




RESEARCH ARTICLE

Annotated mitochondrial genome with Nanopore R9 signal for *Nippostrongylus brasiliensis* [version 1; referees: 1 approved, 2 approved with reservations]

Jodie Chandler¹, Mali Camberis¹, Tiffany Bouchery², Mark Blaxter³, Graham Le Gros¹, David A Eccles ¹

¹Malaghan Institute of Medical Research, Wellington, New Zealand

²École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

³Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

v1 First published: 19 Jan 2017, 6:56 (doi: [10.12688/f1000research.10545.1](https://doi.org/10.12688/f1000research.10545.1))
 Latest published: 19 Jan 2017, 6:56 (doi: [10.12688/f1000research.10545.1](https://doi.org/10.12688/f1000research.10545.1))

Abstract

Nippostrongylus brasiliensis, a nematode parasite of rodents, has a parasitic life cycle that is an extremely useful model for the study of human hookworm infection, particularly in regards to the induced immune response. The current reference genome for this parasite is highly fragmented with minimal annotation, but new advances in long-read sequencing suggest that a more complete and annotated assembly should be an achievable goal. We *de-novo* assembled a single contig mitochondrial genome from *N. brasiliensis* using MinION R9 nanopore data. The assembly was error-corrected using existing Illumina HiSeq reads, and annotated in full (i.e. gene boundary definitions without substantial gaps) by comparing with annotated genomes from similar parasite relatives. The mitochondrial genome has also been annotated with a preliminary electrical consensus sequence, using raw signal data generated from a Nanopore R9 flow cell.




This article is included in the **Nanopore analysis** channel.

Open Peer Review

Referee Status: ? ? ✓

	Invited Referees		
	1	2	3
version 1 published 19 Jan 2017	? report	? report	✓ report

- 1 **Matthias Bernt**, Leipzig University
Germany
- 2 **Christian Rödel** ,
Max-Planck-Institute for Developmental
Biology Germany
- 3 **Jianbin Wang**, University of Colorado
School of Medicine USA

Discuss this article

Comments (0)

Corresponding authors: Jodie Chandler (jchandler@malaghan.org.nz), David A Eccles (bioinformatics@gringene.org)

How to cite this article: Chandler J, Camberis M, Bouchery T *et al.* **Annotated mitochondrial genome with Nanopore R9 signal for *Nippostrongylus brasiliensis* [version 1; referees: 1 approved, 2 approved with reservations]** *F1000Research* 2017, **6**:56 (doi: [10.12688/f1000research.10545.1](https://doi.org/10.12688/f1000research.10545.1))

Copyright: © 2017 Chandler J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This study was supported by program grant funding from the Health Research Council of New Zealand, the Marjorie Barclay Trust, and the Glenpark Foundation.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The present project has been fully-funded, but we have in the past received complimentary deliveries of flow cells and sequencing reagents from Oxford Nanopore Technologies, as part of the MinION Access Program. The authors declare that there are no other competing interests.

First published: 19 Jan 2017, **6**:56 (doi: [10.12688/f1000research.10545.1](https://doi.org/10.12688/f1000research.10545.1))

Introduction

Nippostrongylus brasiliensis is a parasitic nematode that naturally infects rodents. Its life cycle and morphology is comparable to *Necator americanus* and *Ancylostoma duodenale*, and it is thus an excellent murine model of human hookworm infection, a disease that affects approximately 700 million people worldwide¹. Like its human counterparts, *N. brasiliensis* L3 larvae infect the host through the skin and migrate to the lungs where they feed on red blood cells (unpublished study; Haem metabolism is a check-point in blood-feeding nematode development and resulting host anaemia; Bouchery T, Filbey K, Shepherd A, Chandler J, Patel D, Schmidt A, Camberis A, Peignier A, Smith AAT, Johnston K, Painter G, Pearson M, Giacomini P, Loukas A, Bottazzi M-E, Hotez P, Le Gros G), causing extensive haemorrhage and anaemia – both hallmarks of hookworm infections. The larvae are coughed up and swallowed to enter the gastrointestinal tract. The nematode matures into a sexually active adult in the small intestine where it secretes eggs that enter the environment via the hosts' faeces. Larvae hatch, undergo two molts to become infective L3 larvae, which propagates the lifecycle². The immunology of *N. brasiliensis* infection has been studied extensively, and the parasite has been utilised as an inducer of potent Th2 responses in the lung and intestine, yielding important discoveries into cellular and molecular immune responses^{3–6}. The *N. brasiliensis* model allows delineation of hookworm-induced immune profiles that could be targeted in drug or vaccine design, and provides a simple and well-characterised murine model in which to test these interventions for efficacy. To underpin these studies, a high-quality reference genome is needed.

Current reference genome

The most recent NCBI reference genome sequence for *N. brasiliensis* is a draft generated from Illumina HiSeq reads as part of the Wellcome Trust Sanger Institute (WTSI) 50 Helminth Genomes initiative^{7–9}. It is 294.4 Mbp in total length, and highly fragmented (29,375 scaffolds with an N50 length of 33.5kb, and a longest scaffold of under 400kb). The *N. brasiliensis* reference genome would benefit from improvement, a goal that may be readily achieved with the advent of affordable long-read sequencing technologies.

MinION sequencing

The Oxford Nanopore Technologies' (ONT) **MinION platform** is improving at a rapid pace, with improvements in flow cell chemistry and base calling software announced frequently. In 2015, the median accuracy of double-stranded MinION reads, using R7.3 sequencing pores, was about 89% pores, sequenced at 60 bases per second with a yield of about 200 Mb¹⁰. The quality and length of sequences generated from R7.3 pores was sufficient to create a single-contig assembly of the *Escherichia coli* K-12 MG1655 chromosome using nanopore reads alone, with consensus accuracy of 99.5%¹¹. An equimolar sample of *Mus musculus*, *E. coli* and *Enterobacteriophage lambda* DNA was sequenced in September 2016 on the International Space Station using R7.3 flow cells, producing approximately equal read counts for the different samples with a median accuracy of 83–92% for 2D reads across four runs^{12,13}.

The recent introduction of R9 sequencing pores in June 2016, together with improved software for base-calling the generated

signal trace at 250 bases per second¹⁴, has improved the median accuracy of high-quality double-stranded reads to 95%, and yield to 800 Mb (personal communication, September 2016; MinION Analysis and Reference Consortium). Consensus accuracy for an *E. coli* K12 assembly consequently also increased to 99.96%¹⁵. A rapid single-stranded sequencing kit was introduced in August 2016, reducing post-extraction sample preparation time to less than 15 minutes (see 16).

The R9.4 flow cell was commercially released by ONT a few months later in October 2016. This release brought together software and chemistry improvements that increased run flow cell yield into the gigabase range, and increased sequencing speed to 450 bases per second. Additional use cases for the MinION are evident with this increased yield: the R9.4 flow cells have already been used for sequencing human genomes using multiple flow cells, with observed yields of about 1–4Gb for each individual sequencing run^{17,18}.

Scientific justification

The mitochondrial genome is useful for epidemiology and population genetic analysis in nematodes, as it is rapidly evolving^{19,20}. An average cell has 100–1000 mtDNA molecules, compared to two nuclear DNA molecules²¹, and this stoichiometric excess facilitates analyses, especially where starting materials are limited. The strict maternal inheritance of the mitochondrial chromosome, coupled with a general lack of recombination in this haploid replicon permits inference of maternal lineages^{21–23}. The ONT MinION can be deployed in infectious disease outbreak scenarios, and a “read until” methodology promises to make rapid, specific identification of known infectious agents possible. The technology has obvious utility in other areas of epidemiology and infection surveillance, and to enhance these applications it will be useful to develop the “read until” methodology to be able to detect a wider range of infectious agents from metagenomic sequencing. To do this, electronic signatures representing the MinION nanopore event signals could be used as a reference library to pre-screen raw signals from the pores before base calling. Here we present a complete mitochondrial genome for *N. brasiliensis*, assess its quality by gene prediction and phylogenetic analyses, and provide a validated electronic signal trace for the sequence.

This annotation represents the first hurdle in generating a complete genomic sequence for this model organism and provides crucial information for evolutionary and immunological studies. The rapid advancement of molecular technologies, such as qPCR, RNAseq, nanostring and high through-put sequencing, has given researchers the capacity to acquire an expansive array of new knowledge and insight into how genetic pathways function and interact at a molecular level. However, the lack of a complete annotated reference genome for *N. brasiliensis* thus far has restricted the full exploration into this important helminth.

Methods and results

Genomic DNA was extracted from adult *N. brasiliensis* and sequenced on a MinION R9 flow cell. Reads from this sequencing run were then assembled, and the highest-coverage contig (mitochondrial DNA) was error-corrected and circularised for further analysis.

DNA extraction and library preparation

N. brasiliensis was originally sourced from Lindsey Dent of the University of Adelaide, South Australia and has been maintained for 22 years by serial passage at the Malaghan Institute. Female Lewis rats were bred and used for the maintenance of the *N. brasiliensis* life cycle at 4 months of age (weight over 150g; housed in IVC caging and given *ad libitum* access to food and water). For the purposes of this study, one rat was infected with 4000 infective larvae. After 7 days, to allow the worms to mature to the adult stage in the small intestine, the rat was euthanized, and the small intestine dissected and flushed with PBS to harvest worms, as outlined in Camberis *et al.*². Ethics approval for the maintenance of the *N. brasiliensis* life cycle is overseen and approved by the Victoria University of Wellington Animal Ethics Committee.

The harvested *N. brasiliensis* were washed in PBS by centrifugation to remove cellular debris. The nematodes were frozen at -80°C bead-beaten, and DNA extracted using Qiagen DNeasy Blood and Tissue DNA extraction kit, yielding approximately 4μg of high molecular weight double-stranded DNA (determined by the Quantus QuantiFluor dsDNA System). This DNA was treated with RNase. Two sequencing libraries were made using the Oxford Nanopore 2D genomic DNA sequencing kit, yielding in total about 70ng of adapter-ligated sequencing library. No effort was made to specifically isolate mitochondrial DNA. The first preparation was loaded onto an R9 MinION flow cell and sequenced for 6 hours, and the second preparation was loaded onto the same flow cell and sequenced for an additional 36 hours. Pore occupancy at 30 minutes into the first run was about 25%, while pore occupancy at 30 minutes into the second run was about 80%.

Whole-genome assembly with Canu

All FASTQ sequences (i.e. both 1D and 2D reads) were extracted from the base-called FAST5 files. These sequences were fed into Canu v1.3²⁴ to generate assembled contigs. The contig with the highest coverage was a 19907 bp sequence with similarity to other nematode mitochondrial genomes (see Supplementary File 1). This sequence had 98% identity to an unannotated *N. brasiliensis* contig in the Wellcome Trust Sanger Institute (WTSI) *N. brasiliensis* assembly⁷.

Error correction and circularisation

Reads generated by WTSI (SRA ID: ERR063640) were mapped as pairs to the MinION mitochondrial contig using Bowtie2²⁵ in local mode. At each location, one read was randomly sampled from those that mapped to that location, representing a reference-based digital normalisation to approximately 100X coverage (see Supplementary File 2). The differences between these normalised reads and the MinION contig were evaluated using a custom script, producing a corrected sequence based on the consensus read alignments. The mapping and correction process was repeated with BWA-MEM²⁶ on the corrected sequence (see Supplementary File 3) to identify additional variants that were missed by Bowtie2, due to multiple matches to duplicated regions.

Repeated sections of the linear contig (representing duplicated regions of the circular sequence) were merged to generate a circular consensus sequence, and the resultant sequence adjusted

(by shifting sequence from the end to the start of the circular genome) so that the first base in the genome was set to the beginning of the COX1 gene (following the convention of OGRE²⁷, see <http://drake.physics.mcmaster.ca/ogre>). A final round of error correction was carried out on the circularised genome using Bowtie2-aligned reads from ERR063640 (see Supplementary File 4), producing a final mitochondrial genome length of 13,355 bp. The original 19 kbp contig thus contained about 6 kbp of duplicated sequence. MinION reads were mapped to the assembled genome to identify variants not present in the WTSI reads.

Comparison of WTSI and MIMR *N. brasiliensis* strains

After remapping the original R9 MinION reads back to the assembled and corrected genome with GraphMap²⁸, four locations were found with variant calls that contributed to more than 50% of the read coverage. Three of these variants involved transition mutations: $T \rightarrow C$ at 5742, $G \rightarrow A$ at 6102, and $T \rightarrow C$ at 11460. One additional complementary mutation was found: $T \rightarrow A$ at 2860 (see Figure 1).

Mitochondrial genome annotation

Approximate gene boundaries were determined by a local NCBI BLASTx search, mapping the contig to mitochondrial protein sequences from *Necator americanus* (see Table 1; Supplementary File 5 and Supplementary File 6). Regions between genes were

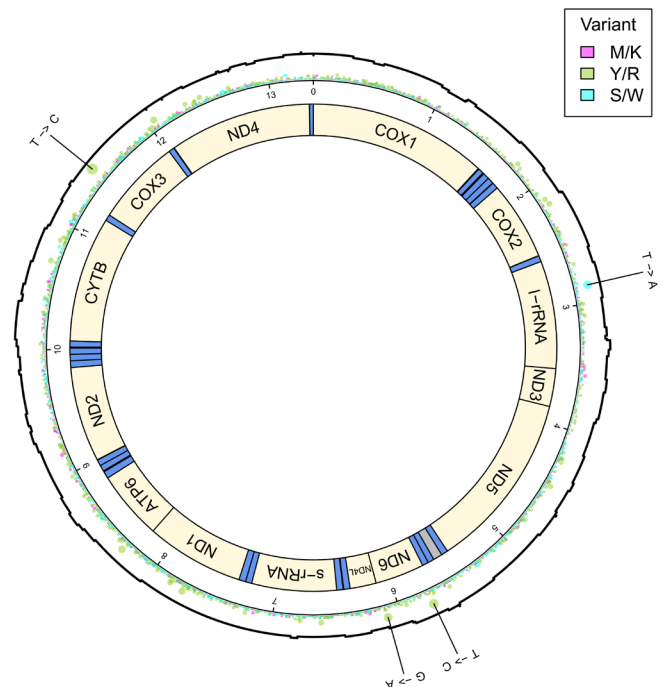


Figure 1. Diagram of mtDNA with mapped MinION read coverage. Gene regions are displayed in this circular mitochondrial DNA diagram in yellow, with tRNA regions in blue. The AT-rich region between the ND5 and ND6 genes is shaded grey. A combined coverage/variant plot is also displayed, showing MinION read coverage (in black), and base-called transition, transversion, and complementary variants (in chartreuse, magenta and cyan, respectively). Variant differences between Wellcome Trust Sanger Institute and Malaghan Institute of Medical Research strains of *Nippostrongylus brasiliensis* are indicated on the perimeter of the diagram.

Table 1. mtDNA gene regions. Predicted gene features from the *Nippostrongylus brasiliensis* mitochondrial genome. Stop codons that end in hyphens (-) are completed by the addition of polyA sequence.

Start	End	Name	Start Codon	Stop Codon
1	1575	COX1	ATT	TAG
1820	2514	COX2	TTG	TA-
2571	3522	l-rRNA	—	—
3523	3857	ND3	ATA	TAA
3858	5438	ND5	ATT	TTA
5498	5578	AT-rich	—	—
5689	6123	ND6	ATA	TAA
6124	6356	ND4L	ATT	TA-
6474	7223	s-rRNA	—	—
7339	8206	ND1	ATT	T--
8207	8806	ATP6	ATT	TAA
9003	9842	ND2	ATT	TAA
10076	11186	CYTb	ATA	T--
11242	12007	COX3	ATA	T--
12062	13291	ND4	GTT	TAA

then scanned using Infernal cmscan²⁹ to identify exact tRNA gene boundaries and codon sequences (see Table 2). The amino acid associated with each tRNA was identified using BWA-MEM to map annotated tRNA sequences from *Oesophagostomum columbianum*, *N. americanus*, *Strongylus vulgaris*, and *A. duodenale*. One tRNA region found by cmscan (between the ND4 and COX1 genes) could not be matched to any existing tRNA sequences. When this sequence was fed into *RNAstructure*³⁰, the predicted secondary structure had no T-loop or D-loop, and an anticodon loop of 8 bases (Figure 2). The anticodon for this structure pairs with one of the two most common gene start codons (i.e. ATT), and could potentially pair with the other most common start codon through a wobble A-A pairing on the third base (see 31).

Precise gene start boundaries were determined by mapping open reading frames (ORFs) between the tRNA genes (codon translation table 5: *Invertebrate Mitochondrial*) with NCBI SmartBLAST (<https://blast.ncbi.nlm.nih.gov/smartblast/smart-blast.cgi?CMD=Web>). Stop boundaries were determined by looking for plausible in-frame stop sequences surrounding the end region of matching SmartBLAST hits. The boundaries for the ribosomal RNA genes were determined by a BLAST search against the four previously compared parasite species. Finally, the AT-rich region was identified as the region between tRNA-Ala and tRNA-Pro.

Phylogenetic analyses

We identified orthologues of cytochrome oxidase 1 (COX1), cytochrome B (CytB), and the large ribosomal RNA subunit (l-rRNA) in other rhabditid nematodes using BLAST, and collated a dataset

Table 2. mtDNA tRNA sites. Predicted tRNA sites in the *Nippostrongylus brasiliensis* mitochondrial genome. One truncated tRNA site between the ND4 and COX1 genes (detected by cmscan) could not be fully annotated.

Start	End	Amino Acid	Codon
1589	1638	Cys	GCA
1649	1705	Met	CAU
1706	1760	Asp	GUC
1764	1819	Gly	UCC
2517	2570	His	GUG
5439	5497	Ala	UGC
5579	5633	Pro	UGG
5634	5688	Val	UAC
6356	6411	Trp	UCA
6417	6473	Glu	UUC
7224	7278	Asn	GUU
7279	7338	Tyr	GUA
8818	8880	Lys	UUU
8890	8944	Leu	UAA
8944	8997	Ser	UCU
9843	9901	Ile	GAU
9902	9959	Arg	ACG
9959	10013	Gln	UUG
10022	10075	Phe	GAA
11187	11241	Leu	UAG
12008	12057	Thr	UGU
13322	13355	—	AUU

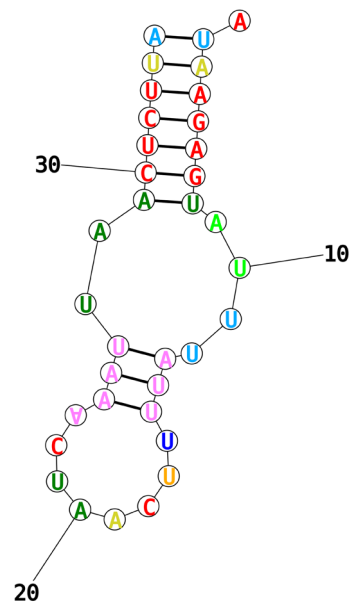


Figure 2. Predicted truncated tRNA structure. RNA structure for the truncated tRNA between ND4 and COX1, predicted by *RNAstructure*.

from 49 taxa. Nucleotide sequences were aligned using clustal³², trimmed with trimAL, and phylogenies estimated using RAXML using the GTRGAMMA model. Bootstrap values were calculated from 100 iterations. Figures were generated using FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). The *Nippostrongylus brasiliensis* sequences were placed within Strongyloomorpha, as expected, and *N. brasiliensis* was found to be sister to *Heligmosomoides polygyrus*, a finding in keeping with morphological systematics. Many internal nodes have very low bootstrap values, suggesting either low or conflicting signal in the data. Some groups were well supported, but these tend to be within rather

than between genera. Overall the tree conforms to the classical morphological and global molecular phylogenies of the suborder, but cannot stand as indicators of those relationships independently (Figure 3).

Park and colleagues³² used whole mitochondrial genomes (i.e. all 12 protein coding loci) to develop a phylogeny of Nematoda, with the goal of analysing the placement of some unusual mitochondria from Ascaridia species, but including many strongyles. Our analyses are largely congruent with theirs, albeit with lower support (as noted above).

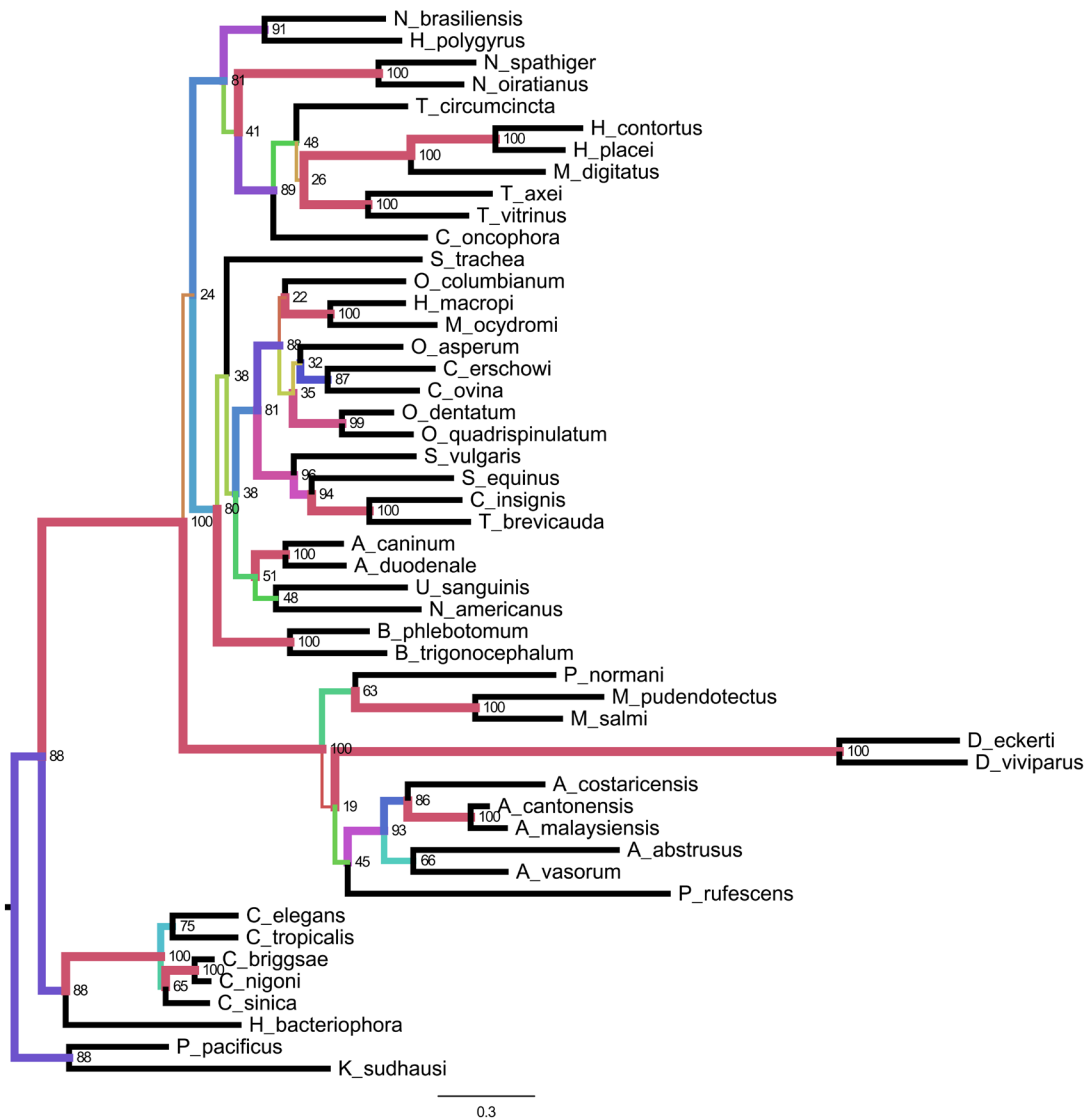


Figure 3. Phylogenetic tree for mtDNA. Phylogenetic tree based on evidence from three mitochondrial-encoded genes: *cytochrome oxidase 1*, *1-rRNA*, and *cytochrome B*. This tree demonstrates sequence similarities for 47 species from the *Rhabditida* together with two outgroups (*Pristionchus pacificus* and *Koerneria sudhausi*). Branch lengths are nucleotide substitutions per bp. Nodes are labelled with sub-sequence deletion bootstrap values. Branch colours and width are representative of bootstrap proportion.

Read mapping

The template and complement raw signal from the MinION reads mapped by GraphMap²⁸ were extracted from the FAST5 files, and sorted into four groups:

1. Template sequence, mapped to coding strand
2. Template sequence, mapped to non-coding strand
3. Complement sequence, mapped to coding
4. Complement sequence, mapped to non-coding

A summary of mapping counts can be found in Table 3. Reads where the template fragment mapped to the non-coding strand were

Table 3. mtDNA read groups. Statistics for the four different read mapping groups, showing reads that mapped to the *Nippostrongylus brasiliensis* mitochondrial genome with over 50% coverage.

Direction	Strand	Count	Mean Length
Template	Coding	26	5.0 kbp
Complement	Non-coding	25	4.8 kbp
Template	Non-coding	17	5.3 kbp
Complement	Coding	16	5.1 kbp

about two-thirds that of coding strand-mapped reads, with a similar proportion of reads distributed between the template and complement read fragments.

Event mapping

Event information (generated by the ONT cloud base caller *Metrichor* dragonet, version 1.22.4) was extracted for these sorted reads, and per-group median event currents were calculated for each pentamer found in the reference mitochondrial genome. An ideal signal trace of the mitochondrial genome was generated using these statistics for the four different signal groups (see Figure 4; Supplementary File 7).

Median complement events mapped to coding strand pentamers had a slightly higher event current when compared to template events (median difference = 3.94 pA, 90% range: 1.2 ~ 6.7, MAD = 1.53), and were lower in events mapped to non-coding pentamers (median difference = -2.08 pA, 90% range: -5.7 ~ 1.6, MAD = 2.93).

The median signal level for pentamers found in the *N. brasiliensis* mitochondrial genome has a very strong positive correlation between read direction for the coding strand ($r = 0.982$, 90% range: 0.980 ~ 0.984) and the non-coding strand ($r = 0.974$, 90% range: 0.972 ~ 0.978), whereas there is weaker negative correlation between strands for the template direction ($r = -0.67$, 90% range: -0.70 ~ -0.63) and the complement direction ($r = -0.66$, 90% range: -0.69 ~ -0.62).

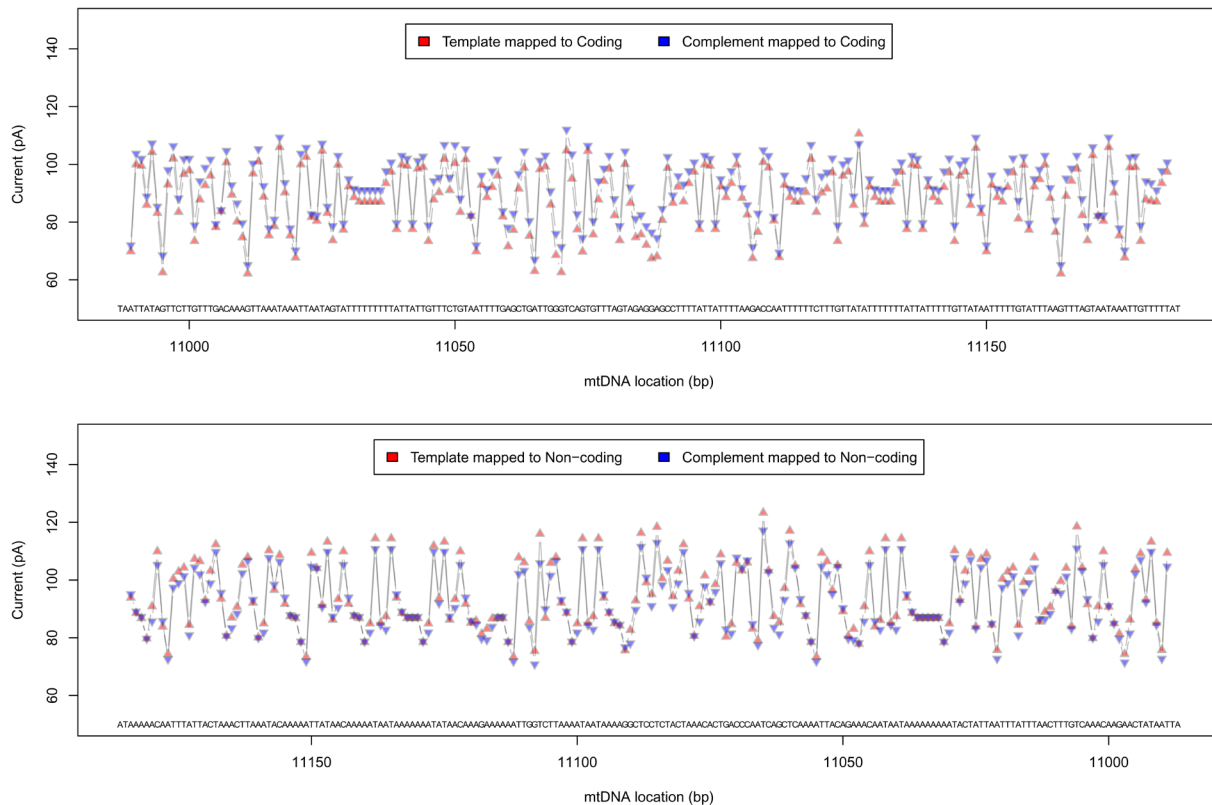


Figure 4. Ideal event plot, CytB gene tail. Ideal event trace for 200 pentamers at the tail end of the Cytochrome B gene. The complement sequence has a slightly lower current than the template sequence for reads mapped to the coding strand, and also a slightly lower current for reads mapped to the non-coding strand.

Raw signal mapping

Raw signal traces from both template and complement strands were converted to pA using scaling metadata in the FAST5 files, mapped to the GraphMap-aligned reference base positions using event metadata, and linearly interpolated to 11 samples per base using the R *approx* function (R version 3.3.1). Median signal traces (at a sub-base resolution) were generated by summarising the mapped signal at each interpolated location (Figure 5; Supplementary File 8).

The event data signal for template sequence mapped to the coding strand was loosely correlated with median raw signals in the middle of the interpolated region ($r = 0.52$, 90% range: $0.51 \sim 0.53$), with other read groups demonstrating lower correlations ($r = 0.29 \sim 0.44$). This correlation disappeared when shifting the compared signal by one base in either direction ($r = 0.03 \sim 0.09$).

Discussion

Using a long-read assembler, and three passes of error correction with publicly-available data, we have created a full-length, error-free, *de novo* assembly of the mitochondrial genome of *N. brasiliensis*. This genome has been annotated with gene and tRNA boundaries, and compared with other related parasite species. An additional preliminary “electrical” annotation was generated from mapped nanopore read sequences.

Mitochondrial genome assembly

Low-cost long-read sequencing has made possible full-length assemblies of a number of different megabase-length genomes from nanopore data alone (e.g 11,33–35), so it is not surprising that a full-length mitochondrial assembly was also possible using

nanopore reads. The vast wealth of publicly-available data allows fast and low cost assembly, correction, and annotation of genomes, producing high-quality reference sequences that are of great benefit to medical research.

We were able to assemble the *N. brasiliensis* mitochondrial genome from a whole-genome sequencing nanopore dataset, by identifying assembly contigs with high relative coverage. The assembly is of high quality, based on read coverage, mapping of Illumina short reads, and annotation. The gene order is identical to that of *Caenorhabditis elegans* and other strongylomorph nematodes (see 36). Despite this shared structure, there is sufficient variation in sequences between species to generate resolved phylogenies³².

WTSI assembly of mtDNA

During the final preparation of this paper for publication, the WTSI deposited an annotated mitochondrial genome for *N. brasiliensis* (accession id: AP017690.1). This complements the introduction of the WormBase ParaSite resource for helminth genomics⁹. While the associated reference for the WTSI *N. brasiliensis* mitochondrial genome is not yet published, it is expected that this mitochondrial genome was assembled using a similar method to the WTSI’s previous work³⁷ (i.e. a reference-based iterative mapping procedure using MITObim³⁸).

The sequence of this assembly differs only in an additional T insertion into a 10 base poly-T tract in the 1-rRNA gene. While such polynucleotide tracts are problematic for MinION, the polyT region appears to be polymorphic, with some support for both variants in the WTSI reads (ERR063640). In addition the WTSI annotation excludes the AT-rich region.

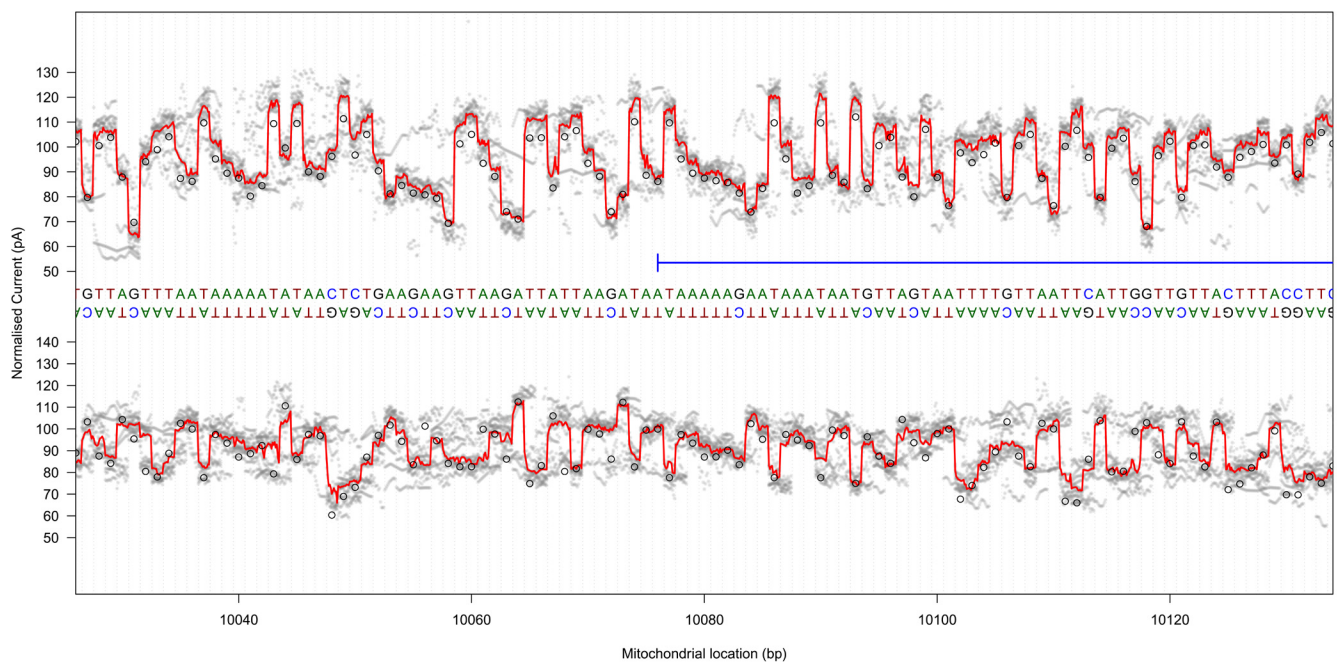


Figure 5. Raw signal plot, CytB gene head. Raw signal plot for 100 bases at the start of the Cytochrome B gene for template read directions (top) and complement read directions (bottom). Median raw signal current is shown as a thick red line, with individual raw signal observations shown in grey. Ideal event current for the observed pentamers is shown as black circles.

MinION whole genome sequencing data from a metazoan can be used for taxon identification

At the time of sequencing, no mitochondrial genome for *N. brasiliensis* was available. We thus explored the utility of the MinION data in species identification. As the mitochondrial genome is at a higher molarity than the nuclear genome, low-coverage sequencing of a target genome can yield deep coverage of the mitochondrion. Assembly of this replicon, and then analysis in a phylogenetic context was successful in placing *N. brasiliensis* in the Strongylomorpha. We suggest that this approach would be a useful technology for identification of unknown specimens in clinical practice, biosurveillance or biodiversity research programmes. In addition the nanopore electronic signal of the mitochondrial sequence could be used in a “read until” approach³⁹ to diagnosis, using live monitoring to identify reads that likely derive from this, or a very similar genome. Usually, identification through sequencing is applied to amplification of specific target loci in a specimen or sample, an approach known as DNA barcoding. Direct sequencing of the whole genome of a specimen on MinION would allow both barcoding and produce additional sequences that could be used for, for example, population genetic diversity analysis.

Nanopore read analyses

Nanopore reads were separated into four different read groups to provide information that could be used to establish whether or not there are different sequencing features associated with template and complement strands. In general, the coding and non-coding strands had similar electrical profiles, as demonstrated by the event data (e.g. see [Figure 4](#)).

As this investigation is the first attempt to categorise the electrical properties of a complete mitochondrial genome, errors in the data analysis (e.g. due to incorrect mapping, low read coverage, and incorrect scaling parameters) cannot be excluded as an explanation for the difference in current that were observed between event data and raw signal. A comparison of raw signal current to the ideal current suggests that the pentamer model is probably sufficient to fully describe variation in signal in the mitochondrial genome. Although correlation between the signal and the ideal pentamer model is low for all four sequencing groups (template coding, template non-coding, complement coding, complement non-coding), this variation could be explained by errors in the raw signal mapping process, and other alternative mapping techniques (e.g. nanoraw⁴⁰) may give better performance for linking raw signal to sequenced bases.

It is possible that the observed difference between the raw and ideal event signal may be due to methylation and other epigenetic modification of the mitochondrial genome. Methylation is a known feature of mitochondrial DNA (see [41](#)), and methylation patterns can be observed as changes in the nanopore electrical signal⁴². Due to the lack of information about epigenetic patterns from *de novo* nanopore sequencing, this dataset is provided without additional epigenetic analysis as a source of discovery for other researchers.

Conclusions

The data presented here have been created from a minimally-prepared whole-genome DNA from *N. brasiliensis*, combining

nanopore reads with publicly-available datasets. Using non-targeted sequencing, we have been able to generate a fully-annotated (gap-free) mitochondrial genome, with an initial electrical signal annotation having a resolution that is finer than a single base. The analysis proves that the efficiently MinION-generated mitochondrial genome of *N. brasiliensis* is of high enough quality for phylogenetic use.

We hope that the procedures discussed here will be sufficient to guide other researchers in annotating mitochondrial genomes and generating consensus signal traces, and that these data will contribute more generally towards improving the sequence base calling algorithms in the future for devices that implement sequencing by observation.

Data availability

Sequences have been deposited into [NCBI Genbank](#), with accession number KY347017. Reads used to produce this assembly are associated with [BioProject PRJNA328296](#). The assembly was error corrected using Illumina reads from a Wellcome Trust Sanger Institute sequencing run ([ERR063640](#)).

The *mpileup2proportion.pl* custom script that was used for error-correcting nanopore reads using Bowtie2-mapped short reads, as well as for generating count data for the read coverage plot, is available from David Eccles' github repository (DOI, [10.5281/zenodo.164193](#))⁴³. Read mapping group statistics were generated using the *fastx-grep.pl* and *fastx-length.pl* scripts also from this repository. These scripts have also been included here as a supplementary file ([Supplementary File 9](#)).

Author contributions

- JC** Preparation & extraction of DNA, phylogenetic analyses, manuscript preparation
- MC** Maintenance & propagation of *N. brasiliensis* larvae
- TB** *N. brasiliensis* sequencing project conception
- MB** Interpretation of phylogenetic results
- GLG** Project design & oversight
- DAE** Project design, MinION sequencing, assembly, data analysis, manuscript preparation

All authors have read the paper and provided edit suggestions where appropriate.

Competing interests

The present project has been fully-funded, but we have in the past received complimentary deliveries of flow cells and sequencing reagents from Oxford Nanopore Technologies, as part of the MinION Access Program.

The authors declare that there are no other competing interests.

Grant information

This study was supported by program grant funding from the Health Research Council of New Zealand (14/003), the Marjorie Barclay Trust, and the Glenpark Foundation.

Acknowledgements

We would like to thank Dr. Matt Berriman and the Wellcome Trust Sanger Institute for the unpublished draft genome data of

N. brasiliensis, Kara Filbey for providing editorial suggestions for this manuscript and to G. Koutsovoulos and A. Buck for prepublication access to *H. polygyrus* genome data.

Supplementary files

Supplementary File 1 Original single-contig assembly generated by Canu v1.3.

[Click here to access the data.](#)

Supplementary File 2 ERR063640 reads mapped to Canu-assembled genome, digitally normalised to one read per base.

[Click here to access the data.](#)

Supplementary File 3 ERR063640 reads mapped to the first error corrected-genome, digitally normalised to one read per base.

[Click here to access the data.](#)

Supplementary File 4 ERR063640 reads mapped to the second error corrected-genome, digitally normalised to one read per base.

[Click here to access the data.](#)

Supplementary File 5 BED-format file of discovered mtDNA features (prior to correction of boundaries following protein translation).

[Click here to access the data.](#)

Supplementary File 6 FASTA file containing subsequences of the mitochondrial genome representing discovered features.

[Click here to access the data.](#)

Supplementary File 7 Event-level data aggregated for each base in the mitochondrial genome, including ideal current derived from pentamer signals.

[Click here to access the data.](#)

Supplementary File 8 Data file containing interpolated raw signal-level data.

[Click here to access the data.](#)

Supplementary File 9 Compressed file containing all Perl and R scripts used for data processing and analysis.

[Click here to access the data.](#)

References

- Hotez PJ, Bethony JM, Diemert DJ, *et al.*: **Developing vaccines to combat hookworm infection and intestinal schistosomiasis.** *Nat Rev Microbiol.* 2010; **8**(11): 814–826.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Camberis M, Le Gros G, Urban J Jr: **Animal model of *Nippostrongylus brasiliensis* and *Heligmosomoides polygyrus*.** *Curr Protoc Immunol.* 2003; Chapter 19: Unit 19.12.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bouchery T, Kyle R, Camberis M, *et al.*: **ILC2s and T cells cooperate to ensure maintenance of M2 macrophages for lung immunity against hookworms.** *Nat Commun.* 2015; **6**: 6970.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ohnmacht C, Schwartz C, Panzer M, *et al.*: **Basophils orchestrate chronic allergic dermatitis and protective immunity against helminths.** *Immunity.* 2010; **33**(3): 364–374.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chen F, Wu W, Millman A, *et al.*: **Neutrophils prime a long-lived effector macrophage phenotype that mediates accelerated helminth expulsion.** *Nat Immunol.* 2014; **15**(10): 938–946.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Neill DR, Wong SH, Bellosi A, *et al.*: **Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity.** *Nature.* 2010; **464**(7293): 1367–1370.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Holroyd N, Sanchez-Flores A: **Producing parasitic helminth reference and draft genomes at the wellcome trust sanger institute.** *Parasite Immunol.* 2012; **34**(2–3): 100–107.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wellcome Trust Sanger Institute: ***Nippostrongylus brasiliensis* genome sequencing.** NCBI BioProject. 2014.
[Reference Source](#)
- Howe KL, Bolt BJ, Shafie M, *et al.*: **WormBase ParaSite - a comprehensive resource for helminth genomics.** *Mol Biochem Parasitol.* 2016; pii: S0166-6851(16)30160-8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ip CL, Loose M, Tyson JR, *et al.*: **MinION Analysis and Reference Consortium: Phase 1 data release and analysis [version 1; referees: 2 approved].** *F1000Res.*

- 2015; 4: 1075.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Loman NJ, Quick J, Simpson JT: **A complete bacterial genome assembled de novo using only nanopore sequencing data.** *Nat Methods*. 2015; 12(8): 733–5.
[PubMed Abstract](#) | [Publisher Full Text](#)
 12. Castro-Wallace SL, Chiu CY, John KK, *et al.*: **Nanopore dna sequencing and genome assembly on the international space station.** *bioRxiv*. 2016.
[Publisher Full Text](#)
 13. JSCNASA and r/Science: **NASA AMA: We just sequenced DNA in space for the first time. ask us anything!** *The Winnower*. 2016.
[Publisher Full Text](#)
 14. Brown C: **Inside the skunkworx.** In: *London Calling*. Oxford Nanopore Technologies. 2016.
[Reference Source](#)
 15. Simpson J: **Supporting R9 data in nanopolish.** Simpson Lab Blog. 2016.
[Reference Source](#)
 16. Edwards A, Debonnaire AR, Sattler B, *et al.*: **Extreme metagenomics using nanopore DNA sequencing: a field report from svalbard, 78 n.** *bioRxiv*. 2016.
[Publisher Full Text](#)
 17. Brown C: **Cliveome onth1 data release.** Github repository. 2016.
[Reference Source](#)
 18. Akesson M, Beggs AD, Nieto T, *et al.*: **Na12878: Data and analysis for na12878 genome on nanopore.** Github repository. 2016.
[Reference Source](#)
 19. Brown WM, George M Jr, Wilson AC: **Rapid evolution of animal mitochondrial DNA.** *Proc Natl Acad Sci U S A*. 1979; 76(4): 1967–1971.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. Martin SA: **Mitochondrial DNA repair.** In: *DNA Repair – On the Pathways to Fixing DNA Damage and Error*. InTech, 2011.
[Publisher Full Text](#)
 21. Pakendorf B, Stoneking M: **Mitochondrial DNA and human evolution.** *Annu Rev Genomics Hum Genet*. 2005; 6: 165–183.
[PubMed Abstract](#) | [Publisher Full Text](#)
 22. Cann RL, Stoneking M, Wilson AC: **Mitochondrial DNA and human evolution.** *Nature*. 1987; 325(6099): 31–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
 23. Harrison RG: **Animal mitochondrial DNA as a genetic marker in population and evolutionary biology.** *Trends Ecol Evol*. 1989; 4(1): 6–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
 24. Berlin K, Koren S, Chin CS, *et al.*: **Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.** *Nat Biotechnol*. 2015; 33(6): 623–630.
[PubMed Abstract](#) | [Publisher Full Text](#)
 25. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods*. 2012; 9(4): 357–359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv*, 1303.3997v2. 2013.
[Reference Source](#)
 27. Jameson D, Gibson AP, Hudelot C, *et al.*: **Ogre: a relational database for comparative analysis of mitochondrial genomes.** *Nucleic Acids Res*. 2003; 31(1): 202–206.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 28. Sović I, Šikić M, Wilm A, *et al.*: **Fast and sensitive mapping of nanopore sequencing reads with GraphMap.** *Nat Commun*. 2016; 7: 11307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 29. Nawrocki EP, Kolbe DL, Eddy SR: **Infernal 1.0: inference of RNA alignments.** *Bioinformatics*. 2009; 25(10): 1335–1337.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 30. Reuter JS, Mathews DH: **RNAstructure: software for RNA secondary structure prediction and analysis.** *BMC Bioinformatics*. 2010; 11(1): 129.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. Murphy FV 4th, Ramakrishnan V: **Structure of a purine-purine wobble base pair in the decoding center of the ribosome.** *Nat Struct Mol Biol*. 2004; 11(12): 1251–1252.
[PubMed Abstract](#) | [Publisher Full Text](#)
 32. Park JK, Sultana T, Lee SH, *et al.*: **Monophyly of clade III nematodes is not supported by phylogenetic analysis of complete mitochondrial genome sequences.** *BMC Genomics*. 2011; 12(1): 392.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. Risse J, Thomson M, Patrick S, *et al.*: **A single chromosome assembly of *Bacteroides fragilis* strain BE1 from illumina and MinION nanopore sequencing data.** *Gigascience*. 2015; 4(1): 60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 34. Istace B, Friedrich A, d'Agata L, *et al.*: **de novo assembly and population genomic survey of natural yeast isolates with the oxford nanopore minion sequencer.** *bioRxiv*. 2016; 066613.
[Publisher Full Text](#)
 35. Davis AM, Iovinella M, James S, *et al.*: **Using minion nanopore sequencing to generate a de novo eukaryotic draft genome: preliminary physiological and genomic description of the extremophilic red alga galdieria sulphuraria strain sag 107.79.** *bioRxiv*. 2016.
[Publisher Full Text](#)
 36. Hu M, Gasser RB: **Mitochondrial genomes of parasitic nematodes--progress and perspectives.** *Trends Parasitol*. 2006; 22(2): 78–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
 37. Hunt VL, Tsai IJ, Coghlan A, *et al.*: **The genomic basis of parasitism in the *Strongyloides* clade of nematodes.** *Nat Genet*. 2016; 48(3): 299–307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 38. Hahn C, Bachmann L, Chevreux B: **Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach.** *Nucleic Acids Res*. 2013; 41(13): e129.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Loose M, Malla S, Stout M: **Real-time selective sequencing using nanopore technology.** *Nat Methods*. 2016; 13(9): 751–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. Stoiber MH, Quick J, Egan R, *et al.*: **De novo identification of dna modifications enabled by genome-guided nanopore signal processing.** *bioRxiv*. 2016.
[Publisher Full Text](#)
 41. Ghosh S, Singh KK, Sengupta S, *et al.*: **Mitoepigenetics: the different shades of grey.** *Mitochondrion*. 2015; 25: 60–66.
[PubMed Abstract](#) | [Publisher Full Text](#)
 42. Simpson JT, Workman R, Zuzarte PC, *et al.*: **Detecting DNA methylation using the oxford nanopore technologies MinION sequencer.** *bioRxiv*. 2016; 047142.
[Publisher Full Text](#)
 43. Eccles D: **Bioinformatics scripts: Initial citable release.** *Zenodo*. 2016.
[Data Source](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 12 April 2017

doi:10.5256/f1000research.11363.r21379



Jianbin Wang

Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO, USA

In this manuscript, Chandler *et al* described in detail how they used the Nanopore sequencing technology to assemble the mitochondrial genome of *Nippostrongylus brasiliensis*. They also annotated the mitochondrial genome and did phylogenetic analysis among a selected group of nematodes. In addition, they characterized the Nanopore sequencing features for this genome. Overall, the authors have demonstrated that they can produce the complete mitochondrial genome from their Nanopore sequencing dataset.

The authors were able to recover the mitochondrial genome from a genomic DNA library due to the much higher copy number (often hundreds or thousands of times) of the mitochondrial DNA when compared to the nuclear genome. This approach has also been extensively used to recover mitochondrial and chloroplast genomes in whole genome shotgun libraries. In principle, it should work for any type of sequencing technology. The Nanopore sequencing technology is relatively new and is still fast evolving. In this case the technology does not seem to me to have a clear advantage over the Illumina or other sequencing approaches on mitochondrial genome assembly. In additional, the authors eventually used the Illumina data to do the error correction to make the final assembly. Nevertheless, the authors have presented a complete genome assembled from a combination of Nanopore and Illumina data with a full description of how they did this.

Not considering the novelty or significance of the work, I think the mitochondrial genome is properly assembled and annotated. The results are clear and the manuscript is well written.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 12 Apr 2017

David Eccles, Malaghan Institute of Medical Research, New Zealand

Thank you very much for your review of our paper on mitochondrial genome sequencing with the MinION sequencer. We are currently working on updating the paper as per the reports of Christian Rödelsperger and Matthias Bernt, and intend to deliver a full response to them at that time.

This paper was intended as a stepping stone for investigating techniques that could be used to assemble a parasite genome from *unamplified* genomic DNA using the MinION. We discovered that the run yield in this case was not sufficient for assembling the entire *N. brasiliensis* genome, but being able to assemble a *mitochondrial* genome as a single contig has given us confidence that the technology is capable of improving on the existing Illumina-derived whole genome assembly. We did not intend to wow the world with this paper, rather it was an attempt to demonstrate methods and show how easy and quick it can now be to assemble a genome. Thank you for understanding this aspect of our paper.

At the time of sequencing, the base-calling software was not sufficiently accurate to generate a reliable sequence at a single base level. Understanding this, we used MinION reads for scaffolding, and Illumina reads (from a different strain) to correct the abundant base call errors. This approach has allowed a relatively cheap and fast assembly of the mitochondrial genome, such that comprehensive phylogenetic analyses can be carried out on the mitochondrial genes.

As you have mentioned, the nanopore sequencing technology is evolving fast. It is likely the case that updated base-calling software has improved base calling accuracy sufficiently that this approach can be carried out using MinION reads alone. I would like to carry out additional investigations on these data to discover if that is indeed the case, but would rather hold off on that until after we have published our attempts at whole genome assembly. Regardless, the mitochondrial sequences (including raw signal) are available for anyone else to determine themselves whether or not a high-quality MinION-only assembly is possible using re-called (but otherwise identical) nanopore sequence data.

Competing Interests: No competing interests were disclosed.

Referee Report 11 April 2017

doi:10.5256/f1000research.11363.r21560



Christian Rödelberger 

Department for Evolutionary Biology, Max-Planck-Institute for Developmental Biology, Tübingen, Germany

The manuscript by Chandler *et al* describes the sequencing, assembly and annotation of the *Nippostrongylus brasiliensis* mitochondrial genome using Nanopore sequencing technology. The comparison of the resulting assembly with other *N. brasiliensis* data from the parasite sequencing initiative of the Sanger Institute and also with mitochondrial genomes from other nematodes support that the produced assembly is of high quality.

In general, the structure of the article is a bit unusual. Methods and Results sections are combined, each part has multiple subsections that are not really connected. Some parts of the paper deal with the mitochondrial genome of *N. brasiliensis* other parts focus on very specific aspects of Nanopore sequencing. I would recommend to concentrate on the mitochondrial genome of *N. brasiliensis* and keep the nanopore-specific questions for a separate methodological paper.

Section: Introduction

The Introduction basically describes the lifecycle of *N. brasiliensis* and the mode of infection. The authors might consider writing a more general introduction about nematodes, parasites, .. that it is important to study these parasites to develop treatments. In addition, there are multiple related parasites that are later part of the phylogenetic analysis. It would be good to give some information about those ones as well. e.g. what are their hosts?

Section: Current reference genome

Please provide the Genbank entry for the NCBI reference genome or provide the assembly that has been used for this study as supplemental data. Otherwise, it will be hard to reproduce the results.

Section: Scientific justification

please explain what a "read until" methodology is and provide some reference for the use of ONT MinION in studies of infectious disease outbreak.

Is the *N. brasiliensis* isolate that was used for sequencing have a strain ID? If yes, please specify and at least register a biosample for it and give the accession number. Was it the same isolate that was used for the NCBI reference genome.

Section: Whole-genome assembly with Canu

How much sequencing data was obtained? Please provide some more details about the assembly results. How many Contigs, total size.

For readers, that would like to use Nanopore technology to sequence their genomes it would be interesting to compare the quality of of the mitochondrial genome with nuclear contigs. I guess, that the lower coverage of nuclear contigs should also result in higher number mismatches with regard to the reference genome. A major finding of the paper could be that based on current nanopore technology, it only makes sense to do the multicopy mitochondrial genome. Such a statement could help people to plan their projects.

Section: Error correction and circularisation

How many sites had to be corrected. Error correction only makes sense, if the WTSI data is from the same isolate. Please clarify if this is the case. If it is the same isolate, there does the 2% mismatches come from in the "Whole-genome assembly with Canu" section.

Section: Mitochondrial genome annotation

"The amino acid associated with each tRNA was identified using BWA-MEM to map annotated tRNA sequences from *Oesophagostomum columbianum*, *N. americanus*, *Strongylus vulgaris*, and *A. duodenale*." Using BWA-mem to annotate tRNAs from other species sounds unusual. Do you have a reference where the performance of this methodology has ever been evaluated?

Section: Phylogenetic analyses

Please provide more information about the alignment, how many sites? amount of missing data? Please provide references for what is called "the classical morphological and global molecular phylogenies"

Section: Read mapping/ Event mapping /Raw signal mapping (Table 3, Fig 4 and 5)

These sections seem to examine very specific aspects of the Nanopore sequencing technology and do not add any additional insights for the presented mitochondrial genome. I also have problems in understanding what kind of questions are asked. It seems to me, as if the authors try to examine whether Nanopore data has a preference for the template or complement strand or whether there is a bias for

coding or non-coding sequences. How well the sequencing signal corresponds to the basecalls in the final assembly and what features correlate with variation in sequencing signals. The presented results are not conclusive (no statistical tests have been done to assess the significance of the results) and are not really related to the rest of the manuscript. I would recommend to use this and other comparable data for a separate more methodological paper. One additional feature that could be tested would be how differences raw and ideal event current, sequencing coverage depend on GC content.

Minor comments

Section: DNA extraction and library preparation

The first paragraph should probably be labeled as "Worm culturing" or something else. It has nothing to do with DNA extraction or library preparation

Section: MinION sequencing

"sequenced at 60 bases per second with a yield of about 200 Mb" does that mean per sequencing run?

This section sounds a bit like a promotion of MinION sequencing. I would recommend to reduce it only to the parts that are relevant for the current paper.

high through-put sequencing -> high-throughput

I wonder why the title has to have the information that R9 signal has been used. Probably most readers have heard about Nanopore sequencing but do not have a clue what R9 signal is. I would recommend to put this detailed information into the methods section but remove it from the title.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 Apr 2017

David Eccles, Malaghan Institute of Medical Research, New Zealand

Thank you very much for your review of our paper on mitochondrial genome sequencing with the MinION sequencer. We are currently working on updating the paper as per your report (and the report of Matthias Bernt), and will deliver a full response once the next revision of the paper is ready.

Competing Interests: No competing interests were disclosed.

Referee Report 20 February 2017

doi:10.5256/f1000research.11363.r19519



Matthias Bernt

Department of Computer Science, Leipzig University, Leipzig, Germany

The paper describes the sequencing of the mitochondrial genome of the *Nippostrangylus brasiliensis* with the novel Nanopore sequencing technique. To the best of my knowledge this seems to be one of the first mitogenomes that have been sequenced with this technology. The annotation of the genome and its use for phylogeny and taxonomic identification have been discussed.

Another group has sequenced the genome (including the mitogenome) has been sequenced using another NGS strategy. While this seem unfortunate its actually good for this study otherwise no reference data would have been available for comparison and error correction. I'm missing an analysis of the error rates of the sequencing without the correction that has used the read data from the other study. I'm wondering if the combination of data from MiniON sequencing and short read sequencing strategies might be a good general strategy?

The paper is well written and needs only a few corrections and additions. Details are given below.

Abstract:

=====

The term "electrical consensus sequence" might be puzzling for uninformed readers.

Introduction:

=====

"L3" is also difficult to understand for non experts. Maybe add 'stage'?

"highquality" missing space

MiniON sequencing

=====

"R7.3" Can you explain what this means?

"89% pores" is unclear to me.

What are "2D reads"?

Scientific justification

=====

"strict maternal inheritance": nothing in biology is strict. Check for paternal leakage or doubly-uniparental inheritance.

The term "read until" methodology is unclear.

DNA extraction and library preparation

=====

Explain the abbreviation PBS

Error correction and circularisation

=====

It needs to be explained what the custom script is doing.

"Repeated sections of the linear contig were merged..." What happens with true repeats?

Since not all readers might know the color chartreuse I would suggest to order the colors as in the legend.

Mitochondrial genome annotation

=====

I'm wondering why automatic methods for genome annotation have been ignored. Not saying that the applied approach is wrong.

When you use cmscan you need to state the used model as well.

"tRNA... codon sequences" Do you mean anticodon?

How about non-canonical start codons? How do you define "plausible" in frame stop?

For the truncated tRNAs there are examples known for Enoplea: see <http://dx.doi.org/10.4161/rna.21630> and <http://dx.doi.org/10.1016/j.biochi.2013.07.034>

Phylogenetic Analyses

=====

References for RAxML and trimAL are missing.

Event Mapping

=====

"Event information" Specify what an event is.

"per-group" and later on "signal groups" You should reformulate this. Currently its a bit confusing.

Why pentamer?

What is an ideal signal trace?

Raw Signal Mapping

=====

Has Graph Map been referenced?

Discussion

=====

Are you really sure that the sequence is "error free"? In the end of the paper you write that its of "high enough quality...".

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 Apr 2017

David Eccles, Malaghan Institute of Medical Research, New Zealand

Thank you very much for your review of our paper on mitochondrial genome sequencing with the MinION sequencer. We are currently working on updating the paper as per your report (and the report of Christian Rödelsperger), and will deliver a full response once the next revision of the paper is ready.

Competing Interests: No competing interests were disclosed.