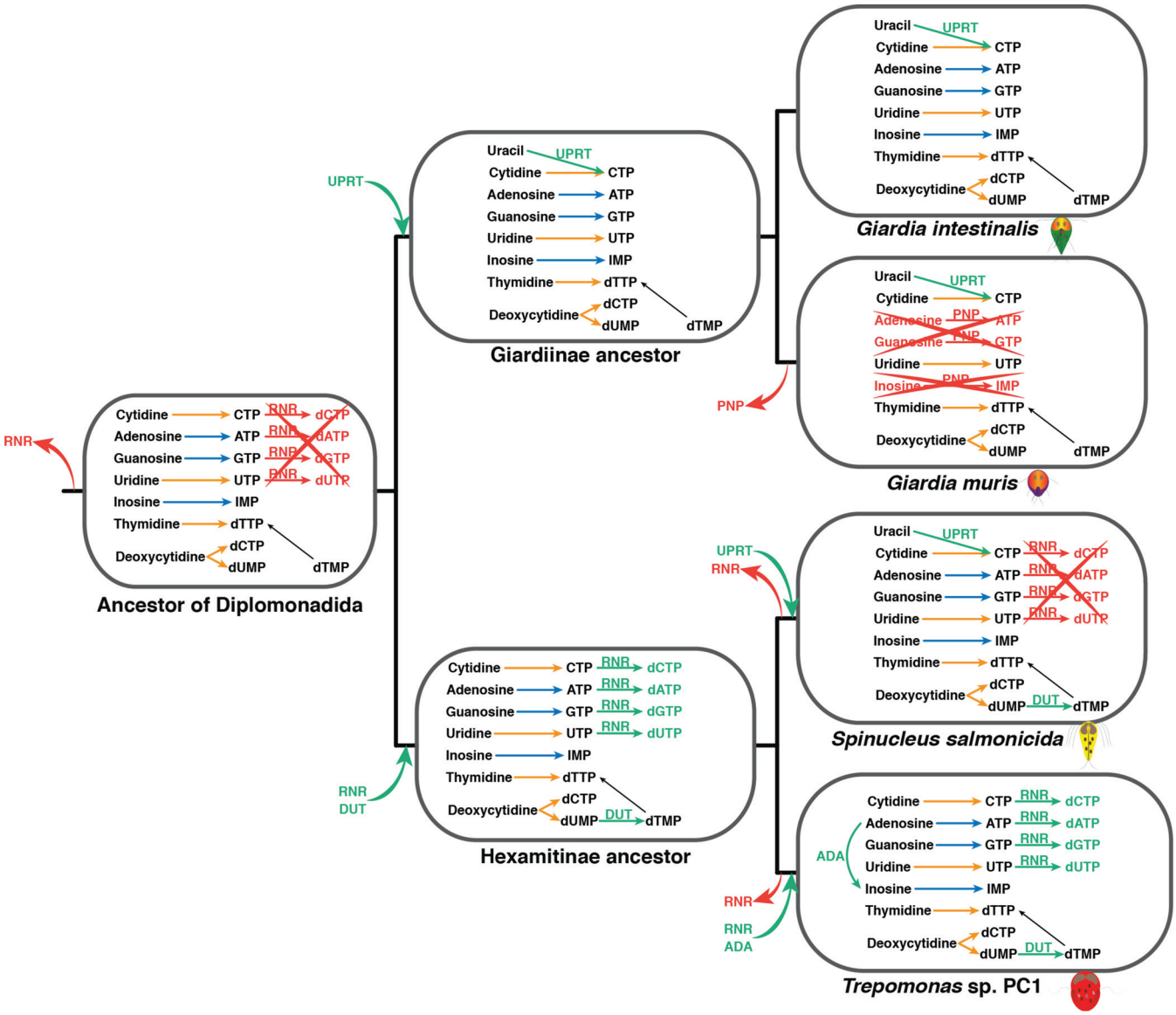

FIG 3 Evolution of the synthesis of nucleotides and nucleosides. Key reactions vertically inherited are in black. Pathways and reactions that clustered with *K. bialata* and were classified as LGT candidates are in orange. Pathways and reactions that did not cluster with *K. bialata* and were classified as LGT candidates are in blue. Pathways and reactions classified as LGT after the last Diplomonadida common ancestor are in green. Pathways and reactions lost are in red. RNR, anaerobic ribonucleoside-triphosphate reductase; UPRT, uracil phosphoribosyltransferase; PNP, purine nucleoside phosphorylase; DUT, deoxyuridine 5′-triphosphate nucleotidohydrolase; ADA, adenosine deaminase.

that the Giardiinae ancestor acquired the oxidative branch of the pentose phosphate pathway (Fig. S2B). This pathway synthesizes ribulose 5-phosphate and NADH from glucose 6-phosphate. The NADH is used in different reactions, including oxygen detoxification. The ribulose 5-phosphate is used to synthesize glyceraldehyde 3-phosphate through the partial pentose phosphate pathway already present in the last Diplomonadida common ancestor. The glyceraldehyde 3-phosphate can be used to synthesize pyruvate in the last steps of glycolysis. Interestingly, our analyses suggested that the Giardiinae ancestor lost the transporters of glyceraldehyde 3-phosphate. This loss could be related to the acquisition of this oxidative branch. This ancestor could take glucose from glycogen, a trait shared with *Trepomonas* sp. strain PC1. However, the genes in the two diplomonad lineages have independent LGT origins.

Only two reactions were classified as LGT candidates in the *G. intestinalis* ancestor since the Giardiinae ancestor, MsrA and flavohemoprotein, both of which have been

reported before (28) (Fig. 2 and Table S2). Our analyses identified that the *G. intestinalis* ancestor lost five reactions (Fig. 2 and Table S2). This ancestor lacked the protein tryptophanase, which degrades tryptophan to pyruvate, indole, and ammonium. While pyruvate is used to produce energy, indole could interact with the microbiota of the host, interfering with the quorum sensing (29). Our analyses also detected that the *G. intestinalis* ancestor lacked an amino acid transporter. We could not identify the nature of this transporter, but its loss could be connected with the loss of the tryptophanase activity.

Our analysis showed that *G. intestinalis* WB and *G. intestinalis* GS B are not metabolically identical (Fig. 1). We identified six reactions lost in *G. intestinalis* WB and nine lost in *G. intestinalis* GS B (Fig. 2 and Table S2). At the same time, we classified two reactions as LGT candidates in *G. intestinalis* GS B since the *G. intestinalis* common ancestor and none in *G. intestinalis* WB (Fig. 2 and Table S2). These differences made these two isolates metabolically distinct. *Giardia intestinalis* WB lost the ability to degrade arginine to L-ornithine and urea, making the deimination of arginine to L-citrulline the only option to degrade this amino acid. *Giardia intestinalis* GS B shared the ability to degrade arginine through two different pathways with *G. muris* (30). On the other hand, *G. intestinalis* GS B lost the synthesis of glycine from glyoxylate.

Another significant difference between the two *G. intestinalis* isolates is the absence of the protein quorum-quenching *N*-acyl-homoserine lactonase in *G. intestinalis* WB (Table S2). This protein interferes with the quorum sensing of different bacteria and was previously reported as laterally acquired from bacteria in the Giardiinae ancestor (30).

Giardia muris showed the most reduced metabolism among the analyzed diplomonads (Fig. 2). *Giardia muris* lacked most of the pathways related to the salvage and degradation of nucleosides and nucleotides (Fig. S3A). Only the salvage and degradation of pyrimidines are retained (Fig. 3). The absence of those pathways suggested that *G. muris* is highly dependent on a supply of nucleosides and nucleotides, especially purines, from the environment within the host. On the other hand, *G. muris* can synthesize coenzyme A *de novo* (30). Our analysis suggested that the genes responsible for the synthesis of coenzyme A in *G. muris* were acquired independently in the lineages leading to *G. muris* and *S. salmonicida* (Fig. S3B and Table S1)

Hexamitinae ancestor, *Spironucleus salmonicida*, and *Trepomonas* sp. strain PC1. The lineage from the last Diplomonadida common ancestor to the last Hexamitinae common ancestor has experienced 35 gains and 40 losses of reactions, numbers that are similar to those of the Giardiinae lineage (Fig. 2). Among the losses are the ability to interconvert phosphoenolpyruvate and oxaloacetate via phosphoenolpyruvate carboxykinase and the degradation of arginine through the enzyme arginase (Fig. S4A). This ancestor also lost the synthesis of farnesyl and geranyl diphosphate from isopentenyl pyrophosphate. These losses most likely are related to the absence of the mevalonate pathway, an intermediary pathway for the synthesis of both compounds. In contrast, the capacity for nucleotide metabolism was extended. The enzyme anaerobic ribonucleoside-triphosphate reductase was acquired via LGT (Fig. 3 and Fig. S4B and S7), allowing the organism to synthesize deoxynucleotide triphosphates (dNTPs) from NTPs. Our analysis also suggested that this ancestor acquired the enzyme deoxyuridine 5′-triphosphate nucleotidohydrolase, which converts dUMP into dTMP, which later is converted to dTTP (Fig. 3 and Fig. S4B). Several enzymes classified as LGT candidates are related to the degradation of sugars and proteins (Fig. S4B), suggesting an adaptation to use compounds that this ancestor could find within the host.

Both analyzed Hexamitinae species showed more pathways than any of the Giardiinae species, although the number of reactions was slightly lower in *S. salmonicida* than *G. intestinalis* (Fig. 2). Notable losses in *S. salmonicida* are the lack of anaerobic ribonucleoside-triphosphate reductase, enzymes for degradation of adenosine, guanosine, and inosine 5-phosphate to urate, and enzymes to repair NADHX (Fig. 4A and Fig. S5A). NADH can be hydrated into NADHX due to the action of some

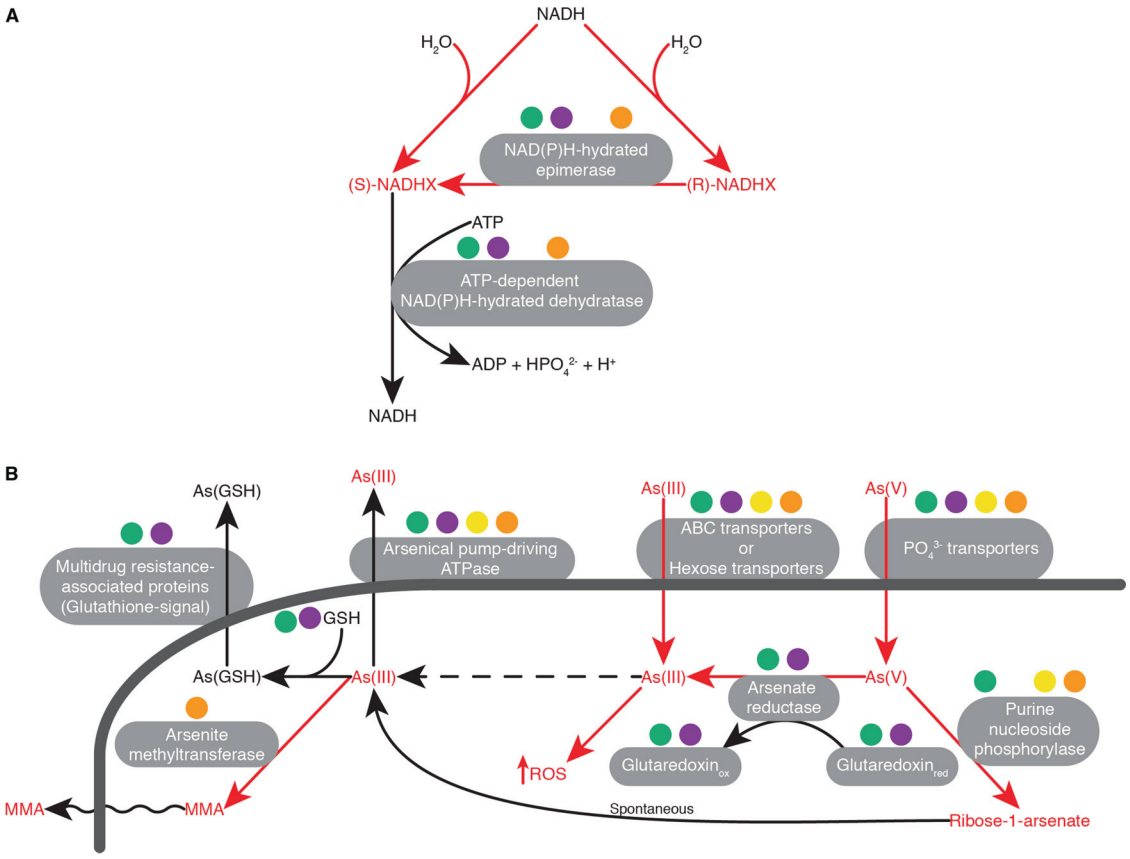

FIG 4 Distribution of proteins across diplomonads for the NADH repair pathway (A) and the arsenic detoxification pathway (B). The presence of proteins and molecules in the studied species are indicated by filled circles: *G. intestinalis* (green), *G. muris* (purple), *S. salmonicida* (yellow), and *Trepomonas* sp. strain PC1 (orange). Toxic compounds and reactions for the cell are in red. NADHX, NADH hydrated; As(V), arsenate; As(III), arsenite; GSH, glutathione; As(GSH), glutathione-chelated-arsenic; MMA, monomethylarsonic acid.

dehydrogenases or spontaneously produced under acidic conditions. NADHX is a toxic compound that inhibits several dehydrogenases (31), and the lack of repair enzymes likely makes *S. salmonicida* sensitive to this compound.

Our analyses found differences in the arsenic detoxification pathway within diplomonads. Arsenic has two biologically relevant oxidation states, arsenite [As(III)] and arsenate [As(V)] (32). Both states are toxic to the cell. While As(III) interacts with thiols, increasing the reactive oxygen species (ROS), As(V) replaces phosphate-producing unstable arsenical compounds (33). For these reasons, all species have mechanisms to avoid the damage of these elements (33, 34). Our analyses suggested that Giardiinae species use glutathione for the detoxification of As(III) and As(V) and *Trepomonas* sp. strain PC1 methylates As(III), making it volatile (Fig. 4B). Curiously, we identified the arsenical pump-driving ATPase as the only direct protein responsible for the detoxification of arsenic in *S. salmonicida* (Fig. 4B). This lack of mechanisms suggested that *S. salmonicida* is very sensitive to stress by arsenic.

Our analyses classified several important proteins as LGT candidates in *S. salmonicida* since the divergence from the lineage leading to *Trepomonas* (Fig. 2 and Table S1). One of them is the protein uracil phosphoribosyltransferase. This protein converts uracil into UMP, making *S. salmonicida* more versatile in the salvage of pyrimidines (Fig. 3 and Fig. S5B). Interestingly, mannose-6-phosphate isomerase was acquired three times independently within diplomonads, making *G. muris*, *Trepomonas* sp. strain PC1, and *S. salmonicida* able to utilize mannose (Table S1). The capacity for melibiose degradation showed a similar

pattern. Our analyses suggested that this reaction was acquired independently in *S. salmonicida* and *Trepomonas* sp. strain PC1 (Table S1).

Trepomonas sp. strain PC1, the only transcriptome within this study, showed the most complex metabolism (Fig. 2). Although the transcriptome appeared to be fairly complete, the 13 losses need to be considered carefully (Fig. S6A). The 80 gains, on the other hand, should be treated as true gains and not putative contaminations, because the transcriptome data have been carefully curated using the fact that Hexamitinae species utilize an alternative genetic code (14).

One of the most significant metabolic changes in *Trepomonas* sp. strain PC1 was related to the synthesis and degradation of nucleosides and nucleotides (Fig. 3 and Fig. S6B). The acquisition of the protein anaerobic ribonucleoside-triphosphate reductase was a clear advantage for *Trepomonas* sp. strain PC1 (14). This protein uses formate and NTPs for the synthesis of dNTPs (Fig. S6B). Our analysis suggested that genes of *Trepomonas* sp. strain PC1 encoded a formate C-acetyltransferase that converts pyruvate into acetyl-CoA and formate. However, a deeper analysis of the sequence revealed that this protein most likely is a 4-hydroxyphenylacetate decarboxylase (35).

We identified the acquisition of a number of reactions related to nucleotide metabolism and degradation of larger molecules, which most likely played important roles in the development of the secondary free-living lifestyle of *Trepomonas* sp. strain PC1, as previously reported (14). The enzymes adenosine deaminase and dihydropyrimidinase made *Trepomonas* less dependent on the salvage of specific intermediates in the nucleotide metabolism, whereas the acquisitions of α-galactosidase, β-D-glucoside glucohydrolase, β-1,2-mannosidase, glucoamylase, and endoglycosylceramidase made it able to utilize a wide range of carbohydrates putatively available for a free-living bacterivore (Fig. S6B). The acquisitions of the enzymes peptidoglycan DL-endopeptidase, *N*-acetylmuramic-L-alanine amidase, and *N*-acetylmuramic acid 6-phosphate etherase were likely directly related to the ability to digest bacteria (14) (Fig. S6B).

DISCUSSION

We have combined several metabolic prediction tools and created manually curated metabolic databases of four diplomonads. This approach allows us to minimize the errors due to misannotations and to present the most complete metabolic study in diplomonads to date (Fig. 2 and Fig. S1 to S6). We mapped the gains and losses of functions on the phylogenetic tree and show that all branches leading to parasitic diplomonads show a net loss of reactions, suggesting an ongoing reduction in the metabolic capacity (Fig. 2). Even though this reduction is not very pronounced, it is most likely a consequence of a parasitic or host-associated lifestyle (36).

Diplomonads, a metabolically diverse group. Our metabolic analysis shows that diplomonads are a metabolically diverse group of organisms. We identified differences between species that indicate that they are adapted to the environments where they are found. For example, *Spironucleus salmonicida* degrades melibiose into galactose and glucose, a capacity shared with *Trepomonas* sp. strain PC1 (Fig. S5B and S6B). However, our analysis suggested that this represents a convergence. Most likely, both species acquired this capacity via LGT independently. Cases of convergence via LGT have been reported before in diplomonads and other eukaryotes (28, 37). This activity is absent in *Giardia intestinalis* and *Giardia muris*, indicating either that no successful LGT of genes needed to degrade melibiose has occurred yet in these lineages or that this compound is less frequent in the intestinal tract of their hosts.

Detoxification of arsenic is a second example of the metabolic diversity in diplomonads (Fig. 4B). Giardiinae species can degrade both As(V) and As(III), suggesting that both elements are present at dangerous levels in the intestinal tract of their host. Our analysis shows that this detoxification most likely is dependent on glutathione, a thiol absent in *S. salmonicida* and *Trepomonas* sp. strain PC1. In contrast, Hexamitinae species can only detoxify As(III). As a free-living organism, *Trepomonas* sp. strain PC1 most likely is able to escape if it encounters dangerous levels of As(V). Our analyses showed

that *S. salmonicida* also could be sensitive to stress by NADHX. Interestingly, NADHX and As(V) are more abundant in acidic environments (31–34), suggesting that *S. salmonicida* has a preference for neutral or alkaline environments (e.g., pH in the fish blood is between 7.4 and 7.5 [38]).

The last Diplomonadida common ancestor. We need to trace the origin of traits associated with parasitism and host association in diplomonads to understand if the last common Diplomonadida ancestor already was a host-associated organism. Previous studies have shown that parasitic diplomonads avoid the immune system of the host via the expression of cysteine-rich proteins (12, 39, 40). These proteins have been studied functionally in *G. intestinalis*, where they are present as a large protein family that functions as variant-specific surface proteins (VSPs) exposed on the surface of the trophozoite (22, 25, 41). A large family of cysteine-rich proteins with a domain structure similar to that of VSPs was described in *S. salmonicida*. Even though the proteins are highly divergent between the two parasitic diplomonads, their structural similarities suggest that they have a common origin (12). The draft genome of *Kipferlia bialata* did not contain any similar cysteine-rich proteins (15), indicating that this protein family represents a parasitic innovation present in the last Diplomonadida ancestor.

We have recently shown that the last Diplomonadida common ancestor was well adapted to low-oxygen environments (28). Even though this characteristic evolved in a free-living lifestyle, being able to survive in low-oxygen environments is a requirement to colonize the intestinal tract of the host. Similarly, there is a set of putative virulence factors identified in *G. intestinalis* (23, 42–44). Here, we show that most of these were present in free-living relatives of diplomonads, and they should be viewed as a preparasitic function and not as a parasitic innovation (45). The formation of cysts is also an important trait in host-associated diplomonads. The pathway to synthesize the cyst wall is shared between diplomonads. Experimental tests have shown that cyst wall proteins from *S. salmonicida* are functional in *G. intestinalis*, suggesting that the ancestor had a similar life cycle (12). The formation of the cyst wall also requires UDP-*N*-acetyl-D-galactosamine (10). Here, we show that the synthesis of this sugar was present already in the last common ancestor of *K. bialata* and diplomonads.

Our analyses show that the last Diplomonadida common ancestor shared important traits with free-living relatives, and several enzymes have been classified as acquired via LGT. However, the metabolic reconstruction suggests that this organism had an overall reduced metabolism (Fig. S1). It likely lacked pathways for the synthesis of essential cellular components, like lipids and most amino acids. Our analyses identified several transporters of both compounds and only pathways for lipid modification. This ancestor also had a reduced nucleotide and nucleoside metabolism and likely lacked the capacity for *de novo* synthesis of dNTPs. Instead, it had acquired the ability to salvage nucleotides and nucleosides. The genes for the enzymes responsible for these reactions were acquired most likely via LGT from bacterial donors. The acquisition via LGT of these reactions is not exclusive of diplomonads. The human parasite *Cryptosporidium parvum* also acquired several enzymes involved in the salvage of nucleotides and nucleosides from bacteria via LGT (46). Studies of the timing of the eukaryotic diversification show that the last common Diplomondida ancestor probably coexisted with bilateral animals (the most probable host) (47). Taken together, all the present and absent traits in the last Diplomonadida common ancestor strongly suggest that this ancestor was already an obligate host-associated organism, if not already a parasite of some animal.

MATERIALS AND METHODS

Sources of data. Protein sequences from genome data sets of *Giardia intestinalis* WB, *G. intestinalis* GS B, *G. muris*, *Spironucleus salmonicida*, and *Kipferlia bialata* and the transcriptome data set of *Trepomonas* sp. strain PC1 were downloaded from NCBI (48).

Identification of metabolic capacities. The metabolic capacities of each genome and transcriptome were predicted with the GhostKOALA tool (genus_prokaryotes + family_eukaryotes + viruses database) implemented in KEGG (49), EggNOG-mapper (DIAMOND mapping mode) (50), and PathwayTool v. 21.5 (default setup) (51). Every genome and transcriptome was manually curated, combining the

prediction of these three pieces of software under the Pathway-Tool framework (52). In the case of *G. intestinalis* and *S. salmonicida*, the information contained in GiardiaDB (53) was also added at this step.

Each curated database was improved using the Pathway Hole Filler implemented in Pathway-Tool (54) using the different diplomonad databases as training data sets. When possible, the function for transporters was assigned using Transport Inference Parser (55) implemented in Pathway-Tool and verified with the Conserved Domain Database (56).

Clustering analysis. Reactions with at least one protein assigned in one curated database were retrieved from Pathway-Tool. Reactions that were predicted to be present in a database but no enzyme could be assigned were not considered for this analysis (i.e., a gap in a pathway). All proteins from the databases of the different organisms were combined in the same data set when they catalyze the same reaction, creating a reaction data set. Every protein in the reaction data set was used as a query in BLASTp searches against a custom protein sequence database made of the diplomonads used in this analysis, the *K. bialata* (15) proteins, and the UniRef90 database (July 2019) (57). For every query, 500 hits were kept with E values of $\leq$1e$-$10. For every query, the first 100 hits were extracted for posterior analyses. We tested different numbers of extracted top hits and found that the first 100 hits gave congruent results with most of the previous phylogenetic analyses performed in studies of diplomonad proteins (28, 30).

A pair of diplomonad sequences were considered to cluster together (i.e., having a common evolutionary origin) if they reciprocally were found among the first 100 hits in the respective BLASTp search. Similarly, a diplomonad sequence was considered to cluster with *K. bialata* if a sequence from that species was among the first 100 hits.

Classification of the reactions and construction of the last Diplomonadida common ancestor. Based on the BLASTp searches, we used a parsimonious approach to classify the reactions. A reaction was considered potentially present in the last Diplomonadida common ancestor (ancestral) if it was present in all diplomonad species and they clustered together, the reaction was missing from at least one of the diplomonad lineages but the proteins clustered with *K. bialata*, or, in the absence of clustering with *K. bialata*, the majority of hits were eukaryotic homologs. A reaction was classified as potentially gained in the lineages leading to the Giardiinae, *G. intestinalis*, and/or Hexamitiae ancestors if the proteins from each lineage cluster independently of each other and independently of *K. bialata* and the majority of hits were from prokaryote homologs. Any reaction with a majority of hits from prokaryotes was considered an LGT candidate. Reactions whose evolutionary history has been described previously were manually curated to be consistent (14, 28, 30). Classifications of all reactions are listed in Table S1 in the supplemental material.

The pathways in the different ancestors were predicted under Pathway-Tool v. 21.5 based on the reactions classified to be present in that particular ancestor. The pathways were manually curated and kept or removed based on the number of gaps and taxonomic distribution of the pathway.

Phylogenetic analysis of anaerobic ribonucleoside-triphosphate reductase. Nucleotide sequences from the transcriptome data sets of the nondiplomonad Fornicata species *Aduncisulcus paluster*, *Carpediemonas membranifera*, *Chilomastix caulleryi*, *Chilomastix cuspidata*, *Dysnectes brevis*, and *Ergobibamus cyprinoides* (6) were downloaded from the Dryad Digital Repository. tBLASTp searches against these transcriptomes were made using *Trepomonas* sp. strain PC1 and *K. bialata* anaerobic ribonucleoside-triphosphate reductase as a query. The obtained hits were translated into amino acid sequences using EMBOSS Sixpack (58) and evaluated using the Conserved Domain Database. Previously used sequences from *S. barkhanus* and *S. vortens* were included in this analysis (14). The result of these procedures was the creation of a curated Metamonada anaerobic ribonucleoside-triphosphate reductase database.

We performed a phylogenetic analysis by following the approach previously described (28). One sequence from *Trepomonas* sp. strain PC1 and one sequence from *K. bialata* homologs were used as a BLASTp query against the NCBI nr database (October 2019). In this case, the optimal number of hits was 10,000. The number of hits in common between both BLASTp searches was calculated with CD-HIT-2D (59) with the default settings. In this case, the proportion of hits in common was 70%, and both BLASTp searches were merged into a single diversity matrix. This matrix was filtered using CD-HIT (59) by keeping only sequences with $<$90% sequence identity to another sequence in the data set. This filtered matrix then was merged with homologous proteins from the curated Metamonada anaerobic ribonucleoside-triphosphate reductase database that we had created previously (described above) and aligned using MAFFT v6.603b (60) with the default settings. The resulting alignment was trimmed using BMGE v1.12 (BLOSUM30 with a block size of 2) (61). A preliminary phylogenetic tree was computed using FastTree v2.1.8 SSE3, with OpenMP (62) with default settings, and sequences with a phylogenetic distance of $<$0.3 were removed in an iterative process to further reduce the size of the matrices until the final matrix was generated.

The final matrix was aligned using MAFFT and trimmed using BMGE, as described above. Maximum likelihood trees were computed using IQtree v. 1.5.3 (63) under the LGX substitution model. Branch supports were assessed using ultrafast bootstrap approximation (UFboot) with 1,000 bootstrap replicates (64) and SH-like approximate likelihood ratio test (SH-aLRT) (65), for which 1,000 replicates were used.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.3 MB.

FIG S2, PDF file, 0.5 MB.

FIG S3, PDF file, 0.5 MB.

FIG S4, PDF file, 0.5 MB.

FIG S5, PDF file, 0.5 MB.
FIG S6, PDF file, 0.6 MB.
FIG S7, PDF file, 0.6 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

Phylogenetic analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

A.J.G. and J.O.A. conceived and designed the analyses; A.J.G. performed the analyses; J.O.A. supervised the analyses; A.J.G. and J.O.A. analyzed the data; and A.J.G. and J.O.A. wrote the paper.

REFERENCES

1. Hofreuter D. 2014. Defining the metabolic requirements for the growth and colonization apacity of *Campylobacter jejuni*. Front Cell Infect Microbiol 4:1–19. https://doi.org/10.3389/fcimb.2014.00137.
2. Nobu MK, Narihiro T, Rinke C, Kamagata Y, Tringe SG, Woyke T, Liu WT. 2015. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. ISME J 9:1710–1722. https://doi.org/10.1038/ismej.2014.256.
3. Gabaldón T, Huynen MA. 2003. Reconstruction of the proto-mitochondrial metabolism. Science 301:609. https://doi.org/10.1126/science.1085463.
4. Adam RD. 2017. Diplomonadida, p 1219–1246. In Archibald JM, Simpson AGB, Slamovits CH (ed), Handbook of the protists, 2nd ed. Springer, Cham, Switzerland.
5. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol 66:4–119. https://doi.org/10.1111/jeu.12691.
6. Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson JO, Xu F, Yabuki A, Eme L, Zhang Q, Takishita K, Inagaki Y, Simpson AGB, Hashimoto T, Roger AJ. 2017. Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes. Nat Ecol Evol 1:92. https://doi.org/10.1038/s41559-017-0092.
7. Andersson JO. 2012. Double peaks reveal rare diplomonad sex. Trends Parasitol 28:46–52. https://doi.org/10.1016/j.pt.2011.11.002.
8. Shiflett AM, Johnson PJ. 2010. Mitochondrion-related organelles in eukaryotic protists. Annu Rev Microbiol 64:409–429. https://doi.org/10.1146/annurev.micro.62.081307.162826.
9. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 317:1921–1926. https://doi.org/10.1126/science.1143837.
10. Adam RD. 2001. Biology of *Giardia lamblia*. Clin Microbiol Rev 14:447–469. https://doi.org/10.1128/CMR.14.3.447-475.2001.
11. Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, Van Der Giezen M, Hernández M, Müller M, Lucocq JM. 2003. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature 426:172–176. https://doi.org/10.1038/nature01945.
12. Xu F, Jerlström-Hultqvist J, Einarsson E, Ástvaldsson Á, Svärd SG, Andersson JO. 2014. The genome of *Spironucleus salmonicida* highlights a fish pathogen adapted to fluctuating environments. PLoS Genet 10:e1004053. https://doi.org/10.1371/journal.pgen.1004053.
13. Stairs CW, Kokla A, Ástvaldsson Á, Jerlström-Hultqvist J, Svärd S, Ettema TJG. 2019. Oxygen induces the expression of invasion and stress response genes in the anaerobic salmon parasite *Spironucleus salmonicida*. BMC Biol 17:19. https://doi.org/10.1186/s12915-019-0634-8.
14. Xu F, Jerlström-Hultqvist J, Kolisko M, Simpson AGB, Roger AJ, Svärd SG, Andersson JO. 2016. On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. BMC Biol 14:62. https://doi.org/10.1186/s12915-016-0284-z.
15. Tanifuji G, Takabayashi S, Kume K, Takagi M, Nakayama T, Kamikawa R, Inagaki Y, Hashimoto T. 2018. The draft genome of *Kipferlia bialata* reveals reductive genome evolution in fornicate parasites. PLoS One 13:e0194487. https://doi.org/10.1371/journal.pone.0194487.
16. Leiva L. 1921. Observations on *Chilomastix intestinalis* Kuczinski. J Parasitol 8:49–57. https://doi.org/10.2307/3270749.
17. Müller M. 1988. Energy metabolism of protozoa without mitochondria. Annu Rev Microbiol 42:465–488. https://doi.org/10.1146/annurev.mi.42.100188.002341.
18. Lindmark DG. 1980. Energy metabolism of the anaerobic protozoan *Giardia lamblia*. Mol Biochem Parasitol 1:1–12. https://doi.org/10.1016/0166-6851(80)90037-7.
19. Lloyd D, Williams CF. 2014. Comparative biochemistry of *Giardia*, *Hexamita* and *Spironucleus*: enigmatic diplomonads. Mol Biochem Parasitol 197:43–49. https://doi.org/10.1016/j.molbiopara.2014.10.002.
20. Biagini GA, Rutter AJ, Finlay BJ, Lloyd D. 1998. Lipids and lipid metabolism in the microaerobic free-living diplomonad *Hexamita* sp. Eur J Protistol 34:148–152. https://doi.org/10.1016/S0932-4739(98)80025-4.
21. Biagini GA, Yarlett N, Ball GE, Billetz AC, Lindmark DG, Martinez MP, Lloyd D, Edwards MR. 2003. Bacterial-like energy metabolism in the amitochondriate protozoon *Hexamita inflata*. Mol Biochem Parasitol 128:11–19. https://doi.org/10.1016/S0166-6851(03)00025-2.
22. Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K, Svärd SG. 2010. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat Rev Microbiol 8:413–422. https://doi.org/10.1038/nrmicro2317.
23. Ma'ayeh SY, Liu J, Peirasmaki D, Hörnaeus K, Bergström Lind S, Grabherr M, Bergquist J, Svärd SG. 2017. Characterization of the Giardia intestinalis secretome during interaction with human intestinal epithelial cells: the impact on host cells. PLoS Negl Trop Dis 11:e0006120. https://doi.org/10.1371/journal.pntd.0006120.
24. Karnkowska A, Treitli SC, Brzoň O, Novák L, Vacek V, Soukal P, Barlow LD, Herman EK, Pipaliya SV, Pánek T, Žihala D, Petrželková R, Butenko A, Eme L, Stairs CW, Roger AJ, Eliáš M, Dacks JB, Hampl V, Battistuzzi FU. 2019. The oxymonad genome displays canonical eukaryotic complexity in the absence of a mitochondrion. Mol Biol Evol 36:2292–2312. https://doi.org/10.1093/molbev/msz147.
25. Jerlström-Hultqvist J, Franzén O, Ankarklev J, Xu F, Nohýnková E, Andersson JO, Svärd SG, Andersson B. 2010. Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. BMC Genomics 11:543. https://doi.org/10.1186/1471-2164-11-543.
26. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UCM, Besteiro S, Sicheritz-Ponten T, Noel CJ, Dacks JB, Foster PG, Simillion C, Van De Peer Y, Miranda-Saavedra D, Barton GJ, Westrop GD, Müller S, Dessi D, Fiori PL, Ren Q, Paulsen I, Zhang H, Bastida-Corcuera FD, Simoes-Barbosa A, Brown MT, Hayes RD, Mukherjee M, Okumura CY, Schneider R, Smith AJ, Vanacova S, Villalvazo M, Haas BJ, Pertea M, Feldblyum TV, Utterback TR, Shu CL, Osoegawa K, De Jong PJ, Hrdy I, Horvathova L, Zubacova Z, Dolezal P, Malik SB, Logsdon JM, Henze K, Gupta A, Wang CC, et alet al. 2007. Draft genome sequence of the

sexually transmitted pathogen *Trichomonas vaginalis*. Science 315:207–212. https://doi.org/10.1126/science.1132894.
27. Lundin D, Torrents E, Poole AM, Sjöberg BM. 2009. RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. BMC Genomics 10:589–598. https://doi.org/10.1186/1471-2164-10-589.
28. Jiménez-González A, Xu F, Andersson JO. 2019. Lateral acquisitions repeatedly remodel the oxygen detoxification pathway in diplomonads and relatives. Genome Biol Evol 11:2542–2556. https://doi.org/10.1093/gbe/evz188.
29. Lee JH, Wood TK, Lee J. 2015. Roles of indole as an interspecies and interkingdom signaling molecule. Trends Microbiol 23:707–718. https://doi.org/10.1016/j.tim.2015.08.001.
30. Xu F, Jiménez-González A, Einarsson E, Ástvaldsson Á, Peirasmaki D, Eckmann L, Andersson JO, Svärd SG, Jerlström-Hultqvist J. 2020. The compact genome of *Giardia muris* reveals important steps in the evolution of intestinal protozoan parasites. Microb Genom 6:mgen000402. https://doi.org/10.1099/mgen.0.000402.
31. Marbaix AY, Noël G, Detroux AM, Vertommen D, Van Schaftingen E, Linster CL. 2011. Extremely conserved ATP- or ADP-dependent enzymatic system for nicotinamide nucleotide. J Biol Chem 286:41246–41252. https://doi.org/10.1074/jbc.C111.310847.
32. Rosen BP. 2002. Biochemistry of arsenic detoxification. FEBS Lett 529:86–92. https://doi.org/10.1016/S0014-5793(02)03186-1.
33. Tsai SL, Singh S, Chen W. 2009. Arsenic metabolism by microbes in nature and the impact on arsenic remediation. Curr Opin Biotechnol 20:659–667. https://doi.org/10.1016/j.copbio.2009.09.013.
34. Slyemi D, Bonnefoy V. 2012. How prokaryotes deal with arsenic. Environ Microbiol Rep 4:571–586. https://doi.org/10.1111/j.1758-2229.2011.00300.x.
35. Nývltová E, Šut'ák R, Žárský V, Harant K, Hrdý I, Tachezy J. 2017. Lateral gene transfer of *p*-cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba balamuthi*. Environ Microbiol 19:1091–1102. https://doi.org/10.1111/1462-2920.13636.
36. Frank AC, Amiri H, Andersson SGE. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. Genetica 115:1–12. https://doi.org/10.1023/a:1016064511533.
37. Cenci U, Moog D, Curtis BA, Tanifuji G, Eme L, Lukeš J, Archibald JM. 2016. Heme pathway evolution in kinetoplastid protists. BMC Evol Biol 16:109. https://doi.org/10.1186/s12862-016-0664-6.
38. Borvinskaya E, Gurkov A, Shchapova E, Baduev B, Shatilina Z, Sadovoy A, Meglinski I, Timofeyev M. 2017. Parallel *in vivo* monitoring of pH in gill capillaries and muscles of fishes using microencapsulated biomarkers. Biol Open 6:673–677. https://doi.org/10.1242/bio.024380.
39. Adam RD, Aggarwal A, Lal AA, De La Cruz VF, McCutchan T, Nash TE. 1988. Antigenic variation of a cysteine-rich protein in *Giardia lamblia*. J Exp Med 167:109–118. https://doi.org/10.1084/jem.167.1.109.
40. Nash TE, Aggarwal A, Adam RD, Conrad JT, Merritt JW, Jr. 1988. Antigenic variation in *Giardia lamblia*. J Immunol 141:636–641.
41. Li W, Saraiya AA, Wang CC. 2013. Experimental verification of the identity of variant-specific surface. mBio 4:1–13. https://doi.org/10.1128/mBio.00321-13.
42. Jiménez JC, Fontaine J, Grzych JM, Dei-Cas E, Capron M. 2004. Systemic and mucosal responses to oral administration of excretory and secretory antigens from *Giardia intestinalis*. Clin Diagn Lab Immunol 11:152–160. https://doi.org/10.1128/CDLI.11.1.152-160.2004.
43. Stadelmann B, Hanevik K, Andersson MK, Bruserud O, Svärd SG. 2013. The role of arginine and arginine-metabolizing enzymes during *Giardia*—host cell interactions in vitro. BMC Microbiol 13:256. https://doi.org/10.1186/1471-2180-13-256.
44. Emery SJ, Mirzaei M, Vuong D, Pascovici D, Chick JM, Lacey E, Haynes PA. 2016. Induction of virulence factors in *Giardia duodenalis* independent of host attachment. Sci Rep 6:1–16. https://doi.org/10.1038/srep20765.
45. Janouskovec J, Keeling PJ. 2016. Evolution: causality and the origin of parasitism. Curr Biol 26:174–177. https://doi.org/10.1016/j.cub.2015.12.057.
46. Striepen B, Pruijssers AJP, Huang J, Li C, Gubbels MJ, Umejiego NN, Hedstrom L, Kissinger JC. 2004. Gene transfer in the evolution of parasite nucleotide biosynthesis. Proc Natl Acad Sci U S A 101:3154–3159. https://doi.org/10.1073/pnas.0304686101.
47. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A 108:13624–13629. https://doi.org/10.1073/pnas.1110633108.
48. NCBI Resource Coordinators. 2017. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 45:D12–D17. https://doi.org/10.1093/nar/gkx1095.
49. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
50. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.
51. Karp PD, Latendresse M, Caspi R. 2011. The pathway tools pathway prediction algorithm. Stand Genomic Sci 5:424–429. https://doi.org/10.4056/sigs.1794338.
52. Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, Kothari A, Weaver D, Lee T, Subhraveti P, Spaulding A, Fulcher C, Keseler IM, Caspi R. 2016. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. Brief Bioinform 17:877–890. https://doi.org/10.1093/bib/bbv079.
53. Aurrecoechea C, Brestelli J, Brunk BP, Carlton JM, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Morrison HG, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert Christian JJ, Sullivan S, Treatman C, Wang H. 2009. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. Nucleic Acids Res 37:D526–D530. https://doi.org/10.1093/nar/gkn631.
54. Green ML, Karp PD. 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinformatics 5:76–16. https://doi.org/10.1186/1471-2105-5-76.
55. Lee TJ, Paulsen I, Karp P. 2008. Annotation-based inference of transporter function. Bioinformatics 24:i259–267. https://doi.org/10.1093/bioinformatics/btn180.
56. Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:327–331. https://doi.org/10.1093/nar/gkh454.
57. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, the UniProt Consortium. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932. https://doi.org/10.1093/bioinformatics/btu739.
58. Madeira F, Park Y, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641. https://doi.org/10.1093/nar/gkz268.
59. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
60. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
61. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210.
62. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
63. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQTree: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
64. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
65. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.