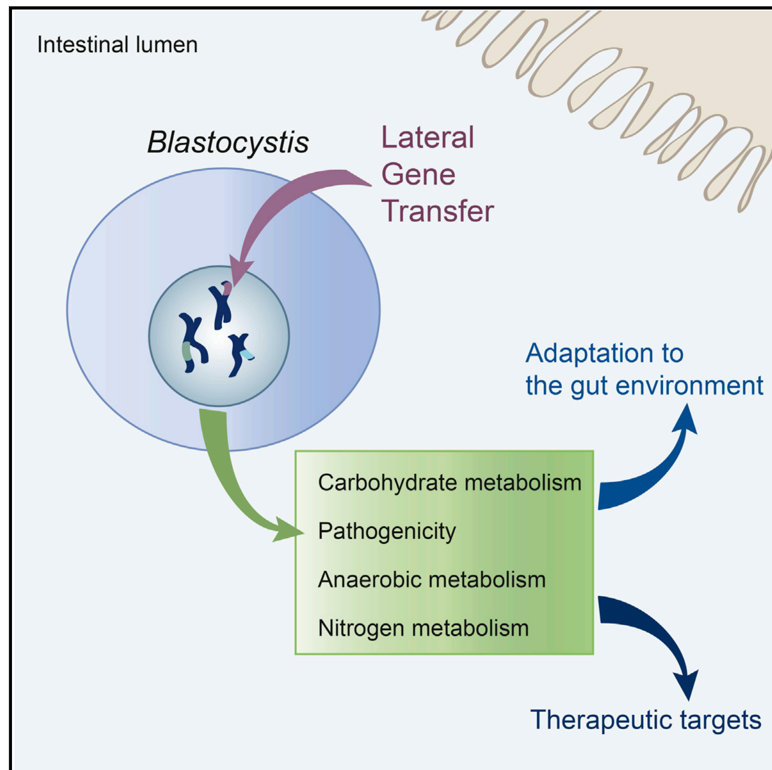
Article

Current Biology

Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut

Graphical Abstract



Authors

Laura Eme, Eleni Gentekaki, Bruce Curtis, John M. Archibald, Andrew J. Roger

Correspondence

andrew.roger@dal.ca

In Brief

Eme et al. investigate lateral gene transfer (LGT) in the genomes of *Blastocystis* spp., the most prevalent eukaryotic gut parasites. They show that ~2.5% of the genes were acquired by LGT and are involved in facilitating infection of the gut and escaping host defenses. Eme et al. identify promising targets for anti-protozoan drug development.

Highlights

- 2.5% of *Blastocystis* sp. ST1 genes originated by recent lateral gene transfer
- LGT in *Blastocystis* was crucial to its adaptation to the gut environment
- Newly acquired functions most likely facilitate infection and evasion of host defenses
- Specific laterally acquired genes represent novel putative drug targets

Eme et al., 2017, Current Biology *27*, 1–14
March 20, 2017 © 2017 Elsevier Ltd.
http://dx.doi.org/10.1016/j.cub.2017.02.003

CellPress

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

Current Biology
Article

CellPress

Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut

Laura Eme,[1] Eleni Gentekaki,[2] Bruce Curtis,[1] John M. Archibald,[1,3] and Andrew J. Roger[1,3,4,*]
[1]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, NS B3H 4R2, Canada
[2]School of Science and Human Gut Microbiome for Health Research Unit, Mae Fah Luang University, Chiang Rai 57100, Thailand
[3]Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, 180 Dundas Street W., Toronto, ON M5G 1Z8, Canada
[4]Lead Contact
*Correspondence: andrew.roger@dal.ca
http://dx.doi.org/10.1016/j.cub.2017.02.003

SUMMARY

***Blastocystis* spp. are the most prevalent eukaryotic microbes found in the intestinal tract of humans. Here we present an in-depth investigation of lateral gene transfer (LGT) in the genome of *Blastocystis* sp. subtype 1. Using rigorous phylogeny-based methods and strict validation criteria, we show that ~2.5% of the genes of this organism were recently acquired by LGT. We identify LGTs both from prokaryote and eukaryote donors. Several transfers occurred specifically in ancestors of a subset of *Blastocystis* subtypes, demonstrating that LGT is an ongoing process. Functional predictions reveal that these genes are involved in diverse metabolic pathways, many of which appear related to adaptation of *Blastocystis* to the gut environment. Specifically, we identify genes involved in carbohydrate scavenging and metabolism, anaerobic amino acid and nitrogen metabolism, oxygen-stress resistance, and pH homeostasis. A number of the transferred genes encoded secreted proteins that are potentially involved in infection, escaping host defense, or most likely affect the prokaryotic microbiome and the inflammation state of the gut. We also show that *Blastocystis* subtypes differ in the nature and copy number of LGTs that could relate to variation in their prevalence and virulence. Finally, we identified bacterial-derived genes encoding NH_3-dependent nicotinamide adenine dinucleotide (NAD) synthase in *Blastocystis* and other protozoan parasites, which are promising targets for drug development. Collectively, our results suggest new avenues for research into the role of *Blastocystis* in intestinal disease and unequivocally demonstrate that LGT is an important mechanism by which eukaryotic microbes adapt to new environments.**

INTRODUCTION

Members of the stramenopile genus *Blastocystis* are the most frequently occurring microbial eukaryotes in the human intestinal tract. It is estimated that one billion individuals are infected worldwide, with higher prevalence in developing countries [1]. Although these anaerobic protists were described over 100 years ago, they are poorly studied and their pathogenicity is still debated [2]. *Blastocystis* infections in humans can be either asymptomatic or symptomatic and have been linked to a number of bowel diseases, including irritable bowel syndrome (IBS) [2]. Knowledge of the pathobiology of *Blastocystis* is hindered by our incomplete understanding of its exceptional genetic diversity as well as its host specificity, geographic distribution, and interactions with the prokaryotic gut flora [2]. Although the various *Blastocystis* strains are indistinguishable microscopically, small subunit (SSU) rDNA analyses have identified 17 subtypes (STs). They diverge in sequence by up to 3% within STs and 5%–15% between them. Only ST1–ST9 have been found in humans, with ST1–ST4 being the most common [2]; most STs have also been found infecting other mammals and birds.

In 2011, the first *Blastocystis* genome was sequenced from the ST7 clade [3]. Recently, the genome sequence of *Blastocystis* ST4 was also published, although a detailed analysis of gene content was not reported [4]. Incomplete draft assemblies for the genomes of ST2, ST3, ST6, ST8, and ST9 were recently used in a microbiomic analysis [5], but they are not annotated and transcriptomic data are lacking. We have sequenced the genome and transcriptome of the NandII strain of *Blastocystis* sp. ST1 (hereafter referred to as ST1) (deposited in GenBank BioProject: PRJNA308101; E.G., B.C., J.M.A., and A.J.R., unpublished data). Here, we have investigated the role of lateral gene transfer (LGT) in shaping the gene repertoire of this strain. As part of their analyses of *Blastocystis* sp. ST7, Denoeud and colleagues briefly reported candidate LGTs from prokaryotes and proposed that some of these are potentially involved in anaerobic fermentation and virulence [3]. However, there is tremendous genetic diversity between STs—the SSU rDNAs of ST1 and ST7 are only 86.6% identical—and the impact of recent, strain-specific LGTs on the pathogenicity of *Blastocystis* has not been investigated.

Lateral gene transfer is a fundamentally important evolutionary mechanism affecting bacteria and archaea. Although long thought to be largely restricted to prokaryotes, the last decade has seen an increasing number of reports of LGT into eukaryotic genomes from both endosymbionts and free-living organisms (e.g., [6, 7]). Some extensive studies of LGT, for example, in oomycete plant pathogens [7], and in diverse microbial eukaryotes

 © 2017 Elsevier Ltd.

[8] suggested that laterally acquired genes contribute to adaptation of eukaryotic microbes to their environment. However, a group of researchers has recently sparked debate by publishing several large-scale studies purporting to show that eukaryotes have remained genetically isolated from prokaryotes (except during the endosymbiotic origins of organelles) [9], that gene inheritance in eukaryotes is vertical [9], and that eukaryotic genomes show no evidence for either continuous or recent gene acquisitions from prokaryotes [10]. This has renewed controversy regarding the impact of LGT on eukaryote evolution.

The most reliable approach to LGT detection involves reconstruction of phylogenies to identify genes showing strongly supported relationships that contradict established species relationships [7]. Additional analyses are also necessary to exclude potential prokaryotic contaminants and phylogenetic artifacts. With this in mind, we developed a phylogenomic pipeline that infers phylogenies of all genes in a genome and identifies those with anomalous branching patterns. We combined these with multiple additional lines of evidence to rigorously investigate the role of LGT in the evolution of *Blastocystis* sp. ST1. Importantly, we searched for LGTs from both prokaryotic and eukaryotic donors, the latter being rarely studied because of the increased difficulty of identifying them with confidence. In total, we report 167 genes (corresponding to 74 transfer events followed by duplications) that are clear cases of "recent" LGT (i.e., acquired in the *Blastocystis* lineage since its divergence from other stramenopiles and alveolates; see the Experimental Procedures). We investigated the putative roles of these genes in the colonization of the intestinal tract by *Blastocystis* and in the interactions of the parasite with the healthy and diseased host and its microbiome. We also investigated differences regarding these laterally acquired genes between STs. Finally, we report a number of novel candidate drug targets that occur not only in *Blastocystis* but also in other protozoan pathogens.

RESULTS AND DISCUSSION

Overall Patterns of LGT

We analyzed 6,544 predicted proteins from the genome and transcriptome of *Blastocystis* sp. ST1 using a phylogenomic approach to identify genes that show well-supported "non-vertical" evolutionary patterns. We focused on LGTs specific to the *Blastocystis* lineage (i.e., either specific to ST1 or shared with other STs), to the exclusion of other stramenopiles or alveolates, their closest relatives. An automated screening procedure followed by thorough manual inspection of candidates (see the Experimental Procedures) yielded 167 genes (2.5% of the protein coding genes of ST1) belonging to 74 gene families that were most likely acquired by LGT (Table S1). Phylogenies can be downloaded from Mendeley Data at http://dx.doi.org/10.17632/pktp3hggf7.1 (hereafter referred to as Mendeley Data).

Several lines of evidence exclude the possibility that our LGT candidates are the result of contamination. All of the gene families identified in ST1 were found in the genome data of at least one other *Blastocystis* ST. In addition, gene members of 66 of the 74 gene families contained introns, and among these, all but four genes have introns with a very narrow size distribution, with a median length of 30 bp, exactly matching the median intron length for the whole genome (Table S1). Polyadenylation is known to create termination codons in approximately 15% of nuclear genes in ST7 [11] and 26% in ST1 (E.G., B.C., J.M.A., and A.J.R, unpublished data). Accordingly, 24% of our LGT candidate genes possess this feature, which is unique to *Blastocystis* spp. All LGT candidates were detected in both genomic and transcriptomic data and are flanked by genes of eukaryotic origin on genomic scaffolds.

It is worth mentioning that, in one protein family (endoribonuclease L-PSP) (Mendeley Data, p. 28), *Blastocystis* homologs showed up to 74% amino acid identity to prokaryotic counterparts, breaking the so-called "70% rule" [10]. A single example might appear insignificant, but it should be remembered that *Blastocystis* genes are generally fast evolving, as illustrated by the fact that ST1 homologs belonging to the 74 families of LGT origin are on average only 62% identical at the protein level to orthologs in ST4 and ST7.

Most acquired genes are of prokaryotic origin (Figure 1), although in many cases, the lack of phylogenetic resolution and rampant LGTs between prokaryotes hampers a precise inference of the donor. Nevertheless, many genes appear to have been acquired from firmicutes, proteobacteria, and bacteroidetes, three bacterial phyla that are predominant in the gut [12]. Surprisingly, we also identified ~20 candidate LGTs involving other eukaryotes. The possibility of transfers between eukaryotes is often neglected and has not been studied systematically. Most LGT candidates (71%) mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways, with 34% and 19% of these mapping to sugar and amino acid (aa) metabolism pathways, respectively (Figure 1), consistent with analyses of other mucosal parasites, such as *Entamoeba histolytica* and *Giardia lamblia* [8]. Finally, more than one-third of LGT candidates are predicted to encode plasma membrane or secreted proteins, possibly mediating interactions with the external environment.

Colonization of the Gut Environment by *Blastocystis* Was Facilitated by LGT

The majority of the LGTs we identified have potential roles in the adaptation to colonization of the intestinal tract (Figure 2A), including genes involved in oxygen-stress response, nutrient scavenging and metabolism, intracellular pH homeostasis, and host immune response evasion and pathogenicity.

Oxygen-Stress Response

Five LGT gene families in *Blastocystis* are predicted to be involved in response to oxidative stress (Figure 2A), such as oxidative bursts generated by the host immune response [13] or metabolites produced by gut microbiota [14]. Of these, two are of particular note. *Blastocystis* acquired a *yqhD* gene encoding the alcohol dehydrogenase ADH3 (GenBank: OAO14540.1) from proteobacteria (Mendeley Data, p. 38). Interestingly, a glycerol dehydrogenase, GldA (OAO14793.1), which acts downstream of YqhD, was also transferred, possibly from proteobacteria, although this is less clear (Mendeley Data, p. 41). In bacteria, these two enzymes are responsible for the degradation of methylglyoxal, a toxic metabolite often produced by intestinal microbes that can cause oxidative stress and is implicated in irritable bowel syndrome [15]. Indeed, detoxification of methylglyoxal by GldA is essential for viability of the enteric pathogen *Clostridium difficile* [16]. Strikingly, *yqhD* appears to have been

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

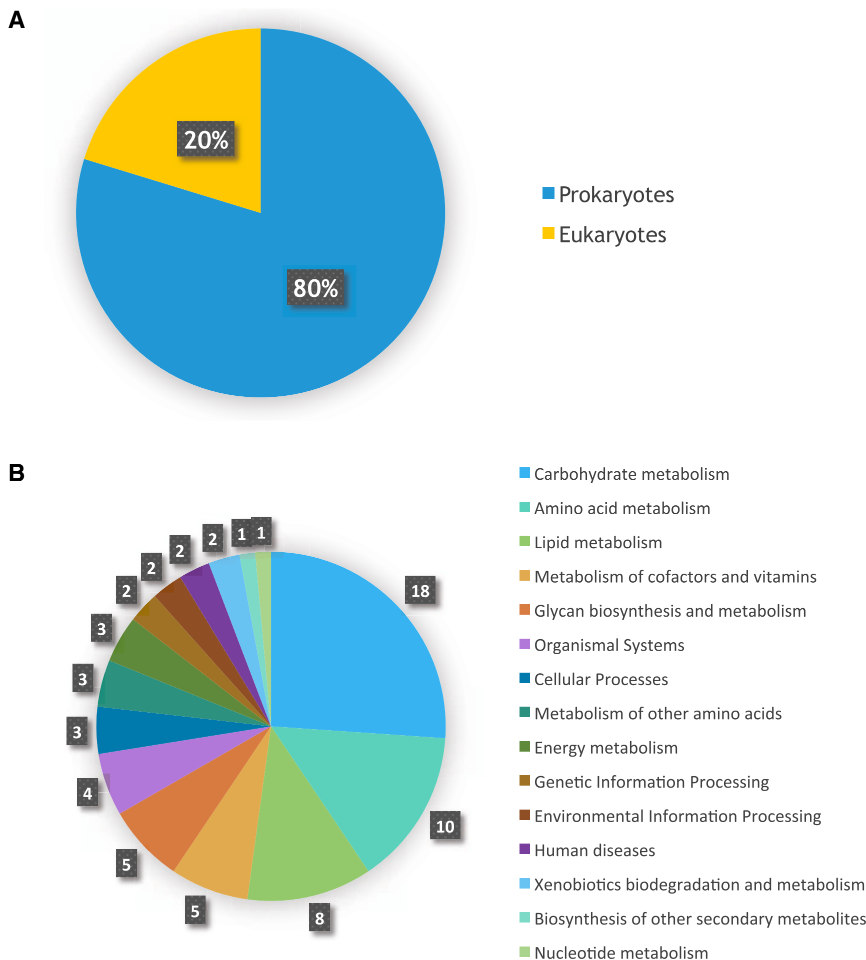


Figure 1. Taxonomic Distribution of LGT Donors and Distribution of KEGG Annotations of LGT Candidates

LGT donors are shown in (A), and KEGG annotations of LGT candidates in (B). The numbers of LGT candidates in each KEGG category are shown in (B). See also Table S1.

laterally transferred to the mucosal protist parasites *Entamoeba* and *Trichomonas vaginalis*, each through independent events. This gene is dramatically more expressed in the virulent strains of *E. histolytica* compared to the commensal *E. dispar* and is most likely responsible for an increased inflammatory response [17]. Collectively, these lines of evidence suggest an important role of YqhD and GldA in the survival—and possibly virulence—of *Blastocystis* in the gut.

Blastocystis has acquired a thioredoxin-domain-containing protein 12 (Txndc12) (OAO12372.1), a metazoan thioredoxin superfamily protein that functions in redox regulation and oxidative stress defense [18]. Txndc12 possibly provides *Blastocystis* with resistance to the oxidative burst generated by the host's immune system in reaction to infection by the parasite [13]. This gene is also a clear example of gene transfer between eukaryotes, perhaps from an ancestral metazoan host to the parasite (Mendeley Data, p. 6). After transfer, this gene underwent several rounds of duplication, yielding six copies in the ST1 genome, consistent with the idea that this enzyme is of particular importance for the parasite's fitness.

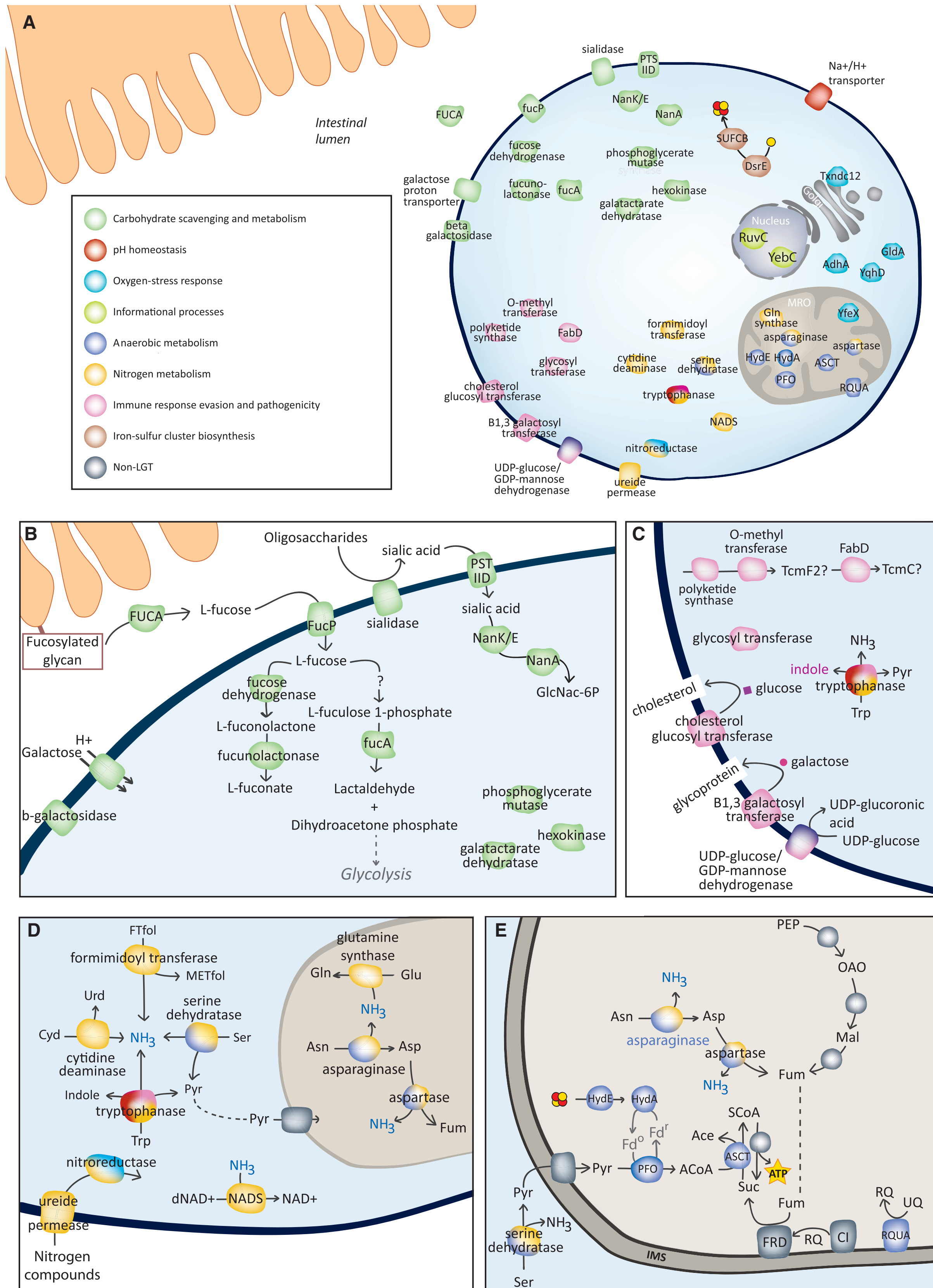
Carbohydrate Scavenging and Metabolism

Remarkably, we identified 17 horizontal acquisitions of genes involved in carbohydrate metabolism (Figure 2B), including five implicated in fucose import and metabolism. Fucose is of major importance for the microbial ecology of the gut: fucosylated glycans exist at the surface of mucosal epithelial cells in the form of H and Lewis antigens and are chemoattractants and carbon sources for many intestinal microbes [19]. As a result, the intestinal microbiota is markedly altered in so-called "non-secretor" individuals who possess inactivated alleles of the FUT2 gene (alpha-(1,2)-fucosyltransferase) and consequently lack fucosylated glycans [20]. Conversely, members of the intestinal microbial community influence the abundance of fucosylated mucosal glycans by modulating FUT2 expression, a process that is important for gut maturation [21]. Gut bacteria secrete fucosidases to release fucose monomers from glycans and internalize them through a fucose permease for metabolism and surface presentation [22]. Mucosal fucosylated glycans are also required for the adherence of many intestinal bacteria, including pathogens like *Campylobacter jejuni* [23].

Our analyses revealed five transferred genes in *Blastocystis* involved in the acquisition and metabolism of L-fucose (Figure 2B). Notably, alpha-L-fucosidase FUCA (OAO14080.1; Mendeley Data, p. 32), L-fucose permease fucP (OAO16100.1; Mendeley Data, p. 53), and L-fuculose-phosphate aldolase fucA (OAO17346.1; Mendeley Data, p. 64) are homologous to enzymes found in *B. thetaiotaomicron* [24], whereas L-fucose dehydrogenase (OAO17151.1; Mendeley Dat, p. 63) and L-fuconolactase (OAO12577.1; Mendeley Data, p. 11) belong to a different metabolic pathway, similar to the one found in *C. jejuni* [25]. The assimilation of fucose is therefore most likely important for the successful colonization of the gut by *Blastocystis*. If so, this may explain why Crohn's disease (CD) patients have a markedly lower *Blastocystis* infection rate than healthy individuals [5]: the strong association between the non-secretor FUT2 genotype and CD [20] could mean that, in non-secretor-type CD patients, *Blastocystis* cannot utilize fucose for growth and colonization. This hypothesis is bolstered by the observation that the non-secretor FUT2 phenotype is associated with reduced susceptibility to infections by *Campylobacter jejuni* [26], a bacterial pathogen with a fucose metabolism pathway similar to *Blastocystis*.

We identified a dozen other gene families involved in scavenging and metabolism of carbohydrates (including sialic acid and chitin), most of bacterial origin (Figure 2B). The presence of α-L-fucose, chitin, and sialic acid on the cell surface of

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003



(legend on next page)

Blastocystis [27] further suggests the importance of laterally acquired genes that play a role in their scavenging and/or metabolism. In fact, it is possible that *Blastocystis* recycles fucose and other host carbohydrates to express glycosylated molecules on its surface, thereby aiding evasion of the host immune system (see below).

Nitrogen Metabolism

Many gut organisms are capable of fixing NH_3, suggesting that this molecule is an important source of nitrogen in the intestine. Although it is unclear whether *Blastocystis* excretes and/or imports NH_3, the parasite has laterally acquired ten genes involved in nitrogen metabolism (Figure 2D).

Among these, we identified a mitochondrion-targeted glutamine synthetase (GS) of bacterial origin (OAO13151.1; Mendeley Data, p. 20), which catalyzes condensation of ammonia and glutamate to form glutamine. Although virtually all eukaryotes possess GS type II—a homo-oligomer of ten ~400-aa subunits [28]—*Blastocystis* apparently does not encode this enzyme. Instead, its genome has a gene for GS type III, an ~700-aa protein that typically functions as a hexamer in anaerobic bacteria [28]. GS III is upregulated under ammonia-limiting conditions in rumen-dwelling bacteria [28], suggesting that *Blastocystis* GSIII could play a role in ammonia assimilation during growth under low-nitrogen conditions in the gut.

Blastocystis has acquired a bacterial *nadE* gene encoding a NH_3-dependent nicotinamide adenine dinucleotide (NAD) synthetase (NH_3-NADS) (OAO17827.1; Mendeley Data, p. 69), an enzyme catalyzing the last step of NAD biosynthesis. Interestingly, phylogenetic analyses show that *Blastocystis* homologs branch within a strongly supported clade containing a few other eukaryotic homologs from distantly related protists, including the amoebozoan parasite *Entamoeba*, the parasitic ichthyosporean *Sphaeroforma*, and many excavate lineages (Figure 3). All of these organisms lack the canonical eukaryotic NAD synthetase, a protein with an extra N-terminal glutamine amidotransferase domain [29]. The latter glutamine-dependent NADS (Gln-NADS) is found in the vast majority of eukaryotes, including the closest relatives of the aforementioned protists. The evolutionary advantage of replacing the eukaryotic NADS with a bacterial one in these protists is unclear, although it is possibly related to the relative availability of ammonia versus glutamine in their habitats.

The xenologous replacement of the canonical Gln-NADS by a bacterial NH_3-NADS in these protists has never been reported and is of considerable therapeutic significance. NAD is an indispensable cofactor required for redox balance and a multitude of metabolic pathways. Indeed, NadE is essential in *Mycobacterium tuberculosis* [30] and is an actively pursued broad-spectrum target against bacterial pathogens. Compounds that inhibit NadE have been identified as effective antibiotics against *Bacillus anthracis* and *M. tuberculosis* [31, 32]. In addition, although Gln- and NH_3-NADS share the C-terminal domain responsible for the NAD synthesis activity, there are structural differences that might prevent these NadE inhibitors from disabling the host's NADS [33]. These NadE-inhibiting compounds could represent equally promising broad-spectrum anti-protozoan drugs—potentially inhibitory not only to *Blastocystis* but also *E. histolytica* and a wide range of excavate pathogens, including *Giardia intestinalis*, *T. vaginalis*, and kinetoplastid parasites (the latter causing devastating diseases, such as leishmaniasis, Chagas disease, and African sleeping sickness).

Among the *Blastocystis* LGTs, we also identified a bacterial nitroreductase (OAO17950.1; Mendeley Data, p. 72). Nitroreductases in prokaryotes are involved in the reduction of nitrogen-containing compounds [34], allowing them to degrade toxic compounds. Nitroreductases are of particular interest because of their role in activating antimicrobials, such as nitrofurans, and nitroimidazoles, such as metronidazole, which they convert into mutagenic metabolites, ultimately resulting in cell death [35]. The homolog acquired by *Blastocystis* is a member of the oxygen-insensitive "type 1" nitroreductases. Although metronidazole is commonly used to treat *Blastocystis* infections (and those of other anaerobic parasites), the mechanism of its activation has not been studied in this organism. It is thought to be activated by pyruvate:ferredoxin oxidoreductase (PFO), as in *Giardia* [36], but so far, no PFO biochemical activity has been detected in *Blastocystis* [37]. In addition, there are reports of metronidazole-resistant *Blastocystis* isolates, some of them being sensitive to a nitrofuran compound (furazolidone) [38]. This suggests that at least another enzyme besides PFO could play a role in activating these drugs; we hypothesize that it is the nitroreductase identified herein. Further investigation of this enzyme may help us better understand the efficacy of, and resistance to, these nitroaromatic drugs in *Blastocystis*.

Immune Response Evasion and Pathogenicity

Our analyses yielded at least seven laterally acquired genes with potential roles in evasion of the host immune response and pathogenicity (Figure 2C). Notably, a gene encoding a beta-1,3-galactosyltransferase (β1,3GalT) appears to have been acquired from animals and was subsequently duplicated several times in the *Blastocystis* genome (OAO16986.1). In humans, β1,3GalT is responsible for the biosynthesis of the Lewis blood group antigens and the corresponding ABH determinants that are expressed in various cell types, including the epithelial cells lining the digestive tract [39]. We hypothesize that, in *Blastocystis*, these genes encode proteins that are involved in "molecular mimicry," potentially serving to camouflage its cell surface from

Figure 2. Cellular Processes Affected by LGT in *Blastocystis* ST1

(A) Overview of the main cellular processes affected by LGT. Proteins are color coded depending on the process they are involved in. Those labeled with several colors are involved in multiple processes. Not all LGT candidates are depicted.

(B–E) Close-ups of LGT affecting carbohydrate uptake and metabolism (B), immune response evasion and pathogenicity (C), nitrogen metabolism (D), and anaerobic energy metabolism (E).

Ace, acetate; AcoA, acetyl-CoA; Asn, asparagine; Asp, aspartate; CI, complex I; Cyd, cytidine; dNAD+, deamido-NAD+; Fd^r, ferredoxin (reduced); Fd°, ferredoxin (oxidized); Fum, fumarate; FRD, fumarate reductase; FTfol, 5-formimidoyltetrahydrofolate; GlcNac-6P, N-acetylglucosamine 6-phosphate; Gln, glutamine; Glu, glutamate; IMS, inter-membrane space; Mal, malate; METfol, 5,10-methenyltetrahydrofolate; OAO, oxaloacetate; PEP, phosphoenolpyruvate; Pyr, pyruvate; RQ, rhodoquinone; Ser, serine; SCoA, succinyl-CoA; Suc, succinate; TcmC, tetracenomycin C; TcmF2, tetracenomycin F2; Trp, tryptophane; UQ, ubiquinone; Urd, uridine. See also Figure S1, Table S1, and Mendeley Data.

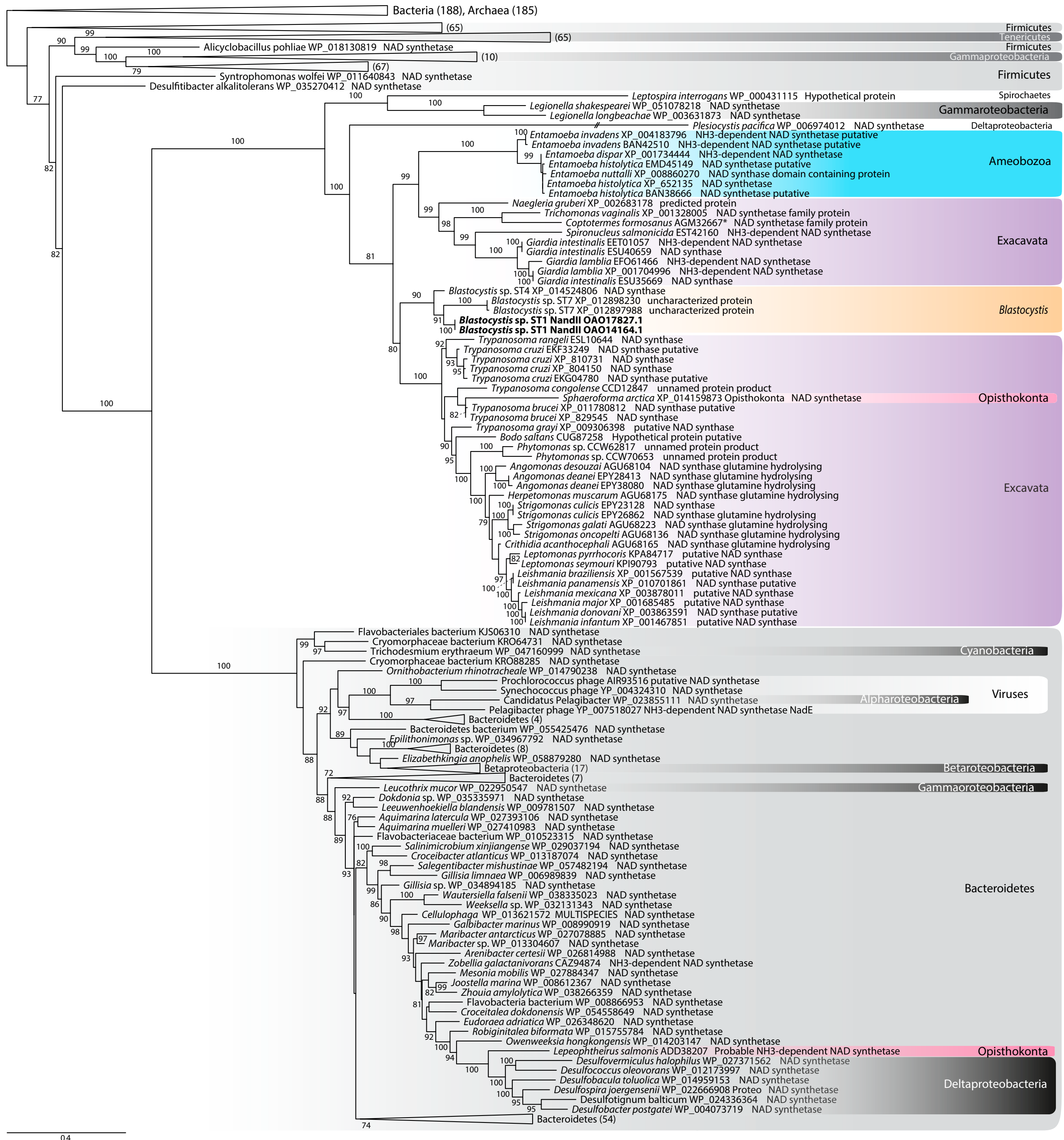


Figure 3. Maximum-Likelihood Phylogeny of NAD Synthase

Shades of gray highlight bacterial, archaeal, and viral homologs. Other colors represent eukaryotic homologs. GenBank accession numbers and annotation are provided next to each taxon. Numbers in brackets indicate the number of homologs collapsed. Numbers at branches are UltraFast bootstrap support (BS); only values ≥70 are indicated. *Blastocystis* sp. ST1 homologs are bolded. The branch leading to *Pleisiocystis pacifica* was reduced to two-thirds of its original length (double dashes). The *Coptotermes formosanus* sequence is marked with an asterisk to indicate that it is most likely a contaminating sequence from a parabasalid gut symbiont.

the host. Indeed, mimicry of host structures is critically important for gastric colonization, adhesion, and immune evasion in mucosal bacteria [40]. For example, structural similarity has been shown to exist between human glycosphingolipids and the surface lipopolysaccharides (LPSs) of *C. jejuni* and *Helicobacter pylori* [41]. Indeed, β1,3GalT homologs in *C. jejuni* are

CellPress

involved in synthesis of ganglioside-mimicking LPS [40, 41]. If *Blastocystis* uses β1,3GalT for antigenic mimicry, this is the first known case of this mechanism occurring in a eukaryotic intestinal parasite.

The most striking example of LGTs involved in immune response evasion is the acquisition of a cholesterol-alpha-glucosyltransferase (OAO12712.1; Mendeley Data, p. 17) from bacteria. Notably, *Blastocystis* homologs are closely related to the well-studied protein encoded by *capJ* in *H. pylori*. Increased cell membrane cholesterol content stimulates phagocytosis of *H. pylori* by macrophages [42]. However, by alpha-glucosylating cholesterol, *H. pylori* escapes phagocytosis, T cell activation, and bacterial clearance in vivo [42]. We hypothesize that the laterally acquired cholesterol glucosyltransferase in *Blastocystis* serves the same purpose.

Host mucus-associated O-glycans are known to have antimicrobial activity through the inhibition of cholesteryl glucoside biosynthesis [43]. The secreted mucin glycoprotein responsible for this antimicrobial, MUC6, has not been detected in healthy small intestine mucosae; it is normally only expressed in the stomach. However, it has been shown to be expressed in the ulcerated small intestinal tissue of CD patients [44]. This abnormal presence of MUC6 in the gut could lead to inhibition of cholesterol-alpha-glucosyltransferase in *Blastocystis*, leading to increased phagocytosis and clearance of the parasite, perhaps contributing to its lowered prevalence in CD patients [5, 45].

Denoeud and colleagues described a putative type I polyketide synthase (PKS) of bacterial origin in *Blastocystis* sp. ST7, which we confirm is also present in ST1 (OAO12646.1; Mendeley Data, p. 12) and all other STs. Polyketides are a vast array of structurally diverse secondary metabolites, including many antibiotics and antiparasitics [46]. It was hypothesized that the polyketides synthesized by *Blastocystis* could contribute to host dysbiosis and inflammation occurring during IBS [3, 47]. In addition, we have identified an LGT of a bacterial O-methyltransferase also involved in polyketide biosynthesis (OAO13452.1; Mendeley Data, p. 24); this is the only case of transfer involving ST1, but not ST4 and ST7. Interestingly, we also detected an ortholog in the ST2 draft genome sequence data, which branches sister to the ST1 sequence in phylogenetic reconstructions (data not shown). We were unable to detect homologs in any of the other six *Blastocystis* STs (ST4 and ST7, as well as the draft genomes of ST3, ST6, ST8, and ST9), strongly suggesting that these STs truly lack the gene. It appears this gene was acquired relatively recently in the branch leading to the ancestor of ST1 and ST2. It is interesting to note that the mucosal pathogen *Trichomonas vaginalis* has also acquired an ortholog of this gene. Whereas the precise function(s) of these methyltransferases in *Blastocystis* and *Trichomonas* remains to be determined, the closely related *Streptomyces* homolog, tcmP, is involved in the biosynthesis of the antibiotic tetracenomycin C [48]. Whether *Blastocystis* ST1 and ST2 are able to synthesize this antibiotic or a related compound remains to be experimentally investigated.

Also related to polyketide synthesis, we found a homolog of FabD (OAO16590.1; Mendeley Data, p. 57), a malonyl transferase that has been suggested to participate in polyketide synthesis processes in *Streptomyces* [49]. In particular, it has been shown to be involved in the biosynthesis of tetracenomycin F2, a precursor of tetracenomycin C [50].

Finally, we identified a *Blastocystis* gene encoding a bacterial tryptophanase that is most likely producing indole, a compound previously thought to be only produced by bacteria. Indole passively diffuses across membranes and influences many physiological processes, including biofilm and spore formation, antibiotic resistance, and virulence [51]. This molecule has inter-species effects in multidrug resistance, virulence factor expression, and host cell invasion by indole-non-producers as well as mitigating inflammation in the intestinal tract [52]. The acquisition of tryptophanase by *Blastocystis* thus possibly has important consequences for its interaction with the gut microbiota and on overall intestinal health.

Anaerobic Metabolism Was Shaped by LGT

Probable pathways of anaerobic ATP generation occurring within the *Blastocystis* mitochondrion-related organelles (MROs) have been previously described [53]. In brief, pyruvate can most likely be decarboxylated into acetyl-coenzyme A (CoA) by two mechanisms: (1) by a mitochondrial-type pyruvate dehydrogenase complex or (2) a pyruvate:ferredoxin and/or pyruvate:NADP oxidoreductase [37]. The latter (OAO12944.1; Mendeley Data, p. 18) appears to be of LGT origin. Acetyl-CoA is then converted to succinyl-CoA by a laterally acquired acetate:succinate CoA transferase (ASCT) (OAO17397.1; Mendeley Data, p. 67). A typical mitochondrial succinyl-CoA synthetase then makes ATP by substrate level phosphorylation (Figure 2E). As in anaerobic helminths and ciliates, it is likely that succinate is generated in *Blastocystis* MROs by running mitochondrial complex II (succinate dehydrogenase) "backward" as a fumarate reductase [53], a reaction that requires a quinone of lower redox potential, such as rhodoquinone. Consistent with this, we identified a laterally acquired, MRO-targeted putative RQUA (OAO14277.1; Mendeley Data, p. 35)—the only enzyme identified so far to be implicated in the biosynthesis of rhodoquinone [54]. The fumarate feeding the fumarate reductase (FRD) reaction most likely comes from the "malate dismutation" pathway in which malate is converted to fumarate by running part of the Kreb's cycle backward [54] (Figure 2E). However, we have also identified two LGTs—a bacterial L-asparaginase II (OAO12421.1; Mendeley Data, p. 9) and a mitochondrion-targeted aspartase (OAO16085.1; Figure S1)—that we suggest produce fumarate from amino acids abundant in the gut (Figure 2E). In *E. coli*, aspartate and asparagine are interconverted by the periplasmic L-asparaginase II, which is expressed under anaerobic conditions in the presence of high levels of amino acids, such as in the intestinal tract [55]. Aspartate can then be converted into fumarate by aspartase. The importance of aspartase for survival in the gut was shown in *Campylobacter jejuni*, in which aspartase is upregulated by oxygen limitation, and enhances persistence of *C. jejuni* in the intestine [56]. The horizontal acquisition of asparaginase and aspartase may have led to similar metabolic modifications (Figure 2E), facilitating adaptation of *Blastocystis* to low-oxygen conditions in the gut.

The above-mentioned, MRO-localized, ATP-generating pathways catabolize pyruvate that is presumably generated by glycolysis. However, we have identified another potential source of pyruvate: *Blastocystis* has acquired a bacterial L-serine dehydratase (OAO12304.1; Mendeley Data, p. 5), which converts serine to pyruvate and ammonia (Figure 2E). Our phylogenetic

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

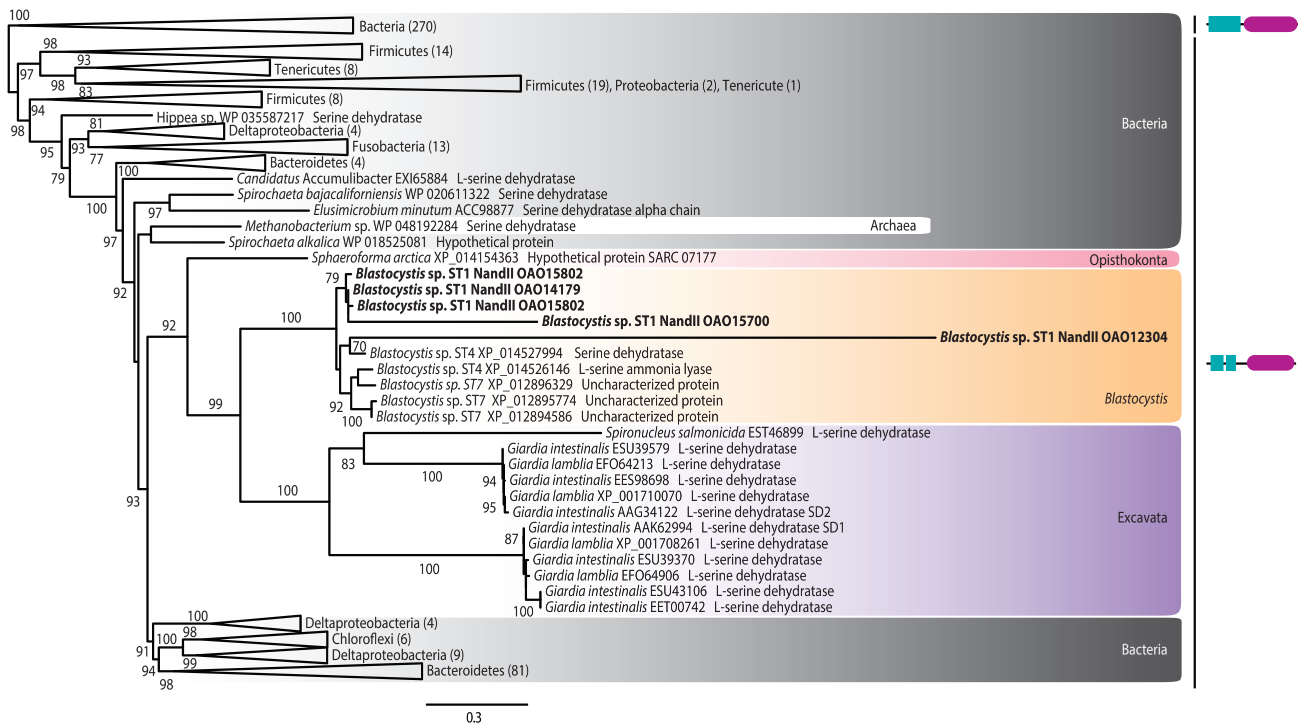


Figure 4. Maximum-Likelihood Phylogeny of Serine Dehydratase

The legend from Figure 3 applies unless otherwise specified hereafter. The tree was arbitrarily rooted based on functional domain composition, as displayed on the right end of the figure: the ingroup sequences are predicted to be composed of a split serine dehydratase alpha chain (teal) and a serine dehydratase beta chain (purple), whereas the outgroup sequences are predicted to be composed of a single alpha chain domain followed by a beta chain longer than in the ingroup.

analyses reveal that homologs of L-serine dehydratase in *Giardia* and its close relative *Spironucleus* branch sister to the *Blastocystis* ones, indicating an LGT between these two lineages after a single transfer event from bacteria (Figure 4). Many intestinal bacteria harbor a L-serine dehydratase, and in *C. jejuni*, it is upregulated in response to fucose [25] and is indispensable to intestinal colonization [57].

Recurring Patterns in the Evolutionary History of Laterally Transferred Genes

Genes of Prokaryotic Origin Are Commonly Transferred Multiple Times between Protists

We found 19 clear cases of bacterial-derived genes that have been subsequently transferred between eukaryotes. In these cases, *Blastocystis* spp. homologs often branch together with a small number of sequences from other protists—frequently anaerobes and/or pathogens. This pattern is observed not only for anaerobic ATP-generating enzymes, such as ASCT, SUFCB, and RQUA [54], but also for other proteins involved in adaptation to low-oxygen conditions, such as serine dehydratase (transferred between *Blastocystis*, the anaerobic diplomonad gut pathogens *Giardia* and *Spironucleus*, and the ichthyosporean parasite *Sphaeroforma*; Figure 4), aspartase (where *Blastocystis* represents the sister lineage of the prasinophyte *Micromonas* and *Ostreococcus*, which both experience transient anoxia; Figure S1) [58], and NAD synthase (where *Blastocystis*, *Sphaeroforma*, and excavates are monophyletic; Figure 3). Another striking example is the dipeptidase PepD, acquired from bacteria and subsequently transferred between the parasitic lineages of *Giardia*, *Sphaeroforma*, and *Blastocystis* (Figure 5). Homologs of this protein have been shown to be involved in *Mycobacterium* virulence [59] and resistance to antimicrobial peptides in *E. coli* [60]. In addition, the *Giardia* homolog is one of 40 genes upregulated upon interaction with intestinal epithelial cells [61], suggesting its importance in establishing infection. Interestingly, our transcriptome data show that it is the fifth most expressed protein-coding gene in *Blastocystis* sp. ST1. Overall, lateral gene exchange with other eukaryotic microbes seems to have been a significant mechanism mediating the evolutionary adaptation of *Blastocystis* to its environment.

Gene Duplication after Transfer

LGT followed by extensive gene duplication is a common occurrence in *Blastocystis* and is consistent with reports for other organisms [8]. For 48 of the LGT candidates identified herein, at least one gene duplication occurred in *Blastocystis* ST1, ST4, or ST7, including 13 cases in which there are at least five copies in one or more ST (Table S1). One of the largest family expansions involves the fucose transporter genes that are present in seven to nine copies in all STs. Given the number of recently acquired enzymes involved in fucose metabolism or that are regulated by fucose concentration, these transporters are likely to be of key importance to the parasite's ability to colonize the gut.

Several LGT gene families seem to have differentially expanded in different STs. For example, *Blastocystis* sp. ST1 has five copies of the cholesterol glucosyltransferase, whereas

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

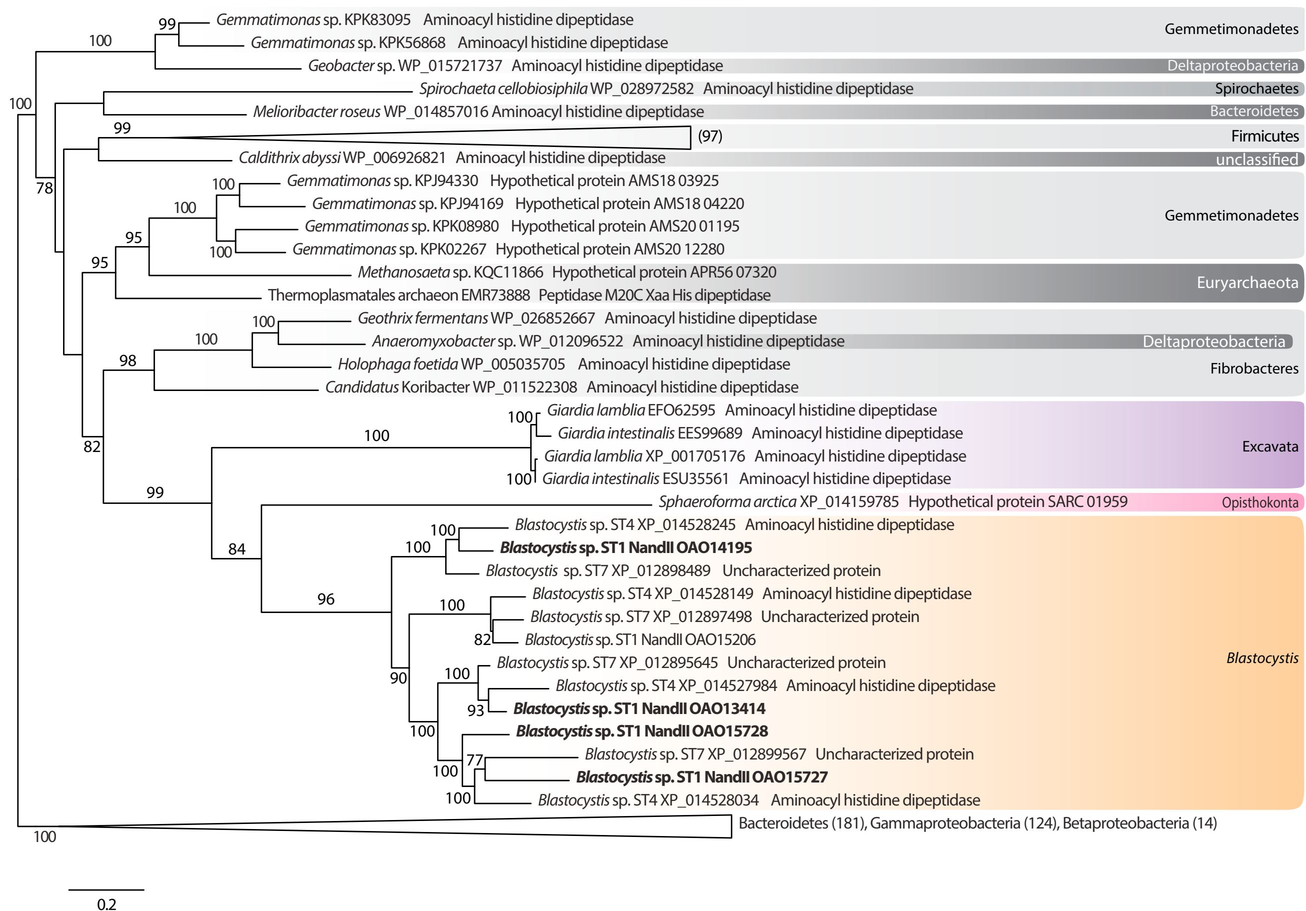


Figure 5. Maximum-Likelihood Phylogeny of PepD-like Dipeptidase
See the legend from Figure 3.

only two and one are found in ST7 and ST4, respectively. Even more striking is the uridine diphosphate (UDP)-glucose/guanosine diphosphate (GDP)-mannose dehydrogenase family, which is found in one and two copies in ST4 and ST7 but in ten copies in ST1 (Figure 6). This is particularly interesting given that the *Mycobacterium* homolog of the latter enzyme is involved in virulence regulation [62] (Mendeley Data). Gene families with large copy-number variation between STs represent good candidates for future experimental investigations of ST virulence differences.

LGT and Differential Loss Are Ongoing in the Blastocystis *Lineage*

Our analyses suggest that a number of the gene families identified as LGTs have either been acquired recently in the ancestor of some STs only or were secondarily lost (Figure 7). We not only identified homologs of ST1 LGT candidates in ST4 and ST7, but we also investigated the draft genomes of ST2, ST3, ST6, ST8, and ST9. One gene family (encoding the O-methyltransferase involved in polyketide biosynthesis) is only found in ST1 and ST2, suggesting that it was laterally acquired after the divergence of their ancestor from other STs. The presence/absence patterns of genes acquired by LGT should be further investigated when complete genome sequences become available for the other STs.

Conclusions

We have shown that LGT has played an important role in the adaptation of *Blastocystis* to the gut. In some cases, laterally acquired genes encode enzymes that feed metabolites into pre-existing metabolic systems (e.g., aspartase and fumarate reduction), showcasing the role of LGT in forging novel biochemical pathways comprising ancestral and newly acquired enzymes. However, the majority of the identified LGTs appear to represent gain of completely new single-protein functions or entire metabolic pathways.

Our results are unexpected in several respects. First, although eukaryote-to-eukaryote LGT is often dismissed [20], it appears to have had a significant role in the evolution of *Blastocystis*. Among the cases we report are several genes transferred from animals (e.g., Txndc12 and β1,3GalT), which could represent acquisitions from host lineages of the parasite. Second, the fact that at least 2.5% of the protein-coding genes in *Blastocystis* ST1 are recent LGTs is noteworthy given that these organisms are not phagotrophs, nor are they known to harbor endosymbionts. Together with the wide range of bacterial and eukaryotic LGT donors, this fact strongly suggests that transfers occurred from free-living organisms or DNA from their environment. Finally, the discovery of LGTs specific to the *Blastocystis* lineage,

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

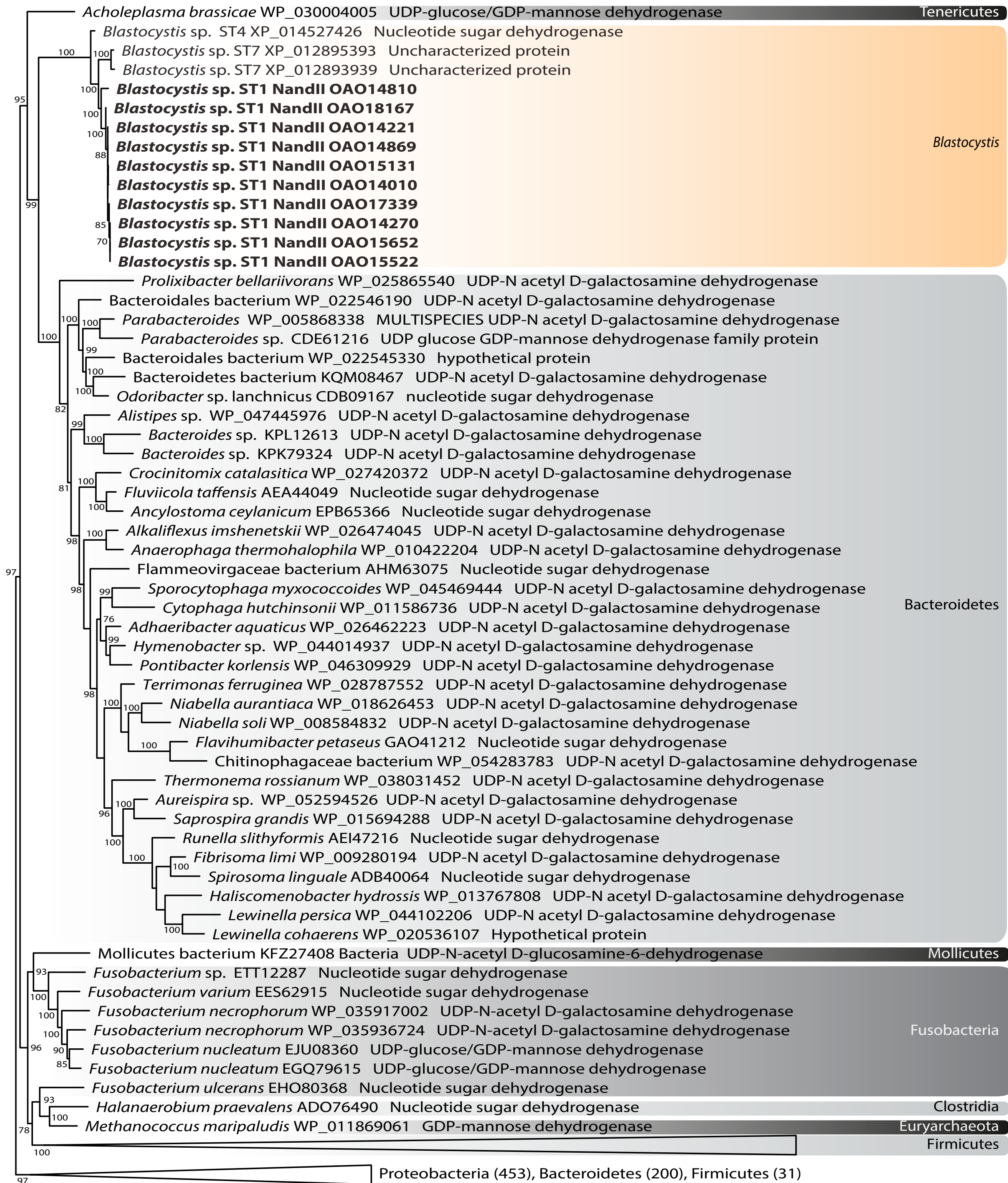


Figure 6. Maximum-Likelihood Phylogeny of UDP-Glucose/GDP-Mannose Dehydrogenase
See the legend from Figure 3.

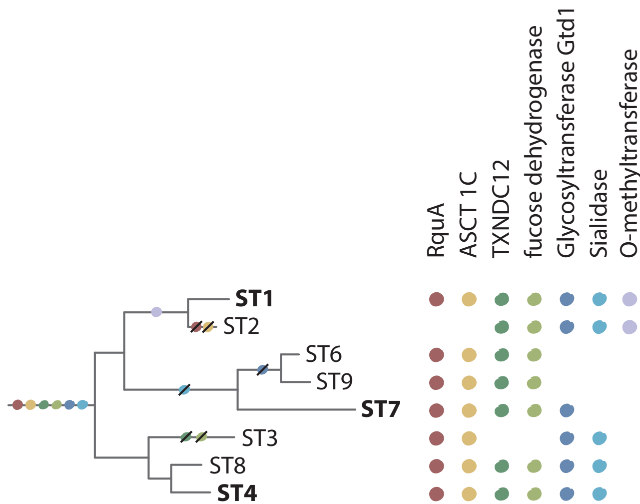


Figure 7. Differential Presence of Genes of LGT Origin in *Blastocystis* STs

Only STs for which complete (bolded) or draft genome data are available are represented. Colored and crossed-out circles on branches represent LGT events and gene losses, respectively. Only laterally acquired genes detected in ST1 and seemingly absent in one or more other subtypes are represented. Additional cases of genes acquired at the base of the lineage and subsequently lost in individual STs have been detected but have not been reported because the genome data for ST2, ST3, ST6, ST8, and ST9 are still incomplete.

including some that are restricted to particular STs, shows that LGT is an ongoing process in eukaryotic genomes, contrary to recent claims [10]. LGT clearly plays an important role in the evolution and adaptation of eukaryotic microbes.

Our results are also relevant to host-parasite interactions. The acquired effector proteins in *Blastocystis* most likely facilitate infection and evasion of host defenses. The genes involved in antigenic mimicry, if confirmed experimentally, will be the first example of such a mechanism in an intestinal protist. *Blastocystis* also most likely impacts the prokaryotic microbiome: its metabolic products, such as indole, and antibiotic polyketides might play a role in the dysbiosis and inflammation associated with specific cases of IBS [47]. Conversely, the gene repertoire of *Blastocystis* spp. holds clues that could explain the low prevalence of the parasite in Crohn's disease patients. *Blastocystis* has thus far proven recalcitrant to standard chemotherapeutic interventions [2]; we have identified a number of biochemical pathways that could represent new directions for drug development. Collectively, our results open up new avenues of research for better understanding the roles of *Blastocystis* in intestinal health and disease.

EXPERIMENTAL PROCEDURES

Protein Family Dataset Assembly

The 6,544 *Blastocystis* sp. ST1 predicted proteins (GenBank BioProject: PRJNA308101) were used as queries for BLASTp searches against the *nr* database (as of January 2016), with an e value cutoff of 1×10^{-10} and a maximum of 2,000 hits. Datasets with >50% overlap in sequence content were considered to represent close paralogs and merged.

Phylogenetic Reconstruction

We first reduced the dataset size by removing similar sequences, thereby allowing more rigorous downstream analyses. We split each dataset into subsets containing homologs from (1) Bacteria, (2) Archaea, (3) Archaeplastida, (4) Fungi, (5) Metazoa, and (6) all other lineages. From these, we selected representative sequences with the cd-hit program [63] using an identity threshold of 90% for groups 1–3 and 95% for groups 4, 5, and 6. Subset 6 was left untouched in order to maximize the representation of sequences from protists. All six subsets were then concatenated into a single file and aligned using mafft v6.903b [64]. Regions of doubtful alignment were removed using block mapping and gathering with entropy (BMGE) using the BLOSUM30 substitution matrix and a block size of 4 [65]. Maximum-likelihood (ML) phylogenetic analyses were then performed using IQtree v. 1.4.4 [66] under the LG4X substitution model with 1,000 ultrafast bootstraps.

Phylogenetic Screening for LGT

We developed an automated "pipeline" to phylogenetically screen for LGT using the following criteria. First, candidate LGTs in *Blastocystis* sp. ST1 cannot branch sister to homologs from any stramenopile (with the exception of other *Blastocystis* STs) or alveolate. If there are stramenopile or alveolate homologs elsewhere in the tree, they must be separated from *Blastocystis* by at least two bipartitions supported by >70% bootstrap support. Given that the topologies were unrooted, the above test was conducted for all possible root placements.

The 125 LGT candidates obtained using the above criteria were manually examined. We used a stringent approach, taking into account information contained in BLAST results, multiple sequence alignments, sequence domain composition (see below), and phylogeny. For example, we excluded some candidates with extensive sequence repeats, making the corresponding alignments dubious. We also discarded cases where the *Blastocystis* sequences were too divergent from their homologs to confidently determine their evolutionary origins. All remaining candidates were verified to be present in both genomic and transcriptomic data and screened for introns (Table S1), and flanking genes on scaffolds were confirmed to be of vertical descent. After examination, we report 74 clear transfer events followed by duplications, corresponding to 167 genes in *Blastocystis* sp. ST1. Orthologs from *Blastocystis* ST4 and 7 were also included in our phylogenetic analyses. Presence of homologs in ST2, ST3, ST6, ST8, and ST9 was established using TblastN against the draft genome scaffolds; these were not included in our trees because gene models were not available.

Functional Annotation and Predicted Cellular Localization

KEGG annotations for *Blastocystis* genes of interest were retrieved with BlastKOALA [67]. Functional domains for all homologs in each dataset were identified using HMMER 3 searches [68] against the PFAM 26.0 database [69] and were mapped onto phylogenetic trees with the ETE2 Python toolkit [70]. Cellular localizations were predicted using TargetP 1.1 [71], SignalP 4.1 [72], and MitoFates [73]. Transmembrane domains were predicted using the TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). All are reported in Table S1.

Phylogenies and additional data and discussion can be found at http://emelaura.com/data/.

ACCESSION NUMBERS

The accession number for the genome assembly reported in this paper is GenBank: LXWW00000000.1. Phylogenies have been deposited to Mendeley Data at http://dx.doi.org/10.17632/pktp3hggf7.1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.02.003.

AUTHOR CONTRIBUTIONS

Conceptualization, L.E., E.G., and A.J.R.; Methodology and Software, L.E. and B.C.; Investigation and Data Curation, L.E.; Writing – Original Draft, L.E.;

Writing – Review & Editing, A.J.R., E.G., B.C., and J.M.A.; Funding Acquisition, A.J.R. and J.M.A.; Resources, A.J.R. and J.M.A.; Supervision, A.J.R.

ACKNOWLEDGMENTS

This work was supported by operating grant MOP-142349 from the Canadian Institutes of Health Research (CIHR) awarded to A.J.R. and a Discovery Grant (RGPIN-2014-05871) from the Natural Sciences and Engineering Research Council of Canada to J.M.A. A.J.R. acknowledges support from the Canada Research Chairs Program. E.G. wishes to thank the Human Gut Microbiome for Health research unit for their support. A.J.R. and J.M.A. acknowledge support from the Canadian Institute for Advanced Research Program in Integrated Microbial Biodiversity. We wish to thank the anonymous reviewers for their constructive comments. We apologize to colleagues whose work could not be cited because of space constraints.

Received: December 15, 2016
Revised: January 26, 2017
Accepted: February 1, 2017
Published: March 2, 2017

REFERENCES

1. Clark, C.G., van der Giezen, M., Alfellani, M.A., and Stensvold, C.R. (2013). Recent developments in Blastocystis research. Adv. Parasitol. *82*, 1–32.
2. Stensvold, C.R., and Clark, C.G. (2016). Current status of Blastocystis: a personal view. Parasitol. Int. *65* (6 Pt B), 763–771.
3. Denoeud, F., Roussel, M., Noel, B., Wawrzyniak, I., Da Silva, C., Diogon, M., Viscogliosi, E., Brochier-Armanet, C., Couloux, A., Poulain, J., et al. (2011). Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite. Genome Biol. *12*, R29.
4. Wawrzyniak, I., Courtine, D., Osman, M., Hubans-Pierlot, C., Cian, A., Nourrisson, C., Chabe, M., Poirier, P., Bart, A., Polonais, V., et al. (2015). Draft genome sequence of the intestinal parasite Blastocystis subtype 4-isolate WR1. Genom. Data *4*, 22–23.
5. Andersen, L.O.B., Bonde, I., Nielsen, H.B.H.B., and Stensvold, C.R. (2015). A retrospective metagenomics approach to studying Blastocystis. FEMS Microbiol. Ecol. *91*, fiv072.
6. Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753–1756.
7. Soanes, D., and Richards, T.A. (2014). Horizontal gene transfer in eukaryotic plant pathogens. Annu. Rev. Phytopathol. *52*, 583–614.
8. Alsmark, C., Foster, P.G., Sicheritz-Ponten, T., Nakjang, S., Martin Embley, T., and Hirt, R.P. (2013). Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biol. *14*, R19.
9. Ku, C., Nelson-Sathi, S., Roettger, M., Garg, S., Hazkani-Covo, E., and Martin, W.F. (2015). Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. Proc. Natl. Acad. Sci. USA *112*, 10139–10146.
10. Ku, C., and Martin, W.F. (2016). A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. BMC Biol. *14*, 89.
11. Klimeš, V., Gentekaki, E., Roger, A.J., and Eliáš, M. (2014). A large number of nuclear genes in the human parasite blastocystis require mRNA polyadenylation to create functional termination codons. Genome Biol. Evol. *6*, 1956–1961.
12. Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The long-term stability of the human gut microbiota. Science *341*, 1237439.
13. Chandramathi, S., Suresh, K., Shuba, S., Mahmood, A., and Kuppusamy, U.R. (2010). High levels of oxidative stress in rats infected with Blastocystis hominis. Parasitology *137*, 605–611.
14. Kalapos, M.P., Desai, K.M., and Wu, L. (2010). Methylglyoxal, oxidative stress, and aging. In Aging and Age-Related Disorders, S. Bondy and K. Maiese, eds. (Humana), pp. 149–167.
15. Campbell, A.K., Matthews, S.B., Vassel, N., Cox, C.D., Naseem, R., Chaichi, J., Holland, I.B., Green, J., and Wann, K.T. (2010). Bacterial metabolic 'toxins': a new mechanism for lactose and food intolerance, and irritable bowel syndrome. Toxicology *278*, 268–276.
16. Liyanage, H., Kashket, S., Young, M., and Kashket, E.R. (2001). Clostridium beijerinckii and Clostridium difficile detoxify methylglyoxal by a novel mechanism involving glycerol dehydrogenase. Appl. Environ. Microbiol. *67*, 2004–2010.
17. Davis, P.H., Chen, M., Zhang, X., Clark, C.G., Townsend, R.R., and Stanley, S.L., Jr. (2009). Proteomic comparison of Entamoeba histolytica and Entamoeba dispar and the role of E. histolytica alcohol dehydrogenase 3 in virulence. PLoS Negl. Trop. Dis. *3*, e415.
18. Liu, F., Rong, Y.P., Zeng, L.C., Zhang, X., and Han, Z.G. (2003). Isolation and characterization of a novel human thioredoxin-like gene hTLP19 encoding a secretory protein. Gene *315*, 71–78.
19. Hugdahl, M.B., Beery, J.T., and Doyle, M.P. (1988). Chemotactic behavior of Campylobacter jejuni. Infect. Immun. *56*, 1560–1566.
20. Pickard, J.M., Maurice, C.F., Kinnebrew, M.A., Abt, M.C., Schenten, D., Golovkina, T.V., Bogatyrev, S.R., Ismagilov, R.F., Pamer, E.G., Turnbaugh, P.J., and Chervonsky, A.V. (2014). Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. Nature *514*, 638–641.
21. Bry, L., Falk, P.G., Midtvedt, T., and Gordon, J.I. (1996). A model of host-microbial interactions in an open mammalian ecosystem. Science *273*, 1380–1383.
22. Hooper, L.V., and Gordon, J.I. (2001). Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology *11*, 1R–10R.
23. Day, C.J., Tiralongo, J., Hartnell, R.D., Logue, C.A., Wilson, J.C., von Itzstein, M., and Korolik, V. (2009). Differential carbohydrate recognition by Campylobacter jejuni strain 11168: influences of temperature and growth conditions. PLoS ONE *4*, e4927.
24. Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V., and Gordon, J.I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science *299*, 2074–2076.
25. Stahl, M., Friis, L.M., Nothaft, H., Liu, X., Li, J., Szymanski, C.M., and Stintzi, A. (2011). L-fucose utilization provides Campylobacter jejuni with a competitive advantage. Proc. Natl. Acad. Sci. USA *108*, 7194–7199.
26. Ruiz-Palacios, G.M., Cervantes, L.E., Ramos, P., Chavez-Munguia, B., and Newburg, D.S. (2003). Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. J. Biol. Chem. *278*, 14112–14120.
27. Lanuza, M.D., Carbajal, J.A., and Borrás, R. (1996). Identification of surface coat carbohydrates in Blastocystis hominis by lectin probes. Int. J. Parasitol. *26*, 527–532.
28. Amaya, K.R., Kocherginskaya, S.A., Mackie, R.I., and Cann, I.K.O. (2005). Biochemical and mutational analysis of glutamine synthetase type III from the rumen anaerobe Ruminococcus albus 8. J. Bacteriol. *187*, 7481–7491.
29. De Ingeniis, J., Kazanov, M.D., Shatalin, K., Gelfand, M.S., Osterman, A.L., and Sorci, L. (2012). Glutamine versus ammonia utilization in the NAD synthetase family. PLoS ONE *7*, e39115.
30. Rodionova, I.A., Schuster, B.M., Guinn, K.M., Sorci, L., Scott, D.A., Li, X., Kheterpal, I., Shoen, C., Cynamon, M., Locher, C., et al. (2014). Metabolic and bactericidal effects of targeted suppression of NadD and NadE enzymes in mycobacteria. MBio *5*, e00747–13.
31. Moro, W.B., Yang, Z., Kane, T.A., Brouillette, C.G., and Brouillette, W.J. (2009). Virtual screening to identify lead inhibitors for bacterial NAD synthetase (NADs). Bioorg. Med. Chem. Lett. *19*, 2001–2005.
32. Velu, S.E., Mou, L., Luan, C.-H., Yang, Z.W., DeLucas, L.J., Brouillette, C.G., and Brouillette, W.J. (2007). Antibacterial nicotinamide adenine

dinucleotide synthetase inhibitors: amide- and ether-linked tethered dimers with α-amino acid end groups. J. Med. Chem. *50*, 2612–2621.

33. Magni, G., Di Stefano, M., Orsomando, G., Raffaelli, N., and Ruggieri, S. (2009). NAD(P) biosynthesis enzymes as potential targets for selective drug design. Curr. Med. Chem. *16*, 1372–1390.
34. Roldán, M.D., Pérez-Reinado, E., Castillo, F., and Moreno-Vivián, C. (2008). Reduction of polynitroaromatic compounds: the bacterial nitroreductases. FEMS Microbiol. Rev. *32*, 474–500.
35. De Oliveira, I.M., Bonatto, D., Antonio, J., and Henriques, J.A.P. (2010). Nitroreductases: enzymes with environmental, biotechnological and clinical importance. In Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, A. Mendez-Vilas, ed. (Formatex Research Center), pp. 1008–1019.
36. Leitsch, D., Burgess, A.G., Dunn, L.A., Krauer, K.G., Tan, K., Duchêne, M., Upcroft, P., Eckmann, L., and Upcroft, J.A. (2011). Pyruvate:ferredoxin oxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in Giardia lamblia. J. Antimicrob. Chemother. *66*, 1756–1765.
37. Lantsman, Y., Tan, K.S.W., Morada, M., and Yarlett, N. (2008). Biochemical characterization of a mitochondrial-like organelle from Blastocystis sp. subtype 7. Microbiology *154*, 2757–2766.
38. Mirza, H., Teo, J.D.W., Upcroft, J., and Tan, K.S.W. (2011). A rapid, high-throughput viability assay for Blastocystis spp. reveals metronidazole resistance and extensive subtype-dependent variations in drug susceptibilities. Antimicrob. Agents Chemother. *55*, 637–648.
39. Isshiki, S., Togayachi, A., Kudo, T., Nishihara, S., Watanabe, M., Kubota, T., Kitajima, M., Shiraishi, N., Sasaki, K., Andoh, T., and Narimatsu, H. (1999). Cloning, expression, and characterization of a novel UDP-galactose:beta-N-acetylglucosamine beta1,3-galactosyltransferase (beta3Gal-T5) responsible for synthesis of type 1 chain in colorectal and pancreatic epithelia and tumor cells derived therefrom. J. Biol. Chem. *274*, 12499–12507.
40. Moran, A.P. (2008). Relevance of fucosylation and Lewis antigen expression in the bacterial gastroduodenal pathogen Helicobacter pylori. Carbohydr. Res. *343*, 1952–1965.
41. Yuki, N., Susuki, K., Koga, M., Nishimoto, Y., Odaka, M., Hirata, K., Taguchi, K., Miyatake, T., Furukawa, K., Kobata, T., and Yamada, M. (2004). Carbohydrate mimicry between human ganglioside GM1 and Campylobacter jejuni lipooligosaccharide causes Guillain-Barre syndrome. Proc. Natl. Acad. Sci. USA *101*, 11404–11409.
42. Wunder, C., Churin, Y., Winau, F., Warnecke, D., Vieth, M., Lindner, B., Zähringer, U., Mollenkopf, H.-J., Heinz, E., and Meyer, T.F. (2006). Cholesterol glucosylation promotes immune evasion by Helicobacter pylori. Nat. Med. *12*, 1030–1038.
43. Kawakubo, M., Ito, Y., Okimura, Y., Kobayashi, M., Sakura, K., Kasama, S., Fukuda, M.N., Fukuda, M., Katsuyama, T., and Nakayama, J. (2004). Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection. Science *305*, 1003–1006.
44. Buisine, M.P., Desreumaux, P., Leteurtre, E., Copin, M.C., Colombel, J.F., Porchet, N., and Aubert, J.P. (2001). Mucin gene expression in intestinal epithelial cells in Crohn's disease. Gut *49*, 544–551.
45. Petersen, A.M., Stensvold, C.R., Mirsepasi, H., Engberg, J., Friis-Møller, A., Porsbo, L.J., Hammerum, A.M., Nordgaard-Lassen, I., Nielsen, H.V., and Krogfelt, K.A. (2013). Active ulcerative colitis associated with low prevalence of Blastocystis and Dientamoeba fragilis infection. Scand. J. Gastroenterol. *48*, 638–639.
46. Wang, H., Sivonen, K., and Fewer, D.P. (2015). Genomic insights into the distribution, genetic diversity and evolution of polyketide synthases and nonribosomal peptide synthetases. Curr. Opin. Genet. Dev. *35*, 79–85.
47. Poirier, P., Wawrzyniak, I., Vivarès, C.P., Delbac, F., and El Alaoui, H. (2012). New insights into Blastocystis spp.: a potential link with irritable bowel syndrome. PLoS Pathog. *8*, e1002545.
48. Decker, H., Motamedi, H., and Hutchinson, C.R. (1993). Nucleotide sequences and heterologous expression of tcmG and tcmP, biosynthetic genes for tetracenomycin C in Streptomyces glaucescens. J. Bacteriol. *175*, 3876–3886.
49. Florova, G., Kazanina, G., and Reynolds, K.A. (2002). Enzymes involved in fatty acid and polyketide biosynthesis in Streptomyces glaucescens: role of FabH and FabD and their acyl carrier protein specificity. Biochemistry *41*, 10462–10471.
50. Bao, W., Wendt-Pienkowski, E., and Hutchinson, C.R. (1998). Reconstitution of the iterative type II polyketide synthase for tetracenomycin F2 biosynthesis. Biochemistry *37*, 8132–8138.
51. Li, G., and Young, K.D. (2013). Indole production by the tryptophanase TnaA in Escherichia coli is determined by the amount of exogenous tryptophan. Microbiology *159*, 402–410.
52. Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-gut microbiota metabolic interactions. Science *336*, 1262–1267.
53. Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S., van der Giezen, M., Clark, C.G., and Roger, A.J. (2008). Organelles in Blastocystis that blur the distinction between mitochondria and hydrogenosomes. Curr. Biol. *18*, 580–585.
54. Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D.M., van der Giezen, M., and Roger, A.J. (2014). A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. Curr. Biol. *24*, 1176–1186.
55. Willis, R.C., and Woolfolk, C.A. (1974). Asparagine utilization in Escherichia coli. J. Bacteriol. *118*, 231–241.
56. Guccione, E., Leon-Kempis, Mdel.R., Pearson, B.M., Hitchin, E., Mulholland, F., van Diemen, P.M., Stevens, M.P., and Kelly, D.J. (2008). Amino acid-dependent growth of Campylobacter jejuni: key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate. Mol. Microbiol. *69*, 77–93.
57. Velayudhan, J., Jones, M.A., Barrow, P.A., and Kelly, D.J. (2004). L-serine catabolism via an oxygen-labile L-serine dehydratase is essential for colonization of the avian gut by Campylobacter jejuni. Infect. Immun. *72*, 260–268.
58. Atteia, A., van Lis, R., Tielens, A.G.M., and Martin, W.F. (2013). Anaerobic energy metabolism in unicellular photosynthetic eukaryotes. Biochim. Biophys. Acta *1827*, 210–223.
59. Mohamedmohaideen, N.N., Palaninathan, S.K., Morin, P.M., Williams, B.J., Braunstein, M., Tichy, S.E., Locker, J., Russell, D.H., Jacobs, W.R., Jr., and Sacchettini, J.C. (2008). Structure and function of the virulence-associated high-temperature requirement A of Mycobacterium tuberculosis. Biochemistry *47*, 6092–6102.
60. Maria-Neto, S., Cândido, Ede.S., Rodrigues, D.R., de Sousa, D.A., da Silva, E.M., de Moraes, L.M.P., Otero-Gonzalez, Ade.J., Magalhães, B.S., Dias, S.C., and Franco, O.L. (2012). Deciphering the magainin resistance process of Escherichia coli strains in light of the cytosolic proteome. Antimicrob. Agents Chemother. *56*, 1714–1724.
61. Ma'ayeh, S.Y., and Brook-Carter, P.T. (2012). Representational difference analysis identifies specific genes in the interaction of Giardia duodenalis with the murine intestinal epithelial cell line, IEC-6. Int. J. Parasitol. *42*, 501–509.
62. Sarkar, D. (2012). Mycobacterial glycolipids: pathways to synthesis and role in virulence. PhD thesis (University of Birmingham).
63. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.
64. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.
65. Criscuolo, A., and 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.

Please cite this article in press as: Eme et al., Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite *Blastocystis* to the Gut, Current Biology (2017), http://dx.doi.org/10.1016/j.cub.2017.02.003

66. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
67. Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. *428*, 726–731.
68. Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. *39*, W29–W37.
69. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.
70. Huerta-Cepas, J., Dopazo, J., and Gabaldón, T. (2010). ETE: a python environment for tree exploration. BMC Bioinformatics *11*, 24.
71. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. *300*, 1005–1016.
72. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.
73. Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol. Cell. Proteomics *14*, 1113–1126.