



OPEN ACCESS

EDITED BY

Ascel Samba-Louaka,
University of Poitiers, France

REVIEWED BY

Bharath Kanakapura Sundararaj,
Boston University, United States
Julia Walochnik,
Medical University of Vienna, Austria

*CORRESPONDENCE

Stephen Kwok-Wing Tsui

✉ kwtsui@cuhk.edu.hk

Patsharaporn T. Sarasombath

✉ p.techasintana@gmail.com;

✉ patsharaporn.tec@mahidol.ac.th

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 10 February 2023

ACCEPTED 06 April 2023

PUBLISHED 05 May 2023

CITATION

Law CT-Y, Nivesvivat T, Xiong Q, Kulkeaw K, Shi L, Ruenchit P, Suwanpakdee D, Suwanpakdee P, Tongkrajang N, Sarasombath PT and Tsui SK-W (2023) Mitochondrial genome diversity of *Balamuthia mandrillaris* revealed by a fatal case of granulomatous amoebic encephalitis. *Front. Microbiol.* 14:1162963. doi: 10.3389/fmicb.2023.1162963

COPYRIGHT

© 2023 Law, Nivesvivat, Xiong, Kulkeaw, Shi, Ruenchit, Suwanpakdee, Suwanpakdee, Tongkrajang, Sarasombath and Tsui. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mitochondrial genome diversity of *Balamuthia mandrillaris* revealed by a fatal case of granulomatous amoebic encephalitis

Cherie Tsz-Yiu Law^{1,2†}, Thirapa Nivesvivat^{3†}, Qing Xiong^{1,2†}, Kasem Kulkeaw^{4†}, Ling Shi^{1,2}, Pichet Ruenchit⁴, Detchvijitr Suwanpakdee³, Piradee Suwanpakdee⁵, Nongnat Tongkrajang⁴, Patsharaporn T. Sarasombath^{4*} and Stephen Kwok-Wing Tsui^{1,2*}

¹School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China,

²Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Shatin, Hong Kong SAR,

China, ³Infectious Disease Unit, Department of Pediatrics, Phramongkutklao Hospital, Bangkok,

Thailand, ⁴Siriraj Integrative Center for Neglected Parasitic Diseases, Department of Parasitology,

Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, ⁵Neurology Division,

Department of Pediatrics, Phramongkutklao Hospital, Bangkok, Thailand

Introduction: *Balamuthia (B.) mandrillaris* is a free-living amoeba that can cause rare yet fatal granulomatous amoebic encephalitis (GAE). However, efficacious treatment for GAE is currently unavailable, especially when genomic studies on *B. mandrillaris* are limited.

Methods: In this study, *B. mandrillaris* strain KM-20 was isolated from the brain tissue of a GAE patient, and its mitochondrial genome was *de novo* assembled using high-coverage Nanopore long reads and Illumina short reads.

Results and Discussion: Phylogenetic and comparative analyses revealed a range of diversification in the mitochondrial genome of KM-20 and nine other *B. mandrillaris* strains. According to the mitochondrial genome alignment, one of the most variable regions was observed in the ribosomal protein S3 (*rps3*), which was caused by an array of novel protein tandem repeats. The repeating units in the *rps3* protein tandem region present significant copy number variations (CNVs) among *B. mandrillaris* strains and suggest KM-20 as the most divergent strain for its highly variable sequence and highest copy number in *rps3*. Moreover, mitochondrial heteroplasmy was observed in strain V039, and two genotypes of *rps3* are caused by the CNVs in the tandem repeats. Taken together, the copy number and sequence variations of the protein tandem repeats enable *rps3* to be a perfect target for clinical genotyping assay for *B. mandrillaris*. The mitochondrial genome diversity of *B. mandrillaris* paves the way to investigate the phylogeny and diversification of pathogenic amoebae.

KEYWORDS

granulomatous amoebic encephalitis, *Balamuthia mandrillaris*, mitochondrial genome, free-living amoeba, ribosomal protein S3, neglected diseases, genotyping

1. Background

Balamuthia mandrillaris is one of the free-living amoeba species that can cause brain infections in humans besides *Acanthamoeba* spp., *Naegleria fowleri*, and *Sappinia* spp. (Visvesvara et al., 2007; Visvesvara, 2013, 2014). *B. mandrillaris* can enter the human body through the skin via wound or inhalation, and the infection causes a cutaneous lesion or brain infection called GAE, with up to 95% mortality rate (Matin et al., 2008; Siddiqui and Khan, 2008; Visvesvara, 2013, 2014; Vollmer and Glaser, 2016). This species was first identified from the brain of a pregnant mandrill baboon at the San Diego Wildlife Park in 1986, and the first human brain infection was reported in 1990 (Visvesvara et al., 1990). Since then, more than 200 cases of *Balamuthia* encephalitis have been reported worldwide (Intalapaporn et al., 2004; Cope et al., 2019; Wang et al., 2020). Due to the rarity and non-specific presentations of GAE, the diagnosis is usually delayed and often made postmortem (Schuster et al., 2009).

Unlike other free-living amoebae, which can be cultured on agar overlaid with bacteria, this organism needs to be grown on mammalian cell monolayer cultures such as monkey kidney cells, human lung fibroblasts, and human neuroblastoma cells (Schuster, 2002). Thus, the diagnosis of *B. mandrillaris* infection mainly relies on other laboratory methods including serology and genotyping based on the molecular marker in mitochondrial DNA (da Rocha-Azevedo et al., 2009). Currently, only two complete genomes of *B. mandrillaris* are available, both of which were obtained from patients in the United States (Detering et al., 2015; Greninger et al., 2015). The genome sizes of the two assemblies are 44.3 and 67.6 Mbp, respectively (Detering et al., 2015; Greninger et al., 2015). A total of nine mitochondrial genomes of *B. mandrillaris* are available, but none of them were isolated from Asia (Detering et al., 2015; Greninger et al., 2015). A previous study compared the mitochondrial genome of seven *B. mandrillaris* strains isolated from patients and the environment in the United States (Greninger et al., 2015). In addition to the various lengths of mitochondrial DNA sequences, the phylogenetic tree shows three distinct lineages (Greninger et al., 2015).

Mitochondrial genomes are increasingly used for phylogenetic and epidemiological analyses. In addition, several antiprotozoal drugs including pentamidine exert their functions by interfering with mitochondrial metabolism (de Souza et al., 2009). Mitochondrial genome analysis of *B. mandrillaris* may provide further insights into the diversity within species and shed light on the functions of mitochondrial genes, which could serve as potential drug targets.

Using both Nanopore long-read and Illumina short-read sequencing data, we *de novo* assembled the mitochondrial genome of the *B. mandrillaris* strain isolated from Asia named strain KM-20. Phylogenetic and comparative analyses of KM-20 and nine other strains were performed to investigate the mitochondrial genome diversity among strains. Notably, a previous study has reported the difference in the mitochondrial *rps3* gene, but the authors suggested that the difference is due to a putative intron or intergenic region (Greninger et al., 2015). In this study, our results demonstrated that the diversity of the *rps3* length is attributed to an array of protein tandem repeats, and the number of repeating units is different among *B. mandrillaris* strains.

2. Materials and methods

2.1. *Balamuthia mandrillaris* culture and DNA extraction

Balamuthia mandrillaris strain KM-20 was obtained by inoculating the left frontoparietal brain tissue of the GAE patient reported here in a monolayer of human lung carcinoma A549 cells in Dulbecco's modified Eagle medium plus 10% fetal bovine serum at 37°C with 5% CO₂, following the protocol as previously described (Schuster et al., 2009). The amoeba was first observed in the culture 4 weeks after inoculation and was maintained in a culture with A549 cells at 37°C with 5% CO₂ (Schuster, 2002). *B. mandrillaris* strains V039 (50209) and V416 (PRA-290) were obtained from ATCC (Manassas, VA, United States) and maintained in culture media containing human neuroblastoma SH-SY5Y cells at 37°C with 5% CO₂. DNA extractions of the amoeba were performed using a QIAmp DNA Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol for isolating DNA from cell cultures.

2.2. DNA sequencing, assembly, and annotation

Balamuthia mandrillaris KM-20 genomic DNA was sequenced using Oxford Nanopore GridION Mk1 with a Ligation Sequencing Kit (SQK-LSK109) on an R9.4.1 MinION flow cell and an Illumina NovaSeq 6,000 sequencing system. The mitochondrial genome of *B. mandrillaris* strains 2046, V039, BeN, GAM-19, OK1, RP5, SAM, V188, and V451 (Table 1) was downloaded and used as a reference to map against the raw reads of KM-20 using Minimap2 (v2.20) to identify the mitochondrial DNA sequences of KM-20 (Li, 2018). The identified sequences were assembled into a single contig using Flye (v2.8.3) and subsequently polished by Illumina data using Pilon (v1.24) (Walker et al., 2014; Kolmogorov et al., 2019). The mitochondrial genome of KM-20 was visualized in Proksee (Stothard and Wishart, 2005; Grant and Stothard, 2008). The coding genes, introns, and novel open reading frames were identified by MITOS WebServer and GeSeq (v2.03) (Bernt et al., 2013; Tillich et al., 2017). The transfer RNA (tRNA) annotation was performed by GeSeq with ARAGORN (v1.2.38), and the rRNA subunit genes were checked by RNAses (Laslett and Canback, 2004; Lang et al., 2007). The coverage of the mitochondrial genome of *B. mandrillaris* KM-20 was obtained by mapping the nanopore raw reads to the assembled mitochondrial genome by Minimap2 (v2.20), and the mapping coverage was obtained using SAMtools (v1.5) (Li, 2018; Danecek et al., 2021). The *B. mandrillaris* KM-20 mitochondrial genome has been deposited in the National Center for Biotechnology Information (NCBI) under the accession number OM994889.

2.3. Comparative mitochondrial genome analysis

Concatenated sequence data of cytochrome oxidase subunit 1 (*cox1*), cytochrome oxidase subunit 3 (*cox3*), cytochrome b (*cob*), ATP synthase F0 subunit 6 (*atp6*), ATP synthase subunit alpha (*atpa*), NADH dehydrogenase subunit 1 (*nad1*), NADH

TABLE 1 *Balamuthia mandrillaris* strains used in this study.

Strain	Accession no.	Location	Source	Sequencing technology	References
KM-20	OM994889	Thailand	4-year-old girl, cultured on A549 feeder cells	Oxford Nanopore, Illumina	Current study
2046	KP888565	California	26-year-old man, survivor, cultured on Vero cells	Illumina	Vollmer and Glaser (2016) and Greninger et al. (2015)
CDC-V039	CM003363	California	3-year, 10-month-old pregnant mandrill from the San Diego Zoo Wild Animal Park, axenic culture	PacBio	Visvesvara et al. (1990) and Detering et al. (2015)
V039*	KT175741	California	3-year, 10-month-old pregnant mandrill from the San Diego Zoo Wild Animal Park, cultured on Vero cells	Illumina	Visvesvara et al. (1990) and Greninger et al. (2015)
V451	KT030670	New York	6-year-old girl, cultured on Vero cells	Illumina	Greninger et al. (2015)
GAM-19	KT175739	-	V188-frozen stock	Illumina	Greninger et al. (2015)
RP5	KT030672	California	Environmental sample, cultured on Vero cells	Illumina	Greninger et al. (2015) and Schuster et al. (2003)
SAM	KT030673	California	3-year-old girl, isolated from brain	Illumina	Greninger et al. (2015) and Bakardjiev et al. (2003)
V188	KT175738	Georgia	59-year-old man, isolated from brain/skin lesion, cultured on Vero cells	Illumina	Greninger et al. (2015); Gordon et al. (1992)
V416*	AF477015	Australia	10-year-old girl, isolated from brain	-	Booton et al. (2003)
OK1	KT030671	California	Environmental sample, cultured on Vero cells	Illumina	Dunnebacke et al. (2004)
BeN	NC_027736	-	-	Illumina	Alexander LG (Unpublished)

The accession number, location, and source of the strains used in this study are presented in the table.

*Obtained from ATCC.

dehydrogenase subunit 2 (*nad2*), NADH dehydrogenase subunit 3 (*nad3*), NADH dehydrogenase subunit 4 (*nad4*), NADH dehydrogenase subunit 5 (*nad5*), NADH dehydrogenase subunit 6 (*nad6*), NADH dehydrogenase subunit 7 (*nad7*), and NADH dehydrogenase subunit 9 (*nad9*) of 10 strains of *B. mandrillaris* were aligned with MAFFT (v7.487) (Kuraku et al., 2013). The same set of genes in a concatenated sequence of *A. castellanii* (GenBank Accession number: U12386.1) was chosen as the outgroup. The alignment was imported into MEGA-X (v10.2.6) to perform phylogenetic analysis, and a maximum likelihood phylogenetic tree was computed using a JTT matrix-based model, with a bootstrap value of 1,000 (Jones et al., 1992; Kumar et al., 2018). The phylogenetic data were subsequently visualized using the Interactive Tree of Life (iTOL) (v5) (Letunic and Bork, 2021). The phylogenetic relationship and the mitochondrial sequences of 10 strains of *B. mandrillaris* were imported to AliTV for comparison and

visualization (Ankenbrand et al., 2017). Regions with a low link identity were further aligned and examined by Clustal Omega (v1.2.4) (Sievers et al., 2011).

The *rps3* protein tandem repeat sequences of all *B. mandrillaris* strains were extracted to perform a phylogenetic analysis using MEGA-X (v10.2.6) (Kumar et al., 2018). A parent tree with all protein tandem repeat sequences was constructed by the maximum likelihood method and JTT matrix-based model with a bootstrap value of 1,000 (Jones et al., 1992; Kumar et al., 2018). A subtree of the protein tandem repeats was constructed by removing the most conserved branch of repeating units and the fourth repeating unit of KM-20 *rps3* from the parent tree. The remaining sequences were, then, used to construct a phylogenetic subtree by the maximum likelihood method and JTT matrix-based model and a bootstrap value of 1,000 (Jones et al., 1992). The phylogenetic data were visualized using iTOL (v5) (Letunic and Bork, 2021).

3. Results

3.1. De novo assembly and annotation

The brain debridement of the GAE patient was sent for culture on A549 cell lines. Genomic DNA of *B. mandrillaris* KM-20 was extracted from the culture and sequenced to obtain a total of 7.66 Gb data using Oxford Nanopore long-read sequencing technology and 23Gb data using the Illumina NovaSeq 6,000 system. A circular mitochondrial genome of the *B. mandrillaris* KM-20 was *de novo* assembled in size of 42,630 bp and 35.34% GC content (Figure 1A). A total of 33 protein-coding, two rRNA, and 13 tRNA genes were annotated and located in the plus strand of the mitochondrial genome of *B. mandrillaris* KM-20 (Figure 1A). The protein-coding genes were classified into five groups, namely, ribosomal protein, NADH dehydrogenase, ATP synthase, cytochrome c oxidase, and cytochrome b.

According to a previous report, the mitochondrial genome size of *B. mandrillaris* strains ranges from 39,996 bp to 42,823 bp (Supplementary Table S2) (Greninger et al., 2015). The mitochondrial *cox1* gene of KM-20 is interrupted by a LAGLIDADG endonuclease-containing group IB intron (Supplementary Figure S2), which was also observed in four other strains, namely, 2046, SAM, RP5, and OK1 (Greninger et al., 2015). The length of LAGLIDADG endonuclease in these five strains ranges from 281 to 283 amino acids. For other strains, including, V451, V188, BeN, and GAM-19, the LAGLIDADG endonuclease is in the 23S rRNA gene, instead of the *cox1* gene (Greninger et al., 2015). V039 is the only strain that has no LAGLIDADG endonuclease inserted in protein-coding genes (Greninger et al., 2015).

3.2. Comparative analysis of *Balamuthia mandrillaris* strains

The phylogenetic analysis of *B. mandrillaris* mitochondrial genomes divided 10 strains into two clades and suggested KM-20 as the most distant strain (Figure 1B). In agreement with a previous study (Greninger et al., 2015), the four California strains (RP5, SAM, 2046, and OK1) formed a highly conserved cluster in the phylogenetic tree.

To perform a global mitochondrial genes comparison, the coding genes of 10 *B. mandrillaris* strains were compared in a matrix of pairwise identity percentage using KM-20 as a reference, and *rps3* was the only mitochondrial gene that has a percentage identity lower than 85% (Supplementary Figure S3). To further investigate the genomic diversity of *B. mandrillaris*, the mitochondrial genomic architectures of 10 strains were visualized by AliTV, and the result revealed a generally conserved gene synteny (Figure 1B) (Stajich et al., 2002; Harris, 2007; Ankenbrand et al., 2017). A break in synteny indicated in the purple triangle was observed in *cox1* and 23S rRNA, which corresponds to the introns that contain LAGLIDADG endonuclease. Other mapping gaps were caused by sequences missing in strain V039 and are not intron after manual checking. A region with only 70% link identity at approximately 14,500 bp was identified in the mitochondrial genome, and the variable position was confirmed to be the *rps3* gene. Interestingly, multiple sequence alignment of *rps3* protein sequences of 10 *B. mandrillaris* strains revealed that the variation could

be attributed to an array of protein tandem repeats (Supplementary Figure S4). The tandem repeat unit in *rps3* is named the R unit, and each R unit is composed of 17 amino acid residues, most of which started with four consensus amino acids, namely, arginine (R), proline (P), tryptophan (W), and leucine (L) (Supplementary Figure S4). KM-20 has seven R units, which is the highest number among all strains; V451 and BeN have six; GAM-19, V188, V039, RP5, and OK1 have five; SAM and 2046 have four and three R units, respectively. Despite the amino acid sequence within each R unit being conserved, the nucleotide sequences are highly degenerated and can be differentiated from each other. The identified repeats, therefore, are not due to sequencing error or collapse (Supplementary Figure S5). The length of the protein tandem repeat region in *rps3* ranges from 51 to 121 amino acid residues. The CNVs of the R units account for the difference in *rps3* length, which makes *rps3* a promising target for strain identification and genotyping of *B. mandrillaris* (Greninger et al., 2015).

The distribution of R units in *rps3* was illustrated with colors according to their phylogenetic relationship (Figure 2). A parent phylogenetic tree was constructed with all R units in 10 *B. mandrillaris* strains (Supplementary Figure S6). The last R units of all strains form a highly conserved branch in the parent phylogenetic tree and are colored in gray (Supplementary Figure S6). To better explore the R unit divergence, the gray-colored R units together with the most distant R4 unit of KM-20 were removed to construct a high-quality phylogenetic subtree (Figure 2B). The R units in 10 *B. mandrillaris* strains can be divided into nine main clades (Figure 2B).

A consensus sequence motif was generated to identify the conserved amino acid residues in different clusters by WebLogo (Figure 3) (Schneider and Stephens, 1990; Crooks et al., 2004). The R units of KM-20 are highly variable (Figure 3A), while the R units of the California strains (RP5, SAM, 2046, and OK1) are significantly conserved and start with RPWL amino acid residues (Figure 3B). The consensus sequences of the gray-colored R units in Figure 2A and overall R units from all strains are RPWLMSTWKNWKPGYAD and RPWL-G-RK--Y-EK--, respectively (Figures 3C,D). Positions 2–5 are populated by hydrophobic amino acids, which are colored in black (Figures 3A–D). The R units of all strains start with RPWL except R2, R3, and R4 of KM-20 and R5 of V451 and BeN, but the substituting amino acid residues, such as alanine, isoleucine, phenylalanine, and methionine, are also hydrophobic in nature, suggesting an N-terminal hydrophobic region is important for R units.

The general structure of *rps3* was predicted to have four transmembrane helices and three cytosolic domains by InterPro88.0 (Jones et al., 2014; Blum et al., 2021) (Supplementary Figure S7). The R units and the C-terminal domain of *rps3* are predicted to be in the cytosolic compartments (Supplementary Figure S7). The protein tandem repeats in *rps3* were identified as an intrinsically disordered region (IDR) by IUPred3 (Figure 3E) (Erdős et al., 2021). IDRs have no well-defined three-dimensional structures but are dynamically disordered and can fluctuate rapidly through different conformations (Wright and Dyson, 2015). The disordered region is in position 145–253 and overlaps with the repeating R units of KM-20, which are in position 167–293. Further structural and molecular analysis may assist in understanding the function of this highly variable region of *rps3*.

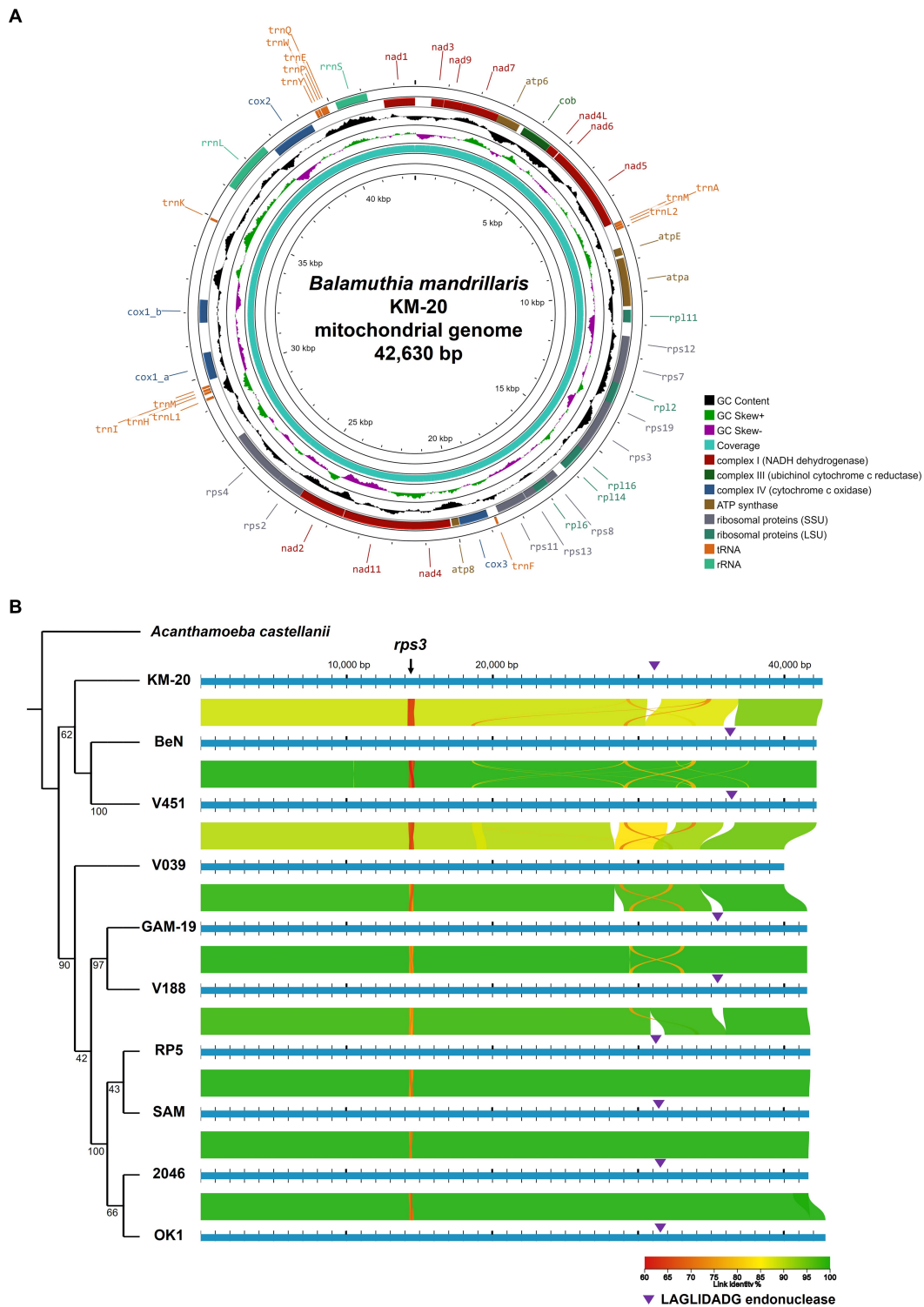


FIGURE 1

The mitochondrial genome of *B. mandrillaris* KM-20 and mitochondrial comparison among 10 strains of *B. mandrillaris*. (A) The whole mitochondrial genome of *B. mandrillaris* KM-20. The circular mitochondrial map depicts 33 protein-coding, two rRNA, and 13 tRNA genes. The average coverage of the mitochondrial genome assembly is 4464.09x, and the highest site coverage is 5,263x. The height of the innermost ring is calculated by dividing the site coverage by 5,263. (B) Phylogenetic relationship and mitochondrial genome alignment of 10 *B. mandrillaris* strains. Syntenic comparisons of linear mitochondrial chromosomal maps of 10 *B. mandrillaris* strains are visualized on AliTV software. Phylogenetic analysis was performed using concatenated sequences of *B. mandrillaris*, with *A. castellanii* chosen as an outgroup. The phylogenetic relationship between 10 strains of *B. mandrillaris* is retained for mitochondrial genome comparison. The line color represents the percentage of linked sequence identity. A red-colored variable region with approximately 70% link identity was identified at approximately 14,500bp in all strains of *B. mandrillaris*, which codes for the *rps3* gene. The purple triangle indicates the LAGLIDADG-endonuclease. Branch lengths are not drawn to proportion, and bootstrap values are shown for each node.

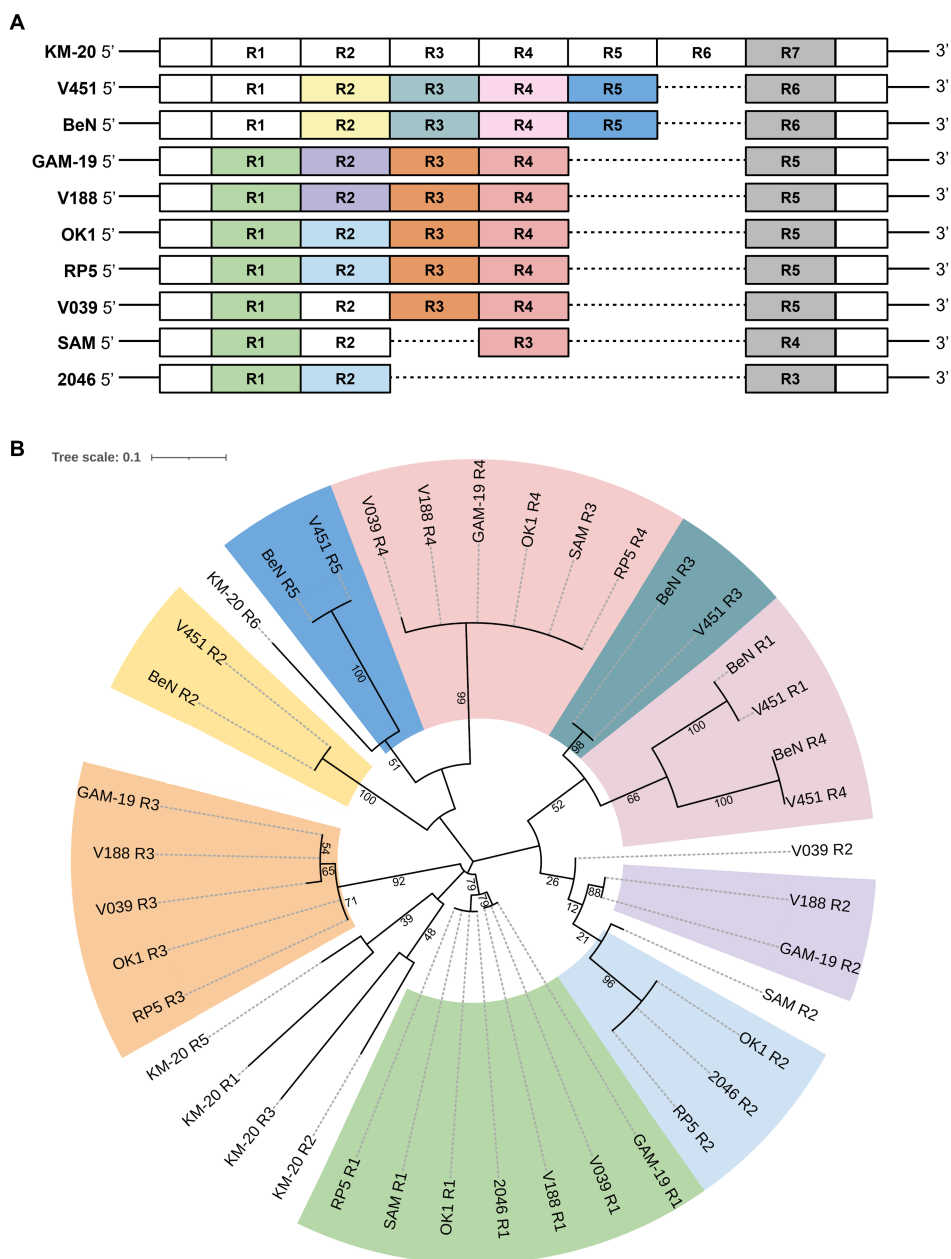
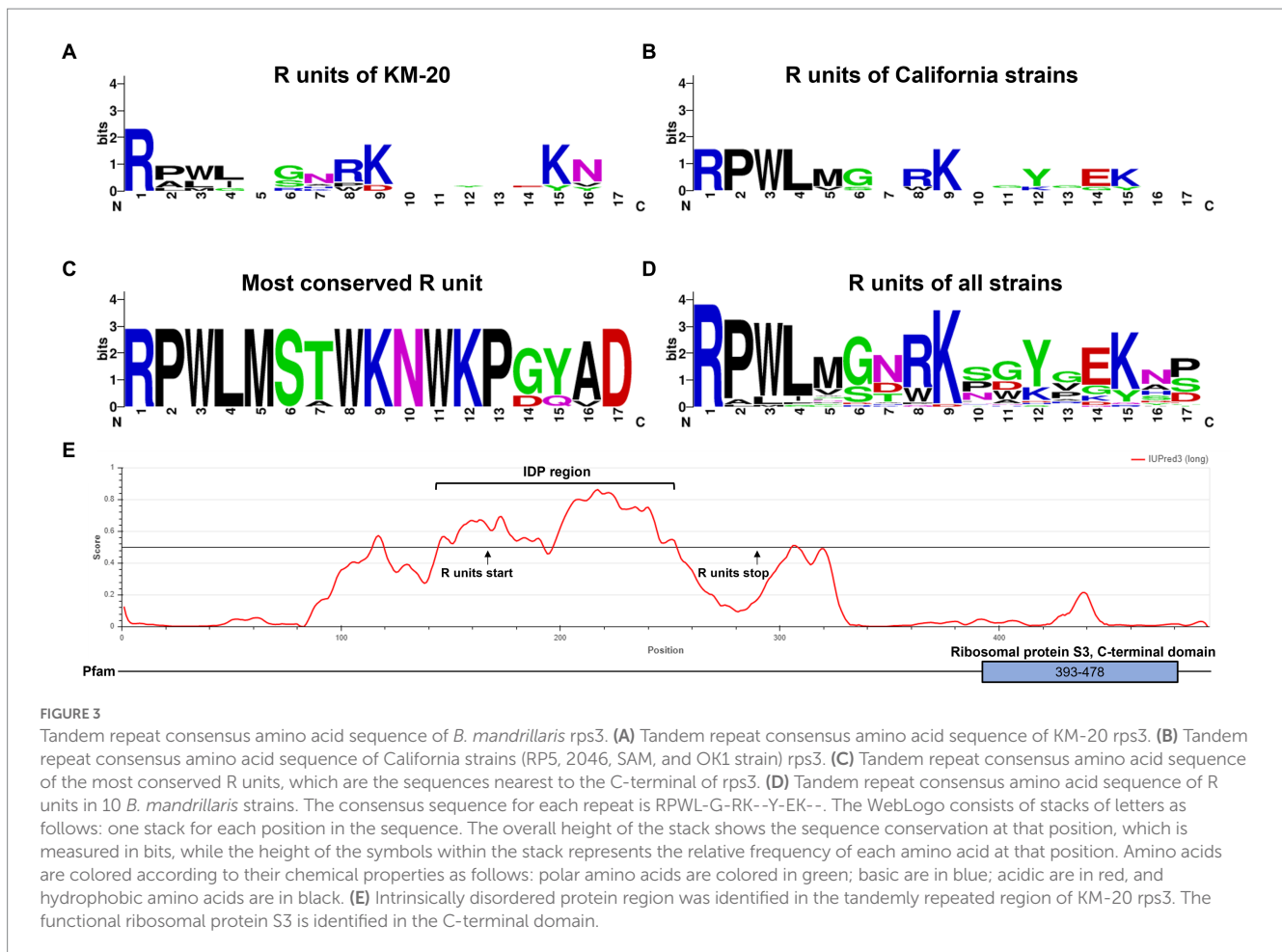


FIGURE 2 Distribution of the protein tandem repeats in *rps3* with their phylogenetic relationship matching the colors. **(A)** Distribution of protein tandem repeats in *rps3*. Sequence segments are not drawn to scale. KM-20 has seven R units; V451 and BeN have six; GAM-19, V188, V039, RP5, and OK1 strain have five; SAM strain has four, and the 2046 strain has three R units in *rps3*. **(B)** Phylogenetic analysis of *rps3* R units. The subtree of *rps3* R units shows nine main clades. R units that are not phylogenetically clustered with other repeating units are shaded in white, such as all R units of KM-20, R2 of SAM, and R2 of V039.

4. Discussion

Balamuthia mandrillaris is one of the free-living amoebae that occasionally cause GAE in humans and animals (Visvesvara et al., 2007; Visvesvara, 2013; Visvesvara, 2014), which is often life-threatening with limited treatment options (Cope et al., 2019). Owing to the restricted number of documented cases worldwide, establishing a conclusive link between genotype variation and clinical manifestation poses a challenge. It is speculated that variations in the genotypes of *B. mandrillaris* may account for the

dissimilar clinical presentations of its infection across diverse regions of the world. Retrospective reports from China and Peru demonstrated that the main clinical manifestations of *B. mandrillaris* infection are cutaneous lesions, which precede neurological involvement that develops several years later. In contrast, reported cases from the US presented solely with neurological symptoms, without any preceding skin lesion (Bravo and Seas, 2012; Wang et al., 2020), which is similar to the clinical presentation of our current case. Thus, the dissimilarity in disease aggressiveness and clinical manifestations could potentially stem in part from the genetic



variability within the species. Unfortunately, mitochondrial genome sequences of the cases reported in Peru and China, apart from those in the US, are unavailable for comparison with the current case. To comprehend the genetic variation that could be associated with clinical manifestations, the mitochondrial genome of *B. mandrillaris* strain KM-20 was *de novo* assembled and annotated in this study. To our knowledge, this study reports the first complete mitochondrial genome of *B. mandrillaris* clinical isolate obtained from Asia, as others were sequenced from samples isolated in the US. By comparing the mitochondria of KM-20 with other strains collected from the non-Asian area, we found the mitochondrial genome diversity can be attributed to the LAGALIDADG-containing intron in either *cox1* or 23S rRNA and a novel array of protein tandem repeats in *rps3*, which raises questions about the functional roles of this region within mitochondria and cells. In addition to clinical genotyping, the CNVs and domain architecture of the *rps3* tandem repeat can also infer the phylogenetic relationship among strains. The close phylogenetic relationship of the last R units among strains suggested that they could be the most ancient R unit (Supplementary Figure S6). *Cox1* gene has been widely used for species identification, phylogeography, and phylogenetic inference studies; however, the efficacy of using other mitochondrial genes has become less explored (Luo et al., 2011). The variable region in *rps3* can provide additional information on the phylogenetic relationship among strains, such as the copy number and divergence of the R units in *rps3*.

The mitochondrial features of *B. mandrillaris* among strains are generally conserved in terms of gene synteny and coding sequence, except for the presence of LAGLIDADG-containing endonuclease in either *cox1* or 23S rRNA, and the CNVs of protein tandem repeats in *rps3*. LAGLIDADG is a homing endonuclease occasionally included in the Group I self-splicing introns, which can cleave an intronless allele, resulting in the insertion of an intron and endonuclease into the previous intronless allele (Heidel and Glöckner, 2008). Group I introns are commonly found in fungi and protist nuclear rRNA genes as well as in organellar genomes, yet other organisms usually have no Group I introns in the genome (Haugen et al., 2007). Electrophoretic mobility shift assay and DNA cleavage assays can be further performed to identify the target sites of the LAGLIDADG-containing endonuclease (Grindl et al., 1998).

It is known that *rps3* plays a critical role in ribosome biogenesis and DNA repair in humans (Kim et al., 2013). Under stress conditions that promote DNA damage in which the cellular reactive oxygen species level increase, *rps3* accumulates in the mitochondria to repair damaged DNA (Kim et al., 2013). The analysis of ribosomal protein genes is currently lacking since metazoan mitochondrial genomes do not carry ribosomal protein genes (Gray et al., 1998; Heidel and Glöckner, 2008). The function of the *rps3* gene and the implication of CNV in the *rps3* tandem region in *B. mandrillaris* are currently unknown. Alterations in ribosomal genes in other species are shown to be related to adaptation and survival (Finken et al., 1993; Chittum and Champney, 1994). For example, amino acid changes in *rps12* in

Mycobacterium tuberculosis are adaptive for streptomycin resistance (Finken et al., 1993). Mutation in the ribosomal protein gene sequence in *Escherichia coli* is related to erythromycin resistance (Chittum and Champney, 1994). The function of CNVs in *rps3* tandem repeats will be interesting to further investigate.

The *rps3* protein sequence of KM-20 was searched in the NCBI non-redundant (NR) database using BLASTp on 26 September 2022. The result showed a significant match to the C-terminal domain of *rps3* with an e-value of 1.88e-05 (Supplementary Figure S8a), confirming the gene annotation. The percentage identities of the *rps3* protein sequence of KM-20 to humans, *Saccharomyces cerevisiae*, *Dictyostelium discoideum*, and *A. castellanii* were calculated by Clustal Omega (v1.2.4), which are 19.69%, 17.28%, 18.37%, and 30.53%, respectively. The C-terminal region of *rps3* of all *B. mandrillaris* strains constituted 113 amino acid residues and shared pairwise identities in the range of 99.12% to 100%, suggesting high conservation of *rps3* C-terminal within the species. In contrast, high variation was observed in the *rps3* sequence between *B. mandrillaris* and other species, including *A. castellanii*. Surprisingly, the *rps3* of *B. mandrillaris* is more similar to bacteria than to other amoebae, as the top BLASTp matches of KM-20 *rps3* are sequences of bacteria such as *Candidatus Calescamantes*, *Caldiserica* bacteria, and *Metallibacterium scheffleri* (e-value: 4e-06, 5e-06, and 7e-06, respectively) (Supplementary Figure S8b). However, the *rps3* protein sequence of humans, mice, *Drosophila melanogaster*, *S. cerevisiae*, and *D. discoideum* is relatively conserved, with pairwise amino acid residues identities in the range of 59.15% to 99.59%. It is speculated that the origin of the *rps3* gene in *B. mandrillaris* is different from other amoeba species. Further investigations can be performed to investigate the evolutionary origin of *rps3* in amoeba species.

Amoebae inhabit a wide range of ecological niches and rapid adaptation to new environments is advantageous to their survival. Varying the number of tandemly arrayed repeating units can increase the genomic sequence diversity and may enable the organisms to adapt to new environments relatively quicker and undergo more rapid and error-prone evolution than non-repeat-containing proteins (Marcotte et al., 1999; Jernigan and Bordenstein, 2015). All 10 strains of *B. mandrillaris* contain an array of protein tandem repeats in the *rps3* gene, which is not found in other amoebae, including *A. castellanii* and *N. fowleri*, raising questions about the function of the protein tandem repeats in *B. mandrillaris* exclusively. Although unicellular organisms can have significant deviations from typical animal mitochondrial genomes (Lavrov and Pett, 2016), a CNV in the mitochondrial coding region with substantial size variation among strains has not been reported in amoeba before. The conserved specific residues within each R unit may be critical for the structure or function, despite the precise number of repeats and the amino acid sequence may vary among strains (Javadi and Itzhaki, 2013).

A total of two genotypes of *rps3* were observed in strain V039, and the genetic difference lies in a 102 bp insertion in *rps3* which accounts for two extra R units (Detering et al., 2015; Greninger et al., 2015) (Supplementary Figure S9a). Both samples of V039 were isolated from the brain of a pregnant baboon that died from meningoencephalitis at the San Diego Zoo Wild Animal Park in 1990 but was subsequently cultured in different culture media (Detering et al., 2015; Greninger et al., 2015). The mitochondrial genome of the axenic cultivated CDC-V039 has a size of 39,894 bp while the other published

mitochondrial genome of V039 was cultured on Vero cells and has a size of 39,996 bp (Detering et al., 2015; Greninger et al., 2015) (Supplementary Figure S9b). Multiple sequence alignment has confirmed that the axenic CDC-V039 has three R units, whereas the Vero cell-cultured V039 has five, suggesting the possibility of mitochondrial heteroplasmy in *B. mandrillaris*. To verify whether the mitochondrial heteroplasmy of *B. mandrillaris* can be observed under certain conditions, we examined whether the change in culture media or temperature would induce mitochondrial heteroplasmy. However, we did not observe any evidence of mitochondrial heteroplasmy in the *rps3* gene of three *B. mandrillaris* strains including KM-20, V039, and V416 under various culture temperatures and culture medium conditions (Supplementary Figure S10).

5. Conclusion

In this study, we *de novo* assembled and annotated the complete mitochondrial genome of *B. mandrillaris* KM-20 using long-read and short-read sequencing data. Our comparative results explored the mitochondrial genome diversity among *B. mandrillaris* strains and revealed that one of the mitochondrial variations arises from an array of protein tandem repeats in the *rps3* gene, which has not been reported in other amoebae before. The copy number and sequence variations of the protein tandem repeats enable *rps3* to be a promising gene target for genotyping *B. mandrillaris* and can provide additional phylogenetic information. Collectively, this comparative mitochondrial genome analysis paves the way to investigate the evolution and genetic diversity of *B. mandrillaris* and other pathogenic amoebae.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving human participants were reviewed and approved by Siriraj Institutional Review Board of research involving human subjects (SIRB) (COA no. Si 806/2020). Risk Management Taskforce, Mahidol University, Thailand with approval no. SI2020-035. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

CL, TN, QX, KK, and PTS contributed to the conception and design of the study. TN, DS, and PS provided clinical data. CL, PTS, KK, LS, PR, and NT were responsible for the laboratory study. CL, TN, and PTS wrote the first draft of the manuscript. CL, QX, PTS, KK, and ST finalized the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants from the General Research Fund from Research Grants Council (Reference numbers: 464710, 475113, 14119219, 14119420, and 14175617), Health and Medical Research Fund from Food and Health Bureau (Reference numbers: 06171016 and 07181266), and Theme-based Research Scheme project from the Research Grants Council (Reference number: T11-712/19N) in Hong Kong. This study was also supported by a grant from the Faculty of Medicine Siriraj Hospital, Mahidol University (Grant number R016433002) and was carried out under the Siriraj Integrative Center for Neglected Parasitic Diseases, Faculty of Medicine Siriraj Hospital, Mahidol University. KK, PR, and PTS also received Chalermphrakiat grants from the Faculty of Medicine Siriraj Hospital, Mahidol University.

Acknowledgments

The authors would like to thank Mathee Ongsiriporn for the brain imaging analysis, and Punpob Lertlaituan for his assistance with the amoeba culture.

References

- Ankenbrand, M. J., Hohlfeld, S., Hackl, T., and Forster, F. (2017). AliTV—interactive visualization of whole genome comparisons. *PeerJ Comput. Sci.* 3:e116. doi: 10.7717/peerj-cs.116
- Bakardjiev, A., Azimi, P. H., Ashouri, N., Ascher, D. P., Janner, D., Schuster, F. L., et al. (2003). Amebic encephalitis caused by *Balamuthia mandrillaris*: report of four cases. *Pediatr. Infect. Dis. J.* 22, 447–452. doi: 10.1097/01.inf.0000066540.18671.f8
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., et al. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. doi: 10.1016/j.ympev.2012.08.023
- Blum, M., Chang, H., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977
- Booton, G. C., Carmichael, J. R., Visvesvara, G. S., Byers, T. J., and Fuerst, P. A. (2003). Identification of *Balamuthia mandrillaris* by PCR assay using the mitochondrial 16S rRNA Gene as a target. *J. Clin. Microbiol.* 41, 453–455. doi: 10.1128/JCM.41.1.453-455.2003
- Bravo, F. G., and Seas, C. (2012). *Balamuthia Mandrillaris* amoebic encephalitis: an emerging parasitic infection. *Curr. Infect. Dis. Rep.* 14, 391–396. doi: 10.1007/s11908-012-0266-4
- Chittum, H. S., and Champney, W. S. (1994). Ribosomal protein gene sequence changes in erythromycin-resistant mutants of *Escherichia coli*. *J. Bacteriol.* 176, 6192–6198. doi: 10.1128/jb.176.20.6192-6198.1994
- Cope, J. R., Landa, J., Nethercut, H., Collier, S. A., Glaser, C., Moser, M., et al. (2019). The epidemiology and clinical features of *Balamuthia mandrillaris* disease in the United States, 1974–2016. *Clin. Infect. Dis.* 68, 1815–1822. doi: 10.1093/cid/ciy813
- Crooks, G. E., Hon, G., Chandonia, J., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- da Rocha-Azevedo, B., Tanowitz, H. B., and Marciano-Cabral, F. (2009). Diagnosis of infections caused by pathogenic free-living amoebae. *Interdiscip. Perspect. Infect. Dis.* 2009, 251406–251414. doi: 10.1155/2009/251406
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. doi: 10.1093/gigascience/giab008
- de Souza, W., Attias, M., and Rodrigues, J. C. F. (2009). Particularities of mitochondrial structure in parasitic protists (Apicomplexa and Kinetoplastida). *Int. J. Biochem. Cell Biol.* 41, 2069–2080. doi: 10.1016/j.biocel.2009.04.007
- Detering, H., Aebischer, T., Dabrowski, P. W., Radonic, A., Nitsche, A., Renard, B. Y., et al. (2015). First draft genome sequence of *Balamuthia mandrillaris*, the causative agent of amoebic encephalitis. *Genome Announc.* 3, e01013–e0101315. doi: 10.1128/genomeA.01013-15
- Dunnebacke, T. H., Schuster, F. L., Yagi, S., and Booton, G. C. (2004). *Balamuthia mandrillaris* from soil samples. *Microbiology (Society for General Microbiology)* 150, 2837–2842. doi: 10.1099/mic.0.27218-0
- Erdős, G., Pajkos, M., and Dosztányi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* 49, W297–W303. doi: 10.1093/nar/gkab408
- Finken, M., Kirschner, P., Meier, A., Wrede, A., and Böttger, E. C. (1993). Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol. Microbiol.* 9, 1239–1246. doi: 10.1111/j.1365-2958.1993.tb01253.x
- Gordon, S. M., Steinberg, J. P., DuPuis, M. H., Kozarsky, P. E., Nickerson, J. F., and Visvesvara, G. S. (1992). Culture isolation of *Acanthamoeba* species and *Leptomyxid* Amebas from patients with amoebic meningoencephalitis, including two patients with AIDS. *Clin. Infect. Dis.* 15, 1024–1030. doi: 10.1093/clind/15.6.1024
- Grant, J. R., and Stothard, P. (2008). CGView server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36, W181–W184. doi: 10.1093/nar/gkn179
- Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., et al. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* 26, 865–878. doi: 10.1093/nar/26.4.865
- Greninger, A. L., Messacar, K., Dunnebacke, T., Naccache, S. N., Federman, S., Bouquet, J., et al. (2015). Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med.* 7:113. doi: 10.1186/s13073-015-0235-2
- Grindl, W., Wende, W., Pingoud, V., and Pingoud, A. (1998). The protein splicing domain of the homing endonuclease PI-SceI is responsible for specific DNA binding. *Nucleic Acids Res.* 26, 1857–1862. doi: 10.1093/nar/26.8.1857
- Harris, R.S. (2007). *Improved pairwise alignment of genomic DNA*. State College, PA: The Pennsylvania State University.
- Haugen, P., Bhattacharya, D., Palmer, J. D., Turner, S., Lewis, L. A., and Pryer, K. M. (2007). Cyanobacterial ribosomal RNA genes with multiple, endonuclease-encoding group I introns. *BMC Evol. Biol.* 7:159. doi: 10.1186/1471-2148-7-159
- Heidel, A. J., and Glöckner, G. (2008). Mitochondrial genome evolution in the social amoebae. *Mol. Biol. Evol.* 25, 1440–1450. doi: 10.1093/molbev/msn088
- Intalaporn, P., Suankratay, C., Shuangshoti, S., Phantumchinda, K., Keelawat, S., and Wilde, H. (2004). *Balamuthia Mandrillaris* meningoencephalitis: the first case in Southeast Asia. *Am. J. Trop. Med. Hyg.* 70, 666–669. doi: 10.4269/ajtmh.2004.70.666
- Javadi, Y., and Itzhaki, L. S. (2013). Tandem-repeat proteins: regularity plus modularity equals design-ability. *Curr. Opin. Struct. Biol.* 23, 622–631. doi: 10.1016/j.sbi.2013.06.011
- Jernigan, K. K., and Bordenstein, S. R. (2015). Tandem-repeat protein domains across the tree of life. *PeerJ (San Francisco, CA)* 3:e732. doi: 10.7717/peerj.732
- Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1162963/full#supplementary-material>

- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kim, Y., Kim, H. D., and Kim, J. (2013). Cytoplasmic ribosomal protein S3 (rpS3) plays a pivotal role in mitochondrial DNA damage surveillance. *Biochim. Biophys. Acta Mol. Cell Res.* 1833, 2943–2952. doi: 10.1016/j.bbamcr.2013.07.015
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Kumar, S., Stecher, G., Li, M., Niyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kuraku, S., Zmasek, C. M., Nishimura, O., and Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* 41, W22–W28. doi: 10.1093/nar/gkt389
- Lang, B. F., Laforest, M., and Burger, G. (2007). Mitochondrial introns: a critical view. *Trends Genet.* 23, 119–125. doi: 10.1016/j.tig.2007.01.006
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152
- Lavrov, D. V., and Pett, W. (2016). Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol. Evol.* 8, 2896–2913. doi: 10.1093/gbe/evw195
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Luo, A., Zhang, A., Ho, S. Y., Xu, W., Zhang, Y., Shi, W., et al. (2011). Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genom.* 12:84. doi: 10.1186/1471-2164-12-84
- Marcotte, E. M., Pellegrini, M., Yeates, T. O., and Eisenberg, D. (1999). A census of protein repeats. *J. Mol. Biol.* 293, 151–160. doi: 10.1006/jmbi.1999.3136
- Matin, A., Siddiqui, R., Jayasekera, S., and Khan, N. A. (2008). Increasing importance of *Balamuthia mandrillaris*. *Clin. Microbiol. Rev.* 21, 435–448. doi: 10.1128/CMR.00056-07
- Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100. doi: 10.1093/nar/18.20.6097
- Schuster, F. L. (2002). Cultivation of pathogenic and opportunistic free-living Amebas. *Clin. Microbiol. Rev.* 15, 342–354. doi: 10.1128/CMR.15.3.342-354.2002
- Schuster, F. L., Dunnebacke, T. H., Martinez, A. J., Visvesvara, G. S., Booton, G. C., Yagi, S., et al. (2003). Environmental isolation of *Balamuthia mandrillaris* associated with a case of amebic encephalitis. *J. Clin. Microbiol.* 41, 3175–3180. doi: 10.1128/JCM.41.7.3175-3180.2003
- Schuster, F. L., Yagi, S., Gavali, S., Michelson, D., Raghavan, R., Blomquist, I., et al. (2009). Under the radar: *Balamuthia amebic encephalitis*. *Clin. Infect. Dis.* 48, 879–887. doi: 10.1086/597260
- Siddiqui, R., and Khan, N. A. (2008). *Balamuthia amebic encephalitis: an emerging disease with fatal consequences*. *Microb. Pathog.* 44, 89–97. doi: 10.1016/j.micpath.2007.06.008
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., et al. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 12, 1611–1618. doi: 10.1101/gr.361602
- Stothard, P., and Wishart, D. S. (2005). Circular genome visualization and exploration using CGView. *Bioinformatics* 21, 537–539. doi: 10.1093/bioinformatics/bti054
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Visvesvara, G. S. (2013). Infections with free-living amebae. *Handb. Clin. Neurol.* 153–168. doi: 10.1016/b978-0-444-53490-3.00010-8
- Visvesvara, G. S. (2014). “Pathogenic and opportunistic free-living amoebae: agents of human and animal disease” in *Anonymous Manson’s tropical diseases*. eds. J. Farrar, P. Hotez, T. Junghanss, G. Kang, D. Lalloo and N. White 23rd ed., (essay, Elsevier Saunders), 683–691.
- Visvesvara, G. S., Martinez, A. J., Schuster, F. L., Leitch, G. J., Wallace, S. V., Sawyer, T. K., et al. (1990). Leptomyxid ameba, a new agent of amebic meningoencephalitis in humans and animals. *J. Clin. Microbiol.* 28, 2750–2756. doi: 10.1128/JCM.28.12.2750-2756.1990
- Visvesvara, G. S., Moura, H., and Schuster, F. L. (2007). Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris*, *Naegleria fowleri*, and *Sappinia diploidea*. *FEMS Immunol. Med. Microbiol.* 50, 1–26. doi: 10.1111/j.1574-695X.2007.00232.x
- Vollmer, M. E., and Glaser, C. (2016). A *Balamuthia* survivor. *JMM Case Rep.* 3:e005031. doi: 10.1099/jmmcr.0.005031
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, L., Cheng, W., Li, B., Jian, Z., Qi, X., Sun, D., et al. (2020). *Balamuthia mandrillaris* infection in China: a retrospective report of 28 cases. *Emerg. Microbes Infect.* 9, 2348–2357. doi: 10.1080/22221751.2020.1835447
- Wright, P. E., and Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29. doi: 10.1038/nrm3920