DNA Barcoding of Nematodes Using the MinION

Knot, I.E.; Zouganelis, G.D.; Weedall, G.D.; Wich, S.A.; Rae, R.

DOI
10.3389/fevo.2020.00100

Publication date
2020

Document Version
Final published version

Published in
Frontiers in Ecology and Evolution

License
CC BY

Citation for published version (APA):
DNA Barcoding of Nematodes Using the MinION

Ineke E. Knot1*, George D. Zouganelis2, Gareth D. Weedall2, Serge A. Wich1,2 and Robbie Rae2

1 Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands, 2 School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, United Kingdom

Many nematode species are parasitic and threaten the health of plants and animals, including humans, on a global scale. Advances in DNA sequencing techniques have allowed for the rapid and accurate identification of many organisms including nematodes. However, the steps taken from sample collection in the field to molecular analysis and identification can take many days and depend on access to both immovable equipment and a specialized laboratory. Here, we present a protocol to genetically identify nematodes using 18S SSU rRNA sequencing using the MinION, a portable third generation sequencer, and proof that it is possible to perform all the molecular preparations on a fully portable molecular biology lab – the Bentolab. We show that both parasitic and free-living nematode species (Anisakis simplex, Panagrellus redivivus, Turbatrix aceti, and Caenorhabditis elegans) can be identified with a 96–100% accuracy compared to Sanger sequencing, requiring only 10–15 min of sequencing. This protocol is an essential first step toward genetically identifying nematodes in the field from complex natural environments (such as feces, soil, or marine sediments). This increased accessibility could in turn improve global information of nematode presence and distribution, aiding near-real-time global biomonitoring.

Keywords: MinION, DNA barcoding, biomonitoring, 18S (SSU) rRNA gene, Anisakis simplex, Panagrellus redivivus, Turbatrix aceti, Caenorhabditis elegans

INTRODUCTION

Nematodes are one of the most abundant groups of metazoan organisms (Seesao et al., 2017). It is estimated that less than 4% of nematode species are currently known to science, with global species richness estimated between $10^6$ and $10^8$ (Lambshead, 2004). Many of these species are parasites that threaten the health of plants and animals, including humans. For example, the World Health Organization estimates that worldwide infections with soil-transmitted nematodes cause a human annual disease burden of 3.8 million years lost to disabilities (YLD), a disease burden in the same range as HIV/AIDS (4 million YLD) and twice as high as malaria (1.7 million YLD).1

Morphological identification is commonly used to identify nematode species, but also has significant drawbacks. For example, easily distinguishable morphological characters are scarce in nematodes, making identification difficult, time-consuming and often unsuccessful to genus or species level (Decraemer and Baujard, 1998; Lawton et al., 1998; Karanastasi et al., 2001; Lambshead, 2004; Hope and Aryuthaka, 2009). As a result, genetic identification is becoming

---

increasingly important in nematology. There have been increased efforts in recent years to resolve the genetic taxonomy of nematodes and barcode nematode species using markers including the 18S small subunit ribosomal RNA gene (18S SSU rRNA), the 28S large subunit ribosomal RNA gene (28S LSU rRNA), the cytochrome oxidase I gene (COI) and the internal transcribed spacer (ITS) regions of the ribosomal RNA locus (Blaxter et al., 1998; Bhadury et al., 2006a,b; Hunt et al., 2016; O’Neil et al., 2017; Seesao et al., 2017; Pařík et al., 2018).

The majority of these studies used Sanger sequencing but currently there are many sequencing technologies that have become more accessible and affordable for a wide array of applications (Kircher and Kelso, 2010; Van Dijk et al., 2014; Goodwin et al., 2016) such as high-throughput sequencing (HTS) and third generation sequencing (TGS). The latter is defined as single-molecule real-time sequencing (Van Dijk et al., 2014). Massive multiplexing of DNA barcode markers generates a great reduction of per sample sequencing costs and labor time compared to Sanger sequencing (Schuster, 2008; Shokralla et al., 2015). In this paper we explore TGS as an exciting opportunity for novel applications, such as near real-time biomonitoring of parasites, particularly nematodes.

A promising TGS platform is the MinION, introduced in 2014 by Oxford Nanopore Technologies (ONT). The MinION is a portable and compact USB-powered sequencer, generating long reads which can be base called in real-time (Jain et al., 2016). It utilizes a nanopore placed in a biological membrane through which DNA fragments are driven (Deamer et al., 2016), generating a difference in electrical current which can be measured and translated to different DNA bases. More in depth explanation of how the MinION works can be found in reviews by Pleshivkova et al. (2019) and Krehenwinkel et al. (2019b). The MinION’s portable nature makes it ideal for field research, proven by sequencing efforts in extreme conditions like the Arctic (Edwards et al., 2016; Goordial et al., 2017), Antarctic (Johnson et al., 2017) and the International Space Station (McIntyre et al., 2016). Shotgun genomic sequencing in a national park in Wales identified closely related plant species (Parker et al., 2017) and DNA barcoding (reviewed in Krehenwinkel et al., 2019b) has allowed for the identification of a variety of vertebrates in a rainforest in Ecuador (Pomerantz et al., 2018) and a rainforest in Tanzania (Menegon et al., 2017), all within hours of collection. Furthermore, the sequencer has successfully been used for real-time detection of Ebola virus during the 2014–2015 Ebola outbreak in West-Africa (Quick et al., 2016), Zika virus in Brazil (Faria et al., 2017; Quick et al., 2017), and the current outbreak of nCoV-2019\(^2\), which is an important step toward actionable clinical diagnostics. The most popular application of the MinION sequencer so far is the identification of viral or bacterial populations through metagenomics of the 16S rRNA gene (e.g., Greninger et al., 2015; Quick et al., 2015; Benítez-Páez et al., 2016; Schmidt et al., 2017), but metazoan parasites such as nematodes have not yet been examined.

In this paper we highlight the first step toward sequencing nematodes in situ, by genetically identifying parasitic and free-living nematode species with the MinION and testing a portable molecular lab. Specifically, we had four objectives and we sought to: (1) optimize existing 18S SSU rRNA primer sets for MinION sequencing of nematodes; (2) genetically identify nematode species with the MinION; (3) compare the MinION sequencing data to Sanger sequencing data, to assess the quality of MinION data and; (4) test whether we could achieve these results using a portable molecular laboratory, the Bentolab.

MATERIALS AND METHODS

Barcode Testing With Known Species

We tested DNA barcoding on four different nematode species, *Anisakis simplex*, *Panagrellus redivivus*, *Turbatrix aceti*, and *Caenorhabditis elegans*. These species represent a subset of parasitic and free-living species with diverse lifestyles.

*Anisakis simplex* was dissected from fresh mackerel and stored in 70% ethanol, and one individual nematode was selected for DNA extraction. *A. simplex* is a marine parasite that uses crustaceans as intermediate hosts to infect teleosts and squids (Anderson, 2000). Although humans are accidental hosts for *Anisakis* spp., there has been a dramatic increase over the last decades in the reported prevalence of anisakiasis, a serious zoonotic disease (Chai et al., 2005).

*Panagrellus redivivus* was harvested from a fresh culture growing on oatmeal medium and used for DNA extraction. *P. redivivus* is a free-living nematode that has been used as a model system to study organ development, signal transduction, and toxicology and recently had its full genome and transcriptome sequenced (Srinivasan et al., 2013). The species is amongst others suggested as a comparative model for Strongyloides, as parasitic taxa are typically difficult to culture and analyze independently of their hosts (Blaxter et al., 1998).

*Turbatrix aceti* was harvested from a fresh culture in an apple cider vinegar medium and used for DNA extraction. The nematodes were washed in distilled water three times before DNA extraction, to mitigate an inhibiting effect of the vinegar medium on the subsequent Polymerase Chain Reaction (PCR). *T. aceti* is a free-living nematode that is mostly researched in relation to aging phenotypes, that are shared with other free-living nematodes such as *Caenorhabditis elegans* (Reiss and Rothstein, 1975). It is also used as live food in the larval stages of many fish species (Brüggemann, 2012). It lacks proper genetic studies, making it an interesting representative for the majority of nematode species that are mostly studied morphologically.

*Caenorhabditis elegans* strain N2 was grown on nematode growth medium (NGM) plates with *E. coli* OP50 for several days using standard procedures (Brenner, 1974) and subsequently harvested for DNA extraction.

DNA Extraction, PCR, and Sequencing

We extracted the DNA using the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific Ltd., Paisley, UK) by first washing the nematodes in distilled water three times before DNA extraction, and using the following protocol:...
United Kingdom) according to manufacturer's instructions for mammalian tissue and rodent tail genomic DNA purification (protocol A), except that samples were lysed overnight (step 3) to ensure complete cuticle break down. DNA purity was measured on a NanoDrop 2000 spectrophotometer (software: NanoDrop2000, version 1.4.2; ThermoFisher Scientific Ltd., Paisley, United Kingdom).

We amplified an internal fragment of the 18S SSU rRNA gene from our DNA samples, using the primers and thermocycler protocol optimized by Floyd et al. (2005). This fragment is ∼900 bp in length and widely used for nematode species identification. According to ONT's instruction we adapted the primers from Floyd et al. (2005) to include an adapter tail at the 5’ end (“MiniION tail,” in lowercase), which is compatible with the MiniION workflows. This resulted in the following forward primer: Nem_18S_F_MiniION: 5’ tttcgttggtctgatatccCGGCAATRGCTCATACAAAGC 3’ and reverse primer: Nem_18S_R_MiniION: 5’ acttgctgtcctcttcGGGCGGTATCTGACGC 3’. A different primer pair, SSU18A and SSU26R (Floyd et al., 2002), was initially tested with the MiniION tails, but resulted in no PCR amplification for these samples. Each 25-µl PCR mix contained 2 µl purified DNA extract, 0.5 µl each forward and reverse primers (10 µM; Sigma-Aldrich/Merck Ltd., Poole, United Kingdom), 9.5 µl nuclease free water (NFW; ThermoFisher Scientific Ltd., Paisley, United Kingdom), and 2X GoTaq Hot Start Colorless Master Mix (Promega, Southampton, United Kingdom). PCR was performed on a Bio-Rad T100 Thermal Cycler (Bio-Rad Laboratories Ltd., Watford, United Kingdom). The PCR protocol for an amplicon length of ∼1,000 bp (including primers) was as follows: initial denaturation 3 min at 95°C, denaturation 15 s at 95°C, annealing 15 s at 62°C, extension 50 s at 65°C; hold at 4°C. The PCR products were cleaned up with 1X Agencourt AMPure XP beads (Beckman Coulter Inc., Indianapolis, IN, United States). Finally, 1 µl per purified second-round PCR product was quantified on the Qubit 1.0 (Thermo Fisher Scientific Ltd., Paisley, United Kingdom) using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific Ltd., Paisley, United Kingdom).

The concentration of A. simplex and P. redivivus DNA was too high for Qubit quantification, so we prepared and quantified a 1/5 dilution in NFW (Thermo Fisher Scientific Ltd., Paisley, United Kingdom) that was taken forward. The second-round PCR products were pooled in roughly equimolar concentrations in 47 µl NFW (Thermo Fisher Scientific Ltd., Paisley, United Kingdom).

Library preparation continued using the reagents from the Ligation Sequencing Kit (SQK-LSK109; ONT Ltd., Oxford, United Kingdom), according to manufacturer's instructions. Briefly, we prepared 325 ng pooled barcoded library in 47 µl NFW (ThermoFisher Scientific Ltd., Paisley, United Kingdom). Amplified product was end-repaired using NEBNext Ultra II End-Repair/da-tailing Module (NEB Inc., Hitchin, United Kingdom) for 5 min at 20°C and 5 min at 65°C, after which it was cleaned up with 1X Agencourt AMPure XP beads (Beckman Coulter Inc., Indianapolis, IN, United States). Adapter ligation was performed using NEB Blunt/TA Ligation Master Mix (NEB Inc., Hitchin, United Kingdom) and reagents provided in the SQK-LSK109 kit. Ligation took place for 10 min at room temperature. DNA was eluted in 15 µl Elution Buffer after being purified with 0.4X AMPure XP beads and washed with the Short Fragment Buffer provided in the SQK-LSK109 kit. 1 µl of prepared library was quantified on the Qubit 1.0 (Thermo Fisher Scientific Ltd., Paisley, United Kingdom) using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific Ltd., Paisley, United Kingdom) and gave a measure of 6.36 ng/µl, which equates to a molarity of 102.9 fmol.

The protocol from ONT recommends loading 5-50 fmol of amplicon product onto the flow cell, so we diluted 5.44 µl of prepared library in 6.56 µl Elution Buffer to load 40
fmol of library onto the flow cell. The flow cell was primed for loading by flushing the flow cell with 1 ml priming mix (30 µl of Flush Tether in one tube of Flush Buffer), taking care to avoid the introduction of air bubbles. The library was prepared for loading by mixing 37.5 µl Sequencing Buffer, 25.5 µl Loading Beads and 12 µl diluted DNA library, after which the sample was added to a flow cell, type R9.5.1, through the SpotON sample port. Total library preparation time was estimated to be ~3 h.

We performed the sequencing run using MinKNOW (version 3.4.5; ONT Ltd., Oxford, United Kingdom) on the MinIT (a small powerful computing unit that eliminates the need for a dedicated laptop; ONT Ltd., Oxford, United Kingdom), indicating the flow cell type and experimental kit used. As a metric of flow cell quality the MinKNOW software assesses flow cell active pore count, in the multiplexer (MUX) scan before each run. Higher active pore counts represent a high flow cell quality, with a maximum of 2,048 and a guaranteed level of 800. Our flow cell had 1,097 pores available for sequencing. The flow cell generated 116,620 reads in 10 min of sequencing, after which the run was stopped. The flow cell was subsequently washed using the Wash Kit (EXP-WSH002; ONT Ltd., Oxford, United Kingdom) with 150 µl Solution A, followed by 500 µl of Storage Solution, and stored in the fridge for re-use.

### Portable DNA Extraction, PCR and Sequencing

In preparation for field work we tested whether the developed MinION procedure could also be performed on a fully portable system. We prepared the model organism *C. elegans* for MinION sequencing using a portable molecular lab, the Bentolab Pro¹ (Nature Biotechnology, 2016; Bento Bioworks Ltd., London, United Kingdom) and a multi tool (CMFTLi 10.8V Li-Ion Cordless Multifunction Tool, Clarke International Ltd., Epping, United Kingdom) as a low-cost handheld vortex. Most of the procedures are similar to above, but working on the Bentolab required some essential adaptations to protocols.

We extracted the DNA using the GeneJet Genomic DNA Purification Kit (ThermoFisher Scientific Ltd., Paisley, United Kingdom) according to manufacturer’s instructions for mammalian tissue and rodent tail genomic DNA purification (protocol A). Adaptations to the procedure to make this protocol work on the Bentolab were as follows: In step 3, the sample was divided over two 0.2 ml PCR tubes and briefly spun down using the Bentolab’s centrifuge (Bento Bioworks Ltd., London, United Kingdom). Subsequently, the two PCR tubes were incubated for 18 h at 56°C, using the Bentolab’s thermocycler (Bento Bioworks Ltd., London, United Kingdom) as a heating block. The thermocycler protocol performed 18 cycles of 1 h at 56°C. In step 4, the lysate was then transferred to a 1.5 ml centrifuge, 20 µl RNase A was added and vortexed on the multi tool². Vortexing on a multi tool can be achieved by attaching the “straight saw blade” to the multi tool. This blade provides enough space for up to four centrifuge tubes at the same time. As a safety measure the sharp end of the saw blade was covered with duck tape. Then the centrifuge tube was added to the blade with duck tape, ensuring a tight fit. The multi tool was turned on at the highest speed (21,000 strokes/minute), creating a similar effect as a lab vortex. The vortexing in step 5 and 6 was also performed using the multi tool. In step 7, the 2 ml collection tube of the GeneJet purification column was replaced by a 1.5 ml centrifuge tube with the cap cut off. The Bentolab’s centrifuge can only handle 1.5 ml tubes; use of 2 ml collection/centrifuge tubes will lead to small plastic particles that can lead to reduced efficiency of the centrifuge lock system. Because of the reduced volume of the collection tube, the lysate was added to the prepared column at a maximum of 350 µl at a time, after which the sample was centrifuged for 1 min at 6,000 × g and the flowthrough discarded (with a total of three repeats necessary to complete step 7). In step 8, 250 µl Wash Buffer I was added at a time and centrifuged for 1 min at 8,000 × g and the flowthrough discarded (with a total of two repeats necessary to complete step 8). In step 9, 250 µl Wash Buffer II was added at a time and centrifuged for 4 min at 8,000 × g and the flowthrough discarded (with a total of two repeats necessary to complete step 9), and increased centrifuge time to compensate for the max 8,000 × g force of the Bentolab’s centrifuge). An additional dry spin of 1 min at 8,000 × g was performed, after which the collection tube was discarded and replaced by a sterile 1.5 ml centrifuge tube. In step 10, 50 µl of Elution Buffer was added to the purification column.

PCR was prepared as described above, but this time performed using the Bentolab’s thermocycler (Bento BioWorks Ltd., London, United Kingdom). Also, aluminum foil was used as a sterile work environment as an alternative for a PCR hood. Aluminum foil was taped to the bench space using masking tape. Bleach (1:10 ratio dilution in water) was sprayed on the surface, letting it sit for 3 min, and wiping the surface with clean paper tissue. This process was repeated twice to decontaminate, after which 70% ethanol was used to remove any residual bleach. The Nem_18S_F/R_MinION primers did not work for *C. elegans*, so the primers from Floyd et al. (2002) were used with MinION tails (in lowercase). The forward primer: SSU18A_MinION: 5’ ttttcgttggctgtatatgcAAAGATTAAGCATGAGT 3’ and reverse primer: SSU26R_MinION: 5’ acctgcctgtcctacttcCATTCTGGCAAATGCTTTCG 3’.

---

1. https://www.bento.bio
2. Inspired by Holly Ganz’ use of a multi tool for bead beating: https://youtu.be/Q7PM1x0MjiU.
(Thermo Fisher Scientific Ltd., Paisley, United Kingdom). For the Bentolab's small gel chamber (Bento Bioworks Ltd., London, United Kingdom) we used 27.5 ml of 1X TBE buffer with 0.5 g agarose. The need for a scale was eliminated by using an Eppendorf tube marked with the needed volume corresponding to 0.5 g agarose. Agarose was melted into the TBE buffer using a traditional coffee pot, which has a typical conical shape, on the hob. We have also found this method to work on a camping stove.

Since we wouldn’t have an accurate way of quantifying DNA in the field we based the used volume of DNA on a previous MiNiON run (Knot, unpublished data), to prepare 33 µl barcoded library in 47 µl NFW (Thermo Fisher Scientific Ltd., Paisley, United Kingdom). Amplified product was end-removed using NEBNext Ultra II End-Repair/dA-tailing Module (NEB Inc., Hitchin, United Kingdom) for 5 min at 20°C and 5 min at 65°C on the Bentolab thermocycler (Bento Bioworks Ltd., London, United Kingdom). The end-removed library was cleaned up with 1X Agencourt AMPure XP beads (Beckman Coulter Inc., Indianapolis, IN, United States) on a 3D-printed magnetic BOMB microtube rack3. Amplified barcode library was used in a 24 h run and once in a 2.5 h run. When reusing a flow cell the library was written onto a different flow cell type and experimental kit used. To test whether old flow cells can still be useful for sequencing small barcoded amplicon libraries, we used a flow cell that was used twice before, once in a 24 h run and once in a 2.5 h run. When reusing a flow cell the starting voltage has to be adjusted and we adjusted this to -225 V, equivalent to ONT’s recommendation after ~26 h previous run time. As mentioned above, higher active pore counts represent a high flow cell quality, with a maximum of 2,048 and a guaranteed level of 800 for new flow cells. The MUX scan indicated our flow cell had 43 pores available for sequencing. The flow cell generated 2,632 reads in 14 min of sequencing, after which the run was stopped.

**Sanger Sequencing**

Each of the samples used for the MinION sequencing was also sent for Sanger sequencing (GATC/Eurofins Genomics).

3https://bomb.bio/protocols/
Both forward and reverse strands were sequenced using the amplification primers as sequencing primers. Sanger sequence electropherograms were visually inspected and edited using 4Peaks version 1.8 (NucleoBytes B.V., Aalsmeer, the Netherlands). Edited forward strand and reverse complemented reverse strand sequences were aligned using Seaview version 4.7 (Gouy et al., 2010). Nucleotide mismatches were checked in the original electropherogram and resolved. A consensus sequence was derived for each sample and primer sequences trimmed from each end of it. The resulting sequences were 885 bp (A. simplex), 887 bp (P. redivivus), 832 bp (T. aceti), and 844 bp (C. elegans) long.

Bioinformatic Analyses

The raw fast5 MinION reads were basecalled and demultiplexed using Guppy version 3.2.4 + d9ed22f (ONT Ltd., Oxford, United Kingdom) to produce fastq files for each sample. Reads were classified as pass/fail based on a minimum quality score of 7. The fastq files were merged into one per sample and explored using Nanoplot (version 1.28.0), creating plots displaying log transformed read length (“--loglength”). Barcode and primer trimming was performed using Porechop (version 0.2.4). A second round of demultiplexing requiring barcodes at both ends of the reads (“--require_two_barcodes”) was performed using Porechop. Subsequently, the MinION reads were processed using the default settings of the ONTrack pipeline (version 1.4.2; Maestri et al., 2019). Briefly, Seqtk seq was used to create fasta files complementary to the fastq files. Reads were clustered using VSEARCH (Rognes et al., 2016), after which the reads in the most abundant cluster were retained. Then 200 randomly sampled reads were used to produce a draft consensus sequence using Seqtk sample and aligned using MAFFT (Katoh et al., 2002). EMBOSS cons was then used to retrieve a draft consensus sequence starting from the MAFFT alignment. Another 200 randomly sampled reads using Seqtk sample, different from the first iteration, were mapped to the draft consensus sequence using Minimap2 (Li, 2018) to polish the obtained consensus sequence. Samtools was used to filter and sort the alignment file and compress it to the bam format (Li et al., 2009). Nanopolish index and nanopolish variants – consensus modules from Nanopolish were used to obtain a polished consensus sequence. The ONTrack pipeline was run with three iterations, the standard value of the pipeline. This resulted in three polished consensus sequences which were aligned with MAFFT to select the consensus sequence that was produced in the majority of times. All scripts of the pipeline were run within a virtual machine (as part of the ONTrack pipeline), emulating an Ubuntu v18.04.2 LTS operating system, on a Mac laptop without using any internet connection. All the code used for the bioinformatics analyses and additional files necessary to replicate the analyses can be found on https://github.com/ieknot/MinION-DNA-barcoding-of-nematodes. MinION fastq and Sanger fasta accession numbers are reported in the results.

To assess sequence accuracy, MinION raw reads and consensus reads were aligned to the corresponding Sanger-derived reference sequence using BLASTn (Altschul et al., 1990), with no sequence complexity masking (“--dust no-soft_masking false”). The consensus sequences were aligned to the corresponding Sanger sequence using the MUSCLE algorithm (Edgar, 2004) in Seaview version 4.7 (Gouy et al., 2010).

RESULTS

Sequencing Run Quality and Yield

The first and multiplexed flow cell had 1,097 pores available for sequencing. The flow cell generated 116,620 reads containing 6,033 Mb in 10 min of sequencing, after which the run was stopped. During basecalling 71.9% of these reads passed the minimum quality threshold. The basecalled reads were demultiplexed, producing 42,304 reads for analysis (Table 1). The mean read length was 1,015 bp for A. simplex, 1,011 bp for P. redivivus and 504 bp for T. aceti.

The second flow cell, used for the library prepared with a fully portable setup, had 43 pores available for sequencing. The flow cell generated 2,632 reads containing 1.94 Mb in 14 min of sequencing, after which the run was stopped. During basecalling 48.9% of these reads passed the minimum quality threshold. The basecalled reads were demultiplexed, producing 205 reads for analysis (Table 1). The mean read length was 833 bp for C. elegans.

To assess the usefulness of the sequence data for taxonomic identification of the samples, the raw reads for each sample were compared to the (Sanger) reference sequences using BLASTn. The distributions of percentage sequence identities for all pairwise read-reference comparisons are shown in Figure 1. The median percent identity was 88.5% for A. simplex, 87.7% for P. redivivus, 89.5% for T. aceti and 82.3% for C. elegans, indicating

<p>| TABLE 1 | Summary of number of reads per sample in subsequent bioinformatics steps. |
|----------------------|----------------------|----------------------|----------------------|----------------------|</p>
<table>
<thead>
<tr>
<th><strong>Species</strong></th>
<th><strong>Sample</strong></th>
<th><strong>Demultiplexed reads</strong></th>
<th><strong>Trimmed reads (%)</strong></th>
<th><strong>Reads used for consensus</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. simplex</td>
<td>BC05</td>
<td>8,059</td>
<td>3,294 (40.9%)</td>
<td>200</td>
</tr>
<tr>
<td>P. redivivus</td>
<td>BC06</td>
<td>13,802</td>
<td>5,515 (40.0%)</td>
<td>200</td>
</tr>
<tr>
<td>T. aceti</td>
<td>BC07</td>
<td>20,443</td>
<td>2,955 (14.5%)</td>
<td>200</td>
</tr>
<tr>
<td>C. elegans</td>
<td>BC10</td>
<td>483</td>
<td>205 (42.4%)</td>
<td>65</td>
</tr>
</tbody>
</table>

Trimmed reads are also represented as percentage of demultiplexed reads. For C. elegans the largest VSEARCH cluster contained only 65 reads, so the consensus sequence was generated with this amount of reads (see main text for explanation of the ONTrack bioinformatics pipeline).
Knot et al. MinION DNA Barcoding of Nematodes

FIGURE 1 | Distribution of percentage sequence identity between raw reads and the (Sanger) reference sequence (A) A. simplex, (B) P. redivivus, (C) T. aceti, and (D) C. elegans. (A–C) were run on a new MinION flow cell, whereas (D) was run on a flow cell that had been used twice before (for a total of 26.5 h, see main text for more details). The median percent identity (indicated by a vertical dotted line) is 88.5% for A. simplex, 87.7% for P. redivivus, 89.5% for T. aceti and 82.3% for C. elegans. Bioinformatic analyses using the ONTrack pipeline generated a consensus sequence for every sample that had a 99.9% (A,B), 100% (C), and 96.6% (D) accuracy compared to their Sanger reads (indicated by a vertical red line).

Bioinformatics Analyses
The second demultiplexing round in Porechop, which included the trimming step to remove primers and ONT barcodes, produced the following number of reads per sample to be taken forward in the ONTrack bioinformatics pipeline (Maestri et al., 2019): 3,294 reads for A. simplex, 5,515 reads for P. redivivus, 2,955 reads for T. aceti and 205 for C. elegans (Table 1). These reads were used in the VSEARCH clustering step of the ONTrack pipeline, where contaminating sequences were removed by only taking the largest cluster of reads forward. T. aceti raw data showed an unexpected short mean read length of 504 bp. The clustering step in the ONTrack pipeline removed contaminating sequences, after which the mean read length of T. aceti improved to 773 bp. From these clusters 200 reads were subsampled per sample for consensus sequence generation. However, for C. elegans the largest VSEARCH cluster contained only 65 reads, so the consensus sequence was generated with this amount of reads (Table 1).

The default setting of the ONTrack pipeline is to run three iterations of the pipeline, generating three consensus sequences per sample. Subsequently, it aligns the consensus sequences generated during each round and selects the final consensus sequence based on the majority rule. Two species, A. simplex and P. redivivus, generated three consensus sequences which were all different. Since they all had the same statistical probability of being correct, the first consensus sequence was randomly selected. T. aceti had a consensus sequence supported by two iterations, and the C. elegans consensus sequence was supported by all three iterations of the pipeline.

Median percent identity between consensus sequences and the reference sequence was significantly improved in all cases (Figure 1). For A. simplex and P. redivivus the accuracy improved to 99.9%, for T. aceti to 100% and for C. elegans to 95.6% (Figure 2). Compared to the raw MinION sequences this is an improvement of 11.4, 12.2, 10.5, and 13.3%, respectively. The MinION datasets generated for this study can be found in the European Nucleotide Archive (ENA) under the project ID PRJEB37489 (samples ERS4397495, ERS4397496, ERS4397497, and ERS4397498). MinION consensus sequences are available in the Supplementary Material.

Sanger sequence read lengths were 885 bp for A. simplex, 887 bp for P. redivivus, 832 bp for T. aceti and 844 bp for C. elegans. The Sanger reads of all samples matched 100% to a sequence of the correct species on NCBI (Figure 2). Sanger consensus sequences are available at GenBank under the accession numbers MT246663, MT246664, MT246665, and MT246666.

DISCUSSION
We successfully genetically identified four nematode species using 18S SSU rRNA barcoding on the MinION. We also proved that this can be accomplished using a fully portable molecular lab. This was possible by successfully adapting 18S SSU rRNA primers (Floyd et al., 2002, Floyd et al., 2005) with MinION tails. The read lengths of both samples fall within the expected range. Our first run yielded three successful species
FIGURE 2 | Species investigated and nucleotide alignments of MinION and Sanger sequences comparing consensus accuracy for (A) A. simplex, (B) P. redivivus, (C) T. aceti, and (D) C. elegans. Sanger sequences have a 100% accuracy. Accuracy shown is the accuracy of the MinION consensus reads. Comparison against accession numbers MF072711.1 (A. simplex), AF083007.1 (P. redivivus), AF202165.2 (T. aceti), and MN519140.1 (C. elegans). The scale bar in the photographs of (A–C) represents 1 mm and in (D) represents 0.5 mm.
identifications from MinION reads that have an accuracy of 99.9–100%, when compared to their respective Sanger reads. Our second run yielded a successful species identification from the MinION consensus read that has an accuracy of 95.6%, when compared to Sanger sequencing. The Sanger reads of all samples matched 100% to a sequence of the correct species on NCBI. The C. elegans data gives a considerably lower accuracy of its MinION consensus sequence than the three other species. We are confident that this is unrelated to the preparation of the sample on a portable setup, as we have applied this setup in a field situation and got MinION consensus sequences that matched closer to the correct species on NCBI (Knot et al., unpublished data). The C. elegans run was set up to test the portable setup before bringing all the equipment out to the field. As such, the primary objective was to prove that everything worked, with data quality being a lower priority. Hence, when the library was loaded and the MinION flow cell MUX scan indicated only 43 working pores, we continued the experiment nonetheless. The MinION generated data for 15 min, after which no active pores remained and the run was stopped. The flow cell that was used twice before, once in a 24 h run and once in a 2.5 h run, multiplexing nine samples in two runs, generating 1,950,657 reads containing 2.11 Gb in total. We therefore feel that the limited accuracy in the C. elegans run has more to do with the limited lifespan of a flow cell, than with the sample preparation or the portable sample preparation. We used the EXP-WSH002 wash kit from ONT, which has now been succeeded by the EXP-WSH003 kit. This latest kit incorporates a nuclease to digest and remove nucleic acid that has been loaded onto a flow cell previously and has proven to be much more efficient in maintaining flow cell quality after a wash than the previous wash kit\(^1\). We therefore do not expect future MinION flow cells to deteriorate as much after using the wash kit as was the case for our C. elegans run.

Tackling a large phylum like Nematoda presents challenges that other phyla might be less affected by Kumar et al. (2012). For example, we started exploring the primers developed by Floyd et al. (2005), because these primers are optimized for a wide phylogenetic range of nematodes. However, addition of the MinION tails seems to alter the efficiency of these nematode primers. The tailed primers amplified A. simplex, P. redivivus and T. aceti successfully, but the addition of MinION tails prohibited the primers to amplify C. elegans successfully. We then switched to a primer optimized specifically for soil nematodes (Floyd et al., 2002), and found that this primer with MinION tails amplified C. elegans without problems. Future work will benefit from testing a wider array of nematode primers.

The potential throughput of a MinION R9 chemistry flow cell is ~20 Gb (Krehenwinkel et al., 2019b), and has been shown to be sufficient to generate new draft genomes of nematodes through shotgun sequencing (Eccles et al., 2018; Fauver et al., 2019). Future barcoding work could focus on maximizing the utility of each flow cell by multiplexing samples in one run or harvesting the long-read potential unique to TGS platforms like PacBio and ONT. For example, Srivathsan et al. (2019) have developed an improved low-cost MinION pipeline where they multiplexed 3,500 samples per flow cell. However, preparing so many samples for sequencing requires significant labor time (Krehenwinkel et al., 2019b; Piper et al., 2019; Srivathsan et al., 2019). Heeger et al. (2018) used PacBio circular consensus sequencing to show the feasibility of long-read metabarcoding of environmental samples using a ~4,500 bp ribosomal DNA marker that included most of the eukaryote SSU and LSU rRNA genes and the complete ITS region. Krehenwinkel et al. (2019a,b) showed that long-read barcoding using the MinION of a similar ribosomal DNA region, spanning ~4,000 bp, has great potential for in situ species identification too, although degraded DNA can be a limiting factor in generating long-read barcodes. Small scale projects that do not require such high throughput could alternatively focus on using the newly released Flongle flow cell instead of a traditional MinION flow cell, at a cost of $90 instead of $475–$900 (depending on number of flow cells purchased), respectively. The membrane in this flow cell contains less nanopores to generate a throughput of 1–2 Gb, to accommodate projects with lower throughput demands (Krehenwinkel et al., 2019b). There is a trade-off between flexibility, where a project can sequence samples whenever they want, and cost-effectiveness, where a project can sequence as many samples as possible to get the lowest possible costs per sample. The latter is highly unlikely to be necessary and achievable in very remote regions, given the previously mentioned time restrictions this places on projects.

CONCLUSION

The use of the MinION opens up exciting possibilities for next-generation biomonitoring. The high efficiency of the MinION consensi compared to the Sanger sequences shows that the MinION can be used to identify diverse nematode species. Extrapolating our results to potential application in a field setting, our results suggest that barcoding with the MinION can generate enough reads for reliable identification within 15 min, assuming good DNA quality and depending on the number of samples that are multiplexed. Our study shows the potential for barcoding eukaryotes and can aid biomonitoring of invertebrate species. Optimizing portable sequencing methods for nematode identification is the first step to sequencing nematode species in the field. One of the challenges ahead for TGS of nematode species lies in the identification of nematodes species from mixed samples from complex natural environments like soil, marine sediments or feces. This challenge could be overcome in several ways. Improvements in the underlying MinION technologies is crucial and will improve accuracy and decrease error rates, just as previous improvements have already shown (Eisenstein, 2019). Further optimization of the bioinformatics analyses is also of high importance. Improved algorithms will lead to higher accuracy in species identification. These improvements will open up possibilities likely near real-time genetic identification of nematodes from e.g., soil or feces, which would allow for analyses.

\(^{1}\)https://store.nanoporetech.com/flow-cell-wash-kit-r9.html
of soil nematodes as indicator of soil environment disturbance or rapid parasite identification.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are publicly available. The MinION dataset can be found in the European Nucleotide Archive under the project ID PRJEB37489 (samples ERS4397495, ERS4397496, ERS4397497, ERS4397498; and runs ERR4030416, ERR4030415, ERR4030417, ERR4030418, respectively). The Sanger sequences are available at GenBank under the accession numbers MT246663, MT246664, MT246665, and MT246666. All the code used for the bioinformatic analyses and additional files necessary to replicate the analyses, including a detailed explanation of dataset content, can be found on https://github.com/ieknot/MinION-DNA-barcoding-of-nematodes.

AUTHOR CONTRIBUTIONS

SW, RR, and IK designed the study. IK performed DNA extraction, prepped the Sanger sequencing samples, and performed the nanopore experiments. GZ, RR, and IK optimized the primers. GZ and IK performed PCRs. GW performed the nanopore experiments. GZ, RR, and IK extraction, prepped the Sanger sequencing samples, and SW, RR, and IK designed the study. IK performed DNA sequencing. RR wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the World Wildlife Fund-Netherlands and the University of Amsterdam.

ACKNOWLEDGMENTS

We thank Anne-Sophie Crunchant and Noémie Bonnin for discussions about and feedback on this manuscript, and Yannis Ancele for his help with some of the bioinformatics. We thank Anna de Luis Otero, Prof. Arthur Edison and Prof. Ian Chin-Sang for permission to use the pictures of A. simplex, P. redivivus, and C. elegans, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2020.00100/full#supplementary-material

The MinION consensus FASTA files correspond to the species as follows: Data Sheet 1 is A. simplex, Data Sheet 2 is P. redivivus, Data Sheet 3 is T. aceti and Data Sheet 4 is C. elegans.

REFERENCES


