View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by LSHTM Research Online

Stensvold, CR; Clark, CG (2016) Molecular Identification and Subtype Analysis of Blastocystis. Current protocols in microbiology, 43. 20A.2.1-20A.2.10. ISSN 1934-8525 DOI: https://doi.org/10.1002/cpmc.17

Downloaded from: http://researchonline.lshtm.ac.uk/3172106/

DOI: 10.1002/cpmc.17

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/

Title: Molecular identification and subtype analysis of *Blastocystis*

Authors: Christen R. Stensvold[1] & C. Graham Clark[2]

Contact information: [1]Department of Microbiology and Infection, Statens Serum Institut, Copenhagen, Denmark; [2]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

Significance Statement: The clinical and public health significance of the single-celled intestinal parasite *Blastocystis* remains unclear. Multiple subtypes of *Blastocystis* are known, and molecular characterization of strains in samples from patients, healthy carriers, non-human hosts, and the environment is critical to obtaining an increased understanding of *Blastocystis* epidemiology, including whether subtypes (or strains within subtypes) are potentially linked to disease. Analysis and comparison of data, however, require that standardized methods be employed for generating such data. Lately, "barcoding" (relying on small subunit ribosomal DNA sequencing) coupled with online sequence analysis has gained popularity, appearing robust and simple to use. Here, we provide the protocol for barcoding and an introduction to using online databases for small subunit ribosomal RNA gene sequence analysis.

ABSTRACT: Several methods have been used in studies aiming to unravel the molecular epidemiology of *Blastocystis*, which is one of the most common intestinal parasites in human and many non-human hosts. Such studies have the potential to add to our knowledge on *Blastocystis* transmission, host specificity, phylogeography, and clinical and public health significance, but rely on robust, standardized methods by which data can be generated and compared directly between studies. One of the most used methods is "barcoding", which involves single-round PCR amplification and sequencing of partial small subunit ribosomal RNA genes of the parasites. Recently, a publicly available online facility was developed for quick and standardized identification of subtypes (ribosomal lineages) and subtype alleles (variation within subtypes) based on sequence data obtained by barcoding PCR. Moreover, a modified barcoding approach is now available using nested PCR, which enables detection of mixed subtype infections.

Keywords: *Blastocystis* • parasite • molecular epidemiology • stool • zoonosis

INTRODUCTION

Blastocystis is one of the most common intestinal parasites in human and many non-human hosts. Based on phylogenetic analysis of ribosomal RNA genes, at least 17 subtypes are known from mammals and birds alone, and these are arguably separate species (Alfellani et al., 2013a). Within each subtype, a variable number of alleles (Stensvold et al., 2012b) have been identified, and analysis of subtype alleles provides a more discriminating tool for investigating the molecular epidemiology of the parasite, including transmission, host specificity, and phylogeography; such tools may also be critical to identifying the clinical and public health significance of the parasite, which are currently controversial.

Several methods have been used for molecular characterization of *Blastocystis*. The two most commonly used methods are i) PCR of unknown DNA targets using subtype-specific primers and ii) PCR amplification of small subunit ribosomal RNA (SSU rRNA) genes followed by Sanger sequencing, one of which is the "barcoding" method (Stensvold CR. 2013; Scicluna et al., 2006). Barcoding is currently the method of choice due to the sensitivity and specificity of this approach (Stensvold CR. 2013). Sequences obtained by barcoding can also be queried using a free-to-use online facility (www.pubmlst.org/blastocystis) in order to identify subtype alleles. Finally, a modified version of the barcoding approach was recently introduced to facilitate better detection and differentiation of mixed subtype infections (Scanlan et al., 2015).

STRATEGIC PLANNING

Barcoding PCR uses genomic DNA extracted directly from fresh, frozen or ethanol-fixed stool or from stool culture (a protocol on *Blastocystis* culture is available XXX). DNA extraction from stool may be accomplished in a variety of ways using more or less automated methods. Contrary to the situation for some protozoa, the extraction of DNA from *Blastocystis* is generally straightforward with no difficulties in terms of lysing of the cells. However, caution should be taken to choose a method that reduces the amount of inhibitory substances as much as possible. Many researchers have used QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) for this purpose, because the kit includes an absorptive resin to remove common inhibitors of the PCR reaction found in stool specimens and DNA can be eluted in 100 µL of AE buffer (Qiagen) to increase its concentration. Automated solutions that also appear to work well include the NucliSENS® easyMag® protocol from bioMérieux.

BASIC PROTOCOL 1

ETHANOL FIXATION OF STOOL PRIOR TO GENOMIC DNA EXTRACTION

Due to the amount of hands-on time involved, ethanol-fixation of stool is recommended only in those cases where it is necessary to store stool samples for a while (weeks/months) prior to DNA extraction, and when freezing of samples at -18 °C or below is not an option. In order to

save time and resources, DNA should be extracted directly from fresh or frozen stool. DNA extracted from formalin-preserved samples will rarely work as a PCR template and is therefore not recommended. So far, there are no data on potential differences in the applicability of different primer pairs to DNA extracted from fresh/frozen stool vs. stool preserved in ethanol.

Materials

Fecal samples of interest
Eppendorf tubes (1.5 mL or 2 mL)
Rack for the Eppendorf tubes
Wooden sticks or similar for transferal of stool from test vial to Eppendorf tube
Ethanol
Phosphate-buffered saline (PBS)
Cupboard/fridge
Centrifuge
Pipette and pipette tips

1. Transfer about 200 µg of fresh stool to a new, clean Eppendorf tube. (Note: It is essential that the Eppendorf tube has not been used before. It is also essential to use a new wooden stick for each individual stool sample so as to prevent contamination from one stool sample to another).

2. To each tube, add three volumes of ethanol to one volume of stool, ensuring a final ethanol concentration of about 70% in the sample. Vortex the tube for 5–10 min, so as to ensure sufficient homogenization and successful preservation of the stool.

3. Place the tubes in a rack and place in a cupboard away from light or in a refrigerator until DNA extraction.

4. Immediately prior to DNA extraction, wash the contents of each tube in PBS three times: First, vortex the sample and then centrifuge for 10 min at 2,000 x *g* in a benchtop microfuge; open the lid, remove the supernatant by pipetting, and add 1 mL or 1.5 mL of PBS if using a 1.5 mL or 2.0 mL Eppendorf tube, respectively. Repeat this step twice without filling up with PBS the final time.

   The PBS wash is to remove the ethanol as much as possible before purification of the DNA; ethanol could affect the extraction process as it affects the binding to the column of the small molecule contaminants as well as the DNA. The last step is without PBS to enable the removal of as much liquid as possible before starting the extraction - the salt in the PBS and any residual ethanol could affect the conditions if it is not all removed. There is always a small amount of liquid after the final spin even though the supernatant was already removed after the previous spin.

5. Proceed to DNA extraction.

BASIC PROTOCOL 2

MOLECULAR DETECTION OF *BLASTOCYSTIS* INFECTION BY TAQ-MAN real-time PCR

For reasons explained elsewhere in this protocol, barcoding should only be performed on genomic DNA extracted from

1) *Blastocystis*-positive cultures

2) Stool identified as *Blastocystis*-positive by one of various diagnostic methods; preferably real-time PCR.

The following Taq-man-based real-time PCR can be used to identify *Blastocystis*-positive stool. It has been shown to have a sensitivity of 2.5 parasites per reaction (Stensvold et al., 2012a) and is able to detect all subtypes found in humans so far (ST1–ST9). Primers and probe target the SSU rRNA gene and has a specificity of 100%. The PCR will not give information on subtype. An internal process control (IPC) can be included (for IPC development, primers and probe, please see Stensvold et al., 2012a).

Materials

For PCR reactions made in 50-µL volumes:

- 1 µM each of the two primers, Blasto FWD F5: 5'-GGTCCGGTGAACACTTTGGATTT-3' and Blasto R F2: 5'-CCTACGGAAACCTTGTTACGACTTCA-3'
- 300 nM of the probe, FAM-TCGTGTAAATCTTACCATTTAGAGGA-MGBNFQ
- 5 U/µL Platinum *Taq* polymerase (Invitrogen, Taastrup, Denmark) in 20 mM Tris-HCl (pH 8.0), 40 mM sodium chloride, 2 mM sodium phosphate, 0.1 mM dithiothreitol (DTT), Stabilizers, 50% (vol/vol) glycerol
- 10X PCR buffer minus $MgCl_2$ (200 mM Tris [pH 8.4], 500 mM KCl) (Invitrogen, Taastrup, Denmark)
- 5 mM $MgCl_2$
- dUTP mix (12.5 mM dUTP, 50 mM dGTP, 50 mM dATP, 50 mM dCTP)
- 5 µL of the appropriate dilution of IPC (if applicable)
- 50% glycerol
- 5 µL of template
- Water enough to produce a final volume of 50 µL.

Samples can be processed on an ABI 7500 real-time PCR system instrument with a 96-well block (e.g., Applied Biosystems, Nærum, Denmark). The PCR profile consists of 50 °C for 1 s, 95 °C for 2 min, and 50 cycles of denaturation at 95°C for 15 s followed by annealing and extension at 60°C for 1 min.

Interpretation of Real-Time PCR Results

An real-time PCR amplification plot is shown in Fig. 1. The controls have cycle threshold (Ct)-values of 22 and 26, respectively. Four patient samples have also been indicated; two with Ct-values of about 30, and two weakly positive ones with Ct values of 37–38. The latter two samples might represent negative samples, depending on the *Blastocystis* DNA concentration and stability of the positive controls. Hence, controls will need to be run to determine the exact sensitivity and specificity (for more information, please look up Stensvold et al., 2012a).

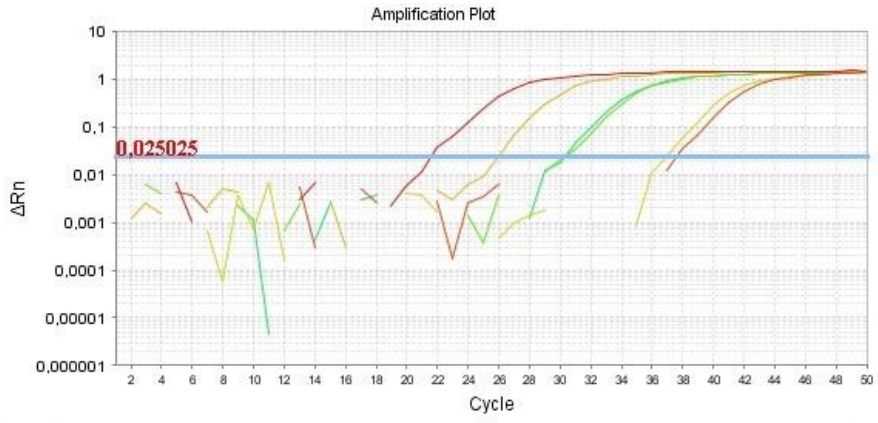

Figure. 1. Amplification plot of a *Blastocystis* real-time PCR. The plot shows the results for two controls (the two left-most curves; Ct-values 22 and 26, respectively), and four patient samples, two of which (the two right-most curves; Ct-values around 37–38) might be interpreted as negative. Rn is the fluorescence of the reporter dye divided by the fluorescence of the reference dye (ROX in this case); ΔRn is Rn at a given cycle minus baseline Rn. Here, the cutoff (blue bar) for a positive signal was set at the beginning of the exponential phase (represented by the linear part of the amplification curve) at a ΔRn of 0.025.

BASIC PROTOCOL 3

BARCODING PCR-DNA SEQUENCING AND SUBTYPING

Barcoding PCR relies on the protocol originally described by Scicluna et al (2006) and involves amplification of partial SSU rRNA genes. When used on genomic DNA from stool, amplification of non-*Blastocystis* DNA may occur, especially so in the absence of *Blastocystis*-specific DNA in the sample. Hence, the primers used in the PCR reaction (Table 1) are not fully specific to *Blastocystis*, and typically, they may amplify fungal DNA, especially in the absence of *Blastocystis* DNA and in non-human samples. This is also one of the reasons why this PCR should not be used diagnostically. Hence, it is recommended that barcoding be used only for molecular characterization of already known positive samples—not for screening; state-of-the-

art screening uses real-time PCR as described by Stensvold et al., 2012a. Used on DNA extracted from *Blastocystis*-positive cultures, the primers are usually very specific.

Materials

PCR is made in 25-uL reactions.

Template DNA (2 µL per reaction; see Basic Protocol 1)
Forward primer (RD5, 1 µL per reaction [stock concentration, 10 µg/µL]; Table 1)
Reverse primer (BhRDr; 1 µL per reaction [stock concentration, 10 µg/µL]; Table 1)
DNase/RNase-free distilled water
PCR tubes (PCR can be performed in plates as well)
PCR master mix (we recommend using commercially available ready-made mixes such as e.g., Extract-N-Amp™ PCR ReadyMix™ [cat. no. E3004; Sigma–Aldrich])
Thermocycler
Agarose gels (1.5% or 2% according to experience)
Gel equipment and gel analyzer
PCR purification kit (e.g., QIAquick PCR Purification Kit [250] cat. no. 28106)

Table 1. Original barcoding primers (Scicluna et al., 2006).

Primer name	Orientation	Target gene	DNA sequence (5´–3´)	Length (bp)	Melting temperature*	GC content
RD5	Forward	SSU rRNA (18S)	ATCTGGTTGATCCTGCCAGT	20	63.5 °C	50%
BhRDr	Reverse	SSU rRNA (18S)	GAGCTTTTTAACTGCAACAACG	22	62.6 °C	40.9%

* According to the algorithm available at https://www.sigmaaldrich.com/configurator/servlet/DesignTool?prod_type=STANDARD#Calcs_0

1. Aliquot 12.5 µL of PCR master mix to each tube (or well if PCR is made in plates).
2. To each tube (well), add 1 µL of each primer.
3. To each tube (well), add 8.5 µL of DNase/RNase-free distilled water.

4. To each tube (well), add 2 µL of DNA template (template may need to be diluted 1:10 depending on DNA concentration).
5. Close the lids (or seal the plate), place tubes/plate in the thermocycler, and use the program listed in Table 2 for PCR.

Table 2. PCR program for barcoding PCR.

Process	Time	Temperature	
Initial denaturation	4 min	94 °C	
Denaturation	15 s	95 °C	X 30 cycles
Annealing	15 s	60 °C	
Elongation	30 s	72 °C	
Final extension	5 min	72 °C	
	"Infinity"	4 °C	

6. Load a total of 5 µL of each PCR product into a well on a 1.5% or 2% agarose gel.
7. Samples producing bands corresponding to 550–600 bp are chosen for PCR product purification.
8. PCR products can be purified with a commercial kit (e.g., QIAquick PCR Purification Kit [250] cat. no. 28106) according to the recommendations of the manufacturer and further processed for Sanger sequencing. The BhRDr primer works well for sequencing, and unidirectional sequencing usually suffices.

BASIC PROTOCOL 3

SEQUENCE ANALYSIS

Usually, the output of Sanger sequences includes ab1 files that can be visualized using the freeware FinchTV from Geospiza that can be downloaded at http://www.geospiza.com/ftvdlinfo.html. Alternatively, the ab1 file can be uploaded into a variety of software for visualization and editing, such as the Lasergene DNA Star suite of programs (see http://www.dnastar.com) or Staden Package (http://staden.sourceforge.net/). Edited ab1 files can be downloaded as fasta files and then queried using the NCBI database one

by one, or more recently, in batches; also here, the option of submitting a whole batch of barcoding sequences for *Blastocystis* subtype analysis is possible using the below designated online tool. Apart from identifying the subtype pertaining to each sequence, an allele is assigned, which enables standardized analysis of intra-subtype genetic variability.

Materials

DNA sequence files in fasta format
Online access to www.pubmlst.org/blastocystis

1. Single sequence query

    a. Open the edited ab1 file, copy the sequence, and paste it into a Notepad or a Word file.

    b. Ensure that the sequence is in fasta format; the format should look like this:

    >sampleID
    CAGTAGTCATACGCTCGTCTCAAAGATTAAGCCATGCATGTGTAAGTATAAATATTTAACTTTGAAAC
    TGCGAATGGCTCATTATATCAGTTATAGTTTATTTGATGAAAAGTACTACTTGGATAACCGTAGTAAT
    TCTAGAGCTAATACATGATAAAATCCTCGACTTTGAAGAGGTGTATTTATTAGAATGAAACCAATAG
    ACTTCGGTCTGTTTGTGAGTAATAATAACTAATCGTATCGCATGCTTAGGTAGCGATATGTCTTTCAA
    GTTTCTGCCCTATCAGCTTTGGATGGTAGTGTATTGGACTACCATGGCAGTAACGGGTAACGAAGAA
    TTTGGGTTCGATTTCGGAGAGGGAGCCTGAGAGATGGCTACCACATCCAAGGAAGGCAGCAGGCG
    CGTAAATTACCCAATCCTGACATAGGGAGGTAGTGACAATAAATCACAATGCGGAACTATTAGTTTT
    GCAATTGGATTGAGAACAATGTACAAATGTTATCGATAAACAATTGGAGGGCAA

    c. Copy the sequence and paste it into the query box available at http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_blastocystis_seqdef&page=sequenceQuery

    d. Press the "Submit" button.

    e. The output of the query is displayed in Fig. 2.

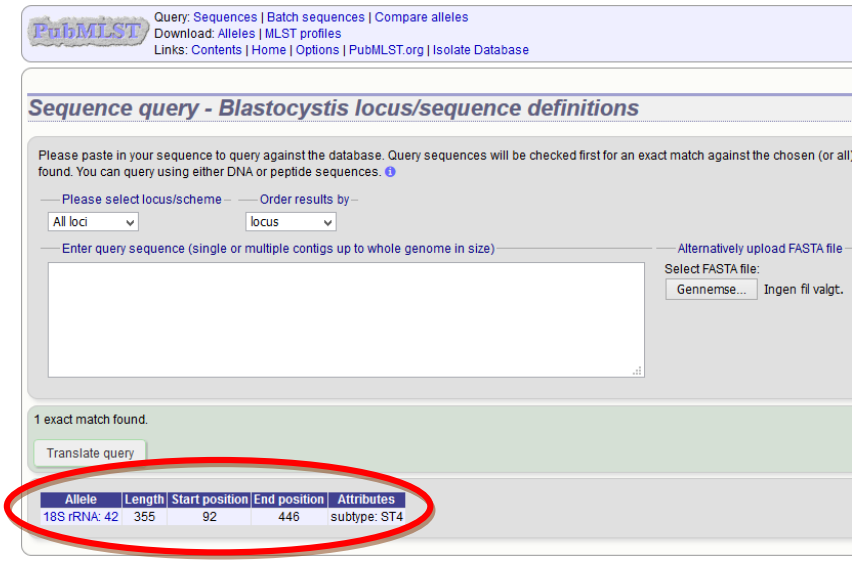

Fig. 2. Example of an output following the query of a single fasta file with 100% similarity to an existing sequence.

f. By clicking the hyperlink “18S rRNA: 42”, you will learn that the sequence is a ST4 sequence, and it is assigned to allele 42 (Fig. 3).

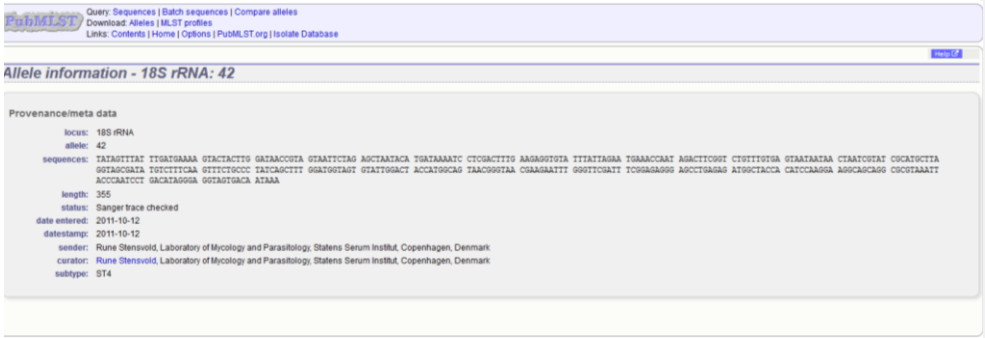

Fig. 3. Information on reference sequence corresponding to the sequence subject to query in Fig. 2.

2. Batch sequence query

   a. Compile fasta files in a single Notepad or Word document, and copy all fasta files to the query box available at http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_blastocystis_seqdef&page=batchSequenceQuery

   b. Press the "Submit" button.

   c. The output will include a list of results, and it is possible to go through each of the results by clicking the individual hyperlinks (please refer to 1e and 1f for more information).

   Please note that the query is independent of sequence direction. Hence, if you have a sequence that has been obtained using BhRDr primers, there is no need to reverse-complement it prior to the sequence query. However, if a batch query is performed, it is advisable to use fasta files with sequences oriented in the 5'–3' direction.

REAGENTS AND SOLUTIONS

COMMENTARY

Background Information

Studies of the molecular epidemiology of *Blastocystis* serve to clarify patterns of transmission and differences in the distribution of subtypes, whether these be differences among host groups or differences across geographical regions. Moreover, the identification of subtypes and alleles is useful for clinical studies aiming at identifying any subtype- or strain-related differences in the clinical presentation of *Blastocystis* carrier state. Subtyping is, however, not a means to distinguishing colonization from infection, and so far, no subtype has been consistently linked to intestinal disease. Meanwhile, subtypes appear to differ remarkably in terms of host specificity and geographical distribution. A subtype such as ST4 is common only in Europeans, while subtypes 1–3 are common in humans all over the world (Alfellani et al., 2013b). On the other hand, subtypes other than those belonging to ST1–ST4 appear to be relatively rare in humans (Alfellani et al., 2013b). ST3 is common in both human and non-human primates; however, ST3 allele analysis has demonstrated that ST3 strains found in humans and non-human primates are different (Stensvold et al., 2012b).

Commented [RUN1]: Acta Trop. 2013 Apr;126(1):11-8. doi: 10.1016/j.actatropica.2012.12.011. Epub 2013 Jan 3.
Variable geographic distribution of Blastocystis subtypes and its potential implications.
Alfellani MA[1], **Stensvold** CR, Vidal-Lapiedra A, Onuoha ES, Fagbenro-Beyioku AF, Clark CG.

Commented [RUN2]: There are two Alfellani et al., 2013 publications, so they should have an a or a b attached.

Commented [RUN3]: Protist. 2013 Jul;164(4):497-509. doi: 10.1016/j.protis.2013.05.003. Epub 2013 Jun 14.
Genetic diversity of blastocystis in livestock and zoo animals.
Alfellani MA[1], Taner-Mulla D, Jacob AS, Imeede CA, Yoshikawa H, **Stensvold** CR, Clark CG.

Commented [RUN4]: Infect Genet Evol. 2012 Mar;12(2):263-73. doi: 10.1016/j.meegid.2011.11.002. Epub 2011 Nov 17.
Levels of genetic diversity vary dramatically between Blastocystis subtypes.
Stensvold CR[1], **Alfellani** M, Clark CG.

Troubleshooting

Interpretation of results

If a sequence has 100% similarity to a sequences in the database, the output of the query will be like the one seen in Fig. 2. However, if there is one or more mismatches, e.g.:

>test_with_two_mismatches
CAGTAGTCATACGCTCGTCTCAAAGATTAAGCCATGCATGTGTAAGTATAAATATTTAACTTTGAAACTGCGAATG
GCTCATTATATCAGTTATAGTTTATTTGATGAAAAGTACTACTTGGATAACCGTAGTAATTCTAGAGCTAATACATG
ATAAAATCCTCGACTTTGAAGAGGTGTATTTATTAGAATGAAACCAATAGACTTCGGTCTGTTTGTGAGTAATAAT
ATCTAATCGTATCGCATGCTTAGGTAGCGATATGTCTTTCAAGTTTCTGCCCTATCAGCTTTGGATGGTAGTGTATT
GGACTACCATGGCAGTAACGGGTAACGAAGAATTTGGGTTCGATTTCGGAGAGGGAGCCTGAGAGATGGCTACC
ACATCCAAGCAAGGCAGCAGGCGCGTAAATTACCCAATCCTGACATAGGGAGGTAGTGACAATAAATCACAATGC
GGAACTATTAGTTTTGCAATTGGATTGAGAACAATGTACAAATGTTATCGATAAACAATTGGAGGGCAA

the output may look something like this:

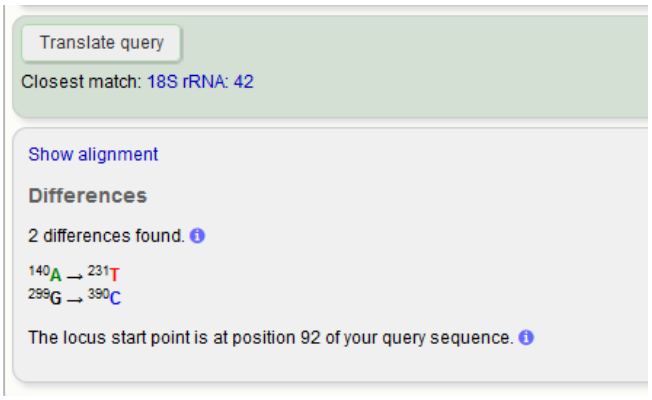

Fig. 4. Example of basepair mismatches as they are displayed when a sequence fails to show 100% identity to a sequence in the database.

Here, there is the option of having the alignment of the sequences displayed (Show alignment [Fig. 5]):

Translate query

Closest match: 18S rRNA: 42

```
                10        20        30        40        50        60        70        80        90       100
Query CAGTAGTCATACGCTCGTCTCAAAGATTAAGCCATGCATGTGTAAGTATAAATATTTAACTTTGAAACTGCGAATGGCTCATTATATCAGTTATAGTTTA
  Ref ----------------------------------------------------------------------------------------------------

               110       120       130       140       150       160       170       180       190       200
Query TTTGATGAAAAGTACTACTTGGATAACCGTAGTAATTCTAGAGCTAATACATGATAAAATCCTCGACTTTGAAGAGGTGTATTTATTAGAATGAAACCAA
  Ref ....................................................................................................

               210       220       230       240       250       260       270       280       290       300
Query TAGACTTCGGTCTGTTTGTGAGTAATAATATCTAATCGTATCGCATGCTTAGGTAGCGATATGTCTTTCAAGTTTCTGCCCTATCAGCTTTGGATGGTAG
  Ref ...............................A....................................................................

               310       320       330       340       350       360       370       380       390       400
Query TGTATTGGACTACCATGGCAGTAACGGGTAACGAAGAATTTGGGTTCGATTTCGGAGAGGGAGCCTGAGAGATGGCTACCACATCCAAGCAAGGCAGCAG
  Ref ........................................................................................G...........

               410       420       430       440       450       460       470       480       490       500
Query GCGCGTAAATTACCCAATCCTGACATAGGGAGGTAGTGACAATAAATCACAATGCGGAACTATTAGTTTTGCAATTGGATTGAGAACAATGTACAAATGT
  Ref ............................................-------------------------------------------------------

               510       520
Query TATCGATAAACAATTGGAGGGCAA
  Ref ------------------------
```

Differences

2 differences found.

$^{140}A \rightarrow {}^{231}T$
$^{299}G \rightarrow {}^{390}C$

The locus start point is at position 92 of your query sequence.

Fig. 5. Alignment of a query sequence with its closest matching reference sequence. A hyphen indicates that the sequence is not covered in the reference sequence. A dot indicates consensus. In those cases where mismatches exist, the differing base (in this case 'A' at position 231 and 'G' at position 390) is indicated.

So, in the event that there is a sequence with a couple of mismatches, the original ab1 sequence file should be scrutinized to ensure that no bases were mis-called. If there is no doubt that the sequence is correct, the sequence probably represents a new allele within a known subtype, and the sequence should be submitted along with the original sequence file to the curator of www.pubmlst.org/blastocystis who will scrutinize the files before uploading the new sequence to the database, assigning a new allele number to this particular sequence.

Sequences that do not represent *Blastocystis*

Occasionally, non-*Blastocystis* DNA sequences will align to a sequence in the database. For instance, sequences belonging to yeasts such as *Saccharomyces* may align to one of the *Blastocystis* sequences (Fig. 6). However, the number of basepair mismatches will be high, and if a high number of mismatches are seen, the sequence should be subject to a BLAST query against the reference sequence database at the NCBI website to identify the organism of origin. http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

>Saccharomyces
CATTAAATCAGTTATCGTTTATTTGATAGTTCCTTTACTACATGGTATAACTGTGGTAATTCTAGAGCTA
ATACATGCTTAAAATCTCGACCCTTTGGAAGAGATGTATTTATTAGATAAAAAATCAATGTCTTCGGACT
CTTTGATGATTCATAATAACTTTTCGAATCGCATGGCCTTGTGCTGGCGATGGTTCATTCAAATTTCTGC
CCTATCAACTTTCGATGGTAGGATAGTGGCCTACCATGGTTTCAACGGGTAACGGGGAATAAGGGTTCGA
TTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATC
CTAATTCAGGGAGGTAGTGACAATAAATAACGATACAGGGCCCATTCGGGTCTTGTAATTGGAATGAGTA
CAATGTAAATACCTTAACGAGGAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCT
CCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACTTTGGGC

```
Translate query
Closest match: 18S rRNA: 42

An alignment between your query and the returned reference sequence is shown rather than a simple list of differences because there are gaps in the alignment.

               10        20        30        40        50        60        70        80        90       100
Query CATTAAATCAGTTATCGTTTATTTGATAGTTCCTTTACTACATGGTATAACTGTGGTAATTCTAGAGCTAATACATGCTTAAAATCTCGACCCTTTGGAA
  Ref ------------...A...........GAAAAG.--.....T...-.....C..A.......................A.A...TC......---..T...

              110       120       130       140       150       160       170       180       190
Query GAGATGTATTTATTAGATAAAAAATCAAT-GTCTTCGGACTCTTTGATGATTCATAATAACTTTTCGAATCGCATGGCCTTGTGCTGGCGATGGTTCATT
  Ref ...G.............-.TG...C....A.A......T..G....-...G.A.........AA...T........--...A-.G.A.....ATG..T..

    200       210       220       230       240       250       260       270       280       290
Query CAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGTGGCCTACCATGGTTTCAACGGGTAACGGGGAATAAGGGTTCGATTCCGGAGAGGGAGCCTGA
  Ref ...G.............G....G........TG..T...A.........CAGT...........AA....TT..........T................

    300       310       320       330       340       350       360       370       380       390
Query GAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCTAATTCAGGGAGGTAGTGACAATAAATAACGATACAGGGCCCATTCGG
  Ref ..G.T................................T...............G.CAT.....................----------------------

    400       410       420       430       440       450       460       470       480       490
Query GTCTTGTAATTGGAATGAGTACAATGTAAATACCTTAACGAGGAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCG
  Ref ------------------------------------------------------------------------------------------------

    500       510       520       530       540
Query TATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAACTTTGGGC
  Ref ---------------------------------------------
```

Fig. 6. The *Saccharomyces* sequence displayed above was queried against the *Blastocystis* reference sequence database. A high level of polymorphism is seen, and although the sequence aligns with a

reference sequence, a BLAST query using the NCBI database would reveal that the sequence belongs to *Saccharomyces*.

Applicability

Barcoding is recommended only as a tool for subtyping of *Blastocystis*; i.e., for molecular characterization. Screening for *Blastocystis* by DNA-based methods could include the use of the real-time PCR described in 2012 (Stensvold et al., 2012a).

ACKNOWLEDGEMENT

This work was funded by Marie Curie Actions (Marie Curie Reintegration Grant for the proposal GUT18S [Call: FP7-PEOPLE-2012-CIG; Grant ID: 321614])

LITERATURE CITED

Alfellani MA, Taner-Mulla D, Jacob AS, Imeede CA, Yoshikawa H, Stensvold CR, Clark CG. 2013a. Genetic diversity of *Blastocystis* in livestock and zoo animals. Protist **164:**497-509.

Alfellani MA, Stensvold CR, Vidal-Lapiedra A, Onuoha ES, Fagbenro-Beyioku AF, Clark CG. 2013b. Variable geographic distribution of *Blastocystis* subtypes and its potential implications. Acta Trop. **126**:11-18.

Scanlan PD, Stensvold CR, Cotter PD. 2015. Development and Application of a *Blastocystis* Subtype-Specific PCR Assay Reveals that Mixed-Subtype Infections Are Common in a Healthy Human Population. Appl Environ Microbiol **81:**4071-4076.

Scicluna SM, Tawari B, Clark CG. 2006. DNA barcoding of *Blastocystis*. Protist **157:**77-85.

Stensvold CR. 2013. Comparison of sequencing (barcode region) and sequence-tagged-site PCR for *Blastocystis* subtyping. J Clin Microbiol **51:**190-194.

Stensvold CR, Ahmed UN, Andersen LO, Nielsen HV. 2012a. Development and evaluation of a genus-specific, probe-based, internal process controlled real-time PCR assay for sensitive and specific detection of *Blastocystis*. J Clin Microbiol **50:**1847-1851.

Stensvold CR, Alfellani M, Clark CG. 2012b. Levels of genetic diversity vary dramatically between *Blastocystis* subtypes. Infect Genet Evol **12:**263-273.

INTERNET RESOURCES

www.pubmlst.org/blastocystis

FIGURE LEGENDS

Fig. 1. Example of an output following the query of a single fasta file with 100% similarity to an existing sequence.

Fig. 2. Information on reference sequence corresponding to the sequence subject to query in Fig. 1.

Fig. 3. Example of basepair mismatches as they are displayed when a sequence fails to show 100% identity to a sequence in the database.

Fig. 4. Alignment of a query sequence with its closest matching reference sequence. A hyphen indicates that the sequence is not covered in the reference sequence. A dot indicates consensus. In those cases where mismatches exist, the differing base (in this case 'A' at position 231 and 'G' at position 390) is indicated.

Fig. 5. The *Saccharomyces* sequence displayed above was queried against the *Blastocystis* reference sequence database. A high level of polymorphism is seen, and although the sequence aligns with a reference sequence, a BLAST query using the NCBI database would reveal that the sequence belongs to *Saccharomyces*.