

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Loftus, B; Anderson, I; Davies, R; Alsmark, UC; Samuelson, J; Amedeo, P; Roncaglia, P; Berriman, M; Hirt, RP; Mann, BJ; +44 more... Nozaki, T; Suh, B; Pop, M; Duchene, M; Ackers, J; Tannich, E; Leippe, M; Hofer, M; Bruchhaus, I; Willhoeft, U; Bhattacharya, A; Chillingworth, T; Churcher, C; Hance, Z; Harris, B; Harris, D; Jagels, K; Moule, S; Mungall, K; Ormond, D; Squares, R; Whitehead, S; Quail, MA; Rabbinowitsch, E; Norbertczak, H; Price, C; Wang, Z; Guillan, N; Gilchrist, C; Stroup, SE; Bhattacharya, S; Lohia, A; Foster, PG; Sicheritz-Ponten, T; Weber, C; Singh, U; Mukherjee, C; El-Sayed, NM; Petri, WA; Clark, CG; Embley, TM; Barrell, B; Fraser, CM; Hall, N; (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature*, 433 (7028). pp. 865-8. ISSN 0028-0836 DOI: <https://doi.org/10.1038/nature03291>

Downloaded from: <http://researchonline.lshtm.ac.uk/13878/>

DOI: <https://doi.org/10.1038/nature03291>

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by-nc-sa/2.5/>

<https://researchonline.lshtm.ac.uk>

**The genome of the protist parasite
*Entamoeba histolytica***

Brendan Loftus¹, Iain Anderson¹, Rob Davies², U. Cecilia M. Alsmark³, John Samuelson⁴, Paolo Amedeo¹, Paola Roncaglia¹, Matt Berriman², Robert P. Hirt³, Barbara J. Mann⁵, Tomo Nozaki⁶, Bernard Suh¹, Mihai Pop¹, Michael Duchene⁷, John Ackers⁸, Egbert Tannich⁹, Matthias Leippe¹⁰, Margit Hofer⁷, Iris Bruchhaus⁹, Ute Willhoeft⁹, Alok Bhattacharya¹¹, Tracey Chillingworth², Carol Churcher², Zahra Hance², Barbara Harris², David Harris², Kay Jagels², Sharon Moule², Karen Mungall², Doug Ormond², Rob Squares², Sally Whitehead², Michael A. Quail², Ester Rabinowitsch², Halina Norbertczak², Claire Price², Zheng Wang¹, Nancy Guillén¹², Carol Gilchrist³, Suzanne E. Stroup³, Sudha Bhattacharya¹¹, Anuradha Lohia¹³, Peter G. Foster¹⁴, Thomas Sicheritz-Ponten¹⁵, Christian Weber¹², Upinder Singh¹⁶, Chandrama Mukherjee¹³, Najib M. El-Sayed¹, William A. Petri Jr³, C. Graham Clark⁸, T. Martin Embley³, Bart Barrell², Claire M. Fraser¹ & Neil Hall^{2*}

- ¹TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA
- ²The Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
- ³School of Biology, University of Newcastle, King George VI Building, Newcastle upon Tyne NE1 7RU, UK
- ⁴Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA
- ⁵Departments of Internal Medicine & Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA
- ⁶Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan
- ⁷Division of Specific Prophylaxis and Tropical Medicine, Center for Physiology and Pathophysiology, Medical University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria
- ⁸Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
- ⁹Department of Molecular Parasitology, Bernhard Nocht Institute for Tropical Medicine, Bernhard Nocht Str. 74, 20359 Hamburg, Germany
- ¹⁰Zoological Institute, University of Kiel, Olshausenstr. 40, 24098 Kiel, Germany
- ¹¹School of Environmental Sciences, Jawaharlal Nehru University, New Delhi 110067, India
- ¹²Unite de Biologie Cellulaire du Parasitisme, INSERM U389, Institut Pasteur 28, rue du Dr Roux 75724, Paris Cedex 15, France
- ¹³Department of Biochemistry, Bose Institute, P1/12 CIT Scheme VIIM, Kolkata 700054, India
- ¹⁴Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
- ¹⁵Center for Biological Sequence Analysis, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark
- ¹⁶Departments of Internal Medicine, Microbiology, and Immunology, Stanford University School of Medicine, Stanford, California 94305-5107, USA

* Present address: TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

Entamoeba histolytica is an intestinal parasite and the causative agent of amoebiasis, which is a significant source of morbidity and mortality in developing countries¹. Here we present the genome of *E. histolytica*, which reveals a variety of metabolic adaptations shared with two other amitochondrial protist pathogens: *Giardia lamblia* and *Trichomonas vaginalis*. These adaptations include reduction or elimination of most mitochondrial metabolic pathways and the use of oxidative stress enzymes generally associated with anaerobic prokaryotes. Phylogenomic analysis identifies evidence for lateral gene transfer of bacterial genes into the *E. histolytica* genome, the effects of which centre on expanding aspects of *E. histolytica*'s metabolic repertoire. The presence of these genes and the potential for novel metabolic pathways in *E. histolytica* may allow for the development of new chemotherapeutic agents. The genome encodes a large number of novel receptor kinases and contains expansions of a variety of

gene families, including those associated with virulence. Additional genome features include an abundance of tandemly repeated transfer-RNA-containing arrays, which may have a structural function in the genome. Analysis of the genome provides new insights into the workings and genome evolution of a major human pathogen.

Genome analysis was carried out on a 12.5-fold coverage genome assembly consisting of 23,751,783 base pairs (bp) distributed among 888 scaffolds. The 9,938 predicted genes average 1.17 kilobases (kb) in size and comprise 49% of the genome. One-quarter of *E. histolytica* genes are predicted to contain introns, with 6% of genes containing multiple introns. No homologues could be identified for a third of predicted proteins (31.8%) from the public databases (see Methods). *E. histolytica* chromosomes do not condense, and the uncertainty surrounding its ploidy and the extensive length variability observed between homologous chromosomes from different isolates makes the exact chromosome number difficult to determine. The chromosome size variation observed may be due to expansion and contraction of subtelomeric repeats, as in other protists^{2,3}, and it is tempting to speculate that in *E. histolytica* these regions consist of tRNA-containing arrays. Comprising almost 10% of the sequence reads, 25 types of long tandem array, each containing between one and five tRNA types per repeat unit, could be identified from the genome data. The full complement of tRNAs required for translation has been identified, and all but four of the tRNA genes are encoded exclusively in arrays. These unique tRNA gene arrays are thus predicted to be functional as well as potentially fulfilling a structural role in the genome. No association could be determined between codon usage and the relative copy numbers of their cognate tRNA species.

The metabolism of *E. histolytica* seems to have been shaped by secondary gene loss and lateral gene transfer (LGT), primarily from bacterial lineages (Fig. 1). *E. histolytica* is an obligate fermenter, using bacterial-like fermentation enzymes and lacking proteins of the tricarboxylic acid cycle and mitochondrial electron transport chain. An atrophic, mitochondrion-derived organelle has been identified in *E. histolytica*⁴, and the genome data support the absence of a mitochondrial genome. Glucose is the main energy source; however, in place of the typical eukaryotic glucose transporters those of *E. histolytica* are related to the prokaryote glucose/ribose porter family, with the amino- and carboxy-terminal domains switched relative to their prokaryotic counterparts.

As a phagocytic resident of the human gut, *E. histolytica* has access to many bacterial and host-derived preformed organic compounds. Most pathways for amino acid biosynthesis have been eliminated, except those for serine and cysteine, which are probably retained for the production of cysteine, the major intracellular thiol. The high levels of cysteine in *E. histolytica* may compensate for the lack of glutathione and its associated enzymes, a major component of oxidative stress resistance in many organisms⁵. *E. histolytica* lacks *de novo* purine, pyrimidine and thymidylate synthesis and must rely on salvage pathways, similar to *G. lamblia* and *T. vaginalis*⁶. In addition, *E. histolytica* appears to lack ribonucleotide reductase, a characteristic that it shares with *G. lamblia*⁷. *E. histolytica* is unable to synthesize fatty acids but retains the ability to synthesize a variety of phospholipids. The absence of identifiable pathways for the synthesis of isoprenoids and the sphingolipid head group aminoethylphosphonate suggest the existence of novel pathways. These pathways, once characterized, might represent attractive drug targets. Two unusual enzymes of fatty acid elongation are shared between *E. histolytica* and *G. lamblia*, including a predicted acetyl-CoA carboxylase with two carboxyltransferase domains⁸. We propose that this enzyme removes a carboxyl group from oxaloacetate and transfers it to acetyl-CoA to form malonyl-CoA and pyruvate. *E. histolytica* also has five members of a fatty acid elongase family, previously identified only in plants, green algae and *G. lamblia*^{9,10}. Folate is

a cofactor essential for thymidylate synthesis and methionine recycling, and genome analysis reveals a complete lack of genes coding for known folate-dependent enzymes and folate transporters. Folate is also required for organelle protein synthesis in mitochondria and chloroplasts, and loss of the mitochondrial genome may have paved the way for the loss of these folate-dependent functions.

LGT is an important force in the evolution of prokaryotes but significantly less is known about its importance in eukaryotic evolution¹¹. We conducted a phylogenetic screen of the *Entamoeba* genome for cases of relatively recent prokaryote to eukaryote LGT (see Methods), and for 96 genes we believe that this is the simplest explanation for the tree topologies obtained (see Supplementary Information). These genes are embedded among typically eukaryotic genes on *E. histolytica* scaffolds and do not seem to represent contaminating prokaryotic sequences. Most (58%) of the LGT genes encode a variety of metabolic enzymes, whereas most of the remaining genes (41%) encode proteins of unknown function (Supplementary Fig. 1). The major impact is in the area of carbohydrate and amino acid metabolism, where they have increased the range of substrates available for energy generation including tryptophanase and aspartase, which contribute to the use of amino acids. Several glycosidases and sugar kinases appear to have been acquired through LGT and would probably enable *E. histolytica* to use sugars other than glucose; for example, fructose and galactose. There is a strong bias in the data for a major donor being in the *Cytophaga-Flavobacterium-Bacteroides* (CFB) group of the phylum Bacteroidetes; however, this should be interpreted with caution, as current sampling of prokaryotic genomes is still relatively incomplete. It is clear that among the 96 genes, some result in significant enhancements to *E. histolytica* metabolism, thus contributing to its biology to a greater extent than indicated by the numbers alone.

E. histolytica feed on bacteria in the lumen of the colon and lyse host epithelial cells after invasion of the intestinal wall¹². A number of amoebic virulence determinants have been characterized, including a multi-subunit GalGalNAc lectin involved in adhesion to host cells, cysteine proteases that degrade host extracellular matrix, and pore-forming peptides (amoebapores) capable of lysing target cells¹². Analysis of the genome reveals redundancy in the genes encoding these virulence factors. Thirty homologues of the intermediate subunit and one homologue of the heavy subunit of the GalGalNAc lectin were identified. Ten new cysteine proteinases with predicted N-terminal transmembrane anchors, which might allow them to be localized on the amoeba cell surface, were identified. In addition to three new amoebapores a homologue of haemolysin III was identified, suggesting that, in addition to amoebapores, haemolysins may have a role in host cell lysis.

Vesicle trafficking has a role in *E. histolytica* pathogenesis through phagocytosis and the delivery of secreted hydrolytic enzymes and amoebapores to the cell surface¹³. *E. histolytica* lacks morphologically identifiable rough endoplasmic reticulum and the Golgi apparatus¹⁴ but encodes the basic elements of the vesicle transport machinery common to other eukaryotic cells, with the coat complexes COPI, COPII, clathrin and retromer all being present. Rab and Arf protein family expansions reflect the increased complexity and number of vesicle fusion and recycling steps that have been associated with phagocytosis and pinocytosis in amoebae¹⁵. The cytoskeleton has a number of important roles in parasite motility, contact-dependant killing and phagocytosis of host intestinal epithelial cells¹⁶. This is reflected in expansions of Rho GTPases and their regulators RhoGAPs and RhoGEFs, which control a number of processes involving the actin cytoskeleton. Five proteins with a unique domain architecture containing both RhoGEF and Arf-GAP domains were identified, suggesting a mechanism for direct

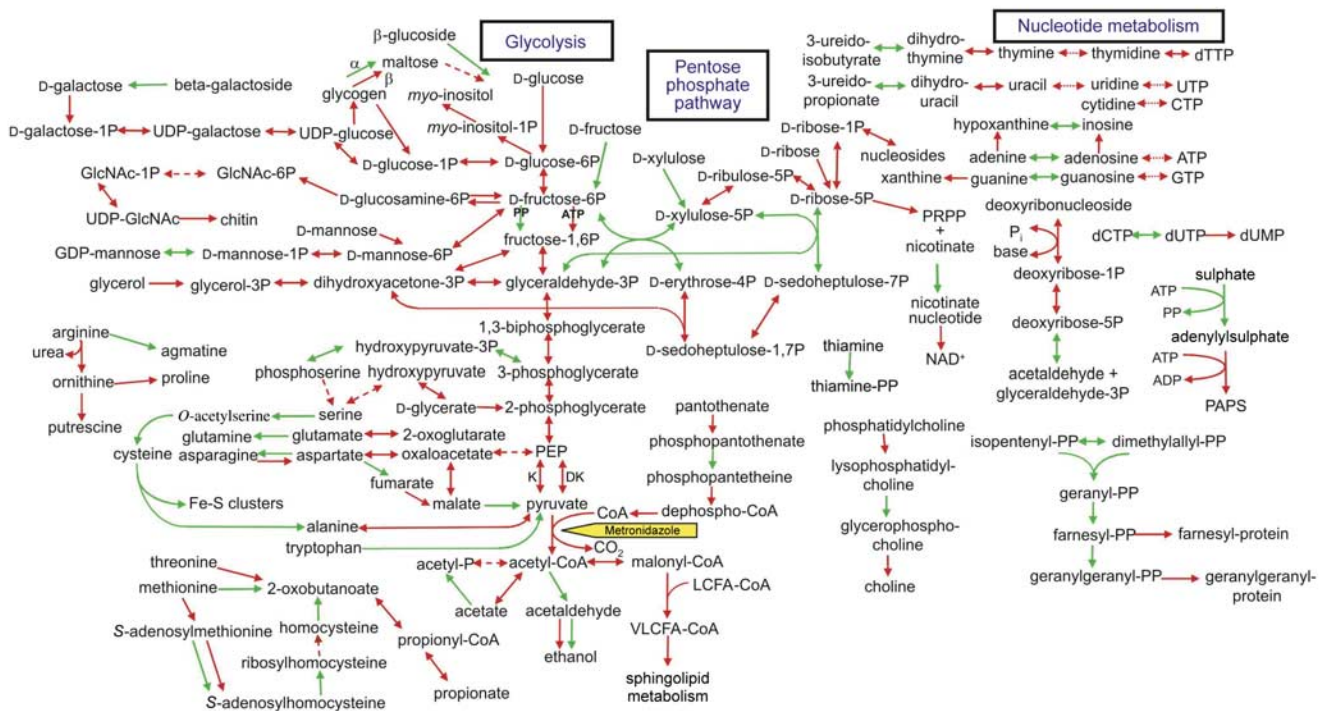


Figure 1 Predicted metabolism of *E. histolytica* based on analysis of the genome sequence data. Arrows indicate enzyme reactions. Glycolysis and fermentation are the major energy generation pathways. Green arrows represent enzymes encoded by genes that are among the 96 candidates for LGT into the *E. histolytica* genome. Broken arrows indicate enzymes for which no gene could be identified using searches of the genome data, although the activity is likely to be present. The yellow arrow points to the source of

electrons for activation of metronidazole, the major drug for treatment of amoebic liver abscess. DK, pyruvate phosphate dikinase; GlnAc, *N*-acetylglucosamine; GPI, glycosylphosphatidylinositol; K, pyruvate kinase; LCFA, long-chain fatty acid; PAPS, phosphoadenosine phosphosulphate; PEP, phosphoenolpyruvate; PP, pyrophosphate; PRPP, phosphoribosyl pyrophosphate; VLCFA, very-long-chain fatty acid.

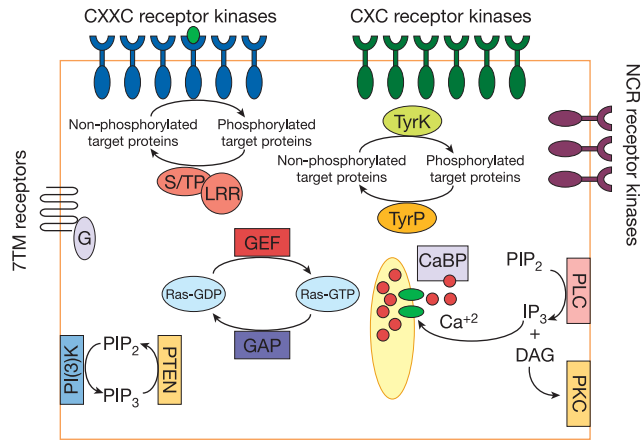


Figure 2 Predicted signal transduction mechanisms of *E. histolytica* based on analysis of the genome sequence data. *E. histolytica* possesses three types of receptor serine/threonine kinases: one group has CXXC repeats in the extracellular domain; a second has CXC repeats; and a third has non-cysteine rich (NCR) repeats. *E. histolytica* has cytosolic tyrosine kinases (TyrK), but not receptor tyrosine kinases. Some serine/threonine phosphatases (S/TP) have an attached LRR domain. CaBP, calcium-binding protein; DAG, diacylglycerol; G, G protein; GAP, GTPase-activating protein; GEF, guanine nucleotide exchange factor; IP₃, inositol-1,4,5-trisphosphate; PI(3)K, phosphatidylinositol-3-OH kinase; PIP₂, phosphatidylinositol-4,5-bisphosphate; PIP₃, phosphatidylinositol-3,4,5-trisphosphate; PKC, protein kinase C; PLC, phospholipase C; PTEN, phosphatase and tensin homologue; TyrP, tyrosine phosphatase; TTM receptors, seven-transmembrane receptors.

communication between the regulators of vesicle budding and cytoskeletal rearrangement.

E. histolytica uses a complex mix of signal transduction systems in order to sense and interact with the different environments it encounters (Fig. 2). Almost 270 putative *E. histolytica* protein kinases representing members of all seven families of the eukaryotic protein kinase superfamily were identified¹⁷. These include tyrosine kinases with SH2 domains, tyrosine kinase-like protein kinases and 90 putative receptor Ser/Thr kinases. These Ser/Thr kinases are uncommon in protists, appear to be absent from *Dictyostelium* and have previously been described only in plants, animals and Choanoflagellates. The *E. histolytica* receptor Ser/Thr kinases all contain an N-terminal signal peptide, a predicted extracellular domain and a single transmembrane helix followed by a cytosolic tyrosine kinase-like domain. The receptor kinases fall into three groups on the basis of differences in their predicted extracellular domains. The first group of 50 receptor kinase proteins contains CXXC-rich repeats similar to those found in the intermediate subunit (Igl) of the Gal/GalNAc lectin and *G. lamblia* variant-specific surface proteins. A second group of 32 proteins encodes cysteine-rich domains containing CXC repeats. The third group of eight receptor kinase-like proteins lacks cysteine-rich extracellular domains. Although no immediate downstream effectors to the amoebic receptor kinases could be identified, *E. histolytica* contains greater than 100 protein phosphatases, which dephosphorylate proteins. An unusual feature of some of the phosphatases is the presence of varying numbers of leucine-rich repeat (LRR) domains that are involved primarily in protein-protein interactions and have not previously been associated with phosphatases. The *E. histolytica* genome encodes numerous putative seven-transmembrane receptors and trimeric G proteins, which are probably involved in mediating autocrine stimulation of encystation¹⁸. In contrast to autocrine stimulation of *Dictyostelium* sporulation, which uses secreted cyclic AMP, *E. histolytica* encystment is self-stimulated by secreted catecholamines¹⁸. Finally, *E. histolytica* has numerous cytosolic proteins involved in signal transduction, including Ras-family proteins,

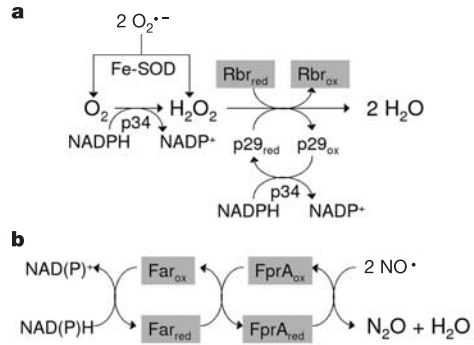


Figure 3 Predicted pathways for oxidative and nitrosative stress resistance in *E. histolytica*. Enzymes boxed and shaded have previously only been identified in anaerobic prokaryotes and amitochondrial protists. **a**, Superoxide is detoxified by an iron-containing superoxide dismutase (Fe-SOD). Molecular oxygen is reduced to hydrogen peroxide by the NADPH-flavin oxidoreductase (p34), which also transfers electrons to peroxiredoxin (p29). Rubrerythrin (Rbr) is predicted to convert hydrogen peroxide to water, although the source of electrons for rubrerythrin in *E. histolytica* is unknown. **b**, A-type flavoproteins (FprA) detoxify nitric oxide to nitrous oxide. FprA receives electrons from flavoprotein A reductase (Far).

EF-hand calcium-binding proteins, phosphatidylinositol-3-OH kinase and MAP kinases. This represents the most varied set of signal-transduction-related proteins yet described in a single-celled eukaryote.

In contrast to life in the anoxic colon, *E. histolytica* encounters a relatively high-oxygen environment during invasive amoebiasis, and coping with this change is therefore an important virulence factor. The importance of this response is underscored by the redundancy of oxygen detoxification mechanisms. *E. histolytica* has four copies of flavoprotein A, which detoxifies nitric oxide and/or oxygen¹⁹ (Fig. 3), and also contains rubrerythrin, which in anaerobic bacteria is protective against intracellular hydrogen peroxide²⁰ (Fig. 3). These oxidative and/or nitrosative stress resistance genes are shared with *G. lamblia* (with the exception of rubrerythrin) and *T. vaginalis*, but have generally been associated with anaerobic prokaryotes (Fig. 3).

E. histolytica is the first amoeba genome to be fully sampled, and comparisons with other genomes will assist in resolving fundamental issues relating to eukaryote and amoeba phylogeny, as well as how LGT affects eukaryotes. Despite a lack of representative genome sampling from amitochondrial protist lineages it is already clear that these unrelated anaerobic eukaryotes seem to use convergent metabolic strategies imposed by their environments. As a first insight into an amitochondrial protist genome, analysis of these data and particularly the bacterial-like proteins contained therein should illuminate future efforts aimed at the development of diagnostics and therapeutics of these luminal parasites. □

Methods

Genome sequencing and assembly

The *E. histolytica* genome sequence was generated by the whole-genome shotgun method. As the chromosomes of *E. histolytica* could not be resolved by pulsed field gel electrophoresis (PFGE) and the A + T content precluded making large or medium insert libraries in bacterial artificial chromosomes (BACs), we were required to use the whole-genome shotgun approach to sequence the genome. Genomic DNA was prepared from *E. histolytica* strain HM-1:IMSS (ATCC number 30459) grown axenically in TYI-S-33 medium²⁰. At TIGR 390,000 reads were produced from a small (1.5–2.0 kb) and a medium insert library (8–10 kb) generated in the pHOS2 vector. At the Sanger Institute, 200,000 reads were generated from a pUC18 library with average insert size of 2.5 kb plus 6,500 reads from a BAC library with an average insert size of 10 kb (the high A + T content of the genomic DNA prevented cloning of larger fragments). To avoid assembly problems, reads containing episomal-derived rDNA or tRNA-containing sequences (170,000 reads (29%)) were excluded from the whole-genome assembly process. The average edited read length was 645 bp, giving an approximate 12.5-fold genome coverage. Genome assembly was carried out at the Sanger Centre using the program phusion²¹. All scaffolds smaller than

2 kb (327) were subsequently removed, leaving 1,425 scaffolds with a combined size of 25,393,225 bp. The remaining scaffolds were analysed to remove redundancy that may have resulted as a consequence of allelic differences or aneuploidy. We removed all scaffolds smaller than 5 kb that shared 98% or more nucleotide sequence identity over greater than 95% of their lengths. Removal of these scaffolds left 888 scaffolds remaining, with a total length of 23,751,783 bp. All scaffolds removed during the clean-up process as well as any singleton reads, although not used in the annotation process, were used in determining the presence or absence of genes in the *E. histolytica* genome. Unfortunately, there is no map to order the scaffolds generated by the assembly; however, the sequence generated by this project should assist in making maps for this genome in the future, and although the large-scale structure of the genome has been lost, the vast majority of the genes that have been predicted are full length with intact 3' and 5' untranslated regions.

Annotation

The Combiner algorithm was used for gene structure identification²² using two genefinder programs, phat²³ and GlimmerHMM²⁴, trained using a set of published *E. histolytica* gene sequences, alignments of protein homologues to the genomic sequence and alignment of a set of *E. histolytica* complementary DNA sequences (provided by N. Guillén) to the genomic sequence. The Combiner gene predictions were then manually curated. Functional annotations for the predicted proteins were automatically generated using a combination of numerous sources of evidence including searches against a non-redundant protein database and identification of functional domains by searches against the Pfam database²⁵. tRNAs were detected using the tRNAscan-SE²⁶ program with default parameters.

Identification of sequence homologues in other species

Sequence homologues from other species were identified by searching the predicted proteins from the *E. histolytica* genome against the publicly available nr database of GenBank using BlastP (<http://www.ncbi.nlm.nih.gov/BLAST/>) and filtering search results with an *e*-value of 10^{-5} or less, which was chosen because of the relatively large divergence between *E. histolytica* and other organisms for which the genomes have been sequenced and for which protein data are available.

Phylogenetic analysis

We modified a published suite of scripts and modules called PyPhy²⁷ to make an automated genome-wide primary screen for LGT. PyPhy was used to make bootstrap (100 replicates) consensus *p*-distance trees from edited alignments of 5,740 *E. histolytica* proteins; that is, those for which there were sufficient homologues (>4) in SwissProt and TrEMBL to make trees. The trees were analysed to identify cases where the nearest neighbour to the *E. histolytica* protein was a prokaryotic sequence. As an additional screen for LGT we identified all proteins for which a prokaryote was the top Blast hit. After manual inspection of the alignments, Blast outputs, tree support values and sequence identities, 279 cases of potential LGT were retained for more detailed phylogenetic analyses. Each candidate LGT was analysed by MrBayes²⁸ using the WAG matrix, a gamma correction for site rate variation and a proportion (pinvar) of invariant sites. The analyses were run for 600,000 generations and sampled every 100 generations, with the first 2,000 samples discarded as burn-in. A consensus tree was made from the remaining samples. Because posterior probabilities—the support values used by bayesian analysis to indicate confidence in groups—have been criticized²⁹, we also used bootstrapping to provide an additional indication of support for relationships. Each data set was bootstrapped (100 replicates) and used to make distance matrices under the same evolutionary model as in the bayesian analysis, using custom (P4) software (available on request). Trees were made from the distance matrices using FastME³⁰ and a bootstrap consensus tree made using P4. On the basis of these analyses we identified 96 genes in which the tree topology is consistent with prokaryote to eukaryote LGT. Blast summary statistics, trees and support values for these 96 candidate LGT are provided as Supplementary Information.

Received 26 October; accepted 2 December 2004; doi:10.1038/nature03291.

- Stanley, S. L. Jr Amoebiasis. *Lancet* **361**, 1025–1034 (2003).
- Patarapotikul, J. & Langsley, G. Chromosome size polymorphism in *Plasmodium falciparum* can involve deletions of the subtelomeric pFrep20 sequence. *Nucleic Acids Res.* **16**, 4331–4340 (1988).
- Melville, S. E., Gerrard, C. S. & Blackwell, J. M. Multiple causes of size variation in the diploid megabase chromosomes of African trypanosomes. *Chromosome Res.* **7**, 191–203 (1999).
- Leon-Avila, G. & Tovar, J. Mitosomes of *Entamoeba histolytica* are abundant mitochondrion-related remnant organelles that lack a detectable organelle genome. *Microbiology* **150**, 1245–1250 (2004).
- Fahey, R. C., Newton, G. L., Arrick, B., Overdank-Bogart, T. & Aley, S. B. *Entamoeba histolytica*: a eukaryote without glutathione metabolism. *Science* **224**, 70–72 (1984).
- Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
- Baum, K. F., Berens, R. L., Marr, J. J., Harrington, J. A. & Spector, T. Purine deoxynucleoside salvage in *Giardia lamblia*. *J. Biol. Chem.* **264**, 21087–21090 (1989).
- Jordan, I. K., Henze, K., Fedorova, N. D., Koonin, E. V. & Galperin, M. Y. Phylogenomic analysis of the *Giardia intestinalis* transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes. *J. Mol. Microbiol. Biotechnol.* **5**, 172–189 (2003).
- James, D. W. Jr *et al.* Directed tagging of the *Arabidopsis* fatty acid elongation1 (FAE1) gene with the maize transposon activator. *Plant Cell* **7**, 309–319 (1995).
- Azachi, M. *et al.* Salt induction of fatty acid elongase and membrane lipid modifications in the extreme halotolerant alga *Dunaliella salina*. *Plant Physiol.* **129**, 1320–1329 (2002).
- Lawrence, J. G. & Hendrickson, H. Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* **50**, 739–749 (2003).
- Huston, C. D. Parasite and host contributions to the pathogenesis of amebic colitis. *Trends Parasitol.* **20**, 23–26 (2004).
- Welter, B. H. & Temesvari, L. A. A unique Rab GTPase, EhrRabA, of *Entamoeba histolytica*, localizes to the leading edge of motile cells. *Mol. Biochem. Parasitol.* **135**, 185–195 (2004).

- Mazzucco, A., Benchimol, M. & De Souza, W. Endoplasmic reticulum and Golgi-like elements in *Entamoeba*. *Micron* **28**, 241–247 (1997).
- Duhon, D. & Cardelli, J. The regulation of phagosome maturation in *Dictyostelium*. *J. Muscle Res. Cell Motil.* **23**, 803–808 (2002).
- Voigt, H. & Guillen, N. New insights into the role of the cytoskeleton in phagocytosis of *Entamoeba histolytica*. *Cell. Microbiol.* **1**, 195–203 (1999).
- Hunter, T. Protein kinase classification. *Methods Enzymol.* **200**, 3–37 (1991).
- Coppi, A., Merali, S. & Eichinger, D. The enteric parasite *Entamoeba* uses an autocrine catecholamine system during differentiation into the infectious cyst stage. *J. Biol. Chem.* **277**, 8083–8090 (2002).
- Gomes, C. M. *et al.* A novel type of nitric-oxide reductase. *Escherichia coli* flavorubredoxin. *J. Biol. Chem.* **277**, 25273–25276 (2002).
- Sztukowska, M., Bugno, M., Potempa, J., Travis, J. & Kurtz, D. M. Jr Role of rubrerythrin in the oxidative stress response of *Porphyromonas gingivalis*. *Mol. Microbiol.* **44**, 479–488 (2002).
- Mullikin, J. C. & Ning, Z. The phusion assembler. *Genome Res.* **13**, 81–90 (2003).
- Allen, J. E., Pertea, M. & Salzberg, S. L. Computational gene prediction using multiple sources of evidence. *Genome Res.* **14**, 142–148 (2004).
- Cawley, S. E., Wirth, A. I. & Speed, T. P. Phat—a gene finding program for *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **118**, 167–174 (2001).
- Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
- Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
- Sicheritz-Ponten, T. & Andersson, S. G. A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* **29**, 545–552 (2001).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Cummings, M. P. *et al.* Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* **52**, 477–487 (2003).
- Desper, R. & Gascuel, O. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol. Biol. Evol.* **21**, 587–598 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institute of Allergy and Infectious Disease and the Wellcome Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.L. (bjloftus@tigr.org). Scaffold sequences have been deposited in GenBank under the project accession number AAFB00000000. Sequences and annotation are available at <http://www.tigr.org/tdb/e2k1/eha1/>.

Excitatory cortical neurons form fine-scale functional networks

Yumiko Yoshimura*, Jami L. M. Dantzker* & Edward M. Callaway

Systems Neurobiology Laboratories, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

* Present addresses: Department of Visual Neuroscience, Research Institute of Environmental Medicine, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan (Y.Y.); Department of Neurology and Neurological Sciences, Stanford University, 300 Pasteur Drive, Room M016, Stanford, California 94305-5122, USA (J.L.M.D.)

The specificity of cortical neuron connections creates columns of functionally similar neurons spanning from the pia to the white matter^{1–6}. Here we investigate whether there is an additional, finer level of specificity that creates subnetworks of excitatory neurons within functional columns. We tested for fine-scale specificity of connections to cortical layer 2/3 pyramidal neurons in rat visual cortex by using cross-correlation analyses of synaptic currents evoked by photostimulation. Recording simultaneously from adjacent layer 2/3 pyramidal cells, we find that when they are connected to each other (20% of all recorded pairs) they share common input from layer 4 and within layer 2/3. When adjacent layer 2/3 neurons are not connected to each other, they share very little (if any) common excitatory input from layers 4 and 2/3. In contrast, all layer 2/3 neurons share common excitatory input