


RESEARCH ARTICLE

Open Access



# Genomics and transcriptomics yields a system-level view of the biology of the pathogen *Naegleria fowleri*

Emily K. Herman<sup>1,2\*†</sup>, Alex Greninger<sup>3,4†</sup>, Mark van der Giezen<sup>5</sup>, Michael L. Ginger<sup>6</sup>, Inmaculada Ramirez-Macias<sup>1,7</sup>, Haylea C. Miller<sup>8,9</sup>, Matthew J. Morgan<sup>10</sup>, Anastasios D. Tsousis<sup>11</sup>, Katrina Velle<sup>12</sup>, Romana Vargová<sup>13</sup>, Kristína Záhonová<sup>1,14,15</sup>, Sebastian Rodrigo Najle<sup>16,17</sup>, Georgina MacIntyre<sup>18</sup>, Norbert Muller<sup>19</sup>, Mattias Wittwer<sup>20</sup>, Denise C. Zysset-Burri<sup>21</sup>, Marek Eliáš<sup>13</sup>, Claudio H. Slamovits<sup>22</sup>, Matthew T. Weirauch<sup>23,24</sup>, Lillian Fritz-Laylin<sup>12</sup>, Francine Marciano-Cabral<sup>25</sup>, Geoffrey J. Puzon<sup>8</sup>, Tom Walsh<sup>10</sup>, Charles Chiu<sup>3</sup> and Joel B. Dacks<sup>1,15,26\*</sup> 

## Abstract

**Background:** The opportunistic pathogen *Naegleria fowleri* establishes infection in the human brain, killing almost invariably within 2 weeks. The amoeba performs piece-meal ingestion, or trophocytosis, of brain material causing direct tissue damage and massive inflammation. The cellular basis distinguishing *N. fowleri* from other *Naegleria* species, which are all non-pathogenic, is not known. Yet, with the geographic range of *N. fowleri* advancing, potentially due to climate change, understanding how this pathogen invades and kills is both important and timely.

**Results:** Here, we report an -omics approach to understanding *N. fowleri* biology and infection at the system level. We sequenced two new strains of *N. fowleri* and performed a transcriptomic analysis of low- versus high-pathogenicity *N. fowleri* cultured in a mouse infection model. Comparative analysis provides an in-depth assessment of encoded protein complement between strains, finding high conservation. Molecular evolutionary analyses of multiple diverse cellular systems demonstrate that the *N. fowleri* genome encodes a similarly complete cellular repertoire to that found in free-living *N. gruberi*. From transcriptomics, neither stress responses nor traits conferred from lateral gene transfer are suggested as critical for pathogenicity. By contrast, cellular systems such as proteases, lysosomal machinery, and motility, together with metabolic reprogramming and novel *N. fowleri* proteins, are all implicated in facilitating pathogenicity within the host. Upregulation in mouse-passaged *N. fowleri* of genes associated with glutamate metabolism and ammonia transport suggests adaptation to available carbon sources in the central nervous system.

**Conclusions:** In-depth analysis of *Naegleria* genomes and transcriptomes provides a model of cellular systems involved in opportunistic pathogenicity, uncovering new angles to understanding the biology of a rare but highly fatal pathogen.

**Keywords:** Illumina, RNA-Seq, Genome sequence, Protease, Cytoskeleton, Metabolism, Lysosomal, Inter-strain diversity, Neuropathogenic

\* Correspondence: [eherman@ualberta.ca](mailto:eherman@ualberta.ca); [dacks@ualberta.ca](mailto:dacks@ualberta.ca)

†Emily K. Herman and Alex Greninger contributed equally to this work.

<sup>1</sup>Division of Infectious Disease, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada

Full list of author information is available at the end of the article



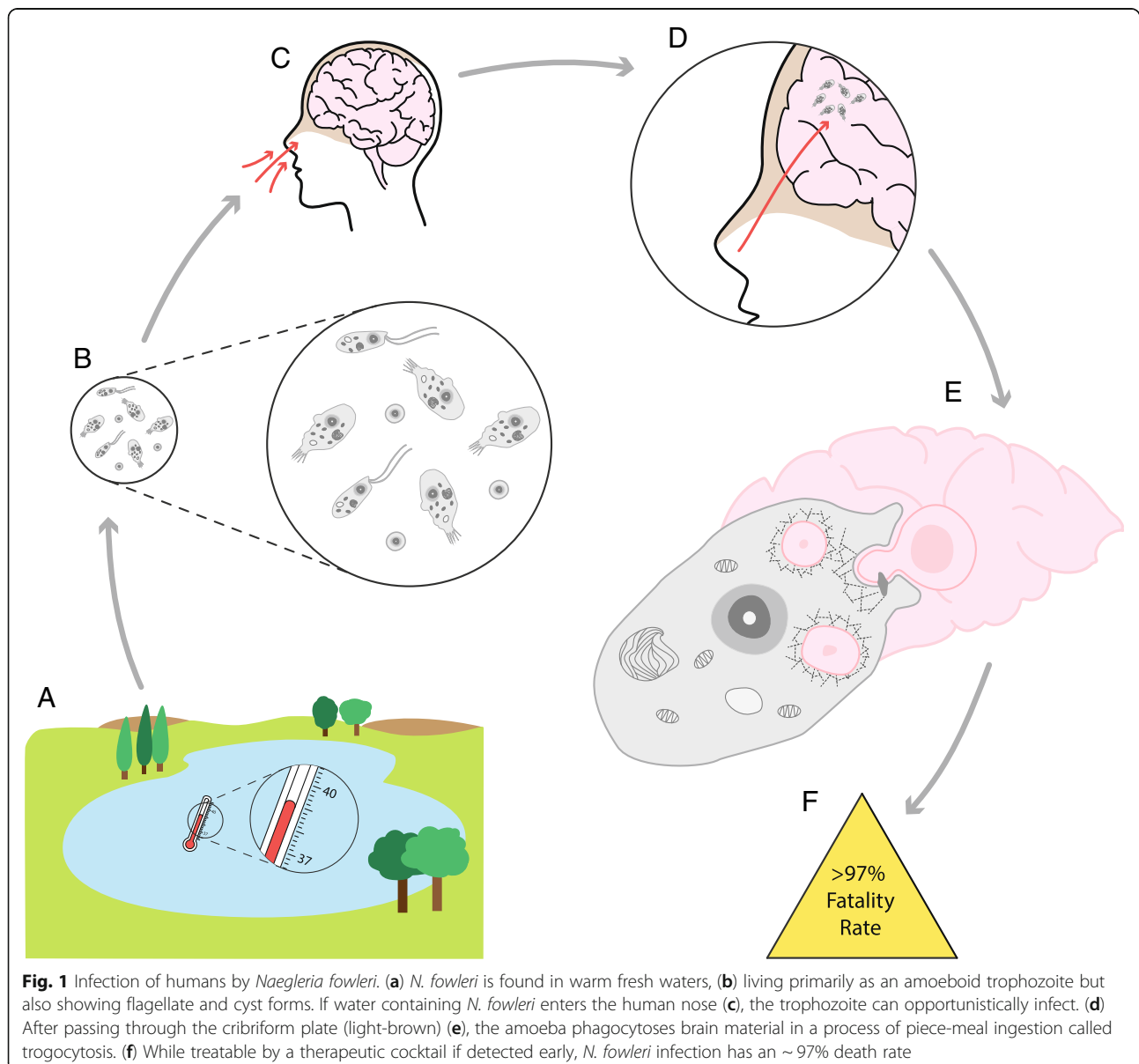
© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

*Naegleria fowleri* is an opportunistic pathogen of humans and animals (Fig. 1), causing primary amoebic meningoencephalitis, and killing up to 97% of those infected, usually within 2 weeks [1]. It is found in warm freshwaters around the world and drinking water distribution systems [2–4] with *N. fowleri*-colonized drinking water distribution systems linked to deaths in Pakistan [5–7], Australia [8], and the USA [9, 10]. Human infection occurs when contaminated water enters the nose (Fig. 1). Opportunistically, *N. fowleri* passes through the cribriform plate to the olfactory bulb in the brain and performs trophocytosis (i.e., piece-meal ingestion) of brain material (Fig. 1), causing physical damage and massive inflammation leading to death. Although successful

treatment with miltefosine and other antimicrobials is becoming more common [11, 12], this relies on appropriate early diagnosis, which is challenging due to the comparatively low incidence of amoebic versus viral or bacterial meningitis.

The known incidence of infection is relatively low, with 381 human cases reported in the literature [13]. However, infection is likely more prevalent. It was recently modelled that there are ~16 undetected cases annually in the US alone [14] and this situation could well be more pronounced in developing countries with warm climates and inconsistent medical reporting. *N. fowleri* has been recently proposed as an emerging pathogen, based on increased case reports in the past decade [15] and an increase in its northward expansion in the USA



[16]. With cases from temperate locations reported in recent years [17, 18], the threat of *N. fowleri* may be exacerbated by range expansion due to climate change [15], which has been associated with a rise in freshwater temperatures, an increase in aquatic recreational activities [19], and extreme weather events. It is likely that we are still not fully aware of the global scale of *N. fowleri* infection [15], or how infection may increase as climate change accelerates.

Of the ~47 species of *Naegleria*, found commonly in soils and fresh waters worldwide, *N. fowleri* is the only species that infects humans, suggesting that pathogenicity is a gained function [20]. Several labs have identified potential *N. fowleri* pathogenicity factors including proteases, lipases, and pore-forming proteins [21–26]. The hypothesized mechanism of *N. fowleri* pathogenicity—tissue degradation for both motility during infection and phagocytosis—is compatible with these factors. However, none appear to be unique to *N. fowleri*, and there are likely proteins yet to be identified that are responsible for pathogenesis. In 2010, the genome of the non-pathogenic *N. gruberi* was published [27], followed by draft genomes of *N. fowleri* [28, 29]. This has set the stage for a thorough in-depth comparative genome-wide perspective on *N. fowleri* diversity and pathogenesis, which is currently lacking.

Here, we report the genome sequences of two *N. fowleri* strains; 986, an environmental isolate from an operational drinking water distribution systems in Western Australia, and CDC:V212, a strain isolated from a patient. We also report a transcriptomic analysis of induced pathogenicity in a third strain of *N. fowleri* (LEE) to identify the genes differentially expressed as a consequence of infection. Guided by these data, our careful curated comparative analysis of cellular machinery provides a comprehensive view of the pathogen *N. fowleri* at a cellular systems level.

## Results

### Two new *N. fowleri* genomes

In order to better understand the cellular basis for *N. fowleri* pathogenesis, we took a combined genomic, transcriptomic, and molecular evolutionary approach. The genomes of *N. fowleri* strains V212 and 986 were sequenced to average coverage of 251X and 250X respectively and the transcriptome of axenically cultured *N. fowleri* V212 was also sequenced to support gene prediction. Assembly statistics for our new genomes, as well as of the *N. fowleri* strain 30863, are shown in Table 1. Comparison between the three strains (Table 2) showed relative conservation of genome statistics. However,

these were remarkably different from those values for *N. gruberi* (Table 2) [27]. The differences in genome statistics are not unreasonable given that genetic diversity within the *Naegleria* clade has been equated to that of tetrapods, at ~95% identity in the 18S rRNA gene ([30], Additional File 1-Table S1). Importantly, there is transcriptomic evidence for 82% of genes in *N. fowleri* V212, suggesting that they are expressed when grown in culture and 100% of the genes with transcriptome evidence were also found in predicted genomic set, meaning that we did not detect any transcribed genes that were not predicted from the genome. Out of 303 near-universal single-copy eukaryotic orthologs, 277 were found in the set of *N. fowleri* V212 predicted proteins, giving a BUSCO score of 88.3%, signifying that the genome and predicted proteome are highly complete.

### *Naegleria fowleri* encodes a complete and canonical cellular complement

In 2010, *N. gruberi* was hailed as possessing extensive eukaryotic cytoskeletal, membrane trafficking, signaling, and metabolic machinery, suggesting sophisticated cell biology for a free-living protist that diverged from other eukaryotic lineages over one billion years ago [27, 31]. Careful manual curation of gene models and genomic analysis, as detailed in the “Methods” and Supplementary Material, of meiotic machinery (Additional File 2-Table S2, Additional File 3-Figure S1), transcription factors (Additional File 4-Supplementary Material 1, Additional File 5-Table S3, [32–35]), sterols (Additional Material 3-Figures S2 and S3, Additional File 4-Supplementary Material 2, Additional File 6-Table S4, [36–46]), mitochondrial proteins (Additional File 7-Table S5), cytoskeletal proteins (Additional File 3-Figure S4, Additional File 4-Supplementary Material 3, Additional File 8-Table S6, [21, 24, 27, 28, 47–64]), membrane trafficking components (Additional File 9-Table S7), and small GTPases (Additional File 3-Figures S5, S6, and S7, Additional File 4-Supplementary Material 4, Additional File 10-Table S8, [27, 65–79]) demonstrates that, like *N. gruberi*, *N. fowleri* possesses a remarkably complete repertoire of cellular machinery.

### Comparative genomics identifies hundreds of genes unique to *N. fowleri*

Pathogenesis is likely a gain-of-function. As *N. fowleri* is the only human pathogenic *Naegleria*, we specifically looked for differences with the non-pathogenic *N. gruberi*. Using OrthoMCL, the proteins from the three *N. fowleri* strains and *N. gruberi* were clustered into 11,399 orthogroups, of which 7656 (67%) appear to be shared by all four *Naegleria* species, and 10,451 (92%) are shared by all three *N. fowleri* strains (Fig. 2a). There are

**Table 1** Assembly statistics for *N. fowleri* strains V212, 986, and ATCC 30863

	<i>N. fowleri</i> V212	<i>N. fowleri</i> 986	<i>N. fowleri</i> 30863
Number of scaffolds	1859	990	1124
Total size of scaffolds	27,711,821	27,495,188	29,619,856
Longest scaffold	387,133	390,775	471,424
Mean scaffold size	14,907	2354	26,352
N50	92,316	101,682	136,406
L50	86	83	63
Number of contigs	1962	1919	2530
Total size of contigs	27,703,916	27,397,881	28,636,847
Longest contig	372,317	272,583	236,403
Mean contig size	14,120	14,277	11,319
N50	86,051	45,674	38,800
L50	93	182	213
BUSCO complete	88.3%	87.9%	87.8%
BUSCO single copy	76.9%	76.5%	73.7%
BUSCO duplicate	11.4%	11.4%	14.1%
BUSCO fragmented	2.4%	3.1%	2.7%
BUSCO missing	9.3%	9.0%	9.5%

2795 groups not identified in *N. gruberi* that are shared by all three *N. fowleri* strains. This number can be further split into orthogroups where BLAST searches against either the *N. gruberi* predicted proteome or the genome retrieve a potential homolog (these may be paralogs, or false negatives from OrthoMCL analysis), and groups that have no hit, and therefore no homolog, in *N. gruberi*. There are 458 orthogroups in *N. fowleri* whose members retrieve no *N. gruberi* homologs (Additional File 11-Table S9). Of these, 80% are unique to *N. fowleri*, with no clearly homologous sequence in any other organisms based on NCBI BLAST, and only 52 (11%) identified either could be functionally annotated based on NR BLAST results or contain a characterized domain. In total, 404 of the 458 genes have transcriptomic evidence (FPKM > 5) in either the axenic or mouse-passaged sample groups,

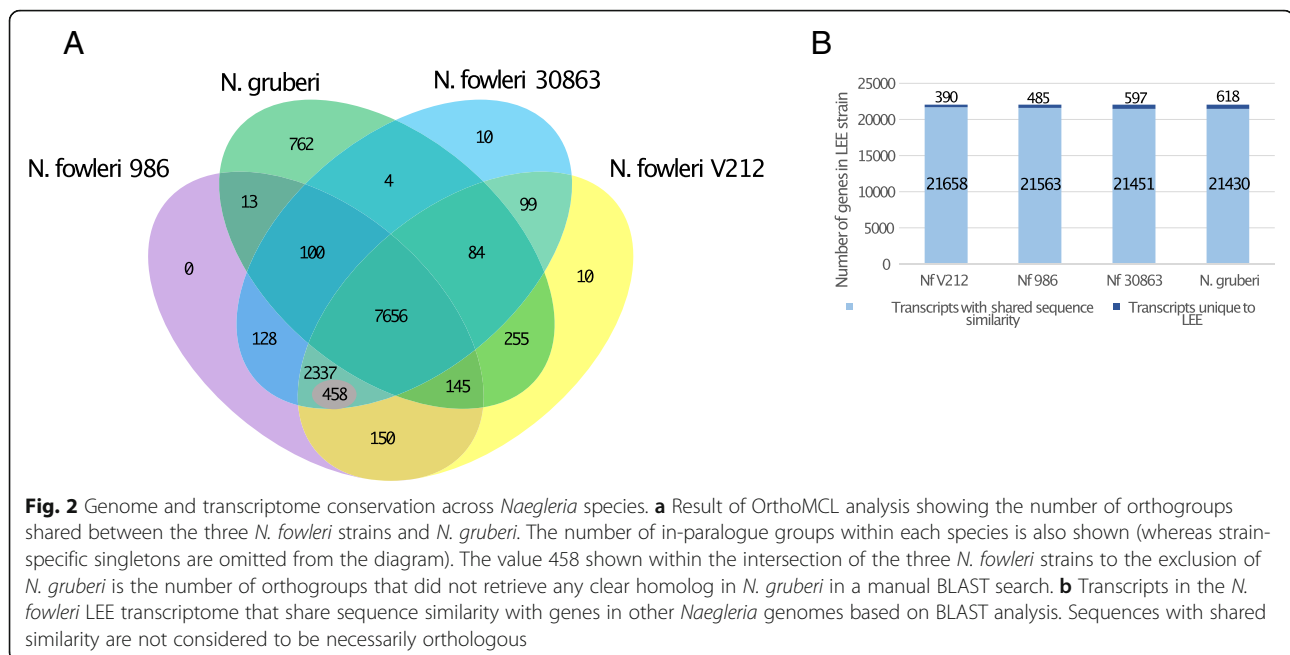
suggesting that most of the gene models are accurate and these genes are expressed.

#### Transcriptomics identifies differentially expressed genes in an animal model of *N. fowleri* pathogenesis

To complement the comparative genomics approach, we also performed transcriptomics, taking advantage of previously established experimentally induced pathogenicity in the *N. fowleri* LEE strain [80]. In this system, not only does the strain that has been continuously passaged in mice have a lower LD<sub>50</sub> in guinea pigs by two orders of magnitude, it is also resistant to complement-mediated killing, while cultured *N. fowleri* LEE and *N. gruberi* are not [80]. De novo assembly of the *N. fowleri* LEE transcriptome shows that there are relatively few gene transcripts that do

**Table 2** Genome statistics for *N. fowleri* strains V212, 986, 30863, and *N. gruberi* strain NEG-M

	<i>N. fowleri</i> V212	<i>N. fowleri</i> 986	<i>N. fowleri</i> ATCC 30863	<i>N. gruberi</i> NEG-M
Total genome size	27.7 Mbp	27.5 Mbp	29.62 Mbp	41.0 Mbp
GC content	36%	36%	35%	33%
Number of genes	12,677	11,599	11,499	15,708
Average gene length	1785 bp	1955 bp	1984 bp	1677 bp
Exons/gene	2	2	2	1.7
Average exon length	777 bp	849 bp	825 bp	894 bp
% coding	71.35%	73.01%	70.79%	57.80%
Average intron length	126 bp	138 bp	144 bp	203 bp



not share any similar sequence with other *N. fowleri* strains (Fig. 2b).

We identified 315 differentially expressed genes in mouse-passaged *N. fowleri* LEE (LEE-MP) compared to *N. fowleri* LEE grown in culture (LEE-Ax) (Additional File 12-Table S10). Of these, 206 are upregulated in mouse-passaged *N. fowleri*, while 109 are downregulated (Additional File 3- Figure S8). In terms of function, these 206 genes span multiple cellular systems with potential links to pathogenesis. Overall analysis of downregulated genes was less informative than that of upregulated genes. Systems represented in the downregulated gene dataset include signal transduction, flagellar motility, genes found to be involved in anaerobiosis in bacteria, and transcription/translation (Additional File 12-Table S10). Nonetheless, approximately 70% of the downregulated genes not in these categories are genes of unknown function.

Building on the manual curation of the encoded cellular machinery and using the comparative genomic and transcriptomic data, we confirmed and extended our knowledge of previously identified categories of pathogenicity factors and identified several novel major aspects of *N. fowleri* biology with implications for pathogenesis and that represent avenues for future investigation.

#### Proteases and endolysosomal proteins are a substantial system in *N. fowleri*

Similar to many other animal parasites, *N. fowleri* is known to secrete proteases that traverse the extracellular matrix during infection, and pore-forming proteins that

kill host cells [21–25]. Prosaposin is the most prominent pore-forming protein (termed Naegleriapore A) and is heavily glycosylated and protease-resistant [22]. The precursor proteins for Naegleriapore A and B were indeed found in our upregulated DE gene set. Similarly, the cathepsin protease (Cathepsin A, or Nf314) is upregulated in mouse-passaged *N. fowleri*, consistent with its previously proposed role in pathogenicity [81].

Comparative genomics found broadly similar complements of proteases between all *N. fowleri* strains and *N. gruberi* (Additional File 3-Figure S9, Additional File 13-Table S11), with one exception. The serine protease S81 was found in all three *N. fowleri* genomes, but not in *N. gruberi*, and is highly (but not differentially) expressed under both axenic and mouse-passaged conditions with FPKM values ranging from 500 to 800. S81 protein has invertebrate-type lysozyme (ilys) and peptidoglycan-binding domains. Sequence-similarity searches against the NCBI non-redundant and EukProt databases to look for homologs identified only two candidates, both from foraminiferans (Additional File 14-Table S12). Subsequent searches using one of these foraminiferan sequences identified further hits in both databases. However, the region of putative homology between S81 and any retrieved sequences was restricted to the ilys domain, with no similarity outside of the region. Ilys domains are found widely across eukaryotes (Additional File 14-Table S12). Therefore, while S81 likely arose from a common ilys of unknown origin, it does not have clear orthologs in other organisms and appears to be *N. fowleri*-specific.

Twenty-eight proteases are upregulated in mouse-passaged *N. fowleri*, making up more than 10% of all upregulated genes (Additional File 12-Table S10). Of the protease families with *N. fowleri* homologs upregulated following mouse passage, half are either localized to lysosomes or are secreted. The most substantially represented types of lysosomal/secreted protease in the upregulated genes are the cathepsin proteases; specifically, the C01 subfamily, with 10 out of 21 genes upregulated in mouse-passaged *N. fowleri*. The C01 subfamily of proteases includes cathepsins B, C, L, Z, and F. Each of these subfamilies has multiple members, up to 10 in the case of Cathepsin B, and members of the B, Z, and F subfamilies are upregulated.

Despite the large number of C01 subfamily cathepsin proteases in *N. fowleri* (20-21 members), *N. gruberi* encodes even more (35). Many of the *N. fowleri* and *N. gruberi* cathepsins have 1:1 orthology (Additional File 3-Figure S10), with at least three expansions that have occurred in the Cathepsin B clade in *N. gruberi*. These expansions account for most of the difference in paralog number between the two species, making up 12 of ~16 *N. gruberi*-specific C01 homologs. Notably, of the *N. fowleri* cathepsin genes that are upregulated in mouse-passaged *N. fowleri*, at least two (i.e., CatB7-*Nfowleri*V212\_g899.t2, CatB8-*Nfowleri*V212\_g10536.t1) lack orthologs in *N. gruberi*, raising the possibility of their specific involvement in pathogenesis.

Protease secretion is underpinned by the membrane trafficking system (the canonical secretion pathway) and autophagy-based unconventional secretion (for proteins lacking N-terminal signal peptides). In both cases, there are few differences in gene presence, absence, and paralog number in the different *Naegleria* genomes (Additional File 9-Table S7 and Additional File 15-Table S13). Strikingly, however, 42% of the upregulated genes are involved in lysosomal processes. In addition to the 22 proteases above, a lysosomal rRNA degradation gene is upregulated, as well as three subunits of the vacuolar ATPase proton pump (16, 21, and 116 kDa) responsible for acidification of both lysosomes and secretory vesicles. Endolysosomal trafficking genes Rab GTPase Rab32 (one of three paralogs) and the retromer component Vps35 were also upregulated.

### Proteins driving actin cytoskeletal rearrangements

While actin is known to drive many cellular processes in eukaryotes, Nf-Actin has been implicated in pathogenicity in *N. fowleri* due to its role in trophocytosis via food cup formation [56]. Furthermore, actin-binding proteins and upstream regulators of actin polymerization were reported to correlate with virulence [28]. While the complements of actin-associated machinery are similar between the *Naegleria* species, we notably identified a

PTEN domain on one of the formin homologs in *N. fowleri* that was not identified in *N. gruberi* (Additional File 4-Supplementary Material 3). Since humans also do not encode formins of the PTEN family, if these PTEN formins are responsible for any vital *Naegleria* processes, they may represent useful drug targets.

Although *Naegleria* actin protein levels do not always correlate with transcript levels [27, 82], a single subunit of the Arp2/3 complex (Arp3) and the WASH complex member strumpellin were both upregulated in the mouse-passaged amoebae (Additional File 4-Supplementary Material 3, Additional File 8-Table S6, Additional File 12-Table S10). Similarly, we identified an upregulated RhoGAP22 gene and the serine/threonine protein kinase PAK3, which are involved in Rac1-induced cell migration in other species as well as an upregulated member of the gelsolin superfamily in the mouse-passaged *N. fowleri*, which may contribute to actin nucleation, capping, or depolymerization [83]. Although the shift from the environment in the mouse brain to tissue culture conditions prior to sequencing may have resulted in an upregulation in macropinocytosis, which can alter cell motility and the transcription of cytoskeletal genes in *Dictyostelium discoideum* [84, 85], the overall modulation of cytoskeletal protein encoding genes highlights the potential importance of cytoskeletal dynamics in promoting virulence.

### Neither LGT nor cell stress have substantially shaped *N. fowleri*'s unique biology

One obvious potential source of pathogenicity factors is lateral gene transfer of bacterial genes into *N. fowleri* to the exclusion of *N. gruberi* and other eukaryotes. However, of the 458 genes exclusive to *N. fowleri*, most are of unknown function (Additional File 11-Table S9) and of these, only 26 have Bacteria, Archaea, or viruses as the largest taxonomic group containing the top five BLAST hits. Furthermore, only one of the genes upregulated in mouse-passaged *N. fowleri* lacks a homolog in *N. gruberi*, and it has a potential homolog in *D. discoideum* and members of the Burkholderiales clade of bacteria (*Nfowleri*V212\_g4665, Additional File 12-Table S10).

Another potential reason for *N. fowleri*'s ability to infect humans and animals is the ability to survive the stresses of infection. However, we observed no obvious differences between the *N. fowleri* and *N. gruberi* complements of the ER-associated degradation machinery and unfolded protein response machinery that would suggest a differential ability to cope with cell stress (Additional File 16-Table S14). Furthermore, our transcriptomics analysis showed a general downregulation of cell stress systems, as well as DNA damage repair (Additional File 12-Table S10). This does not suggest that

these systems are significantly involved in pathogenesis and our transcriptomics experiment reflects an organism that is not under duress.

#### **Beyond Nfa1, adhesion factors remain mysterious in *N. fowleri***

Since infection requires the ability to attach to cells of the nasal epithelium, differences between *N. fowleri* and *N. gruberi* in cell-cell adhesion factors may be relevant to pathogenesis. While we were unable to find evidence of a previously reported integrin-like protein [86] in any of the genomic data (based on sequence similarity to human integrin), we did find that another previously identified attachment protein, Nfa1, was highly expressed in both mouse-passaged and axenically cultured *N. fowleri* LEE (FPKM > 1000), providing further evidence supporting its previously reported role in pathogenicity [87]. It is likely that the integrin-like protein identified by Jamerson and colleagues is unrelated to animal integrins, although it appears to be recognized by antibodies to human  $\beta 1$  integrin.

*N. fowleri* and *N. gruberi* encode relatively few putative adhesion G protein-coupled receptors (AGPCRs), with 10 or fewer in each organism (Additional File 17-Table S15, Additional File 18-Table S16). These were identified by searching for proteins with the appropriate domain organization: sequences that have both an extracellular domain (assuming correctly predicted topology in the membrane) and seven transmembrane regions. A full collection of repeat-containing proteins in *N. fowleri* V212 and *N. gruberi* is shown in Additional File 17-Table S15. Of the proteins involved in adhesion in *D. discoideum* (TM9/Phg1, SadA, SibA, and SibC), only orthologs of the TM9 protein could be reliably identified (Additional File 19-Table S17), which was not found to be differentially regulated in our transcriptomic dataset. It is possible that the *Naegleria* homolog may be involved in cell adhesion, but with other downstream effectors.

#### ***N. fowleri* shows modulation and unusual metabolic pathways**

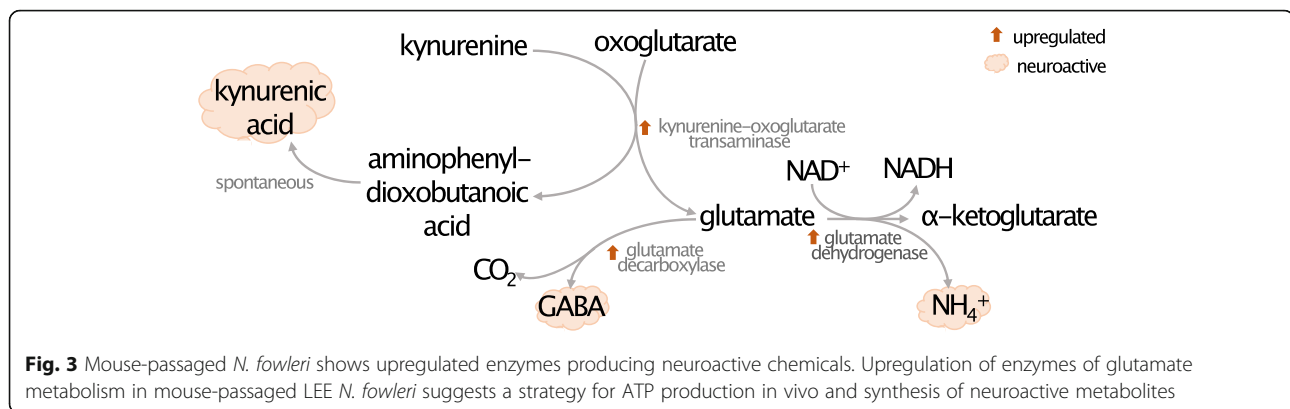
Strikingly, 19% of upregulated genes are involved in metabolism (Additional File 12-Table S10). Both catabolic and anabolic processes are represented; some upregulated genes include phospholipase B-like genes necessary for beta-oxidation, and genes involved in phosphatidate/phosphatidylethanolamine, fatty acid (including long-chain fatty acid elongation), and isoprenoid biosynthesis. Phospholipase B was previously identified as pathogenicity factors in *N. fowleri* [26]. Also identified in our study was a Rieske cholesterol C7(8)-desaturase (Additional File 3-Figure S3, Additional File 4-Supplementary Material 2), a protein involved in sterol production that is absent in mammals, thus potentially representing a drug target.

Recent work shows that *N. gruberi* trophozoites prefer to oxidize fatty acids to generate acetyl-CoA, rather than use glucose and amino acids as growth substrates [88]. Several genes involved in metabolism of both lipids and carbohydrates are upregulated in mouse-passaged *N. fowleri*. Of interest are those that may be involved in metabolizing the polyunsaturated long-chain fatty acids that are abundant in the brain, such as long-chain fatty acyl-CoA synthetase and a delta 6 fatty acid (linoleoyl-CoA) desaturase-like protein. Consistent with possible shifting carbon source usage or increased growth rates, mitochondrial and energy conversion genes are upregulated, such as ubiquinone biosynthesis genes, isocitrate dehydrogenase (TCA cycle), complex I and complex III genes (oxidative phosphorylation), and a mitochondrial ADP/ATP translocase. Eight genes involved in amino acid metabolism are also upregulated.

Intriguingly, we identified several areas of *N. fowleri* metabolism that may impact the human host. Glutamate is found in high millimolar concentrations in the brain and is thought to be the major excitatory neurotransmitter in the central nervous system [89, 90]. Several genes that function in glutamate metabolism are upregulated in mouse-passaged *N. fowleri*, including kynurenine-oxoglutarate transaminase, glutamate decarboxylase, glutamate dehydrogenase, and isocitrate dehydrogenase (Fig. 3, Additional File 12-Table S10). Multiple neurotropic compounds are generated via these enzymes, such as kynurenic acid, GABA, and  $\text{NH}_4^+$ . Kynurenic acid in particular has been linked to neuropathological conditions in tick-borne encephalitis [91].

Ammonia transporters are also upregulated (Additional File 12-Table S10), which may be one way to remove toxic ammonia from the cell, as *N. fowleri*, like *N. gruberi*, has an incomplete urea cycle [92]. Both glutamate and polyamine metabolism pathways discussed above generate  $\text{NH}_4^+$  as a by-product, and it is possible that this is secreted into the host brain and leads to pathological effects.

Finally, agmatine deiminase, which is involved in putrescine biosynthesis, is also upregulated (Additional File 12-Table S10). This enzyme catalyzes the conversion of agmatine to carbamoyl putrescine, an upstream precursor to the polyamines glutathione and trypanothione [93]. Trypanothione provides a major defense against oxidative stress, some heavy metals, and potentially xenobiotics in trypanosomatid organisms (e.g., *Leishmania*, *Trypanosoma*) [94] and has been isolated from *N. fowleri* trophozoites [95]. Although a similar protective role for trypanothione has yet to be confirmed in *N. fowleri*, it is a critical component in trypanosomatid parasites [96, 97], and enzymes in this pathway may represent novel drug targets.



### Upregulated genes of miscellaneous or unknown function

There are many other genes that are upregulated, but do not fall into one of the categories outlined above (Additional File 12-Table S10). Notably, one of these is a transcription factor of the RWP-RK family, which has previously only been identified in plants and *D. discoideum*, functioning in plants to regulate responses in nitrogen availability, including differentiation and gametogenesis. While its role in amoebae is unclear, it may represent a potential drug target, as RWP-RK transcription factors are not present in human cells. Notably, of the 208 genes upregulated in highly pathogenic *N. fowleri*, 49 genes found either in *N. fowleri* alone or in *N. fowleri* and *N. gruberi* (Additional File 12-Table S10). These represent unique potential targets against which anti-*Naegleria* therapeutics may be developed.

### Discussion

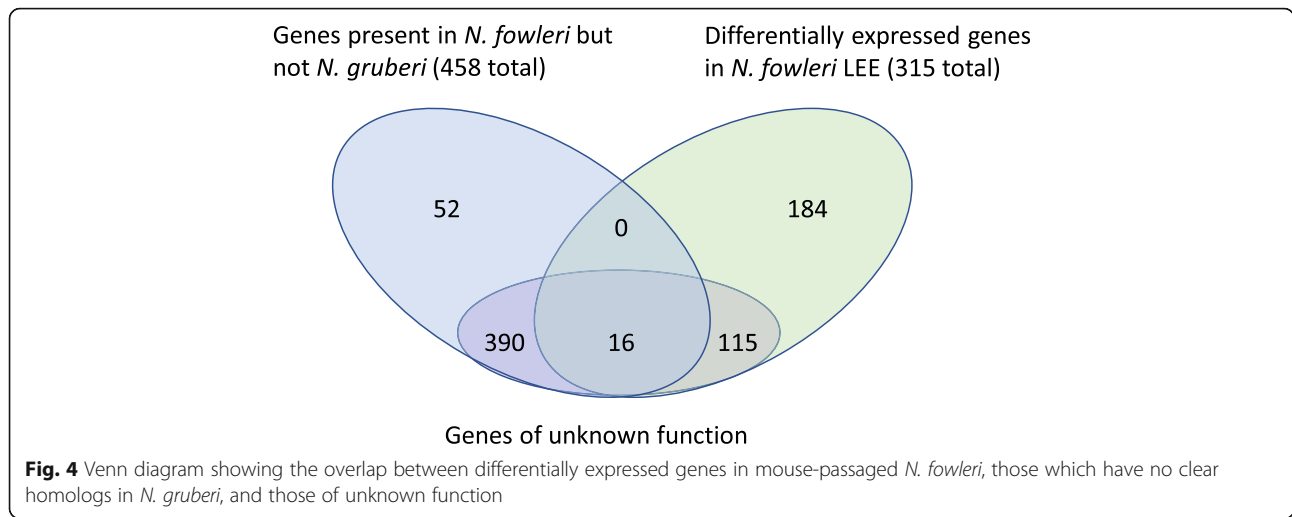
In this study, we provide a comparative assessment of genomic encoded proteomes between *N. fowleri* strains and the first comprehensive system-level analysis to understand why this species of *Naegleria* is a highly fatal human pathogen while other species are essentially benign.

Our work is consistent with previous understanding of *N. fowleri* pathogenicity, but builds extensively upon it. Our transcriptomic analysis revealed increased expression of several genes previously considered as pathogenicity factors (e.g., actin, the prosaposin precursor gene of *Naegleriapore* A and B, phospholipases and Nf314 (Cathepsin A) [22, 56, 81]. In 2014, Zysset-Burri and colleagues published a proteomic screen of highly versus weakly virulent *N. fowleri*, as a function of culturing cells with different types of media [28]. While there were clear differences between these strains, potentially due to genetic differences and the method of virulence induction, there were some shared pathways. This included villin and severin, which were both more abundant in high virulence *N. fowleri* and are involved

in actin cytoskeletal dynamics, as well as a phospholipase D homolog.

However, our comparative genomics and transcriptomics approaches have identified many putative novel pathogenicity factors in *N. fowleri* (Fig. 4). We recognize that our comparison was limited to a single non-pathogenic species (*N. gruberi*) and so observed *N. fowleri*-specific aspects could be due to losses in *N. gruberi*. Given the more extensive predicted proteome of *N. gruberi* (15,708 genes) compared with the three *N. fowleri* strains considered here (average of 11,925 genes), we believe this effect to be minimal, but for any given protein the potential for this artifact exists. When multiple high-quality genomes from additional non-pathogenic *Naegleria* species become available, a re-analysis of the comparative genomic assessment will be worthwhile. Nonetheless, we identified key individual targets, such as the S81 protease and two cathepsin B proteases, which are both missing from *N. gruberi*, and differentially expressed in mouse-passaged *N. fowleri*. Moreover, taking a hierarchical approach of overlapping criteria, we can distill from high-throughput RNA-Seq data a catalog of novel potential pathogenicity factors. A total of 458 genes are shared by *N. fowleri* strains to the exclusion of *N. gruberi*, while 315 are differentially expressed upon pathogenicity inducing conditions; both are logical criteria for their consideration as potential pathogenicity factors. Annotation as “unknown function” was taken as a criterion for novelty, but not necessarily pathogenicity. At the intersection of these criteria, there are 390 genes of unknown function that are specific to *N. fowleri*, and 115 genes that are differentially expressed and are of unknown function. Sixteen genes fulfill all three criteria; they are specific to *N. fowleri*, are upregulated, and have no putative function. Notably, 90 of the upregulated genes do not appear to have a human ortholog and represent potential novel drug targets. Once genetic tools are developed in this lineage, these can be used to functionally characterize the most promising candidate genes and better study *N. fowleri* cell biology. These data

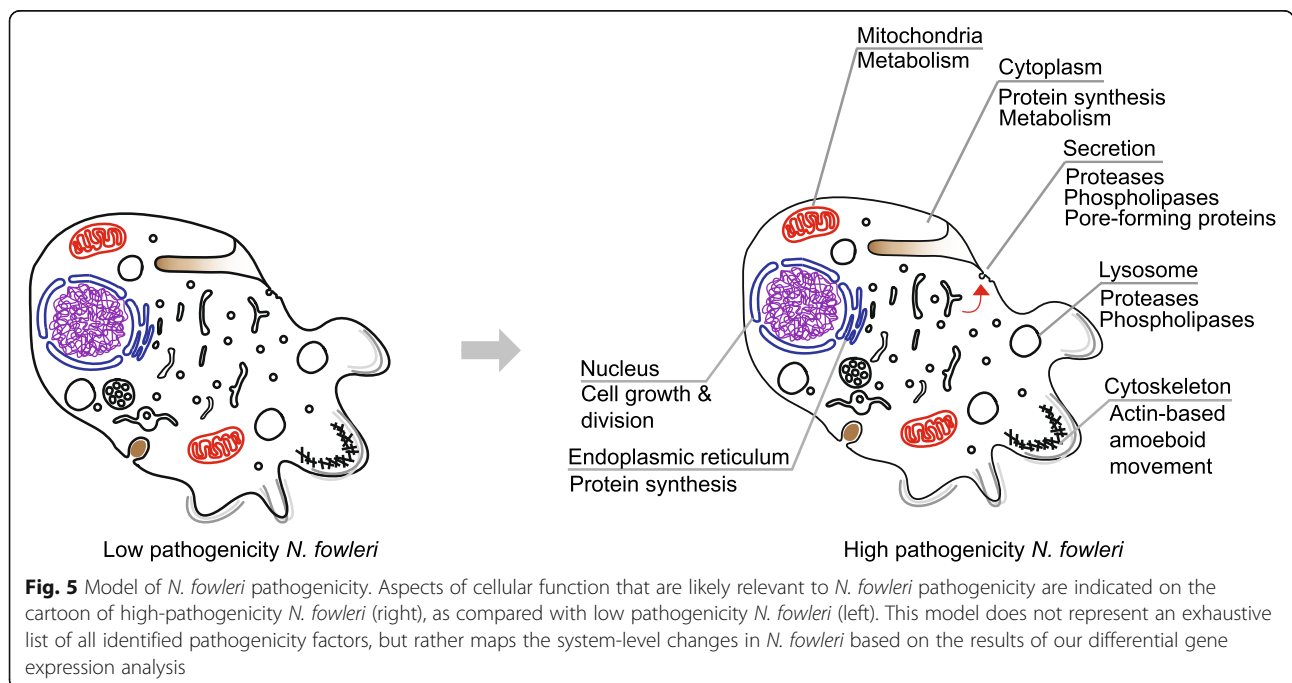




would greatly improve our understanding of why and how it is so virulent.

Our work has allowed us to generate a model for pathogenicity in *N. fowleri* (Fig. 5), hinging on both specific protein factors as well as whole cellular systems. Secreted proteases (e.g., metalloproteases, cysteine proteases, and pore-forming proteases) and phospholipases involved in host tissue destruction are likely to be secreted by the cell’s membrane trafficking system. The lysosomal system clearly serves a major role, potentially both as a secretory route and in the degradation pathway, and is a major source of differentially expressed genes in our analysis. We identified

many upregulated cathepsin proteases, which function in the lysosome, and predict that they are involved in ingestion and breakdown of host material. Increased cellular ingestion goes hand-in-hand with cell growth and division, processes which we also see represented in the upregulated gene dataset: genes involved in protein synthesis, metabolism, and mitochondrial function. This includes several metabolic pathways that produce compounds that could interact with the host immune system or have neurotropic effects. Finally, a major part of *N. fowleri*’s pathogenesis undoubtedly involves cell motility and phagocytosis, which are almost always actin-mediated processes.



## Conclusions

Overall, our comparative molecular evolutionary analysis, and the high-quality curated set of resources that have resulted, provides a rich and detailed cellular understanding of this enigmatic but unquestionably deadly human pathogen.

## Methods

### Culturing

The Australian *Naegleria fowleri* isolate 986 was obtained from an operational drinking water distribution system in rural Western Australia and verified as *N. fowleri* using qPCR-melt curve analysis [3] and DNA sequencing to confirm as Type 5 variant [98]. The Australian *N. fowleri* type 5 was cultured on axenic media, modified Nelson's medium consisting of 1 g/L Oxoid Liver Digest (Oxoid), 1 g/L Glucose, 24.0 g/L NaCl, 0.40 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.402 g/L CaCl<sub>2</sub>·2H<sub>2</sub>O, 14.2 g/L Na<sub>2</sub>HPO<sub>4</sub>, 13.60 g/L KH<sub>2</sub>PO<sub>4</sub>, 10 % v/v Hyclone Fetal Bovine Serum (Thermo Fisher), heat-inactivated, 0.5% v/v Vitamin Mix (0.05 g D- biotin (Sigma), 0.05 g Folic acid (Sigma), and 2.5 g Sodium hydroxide) and Gentomycin (100 µg/mL), and incubated at 37 °C in 75 cm<sup>2</sup> tissue culture flasks (Iwaki, Japan). Total DNA was extracted from the cultures as previously described using the PowerSoil DNA Isolation kit (MO BIO Laboratories, USA) [2, 99–101]. DNA was quantified using a Qubit (Invitrogen, USA) and stored at – 80 °C for subsequent genome sequencing.

*N. fowleri* V212 was cultured as described in Herman et al. [102]. *N. fowleri* LEE was grown axenically in oxoid media. Three replicates of *N. fowleri* LEE were passaged continuously through 50 B6C3F1 male mice. After mouse sacrifice, amoebae were extracted and grown in axenic media for 1 week to clear the culture of human cells prior to mRNA extraction.

### Genome sequencing and assembly

*Naegleria fowleri* V212 DNA was prepared and sequenced as described in Herman et al. [102]. Mitochondrial reads were first removed from 112,479,620 paired-end 100 bp Illumina and 609,044 454 reads using bowtie2 [103] and the remaining reads were de novo assembled using SPAdes v3.1.1 [104] and scaffolded using SSPACE [105], resulting in an assembly of 987 scaffolds > 200 bp totaling 27.7 Mb with an N50/N90 of 92,461/25,477 bp and depth of coverage of 251×. Notably, there are 366 sequences > 200 bp and < 1000 bp, comprising 387,133 bp or 1.4% of the assembly. Using the Geneious read mapper (22543367), 99.5% of 454 reads and 99.7% of Illumina reads mapped to the V212 mitochondrial genome and assembly.

For *N. fowleri* 986, a genomic sequencing library was constructed using the Illumina Nextera sequencing

library kit and sequenced on an Illumina HiSeq 2000 (100 bp paired-end sequencing). Raw sequence reads were imported into CLC Genomics v 10.0.1 (Qiagen), quality trimmed using the default parameters and assembled using the denovo assembly program in CLC (minimum contig length 1000 bp). Using BWA-MEM [106], 99.61% of the reads from 986 matched to the 986 genome. Assembly statistics for V212, 986, and 30863 are provided in Table 1.

In Zysset-Burri [28], the ~ 30 Mb genome assembly of 30863 was taken to be a haploid assembly of the *N. fowleri* diploid genome based in a comparison to obtained flow cytometry data. Given the similarity in size of our assembled V212 and 986 genomes to that of 30863, we assume the same for these strains.

### Transcriptomics

For gene prediction purposes, mRNA was extracted from *N. fowleri* CDC:V212 grown in axenic culture. Sequencing was done using an Illumina HiSeq platform, and ~ 150 million paired-end reads were generated. These were quality filtered using Trimmomatic [107] and transcripts were de novo assembled using Trinity software [108] with default parameters. These transcripts were then used as hints to generate gene models using Augustus, as described below.

For differential expression of genes associated with pathogenicity in *N. fowleri*, mRNA was extracted from three independent cultures of *N. fowleri* LEE-MP (mouse passaged), and from *N. fowleri* LEE-AX (grown only in axenic culture). One microgram of RNA was converted to first-strand cDNA in a 25 µl volume (USB M-MLV RT). Twenty microliters (from 25 µl total volume) from first-strand synthesis was converted to double-strand cDNA (dscDNA; NEBNext mRNA Second Strand Synthesis Module, #E6111S). Twenty microliters of dscDNA was then used for library generation (Illumina Nextera XT library preparation kit). Ampure beads were used for sample clean-up throughout. Libraries were sequenced on an Illumina MiSeq (2 × 300 cycles; 600 V3 sequencing kit). Illumina MiSeq sequencing was performed at The Applied Genomics Centre at the University of Alberta, generating paired-end 2 × 300 reads. Reads were pre-processed using Trimmomatic v0.32400 [107], by adaptor trimming, 5' end trimming (15 bp), trimming of regions where the average Phred score was < 20, and removal of short reads (< 50 bp). Remaining read set quality and characteristics were visualized using FastQC v0.11.2.401 [109].

The *N. fowleri* LEE-AX and LEE-MP reads were mapped to the *N. fowleri* V212 genome using TopHat v2.0.10 [110], with minimum intron length set to 30 bp, based on an assessment of predicted genes. Transcripts were then generated using Cufflinks with the

Reference Annotation Based Transcript option, using the predicted genes as a reference dataset [111, 112]. The reference transcripts were tiled with “faux-reads” to aid in the assembly, and these sequences were added to the final dataset containing the newly assembled transcripts. In order to obtain transcripts not represented by genes in the *N. fowleri* V212 genome, Trinity (release 2013-02-25) was used for purely de novo transcriptome assembly, using a genome-guided approach, with a `--genome_guided_max_intron` value of 5000 [113, 114]. Novel Trinity-generated transcripts were added to the Cufflinks-generated transcriptome prior to downstream differential expression analyses. The reads were then re-assembled by Trinity de novo using default parameters with the exception of `--jaccard_clip` in order to ensure that no LEE-specific transcripts were missed.

To identify genes in the *N. fowleri* LEE transcriptome that do not have any sequence similarity with any predicted proteins in the other *N. fowleri* strains, the *N. fowleri* LEE transcripts were used as BLASTX queries to search the predicted proteomes of *N. fowleri* V212, 30863, and 986. In this analysis, sequences that retrieved any hit were considered to share sequence similarity with one or more genes in the other strains, and not necessarily directly orthologous.

#### Differential expression analysis

Differential expression (DE) analyses were performed for the pathogenicity transcriptomic data using the programs Cuffdiff [114] and Trinity [113]. Cuffdiff was first used to map reads to the final transcriptome assembly, with high-abundance (> 10,000 reads) mitochondrial and extrachromosomal plasmid genes masked, and differential expression was calculated using geometric normalization. The Trinity Perl-to-R (PtR) toolkit was used to assess variation between replicates. One mouse passage replicate, MP2, was highly dissimilar to other mouse-passaged and axenically grown samples and was therefore excluded from further analysis. Using the Trinity suite of scripts, reads were aligned and transcript abundance estimated using RSEM [115] with TMM library normalization. Differentially expressed transcripts were then identified using EdgeR [116]. Comparison of these transcripts with those identified by Cuffdiff revealed highly overlapping datasets, with only a few genes considered to be differentially expressed by EdgeR but not Cuffdiff. Because this work is exploratory and we sought to minimize the potential for false negatives, we chose a lax false discovery rate cutoff of 0.1. However, the majority of upregulated genes have FDR values less than 0.05.

#### Gene prediction and annotation

Gene prediction was performed using the program Augustus v.2.5.5 [117, 118] incorporating the HiSeq-generated transcript dataset as extrinsic evidence termed “hints”. Furthermore, Augustus was also trained using a manually annotated 60 kb segment from the *N. fowleri* V212 genome published previously [102]. This region was annotated by using it as a tBLASTn query to search the non-redundant database. Gene boundaries were identified using the alignments of the top hits, and genes were annotated based on top BLAST hit identities. Gene prediction was performed for the *N. fowleri* V212 and 986 genomes generated by this study, as well as on the publicly available genome for strain 30863. For each genome, the parameter `--alternatives-from-evidence` was set to true, as this reports alternative gene transcripts if there is evidence for them (i.e., from transcriptome dataset). The parameter `--alternatives-from-sampling` was also set to true, as this outputs additional suboptimal transcripts. Parameters for determining the importance of different hint data were kept as default.

Annotation of genes in specific subsystems of focus in this manuscript was done using sequence similarity searching to assess putative homology. Functionally characterized homologs of proteins of interest were used as BLASTP [119] queries to search the predicted proteins from the *N. fowleri* strain genomes. At a minimum, putative *N. fowleri* homologs must be retrieved with an E-value < 0.05. To be considered true homologs, the *N. fowleri* protein must retrieve the original query or a differently named version thereof in a reciprocal BLAST search also with an E-value < 0.05.

Additionally, specific methods to assess the repertoire of several cellular systems were additionally used. For the cytoskeletal components of the three *N. fowleri* isolates, we first compared *N. fowleri* proteins to previously identified *N. gruberi* actin and microtubule associated proteins using BLAST [119] against the protein databases generated in Augustus for each of the three *N. fowleri* isolates, using default search parameters. The top *N. fowleri* hits identified by comparisons to *N. gruberi* were then compared using BLAST to the full *N. gruberi* protein library to establish the Reciprocal Best Hit which are included in Additional File 8-TableS6. We then searched for additional proteins not found in *N. gruberi* using human or *Dictyostelium discoideum* protein sequences obtained from PubMed and dictyBase (<http://dictybase.org>), respectively. To identify *N. fowleri* homologs, we again used BLAST with the default parameters except replacing the default scoring matrix with BLOSUM45. We further validated protein identities through hmmscan searches using a gathering threshold and Pfam domains on the HMMER website (<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>).

Protein domains of *N. fowleri* S81 protease (*Nfowleri*V212\_g9810.t1) were predicted using InterProScan [120] implemented in the Geneious Prime v2020.2.3 software [121]. Full-length sequence and sequence with the Invertebrate-type lysozyme (Lysozyme\_I; IPR008597) domain removed were used as queries in BLASTp [122] searches in NCBI non-redundant and EukProt [123] databases. This analysis was repeated using one of the two retrieved sequences *Ammonia* sp. (CAMPEP\_0197058342) as a query and both aforementioned databases. The validity of the retrieved sequences was verified by reverse BLASTp searches conducted against the *N. fowleri* protein database.

Mitochondrial protein location for all *N. fowleri* strains and the *N. gruberi* published genome [27] was determined based on a pipeline that tested several features. All predicted proteins were assessed for the following: (1) the presence of an N-terminal extension relative to bacterial/cytosolic homologs which was predicted as mitochondrial by Mitoprot [124], TargetP [125], and WoLFPSORT [126] and (2) a BLAST hit that was most significant against known mitochondrial proteins (and not against the non-mitochondrial paralogs from the same species). Biochemically confirmed mitochondrial proteomes from the following species were used for these BLAST searches: *Arabidopsis thaliana* [127], *Chlamydomonas reinhardtii* [128], *Homo sapiens* [129], *Mus musculus* [129], *Tetrahymena thermophila* [130], and *Saccharomyces cerevisiae* [131]. In addition, mitochondrial and hydrogenosomal predicted proteomes from the following species were also used in BLAST searches: *Giardia intestinalis* [132], *Entamoeba histolytica* [133], and *Trichomonas vaginalis* [134]. All hits were subsequently assessed against the Pfam [135] and SwissProt [136] databases. Proteins were ranked according to number of positive hits (e.g., mitochondrial targeting signal predicted by all predictors would be +3 points, a positive hit against confirmed mitochondrial proteomes: maximum +5, and the mitochondria/hydrogenosomal proteomes: maximum +3). All predictions were manually curated, compared with the previously [27] and the newly curated *N. gruberi* mitochondrial proteins and subsequently spurious predictions were removed. Proteins most similar to unidentified proteins despite having a predicted mitochondrial targeting signal were removed as well.

For analysis of Ras super family GTPases, genomic and predicted protein sequences of *N. fowleri* strain V212, 986 and 30863, as well as *N. gruberi* were analyzed. For comparative analyses, we also identified Ras superfamily genes from a set of reference, phylogenetically diverse species using data available in the NCBI database (<http://www.ncbi.nlm.nih.gov>). Ras superfamily gene sequences were detected and preliminarily

identified using the program BLAST and its variants [122]. The protein sequences were manually inspected and prediction of the underlying gene models corrected whenever necessary. Multiple protein alignments were constructed using the program MAFFT [137]. All alignments were checked and if necessary, further edited manually using BioEdit [138]. Phylogenetic trees were computed using the maximum likelihood method implemented in the program RAxML-HPC with the LG +  $\Gamma$  model and branch support assessed by the rapid bootstrapping algorithm [139] at the CIPRES Science Gateway portal [140]. The robustness was additionally assessed by the maximum likelihood method implemented in the IQ-tree program [141] with branch supports assessed by the ultrafast bootstrap approximation [142] and SH-aLRT test [143]. Genes with one-to-one orthology relationships were detected by maximum likelihood phylogenetic analysis (IQ-tree) or BLASTp searches. Two genes were considered as one-to-one orthologs if they form a monophyletic group to the exclusion of other analyzed sequences with support at least 80/95 (SH-aLRT support (%) / ultrafast bootstrap support (%)), or if they represent reciprocally best hits in BLASTp searches in our in-house Ras superfamily GTPase database containing also Ras superfamily genes from various other taxa. For the analysis of Rab proteins, the dataset from Elias et al. [72] was used and expanded with sequences from *Naegleria* spp. Protein domains were searched using SMART [144], Pfam [34], and NCBI's conserved domain database [145]. Several suspicious domains detected only by NCBI's conserved domain with low e-value and/or representing only a small part of detected domain were excluded from the results.

### Phylogenetic analysis

Bayesian and maximum likelihood phylogenetic analyses were performed to assign orthology to proteins from highly paralogous families. Sequences were aligned using MUSCLE v.3.8.31 [146], alignments were visualized in Mesquite v.3.2 [147], and manually masked and trimmed to remove positions of uncertain homology. ProtTest v3.4 [148] was used to determine the best-fit model of sequence evolution. PhyloBayes v4. 1[149] and MrBAYES v3.2.2 [150] programs were run for Bayesian analysis, and RAxML v8.1.3 [139] was run for maximum likelihood analysis. Phylobayes was run until the largest discrepancy observed across all bipartitions was less than 0.1 and at least 100 sampling points were achieved. MrBAYES was used to search treespace for a minimum of one million MCMC generations, sampling every 1000 generations, until the average standard deviation of the split frequencies of two independent runs (with two chains each) was less than 0.01. Consensus trees were generated using a burn-in value of 25%, well above the

likelihood plateau in each case. RAxML was run with 100 pseudoreplicates.

### Genetic diversity analysis

As the highly cited analysis proposing that the genus *Naegleria* shows equivalent diversity to tetrapods [30] was based on only 281 nucleotides comparing four *Naegleria* species and five tetrapod sequences, we performed a more extensive analysis. For comparison purposes of *Naegleria* small subunit ribosomal RNA (18S rRNA) gene sequences, *Naegleria fowleri* 18S rRNA sequence (accession number NFT80059) was used as a query in BLASTN search [122] against nucleotide database on NCBI restricting search to the genus *Naegleria* and excluding *N. fowleri*. Only full-length sequences were used in further analysis. For comparison of tetrapods, 18S rRNA sequences were retrieved randomly with the aim to represent all major clades of tetrapods. 18S rRNA sequences were aligned using MAFFT v7.458 [137] under L-INS-i strategy. The resulting alignments were imported into the Geneious Prime v2020.2.3 software [121] and percent identities were obtained directly under “Distances” tab. Geometric mean and median were calculated using built-in functions in Excel. Overall, the hypothesis proposed by Baverstock [30] is supported but now with much more comprehensive data.

### BUSCO analysis

BUSCO v4.0.4 software [151, 152] was used to assess genome completeness. Predicted proteins were used as input, with the eukaryote dataset eukaryota\_odb10 set of hidden Markov models.

### Orthologous groups analysis

To identify orthologous groups of sequences between the four *Naegleria* predicted proteomes, the program OrthoMCL v.2.0.9424 was used [153]. The Markov Clustering algorithm is an unsupervised clustering algorithm that clusters graphs based on pairwise scores (in this case, normalized E-values following an all-versus-all BLAST) and an inflation value. This latter value controls the clustering tightness and, for these analyses, was kept at the suggested 1.5.

For further manual analysis of potential orthogroups in *N. fowleri* strains but not *N. gruberi*, the above BLAST search criteria were used. Protein sequences from *N. fowleri* V212 were used to search the predicted proteome of *N. gruberi* using BLASTp, as well as the *N. gruberi* genome using tBLASTn. Any hits with an E-value less than 0.05 were used as queries in a reciprocal BLAST search against the *N. fowleri* V212 predicted proteome. *N. gruberi* sequences were considered homologous to the orthogroup protein if they retrieved the

original query with an E-value < 0.05 (although the observed E-values were typically much lower). These relaxed BLAST criteria were designed to capture even highly divergent sequence, in order to be confident that the manually curated orthogroup dataset does not contain *N. gruberi* homologs. A total of 458 genes were identified as shared by all three *N. fowleri* strains and absent from *N. gruberi*.

### Transmembrane domain prediction

The program TMHMM v2.0426 [154, 155] was used to detect transmembrane helices in all *N. fowleri* V212 proteins. This software uses an HMM generated from 160 cross-validated membrane proteins, and outputs all predicted helices and protein orientation in the membrane. Because transmembrane helix prediction was done to generate a list of potential G protein-coupled receptors investigated further by domain prediction, scoring cutoffs were not used.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01078-1>.

**Additional file 1: Table S1.** % identity of SSU rRNA genes. Sheet 1: *Naegleria* species. Sheet 2: Selected tetrapod organisms.

**Additional file 2: Table S2.** Meiosis gene comparative genomics.

**Additional file 3. Figures S1-S10.**

**Additional file 4: Supplementary Material 1.** Transcription factor identification. **Supplementary Material 2.** Sterol metabolism genes from *Naegleria fowleri*. **Supplementary Material 3.** Analysis of the cytoskeletal protein complement in *N. fowleri*. **Supplementary Material 4.** Ras superfamily GTPases in *Naegleria* spp.

**Additional file 5: Table S3.** Transcription factors in *Naegleria* spp.

**Additional file 6: Table S4.** Genes involved in sterol production in *N. fowleri*.

**Additional file 7: Table S5.** Comparative genomic assessment of mitochondrial genes in *Naegleria* spp.

**Additional file 8: Table S6.** Cytoskeletal genes in *N. fowleri*.

**Additional file 9: Table S7.** Results of homology searching for membrane trafficking components in the predicted proteomes and genomes of *N. fowleri* V212, 39863, and 986.

**Additional file 10: Table S8.** Comparative genomics analysis of Ras superfamily GTPases in *Naegleria* spp.

**Additional file 11: Table S9.** Putative functional annotation of genes identified in *N. fowleri* but not *N. gruberi*.

**Additional file 12: Table S10.** Differentially expressed genes following mouse-passage of *N. fowleri* LEE. Grey shading indicates genes found in *N. fowleri* alone or *N. fowleri*.

**Additional file 13: Table S11.** Comparative genomic analysis of proteases identified in *Naegleria* spp.

**Additional file 14: Table S12.** Homology searching results for protease S81.

**Additional file 15: Table S13.** Comparative genomic analysis of genes involved in autophagy in *N. fowleri*.

**Additional file 16: Table S14.** Comparative genomic analysis of genes involved in the Unfolded Protein Response and the ER-Associated Degradation machinery in *N. fowleri*.

**Additional file 17: Table 15.** Repeat domain-containing proteins in *Naegleria* spp. identified by BLAST search using human adhesion G protein-coupled receptors as queries.

**Additional file 18: Table S16.** Predicted adhesion G protein-coupled receptors in *Naegleria* spp.

**Additional file 19: Table S17.** TM9 orthologues identified in *Naegleria* spp.

### Acknowledgements

JBD would like to thank M. Jamerson (VCU) for helpful discussion. Moreover, all authors wish to thank the many lab members, administrative, and technical staff who have supported us through this project. There are too many to name, but we hope that you know who you are and how greatly you are valued.

### Authors' contributions

Generated unpublished primary data included in the manuscript: EKH, AG, HCM, MJM, DCZ-B, FMC. Performed curated analysis of an encoded genomic system or additional analyses and provided display items (includes supervision of trainee): EKH, IM-R, RV, MvdG, MLG, AT, KZ, KV, SRN, ME, CS, MW, LF-L, JBD. Edited or provided intellectual input on the manuscript: EKH, AG, MvdG, MLG, AT, KV, SRN, GM, KZ, NM, ME, MW, LF-L, GJP, TW, CC, JBD. Initial conception of the project: FMC, GJP, MWI TW, NM, CC, JBD. All authors read and approved the final manuscript

### Funding

EKH was supported by an AHFMR Fulltime Graduate Studentship and a Vanier Canada Graduate Scholarship. JBD is the Canada Research Chair in Evolutionary Cell Biology. Work in the Eliáš Lab is supported by CePaViP (CZ.02.1.01/0.0/0.0/16\_019/0000759) and Czech Science Foundation project 18-18699S. This work was supported by CSIRO Land and Water (Australia).

### Availability of data and materials

Genomic sequence data for the three *N. fowleri* genomes (V212 [156], 986 [157], 30863 [28]) and the associated gff files are available on Zenodo <https://zenodo.org/record/3951746#.XxW7xyhKhPZ> [158]. Reads associated with the transcriptomics experiment of the LEE strain are deposited in the SRA (SRX8770894-SRX8770899) with the associated BioProject ID PRJNA647238 [159].

### Declarations

#### Ethics approval and consent to participate

Care of animals was in compliance with the standards of the National Institutes of Health and the Institutional Animal Care and Use Committee at Virginia Commonwealth University.

#### Consent for publication

Not applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Division of Infectious Disease, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada. <sup>2</sup>Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada. <sup>3</sup>Laboratory Medicine and Medicine / Infectious Diseases, UCSF-Abbott Viral Diagnostics and Discovery Center, UCSF Clinical Microbiology Laboratory UCSF School of Medicine, San Francisco, USA. <sup>4</sup>Department of Laboratory Medicine, University of Washington Medical Center, Montlake, USA. <sup>5</sup>Centre for Organelle Research, Department of Chemistry, Bioscience and Environmental Engineering, University of Stavanger, Stavanger, Norway. <sup>6</sup>School of Applied Sciences, Department of Biological and Geographical Sciences, University of Huddersfield, Huddersfield, UK. <sup>7</sup>Department of Cardiology, Hospital Clínico Universitario Virgen de la Arrixaca. Instituto Murciano de Investigación Biosanitaria. Centro de Investigación Biomedica en Red-Enfermedades Cardiovasculares (CIBERCV), Madrid, Spain. <sup>8</sup>CSIRO Land and Water, Centre for Environment and Life Sciences, Private Bag No.5, Wembley, Western Australia 6913,

Australia. <sup>9</sup>CSIRO, Indian Oceans Marine Research Centre, Environomics Future Science Platform, Crawley, WA, Australia. <sup>10</sup>CSIRO Land and Water, Black Mountain Laboratories, Canberra, Australia. <sup>11</sup>School of Biosciences, University of Kent, Canterbury, UK. <sup>12</sup>Department of Biology, University of Massachusetts, Amherst, UK. <sup>13</sup>Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic. <sup>14</sup>Faculty of Science, Charles University, BIOCEV, Prague, Czech Republic. <sup>15</sup>Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic. <sup>16</sup>Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain. <sup>17</sup>Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Catalonia, Spain. <sup>18</sup>Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada. <sup>19</sup>Institute of Parasitology, Vetsuisse Faculty Bern, University of Bern, Bern, Switzerland. <sup>20</sup>Spiez Laboratory, Federal Office for Civil Protection, Austruasse, Spiez, Switzerland. <sup>21</sup>Department of Ophthalmology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. <sup>22</sup>Department of Biochemistry and Molecular Biology, Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Canada. <sup>23</sup>Center for Autoimmune Genomics and Etiology and Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>24</sup>Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, USA. <sup>25</sup>Department of Microbiology and Immunology, Virginia Commonwealth University School of Medicine, Richmond, Virginia, USA. <sup>26</sup>Department of Life Sciences, The Natural History Museum, London, UK.

Received: 16 December 2020 Accepted: 24 June 2021

Published online: 22 July 2021

### References

- Carter RF. Description of a *Naegleria* sp. isolated from two cases of primary amoebic meningo-encephalitis, and of the experimental pathological changes induced by it. *J Pathol.* 1970;100:217–44.
- Puzon GJ, Wylie JT, Walsh T, Braun K, Morgan MJ. Comparison of biofilm ecology supporting growth of individual *Naegleria* species in a drinking water distribution system. *FEMS Microbiol Ecol.* 2017;93:1–8.
- Puzon GJ, Lancaster JA, Wylie JT, Plumb JJ. Rapid detection of *Naegleria fowleri* in water distribution pipeline biofilms and drinking water samples. *Environ Sci Technol.* 2009;43:6691–6.
- Morgan MJ, Halstrom S, Wylie JT, Walsh T, Kaksonen AH, Sutton D, et al. Characterization of a drinking water distribution pipeline terminally colonized by *Naegleria fowleri*. *Environ Sci Technol.* 2016;50(6):2890–8. <https://doi.org/10.1021/acs.est.5b05657>.
- Kazi AN, Riaz T. Deaths from rare protozoan encephalitis in Karachi blamed on unchlorinated water. *BMJ.* 2013;346:4461.
- Mahmood K. *Naegleria fowleri* in Pakistan - an emerging catastrophe. *J Coll Physicians Surg Pak.* 2015;25:159–60.
- Naqvi AA, Yazdani N, Ahmad R, Zehra F, Ahmad N. Epidemiology of primary amoebic meningoencephalitis-related deaths due to *Naegleria fowleri* infections from freshwater in Pakistan: An analysis of 8-year dataset. *Arch Pharm Pract.* 2016;7:119–29.
- Dorsch MM. Primary amoebic meningoencephalitis: an historical and epidemiological perspective with particular reference to South Australia. Adelaide: Epidemiology Branch, South Australian Health Commission; 1982.
- Cope JR, Ratard RC, Hill VR, Sokol T, Causey JJ, Yoder JS, et al. The first association of a primary amoebic meningoencephalitis death with culturable *Naegleria fowleri* in tap water from a US treated public drinking water system. *Clin Infect Dis.* 2015;60(8):e36–42. <https://doi.org/10.1093/cid/civ017>.
- Yoder JS, Straif-Bourgeois S, Roy SL, Moore TA, Visvesvara GS, Ratard RC, et al. Primary amoebic meningoencephalitis deaths associated with sinus irrigation using contaminated tap water. *Clin Infect Dis.* 2012;55:79–85.
- Linam WM, Ahmed M, Cope JR, Chu C, Visvesvara GS, Da Silva AJ, et al. Successful treatment of an adolescent with *Naegleria fowleri* primary amoebic meningoencephalitis. *Pediatrics.* 2015;135(3):e744–8. <https://doi.org/10.1542/peds.2014-2292>.
- Cope JR, Conrad DA, Cohen N, Cotilla M, Dasilva A, Jackson J, et al. Use of the novel therapeutic agent miltefosine for the treatment of primary amoebic meningoencephalitis: report of 1 fatal and 1 surviving case. *Clin Infect Dis.* 2016;62(6):774–6. <https://doi.org/10.1093/cid/civ1021>.

13. Gharpure R, Bliton J, Goodman A, Ali IKM, Yoder JS, Cope JR. Epidemiology and clinical characteristics of primary amebic meningoencephalitis caused by *Naegleria fowleri*: a global review. *Clin Infect Dis*. 2020;ciaa520.
14. Matanock A, Mehal JM, Liu L, Blau DM, Cope JR. Estimation of undiagnosed *Naegleria fowleri* primary amebic meningoencephalitis, United States. *Emerg Infect Dis*. 2018;24(1):162–4. <https://doi.org/10.3201/eid2401.170545>.
15. Maciver SK, Piñero JE, Lorenzo-Morales J. Is *Naegleria fowleri* an emerging parasite? *Trends Parasitol*. 2019.
16. Gharpure R, Gleason M, Salah Z, Blackstock A, Hess-Homeier D, Yoder J, et al. Geographic range of recreational water-associated primary amebic meningoencephalitis, United States, 1978–2018. *Emerg Infect Dis J*. 2021; 27(1):271–4. <https://doi.org/10.3201/eid2701.202119>.
17. Baral R, Vaidya B. Fatal case of amoebic encephalitis masquerading as herpes. *Oxford Med Case Rep*. 2018;2018:134–7.
18. Kemble SK, Lynfield R, DeVries AS, Drehner DM, Pomputius WF, Beach MJ, et al. Fatal *Naegleria fowleri* infection acquired in Minnesota: possible expanded range of a deadly thermophilic organism. *Clin Infect Dis*. 2012; 54(6):805–9. <https://doi.org/10.1093/cid/cir961>.
19. Siddiqui R, Khan NA. Primary amoebic meningoencephalitis caused by *Naegleria fowleri*: an old enemy presenting new challenges. *PLoS Negl Trop Dis*. 2014;8.
20. De Jonckheere JF. What do we know by now about the genus *Naegleria*? *Exp Parasitol*. 2014;145:52–9. <https://doi.org/10.1016/j.exppara.2014.07.011>.
21. Aldape K, Huizinga H, Bouvier J, McKerrow J. *Naegleria fowleri*: characterization of a secreted histolytic cysteine protease. *Exp Pathol*. 1994; 78:230–41.
22. Herbst R, Ott C, Jacobs T, Marti T, Marciano-Cabral F, Leippe M. Pore-forming polypeptides of the pathogenic protozoan *Naegleria fowleri*. *J Biol Chem*. 2002;277(25):22353–60. <https://doi.org/10.1074/jbc.M201475200>.
23. Hu WN, Band RN, Kopachik WJ. Virulence-related protein synthesis in *Naegleria fowleri*. *Infect Immun*. 1991;59(11):4278–82. <https://doi.org/10.1128/iai.59.11.4278-4282.1991>.
24. Serrano-Luna J, Cervantes-Sandoval I, Tsutsumi V, Shibayama M. A biochemical comparison of proteases from pathogenic *Naegleria fowleri* and non-pathogenic *Naegleria gruberi*. *J Eukaryot Microbiol*. 2007;54(5):411–7. <https://doi.org/10.1111/j.1550-7408.2007.00280.x>.
25. Toney DM, Marciano-Cabral F. Alterations in protein expression and complement resistance of pathogenic *Naegleria* amoebae. *Infect Immun*. 1992;60(7):2784–90. <https://doi.org/10.1128/iai.60.7.2784-2790.1992>.
26. Barbour SE, Marciano-Cabral F. *Naegleria fowleri* amoebae express a membrane-associated calcium-independent phospholipase A2. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2001;1530(2-3):123–33. [https://doi.org/10.1016/S1388-1981\(00\)00069-X](https://doi.org/10.1016/S1388-1981(00)00069-X).
27. Fritz-Laylin LK, Prochnik SESE, Ginger MLML, Dacks JB, Carpenter MLML, Field MCMC, et al. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell*. 2010;140:631–42.
28. Zysset-Burri DC, Müller N, Beuret C, Heller M, Schürch N, Gottstein B, et al. Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. *BMC Genomics*. 2014;15:496.
29. Liechti N, Schürch N, Bruggmann R, Wittwer M. Nanopore sequencing improves the draft genome of the human pathogenic amoeba *Naegleria fowleri*. *Sci Rep*. 2019;9:16040.
30. Baverstock PR, Illana S, Christy PE, Robinson BS, Johnson AM. srRNA evolution and phylogenetic relationships of the genus *Naegleria* (Protista: Rhizopoda). *Mol Biol Evol*. 1989;6(3):243–57. <https://doi.org/10.1093/oxfordjournals.molbev.a040549>.
31. Koonin EV. The Incredible Expanding Ancestor of Eukaryotes. *Cell*. 2010; 140(5):606–8. <https://doi.org/10.1016/j.cell.2010.02.022>.
32. Weirauch MT, Hughes TR. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell Biochem*. 2011;52:25–73.
33. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. 2014;158:1431–43.
34. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. *Nucleic Acids Res*. 2014;42(D1):D222–30. <https://doi.org/10.1093/nar/gkt1223>.
35. Eddy SR. A new generation of homology search tools based on probabilistic inference. In: *Genome Informatics*; 2009. p. 2009.
36. Desmond E, Gribaldo S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol Evol*. 2009;1:364–81. <https://doi.org/10.1093/gbe/evp036>.
37. Choi JY, Podust LM, Roush WR. Drug strategies targeting CYP51 in neglected tropical diseases. *Chem Rev*. 2014;114(22):11242–71. <https://doi.org/10.1021/cr5003134>.
38. Fu C, Xiong J, Miao W. Genome-wide identification and characterization of cytochrome P450 monooxygenase genes in the ciliate *Tetrahymena thermophila*. *BMC Genomics*. 2009;10(1):208. <https://doi.org/10.1186/1471-2164-10-208>.
39. Raederstorff D, Rohmer M. Sterol biosynthesis via cycloartenol and other biochemical features related to photosynthetic phyla in the amoeba *Naegleria lovaniensis* and *Naegleria gruberi*. *Eur J Biochem*. 1987;164(2):427–34. <https://doi.org/10.1111/j.1432-1033.1987.tb11075.x>.
40. Debnath A, Calvet CM, Jennings G, Zhou W, Aksenov A, Luth MR, et al. CYP51 is an essential drug target for the treatment of primary amoebic meningoencephalitis (PAM). *PLoS Negl Trop Dis*. 2017;11(12):e0006104. <https://doi.org/10.1371/journal.pntd.0006104>.
41. Yoshiyama-Yanagawa T, Enya S, Shimada-Niwa Y, Yaguchi S, Haramoto Y, Matsuya T, et al. The conserved Rieske oxygenase DAF-36/Neverland is a novel cholesterol-metabolizing enzyme. *J Biol Chem*. 2011;286:25756–62.
42. Wollam J, Magomedova L, Magner DB, Shen Y, Rottiers V, Motola DL, et al. The Rieske oxygenase DAF-36 functions as a cholesterol 7-desaturase in steroidogenic pathways governing longevity. *Aging Cell*. 2011;10(5):879–84. <https://doi.org/10.1111/j.1474-9726.2011.00733.x>.
43. Najle SR, Nusblat AD, Nudel CB, Uttaro AD. The sterol-C7 desaturase from the ciliate *tetrahymena thermophila* is a rieske oxygenase, which is highly conserved in animals. *Mol Biol Evol*. 2013;30:1630–43.
44. Najle SR, Molina MC, Ruiz-Trillo I, Uttaro AD. Sterol metabolism in the filasterean *Capsaspora owczarzaki* has features that resemble both fungi and animals. *Open Biol*. 2016;6(7). <https://doi.org/10.1098/rsob.160029>.
45. Kodner RB, Summons RE, Pearson A, King N, Knoll AH. Sterols in a unicellular relative of the metazoans. *Proc Natl Acad Sci U S A*. 2008;105(29): 9897–902. <https://doi.org/10.1073/pnas.0803975105>.
46. Najle SR, Hernandez J, Ocana-Pallares E, Garcia Siburu N, Nusblat AD, Nudel CB, et al. Genome-wide transcriptional analysis of *tetrahymena thermophila* response to exogenous cholesterol. *J Eukaryot Microbiol*. 2019.
47. Lai EY, Walsh C, Wardell D, Fulton C. Programmed appearance of translatable flagellar tubulin mRNA during cell differentiation in *Naegleria*. *Cell*. 1979;17(4):867–78. [https://doi.org/10.1016/0092-8674\(79\)90327-1](https://doi.org/10.1016/0092-8674(79)90327-1).
48. Patterson M, Woodworth TW, Marciano-Cabral F, Bradley SG. Ultrastructure of *Naegleria fowleri* enflagellation. *J Bacteriol*. 1981;147(1):217–26. <https://doi.org/10.1128/jb.147.1.217-226.1981>.
49. González-Robles A, Cristóbal-Ramos AR, González-Lázaro M, Omaña-Molina M, Martínez-Palomo A. *Naegleria fowleri*: light and electron microscopy study of mitosis. *Exp Parasitol*. 2009;122:212–7.
50. Walsh CJ. The structure of the mitotic spindle and nucleolus during mitosis in the amoeba-flagellate *Naegleria*. *PLoS One*. 2012;7:e34763.
51. Jamerson M, Schmoyer JA, Park J, Marciano-Cabral F, Cabral GA. Identification of *Naegleria fowleri* proteins linked to primary amoebic meningoencephalitis. *Microbiology*. 2017;163(3):322–32. <https://doi.org/10.1099/mic.0.000428>.
52. Campellone KG, Welch MD. A nucleator arms race: cellular control of actin assembly. *Nat Rev Mol Cell Biol*. 2010;11(4):237–51. <https://doi.org/10.1038/nrm2867>.
53. Dominguez R. The WH2 domain and actin nucleation: necessary but insufficient. *Trends Biochem Sci*. 2016;41:478–90.
54. Rotty JD, Wu C, Bear JE. New insights into the regulation and cellular functions of the ARP2/3 complex. *Nat Rev Mol Cell Biol*. 2013;14(1):7–12. <https://doi.org/10.1038/nrm3492>.
55. Fritz-Laylin LK, Lord SJ, Mullins RD. WASP and SCAR are evolutionarily conserved in actin-filled pseudopod-based motility. *J Cell Biol*. 2017;216: 1673–88.
56. Sohn HJ, Kim JH, Shin MH, Song KJ, Shin HJ. The Nf-actin gene is an important factor for food-cup formation and cytotoxicity of pathogenic *Naegleria fowleri*. *Parasitol Res*. 2010;106:917–24.
57. Walsh CJ. The role of actin, actomyosin and microtubules in defining cell shape during the differentiation of *Naegleria* amoebae into flagellates. *Eur J Cell Biol*. 2007;86:85–98.
58. Breitsprecher D, Goode BL. Formins at a glance. *J Cell Sci*. 2013;126(Pt 1):1–7. <https://doi.org/10.1242/jcs.107250>.
59. Siddiqui R, Ali IKM, Cope JR, Khan NA. Biology and pathogenesis of *Naegleria fowleri*. *Acta Trop*. 2016;164:375–94.
60. Rojas-Hernández S, Jarillo-Luna A, Rodríguez-Monroy M, Moreno-Fierros L, Campos-Rodríguez R. Immunohistochemical characterization of the initial

- stages of *Naegleria fowleri* meningoencephalitis in mice. *Parasitol Res.* 2004; 94(1):31–6. <https://doi.org/10.1007/s00436-004-1177-6>.
61. Brown T. Observations by light microscopy on the cytopathogenicity of *Naegleria fowleri* in mouse embryo-cell cultures. *J Med Microbiol.* 1978; 11(3):249–59. <https://doi.org/10.1099/00222615-11-3-249>.
  62. Visvesvara GS, Callaway CS. Light and electron microscopic observations on the pathogenesis of *Naegleria fowleri* in mouse brain and tissue culture. *J Protozool.* 1974;21:239–50.
  63. Martínez-Castillo M, Cárdenas-Guerra RE, Arroyo R, Debnath A, Rodríguez MA, Sabanero M, et al. Nf-GH, a glycosidase secreted by *Naegleria fowleri*, causes mucin degradation: an in vitro and in vivo study. *Future Microbiol.* 2017;12(9):781–99. <https://doi.org/10.2217/fmb-2016-0230>.
  64. Fritz-Laylin LK, Cande WZ. Ancestral centriole and flagella proteins identified by analysis of *Naegleria* differentiation. *J Cell Sci.* 2010;123(Pt 23):4024–31. <https://doi.org/10.1242/jcs.077453>.
  65. Wennerberg K, Rossman KL, Der CJ. The Ras superfamily at a glance. *J Cell Sci.* 2005;118(Pt 5):843–6.
  66. Görlich D, Mattaj JW. Nucleocytoplasmic transport. *Science.* 1996;271(5255): 1513–8. <https://doi.org/10.1126/science.271.5255.1513>.
  67. Vlahou G, Eliáš M, von Kleist-Retzow J-C, Wiesner RJ, Rivero F. The Ras related GTPase Miro is not required for mitochondrial transport in *Dictyostelium discoideum*. *Eur J Cell Biol.* 2011;90(4):342–55. <https://doi.org/10.1016/j.ejcb.2010.10.012>.
  68. Kipreos ET, Pagano M. The F-box protein family. *Genome Biol.* 2000;1:REVIEWS3002.
  69. Perez-Torrado R, Yamada D, Defossez P-A. Born to bind: the BTB protein-protein interaction domain. *Bioessays.* 2006;28(12):1194–202. <https://doi.org/10.1002/bies.20500>.
  70. Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, et al. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome Biol.* 2011;12(2):R20. <https://doi.org/10.1186/gb-2011-12-2-r20>.
  71. Liu Y, Lecal J, Firtel RA, Kortholt A. Connecting G protein signaling to chemoattractant-mediated cell polarity and cytoskeletal reorganization. *Small GTPases.* 2018;9(4):360–4. <https://doi.org/10.1080/21541248.2016.1235390>.
  72. Elias M, Brighthouse A, Gabernet-Castello C, Field MC, Dacks JB. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J Cell Sci.* 2012;125(Pt 10):2500–8. <https://doi.org/10.1242/jcs.101378>.
  73. Marín I, van Egmond WN, van Haastert PJM. The Roco protein family: a functional perspective. *FASEB J Off Publ Fed Am Soc Exp Biol.* 2008;22: 3103–10.
  74. van Dam TJP, Zwartkruis FJT, Bos JL, Snel B. Evolution of the TOR pathway. *J Mol Evol.* 2011;73:209–20.
  75. Záhonová K, Petříčková R, Valach M, Yazaki E, Tikhonov DV, Butenko A, et al. Extensive molecular tinkering in the evolution of the membrane attachment mode of the Rheb GTPase. *Sci Rep.* 2018;8(1):5239. <https://doi.org/10.1038/s41598-018-23575-0>.
  76. Leung KF, Baron R, Seabra MC. Thematic review series: lipid posttranslational modifications. Geranylgeranylation of Rab GTPases. *J Lipid Res.* 2006;47:467–75.
  77. Elias M, Novotny M. cPRAS: a novel circularly permuted RAS-like GTPase domain with a highly scattered phylogenetic distribution. *Biol Direct.* 2008; 3:21.
  78. van Dam TJP, Bos JL, Snel B. Evolution of the Ras-like small GTPases and their regulators. *Small GTPases.* 2011;2(1):4–16. <https://doi.org/10.4161/sqtp.2.1.15113>.
  79. Ji W, Rivero F. Atypical Rho GTPases of the RhoBTB subfamily: roles in vesicle trafficking and tumorigenesis. *Cells.* 2016;5.
  80. Whiteman LY, Marciano-Cabral F. Susceptibility of pathogenic and nonpathogenic *Naegleria* spp. to complement-mediated lysis. *Infect Immun.* 1987;55:2442–7.
  81. Hu WN, Kopachik W, Band RN. Cloning and characterization of transcripts showing virulence-related gene expression in *Naegleria fowleri*. *Infect Immun.* 1992;60:2418–24.
  82. Sussman DJ, Lai EY, Fulton C. Rapid disappearance of translatable actin mRNA during cell differentiation in *Naegleria*. *J Biol Chem.* 1984;259(11): 7355–60. [https://doi.org/10.1016/S0021-9258\(17\)39879-4](https://doi.org/10.1016/S0021-9258(17)39879-4).
  83. Nag S, Larsson M, Robinson RC, Burnick LD. Gelsolin: the tail of a molecular gymnast. *Cytoskeleton.* 2013;70:360–84.
  84. Kayman SC, Clarke M. Relationship between axenic growth of *Dictyostelium discoideum* strains and their track morphology on substrates coated with gold particles. *J Cell Biol.* 1983;97(4):1001–10. <https://doi.org/10.1083/jcb.97.4.1001>.
  85. Sillo A, Bloomfield G, Balest A, Balbo A, Pergolizzi B, Peracino B, et al. Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*. *BMC Genomics.* 2008;9:291.
  86. Jamerson M, da Rocha-Azevedo B, Cabral GA, Marciano-Cabral F. Pathogenic *Naegleria fowleri* and non-pathogenic *Naegleria lovaniensis* exhibit differential adhesion to, and invasion of, extracellular matrix proteins. *Microbiology.* 2012;158(3):791–803. <https://doi.org/10.1099/mic.0.055020-0>.
  87. Kang S-Y, Song K-J, Jeong S-R, Kim J-H, Park S, Kim K, et al. Role of the Nfa1 protein in pathogenic *Naegleria fowleri* cocultured with CHO target cells. *Clin Vaccine Immunol.* 2005;12:873–6.
  88. Bexkens ML, Zimorski V, Sarink MJ, Wienk H, Brouwers JF, De Jonckheere JF, et al. Lipids are the preferred substrate of the protist *Naegleria gruberi*, relative of a human brain pathogen. *Cell Rep.* 2018;25:537–543.e3.
  89. Danbolt NC. Glutamate uptake. *Prog Neurobiol.* 2001;65:1–105.
  90. Featherstone DE, Shippy SA. Regulation of synaptic transmission by ambient extracellular glutamate. *Neuroscientist.* 2008;14(2):171–81. <https://doi.org/10.1177/1073858407308518>.
  91. Holtze M, Mickiené A, Atlas A, Lindquist L, Schwieler L. Elevated cerebrospinal fluid kynurenic acid levels in patients with tick-borne encephalitis. *J Intern Med.* 2012;272:394–401.
  92. Opperdoes FR, De Jonckheere JF, Tielens AGM. *Naegleria gruberi* metabolism. *Int J Parasitol.* 2011;41:915–24.
  93. Colotti G, Ilari A. Polyamine metabolism in *Leishmania*: from arginine to trypanothione. *Amino Acids.* 2011;40(2):269–85. <https://doi.org/10.1007/s00726-010-0630-3>.
  94. Fairlamb AH, Cerami A. Metabolism and functions of trypanothione in the Kinetoplastea. *Annu Rev Microbiol.* 1992;46(1):695–729. <https://doi.org/10.1146/annurev.mi.46.1.00192.003403>.
  95. Ondarza RN, Hurtado G, Tamayo E, Iturbe A, Hernández E. *Naegleria fowleri*: a free-living highly pathogenic amoeba contains trypanothione/trypanothione reductase and glutathione/glutathione reductase systems. *Exp Parasitol.* 2006;114(3):141–6. <https://doi.org/10.1016/j.exppara.2006.03.001>.
  96. Steiger RF, Steiger E. Cultivation of *Leishmania donovani* and *Leishmania braziliensis* in defined media: nutritional requirements. *J Protozool.* 1977;24: 437–41.
  97. Krassner SM, Flory B. Essential amino acids in the culture of *Leishmania tarentolae*. *J Parasitol.* 1971;57:917–20.
  98. de Jonckheere JF. Origin and evolution of the worldwide distributed pathogenic amoeboid flagellate *Naegleria fowleri*. *Infect Genet Evol.* 2011; 11(7):1520–8. <https://doi.org/10.1016/j.meegid.2011.07.023>.
  99. Miller HC, Wylie J, Dejean G, Kaksonen AH, Sutton D, Braun K, et al. Reduced efficiency of chlorine disinfection of *Naegleria fowleri* in a drinking water distribution biofilm. *Environ Sci Technol.* 2015;49(18):11125–31. <https://doi.org/10.1021/acs.est.5b02947>.
  100. Miller HC, Morgan MJ, Wylie JT, Kaksonen AH, Sutton D, Braun K, et al. Elimination of *Naegleria fowleri* from bulk water and biofilm in an operational drinking water distribution system. *Water Res.* 2017;110:15–26. <https://doi.org/10.1016/j.watres.2016.11.061>.
  101. Miller HC, Wylie JT, Kaksonen AH, Sutton D, Puzon GJ. Competition between *Naegleria fowleri* and free living amoeba colonizing laboratory scale and operational drinking water distribution systems. *Environ Sci Technol.* 2018; 52:2549–57.
  102. Herman EK, Greninger AL, Visvesvara GS, Marciano-Cabral F, Dacks JB, Chiu CY. The mitochondrial genome and a 60-kb nuclear DNA segment from *Naegleria fowleri*, the causative agent of primary amoebic meningoencephalitis. *J Eukaryot Microbiol.* 2013;60(2):179–91. <https://doi.org/10.1111/jeu.12022>.
  103. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>.
  104. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
  105. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27(4):578–9. <https://doi.org/10.1093/bioinformatics/btq683>.



106. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM: arXiv e-prints; 2013. p. arXiv:1303.3997.
107. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
108. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013.
109. Andrews S. FastQC: a quality control tool for high throughput sequence data; 2010.
110. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
111. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5. <https://doi.org/10.1038/nbt.1621>.
112. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics*. 2011;27: 2325–9.
113. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512. <https://doi.org/10.1038/nprot.2013.084>.
114. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–78. <https://doi.org/10.1038/nprot.2012.016>.
115. Li B, Dewey C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011;12(1):323. <http://www.biomedcentral.com/1471-2105/12/323>. <https://doi.org/10.1186/1471-2105-12-323>.
116. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
117. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res*. 2004;32(suppl 2):W309–12. <https://doi.org/10.1093/nar/gkh379>.
118. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19(suppl 2):ii215–25. <https://doi.org/10.1093/bioinformatics/btg1080>.
119. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
120. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30: 1236–40.
121. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28: 1647–9.
122. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402. <https://doi.org/10.1093/nar/25.17.3389>.
123. Richter DJ, Berney C, Strassert JFH, Burki F, de Vargas C. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. *bioRxiv*. 2020:2020.06.30.180687. <https://doi.org/10.1101/2020.06.30.180687>.
124. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem*. 1996;241:779–86.
125. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 2000;300(4):1005–16. <https://doi.org/10.1006/jmbi.2000.3903>.
126. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007; 35(suppl\_2):W585–7. <https://doi.org/10.1093/nar/gkm259>.
127. Millar AH, Sweetlove LJ, Giegé P, Leaver CJ. Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol*. 2001;127(4):1711–27. <https://doi.org/10.1104/pp.010387>.
128. Atteia A, Adrait A, Brugière S, Tardif M, van Lis R, Deusch O, et al. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. *Mol Biol Evol*. 2009;26(7): 1533–48. <https://doi.org/10.1093/molbev/msp068>.
129. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong S-E, et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell*. 2008;134(1):112–23. <https://doi.org/10.1016/j.cell.2008.06.016>.
130. Smith DGS, Gawryluk RMR, Spencer DF, Pearlman RE, Siu KWM, Gray MW. Exploring the mitochondrial proteome of the ciliate protozoan *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. *J Mol Biol*. 2007;374(3):837–63. <https://doi.org/10.1016/j.jmb.2007.09.051>.
131. Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, et al. The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci U S A*. 2003;100(23):13207–12. <https://doi.org/10.1073/pnas.2135385100>.
132. Jedelský PL, Doležal P, Rada P, Pyrih J, Smíd O, Hrdý I, et al. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS One*. 2011;6(2):e17285. <https://doi.org/10.1371/journal.pone.0017285>.
133. Mi-ichi F, Abu Yousuf M, Nakada-Tsukui K, Nozaki T. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc Natl Acad Sci U S A*. 2009;106(51):21731–6. <https://doi.org/10.1073/pnas.0907106106>.
134. Schneider RE, Brown MT, Shiflett AM, Dyal SD, Hayes RD, Xie Y, et al. The *Trichomonas vaginalis* hydrogenosome proteome is highly reduced relative to mitochondria, yet complex compared with mitosomes. *Int J Parasitol*. 2011;41(13–14):1421–34. <https://doi.org/10.1016/j.ijpara.2011.10.001>.
135. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Res*. 2004;32(Database issue): D138–41. <https://doi.org/10.1093/nar/gkh121>.
136. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its new supplement TrEMBL. *Nucleic Acids Res*. 1996;24:21–5.
137. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
138. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41: 95–8.
139. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
140. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE) 2010*. New Orleans; 2010. p. 1–8.
141. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285>.
142. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;30:1188–95.
143. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21. <https://doi.org/10.1093/sysbio/syq010>.
144. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015;43(Database issue):D257–60. <https://doi.org/10.1093/nar/gku949>.
145. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015; 43(Database issue):D222–6.
146. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
147. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. 2015. <http://mesquiteproject.org>.
148. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest-HPC: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;6586(LNCS):177–84.
149. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 2009; 25:2286–8. <https://doi.org/10.1093/bioinformatics/btp368>.
150. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42. <https://doi.org/10.1093/sysbio/sys029>.

151. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 2018;35(3):543–8. <https://doi.org/10.1093/molbev/msx319>.
152. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
153. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 2003;13(9):2178–89. <https://doi.org/10.1101/gr.1224503>.
154. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes<sup>11</sup>Edited by F. Cohen. *J Mol Biol.* 2001;305(3):567–80. <https://doi.org/10.1006/jmbi.2000.4315>.
155. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Sixth Int Conf Intell Syst Mol Biol.* 1998;6:175–82.
156. Herman EK, Greninger A, van der Giezen M, Ginger ML, Ramirez-Macias I, Miller HC, et al. BioProject PRJNA643799-N. fowleri strain V212 Genomic sequence and predicted proteins. Genomics and transcriptomics yields a systems-level view of the biology of the pathogen *Naegleria fowleri*. 2021.
157. Herman EK, Greninger A, van der Giezen M, Ginger ML, Ramirez-Macias I, Miller HC, et al. BioProject PRJNA734907 -N. fowleri strain 986 Genomic sequence and predicted proteins. Genomics and transcriptomics yields a systems-level view of the biology of the pathogen *Naegleria fowleri*.
158. Herman EK, Greninger A, van der Giezen M, Ginger ML, Ramirez-Macias I, Miller HC, et al. Genomes and predicted proteins of *Naegleria fowleri* strains CDC:V212, 986, and ATCC30863; 2021. <https://doi.org/10.1101/2020.01.16.908186>.
159. Herman EK, Greninger A, van der Giezen M, Ginger ML, Ramirez-Macias I, Miller HC, et al. BioProject PRJNA647238 -LEE RNASeq Reads. A comparative 'omics approach to candidate pathogenicity factor discovery in the brain-eating amoeba *Naegleria fowleri*. 2021. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA647238>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

