

RESEARCH

Open Access



Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing

Alexander L. Greninger^{1,2†}, Kevin Messacar^{3†}, Thelma Dunnebacke⁴, Samia N. Naccache^{1,2}, Scot Federman^{1,2}, Jerome Bouquet^{1,2}, David Mirsky³, Yosuke Nomura³, Shigeo Yagi⁴, Carol Glaser⁵, Michael Vollmer⁶, Craig A. Press³, Bette K. Kleinschmidt-DeMasters³, Samuel R. Dominguez³ and Charles Y. Chiu^{1,2,7*}

Abstract

Background: Primary amoebic meningoencephalitis (PAM) is a rare, often lethal, cause of encephalitis, for which early diagnosis and prompt initiation of combination antimicrobials may improve clinical outcomes.

Methods: In this study, we sequenced a full draft assembly of the *Balamuthia mandrillaris* genome (44.2 Mb in size) from a rare survivor of PAM, and recovered the mitochondrial genome from six additional *Balamuthia* strains. We also used unbiased metagenomic next-generation sequencing (NGS) and SURPI bioinformatics analysis to diagnose an ultimately fatal case of *Balamuthia mandrillaris* encephalitis in a 15-year-old girl.

Results and Discussion: Comparative analysis of the mitochondrial genome and high-copy number genes from six additional *Balamuthia mandrillaris* strains demonstrated remarkable sequence variation, and the closest *Balamuthia* homologs corresponded to other amoebae, hydroids, algae, slime molds, and peat moss. Real-time NGS testing of hospital day 6 CSF and brain biopsy samples detected *Balamuthia* on the basis of high-quality hits to 16S and 18S ribosomal RNA sequences present in the National Center for Biotechnology Information (NCBI) nt reference database. The presumptive diagnosis of PAM by visualization of amoebae on brain biopsy histopathology and NGS analysis was subsequently confirmed at the US Centers for Disease Control and Prevention (CDC) using a *Balamuthia*-specific PCR assay. Retrospective analysis of a day 1 CSF sample revealed that more timely identification of *Balamuthia* by metagenomic NGS, potentially resulting in a better clinical outcome, would have required availability of the complete genome sequence.

Conclusions: These results underscore the diverse evolutionary origins of *Balamuthia mandrillaris*, provide new targets for diagnostic assay development, and will facilitate further investigations of the biology and pathogenesis of this eukaryotic pathogen. The failure to identify PAM from a day 1 sample without a fully sequenced *Balamuthia* genome in the database highlights the critical importance of whole-genome reference sequences for microbial detection by metagenomic NGS.

* Correspondence: charles.chiu@ucsf.edu

†Equal contributors

¹Department of Laboratory Medicine, University of California, 185 Berry Street, Box 0134, San Francisco, CA 94107, USA

²UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA 94107, USA

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

Background

Balamuthia mandrillaris is a free-living amoeba that is a rare, almost uniformly lethal, cause of primary amoebic meningoencephalitis (PAM) in humans [1]. Originally isolated from the brain of a baboon at the San Diego Zoo in 1986, *Balamuthia mandrillaris* has since been reported in over 100 cases of PAM worldwide [2–4], and amoebae associated with fatal encephalitis in a child have been cultured directly from soil [5]. The vast majority of cases are fatal, although there are a few published case reports of patients surviving *Balamuthia* encephalitis after receiving combination antimicrobial therapy and *in vitro* data supporting the potential efficacy of several antimicrobial agents [3, 6–10]. Despite the availability of validated RT-PCR assays for the detection of free-living amoebae [11, 12], PAM is not often clinically suspected and the diagnosis is most commonly made around the time of death or postmortem on brain biopsy [13, 14].

Our lab has demonstrated the capacity of metagenomic next-generation sequencing (NGS) to provide clinically actionable information in a number of acute infectious diseases, most notably encephalitis [15, 16]. This approach enables the rapid and simultaneous detection of viruses, bacteria, and eukaryotic parasites in clinical samples [17]. Encephalitis is a critical illness with a broad differential, for which unbiased diagnostic tools such as metagenomic NGS can make a significant impact [18]. However, the utility of diagnostic NGS is highly dependent on the breadth and quality of databases that contain whole-genome sequence information of reference strains needed for alignment [19].

In this study, we describe the first draft genome sequence of a strain of *Balamuthia mandrillaris* from a rare survivor of PAM and comparative sequence analysis of six additional mitochondrial genomes. We also demonstrate the ability of metagenomic NGS to rapidly detect *Balamuthia mandrillaris* from the cerebrospinal fluid (CSF) of a critically ill 15-year-old girl, and highlight the importance of genomic reference sequences by retrospective analysis of a hospital day (HD) 1 sample.

Methods

Ethics

Written informed consent was obtained from the parents of the patient for analysis of her clinical samples, in accordance with a study approved by the Colorado Multiple Institutional Review Board (IRB). Written informed consent was also obtained from the parents for publication of this research. Coded samples from the patient were analyzed for pathogens by NGS under study protocols approved by the University of California, San Francisco IRB.

Metagenomic sequencing of CSF and brain biopsy

Total nucleic acid was extracted from 200 μ L of CSF using the Qiagen EZ1 Viral kit. Half of the nucleic acid from CSF was treated with Turbo DNase (Ambion). Total RNA was extracted from 2 mm³ brain biopsy tissue using the Direct-zol RNA MiniPrep Kit (Zymo Research), followed by mRNA purification using the Oligotex mRNA Mini Kit (Qiagen). Total RNA from CSF and mRNA from brain biopsy were reverse-transcribed using random hexamers and randomly amplified as previously described [20]. The resulting double-stranded cDNA or extracted DNA from CSF (the fraction not treated with Turbo DNase) was used as input into Nextera XT, following the manufacturer's protocol except with reagent volumes cut in half for each step in the protocol. After 14–18 cycles of PCR amplification, barcoded libraries were cleaned using Ampure XP beads, quantitated on the BioAnalyzer (Agilent), and run on the Illumina MiSeq (1 \times 160 bp run). Metagenomic NGS data were analyzed for pathogens via SURPI using NCBI nt/nr databases from June 2014 [17].

A rapid taxonomic classification algorithm based on the lowest common ancestor was incorporated into SURPI, as previously described [20], and used to assign viral, bacterial, and non-chordate eukaryotic NGS reads to the species, genus, or family level. For the SNAP nucleotide aligner [21], an edit distance cutoff of 12 was used for viral reads [17], but adjusted to a more stringent edit distance of 6 for bacterial and non-chordate eukaryotic reads to increase specificity.

Rigorous measures were taken to prevent laboratory contamination and cross-contamination in this study. This included: (1) library preparation and sequencing of day 1 clinical samples, day 6 samples, and *Balamuthia* cultures on different weeks (with at least three sequencing runs in between to minimize carryover); (2) unidirectional laboratory workflow, with separate rooms for the pre-PCR and post-PCR steps; (3) cleanup of the work area prior to and after specimen processing with 10 % sodium hypochlorite; (4) use of disposable gloves and biohazard pouches that are frequently changed; and (5) dual-index barcoding of multiplexed clinical samples within individual runs. A negative control clinical CSF sample from an unrelated patient was processed in parallel with the case patient's CSF and brain biopsy samples at day 1, and was negative for any reads aligning to *Balamuthia*. Furthermore, the nine *Balamuthia* DNA library sequences observed in the patient's day 1 CSF were unique and did not overlap any of the 926 DNA library sequences in day 6 CSF, excluding the possibility of amplicon contamination.

Propagation of *Balamuthia mandrillaris* in culture

Trypsin-treated cultures from Vero or BHK (baby hamster kidney) cell monolayers and *Balamuthia mandrillaris*

amoebas were placed into T25 culture flasks in Dulbecco's Modified Eagle Medium (DMEM) plus 10 % bovine serum, 1 % of the antibiotics 100 U/mL penicillin, 0.1 mg/mL streptomycin, and 0.25 µg/mL fungizone, and incubated at 37 °C for 7 to 10 days plus 2 days at room temperature until the underlying cell sheet was completely destroyed and only actively dividing amoebae were seen floating and/or attached to the surface. Attached amoebas were freed by gently tapping the side of the flask or putting the flask on a bed of ice for 20 min. The amoebas were concentrated by centrifugation for 5 min. After removal of supernatant, the amoeba pellet was washed by addition of PBS, centrifuged again, and then placed in a flask with Bacto-casitone axenic medium and allowed to grow for another 7–10 days, after which the amoebas were concentrated again, washed, and placed into fresh axenic medium [22]. After a final centrifugation step, the amoebae were collected, washed 3× in PBS, pelleted, and stored at –80 °C.

Sequencing and annotation of cultured *Balamuthia mandrillaris* 2046 strain

DNA from individual strains of *Balamuthia mandrillaris* was extracted using the Qiagen EZ1 Tissue Kit and used as input for the Nextera Mate Pair Kit (Illumina) and Nextera XT Kit (Illumina), following the manufacturer's instructions. Mate-pair libraries were sequenced on an Illumina MiSeq (2 × 80 nt and 2 × 300 nt runs), while the Nextera XT library was sequenced on an Illumina HiSeq (2 × 250 bp paired-end sequencing) (Table 2). Mate-pair reads from run MP1 were adapter-trimmed with NxTrim [23], and the mitochondrial genome of strain 2046 and high-copy number contigs were assembled using SPAdes v3.5 [24, 25]. The average insert size of the mate-pair library was 2,187 nucleotides. Prediction of tRNA and rRNA genes was performed using tRNAscan-SE and RNAmmer v1.2, respectively [26, 27]. ORFs were predicted in translation code 4 with the Glimmer gene predictor, and all predicted ORF sequences were confirmed using BLASTx and HHPred [28, 29].

Reads from runs MP2 and MP3 were mate-pair adapter-trimmed using NxTrim, while reads from all runs were quality-filtered (q30) and adapter-trimmed using cutadapt [30]. Reads that aligned to the *Balamuthia* mitochondrial genome and golden hamster (*Mesocricetus auratus*) were identified using SNAP [21] and removed prior to *de novo* assembly using platanus [31]. Any scaffold of length less than 500 bp along with 62 scaffolds that aligned to *Mesocricetus auratus*, *Chlorocebus sabaesus*, *Waddlia chondrophila*, and *Enterobacteria phage phiX174* (all likely deriving from cell culture contamination) were removed.

Rps3 PCR confirmation

Genomic DNA was PCR-amplified using 0.5 µM final concentration of primers rps3-F (5'-CTGYTCGATTTT CGAAAAATAAAGTAG-3') and rps3-R (5'-TGAAA-GAAGAACATTTAGATCAGACT-3') using 2X iProof HF Master Mix (Bio-Rad) in 20 µL total volume. Conditions were 95 °C × 2 min, followed by 35 cycles of 95 °C × 30 s, 52 °C × 30 s, 72 °C × 40 s, and a final incubation at 72 °C × 2 min. PCR amplicons were visualized by 3 % agarose gel electrophoresis.

Data access

The *Balamuthia mandrillaris* mitochondrial genomes have been deposited in NCBI under the following accession numbers: 2046 axenic (KP888565), 2046–1 (KT175740), V451 (KT030670), OK1 (KT030671), RP-5 (KT030672), SAM (KT030673), V188-axenic (KT175738), V188-frozen stock (KT175739), and V039 (KT175741). The *Balamuthia mandrillaris* scaffolds have been deposited in the NCBI Whole Genome Shotgun (WGS) database under accession number LEOU00000000. Metagenomic NGS data corresponding to the patient's brain biopsy and CSF fluid, as well as corresponding to a negative control CSF sample from an unrelated patient, have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRP064607. NGS reads were filtered for exclusion of human sequences using Bowtie2 high-sensitivity local alignment to the human hg38 reference database at default parameters followed by BLASTn alignment to all primate sequences in NCBI nt at an e-value cutoff of 10⁻⁸ prior to deposition into the NCBI SRA database.

Results

First mitochondrial genome of *Balamuthia mandrillaris*

We cultured *Balamuthia mandrillaris* strain 2046 in axenic media from a brain biopsy corresponding to a rare survivor of PAM (Vollmer and Glaser, manuscript in review). Sequencing of DNA extracted from the axenic culture generated 3.8 million 75 base pair (bp) mate-pair reads with an average insert size of 2,187 bp. *De novo* assembly yielded a circular mitochondrial genome of 41,656 bases that was comprised of 64.8 % AT at 2,082× coverage (Fig. 1a). The overall size and AT content of the *Balamuthia mandrillaris* mitochondrial genome was closer to that of *Acanthamoeba castellanii* (41,591 bp, 70.6 % AT) [32] than *Naegleria fowleri* (49,531 bp, 74.8 % AT) [33], although overall average nucleotide identity with *Balamuthia* was found to be low for both amoebae (approximately 68 %).

The *Balamuthia mandrillaris* 2046 mitochondrial genome contained two ribosomal RNA (rRNA) genes, 18 transfer RNA (tRNA) genes, and 38 coding sequences, with five of those being hypothetical proteins. The organization of the mitochondrial genome retained

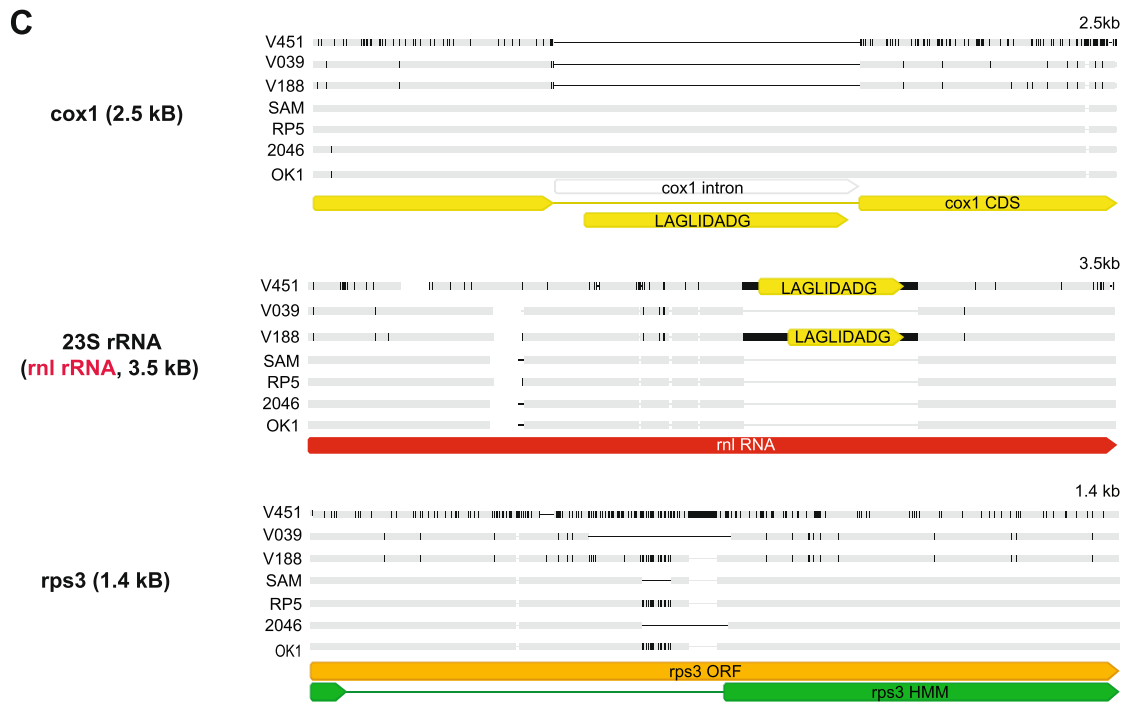
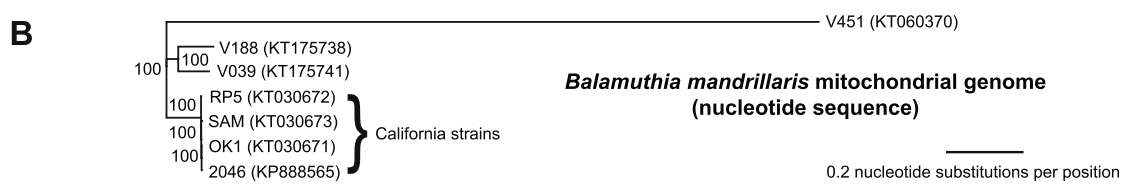
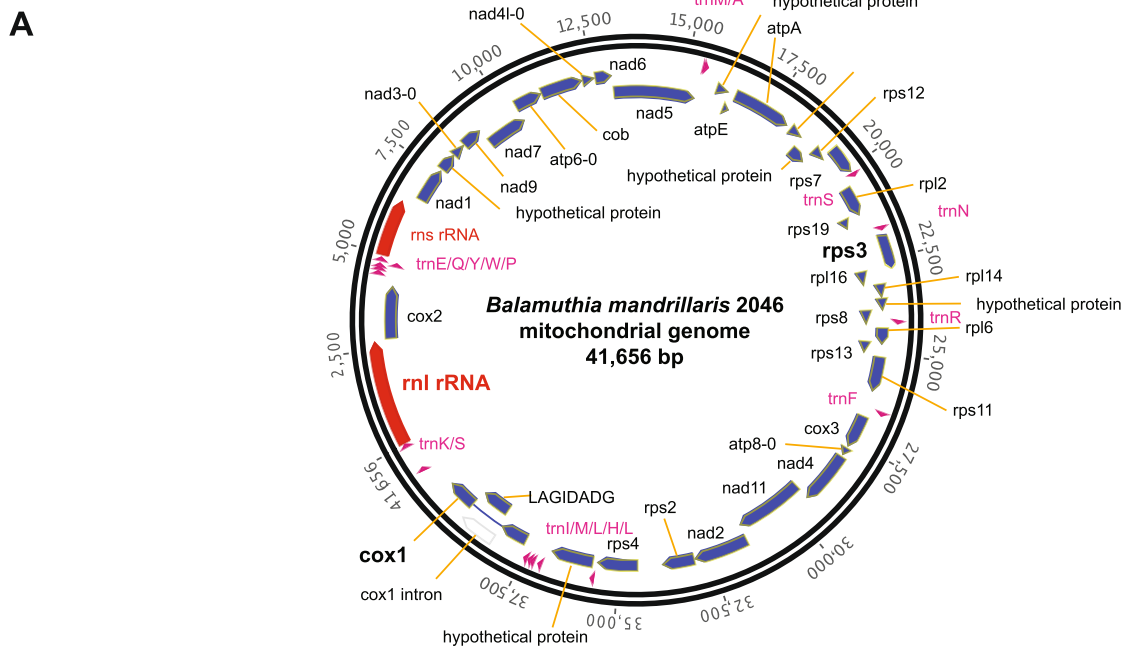


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Sequencing and comparative phylogenetic analysis of the mitochondrial genome of *Balamuthia mandrillaris*. **a** *Balamuthia mandrillaris* 2046 mitochondrial genome. Annotation of the 41,656 bp genome was performed using RNAmmer, tRNAscan-SE, and Glimmer gene predictor, with all ORFs manually verified using BLASTx alignment. **b** Phylogenetic analysis of seven newly sequenced mitochondrial genomes from different strains of *Balamuthia mandrillaris*. An outgroup (for example, *Acanthamoeba castellanii*) is not shown given the lack of gene synteny. Branch lengths are drawn proportionally to the number of nucleotide substitutions per position, and support values are shown for each node. **c** Differences in individual gene features (cox1, 23S rRNA, and rps3), among the 7 mitochondrial genomes, as detailed in the text. The mitochondrial cox1, 23S rRNA / rnl RNA, and rps3 genes are highlighted in boldface

several syntenic blocks with the *Acanthamoeba castellanii* genome, including tnad3-9-7-atp6 and rpl11-rps12-rps7-rpl2-rps19-rps3-rpl16-rpl14. However, many other features of the genome were unique, such as the order of the remaining coding blocks, the lack of a combined cox1/cox2 gene, as present in *Acanthamoeba castellanii* [32], and the lack of intron splicing in 23S rRNA. The *Balamuthia mandrillaris* mitochondrial cox1 gene was interrupted by a LAGLIDADG endonuclease open reading frame (ORF) containing a group I intron, as has been reported for a wide variety of other eukaryotic species [34, 35]. The putative rps3 gene was encoded within a 1,290 bp ORF that, when translated, aligned by hidden Markov model (HMM) analysis to rps3 proteins from *Escherichia coli* (PDB 4TP8/4U26) and *Thermus thermophilus* (PDB 4RB5/4W2F), and only in base positions 1–66 and 583–1290 of the ORF. This finding was consistent with the presence of a putative > 500 bp intron in the *Balamuthia mandrillaris* rps3 gene that to date has only been described to in plants [36, 37]. Alternatively, the ORF was also found to encode a putative tRNA (Asn) such that the 5' end of the ORF could represent a large intergenic sequence.

Mitochondrial genome diversity of *Balamuthia mandrillaris*

To investigate the extent of sequence diversity in *Balamuthia mandrillaris*, we sequenced the mitochondrial genomes from six additional *Balamuthia* strains obtained from the California Department of Public Health and

the American Tissue Culture Collection (Table 1). The seven total circular mitochondrial genomes averaged 41,526 bp in size (range, 39,996 to 42,217 bp), and shared pairwise nucleotide identities in the range of 82.6 % to 99.8 %. The phylogeny revealed the presence of at least three separate lineages of *Balamuthia mandrillaris*, with all four strains from human cases in California clustering together in a single clade (Table 1; Fig. 1b). Consistent with a previous report [38], we found that the mitochondrial genome of strain V451 was the most divergent among tested strains, and possessed an additional 1,149 bp ORF downstream of the cox1 gene that did not align significantly to any sequence in the NCBI nt or nr reference database. Of note, our V039 mitochondrial genome assembly showed > 99.9 % identity to a recently deposited PacBio / Illumina assembly of the V039 mitochondrial genome. The only difference between the two mitochondrial genomes was a 102-bp insertion that was located in the unique rps3 locus [39].

Putative introns constituted the major source of variation among the mitochondrial genomes (Fig. 1c). Four strains out of seven, including strain 2046 from the rare survivor of *Balamuthia* infection, contained a 975 bp LAGLIDADG-containing intron in the cox1 gene, whereas no such intron was present in the remaining three strains. Two of the remaining three strains, strains V451 and V188, instead had an approximately 790 bp insert in the 23S rRNA gene (Fig. 1a, 'rnl RNA') that contained a 530 bp or 666 bp LAGLIDADG-containing ORF,

Table 1 Strains used in this study

Strain	Date	Location	Note	Citation
SAM ^a	Spring 2001	California	3-year-old girl, cultured on Vero cells	Bakardjiev A <i>et al.</i> , 2002 [55]
RP-5 ^a	Spring 2001	California	Environmental sample associated with SAM, cultured on Vero cells	Schuster FL <i>et al.</i> , 2003 [5]
OK1 ^a	2002	California	Environmental sample in NorCal unrelated to SAM, cultured on Vero cells	Dunnebacke TH <i>et al.</i> , 2004 [56]
V039 ^b	1990	California	Type strain isolated from pregnant mandrill at San Diego Zoo, cultured on Vero cells	Visvesvara GS <i>et al.</i> , 1990 [4]
V188 ^a	1989	Georgia	59-year-old man, amoeba isolated from brain/skin lesion, cultured on Vero cells	Gordon SM <i>et al.</i> , 1992 [57]
V451 ^a	-	New York	6-year-old girl, cultured on Vero cells	Yang XH <i>et al.</i> , 2001 [58]
2046 ^a	March 2010	California	26-year-old man, survivor, cultured on Vero cells	Vollmer and Glaser, in review
CSF108 ^c	May 2015	Colorado	15-year-old girl, direct metagenomic detection from CSF	

^aStrain provided by Dr. Dunnebacke at the California Department of Public Health

^bStrain obtained from ATCC (V39)

^cStrain obtained from recent clinical case in Colorado

respectively, and that coded for a putative endonuclease. The LAGLIGDADG-containing endonucleases in the two strains shared 84 % amino acid pairwise identity with each other, but approximately 50 % amino acid identity to a corresponding LAGLIGDADG-containing endonuclease in the 23S rRNA gene of *Acanthamoeba castellanii*, and <12 % amino acid identity to the LADGLIDADG-containing *cox1* introns in the four other *Balamuthia* strains. The final remaining strain, ATCC-V039, lacked an intron in either the *cox1* or 23S rRNA gene.

The ORF containing the *rps3* gene, found to contain a possible *rps3* intron / intergenic region by analysis of the strain 2046 mitochondrial genome, varied in length among the seven sequenced mitochondrial genomes from 1,290 bp to 1,425 bp. Notably, the length of the putative intron / intergenic region accounted for all of the differences in overall length of the *rps3* gene. Confirmatory PCR and sequencing of this locus using conserved outside primers revealed that each strain tested had a unique length and sequence (Fig. 2), raising the possibility of targeting this region for *Balamuthia mandrillaris* strain detection and genotyping.

First draft genome of *Balamuthia mandrillaris*

Because of the high-copy number of mitochondrial sequences in the *Balamuthia* sequencing run, as noted previously for *Naegleria fowleri* [33], we performed an additional NGS run of 14.1 million 250 bp single-end reads, and computationally subtracted reads aligning to the mitochondrial genome. Assembly of the remaining 4.4 million high-quality reads yielded 31,194 contiguous sequences (contigs) with an N50 of 3,411 bp. Scaffolding and gap closure using an additional 57.4 million NGS reads and computational removal of exogenous sequence contaminants yielded a final assembly of approximately 44.3 Mb comprising 14,699 scaffolds with an N50 of 19,012 bp (Table 2). Direct BLASTn alignment of the scaffolds to

the National Center for Biotechnology Information (NCBI) nt database revealed that the most common organism aligning to *Balamuthia mandrillaris* was *Mus musculus* (house mouse) (2,067/14,699 = 14.1 % of scaffolds), nearly entirely due to low-complexity sequences, followed by high-significance hits to *Acanthamoeba castellanii* (627/14,699 = 4.3 % of scaffolds).

Phylogenetic analysis of individual genes from the *Balamuthia* mitochondrial genome revealed the presence of significant diversity across all kingdoms of life (Fig. 3). The 18S-28S rRNA locus in the *Balamuthia* mitochondrial genome corresponded to a 12.5 kB contig sequenced at high coverage (>400× over rRNA regions). The previously sequenced 18S gene (2,017 bp) demonstrated 99.5 % identity to existing *Balamuthia mandrillaris* 18S rRNA sequences in the NCBI nt database, while the 28S gene (4,999 bp) had homology across multiple diverse species, with only 68.5 % pairwise identity to its closest phylogenetic relative, *Acanthamoeba castellanii* (Fig. 3b). From the nuclear genome, one high-copy contig contained a truncated 5,250 nucleotide ORF exhibiting only 33 % amino acid identity to *Rhizopus delemar* (pin mold), and harboring elements consistent with a retrotransposon [40], including an RNase HI from Ty3/Gypsy family retroelements, a reverse transcriptase, a chromodomain, and a retropepsin. Two high-copy, approximately 1,600 bp ORFs that failed to match any sequence by BLASTx alignment to the NCBI nr protein database were found to align significantly to *Escherichia coli* site-specific recombinase by remote homology HMM analysis.

A case of *Balamuthia* encephalitis diagnosed by metagenomic NGS

Concurrent with assembly of the *Balamuthia* genome, metagenomic NGS was performed to investigate a case of meningoencephalitis in a 15-year-old girl with insulin-

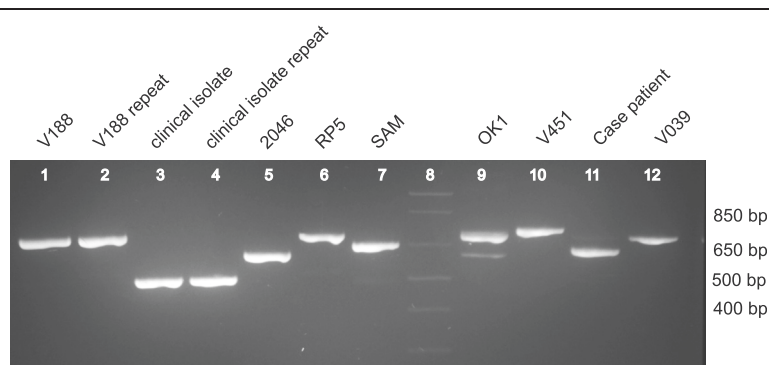


Fig. 2 PCR amplification of the *Balamuthia rps3* mitochondrial gene. The variable length of the *rps3* intron among eight different *Balamuthia* strains (seven newly sequenced mitochondrial genomes and the case patient) suggests that this gene may be an attractive target for development of a molecular genotyping assay. Lane 8 corresponds to the DNA ladder (faint appearance), while lanes 3 and 4 correspond to an additional clinical *Balamuthia* isolate whose mitochondrial genome was not sequenced

Table 2 Sequencing runs and genome assembly details

Sequencing run	Raw reads	Filtered reads	Read length	Library prep
MP1 ^a	3845734	N/A	2 × 80 bp	Nextera Mate-Pair
MP2 ^a	31223454	12095792	2 × 80 bp	Nextera Mate-Pair
MP3 ^a	14121900	4397510	2 × 250 bp	Nextera Mate-Pair
PE1 ^b	213604902	29314946	2 × 250 bp	Nextera XT
Genome Assembly ^c	Contigs	N50 (bp)	Assembly Size	
Mitochondria	1	41,656	41,656	
Contigs	31,194	3,411	48,938,887	
Scaffolds	26,811	19,415	49,120,517	
Gap-fill	26,811	19,012	48,939,625	
Final assembly (minus <500 bp, contaminants)	14,699	26,144	44,270,879	

^aThree runs of a Nextera mate-pair library of *Balamuthia* strain 2046 from axenic culture sequenced on an Illumina MiSeq

^bNextera XT paired-end library of strain 2046 sequenced on an Illumina HiSeq

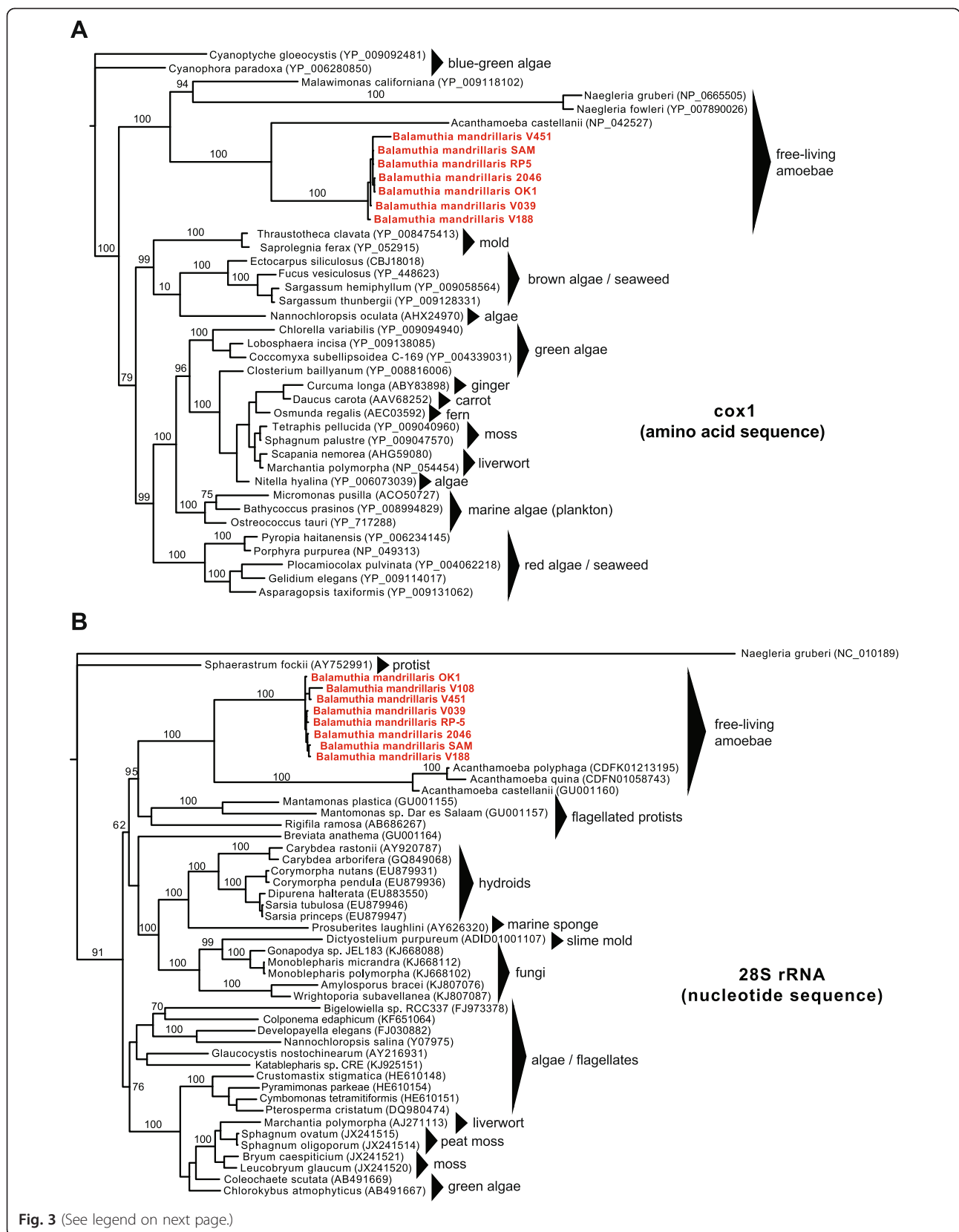
^cAssembly of mitochondrial genome from MP1 library, of whole-genome from all four libraries

dependent diabetes mellitus and celiac disease. The patient initially presented to a community emergency room with 7 days of progressive symptoms including right arm weakness, headache, vomiting, ataxia, and confusion. Her diabetes was well controlled with an insulin pump, and she did not take any additional medications. Exposure history was significant for contact with alpacas at a family farm and swimming in a freshwater pond 9 months prior. She had no international travel, sick contacts, or insect bites. She was given 10 mg dexamethasone with symptomatic improvement in her headaches, but was subsequently transferred to a tertiary care children's hospital after a computed tomography (CT) scan revealed left occipital and frontal hypodensities.

On HD 1, peripheral white blood count was 11.6×10^3 cells/ μ L (89 % neutrophils, 6 % lymphocytes, 4 % monocytes), erythrocyte sedimentation rate was 13 mm/h (normal range, 0–20 mm/h), C-reactive protein was 3 mg/dL (normal range, 0–1 mg/dL), and procalcitonin 0.05 ng/mL (normal range, 0–0.5 ng/mL). CSF analysis demonstrated 377 leukocytes/ μ L (2 % neutrophils, 53 % lymphocytes, 39 % monocytes, and 6 % eosinophils), glucose of 122 mg/dL (normal range, 40–75 mg/dL), and protein of 59 mg/dL (normal range, 12–60 mg/dL). Viral polymerase chain reaction (PCR) testing for herpes simplex virus (HSV) and bacterial cultures were negative. Magnetic resonance imaging (MRI) scan of the brain on HD 1 showed hemorrhagic lesions with surrounding edema in the superior left frontal lobe and left occipital lobe with a small focus of edema in the right cerebellum (Fig. 4a).

Given the patient's autoimmune predisposition and hemorrhagic appearance of the brain lesions, acute hemorrhagic leukoencephalitis, an autoimmune disease,

was initially suspected and intravenous methylprednisolone (1,000 mg daily) was given HD 2–5. The patient clinically deteriorated with worsening headache, increasing weakness, and altered mental status on HD 5. Repeat MRI on HD 5 demonstrated enlargement of the previous hemorrhagic lesions with interval development of multiple rim-enhancing lesions (Fig. 4b). Steroids were discontinued and broad-spectrum antimicrobial therapy with vancomycin, cefotaxime, metronidazole, amphotericin B, voriconazole, and acyclovir was initiated. On HD 6, she underwent craniotomy for brain biopsy, revealing partially necrotic white matter, and had an external ventricular drain placed. CSF wet mount and gram stain, bacterial and fungal cultures, PCR testing for HSV and varicella-zoster virus (VZV), and oligoclonal bands were negative. Pathology of the brain biopsy sample showed a hemorrhagic necrotizing process with neutrophils, tissue necrosis, vasculitis, and numerous amoebae. She was additionally started on azithromycin, sulfadiazine, pentamidine, and flucytosine on HD 7. On HD 8, she developed intracranial hypertension, cardiac arrest and died. Miltefosine had been requested and was *en route* from the CDC [3, 9, 41]; however this medication did not arrive in time to administer before the patient died. Parallel metagenomic NGS testing of CSF and brain biopsy samples around the time of death confirmed the presence of sequences from *Balamuthia mandrillaris* (see below). Autopsy was not performed according to the wishes of the family. Subsequent clinical laboratory testing of the brain biopsy sample at the US Centers for Disease Control and Prevention (CDC), using a specific PCR targeting the RNase P gene [42], confirmed the diagnosis of *Balamuthia mandrillaris* encephalitis.



(See figure on previous page.)

Fig. 3 Phylogenetic trees of the mitochondrial *cox1* protein and 28S rRNA gene reveal the close phylogenetic relationship between *Balamuthia* and *Acanthamoeba*. **a** Phylogeny of seven *Balamuthia* *cox1* amino acid sequences along with the top complete sequence hits in NCBI nr ranked by BLASTp E-score. **b** Phylogeny of seven *Balamuthia* 23S rRNA nucleotide sequences along with the top complete sequence hits in NCBI nt ranked by BLASTn E-score. Sequences were aligned using MUSCLE and a phylogenetic tree constructed using MrBayes. Branch lengths are drawn proportionally to the number of nucleotide substitutions per position, and support values are shown for each node

Identification of *Balamuthia* in CSF and brain biopsy material

Metagenomic NGS and SURPI bioinformatics analysis were used to analyze the patient's HD 6 CSF and brain biopsy for potential pathogens. Analysis of the viral portion of RNA or DNA derived reads revealed only phages or misannotated sequences (Additional file 1: Table S1), while most of the bacterial reads mapped to common skin / environmental contaminants such as *Propionibacterium* and *Staphylococcaceae* in the HD 6 CSF RNA library. In contrast, a total of 22,506 reads aligned to *Balamuthia* (Fig. 5a and 5b), of which 20,145 were taxonomically assigned to *Balamuthia mandrillaris* (Additional file 1: Table S1). These 20,145 reads comprised 79 % of all species-level non-chordate (lacking a backbone) eukaryotic reads in the HD 6 CSF dataset, and mapped to available 16S and 18S sequences of *Balamuthia mandrillaris* in the NCBI nt reference database [38]. A minority of the non-chordate eukaryotic reads aligned to *Acanthamoeba* spp. (145 reads). Reads to *Balamuthia* were also detected in the DNA (13 reads) and brain biopsy RNA libraries (9 reads). The coverage of the 16S rRNA gene in the RNA library was sufficiently high to assemble a 1,405 bp full-length contig sharing 99.9 % identity with the 2046 strain of *Balamuthia*. In the 18S locus, mapped NGS reads from the patient spanned 98.1 % of the gene and were 99.1 % identical by nucleotide. No NGS hits were detected to the RNase P gene, the only additional *Balamuthia* gene represented in the NCBI nt reference database as of August 2015.

We then sought to determine in retrospect whether earlier detection and diagnosis of *Balamuthia* infection in the case patient by NGS would have been feasible. Metagenomic NGS of a day 1 CSF sample followed by SURPI analysis using the June 2014 NCBI nt reference database generated no sequence hits to *Balamuthia* (Fig. 5b; Additional file 1: Table S1). However, repeating the analysis after adding the draft genome sequence of *Balamuthia mandrillaris* to the reference database resulted in the detection of additional *Balamuthia* reads (Fig. 5b and c). Importantly, nine species-specific DNA reads were detected from day 1 CSF (Fig. 5b, red text; Table 3). Although only two of nine putative *Balamuthia* reads had identifiable translated nucleotide identity to any protein in the NCBI nr database, one of those reads was found to share 77 % amino acid identity to the glutathione transferase protein from

Acanthamoeba castellanii, and hence most likely represented a *bona fide* hit to *Balamuthia*. These findings also indicated that the detection of *Balamuthia* reads was not due to errors in the draft genome assembly from incorporation of contaminating sequences from other organisms. Thus, detection of *Balamuthia* from the patient's day 1 sample and a more timely diagnosis by metagenomic NGS would presumably not have been made without the availability of the full draft genome as part of the reference database used for alignment.

Discussion

In this study, we describe a 'virtuous cycle' of clinical sequencing in which the continually increasing breadth of microbial sequences in reference databases improves the sensitivity and accuracy of infectious disease diagnosis, in turn driving the sequencing of additional reference strains. The assembly of the first draft reference genome for *Balamuthia* not only enhances the potential sensitivity of metagenomic NGS for detecting this pathogen, as shown here, but also provides new target sequences such as the *rps3* intron / intergenic region and 28S rRNA gene that can be leveraged for the future development of more sensitive and specific diagnostic assays. Given the lack of proven efficacious treatments for *Balamuthia* encephalitis, it is unclear whether even earlier diagnosis at HD 1 would have impacted the fulminant course of our case patient's infection. However, it has been suggested that timely intervention in cases of *Balamuthia* might lead to improved outcome [43]. In addition, promising new experimental treatments such as miltefosine [3, 9, 41], administered to the survivor infected by the sequenced 2046 strain (Vollmer and Glaser, under review), are now available.

Unbiased metagenomic NGS is a powerful approach for diagnosis of infectious disease because it does not rely on the use of targeted primers and probes, but rather, detects any and all pathogens on the basis of uniquely identifying sequence information [44]. Rapid and accurate bioinformatics algorithms [21, 45–47] and computational pipelines [17] have also been developed, with the capacity to analyze metagenomic NGS data in clinically actionable time frames. We also demonstrate here the critical role of comprehensive reference genomes in the NGS diagnostic paradigm. The availability of pathogen genomes with coverage of all clinically relevant

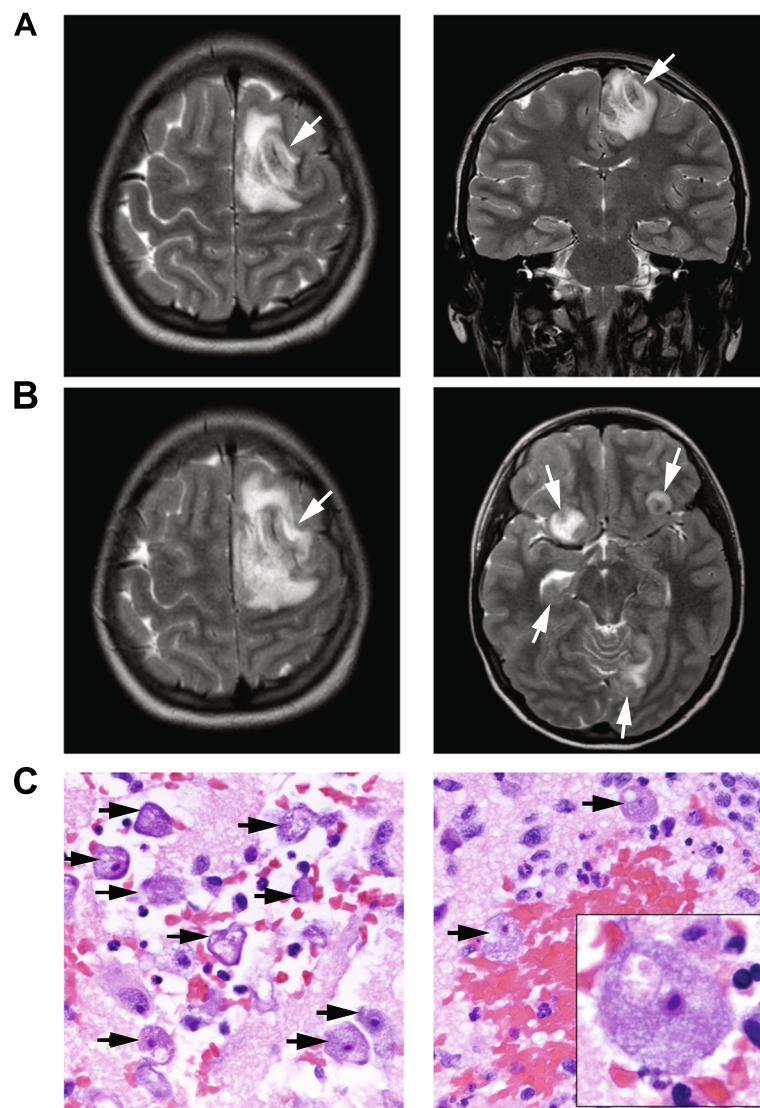


Fig. 4 MRI and histopathology from a 15-year-old patient with a fulminant acute encephalitis. **a** A hospital day (HD) 1 coronal T2-weighted MR image, demonstrating a hemorrhagic lesion with surrounding edema within the superior left frontal lobe (left panel, white arrow) and left occipital lobe (right panel, white arrow). **b** A HD 5 contrast-enhanced T1-weighted MR image, revealing enlargement of the pre-existing left frontal lobe lesion (left panel, white arrow), as well as interval development of numerous additional rim-enhancing lesions in multiple regions (right panel, white arrows). **c** 20 \times (left and right panels) and 100 \times fields of view (right panel, inset) of a brain biopsy specimen from the patient demonstrating numerous viable, large amoebae (black arrows), with abundant basophilic vacuolated cytoplasm, round central nuclei, and prominent nucleoli, consistent with *Balamuthia mandrillaris*. There were areas of extensive hemorrhagic necrosis accompanied by a polymorphic inflammatory cell infiltrate including neutrophils and eosinophils (right panel)

genotypes can maximize the utility of NGS for not only diagnosis of individual patients [15, 16], but also public health applications such as transmission dynamics [48] and outbreak investigation [20, 49]. Current limitations of metagenomic NGS for routine infectious disease diagnosis include: (1) the need to analyze millions of sequences in clinically actionable timeframes of minutes to hours (which has been at least partially addressed by rapid computational pipelines such as SURPI) [17]; (2) challenges in discriminating true pathogens from host flora or laboratory reagent

contaminants in NGS data [50, 51]; and (3) regulatory issues concerning validation of an assay targeting a potentially unlimited number of pathogens.

However, a key advantage of metagenomic NGS is that it does not rely on *a priori* suspicion of the etiology of the infectious agent. In our case patient, PAM was not clinically suspected, so specific PCR testing for *Balamuthia* was not done. Similarly, in another reported case of a woman with endophthalmitis followed by meningoencephalitis [52], PAM also was not clinically suspected, requiring

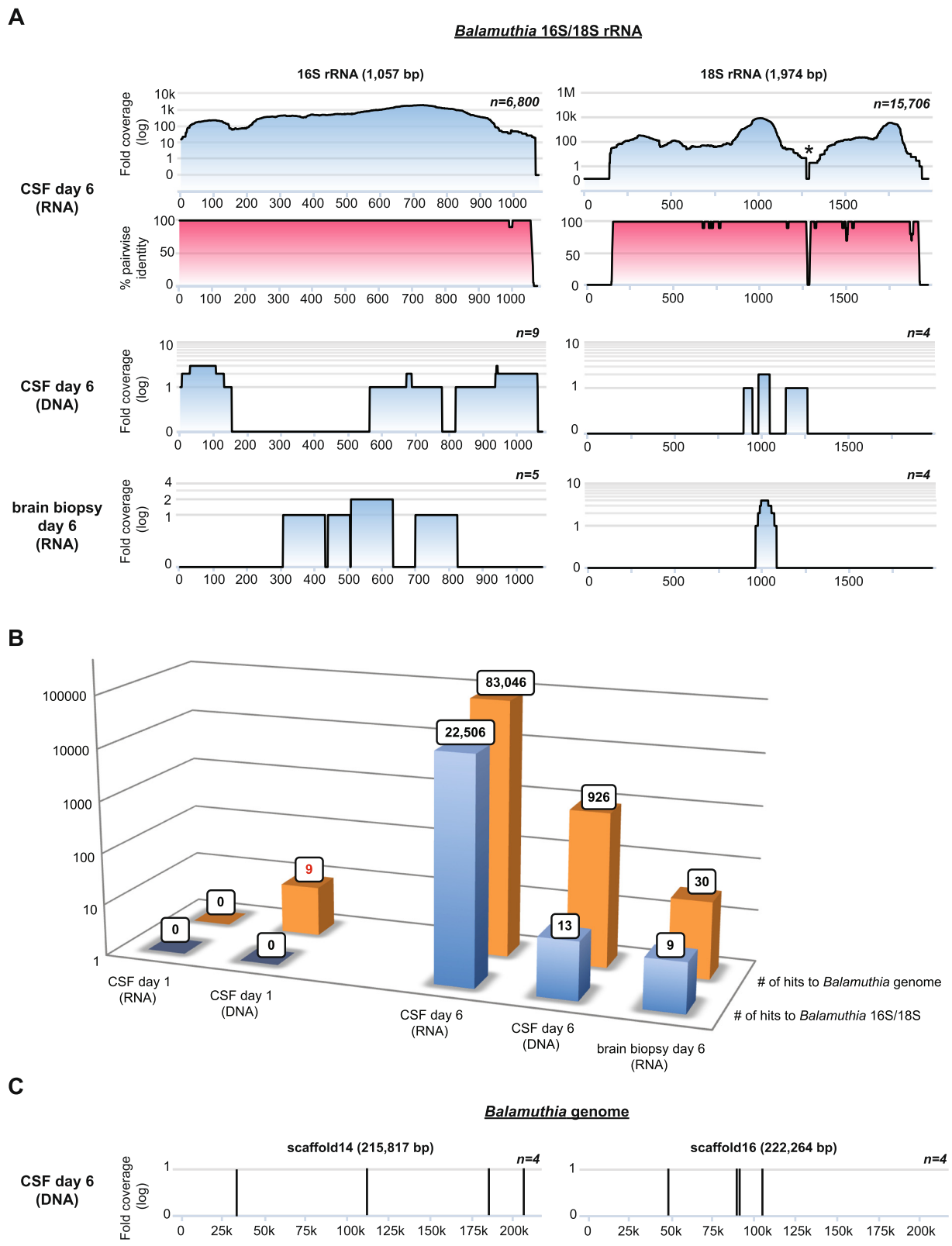


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Identification of *Balamuthia mandrillaris* infection by metagenomic next-generation sequencing (NGS). **a** Coverage maps (blue gradient) and pairwise identity plots (magenta gradient) of two of the three available sequences from *Balamuthia* (16S/18S rRNA genes) in the NCBI nt reference database as of August 2015. Shown are coverage maps corresponding to day 6 DNA and RNA libraries from CSF and a day 6 mRNA library from brain biopsy. No hits to 16S and 18S *Balamuthia* sequences were seen from day 1 samples. The asterisk denotes an area with artificially low coverage after taxonomic classification of the NGS reads due to high conservation among eukaryotic sequences (for example, human, *Balamuthia*, and so on) within that region. **b** A bar graph of the number of species-specific NGS reads aligning to *Balamuthia* 16S/18S rRNA (blue) or the *Balamuthia* genome (orange) in day 1 or day 6 samples. Note that with the availability of the newly assembled 44 Mb *Balamuthia* genome, diagnosis of *Balamuthia mandrillaris* encephalitis at day 1 would have possible by detection of nine species-specific reads (red boldface). **c** Coverage maps of two large scaffolds, approximately 216 kB and 222 kB in size, from the *Balamuthia* draft genome, showing eight out of 926 hits to *Balamuthia* in the day 6 CSF DNA library that are identified by SURPI after the draft genome sequence is added to the reference database (versus only 13 hits previously)

metagenomic NGS for diagnosis. Although broad-range PCR assays targeting the conserved bacterial 16S rRNA gene and eukaryotic 18S rRNA gene are available in select reference laboratories, there have been some concerns raised regarding sensitivity [53]. In addition, it is highly unlikely that 18S PCR sequencing would have been positive in the patient's day 1 CSF given the absence of *Balamuthia* 18S sequences in the case patient's metagenomic NGS data (Fig. 5b).

In the field of amoebic encephalitis, draft genomes are now available for *Acanthamoeba castellanii* [32], *Naegleria fowleri* [54], and *Balamuthia mandrillaris*. However, more sequencing is certainly necessary to better understand the genetic diversity of these eukaryotic pathogens. In particular, shotgun sequencing and comparative analysis of mitochondrial genomes from seven *Balamuthia* strains uncovered at least three unique lineages, one of which was comprised entirely of amoebae isolated from California, revealing that geographic differences likely exist among strains (Fig. 1b). This study also identified a unique locus in a putative rps3 intron / intergenic region of the mitochondrial genome that is an attractive target for a clinical genotyping assay

(Figs. 1c and 2). Given the rarity of the disease, it is unknown whether infection by different strains of *Balamuthia* would affect clinical course or outcome, although the future availability of routine genotyping assays could help in addressing this question.

Limitations to this study include the small number of accessible clinical samples of *Balamuthia mandrillaris* infection and assembly of a draft genome with > 14,000 scaffolds as a result of restricting the sequencing to short reads. Notably, an independently assembled draft genome of *Balamuthia mandrillaris*, comprising only 1,605 total contigs as a result of leveraging both long-read (PacBio) and short read (Illumina) technologies, has recently been published [39]. Additional messenger RNA (mRNA) sequencing of *Balamuthia* will still be needed to predict transcripts, identify splice junctions, and facilitate complete annotation of the genome.

Conclusions

We demonstrate here that the availability of pathogen reference genomes is critical for the sensitivity and success of unbiased metagenomic next-generation sequencing approaches in diagnosing infectious disease. In hindsight,

Table 3 *Balamuthia mandrillaris* reads from the day 1 patient sample

Read no.	Top NCBI NR match by BLASTx	% amino acid identity	E-value	Matched scaffold in <i>Balamuthia</i> genome	% nucleotide identity (<i>Balamuthia</i> genome)	E-value (<i>Balamuthia</i> genome)
1	None	–	–	scaffold46	96 %	5.00E-29
2	None	–	–	scaffold26764	96 %	3.00E-27
3	None	–	–	scaffold45	100 %	3.00E-36
4	None	–	–	scaffold353	98 %	2.00E-24
5	Fumarylacetoacetase (<i>Flavobacteriales bacterium</i> BRH_c54)	82 %	3.00E-05	scaffold203	100 %	3.00E-36
6	None	–	–	scaffold106	96 %	1.00E-26
7	Glutathione transferase protein (<i>Acanthamoeba castellanii</i> str. Neff)	77 %	0.022	scaffold30	99 %	8.00E-34
8	None	–	–	scaffold112	97 %	7.00E-28
9	None	–	–	scaffold511	98 %	5.00E-26

more timely and potentially actionable diagnosis at hospital day 1 in a fatal case of PAM from *Balamuthia mandrillaris* would have required the availability of the full genome sequence. Thus, in addition to revealing a significant amount of evolutionary diversity, the draft genome of *Balamuthia mandrillaris* presented here will improve the sensitivity of sequencing-based efforts for diagnosis and surveillance, and can be used to guide the development of targeted assays for genotyping and detection. The draft genome also constitutes a valuable resource for future studies investigating the biology of this eukaryotic pathogen and its etiologic role in PAM.

Additional file

Additional file 1: Table S1. SURPI clinical metagenomic results (XLSX 390 kb)

Competing interests

CYC is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center (VDDC) and receives research support in pathogen discovery from Abbott Laboratories, Inc. The other authors declare that they have no competing interests.

Authors' contributions

ALG, KM, SD, and CYC conceived of and designed the study. ALG, TD, JB, and SY performed the experiments. KM, DM, YN, MV, CAP, BKK, and SR took care of patients with *Balamuthia* infection and contributed clinical samples. KM, DM, CG, MV, CAP, BKK, SR, and CYC analyzed the clinical and epidemiological data. ALG, SNN, SF, JB, and CYC analyzed the genomic sequencing data. ALG, SNN, SF, and CYC developed and contributed software analysis tools. ALG, KM, and CYC wrote the manuscript. All authors read and approved final manuscript.

Acknowledgements

This study was supported by grants from the National Institutes of Health (NIH) R01-HL105704 (to CYC) and UL1-TR001082 (to KM and SD), the Centers for Disease Control and Prevention (CDC) Emerging Infections Program U50/CCU915546-09 (CG), University of California Discovery Grants Award #180966 from the UC Office of the President, and an Abbott Pathogen Discovery Award (CYC).

Author details

¹Department of Laboratory Medicine, University of California, 185 Berry Street, Box 0134, San Francisco, CA 94107, USA. ²UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA 94107, USA. ³Children's Hospital Colorado and University of Colorado School of Medicine, Aurora, CO, USA. ⁴California Department of Public Health, Richmond, CA, USA. ⁵Kaiser Permanente Hospital, Oakland, CA, USA. ⁶John Muir Hospital, Walnut Creek, CA, USA. ⁷Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, CA 94107, USA.

Received: 18 August 2015 Accepted: 13 October 2015

Published online: 01 December 2015

References

- Visvesvara GS. Epidemiology of infections with free-living amoebas and laboratory diagnosis of microsporidiosis. *Mt Sinai J Med.* 1993;60:283–8.
- Matin A, Siddiqui R, Jayasekera S, Khan NA. Increasing importance of *Balamuthia mandrillaris*. *Clin Microbiol Rev.* 2008;21:435–48.
- Schuster FL, Guglielmo BJ, Visvesvara GS. In-vitro activity of miltefosine and voriconazole on clinical isolates of free-living amoebas: *Balamuthia mandrillaris*, *Acanthamoeba* spp., and *Naegleria fowleri*. *J Eukaryot Microbiol.* 2006;53:121–6.
- Visvesvara GS, Martinez AJ, Schuster FL, Leitch GJ, Wallace SV, Sawyer TK, et al. Leptomyxid amoeba, a new agent of amoebic meningoencephalitis in humans and animals. *J Clin Microbiol.* 1990;28:2750–6.
- Schuster FL, Dunnebacke TH, Booton GC, Yagi S, Kohlmeier CK, Glaser C, et al. Environmental isolation of *Balamuthia mandrillaris* associated with a case of amoebic encephalitis. *J Clin Microbiol.* 2003;41:3175–80.
- Ahmad AF, Heaselgrave W, Andrew PW, Kilvington S. The in vitro efficacy of antimicrobial agents against the pathogenic free-living amoeba *Balamuthia mandrillaris*. *J Eukaryot Microbiol.* 2013;60:539–43.
- Deetz TR, Sawyer MH, Billman G, Schuster FL, Visvesvara GS. Successful treatment of *Balamuthia* amoebic encephalitis: presentation of 2 cases. *Clin Infect Dis.* 2003;37:1304–12.
- Kato H, Mitake S, Yuasa H, Hayashi S, Hara T, Matsukawa N. Successful treatment of granulomatous amoebic encephalitis with combination antimicrobial therapy. *Intern Med.* 2013;52:1977–81.
- Martinez DY, Seas C, Bravo F, Legua P, Ramos C, Cabello AM, et al. Successful treatment of *Balamuthia mandrillaris* amoebic infection with extensive neurological and cutaneous involvement. *Clin Infect Dis.* 2010;51:e7–e11.
- Schuster FL, Visvesvara GS. Axenic growth and drug sensitivity studies of *Balamuthia mandrillaris*, an agent of amoebic meningoencephalitis in humans and other animals. *J Clin Microbiol.* 1996;34:385–8.
- Qvarnstrom Y, Visvesvara GS, Sriram R, da Silva AJ. Multiplex real-time PCR assay for simultaneous detection of *Acanthamoeba* spp., *Balamuthia mandrillaris*, and *Naegleria fowleri*. *J Clin Microbiol.* 2006;44:3589–95.
- Yagi S, Booton GC, Visvesvara GS, Schuster FL. Detection of *Balamuthia* mitochondrial 16S rRNA gene DNA in clinical specimens by PCR. *J Clin Microbiol.* 2005;43:3192–7.
- Perez MT, Bush LM. Fatal amoebic encephalitis caused by *Balamuthia mandrillaris* in an immunocompetent host: a clinicopathological review of pathogenic free-living amoebae in human hosts. *Ann Diagn Pathol.* 2007;11:440–7.
- Schuster FL, Glaser C, Honarmand S, Maguire JH, Visvesvara GS. *Balamuthia* amoebic encephalitis risk, Hispanic Americans. *Emerg Infect Dis.* 2004;10:1510–2.
- Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin Infect Dis.* 2015;60:919–23.
- Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med.* 2014;370:2408–17.
- Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* 2014;24:1180–92.
- Schubert RD, Wilson MR. A tale of two approaches: how metagenomics and proteomics are shaping the future of encephalitis diagnostics. *Curr Opin Neurol.* 2015;28:283–7.
- Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet.* 2014;15:49–55.
- Greninger AL, Naccache SN, Messacar K, Clayton A, Yu G, Somasekar S, et al. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012–14): a retrospective cohort study. *Lancet Infect Dis.* 2015;15:671–82.
- Zaharia M, Bolosky W, Curtis K, Fox A, Patterson D, Shenker S, et al. Faster and more accurate sequence alignment with SNAP. *arXiv.org.* 2011;1111.5572.
- Lares-Jimenez LF, Gamez-Gutierrez RA, Lares-Villa F. Novel culture medium for the axenic growth of *Balamuthia mandrillaris*. *Diagn Microbiol Infect Dis.* 2015;82:286–8.
- O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics.* 2015;31:2035–7.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.
- Greninger AL, Chorny I, Knowles S, Ng VL, Chaturvedi V. Draft genome sequences of four NDM-1-Producing *Klebsiella pneumoniae* strains from a health care facility in northern California. *Genome Announc.* 2015;3:3.
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35:3100–8.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955–64.

28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
29. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33:W244–8.
30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:1.
31. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 2014;24:1384–95.
32. Burger G, Plante I, Lonergan KM, Gray MW. The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *J Mol Biol.* 1995;245:522–37.
33. Herman EK, Greninger AL, Visvesvara GS, Marciano-Cabral F, Dacks JB, Chiu CY. The mitochondrial genome and a 60-kb nuclear DNA segment from *Naegleria fowleri*, the causative agent of primary amoebic meningoencephalitis. *J Eukaryot Microbiol.* 2013;60:179–91.
34. Fukami H, Chen CA, Chiou CY, Knowlton N. Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of Scleractinian corals. *J Mol Evol.* 2007;64:591–600.
35. Zheng Z, Jiang K, Huang C, Mei C, Han R. *Cordyceps militaris* (Hypocreales: Cordycipitaceae): transcriptional analysis and molecular characterization of *cox1* and group I intron with putative LAGLIDADG endonuclease. *World J Microbiol Biotechnol.* 2012;28:371–80.
36. Laroche J, Bousquet J. Evolution of the mitochondrial *rps3* intron in perennial and annual angiosperms and homology to *nad5* intron 1. *Mol Biol Evol.* 1999;16:441–52.
37. Regina TM, Picardi E, Lopez L, Pesole G, Quagliariello C. A novel additional group II intron distinguishes the mitochondrial *rps3* gene in gymnosperms. *J Mol Evol.* 2005;60:196–206.
38. Booton GC, Carmichael JR, Visvesvara GS, Byers TJ, Fuerst PA. Genotyping of *Balamuthia mandrillaris* based on nuclear 18S and mitochondrial 16S rRNA genes. *Am J Trop Med Hyg.* 2003;68:65–9.
39. Detering H, Aebischer T, Dabrowski PW, Radonic A, Nitsche A, Renard BY, et al. First draft genome sequence of *Balamuthia mandrillaris*, the causative agent of amoebic encephalitis. *Genome Announc.* 2015;3:5.
40. Kordis D. A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. *Gene.* 2005;347:161–73.
41. Centers for Disease C, Prevention. Investigational drug available directly from CDC for the treatment of infections with free-living amebae. *MMWR Morb Mortal Wkly Rep.* 2013;62:666.
42. Kiderlen AF, Radam E, Lewin A. Detection of *Balamuthia mandrillaris* DNA by real-time PCR targeting the RNase P gene. *BMC Microbiol.* 2008;8:210.
43. Bakardjiev A, Azimi PH, Ashouri N, Ascher DP, Janner D, Schuster FL, et al. Amebic encephalitis caused by *Balamuthia mandrillaris*: report of four cases. *Pediatr Infect Dis J.* 2003;22:447–53.
44. Chiu CY. Viral pathogen discovery. *Curr Opin Microbiol.* 2013;16:468–78.
45. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12:59–60.
46. Freitas TA, Li PE, Scholz MB, Chain PS. Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic Acids Res.* 2015;43:e69.
47. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46.
48. Grad YH, Lipsitch M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol.* 2014;15:538.
49. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* 2009;5:e1000455.
50. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol.* 2013;87:11966–77.
51. Strong MJ, Xu G, Morici L, Splinter Bon-Durant S, Baddoo M, Lin Z, et al. Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* 2014;10:e1004437.
52. Wilson MR, Shanbhag NM, Reid MJ, Singhal NS, Gelfand JM, Sample HA, et al. Diagnosing *Balamuthia mandrillaris* encephalitis with metagenomic deep sequencing. *Ann Neurol.* 2015. doi: 10.1002/ana.24499. [Epub ahead of print]
53. Petti CA. Detection and identification of microorganisms by gene amplification and sequencing. *Clin Infect Dis.* 2007;44:1108–14.
54. Zysset-Burri DC, Muller N, Beuret C, Heller M, Schurch N, Gottstein B, et al. Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. *BMC Genomics.* 2014;15:496.
55. Bakardjiev A, Azimi PH, Ashouri N, Ascher DP, Janner D, Schuster FL, et al. Amebic encephalitis caused by *Balamuthia mandrillaris*: report of four cases. *Pediatr Infect Dis J.* 2003;22:447–453.
56. Dunnebacke TH, Schuster FL, Yagi S, Booton GC. *Balamuthia mandrillaris* from soil samples. *Microbiology.* 2004;150:2837–2842.
57. Gordon SM, Steinberg JP, DuPuis MH, Kozarsky PE, Nickerson JF, Visvesvara GS et al. Culture isolation of *Acanthamoeba* species and leptomyxid amebas from patients with amoebic meningoencephalitis, including two patients with AIDS. *Clin Infect Dis.* 1992;15:1024–1030.
58. Yang XH, Jones D, Kaepfen A, Qian J. Amebic meningoencephalitis in a 6 year-old girl caused by *Balamuthia mandrillaris*. *J Neuropathol Exp Neurol.* 2001;60:556.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

