Genome Biology

Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling

Genome Biology 2013, **14**:R11 doi:10.1186/gb-2013-14-2-r11

Michael Clarke (michael.clarke@ucd.ie)
Amanda J Lohan (amanda.lohan@ucd.ie)
Bernard Liu (bernard.a.liu@gmail.com)
Ilias Lagkouvardos (lagkouvardos@microbial-ecology.net)
Scott Roy (scottwroy@gmail.com)
Nikhat Zafar (nzafar@jcvi.org)
Claire Bertelli (Claire.Bertelli@chuv.ch)
Christina Schilde (C.Schilde@dundee.ac.uk)
Arash Kianianmomeni (arashkianian@yahoo.com)
Thomas R Burglin (thomas.burglin@ki.se)
Christian Frech (frech.christian@gmail.com)
Bernard Turcotte (bernard.turcotte@mcgill.ca)
Klaus O Kopec (klaus.kopec@tuebingen.mpg.de)
John M Synnott (synnojd@gmail.com)
Caleb Choo (ckchoo@sfu.ca)
Ivan Paponov (ivan.paponov@biologie.uni-freiburg.de)
Aliza Finkler (alizaf@post.tau.ac.il)
Chris Soon Heng Tan (chris.tan.sh@gmail.com)
Andrew P Hutchins (ahutchins@ifrec.osaka-u.ac.jp)
Thomas Weinmeier (thomas.weinmaier@univie.ac.at)
Thomas Rattei (thomas.rattei@univie.ac.at)
Jeffery SC Chu (jeff.sc.chu@gmail.com)
Gregory Gimenez (gregory.gimenez@univmed.fr)
Manuel Irimia (mirimia@gmail.com)
Daniel J Rigden (drigden@liverpool.ac.uk)
David A Fitzpatrick (david.fitzpatrick@nuim.ie)
Jacob Lorenzo-Morales (jacob.lorenzo@gmail.com)
Alex Bateman (agb@sanger.ac.uk)
Cheng-Hsun Chiu (chchiu@adm.cgmh.org.tw)
Petrus Tang (petang@mail.cgu.edu.tw)
Peter Hegemann (hegemann@rz.hu-berlin.de)
Hillel Fromm (HillelF@tauex.tau.ac.il)
Didier Raoult (didier.raoult@gmail.com)
Gilbert Greub (Gilbert.Greub@chuv.ch)
Diego Miranda-Saavedra (diego@ifrec.osaka-u.ac.jp)
Nansheng Chen (chenn@sfu.ca)
Piers Nash (pdnash.uchicago@gmail.com)
Michael L Ginger (m.ginger@lancaster.ac.uk)
Matthias Horn (horn@microbial-ecology.net)

Genome Biology

Pauline Schaap (p.schaap@dundee.ac.uk)
Lis Caler (ecaler@jcvi.org)
Brendan Loftus (brendan.loftus@ucd.ie)

ISSN 1465-6906

Article type Research

Submission date 19 July 2012

Acceptance date 15 January 2013

Publication date 1 February 2013

Article URL http://genomebiology.com/2013/14/2/R11

Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling

Michael Clarke[†1], Amanda J Lohan[†1], Bernard Liu[2], Ilias Lagkouvardos[3], Scott Roy[4], Nikhat Zafar[5], Claire Bertelli[6], Christina Schilde[7], Arash Kianianmomeni[8], Thomas R Bürglin[9], Christian Frech[10], Bernard Turcotte[11], Klaus O. Kopec[12], John M. Synnott[1], Caleb Choo[10], Ivan Paponov[13], Aliza Finkler[14], Chris Soon Heng Tan[15], Andrew P Hutchins[16], Thomas Weinmeier[17], Thomas Rattei[17], Jeffery SC Chu[18], Gregory Gimenez[19], Manuel Irimia[20], Daniel J Rigden[21], David A Fitzpatrick[22], Jacob Lorenzo-Morales[23], Alex Bateman[24], Cheng-Hsun Chiu[25], Petrus Tang[26], Peter Hegemann[8], Hillel Fromm[14], Didier Raoult[19], Gilbert Greub[6], Diego Miranda-Saavedra[16], Nansheng Chen[10], Piers Nash[27], Michael L Ginger[28], Matthias Horn[3], Pauline Schaap[7], Lis Caler[5], Brendan J Loftus[*1]

1. Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland.
2. Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Ave, Room 1081 Toronto, Ontario M5G 1X5, Canada.
3. Department für Mikrobielle Ökologie, Universität Wien, Althanstr. 14, A-1090 Wien, Austria.
4. Department of Biology, San Francisco State University, 1600 Holloway Ave. San Francisco, CA 94132, USA.
5. Bioinformatics Department, J. Craig Venter Institute, Inc., 9704 Medical Center Drive Rockville, MD 20850, USA.
6. Center for Research on Intracellular Bacteria, Institute of Microbiology, Institute of Microbiology, Rue du Bugnon 48, 1011 Lausanne, Switzerland.
7. College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, United Kingdom.
8. Institute of Biology, Experimental Biophysics, Humboldt-Universität zu Berlin, Invalidenstrasse 42, D- 10115 Berlin, Germany.
9. Department of Biosciences and Nutrition and Center for Biosciences, Karolinska Institutet, Hälsovägen 7, Novum, SE 141 83 Huddinge, Sweden.
10. Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive Burnaby, B.C. V5A 1S6, Canada.
11. Department of Medicine, McGill University, McIntyre Medical Building, 3655 Sir William Osler, Montreal, Quebec H3G 1Y6, Canada.
12. Max Planck Institute for Developmental Biology, Spemannstr. 35 – 39, 72076 Tübingen, Germany
13. Institut für Biologie II/Molecular Plant Physiology, Faculty of Biology, Albert-Ludwigs University of Freiburg, Freiburg, Germany.
14. Department of Plant Sciences, Britannia 04, Tel-Aviv University, Tel-Aviv 69978, Israel.
15. CeMM-Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, AKH BT 25.3, A-1090 Vienna, Austria.
16. World Premier International (WPI) Immunology Frontier Research Center (IFReC), Osaka University, 3-1 Yamadaoka, Suita, 565-0871 Osaka, Japan.
17. Department für Computational Systems Biology, Universität Wien, Althanstraße 14, 1090 Wien, Austria.
18. Department of Medical Genetics, Medical Genetics, C201 - 4500 Oak Street,Vancouver, BC, V6H 3N1, Canada
19. Unité des rickettsies, IFR 48, CNRS-IRD UMR 6236, Faculté de médecine, Université de la Méditerranée, Marseille, France.
20. Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, 160 College Street, Room 230,Toronto, Ontario M5S 3E1, Canada.
21. Institute of Integrative Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.
22. Department of Biology, NUI Maynooth, Co Kildare, Ireland.
23. University Institute of Tropical Diseases and Public Health of the Canary Islands, University of La Laguna, Avda. Astrofísico Fco. Sánchez, S/N 38203 La Laguna, Tenerife, Canary Islands, Spain
24. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
25. Divisions of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Children's Hospital and Chang Gung Memorial Hospital, Taoyuan, Taiwan.
26. Department of Parasitology, Chang Gung University, Taoyuan, Taiwan.
27. Ben May Department for Cancer Research and Committee on Cancer Biology, The University of Chicago, Chicago, IL 60637, USA.

28. Faculty of Health and Medicine, Division of Biomedical and Life Sciences, Lancaster University, Lancaster, LA1 4YQ, UK.

†Michael Clarke
michael.clarke@ucd.ie

†Amanda J Lohan
amanda.lohan@ucd.ie

Bernard Liu
bernard.a.liu@gmail.com

Ilias Lagkouvardos
lagkouvardos@microbial-ecology.net

Scott Roy
scottwroy@gmail.com

Nikhat Zafar
nzafar@jcvi.org

Claire Bertelli
Claire.Bertelli@chuv.ch

Christina Schilde
C.Schilde@dundee.ac.uk

Arash Kianianmomeni
arashkianian@yahoo.com

Thomas R Bürglin
thomas.burglin@ki.se

Christian Frech
frech.christian@gmail.com

Bernard Turcotte
bernard.turcotte@mcgill.ca

Klaus O Kopec
klaus.kopec@tuebingen.mpg.de

John M Synnott
synnojd@gmail.com

Caleb Choo
ckchoo@sfu.ca

Ivan Paponov
ivan.paponov@biologie.uni-freiburg.de

Aliza Finkler
mailto:alizaf@post.tau.ac.il

Chris Soon Heng Tan
chris.tan.sh@gmail.com

Andrew P Hutchins
ahutchins@ifrec.osaka-u.ac.jp

Thomas Weinmeier
thomas.weinmaier@univie.ac.at

Thomas Rattei
thomas.rattei@univie.ac.at

Jeffery SC Chu
jeff.sc.chu@gmail.com

Gregory Gimenez
gregory.gimenez@univmed.fr

Manuel Irimia
mirimia@gmail.com

Daniel J Rigden
drigden@liverpool.ac.uk

David A Fitzpatrick
david.fitzpatrick@nuim.ie

Jacob Lorenzo-Morales
jmlorenz@ull.edu.es

Alex Bateman
agb@sanger.ac.uk

Cheng-Hsun Chiu
chchiu@adm.cgmh.org.tw

Petrus Tang
petang@mail.cgu.edu.tw

Peter Hegemann
hegemann@rz.hu-berlin.de

Hillel Fromm
HillelF@tauex.tau.ac.il

Didier Raoult
didier.raoult@gmail.com

Gilbert Greub
Gilbert.Greub@chuv.ch

Diego Miranda-Saavedra
diego@ifrec.osaka-u.ac.jp

Nansheng Chen
chenn@sfu.ca

Piers Nash
pdnash.uchicago@gmail.com

Michael L Ginger
m.ginger@lancaster.ac.uk

Matthias Horn
horn@microbial-ecology.net

Pauline Schaap
p.schaap@dundee.ac.uk

Lis Caler
ecaler@jcvi.org

Brendan J Loftus
brendan.loftus@ucd.ie

Abstract

Background

The Amoebozoa constitute one of the primary divisions of eukaryotes encompassing taxa of both biomedical and evolutionary importance, yet its genomic diversity remains largely unsampled. Here we present an analysis of a whole genome assembly of *Acanthamoeba castellanii* (*Ac*) the first representative from a solitary free-living amoebozoan.

Results

Ac encodes 15,455 compact intron rich genes a significant number of which are predicted to have arisen through interkingdom lateral gene transfer (LGT). A majority of the LGT candidates have undergone a substantial degree of intronization and *Ac* appears to have incorporated them into established transcriptional programs. *Ac* manifests a complex signaling and cell communication repertoire including a complete tyrosine kinase signaling toolkit and a comparable diversity of predicted extracellular receptors to that found in the facultatively multicellular dictyostelids. An important environmental host of a diverse range of bacteria and viruses, *Ac* utilizes a diverse repertoire of predicted pattern recognition receptors many with predicted orthologous functions in the innate immune systems of higher organisms.

Conclusions

Our analysis highlights the important role of LGT in the biology of *Ac* and in the diversification of microbial eukaryotes. The early evolution of a key signaling facility implicated in the evolution of metazoan multicellularity strongly argues for its emergence early in the Unikont lineage. Overall the availability of an *Ac* genome should aid in deciphering the biology of the Amoebozoa and facilitate functional genomic studies in this important model organism and environmental host.

Keywords

Acanthamoeba castellanii, genome, lateral gene transfer, tyrosine kinase, signal transduction, amoeba, evolution

Background

Ac is one of the predominant soil organisms in terms of population size and distribution where it acts both as a predator and an environmental reservoir for a number of bacterial, fungal and viral species [1]. In the rhizosphere selective grazing by *Ac* alters microbial community structure and is an important contributor to the development of root architecture and nutrient uptake by plants [2]. *Ac* can also be isolated from almost any body of water and manifests in a wide variety of man-made water systems including potable water sources, swimming pools, hot tubs, showers and hospital air conditioning units [3, 4]. Acanthamoebae are frequently associated with diverse range of bacterial symbionts [5, 6]. A subset of the microbes which serve as prey for *Ac* have evolved virulence stratagems to use *Ac* as both a replicative niche and as a vector for dispersal are important human intracellular pathogens [7, 8]. These pathogens utilise analogous strategies to infect and persist within mammalian macrophages, illustrating the role of environmental hosts such as *Ac* in the evolution and maintenance of virulence [9, 10]. Commonalities at the level of host response between amoebae and macrophages to such pathogens have led to the use of both *Dd* and *Ac* as model systems to study pathogenesis [11, 12].

Published Amoebozoa genomes from both the obligate parasite *Entamoeba histolytica* (*Eh*) and the facultatively multicellular *Dd* have both highlighted unexpected complexities at the level of cell motility and signaling [13, 14]. As the only solitary free-living representative, the genome of *Ac* establishes a unique reference point for comparisons for the interpretation of other amoebozoan genomes. Experimentally *Ac* has been a more thoroughly studied organism than most other free living amoebae (FLA) acting as a model organism for studies on the cytoskeleton, cell movement, and aspects of gene regulation, with a large body of literature supporting its molecular interactions [15-18].

Results and discussion

Lateral Gene Transfer (LGT)

LGT is considered a key process of genome evolution and a number of studies have indicated that phagotrophs manifest an increased rate of LGT compared to non-phagotrophic organisms [19]. As a geographically dispersed bacteriovorous amoebae with a penchant for harbouring endosymbionts, *Ac* encounters a rich and diverse supply of foreign DNA providing ample opportunity for LGT. Homology based searches of the proteome illustrates the potential for diverse contributions to the genome (Figure 1).

We therefore undertook a phylogenomic analysis to determine cases of predicted inter-domain LGT in the *Ac* genome (Additional file 1, Section 2). Our analysis identified 450 genes or 2.9% of the proteome, predicted to have arisen through LGT (Figure 2; Additional file 1, Section 2). To determine the fate and ultimate utility of the LGT candidates within the *Ac* genome we examined their expression levels across a number of experimental conditions using RNA.seq (Additional file 1, Table S1.6.1). Our results show that most of the LGT candidates are expressed in at least some of the conditions tested (Additional file 2, LGT_analysis.xls).

Genetic exchange is also thought to occur between phylogenetically disparate organisms that reside within the same amoebal host cell [20, 21]. *Ac* contains 3 copies of a miniature transposable element (ISSoc2) of the IS607 family of insertion sequences with a restricted distribution, being present in populations of thermophilic cyanobacteria [22] and in the Mimivirus genome. The mimiviral IS elements are found within islands of genes of bacterial origin a number of which appear to have been contributed by a cyanobacterial donor. *Ac* is both the host of Mimivirus and a predator of cyanobacteria [17] and the presence of these IS elements underscore the complex intermediary role that *Ac* may play in facilitating genetic transfer between these sympatric species.

Comparison of predicted LGT across amoeboid genomes

In order to compare the impact and scale of LGT across *Ac* and other amoeba we applied the same phylogenomic approach used to identify LGT in the *Ac* genome to published genomes of other amoeboid protists including *Dd*, *Eh*, *Entamoeba dispar* (*Ed*) and *Naegleria gruberi* (*Ng*). Our findings predict that *Ac* and the excavate *Ng* encode a notably higher number of laterally acquired bacterial genes than either of the more closely related parasitic *Entamoeba* or the social *Dd* amoebozoans (Figure 2A). The taxonomic distribution of putative LGT donors is broadly similar for both *Entamoeba* species, but surprisingly also between *Ac* and *Ng* (Figure 2B; Figure 2C; Additional file 1, Section 2). The genomes of both *Eh* and *Ed* are predicted to have experienced a proportionately higher influx from anaerobic and host-associated microbes than their free-living counterparts *Ac* and *Ng* (Figure 2C; Additional file 2, LGT_analysis.xls), likely reflecting the composition of microbes within their habitats. Many of the LGT candidates across all of the amoebae have predicted metabolic functions suggesting that in amoebae, LGT is reflective of trophic strategy and driven by the selective pressure of new ecological niches. Our data illustrating LGT as a contributing

factor in shaping the biology of a diversity of amoeboid genomes provides further evidence supporting an underappreciated role for LGT in the diversification of microbial eukaryotes [23].

Introns

Intron-exon structures exhibit complex phylogenetic patterns with orders-of-magnitude differences across eukaryotic lineages, which imply frequent transformations during eukaryotic evolution [24]. Some researchers have argued that intron gain is episodic, with long periods of stasis [25] punctuated by periods of rapid gain while others argue for generally higher rates [26]. Strikingly, *Ac* genes have an average of 6.2 introns per gene, among the highest known in eukaryotes [27]. Genes predicted to have arisen through LGT have slightly lower but broadly comparable intron densities offering an opportunity to study the evidence for proposed mechanisms underpinning post LGT intron gain [28]. An analysis of LGT candidate introns however did not provide support for any of the proposed mechanisms of intron gain (Additional file 1, Section 2). Thus, while the preponderance of introns in predicted LGTs clearly indicates substantial intron gain at some point, it appears that for *Ac* these events have been very rare in more recent times, consistent with a punctate model of intron gain.

Cell signaling

As a unicellular sister grouping to the multicellular Dictyostelids, *Ac* provides a unique point of comparison to gain insight into the molecular underpinnings of multicellular development in Amoebozoa. Cell-cell communication is a hallmark of multicellularity and we looked at putative receptors for extracellular signals and their downstream targets. G-protein-coupled receptors (GPCRs) represent one of the largest families of sensors for extracellular stimuli. Overall *Ac* encodes 35 GPCRs (compared to 61 in *Dd*), representing 4 out of the six major families of GPCRs [29] while lacking metabotropic glutamate-like GPCRs or fungal pheromone receptors. We identified 3 predicted fungal associated glucose sensing Git3 GPCRs [30] and an expansion in the number of frizzled/smoothened receptors [31] (Additional file 1, Figure S3.1.1). We identified seven G-protein alpha subunits and a single putative target, phospholipase C, for GPCR-mediated signaling. The number and diversity of receptors in *Ac* raises the question of what they are likely to be sensing. Nematodes use a large number of GPCRs to detect compounds secreted by their bacterial food sources [32] and given the diversity of *Ac*'s feeding environments we predict that many of the *Ac* GPCRs may fulfill a similar function.

Environmental sensing

We identified 48 sensor histidine kinases (SHKs), of which 17 harbor transmembrane domains and may function as receptors (Additional file 1, Figure S3.2.1). Remarkably, there are also 67 nucleotidyl cyclases consisting of an extracellular receptor domain separated by a single transmembrane helix from an intracellular cyclase domain flanked by two serine/threonine kinase domains. This domain configuration is present in a number of the amoeba-infecting giant viruses but thus far appears unique for a cellular organism (Additional file 1, Figure S3.3.1). *Ac* is able to survive under microaerophilic conditions such as those found in the deeper layers of underwater sediments or within the rhizosphere. The genome encodes a number of prolyl 4-hydroxylases that likely mediate oxygen response however *Ac* also contains a number of Heme-Nitric oxide/Oxygen binding (H-NOX) proteins that, unlike those in other eukaryotes are not found in conjunction with guanylyl cyclases [33]. The *Ac* H-NOXproteins lack a critical tyrosine residue in the non-polar distal heme pocket making it likely that they are for Nitric oxide (NO) rather than oxygen signaling [34]. Both *Dd* and *Ac* are responsive to light, however the photoreceptor that mediates phototaxis in *Dictyostelium* has yet to be identified [35]. We identified two rhodopsins both with C-terminal histidine kinase and response regulator domains with homology to the sensory rhodopsins of the green algae that represent candidates for light sensors in *Ac* (Figure 3).

Cellular response

Modulation of cellular response to environmental cues is enacted by a diversity of protein kinases and *Ac* is predicted to encode 377, the largest number predicted to date for any amoebozoan (Additional file 1, Section 4). In *Ac* MAPK kinase-mediated pathways have been shown to be involved in encystment [36] and its genome encodes homologues of both of *Dd*'s two MAPK proteins ErkA and ErkB [37]. Phosphotyrosine signaling (pTyr) mediated through tyrosine kinases was until recently thought to be generally absent from the amoebozoan lineage [38]. This signaling capacity has been associated with intercellular communication, the evolutionary step towards multicellularity and the expansion of organismal complexity in metazoans [39]. PTyr is thought to depend upon a triad of signaling molecules; tyrosine kinase "writers" (PTKs), tyrosine phosphatases "erasers" (PTPs) and Src Homology 2 (SH2) "reader" domains that connect the phosphorylated ligand containing domains to specify downstream signaling events [39]. Remarkably the genome of *Ac*

encodes 22 PTKs, 12 PTPs, and 48 SH2 domain-containing proteins (Figure 4A) revealing a primordial yet elaborate pTyr signaling system in the amoebozoan lineage (Figure 4B).

The *Ac* PTK domains are highly conserved in key catalytic residues resembling dedicated PTKs found in metazoans (Additional file 1, Figure S4.2.1), and are distinct from *Dd* and *Eh* PTKs that are more tyrosine kinase like (TKL) (Additional file 1, Figure S4.2.2). *Ac* PTK homologues are present in the apusomonad *Thecamonas trahens* and have also recently been described in two described filasterean species, *Capsaspora owczarzaki* and *Ministeria vibrans*, [38]. One unusual feature of the pTyr machinery in *Ac* is the 2:1 ratio of SH2 to PTK domains as comparisons across Opisthokonts show a strong correlation and co-expansion of these two domains with a ratio close to 1:1 (Figure 4C/D) [40]. This increased ratio in *Ac* indicates either an expansion to handle the cellular requirements of pTyr signaling or that aspects of PTK function are accomplished by TKL or dual specificity kinases as appears to be the case in *Dd* [41]. We also found that *Ac* has fewer tyrosine residues in its proteome in comparison to *Dd*, which lacks PTKs (Figure S4.3.1). This result is in line with recent analysis of metazoan genomes suggesting increased pressure for selection against disadvantageous phosphorylation of tyrosine residues in genomes with extensive pTyr signaling [42].

Domain organization and composition of pTyr components reveal the selective pressures for adapting pTyr signaling into various pathways. Seven PTKs have predicted transmembrane domains and may function as receptor tyrosine kinases (RTKs). The presence of transmembrane bound PTKs in *Ac* hints at the potential for intercellular communication, a facility that could prove advantageous in navigating its complex ecology. The majority of PTKs in *Ac* however show unique domain combinations; six PTKs contain a sterile alpha motif (SAM) domain, which is found in members of the ephrin receptor family (Additional file 1, Figure S4.4.3). The *Ac* SH2 proteins are conserved within the pTyr binding pocket and resemble SH2 domains from the SOCS, RIN, CBL and RASA families (Additional file 1, Figure S4.4.2), however the domain composition within these proteins differ between those of *Monosiga brevicollis* and metazoans (Additional file 1, Figure S4.4.3A). Approximately half of the *Ac* SH2 proteins share domain architectures with *Dd* including the STAT family of transcription factors (Additional file 1, Figure S4.4.3B). The presence of homologous SH2 proteins in *Dd* coupled with the complete facility in *Ac* predicts an emergence of the complete machinery for pTyr early in the Unikont lineage. This finding is in contrast with models that posit a complete pTyr signaling machinery emerging late in the Unikont lineage [39] and has important implications

for understanding the relationship between pTyr signaling and the evolution of multicellularity. The lack of clear metazoan orthologues makes it difficult to trace the evolutionary paths of pTyr signaling networks [43] or to accurately predict the cellular functions and adaptations of pTyr in *Ac*. However, with phosphoproteomics and sequence analysis, insights into the ancient pTyr signaling circuits may be revealed through future studies in *Ac* (Additional file 1, Figure S4.5.1).

Cell adhesion

Ac is not known to participate in social activity yet must adhere to a diversity of surfaces within the soil and practice discrimination between self and prey during phagocytosis [44]. *Ac* shares some adhesion proteins with *Dd* (Additional file 1, Table S5.1.1) however homologues of the calcium-dependent, integrin-like Sib cell-adhesion proteins are absent. Surprisingly, *Ac* contains a number of bacterial-like integrin and haemagglutinin domain adhesion proteins that may improve its ability to attach to bacterial cells or biofilms [45]. *Ac* encodes 2 MAM domain-containing proteins, a domain found in functionally diverse receptors with roles in cell-cell adhesion [46]. *Ac* has a copy of the laminin-binding protein (AhLBP) first identified *Acanthamoeba healyi*, which has been shown to act as a non-integrin laminin binding receptor [47]. Remarkably, *Ac* also encodes proteins containing cell adhesion immunoglobulin domains (Additional file 1, Section 5). Both show affinity to the I-set subfamily [48] and contain weakly predicted transmembrane domains (Figure S5.1.1).

Microbial recognition through pattern recognition receptors

Ac grazes on a variety of micro fauna including a number of pathogens, which requires the mobilization of a set of defense responses initiated upon microbial recognition. In vertebrates molecular signatures often termed microbe-associated molecular patterns (MAMPs) [49] are detected by pattern-recognition receptors (PRRs) that activate downstream transcriptional responses. As *Ac* encounters a wide variety of microbial prey and practices selective feeding behaviour we looked for the presence of predicted PRRs in the *Ac* genome (Figure 5). One of the best-studied MAMPs is lipopolysaccharide (LPS) and self non-self recognition via lectin mediated protein-carbohydrate interactions is an important innate immunity strategy in both vertebrates and invertebrates [50]. *Ac* contains 6 members of the bactericidal permeability-increasing protein (BPI) / Lipopolysaccharide-binding protein (LBP) family and 2 peptidoglycan binding proteins (Figure 5), (Additional file 1, Section 6). *Ac* also encodes a membrane bound homologue of an MD-2-related

protein that in vertebrate immunity has been implicated in opsonophagocytosis of Gram-negative bacteria through its interactions with lipopolysaccharide [51].

Receptor-mediated endocytosis of *Legionella pneumophila* in *Ac* is mediated by the c-type lectin mannose binding protein (MBP) [52] and in pathogenic Acanthamoebae MBP is the principal virulence factor [53]. In addition to MBP, the *Ac* genome encodes two paralogues of MBP with similarity to the N-terminal region of the protein that may fulfill similar functions. Rhamnose-binding lectins (RBLs) serve a variety of functions in invertebrates, one of which is in their role as germline-encoded PRRs in innate immunity [54]. RBLs are absent from other Amoebozoa however *Ac* encodes 11 D-galactoside/L-rhamnose binding (SUEL) lectin domain-containing proteins. Approximately half also contain epidermal growth factor (EGF) domains, a combination reminiscent of the selectin family of adhesion proteins found exclusively in vertebrates [55]. An L-rhamnose synthetic pathway has recently been identified in Mimivirus that is thought to contribute to biosynthesis of the LPS-like outer layer of the virus particle and contribute to its uptake by phagocytosis [56, 57]. *Ac* also encodes a protein containing 3 copies of a H-type lectin domain attached to an inhibitor of apoptosis domain. The H-lectin domain is predicted to be N-acetylgalactosamine (GalNAc) binding and is found in *Dictyostelium* discoidin I & II [58] and in a number of invertebrates where it plays a role in antibacterial defence [59]. In the brown algae *Ectocarpus* leucine-rich repeat (LRR) containing GTPases of the ROCO family and NB-ARC-TPR proteins have been proposed to represent PRRs that are involved in immune response [60]. *Ac* encodes a NB-ARC-TPR homologue with a disease resistance interpro domain (IPR000767) and an LRR-ROCO GTPase.

Antimicrobial defense

Ac encodes a number of genes with potential roles in antiviral defense including homologues of the NCLDV major capsid protein [61] as well as homologues of Dicer and Piwi, both of which have been implicated in RNA mediated antiviral silencing [62]. Our data also illustrates early evolution of a number of interferon inducible innate immunity proteins absent from other sequenced Amoebozoa. These include a homologue of the interferon-γ-inducible lysosomal thiol reductase enzyme (GILT), which is an important host factor targeted by *Listeria monocytogenes* during infection in macrophages [63]. In addition *Ac* encodes two Interferon-inducible GTPase homologues, which in vertebrates promote cell-autonomous immunity to vacuolar bacteria including *Mycobacteria* and *Legionella* species [64]. *Ac* also contains a natural resistance-

associated macrophage protein (NRAMP) homologue, has been implicated in protection against *L. pneumophila* and *Mycobacterium avium* infection in both macrophages and *Dd* [65].

Metabolism

Ac has traditionally been considered to be an obligate aerobe, although the recent identification of the oxygen-labile enzymes pyruvate:ferredoxin oxidoreductase and FeFe-hydrogenase perhaps pointed towards a cryptic capacity for anaerobic ATP production [66]. Predictions for nitrite and fumarate reduction, hydrogen fermentation, together with a likely mechanism for acetate synthesis, coupled to ATP production indicate a considerable capacity for anaerobic ATP generation. This clearly sets *Ac* apart from *Dd* which hunts within the aerobic leaf litter, but provides parallels with *Ng*, the alga *Chlamydomonas reinhardtii* and other soil-dwelling protists that are likely to experience considerable variation in local oxygen tensions [67]. These protists achieve their flexible, facultative anaerobic metabolism, however, using different pathways (Additional file 1, Figure S7.1). In addition, the classic anaerobic twists on glycolysis provided by pyrophosphate-dependent phosphofructokinase and pyruvate phosphate dikinase [68] are absent from *Ac*. This suggests that although multiple pathways are available for oxidation of NADH to NAD+ in the absence of oxygen, including a capacity for anaerobic respiration in the presence of NO2-, a shift to a more ATP-sparing form of glycolysis is not necessary under low oxygen-tension. Given genome-led predictions of facultative anaerobic ATP metabolism, as well as extensive use of receptors and signaling pathways classically associated with animal biology, we also considered the possibility of a hypoxia-inducible transcription factor (HIF)-dependent system for oxygen sensing, similar to that seen across the animal kingdom, including the simple animal *Trichoplax adhaerens* [69, 70]. However, despite conservation of a Skp1/HIFα-related prolyl hydroxylase in *Ac* we found no genes encoding proteins with the typical domain architecture of animal HIFα or HIFβ. Currently, therefore HIF-dependent oxygen sensing remains restricted to metazoan lineages.

Ac also retains biosynthetic pathways involved in anabolic metabolism which are absent in *Dd* – (e.g. the shikimic acid pathway and a classic type I pathway for fatty acid biosynthesis (Additional file 1, Table S7.1.) although investment in extensive polyketide biosynthesis [71] is not evident. An autophagy pathway, as defined by genetic studies of yeast, *Dd* and other organisms [72] is present in *Ac* with little paralogue

expansion or loss of known ATG genes evident, (Additional file 1, Figure 7.2) and likely contributes to both intracellular re-modelling in response to environmental cues and the interaction with phagocytosed microbes.

Transcription factors

Ac shares a broadly comparable repertoire of transcription factors with *Dd* excepting a number of lineage specific expansions (Additional file 1, Table S8.1). *Ac* encodes 22 zinc cluster transcription factors compared to the 3 in *Dd* (Additional file 1, Fig S8.2.1) [73]. *Ac* has almost double the number of predicted homeobox genes (25) as compared to the 13 in *Dd* [74]. Two are of the MEIS and PBC class respectively, with an expansion in a homologue of Wariai, a regulator of anterior-posterior patterning in *Dictyostelium* [75] comprising most of the additional members (Additional file 1, Fig S8.3.2). Strikingly we also identified 22 Regulatory Factor X (RFX) genes, which are absent from other sequenced amoebozoan genomes [76]. RFX genes are found solely in unikonts and the *Ac* RFX repertoire is the earliest branching yet identified and forms an out-group to other known RFX genes (Additional file 1, Section 8). *Ac* has been proposed to affect plant root branching in the rhizosphere via its effects on auxin balance in plants [77]. *Ac* encodes a number of genes involved in auxin biosynthesis as well as those involved in free auxin (IAA) de-activation via formation of IAA conjugates (Additional file 1, Table S9.1). This data suggests that *Ac* plays a role in altering the level of IAA in the rhizosphere through a strategy of alternate biosynthesis and sequestration. Ac may also respond transcriptionally to auxin as it encodes a member of the calmodulin-binding transcription activator (CAMTA) family (Figure S8.4.1), which in plants co-ordinate stress responses via effects on auxin signaling [78, 79].

Conclusions

Comparative genomics of the Amoebozoa has until now been restricted to comparisons between the multicellular Dictyostelids and the obligate parasite *Eh* [80, 81]. *Ac* while sharing many of their features enriches the repertoire of amoebozoan genomes in a number of important areas including signaling and pattern recognition. LGT has shaped both the genome and transcriptome of *Ac* and our analysis of LGT across a number of amoeboid genomes reveals unexpected similarities between phylogenetically distant amoebae. *Ac* plays host to some of the earth's most unusual organisms [82] as well a number of important human pathogens [7, 8] and appears likely to play a role in facilitating genetic exchange in sympatric organisms [83].

Through LGT *Ac* has adopted bacterial-like adhesion proteins that may increase its capacity to adhere to bacterial cells and biofilms upon which *Ac* predates. Participation in NO signaling through the presence of H-NOX proteins may aid in the dispersal of biofilms [84]. The adaptive value conferred by many of the LGT candidates is difficult to establish however our data demonstrates diverse expression profiles across a sampling of conditions points to the incorporation of a large majority into new transcriptional networks. Given the feeding behaviour of *Ac* it seems plausible that as yet unidentified eukaryote-to-eukaryote gene transfers may also have provided *Ac* with adaptive advantages [23]. Increased sampling will be necessary to establish the level of eukaryote-to-eukaryote gene transfers into the *Ac* genome and the degree to which "you are what you eat" also applies on a diet of eukaryotes [23].

As a solitary amoeba *Ac* participates in a myriad of as yet unexplored interactions as reflected in the diversity of genes devoted to sensory perception and signal transduction of extracellular stimuli. *Ac*'s survival in the rhizosphere likely resides in interactions not only with the microbes present but also through communication with plant roots through manipulation of the levels of the plant hormone auxin. Some of the components of recognition and environmental sensing in *Ac* may also have been acquired via LGT potentially providing selective advantages to *Ac*. An interesting parallel is the planktonic protozoan *Oxyrrhis marina* which utilizes both MBP and LGT derived sensory rhodopsins, to enable selective feeding behavior through prey detection and biorecognition [85]. We predict that host response of *Ac* to pathogens and symbionts is likely modulated via a diversity of predicted PRRs that act in an analogous manner to effectors of innate immunity in higher organisms. Given the close association of *Ac* with a number of important intracellular pathogens it will be interesting to determine which host-pathogen interactions can trace their origins to encounters with primitive cells such as *Ac*.

Ac devotes large numbers of genes to signal transduction including expansions found in other amoebozoans (e.g. TKLs) while introducing new components based on novel domain architectures (nucleotidyl cyclases) which may act as small molecule receptors [86]. A remarkable feature of the *Ac* genome is the presence of the complete pTyr signaling toolkit including potential RTKs especially when contrasted with its absence in the multicellular Dictyostelids. The role of Tyr kinase signaling in both amoebozoan and mammalian phagocytosis [87-89] indicates that it represents an ancestral function. The most parsimonious interpretation

of our analysis predicts that functions originally carried out by tyrosine kinases were replaced by other kinases within the Amoebozoa. This finding emphasizes the importance of representative sampling and the inherent difficulties in re-constructing ancestral signaling capacities.

Transcriptional response networks can be re-programmed either through expansion of transcription factors or their target genes [90]. *Ac* and *Dd* share a conserved core of regulatory proteins with lineage specific amplifications of single or small numbers of TF family members accounting for the majority of the differences between them. These expansions in a restricted set are predicted to have resulted in sub or neo-functionalization and have contributed to the adaptive radiation of Acanthamoebae into new ecological niches.

Our analysis suggests that many signal processing and regulatory modules of higher animals and plants likely have deep origins and are balanced with subsequent losses in certain lineages e.g. the sensor histidine kinases in metazoa and the tyrosine kinases in fungi, plants and many protists. Comparison of *Ac* with *Dd* highlights a broadly similar apparatus for environmental sensing and cell-cell communication and implies that the molecular elements underpinning the transition to a multicellular life style may be widespread. Such transitions would likely have involved cooption of ancestral functions into multicellular programs and have occurred multiple times.

The availability of an *Ac* genome offers the first opportunity to initiate functional genomics in this important constituent of a variety of ecosystems and should foster a better understanding of the amoebic lifestyle. Utilizing the genome as a basis for unravelling the molecular interactions between *Ac* and a variety of human pathogens will provide a platform for understanding the contributions of environmental hosts to the evolution of virulence.

Materials and methods

DNA isolation

A. castellanii strain Neff (ATCC 30010) was grown at 30°C with moderate shaking to an $O.D_{550}$ of ~1.0. Total nucleic acid preparations were depleted of mitochondrial DNA contamination via differential centrifugation of cell extracts, [91]. High molecular weight DNA was extracted from nuclear pellets either on

Cesium chloride-Hoechst 33258 dye gradients as per [92] or by utilizing the Qiagen Genomic-tip 20/G kit (Qiagen, Hilden, Germany).

Genomic DNA library preparation and sequencing

All genomic DNA libraries were generated according to the protocol Genomic DNA Sample Prep Guide - Oligo Only Kit (1003492 A) – sonication was substituted for the recommended nebulization as the method for DNA fragmentation utilising a Biorupter™ (Diagenode, Liége, Belgium). The library preparation methodology of end repair to create blunt ended fragments, addition of 3′ -A overhang for efficient adapter ligation, ligation of the adapters, size selection of adapter ligated material was carried out utilising enzymes indicated in the protocol. Adapters and amplification primers were purchased from Illumina (Illumina, San Diego, CA, USA) both Single Read Adapters (FC-102-1003) and Paired End Adapters catalogue number PE-102-1003 were used in library construction. All enzymes for library generation were purchased from New England Biolabs (Ipswitch, MA, USA). A limited 14-cycle amplification of size-selected libraries was carried out. To eliminate adapter-dimers libraries were further sized selected 2.5% TAE agarose gels. Purified libraries were quantified using a Qubit™ fluorometer (Invitrogen, Carlsbad, CA, USA) and a Quant-iT™ double-stranded DNA High-Sensitivity Assay Kit (Invitrogen, Carlsbad, CA, USA). Clustering and sequencing of the material was carried out as per manufacturers instructions on the Illumina GAII platform in the UCD Conway Institute (UCD, Dublin Ireland).

RNA extraction and RNA.seq library preparation and sequencing

For all tested conditions refer to Additional file 1, Table S1.6.1 except the infection series, RNA was extracted from a minimum of $1x10^6$ cells using TRIzol® (Invitrogen/Life Technologies, Paisley, UK). For infection material the detailed protocol is published in [93]. Strand specific RNA.seq libraries were generated total RNA using a modified version of [94] which is detailed in [93]. Briefly was total RNA was poly(A) selected, fragmented, reverse transcribed and 2^{nd} strand cDNA marked with the addition of dUTP. Standard Illumina methodology was followed – end-repair, A-addition, adapter ligation and library size selection – with the exception of the use of 'home-brew 6-nt indexed' adapters as per Craig *et al* [95]. Prior to limited amplification of the libraries the dUTP marked 2^{nd} strand was removed via Uracil DNA-Glycosylase (Bioline, London, UK) digestion. Final libraries were quantified using the High Sensitivity DNA Quant-iT™ assay kit and Qubit™ Fluorometer (Invitrogen/ Life Technologies, Paisley, UK). All sequencing was carried out in

UCD Conway Institute on an Illumina GAII as per manufacturers instructions.

Sequencing and assembly

Genome assembly was carried out using a two-step process. Firstly the Illumina reads were assembled using the Velvet [96] short read assembler to generate a series of contigs. These assembled contigs were used to generate a set of pseudo-reads of 400 base pairs (bp) in length. These pseudo reads were then assembled in conjunction with the 454 FLX and Sanger sequences using version 2.3 of the GS De Novo Assembler using default parameters (http://454.com/products/analysis-software/index.asp) (Table S1.1.1). The assembly contained 45.1 Mb of scaffold sequence, of which 3.4 Mb (7.5%) represents gaps and 75% of the genome is contained in less than 100 scaffolds. For assembly statistics see (Additional file 1, Table S1.2.1). In order to determine the coverage of the transcriptome we aligned our genome assembly to a publicly available EST dataset from Genbank (using the entrez query acanthamoeba EST) AND "Acanthamoeba castellanii"[porgn:txid5755]). Of the 13,784 EST sequences downloaded, 12,975 (94%) map over 50% of their length with an average % identity of 99.2% and 12,423 (90%) map over 70% of their length with an average percent identity of 99.26%.

Gene structure prediction

Gene finding was carried out on the largest 384 scaffolds of the *Ac* assembly using an iterative approach by firstly generating gene models directly from RNA.seq to train a gene-finding algorithm using a genome annotation pipeline followed by manual curation. Firstly predicted transcripts were generated using RNA.seq data from a variety of conditions (Table S1.4.1) in conjunction with the G.Mo.R-Se algorithm (Gene Modelling using RNA.seq), an approach aimed at building gene models directly from RNA.seq data [97] running with default parameters. This algorithm generated 20,681 predicted transcripts. We then used these predicted transcripts to train the genefinder SNAP [98] using the MAKER genome annotation pipeline (http://www.yandell-lab.org/software/maker.html) [99]. MAKER is used for the annotation of prokaryotic and eukaryotic genome project. MAKER identifies repeats, aligns ESTs (in this case the transcripts generated by the G.Mo.R-Se algorithm) and proteins from (nr) to a genome, produces *ab-initio* gene predictions and automatically synthesizes these data into gene annotations. The 17,013 gene predictions generated by MAKER were then manually annotated using the Apollo genome annotation curation tool

(apollo.berkeleybop.org/) [100]. Apollo allows the deletion of gene models, creation of gene models from annotations and the editing of gene starts, stops, 3' and 5' splice sites. Models were manually annotated examining at a variety of evidence including expressed sequence data and matches to protein databases (Additional file 1, Section 1). Out of a total of 113,574 exons 32,836 exons are exactly covered and 64,724 are partially covered by transcripts and 7,193 genes have at least 50% of their entire lengths covered by transcript data.

Functional annotation assignments

Functional annotation assignments were carried out using a combination of automated annotation as described previously [101] followed by manual annotation. Briefly gene level searches were performed against protein, domain and profile databases including JCVI in-house non-redundant protein databases, Uniref (http://www.ebi.ac.uk/uniref/), Pfam (http://pfam.sanger.ac.uk/), TIGRfam HMMs (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi), Prosite (http://prosite.expasy.org/), and InterPro (www.ebi.ac.uk/interpro/). After the working gene set had been assigned an informative name and a function, each name was manually curated and changed where it was felt a more accurate name could be applied. Predicted genes were classified using Gene Ontology (GO) [102]. GO assignments were attributed automatically, based on other assignments from closely related organisms using Pfam2GO, a tool that allows automatic mapping of Pfam hits to GO assignments.

Data access

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AHJI00000000. The version described in this paper is the first version, AHJI01000000. The RNA.seq data is available under accessions (SRA061350 and SRA061370-SRA061379).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Experiments were conceived and designed by MC, AJL, and BL. Analyses were carried out by all authors. Cell cultures of *A. castellanii* were grown and DNA isolated by AJL. DNA sequencing libraries were made,

and sequencing carried out, by AJL. The manuscript was drafted by BL, with contributions from all authors. All authors read and approved the final manuscript for publication.

Acknowledgements

This work was funded by the Irish Science Foundation (SFI) grants 05/RP1/B908 & 05/RP1/908/EC07 awarded to BJL. The authors thank the Broad Institute and the investigators of ‘‘the Origins of Multicellularity Sequencing Project, Broad Institute of Harvard and MIT’’ (http://www.broadinstitute.org/) for making data publicly available and the Liverpool Centre for Genomic Research for provision of 454 sequencing data. IL and MH were funded through grants from the Austrian Research Fund (Y277-B03), the European Research Council (Starting Grant EVOCHLAMY 281633), and the University of Vienna (Graduate School “Symbiotic Interactions”). BT was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (grant #184053). TRB is funded by the Centre for Biosciences. EST sequencing was supported by grants (CMRPD1A0581, CMRPG450131, and SMRPD160011) from Chang Gung Memorial Hospital to PT and CHC. J.L.M. was supported by the Ramón y Cajal Subprogramme of the Spanish Ministry of Economy and Competitivity RYC-2011-08863.

Abbreviations

Ac, *Acanthamoeba castellanii*; ATG, AuTophaGy related; BPI, bactericidal permeability-increasing protein; CAMTA, Calmodulin-binding transcription activator; *Dd*, *Dictyostelium discoideum*; *Ed*, *Entamoeba dispar*; EGF, epidermal growth factor; *Eh*, *Entamoeba histolytica*; ERVR, endogenous virus receptor; FLA, free living amoeba; GalNAc, N-acetylgalactosamine; GILT, Gamma-interferon-inducible lysosomal thiol reductase; GPCR, G-protein-coupled receptor; H-NOX, Heme-Nitric oxide/Oxygen binding; HIF, hypoxia-inducible transcription factor; IAA, indole-3-acetic acid; LBP, Lipopolysaccharide-binding protein; LGT, lateral gene transfer; LPS, lipopolysaccharide; LRR, leucine-rich repeat; MAMPs, microbe-associated molecular patterns; MAPK, Mitogen-activated protein kinase; MBP, mannose binding protein; NADH, nicotinamide adenine dinucleotide; NB-ARC-TPR, NB-ARC tetratricopeptide repeat containing protein; NCLDV, nucleocytoplasmic large DNA viruses; *Ng*, *Naegleria gruberi*; NO, Nitric oxide; NRAMP, natural resistance-associated macrophage protein; PRRs, pattern-recognition receptors; PTK, tyrosine kinase “writers”; PTPs, tyrosine phosphatases “erasers”; RBLs, Rhamnose-binding lectins; RFX, Regulatory Factor X; RTK, receptor tyrosine kinases; SAM, sterile alpha motif; SH2, Src Homology 2 “reader” domains; SUEL, D-galactoside/L-rhamnose binding SUEL lectin domain-containing proteins; SHK, sensor histidine kinase pTyr, Phosphotyrosine signaling; TKL, tyrosine kinase like; TF, transcription factor.

References

1. De Jonckheere JF: **Ecology of Acanthamoeba.** *Reviews of Infectious Diseases* 1991, **13 Suppl 5:**S385-387.
2. Rosenberg K, Bertaux J, Krome K, Hartmann A, Scheu S, Bonkowski M: **Soil amoebae rapidly change bacterial community composition in the rhizosphere of Arabidopsis thaliana.** *The ISME Journal* 2009, **3:**675-684.
3. Nwachuku N, Gerba CP: **Health effects of Acanthamoeba spp. and its potential for waterborne transmission.** *Reviews of Environmental Contamination & Toxicology* 2004, **180:**93-131.
4. Thomas V, McDonnell G, Denyer SP, Maillard JY: **Free-living amoebae and their intracellular pathogenic microorganisms: risks for water quality.** *FEMS Microbiology Reviews* 2009.
5. Horn M, Wagner M: **Bacterial endosymbionts of free-living amoebae.** *The Journal of Eukaryotic Microbiology* 2004, **51:**509-514.
6. Horn M: **Chlamydiae as symbionts in eukaryotes.** *Annual Review of Microbiology* 2008, **62:**113-131.
7. Greub G, Raoult D: **Microorganisms resistant to free-living amoebae.** *Clinical Microbiology Reviews* 2004, **17:**413-433.
8. Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y: **Amoebae as training grounds for intracellular bacterial pathogens.** *Applied and Environmental Microbiology* 2005, **71:**20-28.
9. Salah IB, Ghigo E, Drancourt M: **Free-living amoebae, a training field for macrophage resistance of mycobacteria.** *Clinical Microbiology and Infection* 2009, **15:**894-905.
10. Winiecka-Krusnell J, Linder E: **Free-living amoebae protecting Legionella in water: the tip of an iceberg?** *Scandinavian Journal of Infectious Diseases* 1999, **31:**383-385.
11. Dallaire-Dufresne S, Paquet VE, Charette SJ: **[Dictyostelium discoideum: a model for the study of bacterial virulence].** *Canadian Journal of Microbiology* 2011, **57:**699-707.
12. Sandström G, Saeed A, Abd H: **Acanthamoeba-bacteria: a model to study host interaction with human pathogens.** *Current Drug Targets* 2011, **12:**936-941.
13. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, et al: **The genome of the social amoeba Dictyostelium discoideum.** *Nature* 2005, **435:**43-57.
14. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, 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, et al: **The genome of the protist parasite Entamoeba histolytica.** *Nature* 2005, **433:**865-868.
15. Bowers B, Korn ED: **The fine structure of Acanthamoeba castellanii (Neff strain). II. Encystment.** *The Journal of Cell Biology* 1969, **41:**786-805.
16. Marciano-Cabral F, Cabral G: **Acanthamoeba spp. as agents of disease in humans.** *Clinical Microbiology Reviews* 2003, **16:**273-307.
17. Pollard TD, Korn ED: **Acanthamoeba myosin. I. Isolation from Acanthamoeba castellanii of an enzyme similar to muscle myosin.** *The Journal of Biological Chemistry* 1973, **248:**4682-4690.
18. Ulsamer AG, Smith FR, Korn ED: **Lipids of Acanthamoeba castellanii. Composition and effects of phagocytosis on incorporation of radioactive precursors.** *The Journal of Cell Biology* 1969, **43:**105-114.
19. Keeling PJ, Palmer JD: **Horizontal gene transfer in eukaryotic evolution.** *Nature Reviews Genetics* 2008, **9:**605-618.
20. Merhej V, Notredame C, Royer-Carenzi M, Pontarotti P, Raoult D: **The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences.** *Molecular Biology & Evolution* 2011, **28:**3213-3223.
21. Thomas V, Greub G: **Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours.** *Intervirology* 2010, **53:**254-267.
22. Nelson WC, Bhaya D, Heidelberg JF: **Novel miniature transposable elements in thermophilic Synechococcus and their impact on an environmental population.** *Journal of Bacteriology* 2012.
23. Andersson JO: **Gene transfer and diversification of microbial eukaryotes.** *Annual Review of Microbiology* 2009, **63:**177-193.
24. Lynch M, Conery JS: **The origins of genome complexity.** *Science* 2003, **302:**1401-1404.

25. Roy SW, Fedorov A, Gilbert W: **Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100:**7158-7162.
26. Li W, Tucker AE, Sung W, Thomas WK, Lynch M: **Extensive, recent intron gains in Daphnia populations.** *Science* 2009, **326:**1260-1262.
27. Roy SW: **Intron-rich ancestors.** *Trends in Genetics* 2006, **22:**468-471.
28. Roy SW, Irimia M, Penny D: **Very little intron gain in Entamoeba histolytica genes laterally transferred from prokaryotes.** *Molecular Biology & Evolution* 2006, **23:**1824-1827.
29. Fredriksson R, Schiöth HB: **The repertoire of G-protein-coupled receptors in fully sequenced genomes.** *Molecular Pharmacology* 2005, **67:**1414-1425.
30. Hoffman CS: **Glucose sensing via the protein kinase A pathway in Schizosaccharomyces pombe.** *Biochemical Society Transactions* 2005, **33:**257-260.
31. Huang HC, Klein PS: **The Frizzled family: receptors for multiple signal transduction pathways.** *Genome Biology* 2004, **5:**234.
32. Gaillard I, Rouquier S, Giorgi D: **Olfactory receptors.** *Cellular and Molecular Life Sciences* 2004, **61:**456-469.
33. Iyer LM, Anantharaman V, Aravind L: **Ancient conserved domains shared by animal soluble guanylyl cyclases and bacterial signaling proteins.** *BMC Genomics* 2003, **4:**5.
34. Fitzpatrick DA, O'Halloran DM, Burnell AM: **Multiple lineage specific expansions within the guanylyl cyclase gene family.** *BMC Evolutionary Biology* 2006, **6:**26.
35. Jekely G: **Evolution of phototaxis.** *Philosophical Transactions of the Royal Society of London Series B, Biological sciences* 2009, **364:**2795-2808.
36. Dudley R, Jarroll EL, Khan NA: **Carbohydrate analysis of Acanthamoeba castellanii.** *Experimental Parasitology* 2009, **122:**338-343.
37. Segall JE, Kuspa A, Shaulsky G, Ecke M, Maeda M, Gaskins C, Firtel RA, Loomis WF: **A MAP kinase necessary for receptor-mediated activation of adenylyl cyclase in Dictyostelium.** *The Journal of Cell Biology* 1995, **128:**405-413.
38. Suga H, Dacre M, de Mendoza A, Shalchian-Tabrizi K, Manning G, Ruiz-Trillo I: **Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases.** *Science Signaling* 2012, **5:**ra35.
39. Lim WA, Pawson T: **Phosphotyrosine signaling: evolving a new cellular communication system.** *Cell* 2010, **142:**661-667.
40. Liu BA, Shah E, Jablonowski K, Stergachis A, Engelmann B, Nash PD: **The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes.** *Science Signaling* 2011, **4:**ra83.
41. Tan JL, Spudich JA: **Developmentally regulated protein-tyrosine kinase genes in Dictyostelium discoideum.** *Molecular and Cellular Biology* 1990, **10:**3578-3583.
42. Tan CS, Pasculescu A, Lim WA, Pawson T, Bader GD, Linding R: **Positive selection of tyrosine loss in metazoan evolution.** *Science* 2009, **325:**1686-1688.
43. Li L, Tibiche C, Fu C, Kaneko T, Moran MF, Schiller MR, Li SS, Wang E: **The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer.** *Genome Research* 2012, **22:**1222-1230.
44. Stuart LM, Ezekowitz RA: **Phagocytosis: elegant complexity.** *Immunity* 2005, **22:**539-550.
45. Watnick PI, Fullner KJ, Kolter R: **A role for the mannose-sensitive hemagglutinin in biofilm formation by Vibrio cholerae El Tor.** *Journal of Bacteriology* 1999, **181:**3606-3609.
46. Beckmann G, Bork P: **An adhesive domain detected in functionally diverse receptors.** *Trends in Biochemical Sciences* 1993, **18:**40-41.
47. Hong YC, Lee WM, Kong HH, Jeong HJ, Chung DI: **Molecular cloning and characterization of a cDNA encoding a laminin-binding protein (AhLBP) from Acanthamoeba healyi.** *Experimental Parasitology* 2004, **106:**95-102.
48. Harpaz Y, Chothia C: **Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains.** *Journal of Molecular Biology* 1994, **238:**528-539.
49. Ausubel FM: **Are innate immune signaling pathways in plants and animals conserved?** *Nature Immunology* 2005, **6:**973-979.
50. Fujita T: **Evolution of the lectin-complement pathway and its role in innate immunity.** *Nature Reviews Immunology* 2002, **2:**346-353.
51. Tissières P, Pugin J: **The role of MD-2 in the opsonophagocytosis of Gram-negative bacteria.** *Current Opinion in Infectious Diseases* 2009, **22:**286-291.

52. Alsam S, Sissons J, Dudley R, Khan NA: **Mechanisms associated with Acanthamoeba castellanii (T4) phagocytosis.** *Parasitology Research* 2005, **96:**402-409.
53. Garate M, Cubillos I, Marchant J, Panjwani N: **Biochemical characterization and functional studies of Acanthamoeba mannose-binding protein.** *Infection and Immunity* 2005, **73:**5775-5781.
54. Watanabe Y, Tateno H, Nakamura-Tsuruta S, Kominami J, Hirabayashi J, Nakamura O, Watanabe T, Kamiya H, Naganuma T, Ogawa T, Naudé RJ, Muramoto K: **The function of rhamnose-binding lectin in innate immunity by restricted binding to Gb3.** *Developmental and Comparative Immunology* 2009, **33:**187-197.
55. Hynes RO, Zhao Q: **The evolution of cell adhesion.** *The Journal of Cell Biology* 2000, **150:**F89-96.
56. Ghigo E, Kartenbeck J, Lien P, Pelkmans L, Capo C, Mege JL, Raoult D: **Ameobal pathogen mimivirus infects macrophages through phagocytosis.** *PLoS Pathogens* 2008, **4:**e1000087.
57. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, Bernardi C, Damonte G, Tonetti M: **Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses.** *Journal of Virology* 2010, **84:**8829-8838.
58. Poole S, Firtel RA, Lamar E, Rowekamp W: **Sequence and expression of the discoidin I gene family in Dictyostelium discoideum.** *Journal of Molecular Biology* 1981, **153:**273-289.
59. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, Breton C, Imberty A, Mitchell EP: **Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold.** *The Journal of Biological Chemistry* 2006, **281:**20171-20180.
60. Zambounis A, Elias M, Sterck L, Maumus F, Gachon CM: **Highly dynamic exon shuffling in candidate pathogen receptors... What if brown algae were capable of adaptive immunity?** *Molecular Biology & Evolution* 2012, **29:**1263-1276.
61. Koonin EV: **Taming of the shrewd: novel eukaryotic genes from RNA viruses.** *BMC Biology* 2010, **8:**2.
62. Aliyari R, Ding SW: **RNA-based viral immunity initiated by the Dicer family of host immune receptors.** *Immunological Reviews* 2009, **227:**176-188.
63. Singh R, Jamieson A, Cresswell P: **GILT is a critical host factor for Listeria monocytogenes infection.** *Nature* 2008, **455:**1244-1247.
64. MacMicking JD: **Interferon-inducible effector mechanisms in cell-autonomous immunity.** *Nature Reviews Immunology* 2012, **12:**367-382.
65. Peracino B, Wagner C, Balest A, Balbo A, Pergolizzi B, Noegel AA, Steinert M, Bozzaro S: **Function and mechanism of action of Dictyostelium Nramp1 (Slc11a1) in bacterial infection.** *Traffic* 2006, **7:**22-38.
66. Hug LA, Stechmann A, Roger AJ: **Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes.** *Molecular Biology & Evolution* 2010, **27:**311-324.
67. Ginger ML, Fritz-Laylin LK, Fulton C, Cande WZ, Dawson SC: **Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes.** *Protist* 2010, **161:**642-671.
68. Slamovits CH, Keeling PJ: **Pyruvate-phosphate dikinase of oxymonads and parabasalia and the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes.** *Eukaryotic Cell* 2006, **5:**148-154.
69. Loenarz C, Coleman ML, Boleininger A, Schierwater B, Holland PW, Ratcliffe PJ, Schofield CJ: **The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens.** *EMBO Reports* 2011, **12:**63-70.
70. Rytkönen KT, Storz JF: **Evolutionary origins of oxygen sensing in animals.** *EMBO Reports* 2011, **12:**3-4.
71. Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang E, Barry K, Lindquist E, Shapiro H, Bruce D, Schmutz J, Salamov A, Fey P, Gaudet P, Anjard C, Babu MM, Basu S, Bushmanova Y, van der Wel H, Katoh Kurasawa M, Dinh C, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B, et al: **Comparative genomics of the social amoebae Dictyostelium discoideum and Dictyostelium purpureum.** *Genome Biology* 2011, **12:**R20.
72. Duszenko M, Ginger ML, Brennand A, Gualdrón-López M, Colombo MI, Coombs GH, Coppens I, Jayabalasingham B, Langsley G, de Castro SL, Menna-Barreto R, Mottram JC, Navarro M, Rigden DJ, Romano PS, Stoka V, Turk B, Michels PA: **Autophagy in protists.** *Autophagy* 2011, **7:**127-158.
73. MacPherson S, Larochelle M, Turcotte B: **A fungal family of transcriptional regulators: the zinc cluster proteins.** *Microbiology and Molecular Biology Reviews* 2006, **70:**583-604.
74. Bürglin TR: **Homeodomain subtypes and functional diversity.** *Sub-Cellular Biochemistry* 2011, **52:**95-122.

75. Han Z, Firtel RA: **The homeobox-containing gene Wariai regulates anterior-posterior patterning and cell-type homeostasis in Dictyostelium.** *Development* 1998, **125:**313-325.
76. Piasecki BP, Burghoorn J, Swoboda P: **Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**12969-12974.
77. Krome K, Rosenberg K, Dickler C, Kreuzer K, Ludwig-Müller J, Ullrich-Eberius C, Scheu S, Bonkowski M: **Soil bacteria and protozoa affect root branching via effects on the auxin and cytokinin balance in plants.** *Plant and Soil* 2010, **328:**191-201.
78. Finkler A, Ashery-Padan R, Fromm H: **CAMTAs: calmodulin-binding transcription activators from plants to human.** *FEBS Letters* 2007, **581:**3893-3898.
79. Galon Y, Aloni R, Nachmias D, Snir O, Feldmesser E, Scrase-Field S, Boyce JM, Bouché N, Knight MR, Fromm H: **Calmodulin-binding transcription activator 1 mediates auxin signaling and responds to stresses in Arabidopsis.** *Planta* 2010, **232:**165-178.
80. Eichinger L, Noegel AA: **Comparative genomics of Dictyostelium discoideum and Entamoeba histolytica.** *Current Opinion in Microbiology* 2005, **8:**606-611.
81. Song J, Xu Q, Olsen R, Loomis WF, Shaulsky G, Kuspa A, Sucgang R: **Comparing the Dictyostelium and Entamoeba genomes reveals an ancient split in the Conosa lineage.** *PLoS Computational Biology* 2005, **1:**e71.
82. Raoult D, Boyer M: **Amoebae as genitors and reservoirs of giant viruses.** *Intervirology* 2010, **53:**321-329.
83. Lurie-Weinberger MN, Gomez-Valero L, Merault N, Glöckner G, Buchrieser C, Gophna U: **The origins of eukaryotic-like proteins in Legionella pneumophila.** *International Journal of Medical Microbiology* 2010, **300:**470-481.
84. Plate L, Marletta MA: **Nitric Oxide Modulates Bacterial Biofilm Formation through a Multicomponent Cyclic-di-GMP Signaling Network.** *Molecular Cell* 2012, **46:**449-460.
85. Martel CM: **Conceptual bases for prey biorecognition and feeding selectivity in the microplanktonic marine phagotroph Oxyrrhis marina.** *Microbial Ecology* 2009, **57:**589-597.
86. Anantharaman V, Iyer LM, Aravind L: **Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation.** *Annual Review of Microbiology* 2007, **61:**453-475.
87. Aderem A, Underhill DM: **Mechanisms of phagocytosis in macrophages.** *Annual Review of Immunology* 1999, **17:**593-623.
88. Boettner DR, Huston CD, Linford AS, Buss SN, Houpt E, Sherman NE, Petri WA, Jr.: **Entamoeba histolytica phagocytosis of human erythrocytes involves PATMK, a member of the transmembrane kinase family.** *PLoS Pathogens* 2008, **4:**e8.
89. Sun T, Kim L: **Tyrosine phosphorylation-mediated signaling pathways in dictyostelium.** *Journal of Signal Transduction* 2011, **2011:**894351.
90. Turkarslan S, Reiss DJ, Gibbins G, Su WL, Pan M, Bare JC, Plaisier CL, Baliga NS: **Niche adaptation by expansion and reprogramming of general transcription factors.** *Molecular Systems Biology* 2011, **7:**554.
91. Lohan AJ, Gray MW: **Analysis of 5'- or 3'-terminal tRNA editing: mitochondrial 5' tRNA editing in Acanthamoeba castellanii as the exemplar.** *Methods in Enzymology* 2007, **424:**223-242.
92. Spencer DF, Schnare MN, Gray MW: **Isolation of wheat mitochondrial DNA and RNA.** In *Modern Methods of Plant Analysis New Series. Volume* 14. Edited by Linskens HF, Jackson JF: Springer-Verlag, Berlin; 1992: 347-360: *Modern Methods of Plant Analysis*].
93. Weissenmayer BA, Prendergast JG, Lohan AJ, Loftus BJ: **Sequencing illustrates the transcriptional response of Legionella pneumophila during infection and identifies seventy novel small non-coding RNAs.** *PloS One* 2011, **6:**e17570.
94. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A: **Transcriptome analysis by strand-specific sequencing of complementary DNA.** *Nucleic Acids Research* 2009, **37:**e123.
95. Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ: **Identification of genetic variants using bar-coded multiplexed sequencing.** *Nature Methods* 2008, **5:**887-893.
96. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Research* 2008, **18:**821-829.
97. Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F: **Annotating genomes with massive-scale RNA sequencing.** *Genome Biology* 2008, **9:**R175.

98. Korf I: **Gene finding in novel genomes.** *BMC Bioinformatics* 2004, **5:**59.
99. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M: **MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes.** *Genome Research* 2008, **18:**188-196.
100. Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglir L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME: **Apollo: a sequence annotation editor.** *Genome Biology* 2002, **3:**RESEARCH0082.
101. Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, Hall N, Caler EV: **New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information.** *PLoS Neglected Tropical Diseases* 2010, **4:**e716.
102. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R: **The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology.** *Nucleic Acids Research* 2004, **32:**D262-266.
103. Rattei T, Tischler P, Gotz S, Jehl M-A, Hoser J, Arnold R, Conesa A, Mewes H-W: **SIMAP--a comprehensive database of pre-calculated protein sequence similarities, domains, annotations and clusters.** *Nucleic Acids Research* 2010, **38:**D223-226.
104. Frickey T, Lupas AN: **PhyloGenie: automated phylome generation and analysis.** *Nucleic Acids Research* 2004, **32:**5231-5238.
105. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Molecular Biology & Evolution* 2011, **28:**2731-2739.

Figure Legends

Figure 1. Measures of the composition of the *Ac* genome based on sequence similarity. For each protein, the best BLASTP hit to the non-redundant database, i.e. the match with the lowest e-value, was recovered and the classification of the corresponding organism was extracted according to NCBI taxonomy. The central bar represents the full complement of annotated *Ac* genes exhibiting a best BLASTP hit respectively against the four kingdoms: Eukaryota (blue), Bacteria (red), Archaea (green) and viruses (purple) with orphan genes depicted in yellow. Results for Eukaryota are subdivided according to the major taxonomic phyla in varying shades of blue. Subdivisions of phyla within the Bacteria (red shading), Archaea (green shading) and viruses (purple shading) are depicted in the expanded upper and lower sidebars.

Figure 2. Predicted LGT-derived genes from Bacteria, Archaea and viruses encoded in the genomes of free-living and parasitic amoebae. LGT-derived genes were predicted using a phylogenomics approach consisting of an initial similarity-based screening using SIMAP [103], several filtering steps to extract amoebal proteins with prokaryotic best hits, followed by automatic calculation and manual inspection of phylogenetic trees using PhyloGenie and PHAT [104]. (A) Percentage of lineage-specific LGT candidates in each genome; the absolute number of LGT candidates per genome is indicated next to each bar. (B) Heat map illustrating the Bray-Curtis similarity of the taxonomic affiliation (at the level of classes within the domain Bacteria) of putative LGT donors. (C) Ecological classification of putative LGT donors with respect to their oxygen requirement and association with a host. The ecology of putative donors was extrapolated from the lifestyles of the respective closest extant relatives.

Figure 3. Phylogenetic tree of rhodopsins from Amoebozoa, algae, bacteria and fungi. The tree was constructed by the Neighbor-joining method based on amino acid sequence of rhodopsin domain using MEGA version 5 [105]. The scale bar indicates the number of substitutions per site. Detailed rhodopsin information is listed in Additional file 1, Table S3.6.1.

Figure 4. The phosphotyrosine signaling circuitry of *Ac*. (A) Phosphotyrosine signaling is modulated by the writers (PTKs), erasers (PTPs) and readers (Src homology 2, SH2; phosphotyrosine binding, PTB). (B) (B) The total number of PTKs, classical PTPs (total PTPs) and SH2 encoded genes across multiple eukaryote genomes. Highlighted in yellow are branches that compose a complete phosphotyrosine signaling circuit. The branched divergence times and lengths in millions of years (mya) are indicated. (C) The percentage of

the genome devoted to encoding PTKs and SH2 domains. (D) *Ac* displays the greatest ratio of SH2:PTKs compared to other eukaryotes.

Figure 5. Potential PRRs in *Ac* LBP/BPI= lipopolysaccharide binding protein/ bactericidal permeability-increasing protein; C-lectin = C-type lectin: MBP = mannose binding protein; SUEL = D-galactoside/L-rhamnose binding SUEL lectin domain containing; NB-ARC-TPR = NB-ARC tetratricopeptide repeat containing protein; ERVR = endogenous virus receptor.

Additional files:

Supplementary online material: Additional_File_1_SOM.pdf

Supplementary material supporting LGT analysis: Additional_File-2_LGT_Analysis.xls

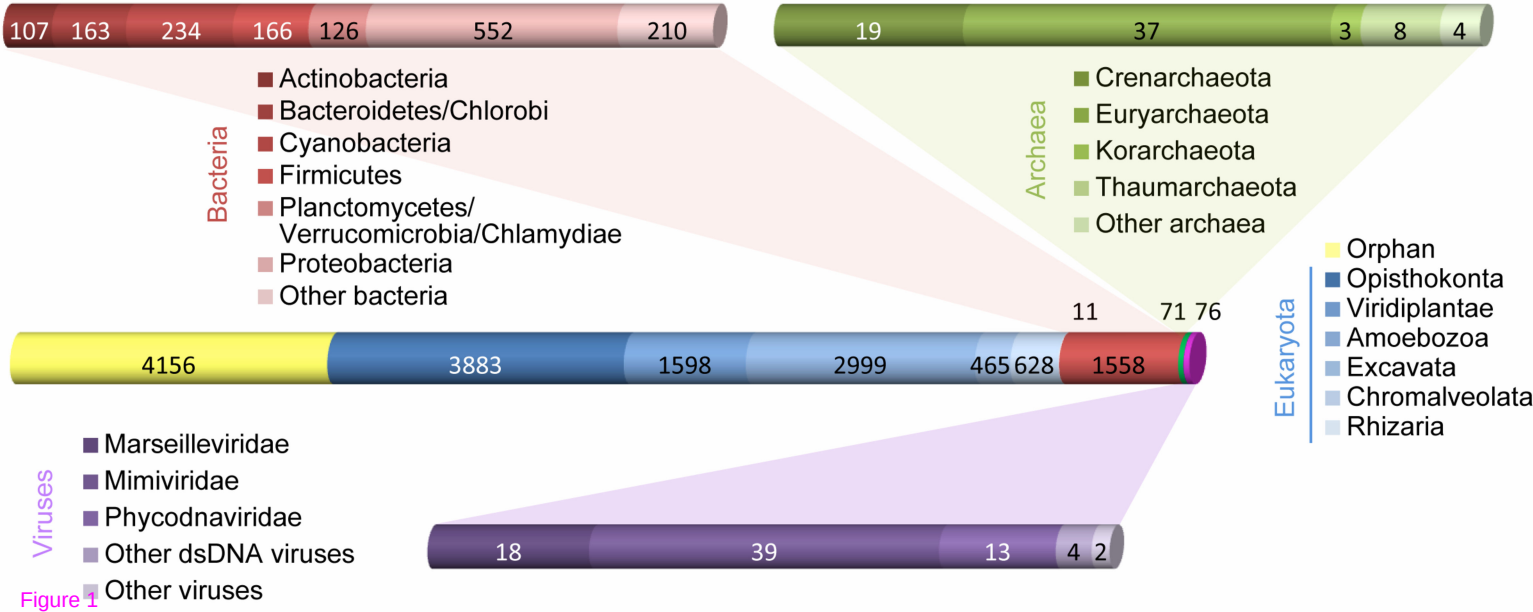
107
163
234
166
126
552
210
Bacteria
Actinobacteria
Bacteroidetes/Chlorobi
Cyanobacteria
Firmicutes
Planctomycetes/
Verrucomicrobia/Chlamydiae
Proteobacteria
Other bacteria
19
37
3
8
4
Archaea
Crenarchaeota
Euryarchaeota
Korarchaeota
Thaumarchaeota
Other archaea
11
71
76
4156
3883
1598
2999
465
628
1558
Eukaryota
Orphan
Opisthokonta
Viridiplantae
Amoebozoa
Excavata
Chromalveolata
Rhizaria
Viruses
Marseilleviridae
Mimiviridae
Phycodnaviridae
Other dsDNA viruses
Other viruses
18
39
13
4
2

Figure 1

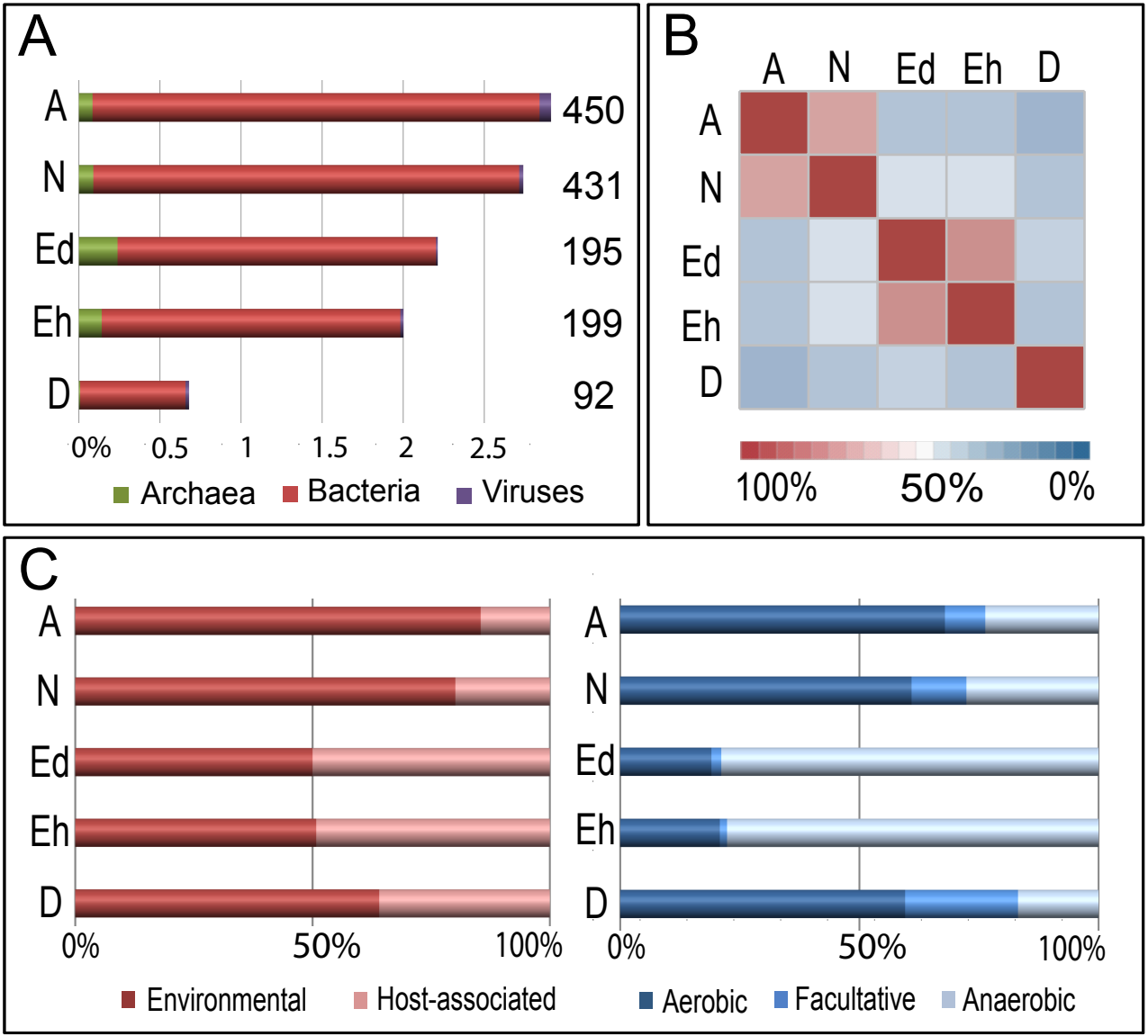
A
A
N
Ed
Eh
D
450
431
195
199
92
0%
0.5
1
1.5
2
2.5
Archaea
Bacteria
Viruses
B
A
N
Ed
Eh
D
A
N
Ed
Eh
D
100%
50%
0%
C
A
N
Ed
Eh
D
0%
50%
100%
Environmental
Host-associated
A
N
Ed
Eh
D
0%
50%
100%
Aerobic
Facultative
Anaerobic

Figure 2

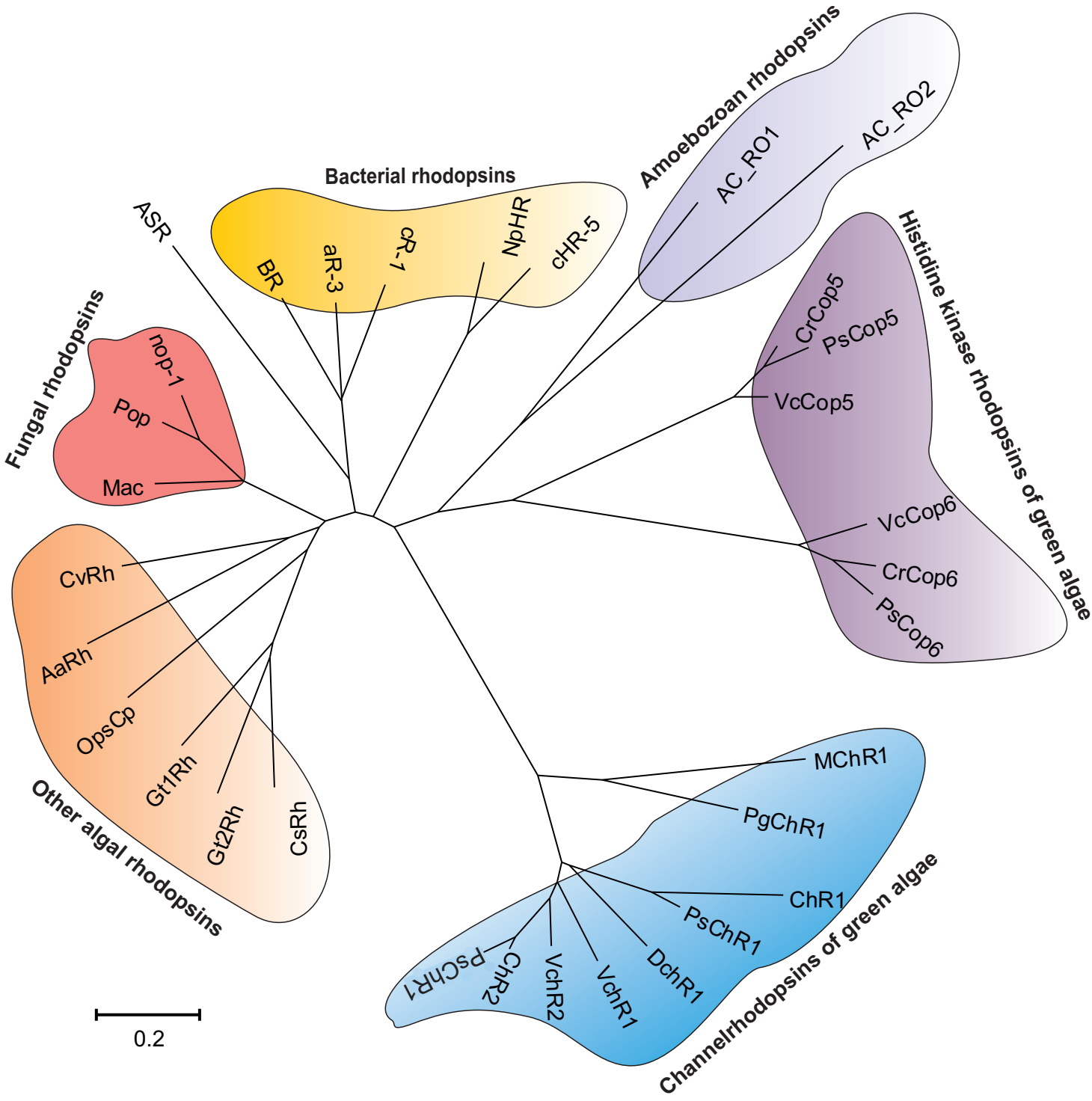
Amoebozoan rhodopsins
AC_RO1
AC_RO2
Bacterial rhodopsins
ASR
BR
aR-3
cR-1
NpHR
cHR-5
Fungal rhodopsins
nop-1
Pop
Mac
Histidine kinase rhodopsins of green algae
CrCop5
PsCop5
VcCop5
VcCop6
CrCop6
PsCop6
CvRh
AaRh
OpsCp
Gt1Rh
Gt2Rh
CsRh
Other algal rhodopsins
MChR1
PgChR1
ChR1
PsChR1
DchR1
VchR1
VchR2
PsChR2
Channelrhodopsins of green algae
0.2

Figure 3

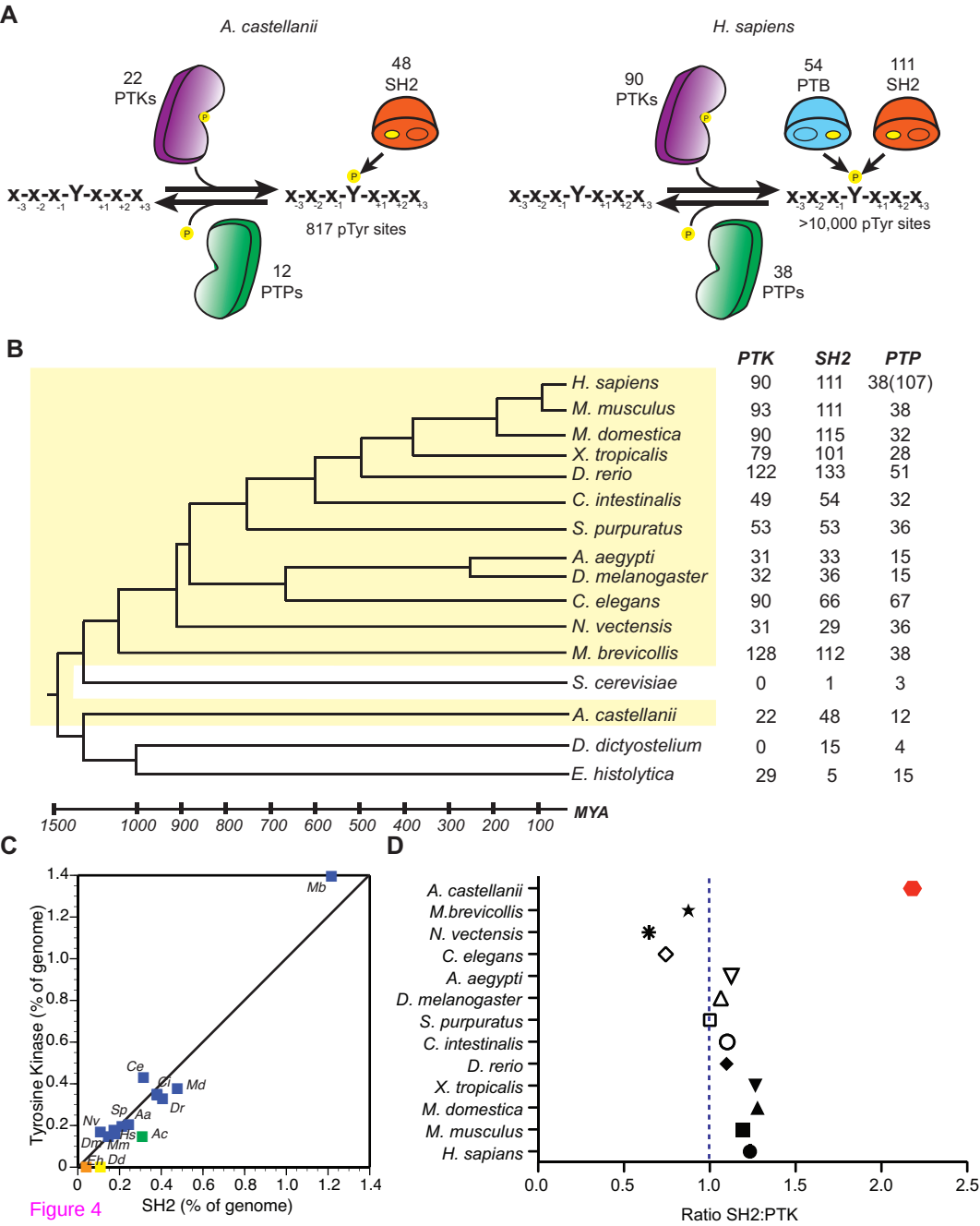
A
A. castellanii
22
PTKs
48
SH2
x-x-x-Y-x-x-x
x-x-x-Y-x-x-x
817 pTyr sites
12
PTPs
H. sapiens
90
PTKs
54
PTB
111
SH2
x-x-x-Y-x-x-x
x-x-x-Y-x-x-x
>10,000 pTyr sites
38
PTPs
B
PTK
SH2
PTP
H. sapiens 90 111 38(107)
M. musculus 93 111 38
M. domestica 90 115 32
X. tropicalis 79 101 28
D. rerio 122 133 51
C. intestinalis 49 54 32
S. purpuratus 53 53 36
A. aegypti 31 33 15
D. melanogaster 32 36 15
C. elegans 90 66 67
N. vectensis 31 29 36
M. brevicollis 128 112 38
S. cerevisiae 0 1 3
A. castellanii 22 48 12
D. dictyostelium 0 15 4
E. histolytica 29 5 15
1500
1000
900
800
700
600
500
400
300
200
100
MYA
C
Tyrosine Kinase (% of genome)
SH2 (% of genome)
Mb
Ce
Ci
Md
Dr
Sp
Aa
Nv
Hs
Ac
Dm
Mm
Eh
Dd
D
A. castellanii
M.brevicollis
N. vectensis
C. elegans
A. aegypti
D. melanogaster
S. purpuratus
C. intestinalis
D. rerio
X. tropicalis
M. domestica
M. musculus
H. sapians
Ratio SH2:PTK

Figure 4

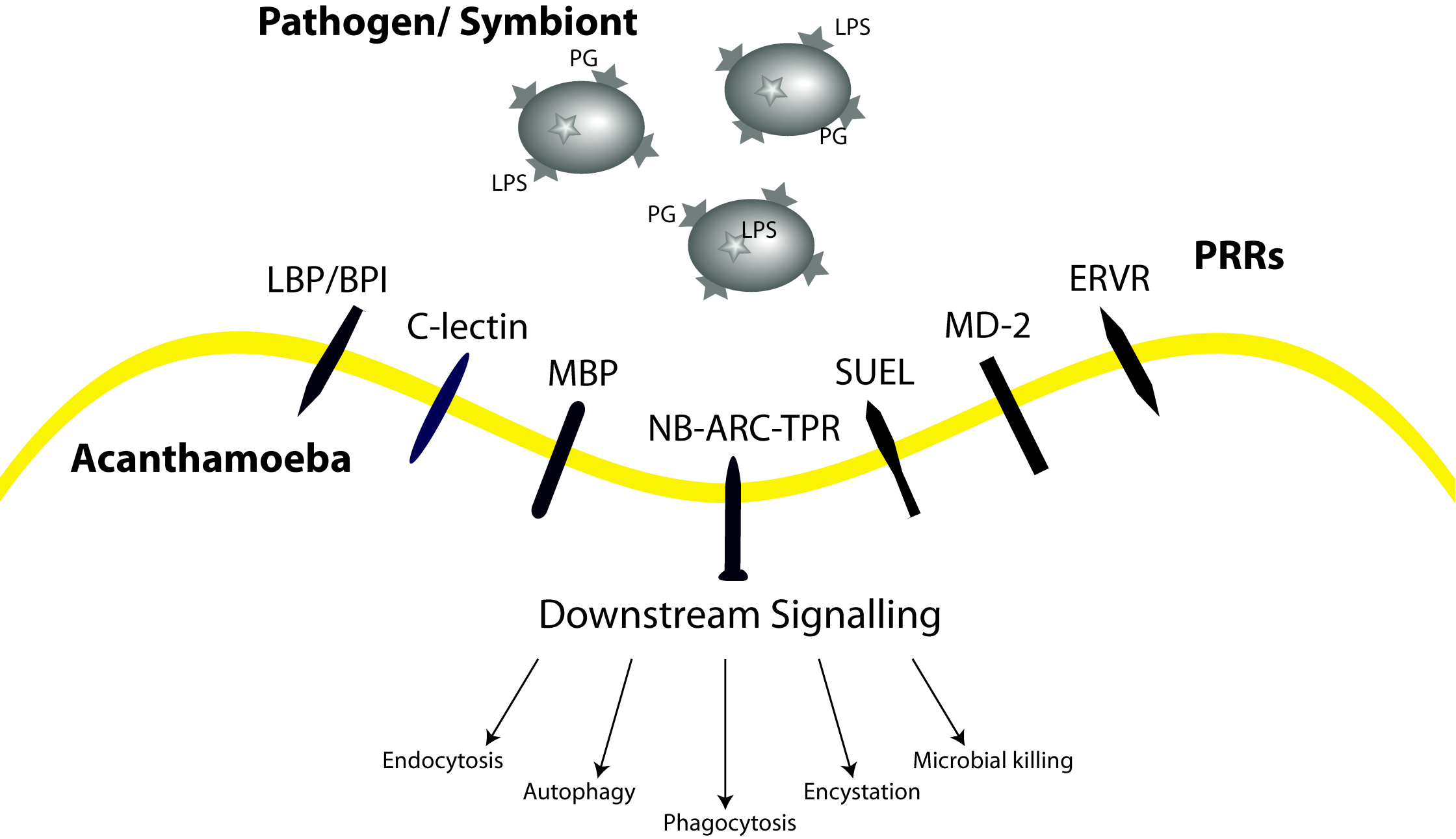
Pathogen/ Symbiont
PG
LPS
LPS
PG
PG
LPS
PRRs
LBP/BPI
C-lectin
MBP
NB-ARC-TPR
SUEL
MD-2
ERVR
Acanthamoeba
Downstream Signalling
Endocytosis
Autophagy
Phagocytosis
Encystation
Microbial killing

Figure 5

Additional files provided with this submission:

Additional file 1: Additional_File_1.pdf, 11033K
http://genomebiology.com/imedia/7811959619020565/supp1.pdf
Additional file 2: Additional_File_2_LGT_Analysis.xls, 400K
http://genomebiology.com/imedia/3633275339020571/supp2.xls