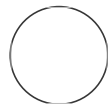


SEP 29, 2023

Nanopore Sequencing for Apicomplexan Haemoparasite 18S rRNA Gene Metabarcoding

hugginsl¹

¹University of Melbourne



hugginsl

ABSTRACT

Here we report the methodology for a novel nanopore sequencing based method for the unbiased characterisation of apicomplexan haemoparasites from mammalian blood, validated on field samples from canines. Through targeting of the apicomplexan 18S rRNA gene the metabarcode of pathogens from this phylum can be elucidated, providing detailed information on haemoparasite single and coinfections with species-level taxonomic classification.

OPEN ACCESS



DOI:

dx.doi.org/10.17504/protocols.io.6qpvr4nqpgmk/v1

Protocol Citation: hugginsl 2023. Nanopore Sequencing for Apicomplexan Haemoparasite 18S rRNA Gene Metabarcoding.

protocols.io

<https://dx.doi.org/10.17504/protocols.io.6qpvr4nqpgmk/v1>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working

Created: Apr 13, 2023

DNA Extraction

- 1 Whole blood was collected via venepuncture into ethylenediaminetetraacetic acid (EDTA) tubes. Samples were temporarily kept at 4 °C in the field before being couriered at -20 °C to the processing laboratory where 200 µl of thawed whole blood was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) using the manufacturer's protocol with a 30-minute proteinase K digestion at 56 °C and two final elution steps in 50 µl total eluent. Extracted DNA was kept at -20°C until use. Haemoparasite positive control samples, that were used for validation of our method had their DNA extracted in a similar way.

All DNA extracts were quantified using a QubitTM4 Fluorometer (Thermo Fisher Scientific, Massachusetts, USA) using the dsDNA HS assay kit.

Nanopore sequencing for apicomplexan 18S rRNA gene metab...

- 2 To conduct the library preparation for metabarcoding of the apicomplexan 18S rRNA gene on the MinION Mk1B sequencer (Oxford Nanopore Technologies, Oxford, UK) we utilised the PCR Barcoding Expansion 1-96 (EXP-PBC096) with ONT' Ligation Sequencing Kit (SQK-LSK110). The protocol followed was 'Ligation sequencing amplicons - PCR barcoding (SQK-LSK110 with EXP-PBC096)' version: PBAC96_9114_v110_revK_10Nov2020, with some modifications to improve yield.

For the first step PCR amplification 20 µl PCRs were conducted using 10 µl of [LongAmp® Hot Start Taq 2× Master Mix](#) (New England Biolabs, Massachusetts, USA) 7 µl [Ambion nuclease-free water](#) (Life Technologies, California, USA) from hereon referred to as water, 1 µl of forward primer APICO_F_Mod_ONT, 1 µl of reverse primer Api_Illum_New-Rev_ONT and 1 µl of genomic blood extracted DNA from Cambodian dogs. The original primer sequences were modified to include the addition of ONT adapter sequences (underlined) that permit the addition of DNA barcodes in a subsequent secondary PCR reaction, hence the primer sequences were APICO_F_Mod_ONT: 5'-TTTCTGTTGGTGCTGATATTGCGCCAGTAGTCATATGCTTGTCT-3' and Api_Illum_New-Rev_ONT: 5'-ACTTGCCGTGTCGCTCTATCTTCCTGTTATTGCCTYAACTTCCTYG-3'.

PCRs were then conducted on a T100TM Thermal Cycler (Bio-Rad, California, USA) using the following conditions: 1 cycle of 94 °C for 1 min, 20 cycles of 94 °C for 30 s, 56 °C for 45 s and 65 °C for 1 min 20 s, with a final extension of 65 °C for 10 min. Separate and different physical laboratory areas were utilised for DNA extraction, pre-PCR and post-PCR experiments with all first-step PCRs prepared in a PCR hood under sterile conditions with filter tips, following UV sterilisation of the workspace.

- 3 PCR product was then added to a 96-well plate and cleaned using a 1× ratio of NucleoMag NGS Clean-up and Size Select Beads (Macherey-Nagel, Duren, Germany) with a 15 min incubation on a HulaMixer (Thermo Fisher Scientific) and two washes with freshly made 75% ethanol, followed by a final elution in 20 µl of Ambion Nuclease-Free Water.
- 4 Next, second step PCRs were conducted to add ONT barcodes to each samples' amplicon and thereby permit multiplexing of up to 96 samples onto a flow cell. These secondary PCRs were 50 µl reactions utilising 25 µl of LongAmp® Taq 2× Master Mix (New England Biolabs), 20 µl of cleaned PCR product from the first PCR reaction, 4 µl of water and 1 µl of a unique barcode from the ONT PCR Barcoding Expansion 1-96 kit. Thermocycling conditions for this second reaction were 1 cycle of 95 °C for 3 min, 15 cycles of 95 °C for 15 s, 62 °C for 15 s and 65 °C for 1 min 35 s, with a final extension of 65 °C for 5 min.
- 5 Secondary PCRs were followed by another clean-up step using a 0.6× ratio of NucleoMag beads to exclude and remove low molecular weight (< 400 bp) PCR product, hence 30 µl of beads were used to clean 50 µl of PCR product with the same incubation and ethanol wash steps used as previously described and elution in 20 µl of water. Next correct amplification of the expected product was assessed using a subset of samples on a 4200 TapeStation System (Agilent Technologies, California, USA) and the final DNA concentrations of this subset analysed using a Qubit™ 4 Fluorometer.
- 6 Subsequently, 2 µl of each barcoded and cleaned PCR product were pooled together and concentrated down using a 2× ratio of NucleoMag beads, washed twice with 75% ethanol and eluted in 50 µl of water. This amplicon pool was then quantified on a Qubit™ 4 Fluorometer to ensure there was adequate DNA, i.e., a minimum requirement of 1,000 ng, to be taken forward.
- 7 Final library preparation steps included DNA repair and end-prep, adapter ligation, clean-up and MinION™ flow cell priming and loading was conducted exactly as described in the aforementioned ONT' protocol, making use of the NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (New England Biolabs) and the ONT' Ligation Sequencing Kit (SQK-LSK110). The final concentration of sequencing library was always between 20 - 50 fmol, i.e., within the recommended 5 – 50 fmol range recommended for loading onto R9.4.1 flow cells.
- 8 Batches were run with four no template PCR negative controls, i.e., water, and four positive controls that were comprised of a uniquely identifiable 1,486 bp gBlock synthetic DNA strand (Integrated DNA Technologies, Iowa, USA) of the 16S rRNA gene from *Aliivibrio fischeri*. This positive control gBlock consisted of the relevant 16S rRNA gene sequence flanked by the appropriate primer binding regions for the APICO_F_Mod and Api_Illum_New-Rev primers as an artificial construct, the design of which, can be seen at the end of this protocol. If flow cells were re-used this was always after a DNase clean-up using the EXP-WSH004 Flow Cell Wash Kit (Oxford Nanopore Technologies), to reduce the possibility of DNA contamination and carry-over from prior sequencing runs.

- 9 Nanopore sequencing was conducted on a MinION Mk1B device using a Legion 7i Gen 6 laptop (Lenovo, Quarry Bay, Hong Kong) that utilises a NVIDIA® GeForce RTX 3070 (8 GB) GPU and 11th Gen Intel® Core™ i7-11800H (8C) processor to permit field-based base-calling. Sequencing was initiated through MinKNOW version 22.12.7 with fast base-calling and a Q-score of ≥ 8 , for between 3.5 - 39 hrs depending on the amount of data required. Once sequencing was stopped, FAST5 reads were base-called using the super high accuracy base-calling model with barcode removal using Guppy version 6.4.6. Upon sequencing commencement, the success of the sequencing run was assessed using MinKNOW to ensure reads were of the expected size and pore activity was healthy. Sequencing that was conducted for methodological comparison was allowed to continue until a mean per sample raw read count of at least 62,500 reads was achieved.

Bioinformatics

- 10 For processing of nanopore sequencing data a bioinformatic pipeline had to be chosen that could correct for the error rate of the utilised ONT' R9.4.1 flow cells, via construction of accurate 18S rRNA gene consensus sequences. The bioinformatic pipeline NanoCLUST was chosen which conducts multiple quality control, read clustering, polishing and consensus forming steps followed by classification of consensus sequences using blastn against a database of the user's choice.
- 11 To create our own apicomplexan database from which NanoCLUST could correctly assign taxonomic classifications we downloaded all NCBI's GenBank 18S rRNA gene sequences greater than 200 bp long and smaller than 10,000 bp long for species within the phylum apicomplexa (txid5794). Our search terms were ((((((18S ribosomal RNA[Title]) OR 18S rRNA[Title]) OR ribosomal RNA[Title]) OR SSU rRNA[Title]) OR SSU ribosomal RNA[Title]) AND txid5794[Organism]) AND 200:10000[Sequence Length]. To this database we also included our positive control sequence for the *A. fischeri* 16S rRNA gene (NCBI accession NR_029255.1). The relevant GitHub page detailing how our database was constructed is available here https://github.com/vetscience/Huggins_NanoCLUST, our database is downloadable from https://melbourne.figshare.com/projects/Huggins_NanoCLUSTdb/160631.
- 12 Optimal NanoCLUST parameters were found to be to be a minimum read length of 1,000 bp, maximum read length of 1,900 bp, minimum cluster size of 30 and 100 reads for polishing, the curated apicomplexan 18S rRNA gene database used as the taxonomic classification database with all other parameters as the pipeline's defaults. Read counts as well as consensus sequence lengths and classifications generated by NanoCLUST were taken as the final dataset generated by our nanopore sequencing assay to which other methods were compared.

Positive Control Design

- 13 **DNA sequence of the unique positive control gBlock construct used for apicomplexan haemoparasite 18S ribosomal RNA gene metabarcoding.**

The gBlock positive control DNA sequence construct (1,486 bp) is comprised of APICO_F_Mod and Api_Illum_New-Rev primer binding sites (underlined) and a region of the 16S ribosomal RNA gene of *Aliivibrio fischeri*. This construct was synthesised by Integrated DNA Technologies (Iowa, USA).

5' –

GCCAGTAGTCATATGCTTGTCTATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAGCGGAA
ACGACTTAACTGAACCTTCGGGGAACGTTAAGGGCGTCGAGCGGCGGACGGGTGAGTAATGCCTG
GGAATATGCCTTAGTGTGGGGGATAACTATTGGAAACGATAGCTAATACCGCATAATGTCTTCGGA
CCAAAGAGGGGGACCTTCGGGCCTCTCGCGCTAAGATTAGCCCAGGTGAGATTAGCTAGTTGGTGA
GGTAAGAGCTACCAAGGCGACGATCTCTAGCTGGTCTGAGAGGATGATCAGCCACACTGGA
ACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGA
TGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTTCAGTAGGGAGGAAGGT
GTTGTAGTTAATAGCTGCAGCATTTGACGTTACCTACAGAAGAAGCACCGGCTAACTCCGTGCCAG
CAGCCGCGGTAATACGGAGGGTGCAGCGTTAATCGGAATTACTGGGCGTAAAGCGCATGCAGGT
GGTTCATTAAGTCAGATGTGAAAGCCCGGGGCTCAACCTCGGAACCGCATTGAACTGGTGA
ACTAGAGTGCTGTAGAGGGGGGTAGAATTTAGGTGTAGCGGTGAAATGCGTAGAGATCTGAAGGAATA
CCAGTGGCGAAGGCGGCCCCCTGGACAGACACTGACACTCAGATGCGAAAGCGTGGGGAGCAAAC
AGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCTACTTGGAGGTTGTTCCCTTGAGGAG
TGGCTTTCGGAGCTAACGCGTTAAGTAGACCGCCTGGGGAGTACGGTCGCAAGATTAA
AACTCAAA TGAATTGACGGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGA
ACCTTACCTACTCTTGACATCCAGAGAATTCGCTAGAGATAGCTTAGTGCCTTCGGGAACTCTGAGACAG
GTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACC
CTTATCCTTGTTTGCCAGCACGTAATGGTGGGAACTCCAGGGAGACTGCCGGTGATAA
ACCGGAGG AAGGTGGGGACGACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCG
CATACAGAGGGCTGCAAGCTAGCGATAGTGAGCGAATCCCAAAAAGTGCGTCGTAGTCCGGATTGG
AGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAAT
ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGCTGCAAAAGAAGTGGGTA
GTTTAACCTTCGGGACGAGGAAGTTTGAGGCAATAACAG– 3'