

Supplemental Text.

Informational genes undergo inter-domain lateral transfer

The majority of LGTs are involved in metabolism, in agreement with the complexity hypothesis which suggests that operational genes (e.g., involved in metabolism) are more likely to be transferred than informational genes (i.e., involved in the maintenance and transfer of genetic information) (1,2). However, we were able to identify a few interdomain lateral transfers of genes involved in informational processes. We have identified the LGT from bacteria to *Blastocystis* of RuvC (OAO16804.1) and YebC (OAO17975.1), two proteins involved in Holliday junctions resolution. Holliday junctions are DNA structures that are formed during homologous recombination – a fundamental cellular process mostly used to rearrange genes and accurately repair DNA double-strand breaks. These four-way DNA junctions need to be resolved to allow proper chromosome segregation. While Holliday junction resolvases have been described in the three domains of life, these systems are not necessarily homologous. RuvC represents a resolvase family present in many bacterial lineages but usually absent from eukaryotic genomes (3,4). It has, in fact, been suggested that the few RuvC eukaryotic homologues have probably been acquired from bacteria via horizontal gene transfer (3). In addition, analyses in bacteria revealed a transcription regulator YebC gene in the neighbourhood of the resolvase RuvC and have suggested that YebC might be involved in DNA recombination and repair (5,6). Interestingly, it has been shown that the intestinal bacterial pathogen *Helicobacter pylori* mutated for *ruvC* was dramatically more sensitivity to antimicrobial drugs such as levofloxacin and metronidazole (7). These mutants also showed to be more susceptible to oxidative stress and exhibited lower virulence. Infection was spontaneously cleared from the gastric mucosa, raising the possibility using *ruvC* as a therapeutic target – a possibility which could also be explored for *Blastocystis*.

Colonization of the gut environment by *Blastocystis* was facilitated by LGT

Oxygen-stress response

We identified a YfeX-like peroxidase (OAO17542.1), a member of the DyP family of peroxidases class B (8), likely acquired from gamma-proteobacteria. The gene piqued our interest for several reasons: (i) it is also found in other eukaryotes, such as the protists *Naegleria*,

Acanthamoeba, *Paramecium* and dictyostelids, but also in some fungi, platyhelminths and molluscs, the vast majority of which have been described to experience low-oxygen conditions (9,10); (ii) many of these homologues are predicted to be targeted to mitochondrion-related organelles (not shown), including the ones in *Blastocystis*; (iii) it has been shown to be upregulated in anaerobic conditions in *Yersinia* (11) and in response to post-infection oxidative stress in *Francisella tularensis* (12); and (iv) it is a heme-containing peroxidase which, in *E. coli*, oxidizes protoporphyrinogen IX and coproporphyrinogen III during heme biosynthesis (13). The loss of heme biosynthesis is common in parasitic protists living in low-oxygen environment, such as *Giardia*, *Entamoeba*, and *Trichomonas*. Indeed, they have lost the majority of their mitochondrial functions, including heme-dependent respiratory complexes (14,15). However, genome analyses most anaerobic parasites have revealed that they retain several genes encoding hemoproteins, suggesting that they are able to import heme, although the mechanism for this is still largely unknown (16). Only two heme-containing proteins have so far been reported in *Blastocystis* (16) and the one identified here comes to complete this list. Since *Blastocystis* does not synthesize protoporphyrinogen IX and coproporphyrinogen III, the substrate of this laterally acquired peroxidase remains to be identified. However, the evidence discussed above suggests that the acquisition of this enzyme is linked to the adaptation to low-oxygen conditions and oxygen-stress resistance.

Carbohydrate scavenging and metabolism

In addition to the fucose pathway, we identified a dozen other genes involved in scavenging and metabolism of carbohydrates, most of bacterial origin. These include a galactose-proton symporter (OAO12387.1), a beta galactosidase (OAO13865.1) and a galactose mutarotase (OAO14790.1); a phosphotransferase sugar transporter subunit IID (OAO12766.1); a hexokinase (OAO14075.1), a phosphoglycerate mutase (OAO12075.1) – an enzyme that catalyzes step 8 of glycolysis; and a galactarate dehydratase (OAO15613.1) which degrades galacturonate into pyruvate and glyceraldehyde 3-phosphate, two intermediates in several central metabolic pathways. We also recovered a chitinase of possible metazoan origin (OAO12908.1). Chitinases hydrolyse chitin, a very abundant polysaccharide found in the fungal cell wall, the exoskeletons of arthropods and the cysts of various protozoans. Chitinases might also serve as virulence

factors targeting glycoproteins and glycolipids of host organisms that contain N-acetylglucosamine, the monomer from which chitin is synthesized (17).

We also noticed three enzymes involved in the metabolism of N-acetylneuraminic acid, the predominant sialic acid found in mammalian cells. It is a particularly important component of mucins. We detected the LGT of a sialidase (OAO16540.1), catalyzing the cleavage of a terminal sialic acid residue from the rest of the oligosaccharide. We also identified a fused epimerase/kinase bifunctional enzyme NanK/NanE (OAO16379.1), and a N-acetylneuraminate lyase NanA (OAO16509.1), both which function in sequence in the metabolic pathway of sialic acid. N-acetylneuraminic acid can thus be used either as a nutrient, providing both carbon and nitrogen, or can be activated and placed on the cell surface, as its been shown to do in pathogenic bacteria (18).

Intracellular pH homeostasis

Enteric pathogens and commensals encounter a wide range of external pHs in the digestive tract fluctuating between 1 and 10 (43). These microbes thus require protective responses to maintain internal pH homeostasis. *Blastocystis* possesses a gene encoding the protein NhaC, a Na⁺/H⁺ antiporter (OAO13479.1) potentially involved in this response, which was likely acquired from a firmicute donor. In bacteria, Na⁺/H⁺ antiporters exchange intracellular Na⁺, Li⁺, or tetracycline for extracellular protons, either to maintain pH or to reduce their toxic effects (44). These antiporters have been implicated in the response to alkaline stress in some bacteria (45) and upregulated in the presence of fucose in others (46), which possibly serves as a signal that the parasite is entering the gut. This Na⁺/H⁺ antiporter might also be important to maintain homeostasis and the proper function of FucP, which acts as a fucose-proton symporter (47).

Nitrogen metabolism

We found three amino-acid ammonia-lyases of bacterial origin: an aspartate ammonia-lyase targeted to the mitochondrial-related organelle (OAO16085.1), a serine ammonia lyase (OAO12304.1), and a tryptophanase (OAO13933.1). These enzymes of bacterial origin turn aspartate into fumarate and ammonia, serine into pyruvate and ammonia, and tryptophan into indole, pyruvate, and ammonia, respectively (Figure 2D). A fourth ammonia-lyase was acquired

horizontally probably from opisthokonts: formimidoyltransferase cyclodeaminase (OAO16360.1) participates in folate metabolism and turns 5-formimidoyltetrahydrofolate into 5,10-methenyltetrahydrofolate and NH_3 (19). Along with these, we identified a cytidine deaminase (OAO17373.1) of probable gamma-proteobacterial origin, an enzyme that catalyzes generation of uridine and NH_3 from cytidine. The *Blastocystis* homologues correspond to the homodimeric form that are found in land plants and some proteobacteria (<http://www.ebi.ac.uk/interpro/entry/IPR006263>), whereas a related but homotetrameric form with a much smaller subunit is found in most bacteria and in animals.

As mentioned briefly in the main text, we identified the horizontal transfer of a bacterial gene belonging to the multidrug DMT transporter superfamily (OAO17832.1). More specifically, most *Blastocystis* homologues and their close bacterial relatives are predicted to contain ureide permease and sugar transporter domains. Ureide permeases represent a family of transporters of a wide spectrum of derivatives of heterocyclic nitrogen compounds, including allantoin, uric acid and xanthine (20). Ureide permeases likely transport other purine degradation products when nitrogen sources are low (20). A search for this domain in Genbank reveals that this family is almost exclusively found in plants and bacteria (mostly in Bacteroidetes and α -proteobacteria), confirming the acquisition of this domain by LGT. In addition, four copies of this gene can be found in the *Blastocystis* genome, suggesting the importance of this transporter for supplying the parasite's need in nitrogen compounds.

Anaerobic metabolism

Serine metabolism

Blastocystis acquired a bacterial L-serine dehydratase (OAO12304.1), which converts serine to pyruvate and ammonia (Figure 2D). Interestingly, while most serine dehydratases from aerobic bacteria are pyridoxal 5'-phosphate (PLP)-dependent enzymes, the L-serine dehydratase in *Blastocystis*, is devoid of PLP, like in *Campylobacter* (21) and other anaerobic bacteria (22,23). Instead, it is an oxygen-sensitive iron-sulfur cluster containing enzyme which can be reactivated by Fe^{2+} under anaerobic conditions (21).

NAD⁺/NADH

As discussed above, nicotinamide adenine dinucleotide (NAD(P)H) is an important coenzyme that needs to be regenerated in its oxidized form (NAD(P)). However, in the host intestine, this has to happen in the challenging absence of oxygen. A compensating mechanism was found in *Blastocystis* through the horizontal acquisition from bacteria of four genes capable of regenerating oxidized NAD: alcohol dehydrogenase YqhD, alcohol dehydrogenase GldA (both mentioned in the section "Oxygen-stress resistance"), NAD-dependant aldehyde reductase/alcohol dehydrogenase AdhA (OAO18004.1), and UDP glucose dehydrogenase (OAO14810.1, see section "Immune response evasion, virulence and pathogenicity"). The acquisition of oxygen-independent bacterial-type dehydrogenases is not uncommon in intestinal parasites and occurred several times independently, such as in *Giardia lamblia* (24,25), *Entamoeba histolytica* (26) and *Cryptosporidium* (27).

Sulfur relay

A particularly interesting LGT case is the one of a DsrE-like protein (OAO12114.1), acquired from bacteria (Fig. X). DsrE is part of a protein superfamily together with two sulfur transferases YchN and TusD. Like these, it harbors two conserved cysteine residues in a C-X-X-C motif (28) and all share the same three-dimensional fold; they are thought to be part of sulfur relay systems (29). In particular, YchN has been proposed to play a sulfur transfer role downstream of the CSD system, a bacterial iron-sulfur (Fe-S) cluster biogenesis pathway (30). Phylogenetic analyses show that the DsrE sequence from *Blastocystis* is forming a monophyletic clade with two homologues from the anaerobic breviate *Pygsuia biforma*, suggesting a single LGT event from bacteria followed by a second transfer event between these two lineages. This is particularly interesting given the recent discovery of a unique Fe-S cluster biosynthesis system in eukaryotes: both *Blastocystis* and *Pygsuia biforma* possess a fused SufC/SufB protein (AFL03357.1 in ST1 Nand II) that has been laterally acquired during a single transfer event from methanoarchaea, followed by a second LGT between these eukaryote lineages (31,32). Little is known about the way the SUF system functions in these anaerobic protists, especially since they lack other SUF components reported in prokaryotes. Notably they lack SufS-SufE which are known to transfer sulfur to SufB in bacteria. Our results now pose the question of a sulfur transferase role of DrsE and its putative interaction with SufCB in *Blastocystis*.

References

1. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A*. 1999;96(7):3801–6.
2. Ford Doolittle W. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Vol. 14, *Trends in Genetics*. 1998. p. 307–11.
3. Aravind L, Makarova KS, Koonin E V. SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res*. 2000;28(18):3417–32.
4. Wyatt HDM, West SC. Holliday junction resolvases. *Cold Spring Harb Perspect Biol*. 2014;6(9):a023192.
5. Galperin MY, Koonin E V. “Conserved hypothetical” proteins: Prioritization of targets for experimental study. Vol. 32, *Nucleic Acids Research*. 2004. p. 5452–63.
6. Liang H, Li N, Dong Z, Surette MG, Duan K. The YebC family protein PA0964 negatively regulates the *Pseudomonas aeruginosa* quinolone signal system and pyocyanin production. *J Bacteriol*. 2008;190(18):6217–27.
7. Loughlin MF, Barnard FM, Jenkins D, Sharples GJ, Jenks PJ. *Helicobacter pylori* mutants defective in RuvC Holliday junction resolvase display reduced macrophage survival and spontaneous clearance from the murine gastric mucosa. *Infect Immun*. 2003;71(4):2022–31.
8. Yoshida T, Sugano Y. A structural and functional perspective of DyP-type peroxidase family. Vol. 574, *Archives of Biochemistry and Biophysics*. 2015. p. 49–55.
9. Kitching JA. The effects of a lack of oxygen and of low oxygen tensions on *Paramecium*. *Biol Bull. JSTOR*; 1939;77(3):339–53.
10. Stairs CW, Leger MM, Roger AJ. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1678):20140326.
11. Babujee L, Balakrishnan V, Kiley PJ, Glasner JD, Perna NT. Transcriptome Changes Associated with Anaerobic Growth in *Yersinia intermedia* (ATCC29909). *PLoS One*. 2013;8(10).
12. Wehrly TD, Chong A, Virtaneva K, Sturdevant DE, Child R, Edwards JA, et al. Intracellular biology and virulence determinants of *Francisella tularensis* revealed by transcriptional profiling inside macrophages. *Cell Microbiol*. 2009;11(7):1128–50.
13. Dailey HA, Septer AN, Daugherty L, Thames D, Gerdes S, Stabb E V., et al. The *Escherichia coli* protein YfeX functions as a porphyrinogen oxidase, not a heme dechelataase. *MBio*. 2011;2(6).
14. Jedelský PL, Doležal P, Rada P, Pyrih J, Smíd O, Hrdý I, et al. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS One*. 2011 Jan;6(2):e17285.
15. Rada P, Doležal P, Jedelsky PL, Bursac D, Perry AJ, Šedinová M, et al. The core

- components of organelle biogenesis and membrane transport in the hydrogenosomes of *Trichomonas vaginalis*. PLoS One. Public Library of Science; 2011;6(9):e24428.
16. Kořený L, Oborník M, Lukeš J. Make It, Take It, or Leave It: Heme Metabolism of Parasites. Knoll LJ, editor. PLoS Pathog. 2013 Jan 17;9(1):e1003088.
 17. Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H, et al. Bacterial chitinases and chitin-binding proteins as virulence factors. Microbiol (United Kingdom). 2013;159(PART 5):833–47.
 18. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens. Vol. 153, Microbiology. 2007. p. 2817–22.
 19. Solans A, Estivill X, de La Luna S. Cloning and characterization of human FTCD on 21q22. 3, a candidate gene for glutamate formiminotransferase deficiency. Cytogenet Genome Res. Karger Publishers; 2000;88(1–2):43–9.
 20. Desimone M, Catoni E, Ludewig U, Hilpert M, Schneider A, Kunze R, et al. A novel superfamily of transporters for allantoin and other oxo derivatives of nitrogen heterocyclic compounds in Arabidopsis. Plant Cell. Am Soc Plant Biol; 2002;14(4):847–56.
 21. Velayudhan J, Jones MA, Barrow PA, Kelly DJ. L-Serine Catabolism via an Oxygen-Labile L-Serine Dehydratase Is Essential for Colonization of the Avian Gut by *Campylobacter jejuni*. Infect Immun. 2004;72(1):260–8.
 22. Grabowski R, Buckel W. Purification and properties of an iron-sulfur-containing and pyridoxal- phosphate-independent L-serine dehydratase from *Peptostreptococcus asaccharolyticus*. Eur J Biochem. 1991;199(1):89–94.
 23. Hofmeister a E, Grabowski R, Linder D, Buckel W. L-serine and L-threonine dehydratase from *Clostridium propionicum*. Two enzymes with different prosthetic groups. Eur J Biochem. 1993;215(2):341–9.
 24. Sánchez LB. Aldehyde dehydrogenase (CoA-acetylating) and the mechanism of ethanol formation in the amitochondriate protist, *Giardia lamblia*. Arch Biochem Biophys. 1998;354(1):57–64.
 25. Field J, Rosenthal B, Samuelson J. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. Mol Microbiol. 2000 Nov;38(3):446–55.
 26. Yang W, Li E, Kairong T, Stanley SL. *Entamoeba histolytica* has an alcohol dehydrogenase homologous to the multifunctional adhE gene product of *Escherichia coli*. Mol Biochem Parasitol. 1994;64(2):253–60.
 27. Huang J, Mullapudi N, Lancto C a, Scott M, Abrahamsen MS, Kissinger JC. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. Genome Biol. 2004 Jan;5(11):R88.
 28. Shin DH, Yokota H, Kim R, Kim SH. Crystal structure of a conserved hypothetical protein from *Escherichia coli*. J Struct Funct Genomics. 2002;2(1):53–66.
 29. Dahl C, Schulte A, Stockdreher Y, Hong C, Grimm F, Sander J, et al. Structural and Molecular Genetic Insight into a Widespread Sulfur Oxidation Pathway. J Mol Biol. 2008;384(5):1287–300.

30. Bolstad HM, Botelho DJ, Wood MJ. Proteomic Analysis of Protein - Protein Interactions within the Cysteine Sulfinase Desulfurase Fe - S Cluster Biogenesis System research articles. 2010;5358–69.
31. Tsaousis AD, Ollagnier S, Choudens D, Gentekaki E, Long S, Gaston D. Evolution of Fe / S cluster biogenesis in the anaerobic parasite Blastocystis. 2012;
32. Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, et al. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygmaia. Curr Biol. Elsevier Ltd; 2014;24(11):1176–86.
33. Denoed F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, Diogon M, et al. Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite. Genome Biol. BioMed Central Ltd; 2011 Jan;12(3):R29.