1   **A comparative 'omics approach to candidate pathogenicity factor discovery in the**

2   **brain-eating amoeba *Naegleria fowleri***

3

4   Emily K. Herman[1,2&] , Alex Greninger[3,4,*], Mark van der Giezen[5*], Michael L.
5   Ginger[6*,] Inmaculada Ramirez-Macias[1,7*], Haylea C. Miller[8*], Matthew J.
6   Morgan[9*,] Anastasios D. Tsaousis[10*], Katrina Velle[11*], Romana Vargová[12], Sebastian
7   Rodrigo Najle[13], Georgina MacIntyre[14], Norbert Muller[15,] Mattias Wittwer[16], Denise C.
8   Zysset-Burri[17], Marek Elias[12], Claudio Slamovits[18], Matthew Weirauch[19], Lillian Fritz-
9   Laylin[11^], Francine Marciano-Cabral[20^], Geoffrey J. Puzon[8^], Tom Walsh[9^], Charles
10   Chiu[3^,] Joel B. Dacks[1, 21,22&]
11
12   1= Division of Infectious Disease, Department of Medicine, Faculty of Medicine and
13   Dentistry, University of Alberta
14   2= Department of Agricultural, Food and Nutritional Science, University of Alberta
15   3= Laboratory Medicine and Medicine / Infectious Diseases, UCSF-Abbott Viral
16   Diagnostics and Discovery Center, UCSF Clinical Microbiology Laboratory UCSF
17   School of Medicine
18   4= Department of Laboratory Medicine, University of Washington Medical Center
19   5= Centre for Organelle Research, University of Stavanger
20   6= School of Applied Sciences, Department of Biological and Geographical Sciences,
21   University of Huddersfield
22   7= Department of Cardiology, Hospital Clinico Universitario Virgen de la Arrixaca.
23   Instituto Murciano de Investigación Biosanitaria. Centro de Investigación Biomedica en
24   Red-Enfermedades Cardiovasculares (CIBERCV)
25   8= CSIRO Land and Water
26   9= CSIRO Land and Water, Black Mountain Laboratories
27   10= Laboratory of Molecular and Evolutionary Parasitology, RAPID group, School of
28   Biosciences, University of Kent, Canterbury, UK
29   11= Department of Biology, University of Massachusetts
30   12= Department of Biology and Ecology, Faculty of Science, University of Ostrava
31   13= Institut de Biologia Evolutiva (UPF-CSIC)
32   14= Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta
33   15= Institute of Parasitology, Vetsuisse Faculty Bern, University of Bern
34   16= Spiez Laboratory, Federal Office for Civil Protection, Austrasse, Spiez, Switzerland
35   17= Department of Ophthalmology, Inselspital, Bern University Hospital, University of
36   Bern
37   18= Department of Biochemistry and Molecular Biology, Centre for Comparative
38   Genomics and Evolutionary Bioinformatics, Dalhousie University
39   19= Center for Autoimmune Genomics and Etiology and Divisions of Biomedical
40   Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center,
41   Cincinnati, Ohio, USA, Department of Pediatrics, University of Cincinnati College of
42   Medicine
43   20= Department of Microbiology and Immunology, Virginia Commonwealth University
44   School of Medicine
45   21= Department of Life Sciences, The Natural History Museum

1 22=Institute of Parasitology, Biology Centre, Czech Academy of Sciences

2 **& Corresponding author**

3

6

Abstract

8 Of the 40 described *Naegleria* species, only *N. fowleri* can establish infection in

9 humans, killing almost invariably within two weeks. In the brain, the amoeba performs

10 piece-meal ingestion, or trogocytosis, of brain material causing massive inflammation.

11 Conversely, its close relative *Naegleria gruberi*, which is used as a laboratory model

12 organism, is non-pathogenic. The exact pathogenicity factors distinguishing *N. fowleri*

13 from its harmless relatives are unclear. We have here taken an -omics approach to

14 understanding *N. fowleri* biology and infection at the system level. We provide the first

15 analysis of genomic diversity between strains, finding little conservation in synteny but

16 high conservation in protein complement. We also demonstrate that the *N. fowleri*

17 genome encodes a similarly complete cellular repertoire to that found in *N. gruberi*. Our

18 comparative genomic analysis, together with a transcriptomic analysis of low versus high

19 pathogenicity *N. fowleri* cultured in a mouse infection model, allowed us to construct a

20 model of cellular systems involved in pathogenicity and furthermore provides ~500 novel

21 candidate pathogenicity factors in this currently rare but highly fatal pathogen.

22

Introduction

24 *Naegleria fowleri* is an opportunistic pathogen of humans and animals, causing

25 primary amoebic meningoencephalitis (PAM), and killing up to 97% of those infected,

26 usually within two weeks (Carter, 1970). It is found in warm freshwaters around the

27 world and drinking water distribution systems (DWDS) (Morgan et al., 2016; Puzon et

28 al., 2009; Puzon et al., 2017) with *N. fowleri*-colonized DWDSs linked to deaths in

29 Pakistan (Kazi and Riaz, 2013; Mahmood, 2015; Naqvi et al., 2016), Australia (Dorsch,

30 1982), and the USA (Cope et al., 2015; Yoder et al., 2012). Infection occurs when

31 contaminated water enters the nose. *N. fowleri* passes through the cribriform plate to the

1   olfactory bulb in the brain and phagocytoses brain material, causing physical damage and

2   massive inflammation leading to death. Although successful treatment with miltefosine

3   and other antimicrobials is becoming more common (Cope et al., 2016; Linam et al.,

4   2015), this relies on appropriate early diagnosis, which is challenging due to the

5   relatively low incidence of amoebic versus viral or bacterial meningitis.

6   The reported incidence of infection is relatively low (Trabelsi et al., 2012);

7   however, infection is likely more prevalent, particularly in developing countries with

8   warm climates and inconsistent medical reporting. *N. fowleri* has been recently proposed

9   as an emerging pathogen based on increased case reports in the past decade (MacIver

10   et.al. 2019). With cases from temperate locations reported in recent years (Baral and

11   Vaidya, 2018; Kemble et al., 2012), the threat of *N. fowleri* may be exacerbated by range

12   expansion due to climate change (MacIver et.al. 2019), which has been associated with a

13   rise in freshwater temperatures, an increase in aquatic recreational activities (Siddiqui and

14   Khan, 2014), and extreme weather events. It is likely that we are still not fully aware of

15   the global scale of *N. fowleri* infection (Maciver et al., 2019), or how infection may

16   increase as climate change accelerates.

17   Of the ~40 species of *Naegleria*, found commonly in soils and fresh waters

18   worldwide, *N. fowleri* is the only species that typically infects humans, suggesting that

19   pathogenicity is a gained function. Several labs have identified potential *N. fowleri*

20   pathogenicity factors including proteases, lipases, and pore-forming proteins. The

21   hypothesized mechanism of *N. fowleri* pathogenicity – tissue degradation for both

22   motility during infection and phagocytosis – is compatible with these factors. However,

23   none appear to be unique to *N. fowleri*, and there are likely proteins factors yet to be

24   identified that are responsible for pathogenesis. In 2010 the genome sequence of the non-

25   pathogenic *Naegleria gruberi* was reported and in 2014, the genome of *N. fowleri* ATCC

26   30863 was sequenced and used to guide comparative proteomic work on highly and

27   weakly pathogenic *N. fowleri* (Zysset-Burri et al., 2014). The genome of a thermotolerant

28   but non-pathogenic species, *Naegleria lovaniensis*, was published in 2018 (Liechti et al.

29   2018). Despite such work, a genome-wide perspective on *N. fowleri* diversity and

30   pathogenesis is lacking.

1   Here we report the genome sequences of two *N. fowleri* strains; 986, an

2   environmental isolate from an operational DWDS in Western Australia, and CDC:V212,

3   a strain isolated from a patient. We also report a transcriptomic analysis of induced

4   pathogenicity in the *N. fowleri* strain LEE to identify the genes differentially expressed as

5   a consequence of infection.

6

7   **Results**

8   The genomes of *N. fowleri* strains V212 and 986 were sequenced to average

9   coverage of 251 x and 250 x respectively and the transcriptome of axenically cultured *N.*

10   *fowleri* V212 was also sequenced to support gene prediction. Analysis of genome

11   architecture shows limited synteny between *N. fowleri* strains but conservation of other

12   genome statistics, as compared to the *N. gruberi* genome (Figure 1, Table 1). Similarly,

13   while GC content and other coding content statistics were similar between *N. fowleri*

14   strains, they were remarkably different from those values for *N. gruberi* (Table 1) (Fritz-

15   Laylin et al., 2010a). The differences in genome statistics are not unreasonable given that

16   genetic diversity within the *Naegleria* clade has been equated to that of tetrapods

17   (Baverstock et al., 1989). Importantly, there is transcriptomic evidence for 82% of genes

18   in *N. fowleri* V212, suggesting that they are expressed when grown in culture and 100%

19   of the genes with transcriptome evidence were contained within the predicted genomic

20   set. Out of 303 near-universal single-copy eukaryotic orthologues, 277 were found in the

21   set of *N. fowleri* V212 predicted proteins, giving a BUSCO score of 91.4%, signifying

22   that the genome and predicted proteome are highly complete.

23

24   *The* Naegleria fowleri *genome encodes a complete and canonical cellular complement*

25   In 2010, *N. gruberi* was hailed as possessing extensive eukaryotic cytoskeletal,

26   membrane-trafficking, signaling, and metabolic machinery, suggesting sophisticated cell

27   biology for a free-living protist that diverged from other eukaryotic lineages over one

28   billion years ago. Careful manual curation and genomic analysis of meiotic machinery

29   (Supplementary Figure 1, Supplementary Table 1) transcription factors (Supplementary

30   Table 2, Supplementary Material 1), sterols (Supplementary Table 3, Supplementary

31   Material 2, Supplementary Figures 2 and 3), mitochondrial proteins (Supplementary

4

Table 4), cytoskeletal proteins (Supplementary Table 5, Supplementary Figure 4, Supplementary Material 3), membrane trafficking components (Supplementary Table 6), and small GTPases (Supplementary Material 4, Supplementary Figures 5-7, Supplementary Table 7) demonstrates that like *N. gruberi*, *N. fowleri* possesses a remarkably complete repertoire of cellular machinery.

A comparative approach to identifying the cellular basis of N. fowleri *pathogenicity*

As pathogenesis is likely a gain-of-function in *N. fowleri*, 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 7,656 (67%) appear to be shared by all four *Naegleria* species, and 10,451 (92%) are shared by all three *N. fowleri* strains (Figure 2). There are 2,795 groups not identified in *N. gruberi* that are shared by all three *N. fowleri* strains. Manual verification of orthology reduced the number of orthogroups with genes in *N. fowleri* but not *N. gruberi* to 458 proteins (Supplementary Table 8.) Of these, 80% are unique to *N. fowleri,* with no clearly homologous sequence in 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. 404 of the 458 genes have substantial transcriptomic evidence (FPKM >5), suggesting that most of the gene models are accurate and these genes are expressed.

In order to prioritize the candidate pathogenesis factors, we took a transcriptomics approach taking advantage of previously established experimentally-induced pathogenicity in the *N. fowleri* LEE strain (Whiteman and Marciano-Cabral, 1987). In this system, not only does mouse passaged *N. fowleri* have a lower LD_{50} 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 (Whiteman and Marciano-Cabral, 1987).

In our analysis, there are 315 differentially expressed genes in mouse-passaged *N. fowleri* LEE (LEE-MP) versus *N. fowleri* LEE grown only in culture (LEE-Ax) (Supplementary Table 9). Of these, 206 are up-regulated in high pathogenicity *N. fowleri*, while 109 are down-regulated (Supplementary Figure 8), and in terms of function, span multiple cellular systems with potential links to pathogenesis.

1   Overall analysis of down-regulated genes was less informative than that of up-
2   regulated genes. Systems represented in the down-regulated gene dataset include signal
3   transduction, flagellar motility, genes found to be involved in anaerobiasis in bacteria and
4   transcription/translation (Supplementary Table 9). Nonetheless, approximately 70% of
5   the down-regulated genes not in these categories are genes of unknown function, and
6   many are specific to *N. fowleri* or *Naegleria* spp.

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

12

13   *Proteases are leading candidates for pathogenicity factors*

14   *N. fowleri* is known to secrete proteases to traverse the extracellular matrix during
15   infection, and pore-forming proteins to kill and digest host cells (Aldape et al., 1994;
16   Herbst et al., 2002; Hu et al., 1991; Serrano-Luna et al., 2007; Toney and Marciano-
17   Cabral, 1992). Perhaps the most prominent example is a prosaposin pore-forming protein
18   (termed Naegleriapore A), which is heavily glycosylated and protease-resistant (Herbst et
19   al., 2002). Consistent with this, the precursor protein for Naegleriapore A and B was
20   found to be in our up-regulated DE gene set. Similarly, the cathepsin protease (Cathepsin
21   A, or Nf314) is up-regulated in highly pathogenic *N. fowleri*, consistent with it previously
22   being identified as a pathogenicity factor (Hu et al., 1992).

23   Comparative genomics found broadly similar complements of proteases between
24   the *N. fowleri* strains and *N. gruberi* (Supplementary Figure 9, Supplementary Table 10),
25   with one exception. The serine protease S81 was found in all three *N. fowleri* genomes
26   but not in *N. gruberi*. The only homologue of this protein is found in *Hirudo medicinalis*,
27   the European medicinal leech. This protein has a destabilase and peptidoglycan-binding
28   domain; it is found in the saliva of the leech and it may have lysozyme activity and be
29   capable of dissolving fibrin (Zavalova et al., 1996; Zavalova et al., 2000). The gene was
30   highly (but not differentially) expressed under both axenic and mouse-passaged
31   conditions with FPKM values ranging from 500-800.

6

1   Twenty-eight proteases are up-regulated in highly pathogenic *N. fowleri*, making
2   up more than 10% of all up-regulated genes (Supplementary Table 9). Of the protease
3   families with *N. fowleri* homologues up-regulated following mouse passage, half are
4   either localized to lysosomes or are secreted, while the others have proteolytic activities
5   in other organelles or in the cytoplasm. The most substantially represented types of
6   lysosomal/secreted protease in the up-regulated genes are the cathepsin proteases;
7   specifically, the C01 subfamily, with 10 out of 21 genes up-regulated in highly
8   pathogenic *N. fowleri*. The C01 subfamily includes cathepsins B, C, L, Z, and F. Each of
9   these subfamilies has multiple members, up to 10 in the case of cathepsin B, and
10   members of the B, Z, and F subfamilies are up-regulated.

11   Despite the large number of C01 subfamily cathepsin proteases in *N. fowleri* (20-
12   21 members), *N. gruberi* encodes even more (35). Many of the *N. fowleri* and *N. gruberi*
13   cathepsins have 1:1 orthology (Supplementary Figure 10) with at least three expansions
14   that have occurred in the cathepsin B clade in *N. gruberi*. These expansions account for
15   most of the difference in paralogue number between the two species, making up 12 of
16   ~16 *N. gruberi*-specific C01 homologues. Notably, of the *N. fowleri* cathepsin genes that
17   are up-regulated in highly pathogenic *N. fowleri* at least two lack orthologues in *N.*
18   *gruberi*, raising the possibility of their specific involvement in pathogenesis. Protease
19   secretion is underpinned by the membrane trafficking system (the canonical secretion
20   pathway) and autophagy-based unconventional secretion (for proteins lacking N-terminal
21   signal peptides). In both cases, there are few differences in gene presence, absence, and
22   paralogue number in the different *Naegleria* genomes (Supplementary Tables 6 and 11).
23   Strikingly, however, 42% of the up-regulated genes are involved in lysosomal processes.
24   In addition to the 22 proteases above, a lysosomal rRNA degradation gene is up-
25   regulated, as well as three subunits of the vacuolar ATPase proton pump (116 kDa, 21
26   kDa, and 16 kDa) responsible for acidification of both lysosomes and secretory vesicles.
27   Endo-lysosomal trafficking genes Rab GTPase Rab32 (one of three paralogues) and the
28   retromer component Vps35 were up-regulated.

29

30   *Proteins driving actin cytoskeletal rearrangements may contribute to pathogenesis*

1   While actin is known to drive many cellular processes in eukaryotes, Nf-Actin has

2   been considered a pathogenicity factor in *N. fowleri* due to its role in trogocytosis via

3   food cup formation (Sohn et al., 2010). Furthermore, actin binding proteins and upstream

4   regulators of actin polymerization were reported to correlate with virulence (Zysset-Burri

5   et al., 2014). While the complements of actin-associated machinery was similar between

6   the Naegleria species, we notably identified a PTEN domain on one of the formin

7   homologues in *N. fowleri,* that was not identified in *N. gruberi* (Supplementary Material

8   3). Human also do not encode formins of the PTEN family.

9   Although *Naegleria* actin protein levels do not always correlate with transcript

10   levels (Fritz-Laylin et al., 2010b; Sussman et al., 1984), a single subunit of the Arp2/3

11   complex (Arp3) and the WASH complex member strumpellin were both up-regulated in

12   the mouse-passaged amoebae (Supplementary Table 5, Supplementary Material 3).

13   Similarly, we identified an up-regulated RhoGAP22 gene and the serine/threonine protein

14   kinase PAK3, which are involved in Rac1-induced cell migration in other species as well

15   as an up-regulated member of the gelsolin superfamily in the mouse-passaged *N. fowleri*,

16   which may contribute to actin nucleation, capping, or depolymerization (Nag et al.,

17   2013). Although the shift from the environment in the mouse brain to tissue culture

18   conditions   prior   to   sequencing   may   have   resulted   in   an   up-regulation   in

19   macropinocytosis, which can alter cell motility and the transcription of cytoskeletal genes

20   in *Dictyostelium discoideum* (Kayman and Clarke, 1983; Sillo et al., 2008) potential

21   importance of cytoskeletal dynamics in promoting virulence.

22

23   *Neither LGT nor cell stress appear to be major drivers of* N. fowleri *pathogenesis*

24   One obvious potential source of pathogenicity factors is lateral gene transfer of

25   bacterial genes into *N. fowleri* to the exclusion of *N. gruberi* and other eukaryotes.

26   However, of the 458 genes exclusive to *N. fowleri* most are of unknown function

27   (Supplementary Table 8) and of these, only 26 have Bacteria, Archaea or virus as the

28   largest taxonomic group containing the top five BLAST hits. Furthermore, only one of

29   the genes up-regulated in highly pathogenic *N. fowleri* lacks a homologue in *N. gruberi*,

30   and it has a potential homologue in *D. discoideum* and members of the Burkholderiales

31   clade of bacteria (NfowleriV212_g4665, Supplementary Table 9).

1   Another potential reason for *N. fowleri*'s ability to infect humans and animals is

2   the ability to survive the stresses of infection. However, we observed no obvious

3   differences between the *N. fowleri* and *N. gruberi* complements of the ER-associated

4   degradation machinery and unfolded protein response machinery that would suggest a

5   differential ability to cope with cell stress (Supplementary Table 12). Furthermore, our

6   transcriptomics analysis showed a general down-regulation of cell stress systems, as well

7   as DNA damage repair (Supplementary Table 9). This does not suggest that these

8   systems are significantly involved in pathogenesis and our transcriptomics experiment

9   reflects an organism that is not under duress.

10

11   *Adhesion factors as candidate pathogenicity factors*

12   Since infection requires the ability to attach to cells of the nasal epithelium,

13   differences between *N. fowleri* and *N. gruberi* in cell-cell adhesion factors may be

14   relevant to pathogenesis. While we were unable to find evidence of a previously reported

15   integrin-like protein (Jamerson et al., 2012) in any of the genomic data (Supplementary

16   Table 13), we did find that another previously identified attachment protein, Nfa1, was

17   highly expressed in both mouse passaged and axenically cultured *N. fowleri* LEE (>1000

18   FPKM), providing further evidence for its previously reported role as a general

19   pathogenicity factor (Kang et al., 2005).

20   *N. fowleri* and *N. gruberi* encode relatively few putative adhesion G protein-

21   coupled receptors (AGPCRs), with 10 or fewer in each organism (Supplementary Table

22   14). These were identified by searching for proteins with the appropriate domain

23   organization: sequences that have both an extracellular domain (assuming correctly

24   predicted topology in the membrane) and seven transmembrane regions. While their

25   specific functions remain unknown, this work provides a list of proteins that can be later

26   characterized by *in vivo* work.

27   Of the proteins involved in adhesion in *D. discoideum* (TM9/Phg1, SadA, SibA,

28   and SibC), only orthologues of the TM9 protein could be reliably identified

29   (Supplementary Table 15), which was not found to be differentially regulated in our

30   transcriptomic dataset. It is possible that the *Naegleria* homologue may be involved in

31   cell adhesion, but with other downstream effectors. Given the roles for these proteins in

1   both signalling and adhesion in other systems, cell biological work in *Naegleria* is
2   required to confirm their role.

3

4   *The metabolic contribution to pathogenicity*

5   Strikingly, 19% of up-regulated genes are involved in metabolism
6   (Supplementary Table 9). Both catabolic and anabolic processes are represented; some
7   up-regulated genes include phospholipase B-like genes necessary for beta-oxidation, and
8   genes involved in phosphatidate/phosphatidylethanolamine, fatty acid (including long
9   chain fatty acid elongation), and isoprenoid biosynthesis. Phospholipase B was
10 previously identified as pathogenicity factors in *N. fowleri* (Barbour and Marciano-
11 Cabral, 2001). Also identified was a Rieske cholesterol C7(8)-desaturase (Supplementary
12 materials 2, Supplementary Figure 3), a protein involved in sterol production that is
13 absent in mammals, thus potentially representing a drug target.

14   Recent work shows that *N. gruberi* trophozoites prefer to oxidize fatty acids to
15 generate acetyl-CoA, rather than use glucose and amino acids as growth substrates
16 (Bexkens et al., 2018). Several genes involved in metabolism of both lipids and
17 carbohydrates are up-regulated in highly pathogenic *N. fowleri*. Of interest are those that
18 may be involved in metabolizing the polyunsaturated long chain fatty acids that are
19 abundant in the brain, such as long chain fatty acyl CoA synthetase and delta 6 fatty acid
20 (linoleoyl CoA) desaturase-like protein. Consistent with possible shifting carbon source
21 usage or increased growth rates, mitochondrial and energy conversion genes are up-
22 regulated, such as ubiquinone biosynthesis genes, isocitrate dehydrogenase (TCA cycle),
23 complex I and complex III genes (oxidative phosphorylation), and a mitochondrial
24 ADP/ATP translocase. Eight genes involved in amino acid metabolism are also up-
25 regulated.

26   Intriguingly, we identified several areas of *N. fowleri* metabolism that may impact
27 the human host. Glutamate is found in high millimolar concentrations in the brain, and is
28 thought to be the major excitatory neurotransmitter in the central nervous system
29 (Danbolt, 2001; Featherstone and Shippy, 2008). Several genes that function in glutamate
30 metabolism are up-regulated following mouse passage, including kynurenine-
31 oxoglutarate transaminase, glutamate decarboxylase, glutamate dehydrogenase, and

10

1   isocitrate dehydrogenase (Figure 3, Supplementary Table 9). Multiple neurotropic

2   compounds are generated via these enzymes, such as kynurenic acid, GABA, and $NH4^+$.

3   Kynurenic acid in particular has been linked to neuropathological conditions in tick-

4   borne encephalitis (Holtze et al., 2012).

5   Ammonia transporters are also up-regulated (Supplementary Table 9), which may

6   be one way to remove toxic ammonia from the cell, as *N. fowleri*, like *N. gruberi*, has an

7   incomplete urea cycle (Opperdoes et al., 2011). Both glutamate and polyamine

8   metabolism pathways discussed above generate $NH4^+$ as a by-product, and it is possible

9   that this is secreted into the host brain and leads to pathological effects.

10   Finally, agmatine deiminase, which is involved in putrescine biosynthesis, is also

11   up-regulated (Supplementary Table 9). This enzyme catalyzes the conversion of agmatine

12   to carbamoyl putrescine, an upstream precursor to the polyamines glutathione and

13   trypanothione (Colotti and Ilari, 2011). Trypanothione provides a major defense against

14   oxidative stress, some heavy metals, and potentially xenobiotics in trypanosomatid

15   organisms (e.g. *Leishmania*, *Trypanosoma*) (Fairlamb and Cerami, 1992), and has been

16   isolated from *N. fowleri* trophozoites (Ondarza et al., 2006). Although a similar protective

17   role for trypanothione has yet to be confirmed in *N. fowleri*, it is a critical component in

18   trypanosomatid parasites (Krassner and Flory, 1971; Steiger and Steiger, 1977), and

19   enzymes in this pathway may represent novel drug targets.

20

21   *Up-regulated genes with unclear roles in pathogenesis*

22   There are many other genes that are up-regulated, but do not fall into one of the

23   categories outlined above (Supplementary Table 9). Notably, one of these is a

24   transcription factor of RWP-RK family, which has previously only been identified in

25   plants and *D. discoideum,* functioning in plants to regulate responses in nitrogen

26   availability, including differentiation and gametogenesis. While its role in amoebae is

27   unclear, it may represent a potential drug target, as RWP-RK transcription factors are not

28   present in human cells. Notably, of the 208 genes up-regulated in highly pathogenic *N.*

29   *fowleri*, 10 are specific to *N. fowleri* (i.e. not found in other organisms), and 47 are found

30   only in *N. fowleri* and *N. gruberi* (Supplementary Table 9). These represent unique

31   potential targets against which anti-*Naegleria* therapeutics may be developed.

1

2

3   **Discussion**

4   In this study, we provide the first genome-wide assessment of *N. fowleri* strain

5   diversity and the first comprehensive systems-level analysis to understand why this

6   species of *Naegleria* is a highly fatal human pathogen while other species are essentially

7   benign.

8   Our work builds extensively upon previous understanding of *N. fowleri*

9   pathogenicity. Our transcriptomic analysis revealed increased expression of several genes

10   previously considered as pathogenicity factors (e.g., actin, the prosaposin precursor gene

11   of Naegleriapore A and B, phospholipases and Nf314 (Cathepsin A) (Herbst et al., 2002;

12   Hu et al., 1992; Sohn et al., 2010). In 2014, Zysset-Burri and colleagues published a

13   proteomic screen of highly virulent versus weakly virulent *N. fowleri*, as a function of

14   culturing cells with different types of media (Zysset-Burri et al., 2014). While there were

15   clear differences between these strains, potentially due to genetic differences and the

16   method of virulence induction, there were some shared pathways. This included villin

17   and severin, which were both more abundant in high virulence *N. fowleri*, and are

18   involved in actin cytoskeletal dynamics, as well as a phospholipase D homologue.

19   Our work, in combination with that of others, has allowed us to generate a model

20   for pathogenicity in *N. fowleri* (Figure 4), hinging on both specific protein factors as well

21   as whole cellular systems. Several previously identified pathogenicity factors are secreted

22   proteases (e.g. metalloproteases, cysteine proteases, and pore-forming proteases) and

23   phospholipases involved in host tissue destruction, which are likely to be secreted by the

24   cell's membrane trafficking system. Also part of this system – and a major source of

25   differentially expressed genes – is the lysosomal degradation pathway. We identified

26   many up-regulated cathepsin proteases, which function in the lysosome, and predict that

27   they are involved in ingestion and breakdown of host material. Increased cellular

28   ingestion goes hand in hand with cell growth and division, processes which we also see

29   represented in the up-regulated gene dataset: genes involved in protein synthesis,

30   metabolism, and mitochondrial function. This includes several metabolic pathways that

31   produce compounds that could interact with the host immune system or have neurotropic

1   effects. Finally, a major part of *N. fowleri*'s pathogenesis undoubtedly involves cell

2   motility and phagocytosis, which are almost always actin-mediated processes. Because

3   the cytoskeleton is central to *N. fowleri* viability and pathogenesis, and because actin-

4   based processes likely dictate pathogenic behaviors, these analyses provide a solid

5   framework for future investigation into *N. fowleri* virulence as well as potential drug

6   targets.

7   Our comparative genomics and transcriptomics approaches have not only yielded

8   the first systems level model of pathogenicity, but it has also identified many putative

9   novel pathogenicity factors in *N. fowleri* (Figure 5). We identified key individual targets,

10 such as the S81 protease and two cathepsin B proteases, which are both missing from *N.*

11 *gruberi*, and differentially expressed in mouse passaged *N. fowleri*. Moreover, taking a

12 hierarchical approach of overlapping criteria we can distill from high-throughput RNA-

13 Seq data a catalogue of novel potential pathogenicity factors. Over 450 genes (458) are

14 shared by *N. fowleri* strains to the exclusion of *N. gruberi*, while 315 are differentially

15 expressed upon pathogenicity inducing conditions; both are logical criteria for their

16 consideration as potential pathogenicity factors. Annotation as "unknown function" was

17 taken as a criterion for novelty, but not necessarily pathogenicity. At the intersection of

18 these criteria, there are 390 genes of unknown function that are specific to *N. fowleri*, and

19 115 genes that are specific to *N. fowleri* and are up-regulated following mouse-passage.

20 Sixteen genes fulfill all three criteria; they are specific to *N. fowleri*, are up-regulated, and

21 have no putative function. Notably, 90 of the up-regulated genes do not appear to have a

22 human orthologue and represent potential novel drug targets. Genetic tools once

23 developed in this lineage, can be used to functionally characterize the most promising

24 candidate genes and better study *N. fowleri* cell biology, improving our understanding of

25 why and how it is so virulent.

26

27 **Methods Summary:** Genomic and transcriptomic data from three strains of *Naegleria*

28 *fowler*i, including one strain passage axenically and through mice to compare DE genes

29 implicated in pathogenicity, were obtained by Illumina sequencing. Gene models,

30 expressions values, and comparative analyses were performed using a variety of

31 computational tools. See Extended Methods for details (Supplementary Material 5).

13

1

Acknowledgements:

Author Contributions

Generated unpublished primary data included in the manuscript: EKH, AG, HCM, MJM,

DCZ-B, FMC. Provided curated analysis of an encoded genomic system and display

items (includes supervision of trainee): EKH, IM-R, RV, MvdG, MLG, AT, KV, SRN,

ME, CS, MW, LF-L, JBD. Edited or provided intellectual input on the manuscript: EKH,

AG, MvdG, MLG, ADT, KV, SRN, GM, NM, ME, MW, LF-L, GJP, TW, CC, JBD.

Initial conception of the project: FMC, GJP, MWi TW, NM, CC, JBD

Figure Legends

Figure 1. An example of genomic synteny conservation among the three *N. fowleri*

isolates. Because of low conserved synteny between *N. fowleri* and *N. gruberi*, the

organization of homologous sequence in *N. gruberi* is shown for only one *N. fowleri*

contig. Sequence lengths are shown above each genome in millions of bases (M). As the

contig and scaffold alignments depicted are a subset representative of the complete

alignment of the assemblies, the scale bars above each assembly are relative to the

complete alignment. They are included to show the relative sizes of the regions of

homologous sequence. Numbering of scaffolds is arbitrary. Blocks of shared sequence

are colour-coded, while contig or scaffold divisions are shown as grey boxes. Blocks that

are syntenic between strains within the visual space are indicated by lines, while blocks

without lines also have homologous sequence elsewhere in the two other strains, but

14

1   cannot be shown due to the large intervening distance. Offset blocks indicate reverse

2   complemented sequence relative to *N. fowleri* 986.

3

4   **Figure 2.** Result of OrthoMCL analysis showing the number of orthogroups shared

5   between the three *N. fowleri* strains and *N. gruberi*. The number of in-paralogue groups

6   within each species is also shown.

7

8   **Figure 3.** Highly pathogenic *N. fowleri* up-regulates enzymes producing neuroactive

9   chemicals. Up-regulation of enzymes of glutamate metabolism following mouse passage

10   of LEE *N. fowleri* suggests a strategy for ATP production *in vivo* and synthesis of

11   neuroactive metabolites.

12

13   **Figure 4.** Model of *N. fowleri* pathogenicity. Aspects of cellular function that are likely

14   relevant to *N. fowleri* pathogenicity are indicated on the cartoon of high pathogenicity *N.*

15   *fowleri* (right), as compared with low pathogenicity *N. fowleri* (left). This model does not

16   represent an exhaustive list of all identified pathogenicity factors, but rather maps the

17   system-level changes in *N. fowleri* based on the results of our differential gene expression

18   analysis.

19

20   **Figure 5.** Venn diagram showing the overlap between differentially expressed genes in

21   highly pathogenic *N. fowleri*, those which have no clear homologues in *N. gruberi*, and

22   those of unknown function.

23

24   **Supplementary Figure 1.** Dot plot showing the presence of meiosis genes in *Naegleria*

25   sp. Meiosis gene in *N. gruberi* are based on Fritz-Laylin *et al.*, 2010. A filled circle

26   indicates the presence of a gene, while an unfilled circle indicates that no clear

27   homologue could be identified. Numbers within the circle indicate multiple paralogues.

28   Like *N. gruberi*, *N. fowleri* encodes a highly complete complement of meiotic genes.

29

30   **Supplementary Figure 2.** Gene repertoire of the sterol synthesis pathway in *N. fowleri*.

31   For comparison, the distribution of orthologues from *S. cerevisiae*, *H. sapiens* and *A.*

15

1   *thaliana*, as well as the non-pathogenic *N. gruberi*, are shown. Double-dots indicate two

2   orthologues performing one step (i.e. SMO1 and SMO2) or two orthologues performing

3   similar reactions (i.e. SMT1 and SMT2). In the case of non-homologous families

4   ERG2/EBP/HYD1 and ERG4/DHCR24/DWF1, the presence of one or the other copy is

5   indicated by orange or blue dots as indicated in the gene names, whereas the presence of

6   both copies is indicated by green dots. The Rieske cholesterol C7(8)-desaturase is shown

7   apart because is not involved in the canonical sterol pathways.

8

9   **Supplementary Figure 3.** Partial alignment of cholesterol C7(8)-desaturases. Red boxes

10   highlight the Rieske [2Fe-2S] motif (CXHX16CX2H) and the non-heme iron binding

11   motif ((D/E)X3DX2HX4H). Asterisks indicate consensus residues important for

12   catalysis.

13

14   **Supplementary Figure 4.** The *N. fowleri* genome encodes an extensive actin

15   cytoskeletal repertoire. Formin family proteins typically nucleate and elongate actin

16   filaments, moving progressively with the barbed end and recruiting profilin-bound actin

17   monomers to the growing end of the filament. Another nucleator, the Arp2/3 complex,

18   typically polymerizes a new filament from the side of a pre-existing filament following

19   activation by a WASP-family protein such as WASP, SCAR/WAVE, and/or WASH. *N.*

20   *fowleri* also encodes myosin motor proteins, including myosin I and II. Finally, *N. fowleri*

21   encodes cofilin and members of the gelsolin/villin superfamily predicted to depolymerize

22   actin filaments. In all, the *N. fowleri* genome encodes at least 22 actins, 14 formins, 11

23   myosins, 4 gelsolin/villin superfamily proteins, and 3 profilins, in addition to all the

24   subunits of the Arp2/3 complex, and 3 WASP family proteins and their respective

25   complexes.

26

27   **Supplementary Figure 5.** Phylogenetic analysis of Ras family genes in *Naegleria*

28   species and other selected eukaryotes. Portrayed is a maximum likelihood tree (RAxML,

29   LG+Γ model). Bootstrap support values were calculated using the rapid bootstrapping

30   (Rboot) algorithm of the RAxML program. The robustness of the tree topology was also

31   assessed by the IQ-tree with LG+F+G4 model (the model selected by the program itself)

1 with the ultrafast bootstrap (UFboot) algorithm (1000 replicates) and the SH-aLRT test

2 (1000 replicates). Circles at branches correspond to bootstrap values indicated in the

3 legend. The bar on the top corresponds to the estimated number of substitutions per site.

4 The identity of the *N. fowleri* genes is provided in Supplementary Table 7, sheet 2.

5 Manually corrected gene models *; Newly created gene models**.

6

7 **Supplementary Figure 6.** Phylogenetic analysis of Rab family genes in *Naegleria*

8 species and other eukaryotes from Elias et al., 2012 (Elias et al., 2012). Portrayed is a

9 maximum likelihood tree (RAxML, LG+Γ model). Bootstrap support values were

10 calculated using the rapid bootstrapping (Rboot) algorithm of the RAxML program. The

11 robustness of the tree topology was assessed also by the IQ-tree with LG+G4 model (the

12 model selected by the program itself) with the ultrafast bootstrap (UFboot) algorithm

13 (1000 replicates) and the SH-aLRT test (1000 replicates). Circles at branches correspond

14 to bootstrap values indicated in the legend. The bar on the top corresponds to the

15 estimated number of substitutions per site. The identity of the *N. fowleri* genes is

16 provided in Supplementary Table 7, sheet 2. Manually corrected gene models *; Newly

17 created gene models**.

18

19 **Supplementary Figure 7.** Multidomain proteins with a Ras superfamily GTPase domain

20 found in *Naegleria* species. Selected proteins containing more than one domain are

21 schematically depicted here. The positions of domains are according to NCBI's conserved

22 domain database, except the C-terminal part of Gpa18 protein which was identified by

23 manual inspection of the multiple sequence alignment of *Naegleria spp.* Gα proteins, as

24 this fragment was too small to be detected by NCBI's conserved domain search. Ras

25 superfamily domains are depicted in orange with a label specifying a subgroup of the Ras

26 superfamily. ZnF UBP = Ubiquitin carboxyl-terminal hydrolase-like zinc finger; ANKs =

27 multiple ankyrin repeats; G – alpha (Gα) = G protein alpha subunit; STKc =

28 serine/threonine protein kinases. Nfo: *N. fowleri*; Ngr: *N. gruberi*; *: manually corrected

29 gene model. The scale bar shows the length of 100 amino acid residues.

30

17

1 **Supplementary Figure 8.** MA plot and Volcano plot of genes differentially expressed in
2 highly pathogenic *N. fowleri* LEE strain. MA plots show the log fold change of each gene
3 relative to the log of the mapped read counts, while volcano plots scale the False
4 Discovery Rate (FDR) to log fold change. Each gene is represented by a dot, and red dots
5 indicate those which meet the differential expression criteria.

6

7 **Supplementary Figure 9.** Protease gene abundance in *Naegleria* per MEROPS protease
8 family. For each MEROPS protease family, the number of predicted proteins in each
9 strain is shown. The S81 family (boxed) is the only family without a homologue in *N.*
10 *gruberi*. The only other known S81 family protease is a destabilase protein in *Hirudo*
11 *medicinalis*, the European medicinal leech. Otherwise, most protease families could be
12 identified in all four genomes, with similar numbers of homologues. The exception is the
13 C01 cysteine protease family, which is analyzed in Supplementary Figure 10.

14

15 **Supplementary Figure 10.** Phylogenetic analysis of the C01 cysteine protease subfamily
16 in *N. fowleri* and *N. gruberi*. Node values are listed as Phylobayes/RAxML (posterior
17 probability/bootstrap), and as symbols indicating a minimum level of support as shown in
18 the inset. Node values are shown on the best Bayesian topology. Sequences with signal
19 peptides have red text, those with potential signal peptides (score near cutoff) have purple
20 text, and those without identifiable signal peptides have blue text. Asterisks (*) indicate
21 genes that are up-regulated in highly pathogenic *N. fowleri*.

22

23

24 **References**

25 **Aldape, K., Huizinga, H., Bouvier, J. and McKerrow, J.** (1994). Naegleria fowleri:
26   Characterization of a secreted histolytic cysteine protease. *Exp. Pathol.* **78**, 230–241.
27 **Baral, R. and Vaidya, B.** (2018). Fatal case of amoebic encephalitis masquerading as
28   herpes. *Oxford Med. Case Reports* **2018**, 134–137.
29 **Barbour, S. E. and Marciano-Cabral, F.** (2001). Naegleria fowleri amoebae express a
30   membrane-associated calcium-independent phospholipase A2. *Biochim. Biophys.*
31   *Acta - Mol. Cell Biol. Lipids* **1530**, 123–133.

1   **Baverstock, P. R., Illana, S., Christy, P. E., Robinson, B. S. and Johnson, A. M.**
2   (1989). srRNA evolution and phylogenetic relationships of the genus Naegleria
3   (Protista: Rhizopoda). *Mol. Biol. Evol.* **6**, 243–257.

4   **Bexkens, M. L., Zimorski, V., Sarink, M. J., Wienk, H., Brouwers, J. F., De**
5   **Jonckheere, J. F., Martin, W. F., Opperdoes, F. R., van Hellemond, J. J. and**
6   **Tielens, A. G. M.** (2018). Lipids Are the Preferred Substrate of the Protist Naegleria
7   gruberi, Relative of a Human Brain Pathogen. *Cell Rep.* **25**, 537–543.e3.

8   **Carter, R. F.** (1970). Description of a Naegleria sp. isolated from two cases of primary
9   amoebic meningo⬚encephalitis, and of the experimental pathological changes
10   induced by it. *J. Pathol.* **100**, 217–244.

11 **Colotti, G. and Ilari, A.** (2011). Polyamine metabolism in Leishmania: From arginine to
12   trypanothione. *Amino Acids* **40**, 269–285.

13 **Cope, J. R., Ratard, R. C., Hill, V. R., Sokol, T., Causey, J. J., Yoder, J. S., Mirani,**
14   **G., Mull, B., Mukerjee, K. A., Narayanan, J., et al.** (2015). The first association
15   of a primary amebic meningoencephalitis death with culturable naegleria fowleri in
16   tap water from a US treated public drinking water system. *Clin. Infect. Dis.* **60**, e36–
17   e42.

18 **Cope, J. R., Conrad, D. A., Cohen, N., Cotilla, M., Dasilva, A., Jackson, J. and**
19   **Visvesvara, G. S.** (2016). Use of the Novel Therapeutic Agent Miltefosine for the
20   Treatment of Primary Amebic Meningoencephalitis: Report of 1 Fatal and 1
21   Surviving Case. *Clin. Infect. Dis.* **62**, 774–776.

22 **Danbolt, N. C.** (2001). Glutamate uptake. *Prog. Neurobiol.* **65**, 1–105.

23 **Dorsch, M. M.** (1982). *Primary amoebic meningoencephalitis: an historical and*
24   *epidemiological perspecitve with particular reference to South Australia*. Adelaide:
25   Epidemiology Branch, South Australian Health Commission.

26 **Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M. C. and Dacks, J. B.** (2012).
27   Sculpting the endomembrane system in deep time: high resolution phylogenetics of
28   Rab GTPases. *J. Cell Sci.* **125**, 2500–8.

29 **Fairlamb, A. H. and Cerami, A.** (1992). Metabolism and Functions of Trypanothione in
30   the Kinetoplastea. *Annu. Rev. Microbiol.* **46**, 695–729.

31 **Featherstone, D. E. and Shippy, S. A.** (2008). Regulation of synaptic transmission by

ambient extracellular glutamate. *Neuroscientist* **14**, 171–181.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredez, A., Chapman, J., Pham, J., et al. (2010a). The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. *Cell* **140**, 631–642.

Fritz-Laylin, L. K., Assaf, Z. J., Chen, S. and Cande, W. Z. (2010b). Naegleria gruberi de novo basal body assembly occurs via stepwise incorporation of conserved proteins. *Eukaryot. Cell* **9**, 860–865.

Herbst, R., Ott, C., Jacobs, T., Marti, T., Marciano-Cabral, F. and Leippe, M. (2002). Pore-forming polypeptides of the pathogenic protozoon Naegleria fowleri. *J. Biol. Chem.* **277**, 22353–22360.

Holtze, M., Mickiené, A., Atlas, A., Lindquist, L. and Schwieler, L. (2012). Elevated cerebrospinal fluid kynurenic acid levels in patients with tick-borne encephalitis. *J. Intern. Med.* **272**, 394–401.

Hu, W. N., Band, R. N. and Kopachik, W. J. (1991). Virulence-related protein synthesis in Naegleria fowleri. *Infect. Immun.* **59**, 4278–4282.

Hu, W. N., Kopachik, W. and Band, R. N. (1992). Cloning and characterization of transcripts showing virulence-related gene expression in Naegleria fowleri. *Infect. Immun.* **60**, 2418–2424.

Jamerson, M., da Rocha-Azevedo, B., Cabral, G. A. and Marciano-Cabral, F. (2012). Pathogenic Naegleria fowleri and non-pathogenic Naegleria lovaniensis exhibit differential adhesion to, and invasion of, extracellular matrix proteins. *Microbiology* **158**, 791–803.

Kang, S.-Y., Song, K.-J., Jeong, S.-R., Kim, J.-H., Park, S., Kim, K., Kwon, M.-H. and Shin, H.-J. (2005). Role of the Nfa1 Protein in Pathogenic Naegleria fowleri Cocultured with CHO Target Cells. *Clin. Vaccine Immunol.* **12**, 873–876.

Kayman, S. C. and Clarke, M. (1983). Relationship between axenic growth of Dictyostelium discoideum strains and their track morphology on substrates coated with gold particles. *J. Cell Biol.* **97**, 1001–1010.

Kazi, A. N. and Riaz, T. (2013). Deaths from rare protozoan encephalitis in Karachi blamed on unchlorinated water. *BMJ* **346**, 4461.
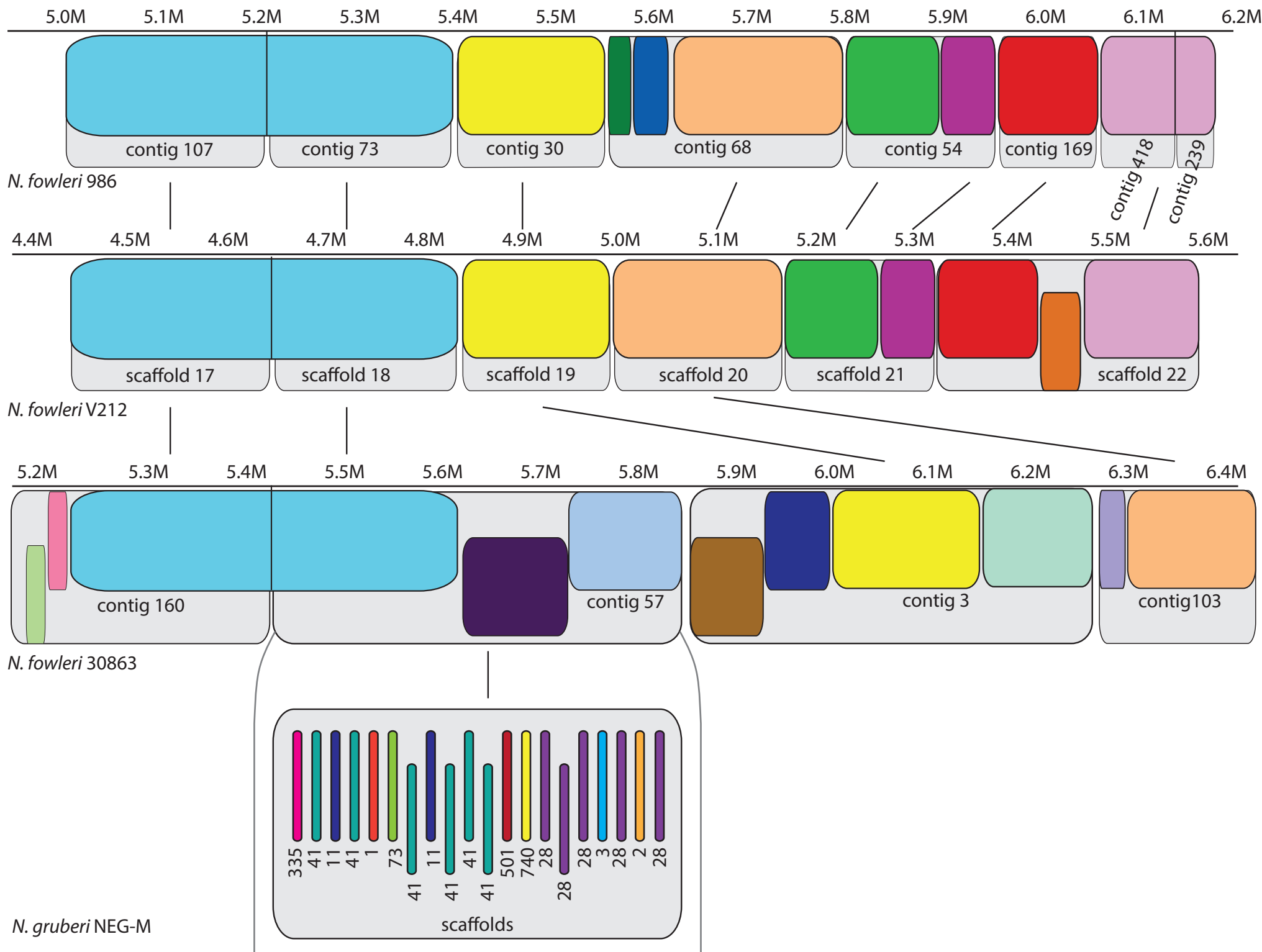
1 **Kemble, S. K., Lynfield, R., DeVries, A. S., Drehner, D. M., Pomputius, W. F.,**
2 **Beach, M. J., Visvesvara, G. S., da Silva, A. J., Hill, V. R., Yoder, J. S., et al.**
3 (2012). Fatal Naegleria fowleri Infection Acquired in Minnesota: Possible Expanded
4 Range of a Deadly Thermophilic Organism. *Clin. Infect. Dis.* **54**, 805–809.

5 **Krassner, S. M. and Flory, B.** (1971). Essential Amino Acids in the Culture of
6 Leishmania tarentolae. *J. Parasitol.* **57**, 917–920.

7 **Linam, W. M., Ahmed, M., Cope, J. R., Chu, C., Visvesvara, G. S., Da Silva, A. J.,**
8 **Qvarnstrom, Y. and Green, J.** (2015). Successful treatment of an adolescent with
9 naegleria fowleri primary amebic meningoencephalitis. *Pediatrics* **135**, e744–e748.

10 **Maciver, S. K., Piñero, J. E. and Lorenzo-Morales, J.** (2019). Is Naegleria fowleri an
11 Emerging Parasite? *Trends Parasitol.* 1–10.

12 **Mahmood, K.** (2015). Naegleria fowleri in Pakistan - An emerging catastrophe. *J. Coll.*
13 *Physicians Surg. Pakistan* **25**, 159–160.

14 **Morgan, M. J., Halstrom, S., Wylie, J. T., Walsh, T., Kaksonen, A. H., Sutton, D.,**
15 **Braun, K. and Puzon, G. J.** (2016). Characterization of a Drinking Water
16 Distribution Pipeline Terminally Colonized by Naegleria fowleri. *Environ. Sci.*
17 *Technol.* **50**, 2890–2898.

18 **Nag, S., Larsson, M., Robinson, R. C. and Burtnick, L. D.** (2013). Gelsolin: The tail
19 of a molecular gymnast. *Cytoskeleton* **70**, 360–384.

20 **Naqvi, A. A., Yazdani, N., Ahmad, R., Zehra, F. and Ahmad, N.** (2016).
21 Epidemiology of primary amoebic meningoencephalitis-related deaths due to
22 Naegleria fowleri infections from freshwater in Pakistan: An analysis of 8-year
23 dataset. *Arch. Pharm. Pract.* **7**, 119–129.

24 **Ondarza, R. N., Hurtado, G., Tamayo, E., Iturbe, A. and Hernández, E.** (2006).
25 Naegleria fowleri: A free-living highly pathogenic amoeba contains
26 trypanothione/trypanothione reductase and glutathione/glutathione reductase
27 systems. *Exp. Parasitol.* **114**, 141–146.

28 **Opperdoes, F. R., De Jonckheere, J. F. and Tielens, A. G. M.** (2011). Naegleria
29 gruberi metabolism. *Int. J. Parasitol.* **41**, 915–924.

30 **Puzon, G. J., Lancaster, J. A., Wylie, J. T. and Plumb, J. J.** (2009). Rapid detection of
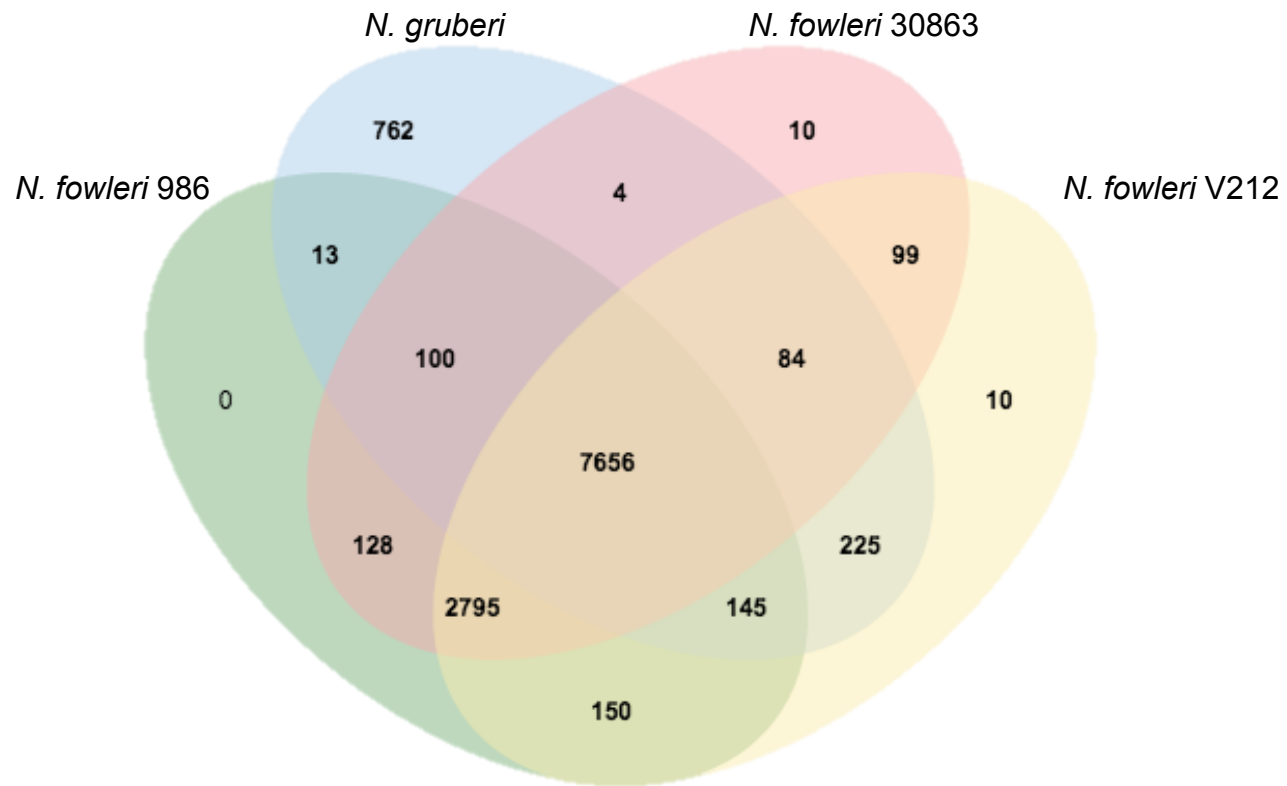31 Naegleria fowleri in water distribution pipeline biofilms and drinking water samples.

1   *Environ. Sci. Technol.* **43**, 6691–6696.

2   **Puzon, G. J., Wylie, J. T., Walsh, T., Braun, K. and Morgan, M. J.** (2017).

3   Comparison of biofilm ecology supporting growth of individual Naegleria species in

4   a drinking water distribution system. *FEMS Microbiol. Ecol.* **93**, 1–8.

5   **Serrano-Luna, J., Cervantes-Sandoval, I., Tsutsumi, V. and Shibayama, M.** (2007).

6   A Biochemical Comparison of Proteases from Pathogenic Naegleria fowleri and

7   Non-Pathogenic Naegleria gruberi. *J. Eukaryot. Microbiol.* **54**, 411–417.

8   **Siddiqui, R. and Khan, N. A.** (2014). Primary Amoebic Meningoencephalitis Caused by

9   Naegleria fowleri: An Old Enemy Presenting New Challenges. *PLoS Negl. Trop.*

10   *Dis.* **8**,.

11   **Sillo, A., Bloomfield, G., Balest, A., Balbo, A., Pergolizzi, B., Peracino, B., Skelton,**

12   **J., Ivens, A. and Bozzaro, S.** (2008). Genome-wide transcriptional changes induced

13   by phagocytosis or growth on bacteria in Dictyostelium. *BMC Genomics* **9**, 1–22.

14   **Sohn, H. J., Kim, J. H., Shin, M. H., Song, K. J. and Shin, H. J.** (2010). The Nf-actin

15   gene is an important factor for food-cup formation and cytotoxicity of pathogenic

16   Naegleria fowleri. *Parasitol Res* **106**, 917–924.

17   **Steiger, R. F. and Steiger, E.** (1977). Cultivation of Leishmania donovani and

18   Leishmania braziliensis in Defined Media: Nutritional Requirements. *J. Protozool.*

19   **24**, 437–441.

20   **Sussman, D. J., Lai, E. Y. and Fulton, C.** (1984). Rapid disappearance of translatable

21   actin mRNA during cell differentiation in Naegleria. *J. Biol. Chem.* **259**, 7355–7360.

22   **Toney, D. M. and Marciano-Cabral, F.** (1992). Alterations in protein expression and

23   complement resistance of pathogenic Naegleria amoebae. *Infect. Immun.* **60**, 2784–

24   2790.

25   **Trabelsi, H., Dendana, F., Sellami, A., Sellami, H., Cheikhrouhou, F., Neji, S.,**

26   **Makni, F. and Ayadi, A.** (2012). Pathogenic free-living amoebae: Epidemiology

27   and clinical review. *Pathol. Biol.* **60**, 399–405.

28   **Whiteman, L. Y. and Marciano-Cabral, F.** (1987). Susceptibility of pathogenic and

29   nonpathogenic Naegleria spp. to complement-mediated lysis. *Infect. Immun.* **55**,

30   2442–2447.

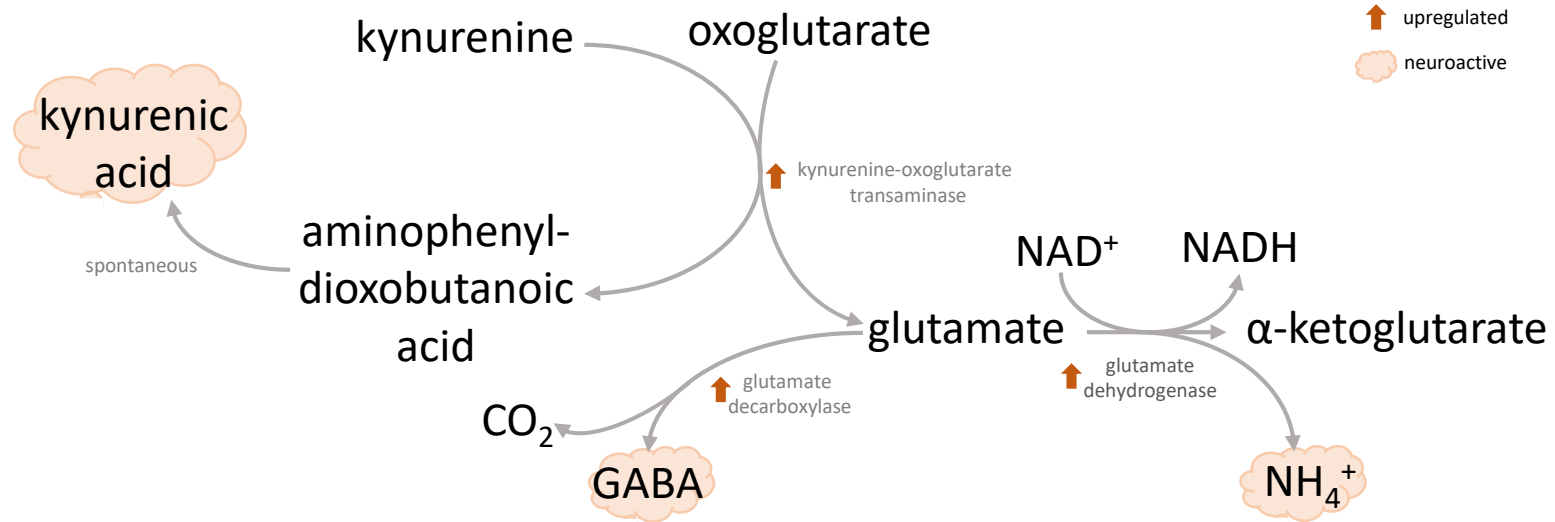31   **Yoder, J. S., Straif-Bourgeois, S., Roy, S. L., Moore, T. A., Visvesvara, G. S.,**

1   **Ratard, R. C., Hill, V. R., Wilson, J. D., Linscott, A. J., Crager, R., et al.** (2012).

2   Primary amebic meningoencephalitis deaths associated with sinus irrigation using

3   contaminated tap water. *Clin. Infect. Dis.* **55**, 79–85.

4   **Zavalova, L., Lukyanov, S., Baskova, I., Snezhkov, E., Akopov, S., Berezhnoy, S.,**

5   **Bogdanova, E., Barsova, E. and Sverdlov, E. D.** (1996). Genes from the medicinal

6   leech (Hirudo medicinalis) coding for unusual enzymes that specifically cleave

7   endo-ε(γ-Glu)-Lys isopeptide bonds and help to dissolve blood clots. *Mol. Gen.*

8   *Genet.* **253**, 20–25.

9   **Zavalova, L. L., Baskova, I. P., Lukyanov, S. A., Sass, A. V., Snezhkov, E. V.,**

10   **Akopov, S. B., Artamonova, I. I., Archipova, V. S., Nesmeyanov, V. A., Kozlov,**

11   **D. G., et al.** (2000). Destabilase from the medicinal leech is a representative of a

12   novel family of lysozymes. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*

13   **1478**, 69–77.

14   **Zysset-Burri, D. C., Müller, N., Beuret, C., Heller, M., Schürch, N., Gottstein, B.**

15   **and Wittwer, M.** (2014). Genome-wide identification of pathogenicity factors of

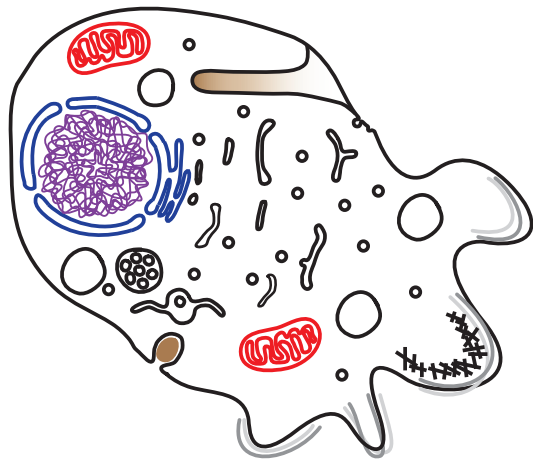16   the free-living amoeba Naegleria fowleri. *BMC Genomics* **15**, 496.

17

Table 1. Genome statistics for *N. fowleri* strains V212, 986, and ATCC 30863, and *N. gruberi* strain NEG-M.

Species	Total genome size	GC content	Number of genes	Average gene length	Exons/gene	Average exon length	% Coding	Average intron length
N. fowleri V212	27.7 Mbp	36%	12,677	1,785 bp	2	777 bp	71.35%	126 bp
N. fowleri 986	27.5 Mbp	36%	11,599	1,955 bp	2	849 bp	73.01%	138 bp
N. fowleri ATCC 30863	29.62 Mbp	35%	11,499	1,984 bp	2	825 bp	70.79%	144 bp
N. gruberi NEG-M	41.0 Mbp	33%	15,708	1,677 bp	1.7	894 bp	57.8%	203 bp

5.0M 5.1M 5.2M 5.3M 5.4M 5.5M 5.6M 5.7M 5.8M 5.9M 6.0M 6.1M 6.2M

contig 107 contig 73 contig 30 contig 68 contig 54 contig 169 contig 418 contig 239

N. fowleri 986

4.4M 4.5M 4.6M 4.7M 4.8M 4.9M 5.0M 5.1M 5.2M 5.3M 5.4M 5.5M 5.6M

scaffold 17 scaffold 18 scaffold 19 scaffold 20 scaffold 21 scaffold 22

N. fowleri V212

5.2M 5.3M 5.4M 5.5M 5.6M 5.7M 5.8M 5.9M 6.0M 6.1M 6.2M 6.3M 6.4M

contig 160 contig 57 contig 3 contig103

N. fowleri 30863

335 41 11 41 1 73 41 11 41 41 501 740 28 28 3 28 2 28

scaffolds

N. gruberi NEG-M

Mitochondria
Metabolism

Cytoplasm

Protein synthesis
Metabolism

Secretion

Proteases
Phospholipases
Pore-forming proteins

Lysosome

Proteases
Phospholipases

Cytoskeleton
Actin-based
amoeboid
movement

Nucleus
Cell growth &
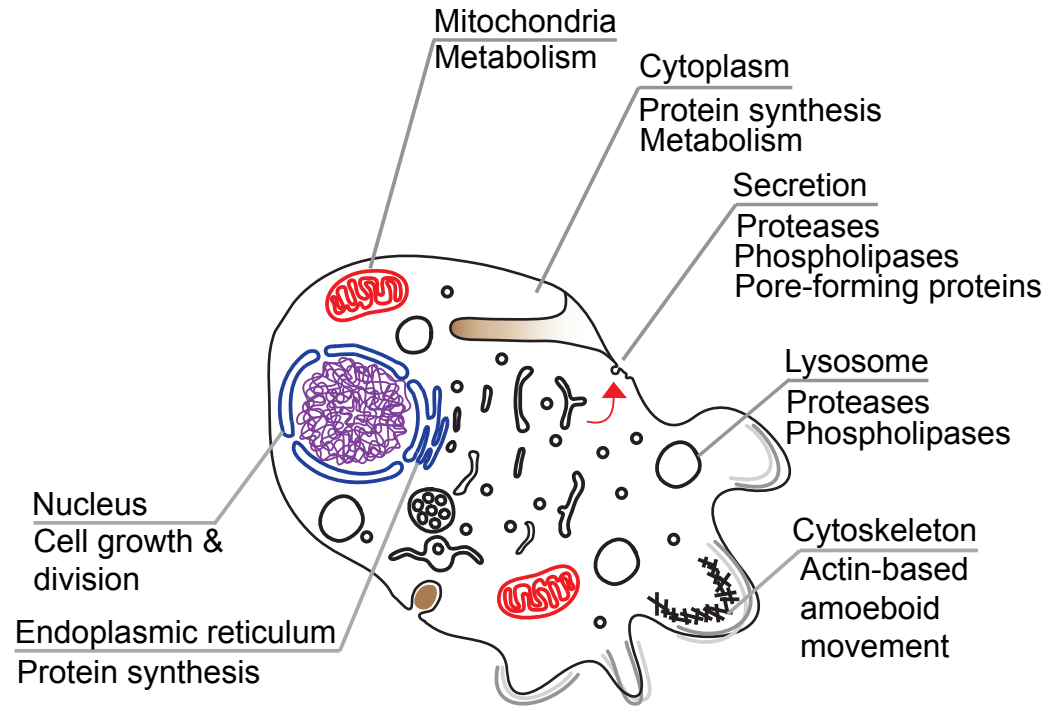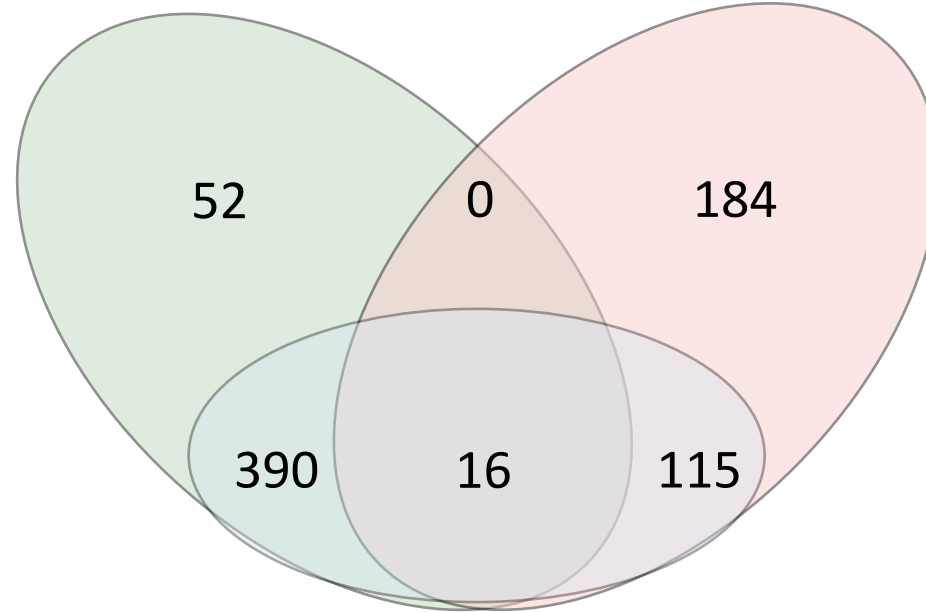division

Endoplasmic reticulum
Protein synthesis

Low pathogenicity *N. fowleri*

High pathogenicity *N. fowleri*

Genes present in *N. fowleri,* but not *N. gruberi* (458 total)

Differentially expressed genes in *N. fowleri* LEE (315 total)

52

0

184

390

16

115

Genes of unknown function