nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0477-7

In the format provided by the authors and unedited.

Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic

John A. Burns*, Alexandros A. Pittis and Eunsoo Kim*

Sackler Institute for Comparative Genomics and Division of Invertebrate Zoology, American Museum of Natural History, New York, NY, USA.
*e-mail: jburns@amnh.org; ekim1@amnh.org

Supplementary Discussion 1

Trophic mode surprises

When the model was tested against against an evolutionarily broad range of eukaryotes (Supplementary Figure 1, Extended Data Table 6), the predictions were largely in agreement with previous observations on the tested organisms; however, two interesting cases were found. First, the choanoflagellates, represented by *Monosiga brevicollis* and *Salpingoeca rosetta*, were predicted to be prototrophic. This was contrary to our prior analysis, which suggested that phagocytotic organisms tend to lose genes that are involved in small-molecule biosynthesis[1]. This curiosity in choanoflagellate genomes nevertheless has been observed before, and is attributed to acquisition of those biosynthesis pathways by lateral gene transfer (LGT) from prokaryotic prey, recovering the functions that were presumably absent in the choanoflagellate ancestor, which was hypothesized to give the choanoflagellates an advantage in oligotrophic waters[2].

Second, our model predicted several oomycetes to be phagocytotic (Supplementary Figure 1, Extended Data Table 6). This is rather unexpected because oomycetes are presumed obligate osmotrophs that do not use phagocytosis for feeding[3]. They have parallels to fungi in morphology, feeding mode, and gene content, with several genes related to osmotrophy apparently derived from fungi via LGT[4]. They also have a complex life cycle with wall-less flagellate zoospore stages[5]. It is clear that many oomycetes from several clades contain genes that are shared with diverse phagocytes and that tend to be absent in fungi, land plants, green and red algae (Supplementary Figure 1, Extended Data Table 6). It is presently unclear why the oomycetes maintain this set of proteins; however, one study discovered that oomycete zoospores use a process of transient fusion of large vesicles with the plasma membrane for selective release of vesicle components, a process that had only previously been known from mammalian cells[6]. The transient fusion process may have mechanistic overlaps with phagocytosis[7].

Supplementary Discussion 2

Observed Phagosome compared to proteins found by comparative genomics

Of those 485 proteins, 54 were also detected in isolated phagosomes (Figure 4, Extended Data Table 7). Notably, the overlap between the two datasets contains four WASP and Scar homologue (WASH) complex proteins[8], two integrins[9], three copper metabolism gene MURR1 (COMM) domain proteins[10], and three dynein motor protein subunits[11] (Extended Data Table 7). Notable differences that are in the predictive set, but not isolated phagosomes include the dynein partner dynactin[12], six Bardet-Biedl

complex proteins[13], and cytoskeleton regulating proteins like alpha-parvin[14] and actin-bundling protein abpB[15] (Extended Data Table 8). Otherwise, the majority of proteins identified in the phagosome compartment (705; 93%) are shared among most eukaryotes. This includes ribosomal proteins, metabolic proteins, and proteins involved in vesicle movement in cells (Extended Data Table 9).

Supplementary Discussion 3

Focus on the Asgard archaea

While the lack of phagocytotic signal in the Asgard archaea is unremarkable, we explored our data further for proteins in the Asgard lineage that are unique among prokaryotes[16]. Among our 14,095 HMMs, there is a set of 54 proteins with homologs in the Asgard archaea that have no homologs in any other prokaryote. Those proteins include previously described members of the ESCRT family of proteins, and also sec23 and sec24 domain proteins (Extended Data Table 13). Six of those 54 proteins (Extended Data Table 14) are also found in the set of proteins that make up the phagocyte models; they are enriched in phagocytes and depleted in non-phagocyte genomes. Those six proteins in the Asgard archaea are a putative homolog of Bardet-Biedl syndrome 2 (BBS2) protein (the homology is driven by the WD40/YVTN domain found in both the eukaryote and archaeal proteins), one CD36 family protein with homology to lysosomal membrane proteins 2 and 2A (LIMPII, LIMPII-1), and three members of the Ragulator complex (Ras-related GTP-binding protein A, and two roadblock domain proteins). The cellular compartments and processes most directly associated with those six proteins and domains are the lysosome[17], eukaryotic flagella/cilia[18], and amino acid sensing[19]. For the other 48 proteins with a broad distribution in eukaryotes, roles in conserved processes like cell division[20], an archaeal endomembrane system[21], or even vesicular membrane trafficking[22,23] are plausible explanations for their presence in the Asgard archaea. Even so, given that the Asgard archaea look like typical prokaryotes in our model of phagocytosis, their membrane dynamics are not likely to encompass the level of complexity seen in eukaryotic phagocytes. It also should be mentioned that when the same analysis was performed on other prokaryotic groups as controls, comparable numbers of eukaryote-like proteins were identified. For instance, there are 269 proteins in the proteobacteria that are not found in other prokaryotes, 9 of which overlap with the phagocyte enriched set (more than the corresponding overlap observed from the Asgard archaea, albeit a smaller proportion), and likewise 46 proteins in PVC group bacteria, 3 of which overlap with the phagocyte enriched set. Thus, while the identified eukaryote-like genes in the Asgard archaea are illuminating, other prokaryote groups also have unique proteins that overlap the functions enriched among phagocytes identified here.

Supplementary Discussion 4

Functional characterization of proteins with bacterial, archaeal, or eukaryote specific affinity

Overrepresented functions in the 567 HMMs with bacterial affinity were determined by considering gene ontology (GO) molecular function (MF) enrichment in those proteins when compared to all 775 HMMs with prokaryote homologs. Nearly all of the proteins (31 out of 34) annotated with calcium binding and signaling modalities have bacterial affinities (Supplementary Table 5, Extended Data Table 16). These include homologs of calcium responsive calpain cysteine proteases[24,25] and annexin[26]. Calcium signaling is a key element in phagocytosis, particularly with regards to phagosome maturation[27,28], and evidence points to eukaryotic calcium signaling having roots in the bacteria[29]. Other modalities with potential bacterial origins include carbohydrate binding (20 out of 22 proteins), and actin binding molecular functions (16 out of 20). Bacterial homology to actin binding proteins is largely derived from a WD-40 repeat (β-propeller) domain, found here in actin interacting protein 1, coronin B, and actin-related protein 2/3 complex subunit 1A. The origin of the WD-40 domain in eukaryotes cannot be definitively determined as having originated in bacteria or archaea[22]; however at least one study suggests the WD-40 domain to be of potential bacterial origin by considering functional gene clustering in prokaryotes and eukaryotes[30]. In the carbohydrate-binding category are several lectins and sugar-binding receptors, including galectin-9, l-type lectin-domain containing receptor kinase IV.1, macrophage mannose receptor 1, malectin-A, and vesicular integral-membrane protein VIP36. Lectins are sugar binding proteins that can act in cell-cell binding in bacterial co-aggregation[31], or as phagocytosis receptors in eukaryotes[32]. The eukaryote lectin domain may also have bacterial origins[30]. Another important class of proteins found among the proteins with bacterial affinity is those with hydrolase function (118 out of 154 annotated proteins). These include proteases, lipases, amidases, and deoxyribonucleases (DNases). Importantly, our phagocyte enrichment analysis detected the enzyme DNase II, a protein that degrades double stranded DNA in the acidic lysosome, whose evolutionary history is linked to the origins of phagocytosis in eukaryotes[33]. Prior analysis of DNase II in eukaryote and prokaryote genomes revealed a single instance of DNase II in a pathogenic betaproteobacterium genus, *Burkholderia*, that was likely the result of lateral gene transfer from host to pathogen[33]. Using current genomic information, we identified eleven additional prokaryote genomes containing DNase II homologs in free-living gamma-proteobacteria, bacteroidetes, beta-proteobacteria, and acidobacteria. A maximum-likelihood phylogenetic tree of DNase II homologs shows that the DNase II sequence *Burkholderia pseudomallei*, forms a well-supported clade with four other bacterial species (Supplementary Figure 3). The other seven bacterial species form a second well-supported, distantly related clade (Supplementary Figure 3) The

presence of those two diverse, distinct bacterial clades opens up the possibility of a bacterial origin of this key phagocytosis protein, contrary to the initial findings that indicated that the bacteria must have obtained DNase II by LGT from eukaryotes[33].

LCA analysis of our data identifies several proteins associated with the phagosome that have archaeal affinity, including ribosomal proteins[34,35], some small GTPases[36], V-type ATPase proteins[37], actin[38,39], tubulin[40], and proteasomal subunits[34,41] (Supplementary Table 6, Extended Data Table 17). Each of these groups is strongly associated with archaeal contributions to eukaryotic genomes. We also found one example of a peptide in the Asgard lineage Heimdallarchaeota with homology to eukaryote CD36 proteins, which are lipid binding glycoproteins with a role in phagocytosis[42]. Proteins unique to eukaryotes include enrichment of members of the Ras superfamily GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs), motor proteins including myosins and kinesin, and actin binding-cytoskeletal regulators (Supplementary Table 7, Extended Data Table 18), suggesting that early eukaryotes added new functions to regulate the cytoskeleton (GAPs and GEFs) and make use of it for membrane deformation[43] and fusion[44] (motor proteins, actin branching/bundling proteins, and SNAP receptors (SNAREs)).

1 Burns, J. A., Paasch, A., Narechania, A. & Kim, E. Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. *Genome biology and evolution* **7**, 3047-3061 (2015).

2 Tucker, R. P. Horizontal gene transfer in choanoflagellates. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **320**, 1-9 (2013).

3 Richards, T. A. & Talbot, N. J. Plant parasitic oomycetes such as Phytophthora species contain genes derived from three eukaryotic lineages. *Plant signaling & behavior* **2**, 112-114 (2007).

4 Richards, T. A. *et al.* Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences* **108**, 15258-15263 (2011).

5 Mendoza, L., Hernandez, F. & Ajello, L. Life cycle of the human and animal oomycete pathogen Pythium insidiosum. *Journal of Clinical Microbiology* **31**, 2967-2973 (1993).

6 Zhang, W., Blackman, L. M. & Hardham, A. R. Transient fusion and selective secretion of vesicle proteins in Phytophthora nicotianae zoospores. *PeerJ* **1**, e221 (2013).

7 Duclos, S. *et al.* Rab5 regulates the kiss and run fusion between phagosomes and endosomes and the acquisition of phagosome leishmanicidal properties in RAW 264.7 macrophages. *Journal of cell science* **113**, 3531-3541 (2000).

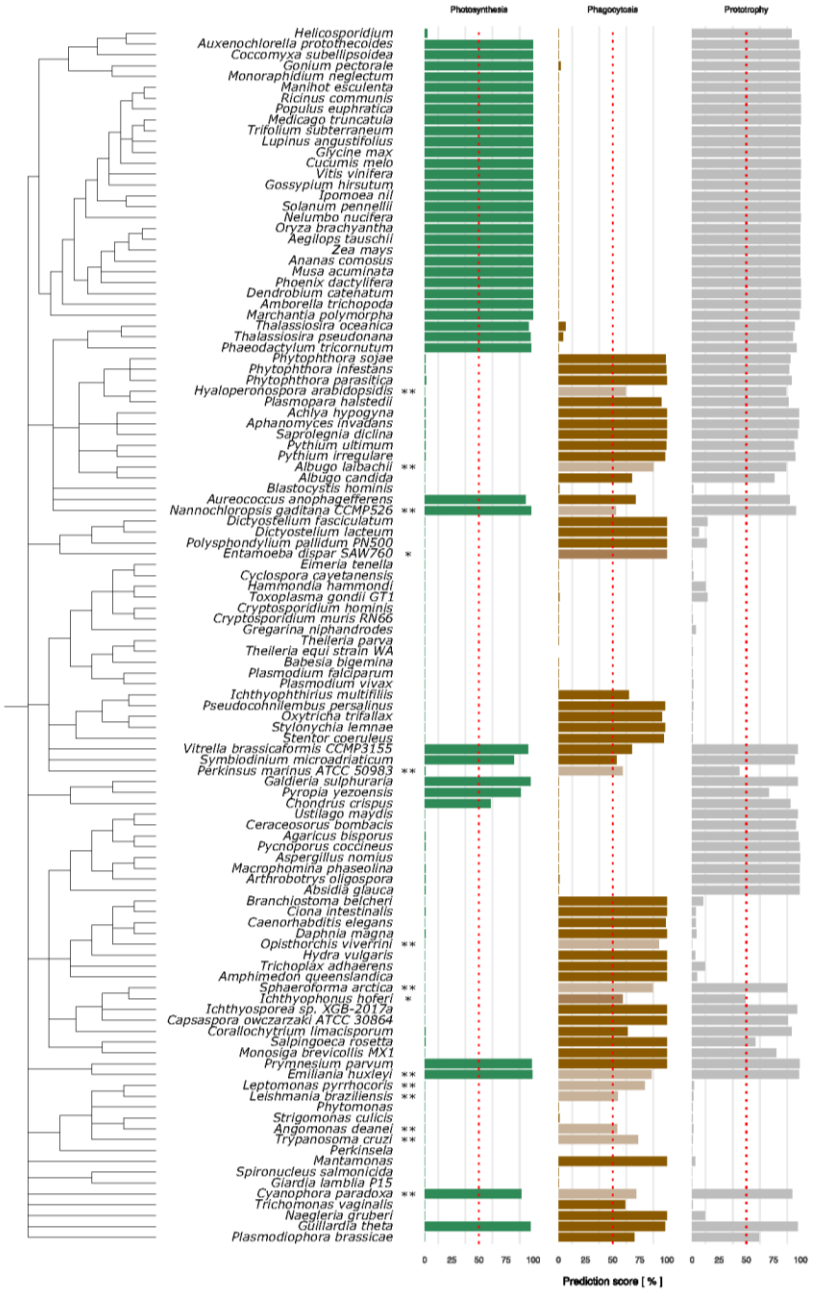
8 Buckley, C. M. *et al.* WASH drives early recycling from macropinosomes and phagosomes to maintain surface phagocytic receptors. *Proceedings of the National Academy of Sciences* **113**, E5906-E5915 (2016).

9 Brown, E. J. Complement receptors and phagocytosis. *Curr Opin Immunol* **3**, 76-82 (1991).

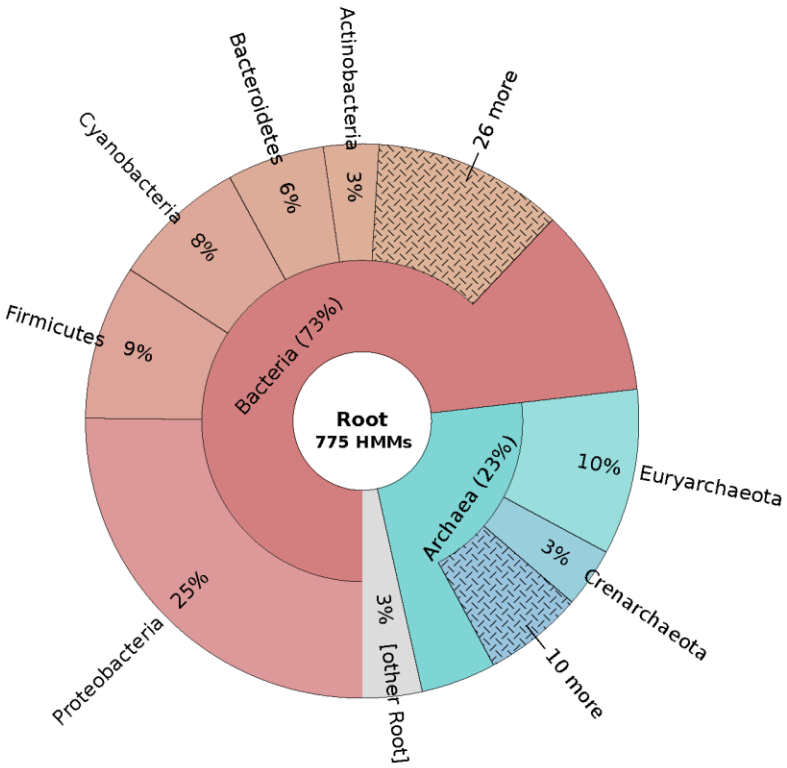
10 Dill, B. D. *et al.* Quantitative proteome analysis of temporally resolved phagosomes following uptake via key phagocytic receptors. *Mol Cell Proteomics* **14**, 1334-1349, doi:10.1074/mcp.M114.044594 (2015).
11 Harrison, R. E., Bucci, C., Vieira, O. V., Schroer, T. A. & Grinstein, S. Phagosomes fuse with late endosomes and/or lysosomes by extension of membrane protrusions along microtubules: role of Rab7 and RILP. *Mol Cell Biol* **23**, 6494-6506 (2003).
12 Driskell, O. J., Mironov, A., Allan, V. J. & Woodman, P. G. Dynein is required for receptor sorting and the morphogenesis of early endosomes. *Nature cell biology* **9**, 113-120 (2007).
13 Langousis, G. *et al.* Loss of the BBSome perturbs endocytic trafficking and disrupts virulence of Trypanosoma brucei. *Proceedings of the National Academy of Sciences* **113**, 632-637 (2016).
14 Wehrle-Haller, B. Structure and function of focal adhesions. *Current opinion in cell biology* **24**, 116-124 (2012).
15 Furukawa, R. & Fechheimer, M. Differential localization of α-actinin and the 30 kD actin-bundling protein in the cleavage furrow, phagocytic cup, and contractile vacuole of Dictyostelium discoideum. *Cytoskeleton* **29**, 46-56 (1994).
16 Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* (2017).
17 Kuronita, T. *et al.* A role for the lysosomal membrane protein LGP85 in the biogenesis and maintenance of endosomal and lysosomal morphology. *Journal of cell science* **115**, 4117-4131 (2002).
18 Nachury, M. V. *et al.* A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* **129**, 1201-1213 (2007).
19 Roustan, V., Jain, A., Teige, M., Ebersberger, I. & Weckwerth, W. An evolutionary perspective of AMPK–TOR signaling in the three domains of life. *Journal of experimental botany* **67**, 3897-3907 (2016).
20 Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L. & Bell, S. D. A role for the ESCRT system in cell division in archaea. *Science* **322**, 1710-1713 (2008).
21 Heimerl, T. *et al.* A complex endomembrane system in the archaeon Ignicoccus hospitalis tapped by Nanoarchaeum equitans. *Frontiers in microbiology* **8**, 1072 (2017).
22 Dacks, J. B. & Robinson, M. S. Outerwear through the ages: evolutionary cell biology of vesicle coats. *Current Opinion in Cell Biology* **47**, 108-116 (2017).
23 Rout, M. P. & Field, M. C. The Evolution of Organellar Coat Complexes and Organization of the Eukaryotic Cell. *Annual Review of Biochemistry* (2017).
24 Hallett, M. B. & Dewitt, S. Ironing out the wrinkles of neutrophil phagocytosis. *Trends in cell biology* **17**, 209-214 (2007).
25 Rawlings, N. D. Bacterial calpains and the evolution of the calpain (C2) family of peptidases. *Biology direct* **10**, 66 (2015).
26 Kodavali, P. K., Dudkiewicz, M., Pikuła, S. & Pawłowski, K. Bioinformatics analysis of bacterial annexins–putative ancestral relatives of eukaryotic annexins. *PloS one* **9**, e85428 (2014).
27 Nunes, P. & Demaurex, N. The role of calcium signaling in phagocytosis. *Journal of leukocyte biology* **88**, 57-68 (2010).
28 Fairn, G. D. & Grinstein, S. How nascent phagosomes mature to become phagolysosomes. *Trends in immunology* **33**, 397-405 (2012).
29 Plattner, H. & Verkhratsky, A. The ancient roots of calcium signalling evolutionary tree. *Cell Calcium* **57**, 123-132 (2015).
30 Dunin-Horkawicz, S., Kopec, K. O. & Lupas, A. N. Prokaryotic ancestry of eukaryotic protein networks mediating innate immunity and apoptosis. *Journal of molecular biology* **426**, 1568-1582 (2014).

31 Abeygunawardana, C., Bush, C. A. & Cisar, J. O. Complete structure of the cell surface polysaccharide of Streptococcus oralis ATCC 10557: a receptor for lectin-mediated interbacterial adherence. *Biochemistry* **30**, 6528-6540 (1991).
32 Ofek, I. & Sharon, N. Lectinophagocytosis: a molecular mechanism of recognition between cell surface sugars and lectins in the phagocytosis of bacteria. *Infection and Immunity* **56**, 539 (1988).
33 Shpak, M., Kugelman, J. R., Varela-Ramirez, A. & Aguilera, R. J. The phylogeny and evolution of deoxyribonuclease II: An enzyme essential for lysosomal DNA degradation. *Molecular phylogenetics and evolution* **47**, 841-854 (2008).
34 Hartman, H. & Fedorov, A. The origin of the eukaryotic cell: a genomic investigation. *Proceedings of the National Academy of Sciences* **99**, 1420-1425 (2002).
35 Ponpuak, M. *et al.* Delivery of cytosolic components by autophagic adaptor protein p62 endows autophagosomes with unique antimicrobial properties. *Immunity* **32**, 329-341 (2010).
36 Wuichet, K. & Søgaard-Andersen, L. Evolution and diversity of the Ras superfamily of small GTPases in prokaryotes. *Genome biology and evolution* **7**, 57-70 (2014).
37 Gogarten, J. P. *et al.* Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proceedings of the National Academy of Sciences* **86**, 6661-6665 (1989).
38 Ettema, T. J., Lindås, A. C. & Bernander, R. An actin-based cytoskeleton in archaea. *Molecular microbiology* **80**, 1052-1061 (2011).
39 Izoré, T., Kureisaite-Ciziene, D., McLaughlin, S. H. & Löwe, J. Crenactin forms actin-like double helical filaments regulated by arcadin-2. *eLife* **5**, e21600 (2016).
40 Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biology direct* **7**, 10 (2012).
41 Lee, W. L., Kim, M.-K., Schreiber, A. D. & Grinstein, S. Role of ubiquitin and proteasomes in phagosome maturation. *Molecular biology of the cell* **16**, 2077-2090 (2005).
42 Ryeom, S. W., Sparrow, J. R. & Silverstein, R. L. CD36 participates in the phagocytosis of rod outer segments by retinal pigment epithelium. *Journal of cell science* **109**, 387-395 (1996).
43 Herant, M., Heinrich, V. & Dembo, M. Mechanics of neutrophil phagocytosis: experiments and quantitative models. *Journal of cell science* **119**, 1903-1913 (2006).
44 Hatsuzawa, K. *et al.* Involvement of syntaxin 18, an endoplasmic reticulum (ER)-localized SNARE protein, in ER-mediated phagocytosis. *Molecular biology of the cell* **17**, 3964-3977 (2006).

Supplementary Figures

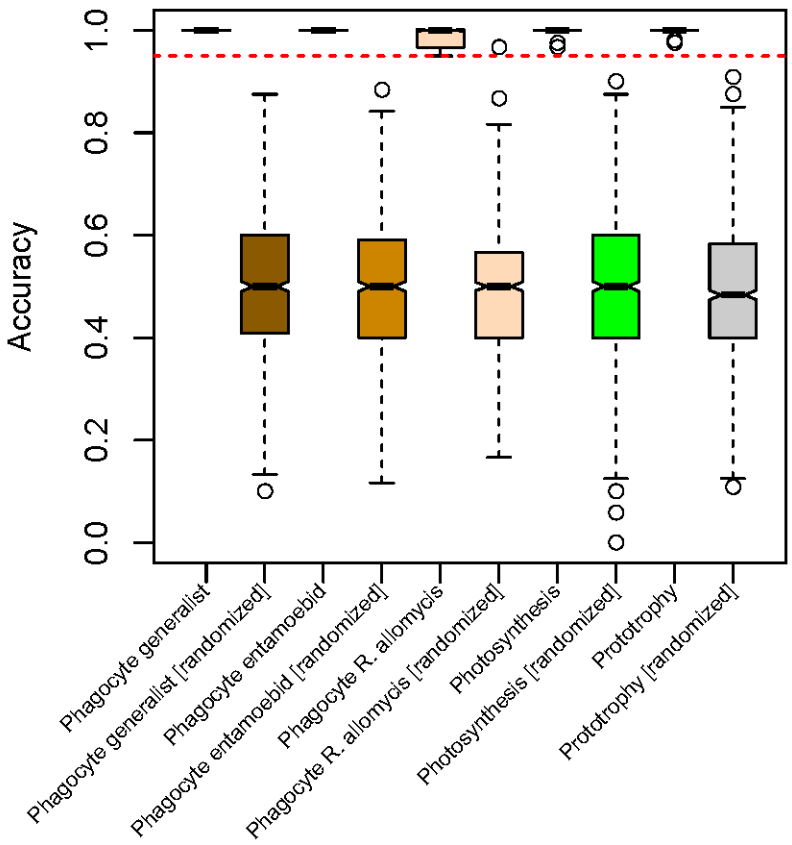


Supplementary Figure 1. Predictions of phagocytosis, photosynthesis, and prototrophy across 112 diverse eukaryotes. Bars indicate the prediction score; organisms with a prediction over 50% for a trophic character are considered positive for that trait. The three phagocytosis predictions are reported as a single score where if the organism is predicted to be a phagocyte generalist, or is not predicted to be a phagocyte in any of the three phagocytosis models, only the phagocyte-generalist prediction is shown. If an organism is predicted to be phagocytotic in the entamoebid model, but not the phagocyte generalist model, that organism has a (*) next to the species name, and the bar is colored a lighter brown. If an organism is predicted to be phagocytotic in the *R. allomycis* model, but not the phagocyte generalist or entamoebid model, that organism has a (**) next to the species name, and the bar is tan colored.



Supplementary Figure 2. Lowest common ancestor analysis of prokaryote homologs of phagosome associated and phagocyte enriched proteins.

Supplementary Figure 3. DNAse II is found in two distinct bacterial clades. Bacterial clades are highlighted in pink. Red circles represent support values for each node. "Bacterial Clade A" contains the previously identified DNase II sequence from *Burkholderia pseudomallei*, plus four other bacterial species. "Bacterial Clade B" contains seven additional bacterial species in a well-supported, distinct, clade.



Supplementary Figure 4. 1000 times repeated, 10-fold cross-validation indicates the accuracy of each model developed for this analysis. For each model, 1000 times repeated, 10-fold cross validation was performed on the training set with biological groupings, and the training set with shuffled groupings. With biological grouping, every model was > 95% accurate in cross-validation tests. When the group labels were randomized, the models were no longer predictive (as expected).

Supplementary Tables

Species	*Group*	*Photosynthesis*	*Phagocytosis*	*Prototrophy*
Acanthamoeba castellani	Amoebozoa	-	✓	?
Acytostelium subglobosum	Amoebozoa	-	✓	-
Allomyces macrogynus	Fungi-Blastocladiomycota	-	-	✓
Arabidopsis thaliana	Chloroplastida	✓	-	✓
Batrachochytrium dendrobatidis	Fungi-Chytridiomycota	-	-	✓
Bigelliowiella natans	Rhizaria	✓	✓	✓
Bodo saltans	Euglenozoa	-	✓	-
Brachypodium distachyon	Chloroplastida	✓	-	✓
Chlamydomonas reinhardtii	Chloroplastida	✓	-	✓
Chlorella variabilis	Chloroplastida	✓	-	✓
Chrysochromulina sp.	Haptophyta	✓	✓	✓
Conidiobolus coronatus	Fungi-Zygomycota	-	-	✓
Cyanidioschyzon merolae	Rhodophyta	✓	-	✓
Cymbomonas tetramitiformis	Chloroplastida	✓	✓	✓
Dictyostelium discoideum	Amoebozoa	-	✓	-
Drosophila melanogaster	Insecta	-	✓	-
Entamoeba histolytica	Amoebozoa	-	✓	-
*Fonticula alba**	Holomycota			
Giardia intestinalis[†]	Excavata	-	-	-
Mimulus guttatus	Chloroplastida	✓		✓
Mus musculus	Mammalia	-	✓	-
Neocallimastix californiae	Fungi-Neocallimastigomycota	-	-	✓
Oryza sativa	Chloroplastida	✓	-	✓
Paramecium tetraurelia	Alveolata	-	✓	-
Physcomitrella patens	Chloroplastida	✓	-	✓
Picea abies	Chloroplastida	✓	-	✓
Puccinia sorghi	Fungi-Basidiomycota	-	-	✓
*Reticulomyxa filosa**	Rhizaria	-	✓	?
Rhizophagus irregularis	Fungi-Glomeromycota	-	-	✓
Rozella allomycis	Rozella	-	✓	?
Saccharomyces cerevisiae	Fungi-Ascomycota	-	-	✓
Schizosaccharomyces pombe	Fungi-Ascomycota	-	-	✓
Selaginella moellendorffii	Chloroplastida	✓	-	-
Tetrahymena thermophila	Alveolata	-	✓	-
Thecamonas trahens	Apusozoa	-	✓	-
Trichomonas vaginalis	Excavata	-	✓	-
Volvox carteri	Chloroplastida	✓	-	✓

Supplementary Table 1. Species used in all versus all BLAST and model training for phagocytosis, photosynthesis, and prototrophy. A checkmark indicates a species has that character, a (-) indicates it does not, a (?) indicates we are uncertain at the time of model building. *Used during model training, but not initial all vs. all BLAST. [†]Used in all vs. all blast, but not during model training.

Organism	Phagocyte-generalist prediction	Phagocyte-entamoebid prediction	Phagocyte-rozellid prediction	Prototrophy prediction	Photosynthesis prediction	Trophic mode
E. histolytica	0.15	1.00	0.29	0.00	0.00	phagocytotic, auxotrophic heterotroph
mantamonad sp.	1.00	0.98	0.97	0.03	0.00	phagocytotic, auxotrophic heterotroph
Ph. tricornutum	0.00	0.00	0.16	0.96	0.98	non-phagocytotic phototroph
Pl. falciparum	0.00	0.00	0.02	0.00	0.00	non-phagocytotic obligate parasite
Pr. parvum[†]	1.00	0.99	0.98	0.99	0.99	phagocytotic-phototroph
R. allomycis	0.01	0.03	1.00	0.03	0.00	phagocytotic, auxotrophic heterotroph
T. vaginalis	0.61	0.96	0.48	0.00	0.00	phagocytotic, auxotrophic heterotroph

Supplementary Table 2. Predictions and trophic mode assignment of select test genomes. Green shading indicates a positive prediction. [†]Data was derived from a transcriptome.

Supplementary Table 3. Analysis of a selection of eukaryotic signature proteins. (See excel file)

Supplementary Table 4. Full predictions table for phylum level pan-prokaryote assemblages. (See excel file)

GO.ID	TERM	ANNOTATED	SIGNIFICANT	P-VALUE
GO:0046872	metal ion binding	158	137	9.4×10^{-5}
GO:0005509	calcium ion binding	34	31	0.0028
GO:0030246	carbohydrate binding	18	16	0.047
GO:0020037	heme binding	8	8	0.058
GO:0003779	actin binding	20	16	0.058
GO:0005102	receptor binding	22	20	0.11
GO:0008201	heparin binding	6	6	0.12
GO:0004252	serine-type endopeptidase activity	6	6	0.12
GO:0030145	manganese ion binding	6	6	0.12
GO:0042803	protein homodimerization activity	19	16	0.13
GO:0009055	electron carrier activity	6	6	0.17
GO:0005507	copper ion binding	6	6	0.17
GO:0046933	proton-transporting ATP synthase activity	5	5	0.17
GO:0016787	hydrolase activity	154	118	0.22
GO:0031418	L-ascorbic acid binding	4	4	0.24

Supplementary Table 5. Molecular function enrichment of 567 proteins from the phagosome (predicted and observed) with LCA affinity in the bacteria. The background set for GO enrichment analysis was the 775 HMMs from the predicted and observed phagosome that have putative homologs in bacteria and/or archaea. For the proteins associated with each category, see Extended Data Table 16.

GO.ID	TERM	ANNOTATED	SIGNIFICANT	P-VALUE
GO:0003735	structural constituent of ribosome	48	43	4.4×10^{-24}
GO:0019843	rRNA binding	12	10	8.3×10^{-6}
GO:0003924	GTPase activity	12	10	1.9×10^{-5}
GO:0005525	GTP binding	21	13	1.6×10^{-4}
GO:0003729	mRNA binding	21	11	0.0036
GO:0003723	RNA binding	86	43	0.031
GO:0046961	proton-transporting ATPase activity , rotational mechanism	4	3	0.044
GO:0004298	threonine-type endopeptidase activity	2	2	0.056
GO:0004579	dolichyl-diphosphooligosaccharide -protein glycotransferase activity	2	2	0.056
GO:0005047	signal recognition particle binding	2	2	0.056
GO:0046983	protein dimerization activity	24	6	0.057
GO:0030515	snoRNA binding	4	3	0.14
GO:0003743	translation initiation factor activity	3	2	0.14
GO:0048306	calcium-dependent protein binding	3	2	0.14
GO:0004867	serine-type endopeptidase inhibitor activity	3	2	0.14

Supplementary Table 6. Molecular function enrichment of 181 proteins from the phagosome (predicted and observed) with LCA affinity in the archaea. The background set for GO enrichment analysis was the 775 HMMs from the predicted and observed phagosome that have putative homologs in bacteria and/or archaea. For the proteins associated with each category, see Extended Data Table 17.

GO.ID	TERM	ANNOTATED	SIGNIFICANT	P-VALUE
GO:0004871	signal transducer activity	31	20	0.0010
GO:0008017	microtubule binding	17	12	0.0014
GO:0005096	GTPase activator activity	15	11	0.0014
GO:0001078	transcriptional repressor activity	5	5	0.0037
GO:0032266	phosphatidylinositol-3-phosphate binding	5	5	0.0037
GO:0000978	RNA polymerase II core promoter proximal promoter sequence-specific DNA binding	5	5	0.0037
GO:0003779	actin binding	51	31	0.0040
GO:0005085	guanyl-nucleotide exchange factor activity	17	15	0.0055
GO:0051015	actin filament binding	28	16	0.0061
GO:0035091	phosphatidylinositol binding	23	17	0.0067
GO:0017112	Rab guanyl-nucleotide exchange factor activity	4	4	0.011
GO:0003774	motor activity	8	7	0.016
GO:0003700	transcription factor activity	18	14	0.016
GO:0005546	phosphatidylinositol-4,5-bisphosphate binding	6	5	0.016

Supplementary Table 7. Molecular function enrichment of 546 proteins from the phagosome (predicted and observed) in eukaryotes, with no prokaryote homologs. The background set for GO enrichment analysis was all 1,321 HMMs from the predicted and observed phagosome. For the proteins associated with each category, see Extended Data Table 18.