



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Rapid in situ identification of biological samples via DNA amplicon sequencing using miniaturized laboratory equipment

Citation for published version:

Pomerantz, A, Sahlin, K, Vasiljevic, N, Seah, A, Lim, M, Humble, E, Kennedy, S, Krehenwinkel, H, Winter, S, Ogden, R & Prost, S 2022, 'Rapid in situ identification of biological samples via DNA amplicon sequencing using miniaturized laboratory equipment', *Nature Protocols*, pp. 1-29.
<https://doi.org/10.1038/s41596-022-00682-x>

Digital Object Identifier (DOI):

[10.1038/s41596-022-00682-x](https://doi.org/10.1038/s41596-022-00682-x)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Protocols

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Rapid *in situ* identification of biological samples via DNA amplicon sequencing using miniaturized laboratory equipment

Aaron Pomerantz^{1,2*}, Kristoffer Sahlin^{3*}, Nina Vasiljevic⁴, Adeline Seah⁵, Marisa Lim⁵, Emily Humble⁶, Susan Kennedy^{7,8}, Henrik Krehenwinkel⁸, Sven Winter⁹, Rob Ogden^{6**} and Stefan Prost^{10,11,12#**}

¹Department of Integrative Biology, University of California, Berkeley, CA-94720, USA

²Marine Biology Laboratory, Woods Hole, MA-02543, USA

³Department of Mathematics, Science for Life Laboratory, Stockholm University, 106 91 Stockholm, Sweden

⁴Zurich Institute of Forensic Medicine, University of Zurich, Winterthurerstrasse 190/52, 8057 Zurich, Switzerland

⁵Wildlife Conservation Society, Zoological Health Program, Bronx Zoo, 2300 Southern Blvd, Bronx, NY, 10460, USA

⁶Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom

⁷Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan

⁸Department of Biogeography, University of Trier, 54296 Trier, Germany

⁹Senckenberg Biodiversity and Climate Research Centre, 60325 Frankfurt am Main, Germany

¹⁰LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, 60325 Frankfurt, Germany

¹¹Natural History Museum, Vienna, Central Research Laboratories, Burgring 7, 1010 Vienna, Austria

¹²South African National Biodiversity Institute, National Zoological Garden, Pretoria 0002, South Africa

* Contributed equally

** Contributed equally

#Correspondence: stefanprost.research@protonmail.com

Editor's Summary This protocol provides steps for DNA amplicon sequencing using miniaturized laboratory equipment for rapid *in situ* identification of biological specimens, at local institutions, field stations, or classrooms.

Tweet DNA amplicon sequencing using miniaturized laboratory equipment for rapid *in situ* identification of biological specimens, at local institutions, field stations, or classrooms. #FieldGenomics

Related Links

Sahlin, K., Lim, M. C. W. & Prost, S. NGSPECIESID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecology and Evolution* 11, 1392–1398 (2021).

Pomerantz, A. et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience* 7, (2018).

Vasiljevic N, Lim M, Humble E, Seat A, Kratzer A, Morf NV, Prost S, Ogden R (2021) Developmental validation of Oxford Nanopore Technology MinION sequence data and the NGSspeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Science International: Genetics*, 53, 102493

Keywords

Nanopore, portable, amplicon, DNA barcoding, biodiversity, biosurveillance

Abstract

In many parts of the world, human-mediated environmental change is depleting biodiversity faster than it can be characterized, while invasive species cause agricultural damage, threaten human health, and disrupt native habitats. Consequently, the application of effective approaches for rapid surveillance and identification of biological specimens is increasingly important to inform conservation and biosurveillance efforts. Taxonomic assignments have been greatly advanced using sequence-based applications, such as DNA barcoding, a diagnostic technique that utilizes polymerase chain reaction (PCR) and DNA sequence analysis of standardized genetic regions. However, in many biodiversity hotspots, endeavors are often hindered by a lack of laboratory infrastructure, funding for biodiversity research, and restrictions on the transport of biological samples. A promising development is the advent of low-cost, miniaturized scientific equipment. Such tools can be assembled into functional laboratories to carry out genetic analyses *in situ*, at local institutions, field stations, or classrooms. Here, we outline the steps required to perform amplicon sequencing applications, from DNA isolation to nanopore sequencing and downstream data analysis, all of which can be conducted outside of a conventional laboratory environment using miniaturized scientific equipment, without reliance on internet connectivity. Depending on sample type, the protocol (from DNA extraction to full bioinformatic analyses) can be completed within 10 hours and with appropriate quality controls can be used for diagnostic identification of samples independent of core genomic facilities that are required for alternative methods.

Introduction

Biodiversity loss has increased rapidly during the past decades. Natural habitats are diminishing at an unprecedented rate and extinctions are predicted for many taxa across the tree of life^{1,2}. Furthermore, the spread of invasive alien species into new environments can cause agricultural damage, jeopardize human health, and negatively impact native biodiversity^{3,4}. Conservation and biosurveillance-focused groups are faced with the immense task of characterizing baseline biodiversity data, documenting how communities change, and detecting harmful invasive pests and pathogens. To do so, it is imperative that the tools and protocols are standardized and are time- and cost-efficient.

The advent of high-throughput sequencing (HTS) technologies, coupled with the development of standardized DNA marker systems, known as DNA barcodes⁵, have greatly facilitated large-scale monitoring and community-level assessment of species diversity. Ever-growing reference databases (such as the Barcode of Life Data System (BOLD)⁶ and the NCBI GenBank database⁷) and the ability to simultaneously analyze high numbers of samples⁸ have further increased the utility of DNA sequencing as a valuable tool for biodiversity characterization and and biosurveillance for invasive alien species.

However, many biodiversity hotspots and inspection sites are located in regions that lack readily available access to traditional HTS platforms. A common alternative, to transport samples abroad for analysis, has become increasingly restricted due to international conventions such as CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora; <https://www.cites.org>) or the Convention on Biological Diversity's Nagoya Protocol on Access and Benefit Sharing (<https://www.cbd.int/abs/>). While these are important measures to control the shipment of natural and potentially protected wildlife, they can increase the time it takes to generate, analyze, and report on the biological information. The ability to sequence biological samples within the country of origin with portable, inexpensive laboratory equipment can bring significant benefits for biodiversity monitoring and explorations, while simultaneously creating opportunities for developing local scientific capacity. The latter is a key aspect of the Nagoya Protocol and the United Nations Sustainable Development Goals (SDGs; <https://sdgs.un.org/goals>). Additionally, rapid characterization of invasive pests and pathogens at or near the site of detection via molecular analyses can be a useful means of mitigating their negative impacts on health, ecosystems, and economies⁹. Thus, these technologies offer a chance to empower local scientists and conservation agencies that currently rely on international research facilities. Miniaturized instruments, such as portable thermocyclers (miniPCR bio, MiniOne Systems) and nanopore-based nucleic acid sequencing devices (Oxford Nanopore Technologies), have gained in popularity, as they are relatively inexpensive compared to traditional, bulky molecular laboratory equipment and allow for *in situ* processing of genetic material¹⁰. Portable genomics laboratories have been deployed around the world in a diverse set of ecosystems and settings, including the rainforests of Tanzania¹¹, Ecuador¹² and Madagascar¹³, out at sea¹⁴ and in the Antarctic¹⁵; as well as being used to monitor disease outbreaks such as Ebola in West Africa¹⁶ or Zika in Brazil¹⁷, and for educational programs^{18,19}.

While several studies have utilized nanopore-based amplicon sequencing, there are various types of portable genomics equipment available, as well as different DNA isolation, amplification, and bioinformatics strategies spread amongst the literature. As such, in this protocol we aim to synthesize the current literature and report leading practices so that anyone, anywhere can carry out their own DNA amplicon sequencing projects using miniaturized laboratory equipment *in situ*.

Development of the protocol

Central to advancements in miniaturized genomics equipment is the small MinION sequencing device (Oxford Nanopore Technologies (ONT)), which uses 'nanopore sequencing', a process by which changes in the ionic current measured when a single-stranded DNA fragment is funneled through a biological pore in the device's membrane, can be converted into a nucleotide sequence²⁰. ONT's MinION is a USB-powered, portable sequencing platform that was launched in 2014. Since that time, there have been substantial improvements made in sequencing yield and quality^{20,21}. ONT also launched an inexpensive low throughput flow cell called 'Flongle', which can further reduce costs of targeted sequencing projects, such as DNA barcoding projects on the MinION platform. Although raw sequence accuracy remains relatively low for the MinION compared to other HTS technologies^{21,22}, highly accurate consensus sequences can be generated for DNA amplicons^{12,23–25}. Furthermore, the development of a new pore with two reader-heads (10.x) has drastically decreased the error rate and increased the consensus barcode accuracy¹⁴. Following several years of extensive experimentation and refinement of MinION methods by the scientific community, combined with the optimization of other commercially available field-deployable laboratory equipment, there is now an opportunity to consolidate a step-by-step protocol for DNA

amplicon sequencing using miniaturized laboratory equipment. This protocol comprises currently available best practices for performing amplicon sequencing experiments outside of conventional laboratory environments based on research groups who have processed samples and applied portable genomics tools under field conditions such as ^{12,18,23,24,26,27}. Additionally, we provide cost-effective strategies for multiplexing high numbers of samples for each sequencing run through user-customized indexing of amplicons, and provide instructions for the loading of both the standard flow cell and the inexpensive lower throughput Flongle flow cell. Finally, as long-read platforms such as ONT result in higher raw-read error relative to other sequencing platforms and require various processing steps, we present a simplified downstream bioinformatics workflow for the demultiplexing, polishing and *de novo* assembly of raw data into accurate consensus amplicon sequences. The presented approach has recently been validated for use in non-human forensic applications and has been shown to be highly reliable²⁵.

Applications

We anticipate that this approach can be adapted for various projects that aim to perform real-time DNA amplicon sequencing, such as within-country biodiversity assessment efforts, on-site invasive species detection, laboratory-based diagnostics and educational genomics programs, as users can tailor the protocol to assess any taxa of interest in a relatively time- and cost-efficient manner.

Limitations

We note that nanopore technology, chemistry, and software are ever-changing and improving at a rapid pace, and thus a user interested in these methods should also consult the most up-to-date literature and manufacturer protocols. We also recognize that with regard to miniaturized field-deployed tools, which are not made for high-throughput processing on samples, there can be a trade-off between quickly characterizing a few to hundreds of biological samples and conducting more thorough biodiversity assessments that may involve several thousands of samples, which could impact productivity and operation costs. As such, the user should take into account the goal of their project, the number of samples to be processed (e.g. processing capacity using miniaturized 16-32 well PCR machines compared to full-sized benchtop 96 well PCR machines), and the available storage options for samples and reagents required to carry out DNA amplicon sequencing experiments outside of a conventional laboratory environment.

~~Experimental design~~ Overview of the Protocol

Here we present a protocol for designing and executing rapid, multiplexed amplicon sequencing using miniaturized laboratory equipment, including ONT's MinION sequencing platform (Figure 1f). The approach can be used in typical molecular laboratories or non-conventional laboratory spaces, such as remote field stations, or classroom settings. In the *Procedure* section, we mark the use of relatively inexpensive mobile equipment and steps that reduce complexity using the **FIELD** flag to enable users to carry out the molecular and bioinformatic processing in areas with limited infrastructure. Schematics of the laboratory processing and the bioinformatic analyses can be found in Figure 2 and Figure 3, respectively.

Experimental design

Commented [HC1]: As well as Applications and limitations, we would usually have a section on 'Alternative methods', in the Introduction before the Experimental Design. In which you can fully compare and contrast your protocol to other available methods with which to achieve similar results, explaining the advantages and disadvantages of your protocol so that readers would know when it is appropriate, or inappropriate to choose this protocol over others, and why? Can you add this sort of information here? Are there other similar protocols people may follow that you can cite and discuss?

Commented [HC2]: Our format will not allow for this flag to be used as real flag (i.e. formatted like the CRITICAL STEP flag is). We can use the word FIELD in normal text and I can ask for it to be made into a different colour and emboldened, but just to make you aware we can't add new 'flags'...

Transportation to the field

Users should consider trade-offs between the time it may take to carry out experiments under field conditions versus carrying out experiments at nearby facilities that may have access to more stable storage and freezer options^{12,15}. If equipment needs to be transported into the field, either to conduct on-site analyses or to perform analyses in a field-station with limited infrastructure, portable equipment can be transported in luggage to the site of interest, and for additional safeguarding, can be loaded into protective cases, such as a Pelican case (Pelican, Torrance, USA)^{12,16}. Cold chain reagents and flow cells can be packed into polystyrene boxes with ice or cool packs and sealed for transportation^{12,16}. Users can consider additional options for maintaining reagents at 2°C-8°C for several days with portable refrigeration systems (e.g. the Crêdo Cube (Pelican BioThermal)). While DNA primers and most non-enzymatic reagents are fairly stable and can be transported at room temperature for hours or sometimes even days, many enzymes, such as ligases require sub-zero storage. Suboptimal transport or storage conditions can lead to a decrease in activity or even loss of functions in the enzymes (see e.g.¹²). We recommend checking the storage conditions for all reagents and ensuring suitable conditions during transport. Under cold environmental conditions, such as the Antarctic, the temperature of the MinION and other electronic equipment can be regulated using hand warmers and insulating materials¹⁵. Long-term storage of molecular biology reagents that require stable cold temperatures can still present challenges for nanopore sequencing projects under non-ideal environmental conditions. Our group, along with other field-focused researchers, have transported a variety of polymerases into the field for DNA barcoding, and here we report the use of either DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) or Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA, USA) for carrying out PCR amplifications. We note that these polymerases are rather expensive and that a variety of commercial polymerases and PCR master mixes are available, and recommend the user should seek out the one most appropriate for their experiment and price range. Several field friendly lyophilized reagents stable for long-term storage at ambient temperature have been reported, including lyophilized PCR reagents¹¹. ONT offers a field sequencing kit (SQK-LRK001) for cold-chain-free library preparation, but this option is not recommended for the presented DNA barcoding protocol, as the kit relies on transposase activity which cleaves template molecules.

Our group, along with other field focused researchers, have transported a variety of polymerases into the field for DNA barcoding, and here we report the use of either DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) or Q5 Hot Start High Fidelity 2X Master Mix (NEB, Ipswich, MA, USA) for carrying out PCR amplifications. We note that these polymerases are rather expensive and that a variety of commercial polymerases and PCR master mixes are available, and recommend the user should seek out the one most appropriate for their experiment and price range. Another interesting development that requires further testing is the use of lyophilized polymerases, which would reduce cold chain requirements. Several field-friendly lyophilized reagents stable for long-term storage at ambient temperature have been reported, including lyophilized PCR reagents¹¹. ONT offers a field sequencing kit (SQK-LRK001) for cold-chain-free library preparation, but this option is not recommended for the presented DNA barcoding protocol, as the kit relies on transposase activity which cleaves template molecules.

Assessment of DNA, either after amplification or during the sequencing library preparation, can be carried out using miniaturized gel-electrophoresis systems, such as the blueGel (miniPCR), MiniOne system (MinOne), or BentoLab (Bento Bioworks Ltd., London, United Kingdom), or quantification instruments such as the Qubit (Thermo Fisher Scientific) or the TapeStation (Agilent) in case these are available. To

Commented [HC3]: Can you comment on how you would achieve this in a non-lab. i.e. field setting?

reduce time and resources running gels, users can check a small subset of reactions on the gel electrophoresis to ensure that there is no widespread amplification failure or presence of amplicons in negative controls.

Sample collection

Sample collection permits should be authorized by the appropriate agencies and samples should be treated ethically in accordance with community guidelines. Whenever possible, samples that are collected should be properly vouchered, meaning that they are labelled, preserved and deposited at publicly accessible research collections, so that they can serve as verifiable and permanent records.

Primer design

~~To amplify DNA barcodes using PCR, one pair of amplification primers per DNA barcode is required. These primers are short oligonucleotides placed in the 5' and 3' flanking regions of the DNA barcode. These flanking regions should ideally be conserved across the taxa of interest. The length of the primers should be between 16 and 36 bp with an annealing temperature (T_m) of 55–70°C and a GC content of 45–60%^{28,29}. These parameters should be well matched for both primers in a pair for the best chance of efficiency and specificity²⁸. In general, annealing temperature differences between the primers, and in case of multiplexing (the simultaneous PCR amplification of several amplicons within the same reaction) also the different primer pairs should not exceed 2–5°C. Furthermore, secondary structures such as hairpin loops or primer dimers should be avoided. Although primer sequences can be designed manually based on a reference sequence, we recommend using primer design tools such as Primer3³⁰ that are designed to identify suitable primer pairs taking multiple parameters into account. Whether the primers amplify only the target region or also other parts of the genome can be tested using Primer-BLAST³¹. For more details on primer design, especially in applications that require multiplex PCR or universal primers for a large variety of organisms, please consult^{28,29}. In order to multiplex several samples in the same sequencing run, universal tails have to be added to the primer sequence see below (Amplification, indexing and pooling) and Procedure Steps 1–3.~~

Commented [HC4]: What types of samples can this Protocol be applied to? Which types have you used the protocol on, and what do you expect it could be applied to in future?

DNA extraction (Procedure Step 4)

The experiment requires the extraction of DNA from the samples of interest. There are numerous methods and commercial options to carry out DNA extraction, several of which have been applied under field conditions, such as QuickExtract™ (Lucigen)^{14,24}, spin column-based nucleotide isolation kits^{12,13,15,23} and the HotSHOT extraction³². We encourage the user to seek out an extraction protocol that is most suitable for their sample type and price range. We have tested a variety of commercially available extraction kits or custom solutions and found them suitable for extraction of DNA from various sample types (such as tissue, fecal, plant and soil samples) under field conditions, which produced sufficient yields and qualities for subsequent amplification of target regions (see reagents list and references^{12,23,27}). These include standard kits such as the Qiagen QIAmp DNeasy Blood & Tissue kit and the Bio-Rad Chelex® 100 resin extraction. The latter is more cost-efficient (approximately \$0.17 USD per sample²⁷) and can be carried out using only a portable thermocycler. Alternatively, to preserve intact specimens for morphological examination, QuickExtract™ (Lucigen) can be used for DNA extractions, which does not require centrifugation^{14,24}. Buffer-based extraction methods such as the HotSHOT protocol³³ are also fast and easy to carry out using a thermocycler. Many of the aforementioned extraction kits and methods have been shown to work with non-invasively sampled DNA sources (e.g from scat, hair, feathers), which may

contain PCR inhibitors or be degraded from environmental exposure²⁷. ~~The presence of contaminants can decrease efficiency of enzymatic steps during ONT library preparation and interfere with sequencing pores in a nanopore flow cell. However, because this protocol performs sequencing of PCR-amplified and cleaned products, rather than material directly from the DNA isolation, we have not experienced reduction in sequencing efficiency or consensus read quality for these extractions²⁰.~~

Commented [HC5]: Is this meant to be deleted? It looks like useful info to keep in?

PCR amplification; primer design, indexing and pooling (Procedure Steps 5 - 28)

After isolating DNA from the samples, genetic regions of interest (so called DNA barcodes) are amplified via polymerase chain reaction (PCR). DNA barcodes can be designed to encompass coding or non-coding regions, and are usually chosen to include a region with high discriminatory power for the taxon group investigated^{5,34}. PCR amplification can be efficiently performed using inexpensive mobile thermocyclers. For instance, researchers have reported success using the miniPCR (<https://www.minipcr.com>), MiniOne (<https://theminione.com>) and BentoLab (<https://www.bento.bio>) for thermocycling steps under field conditions (see e.g. ^{12,14,26,35}). These devices can also serve as miniature heating blocks, be programmed with a mobile phone or laptop, and run off external battery power in remote environments (see Figure 1). While standard amplicons (amplified DNA regions) for DNA barcoding are generally ~300-900 bp in length, sequencing on the MinION also allows the user to generate long-range amplicons (thousands of bases long), which can increase phylogenetic resolution compared with shorter standard barcode sequences (see e.g. ^{23,36}). We have to caution that the presented protocol does not allow for the utilization of amplicons shorter than 100 bp, which is the minimum read length that ONT recommends for their Guppy basecaller (<https://community.nanoporetech.com/support/articles/what-is-the-minimum-length-for-the-basecaller>). Alternatively, methods such as rolling circle PCR³⁷ or amplicon ligation³⁸ have been suggested to overcome this issue.

Commented [HC6]: Can you clarify the use of the term DNA barcodes? Are these just target sequences, i.e. taxon specific sequences that you are trying to identify? I think the use of the term barcode is a bit confusing – are these different from DNA barcoding tags, commonly known as DNA barcodes, used in high throughput pooling and screening sequencing experiments? Apologies if I have missed the point, but would it be more straightforward to just refer to them as the target sequences?

To amplify DNA barcodes using PCR, one pair of amplification primers per DNA barcode is required. These primers are short oligonucleotides placed in the 5'- and 3'- flanking regions of the DNA barcode. These flanking regions should ideally be conserved across the taxa of interest. The length of the primers should be between 16 and 36 bp with an annealing temperature (T_m) of 55-70°C and a GC-content of 45-60%^{28,29}. These parameters should be well matched for both primers in a pair for the best chance of efficiency and specificity²⁸. In general, annealing temperature differences between the primers, and in case of multiplexing (the simultaneous PCR amplification of several amplicons within the same reaction) also between the different primer pairs, should not exceed 3-5°C. Furthermore, secondary structures such as hairpin-loops or primer dimers should be avoided. Although primer sequences can be designed manually based on a reference sequence, we recommend using primer design tools such as Primer3³⁰ that are designed to identify suitable primer pairs taking multiple parameters into account. Whether the primers amplify only the target region or also other parts of the genome can be tested using Primer-BLAST³¹. For more details on primer design, especially in applications that require multiplex PCR or universal primers for a large variety of organisms, please consult ^{28,29}.

In order to multiplex several samples in the same sequencing run, universal tails have to be added to the primer sequence see below (Amplification, indexing and pooling) and Procedure Steps 1-3. In order to pool amplicons from numerous samples on a single sequencing flow cell, indexes must be added to the amplicons. This can either be performed using a 1-Step or 2-Step PCR protocol (Figure 2). The 1-Step protocol (see e.g. ^{23,24}) may be preferable if a single primer set is used for the study, as it is faster than the 2-Step PCR protocol, requires fewer reagents, and reduces the chance of chimeric-amplicon formation **as** it only includes a single PCR step³⁹. Chimeric-amplicons primarily form due to incomplete PCR products

serving as primers on template DNA⁴⁰. They are a particular problem in community metabarcoding experiments, where multiple taxa are amplified in a community sample^{39,41}. Creation of chimera between many different taxa can significantly inflate the recovered sequence diversity. However, in single specimen amplicon sequencing, chimera are of minor importance, as only a single dominant DNA molecule is amplified. The possibility of chimeric sequence formation should [however](#) be kept in mind in single specimen barcoding, when DNA is extracted from complete specimens and for example the gut content of the specimen is included in the DNA extract. In the presented protocol chimeric reads are filtered out during the cluster formation step [in the Bioinformatics section below](#). In the case of 1-Step PCR, unique indexes are added to the primer sequence directly. Alternatively, the 2-Step PCR protocol may be the preferred option for studies that aim to use multiple primer sets (e.g. for multi-locus PCR amplification or the application of primer cocktails, which is common for universal Cytochrome c oxidase subunit 1 (COI) amplification⁵). In the 2-Step PCR approach, universal tail sequences are added to the amplification primers. These universal tails then allow for the addition of index sequences to the ends of the DNA fragment in a second PCR amplification step. The 2-Step PCR protocol allows the user to share indexes between projects and research groups, thereby further reducing the cost of indexing. Furthermore, using different indexes on either end of the amplicon (called dual-indexing) enables cost-effective sequencing for large-scale sampling in both indexing setups (the 1-Step and the 2-Step protocol)²⁴. However, while adding indexes to the primers allows for multiplexing of many samples, it might reduce the amplification efficiency during the PCR⁴². This can lead to lower demultiplexing rates. With decreasing error rates of the newly developed flow cells, index lengths should be reduced to increase the amplification efficiency⁴³. In the presented protocol we provide the option to incorporate ONT's standard universal tails in the primer sequences, which makes the primers compatible with inexpensive custom indexes and/or ONT's DNA barcoding kit (see [Reagents list](#)). In case researchers plan to also sequence the amplicons on an Illumina HTS platform, Illumina's universal tails (see ⁴⁴) can be used for both Illumina and ONT sequencing. Tag jumping (false tag combinations) has been identified as a potential problem when multiplexing multiple samples on a single sequencing run⁴². We recommend including no template control PCRs (negative controls where DNA template is replaced by water) when running the protocol to monitor for cross-contamination, tag jumping and reagent impurities. Ideally, all or a subset of the blanks should also be sequenced along with the actual samples.

~~Our group, along with other field-focused researchers, have transported a variety of polymerases into the field for DNA barcoding, and here we report the use of either DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) or Q5 Hot Start High Fidelity 2X Master Mix (NEB, Ipswich, MA, USA) for carrying out PCR amplifications. We note that these polymerases are rather expensive and that a variety of commercial polymerases and PCR master mixes are available, and recommend the user should seek out the one most appropriate for their experiment and price range. Another interesting development that requires further testing is the use of lyophilized polymerases, which would reduce cold chain requirements. Assessment of DNA, either after amplification or during the sequencing library preparation, can be carried out using miniaturized gel electrophoresis systems, such as the blueGel (miniPCR), MiniOne system (MinOne), or BentoLab (Bento Biosystems Ltd., London, United Kingdom), or quantification instruments such as the Qubit (Thermo Fisher Scientific) or the TapeStation (Agilent) in case these are available. To reduce time and resources running gels, users can check a small subset of reactions on the gel electrophoresis to ensure that there is no widespread amplification failure or presence of amplicons in negative controls.~~

Commented [HC7]: For amplification of?

Commented [HC8]: Please add which Procedure step this is.

Sequencing library preparation (*Procedure Steps 29- 51*)

The presented sequencing library preparation protocol is based on ONT's library preparation manual for the Ligation Sequencing Kit (SQK-LSK110), experiences gained from working in the field (see ^{11,12,18,23,26}) and MinION-based DNA barcoding in the laboratory^{25,27}. This section of the protocol largely follows the ONT Ligation Sequencing Kit 1D (SQK-LSK110) library preparation instructions but omits the NEBNext FFPE Repair Mix (M6630) at the 'DNA repair and end-prep' step. The library preparation can be carried out in approximately one and a half hours in remote conditions, and requires pipettes, a small centrifuge, a miniature thermocycler and a small magnetic rack for bead clean-ups (Figure 1). We have included optional steps that require additional equipment in case the protocol is performed in a fully equipped laboratory.

Sequencing (*Procedure Steps 55 - 63*)

Running the MinION Mk1B sequencer can be performed using external battery power on either a laptop (with sufficient space and speed; see <https://nanoporetech.com/community/lab-it-requirements>) or on standalone devices such as ONT's Mk1C. Alternatively, DIY setups using mini-servers such as the Nvidia Jetson Xavier NX can be used effectively as an inexpensive means to running the MinION and the basecalling (see e.g. https://github.com/sirselim/jetson_nanopore_sequencing). Sequencing using the Mk1B sequencer requires the use of the MinKNOW software, which does not require internet access if the proper offline version is installed. The MinION sequencing run generates data in real-time and depending on the number of samples and read output required, can be run for a few hours (for smaller runs) or in case of high-throughput multiplexing (e.g. hundreds to thousands of samples), up to 72h. Additionally, nuclease-based washes can be applied to used flow cells for their re-use (Flow Cell Wash Kit: EXP-WSH004). In our experience with current versions, the MinION produces about 10,000 sequenced reads per minute for ~500bp long DNA barcodes on the standard MinION flowcell with pores R9.4 and R10.3 (the Flongle has less pores and therefore produces less reads). However, this number should be treated with caution as it depends on the number of active pores and may decrease over time. We recommend calculating the number of reads required after sequencing with ~1,000 per individual amplicon to obtain around 300-500x coverage per sample, ~~to account for slightly uneven pooling of samples, which This should provide the user with a good representation for each DNA barcode even in the presence of slightly uneven pooling of the different amplicons.~~ Thorough testing (using R9.4 flow cells) has shown that 300-500x produces highly reliable consensus sequences²⁵. However, in some cases as little as 10-50x coverage was sufficient to produce consensus sequences with 100% accuracy^{23,25,43}. This will likely improve further with decreasing error rates of the newer flow cell generations. ~~This should provide the user with a good representation for each DNA barcode even in the presence of slightly uneven pooling of the different amplicons.~~ ONT provides updated videos on how to load the different flow cell types and other useful manuals on their YouTube channel (<https://www.youtube.com/channel/UC5yMIYjHSgFfZ37LYq-dzig>).

Commented [HC9]: What do you mean by this here?

Bioinformatics (*Procedure Steps 64 - 70*)

Basecalling. Recent developments in ONT's basecalling algorithms (the conversion of the MinION's ionic current profiles to fastq sequences) allow for live basecalling during sequencing using MinKNOW. We recommend using live basecalling for smaller scale sequencing runs of 1-5h, or when the sequencing is carried out using ONT's Mk1C device. However, basecalling can also be performed at a later stage on a laptop or mini-server if needed, especially for long sequencing runs producing several gigabases of

Formatted: Font: Italic

Formatted: Font: Italic

data. ONT offers several options for basecalling. In general, High-Accuracy models produce highly accurate basecalls at the expense of speed, while Fast models produce basecalls with lower accuracy, but are significantly faster than High-Accuracy models. The Guppy basecaller further offers the option to use graphics card chips (GPUs) instead of CPUs for basecalling, which increases the basecalling speed even further. Today, (small) computers with many GPUs (such as the Nvidia Jetson Xavier NX) can be purchased for relatively low prices.

Demultiplexing. Following basecalling, we show the use of two different demultiplexing tools, which work for ONT indexing kits (e.g. the PCR Barcoding Expansion 1-96 kit) and custom indexing: minibar²³ (<https://github.com/calacademy-research/minibar>) and Guppy (Oxford Nanopore Technologies, UK). These tools were specifically developed for error-prone nanopore sequencing data. In addition to demultiplexing of dual or single index libraries, minibar also offers the option to demultiplex samples based on primer sequences, in case different DNA barcodes were amplified in single reactions (multiplex PCR). There are other options available see e.g. ⁴⁵. We recommend demultiplexing the read data before quality assessment and filtering. This allows the user to pool barcodes of varying sizes in the same multiplex PCR amplification, and carry out filtering for each amplicon separately afterwards.

Quality assessment and further downstream processing. NanoPlot (<https://github.com/wdecoster/Nano-Plot>) outputs summary statistics and plots showing the quality of the sequenced reads. For downstream processing of raw reads, we use NGSspeciesID, a tool developed for the generation of highly accurate consensus sequences from third-generation long-read sequencing technologies⁴⁶. We note that there are a variety of bioinformatic workflows that have been reported for generating accurate consensus amplicon sequences using the MinION ^{12,24,26,27,36,43,45}. Pipelines such as ONTrack²⁶ or Consension³⁶ have been shown to work more reliably than software used in the early stages of MinION-based DNA barcoding (such as the workflow presented in ¹²). NGSspeciesID has been benchmarked against current pipelines, and all have been found to be comparable for generating highly accurate consensus sequences⁴⁶. NGSspeciesID offers many options (including read filtering on quality and size, subsampling, primer site removal) and is easy to run, install and scale up. It can also be used for any type of DNA amplicon and offers the possibility to call consensus sequences from multiplexed samples (where more than one DNA barcode is generated per sample), and is thus not limited to a single specific DNA barcode such as other tools, e.g. ⁴³. Furthermore, it is the only consensus caller for ONT data that has formally been validated for use in non-human forensic applications²⁵. We thus decided to use it in this protocol. It first removes low quality reads and PCR chimeras from the data and subsequently, automatically carries out read clustering, consensus calling, polishing and optional primer removal (see [description](#) below). The final polished consensus sequences can subsequently be compared against different databases, such as the Barcode of Life Data System (BOLD)⁶ or the National Center for Biotechnology Information (NCBI) GenBank nucleotide database⁷. BOLD offers a large compilation of COI sequences from many taxa, while GenBank includes a variety of different markers. In addition to the matching of sequences based on similarity, BOLD uses Neighbor-Joining tree reconstruction with closely related species to infer species identifications. We also outline different statistics used to interpret database hits with BLAST⁷ and provide recommendations and guidelines on how to interpret the results.

NGSpeciesID ([Procedure Step 69](#))

To simplify the bioinformatic processing, we provide commands for NGSspeciesID, a tool designed to generate highly reliable consensus sequences for amplicon sequencing reads (such as DNA barcodes) generated using long read technologies⁴⁶. This tool has been benchmarked against other published pipe-

Formatted: Font: Italic

Formatted: Font: Italic

lines, such as ones that rely on multiple sequence alignment or operational taxonomic unit (OTU) clustering, and all have been shown to generate high accuracy consensus sequences⁴⁶. NGSspeciesID allows the user to first filter the raw demultiplexed read data for (a) read quality using Phred scores and (b) amplicon lengths to filter out chimeras, and then it carries out subsampling to obtain the preferred number of reads after the filtering. It then clusters the reads based on expected sequence similarity using read-specific error rates with the isONclust algorithm⁴⁷. A draft consensus is formed using spoa⁴⁸ for each cluster containing more reads than a user-defined abundance threshold (default: 10% of total sample read depth). Due to consensus calling *being* based on multiple reads, the consensus sequences formed by spoa typically have significantly lower error rates than the individual reads. NGSspeciesID then detects and merges any consensus sequences classified as reverse complements to each other using pairwise alignment with parasail⁴⁹. Two consensus sequences are merged if they have a sequence identity above a user-defined parameter given to NGSspeciesID (default 10%). Sequence identity is calculated as $1 - \text{mismatches}/\text{alignment-length}$, where mismatches can be either indels or substitutions. *If one or more of the consensus sequences are classified as reverse complements to each other, then the reads from all these clusters are merged into a single file.* Finally, all draft consensus sequences passing this step are polished using the original reads and medaka (<https://github.com/nanoporetech/medaka>) or Racon⁴⁸. To do so, reads on which the consensus sequence is based are first mapped to the respective consensus sequence using minimap2⁵⁰, and then processed either with medaka or Racon. The polished sequences are the final output of NGSspeciesID. Optionally, users can specify to automatically trim priming sequences from the consensus, in which case NGSspeciesID will also carry out an additional round of reverse complement detection and polishing.

Commented [HC10]: Deleted as repeated in section above.

Commented [HC11]: Deleted as already said a few sentences earlier.

Materials

REAGENTS

DNA extraction

- QIAmp DNeasy Blood and Tissue kit (Qiagen, cat. no. 69504)
or:
 - HotSHOT method ([after-sec](#)²⁷)
 - Chelex® 100 Resin (Bio-Rad, cat. no. 142-1253)
 - QuickExtract™ (Lucigen, cat. no. QE09050)
 - Quick-DNA Plant/Seed Miniprep Kit (Zymo Research, cat. no. D6020)

CRITICAL FIELD Qiagen's QIAmp DNeasy Blood and Tissue kit is a standard extraction kit, which works well for most tissue types. However, it is not the cheapest option. Inexpensive options such as the Chelex® 100 Resin have been shown to work well for many tissue types²⁷. Alternatively, buffer-based extractions, such as the HotSHOT method³³, are easy and quick to carry out and represent very cheap alternatives (see ³²). The user should take their sample type into consideration (e.g. plant, insect, scat, feather) to select the most appropriate DNA isolation method.

Amplicon generation

- DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, cat. No K9011)

or:

- Q5 Hot Start High-Fidelity 2X Master Mix (NEB, cat. no. M0494S)
- Assemble your own PCR mix (e.g. using Taq DNA polymerase, buffer, MgCl₂, dNTPs, etc.)

CRITICAL The primer annealing temperatures in our protocol are optimized for the two polymerase mixes listed above. While others may work, thermocycling conditions may need to be optimized.

CRITICAL FIELD Here, we present different options for hot-start high-fidelity polymerases but note that the user should decide on the polymerase that is most cost-effective and appropriate for their experiment. Although hot-start master mixes tend to be more expensive than other commercial polymerases, we recommend these options here because (1) they are pre-mixed formulations, allowing PCR to be quickly carried out after the addition of DNA and primers, (2) they can be used to generate either short or long-range amplicons, and (3) amplification only starts after a short heating phase, which can increase shelf life as the enzymes will not be inadvertently activated in cases of unreliable freezing conditions. We recommend testing other polymerases before applying them in the field.

- PCR primers with the option to include ONT's universal tail sequences if performing 2-step PCR, or with custom indexes if using 1-step PCR (ordered from suppliers such as Integrated DNA Technologies Inc. or Sigma Aldrich)
- Custom made indexing oligos (ordered from suppliers such as Integrated DNA Technologies Inc. or Sigma Aldrich)

or:

- **Optional:** PCR Barcoding Expansion 1-96 (Oxford Nanopore Technologies, cat. No. EXP-PBC096). To avoid confusion with DNA barcoding, this kit will be referred to as the ONT indexing kit from here on.

CRITICAL FIELD Custom barcoding using PCR primers that include ONT's universal tail sequences coupled with custom made indexing oligos is a cost-effective alternative to purchasing ONT's PCR Barcoding Expansion kits.

- Agencourt AMPure XP (e.g. Beckman Coulter, cat. no. A63881)

or:

- Home-made magnetic beads (https://s3-us-west-2.amazonaws.com/oww-files-public/f/f8/SPRI_buffers_v2_2.pdf) **CRITICAL FIELD** Home-made magnetic beads have been shown to be highly effective and can be used to reduce costs.

- Molecular Biology Grade Nuclease-free Water (e.g. Sigma-Aldrich cat. no. W4502-1L)
- Ethanol, absolute (e.g. Thermo Fisher Scientific, cat. no. BP28184)
- **Optional:** Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q32850)

Gel electrophoresis

- Gel Loading Dye, Purple (6×) (NEB, cat. no. B7024)
- 1 kb Plus DNA Ladder (NEB, cat. no. N3200S)
- Agarose (e.g. Thermo Fisher Scientific, cat. no. 16500100)

or:

- **CRITICAL FIELD** If no scale is available, pre-weighed and packed quantities of agarose can be used, such as GelGreen® Agarose Tabs™ (MiniPCR™, cat no. RG-1500-10) or home-made agarose packs.
- **CRITICAL FIELD** Gels can also be premade and transported to the field (keep cold and humid).
- 10× TBE Buffer (Thermo Fisher Scientific, cat. no. B52)

Commented [HC12]: Please see my earlier comment – it is confusing what you are referring to as DNA barcoding, and what not? E.g. this kit is called barcoding kit, yet you say you don't want to confuse it with barcoding so will call it something else?... Please try to clarify and use consistent terminology throughout.

- SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, cat. no. S33102)

ONT library preparation

- ONT Sequencing Kit (Oxford Nanopore Technologies, cat. no. SQK-LSK109 or SQK-LSK110)
- NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (cat. no. E7180S), which contains all NEB reagents needed for use with the Ligation Sequencing Kit
 - **Optional:** Alternatively, users can purchase individual NEBNext® products:
 - NEB Next Ultra II End-repair/dA-tailing Module (NEB, cat. no. E7546)
 - NEBNext Quick Ligation Module (NEB, cat. no. E6056)
- Freshly prepared 70% (vol/vol) ethanol in nuclease-free water
- Nuclease-free water (e.g. ThermoFisher, cat # AM9937)
- **Optional:** D1000 ScreenTape and Reagents (Agilent, cat. no. 5067-5582, 5067-5583)
- **Optional:** Flow Cell Wash Kit (Oxford Nanopore Technologies, cat. no. EXP-WSH004)

EQUIPMENT

Standard equipment and consumables

- P1000, P200, P10 pipette and filtered pipette tips
- 1.5 ml Eppendorf DNA LoBind tubes (e.g. Eppendorf, cat. no. 022431021)
- 0.2 ml strip thin-walled PCR tubes (e.g. Thermo Fisher Scientific, cat. no. AB0451)
- Standard thermocycler (e.g. Bio-Rad Laboratories, cat. no. 1861096EDU)
 - or:
 - **CRITICAL FIELD** Inexpensive and miniaturized thermocyclers: 16 or 32-well mobile thermocycler (e.g. miniPCR™ Thermal Cyclers, cat. no. QP-1016-01; or MiniOne PCR System, cat. no. M4000, or BentoLab)
- Benchtop microcentrifuge (e.g. Thermo Fisher Scientific mySPIN 6, cat. no. 75004061).

CRITICAL FIELD We have used microcentrifuge devices for this protocol ~~which that~~ spin at 2,000 x g, which we have found to be sufficient for spin column-based DNA isolation methods. Alternatively, a 3D-printed hand-powered centrifuge device can be used for spin column-based DNA isolations⁵¹.
- **CRITICAL FIELD** Use external battery packs to be able to run miniature PCR machines and micro-centrifuge if no constant external power source is available. We have used the RAVPower (model: RP-PB055) or Poweradd (model: Pilot Pro).
- Magnetic rack (e.g. Thermo Fisher Scientific DynaMag-2, cat. no. 12321D)
 - or:
 - DIY 3D printed magnetic rack (e.g. <https://thingiverse.com/thing:79424>)
- Ice or frozen cool packs
- **Optional:** Qubit™ 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA)

CRITICAL FIELD The protocol can be run without measuring exact DNA concentrations, but pooling amplicons in equimolar ratios will help to generate even sequencing coverage and increase the recovery rate for highly multiplexed libraries.
- **Optional:** TapeStation 2200 (Agilent)
- **Optional:** Hula mixer (gentle rotator mixer)

CRITICAL FIELD Alternatively mixing can be performed by hand.

- **Optional:** Scale. Not needed when agarose tabs are used or if agarose is individually packed in the correct quantities beforehand.
- **Optional:** Portable vortex shaker (see e.g. <https://gistgear.com/industrial/lab-equipment/lab-vortex-shakers>)

Gel electrophoresis

- Electrophoresis chamber such as blueGel™ electrophoresis (MiniPCR™, cat no. RG-1500-01)
or:
 - MiniOne System (MiniOne®, cat. no. M1000)
 - BentoLab (Bento Bioworks Ltd., London, United Kingdom)
- Cell phone with camera
- Standard kitchen Microwave



CRITICAL FIELD Alternatively, agarose gel solutions can be heated up in a water-baths on a camping stove.

- Laboratory bottles (e.g. VWR, cat no. VWRI215-1593-E)

MinION sequencing

- MinION (Oxford Nanopore Technologies, cat. no. MinION Mk1B)
CRITICAL FIELD Make sure to obtain the country permits from ONT for use of the MinION in the respective country.
- Laptop that meets ONT host computer specifications (including solid-state disk drive and sufficient memory and storage) for running MinION sequencing (<https://nanoporetech.com/community/lab-it-requirements>).
or:
 - MinION Mk1C is a standalone device for running MinION sequencing that can be used in place of the MinION Mk1B plus a laptop.
 - Custom built mini-server such as the Nvidia Jetson Xavier NX (<https://developer.nvidia.com/buy-jetson?product=all&location=US>)

MinION flow cell options

- Standard MinION flow cell R9.4 (Oxford Nanopore Technologies, cat. no. FLO-MIN106D)
or:
 - R10.x (Oxford Nanopore Technologies, cat. no. FLO-MIN111)

or:

- Flongle adapter plus flow cells (Oxford Nanopore Technologies, cat. no. FLGIntSP)

CRITICAL FIELD We recommend the use of the Flongle for smaller scale DNA barcoding projects, as it is a more cost-effective (but lower throughput) flow cell compared to the classic MinION flow cell.

SOFTWARE

CRITICAL FIELD Users going into remote field conditions should consider downloading relevant software and reference databases ahead of time.

Index design

- Barcode_Generator (https://github.com/lcomai/barcode_generator)

Converting double stranded DNA from μg to fmol

- Biomath Calculators <https://www.promega.com/resources/tools/biomath/>

MinION sequencing

- MinKNOW (Oxford Nanopore Technologies, UK)

CRITICAL FIELD Make sure you acquire the offline version of MinKNOW for sequencing in remote areas with limited to no internet access from ONT.

Basecalling

- **Optional:** Guppy (Oxford Nanopore Technologies, UK); can be used instead of live basecalling with MinKNOW.

Quality control

- NanoPlot (<https://github.com/wdecoster/NanoPlot>)

Read de-multiplexing

- Guppy (Oxford Nanopore Technologies, UK)
- or:
- minibar (<https://github.com/calacademy-research/minibar>)

DNA barcode read filtering and consensus generation

- NGSpeciesID (<https://github.com/ksahlin/NGSpeciesID>)

Comparison of the sequences against a database

- BLAST (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>)

DNA barcode databases

- BOLD (www.boldsystems.org)
- NCBI nucleotide (<https://www.ncbi.nlm.nih.gov/nucleotide>) or NCBI via BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

Reverse complementing DNA sequences

- http://www.bioinformatics.org/sms/rev_comp.html

REAGENT SETUP

Home-made size-selection beads

In order to prepare inexpensive home-made size selection beads, follow the protocol at: https://s3-us-west-2.amazonaws.com/oww-files-public/f/f8/SPRI_buffers_v2_2.pdf

70% (vol/vol) Ethanol

In order to prepare 70% (vol/vol) Ethanol, mix 70 ml of absolute Ethanol with 30 ml of nuclease-free water. **CRITICAL** Make fresh before use.

Primer and index dilutions and aliquots

To make 1:10 dilution of primer stocks (100 µM), mix 1 µl of primer stock solution with 9 µl of nuclease-free water to obtain a final concentration of 10 µM. **CRITICAL** We strongly recommend aliquoting primer and index dilutions to avoid contaminating the stock solutions.

1xTBE Buffer

To make 1xTBE Buffer, dilute the 10xTBE Buffer stock solution 1:10 ~~by mixing~~ 5ml of 10xTBE Buffer with 45ml distilled or nuclease-free water.

Agarose gels for gel electrophoresis

To make a 2% agarose gel, use 0.4g of agarose and add it to 20ml 1xTBE buffer. Heat the solution up and mix well until the agarose dissolves and you are left with a viscous solution. Then add 2 µl of GelGreen® stain to the 20 ml and swirl until the dye dissolves. Pour the solution in a prepared gel tray and let cool down. **!Caution** The solution might boil over the top and is extremely hot.

Procedure

Section 1: Primer and index design - Adapting amplification primers (2-Step PCR protocol) •

TIMING 5min

1. Design project-specific primers using primer design tools such as Primer3 (ref. 30; see also Experimental Design for more info on primer design parameters).
2. In order to carry out multiplexing of amplicons from multiple samples on a single flow cell sequencing run, design indexes to add to the amplicon in one of two ways: use option (A) to design custom indexes designed by the user or option (B) to use ONT's PCR Barcoding Expansion kits (1-12, cat. no. EXPPBC001 or 1-96, cat. no. EXP-PBC096).

First add ONT's universal tail sequences to your primer sequences. We recommend a custom indexing approach, as it is cost-effective and double index combinations allow the user to pool more than 96 samples on a single flow cell.

- ~~2. Add ONT's universal tail (capitalized nucleotides) to your primer sequences before ordering the oligos:~~

~~5' TTTCTGTTGCTGCTGATATTGC [project specific forward primer sequence] 3'~~

~~5' ACTTGCCTGTCGCTCTATCTTC [project specific reverse primer sequence] 3'~~

~~**CRITICAL** For the 1 Step PCR protocol replace ONT's universal tail sequence here with the index sequences generated in Step 3A (e.g. 5' ctgtagacaaatcaaggctccag [project specific forward primer sequence] 3').~~

- ~~3. Designing custom indexes or selection of indexes from ONT's PCR Barcoding Expansion kits~~

Commented [HC13]: Please amend step numbering throughout the Procedure as necessary, and ensure that all cross references to numbered steps are changed accordingly throughout the text and figures.

Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 1.27 cm + Indent at: 2.54 cm

Formatted: Indent: Left: 0 cm

(A) Designing custom indexes for multiplexing • TIMING 5min

i) We recommend using barcode_generator (https://github.com/lcomai/barcode_generator) to generate the custom indexes, with the following command:
`python3 barcode_generator_3.4.py none 24 40 12 40 50`

~~In order to maximize the number of indexes while minimizing costs, we recommend dual indexing (different indexes on the 5' and the 3' end of the DNA fragment). We describe how to generate custom indexes below.~~

~~CRITICAL Be sure to take into account 5'–3' strand orientation when appending custom indexes on forward and reverse primer sequences.~~

④ Generate custom indexes using barcode_generator:

`python3 barcode_generator_3.4.py none 24 240 2012 40 50`

This tool allows the user to specify a range of settings for the generation of indexes (also referred to as sequencing barcodes). ~~In order to maximize the number of indexes while minimizing costs, we recommend dual indexing (different indexes on the 5' and the 3' end of the DNA fragment).~~ For MinION-based sequencing, we recommend designing indexes with a length of 24 bp to account for sequencing error rate, although due to improving error rates, e.g. with the new R10.3 pore, users can also design shorter indexes if desired (see ²⁴). Shorter index lengths can be used but may lead to lower demultiplexing rates. Furthermore, we recommend specifying a hamming distance of at least one third to half the index length, e.g. 8–12 for a 24 bp index, to avoid cross-contamination due to inaccurate demultiplexing of reads with high errors in the index sequences. The tool also allows the user to set a range of GC content for the indexes. We recommend generating indexes that have a similar GC content and annealing temperature to the universal tails used for efficient amplification (in the example we used a GC content range from 40 to 50, similar to ONT's indexes). The tool automatically removes indexes that show homopolymers >4bp, which could otherwise cause issues due to the MinION's error profile. This number can be increased or decreased as needed. Barcode generator will automatically attach ONT's universal tails to the indexes (forward: TTTCTGTTGGTGCTGATATTGC and reverse: ACTTGCCTGTCGCTCTATCTTC), which enables the addition of index sequences to DNA barcode amplicons ~~using a second in a 2-step PCR amplification~~. Each forward and reverse index can be used in different combinations to increase the amount of experimental multiplexing. ~~In the example, 20 unique forward and 20 unique reverse indexes are generated, which yields 400 sample combinations.~~

Example forward index (lower case) appended to ONT forward universal tail (upper case):

5'-ctgtagacaaatcaaggcctccagTTTCTGTTGGTGCTGATATTGC-3'

Example reverse index (lower case) appended to ONT reverse universal tail (upper case):

Commented [HC14]: Please specify which Step number the indexes are added in.

Commented [HC15]: Which example are you referring to here?

5'-tgctgtagtagctcgttctacctACTTGCCTGTCGCTCTATCTTC-3'

CRITICAL STEP Be sure to take into account 5' - 3' strand orientation when appending custom indexes on forward and reverse primer sequences.

(B) Selection of indexes from ONT's PCR Barcoding Expansion kits • TIMING 5min

i) Select the indexes supplied in from the ONT's PCR Barcoding Expansion kits fit the universal tails added in Step 23. The selected indexes are added to the amplicon in Step 8.

2

3. Add ONT's universal tail (capitalized nucleotides) to your primer sequences before ordering the oligos:

5' TTTCTGTTGGTGCTGATATTGC-[project-specific forward primer sequence] 3'

5' ACTTGCCTGTCGCTCTATCTTC-[project-specific reverse primer sequence] 3'

CRITICAL For the 1-Step PCR protocol replace ONT's universal tail sequence here with the index sequences generated in Step 3A (e.g. 5'-ctgtagacaaatcaaggcctccag-[project-specific forward primer sequence]-3').

Section 2: DNA extraction and purification from collected samples • TIMING 30min-1h

4. Collect and cut samples into small pieces (for instance 10-25 mg for most tissues and plant samples) to ensure rapid lysis and high yields.

CAUTION: Sample collection permits should be authorized by the appropriate agencies and samples should be treated ethically in accordance with community guidelines.

5. such as DNA can be extracted using a variety of methods. Use option A for DNeasy Blood & Tissue Kit (option A), option B for alkaline lysis buffer (option B), or option C for Chelex® resin (option C). **CRITICAL STEP** Sterilize tools for manipulating and lysing tissues with a flame or bleach in between processing samples. Residual ethanol can affect downstream applications, so ensure the tools are dry. **CRITICAL STEP FIELD** The choice of the appropriate DNA extraction method should be based on the sample type, number of samples to process, and available equipment and funds. Extraction methods such as the HotSHOT or the Chelex® 100 Resin-based method are inexpensive and require little equipment, but DNA purity may be lower relative to kits that contain additional wash steps.

(A) DNA extraction using DNeasy Blood & Tissue Kit (Procedure for Tissue Sample)

- (i) Add 180 µl Buffer ATL and 20 µl of proteinase K (25 mg/mL) to each sample. Incubate at 55°C for 1 hr.
- (ii) Mix well for 15 s. Add 200 µl Buffer AL to the sample, and mix thoroughly. Then add 200 µl ethanol (96–100%), and mix again thoroughly.
- (iii) Pipette the mixture into a DNeasy Mini spin column placed in a 2 ml collection tube and centrifuge for 1 min. Discard flow-through and collection tube. **CRITICAL STEP**

Commented [HC16]: I can't easily understand the combinations of primer sequence, ONT tails and Index sequences you need to design and order, for either the 1-step PCR or the 2-step PCR. Please try to make this clear by rearranging steps as necessary.

So for the one step PCR, you need to order only one set of primers with the custom index sequences appended (e.g. 5'-ctgtagacaaatcaaggcctccag-[project-specific forward primer sequence]-3'). No ONT tail sequences attached?

And for the two step PCR?

A project-specific primer with an ONT universal tail AND a primer for the second step consisting of the ONT tail with an index sequence appended? As per Fig. 2?

Perhaps the Procedure would be clearer if you had Step 1. Design project specific primer sequences Step 2. Design index sequences (using options A or B) Step 3. Order Primers (using options A - for one step PCR - or B - for two-step PCR)

Commented [HC17]: Do you need to add a Step for this? To describe in more detail how it is done? How many samples/what types/should they be used immediately or can they be stored and if so for how long and under which conditions?

Formatted: Body Text, Left, Indent: Left: 0 cm, First line: 0 cm, Space Before: 0 pt, Line spacing: single

Formatted: Default Paragraph Font, Font: 12 pt

Commented [HC18]: How?

Commented [HC19]: How?

Commented [HC20]: What speed, temp.? Please give speed, temp. and duration for all centrifugations.

FIELD We have used microcentrifuges for this protocol that spin at 2,000 x g, which we have found to be sufficient for spin-column-based DNA isolation.

- (iv) Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 µl Buffer AW1, and centrifuge for 1 min. Discard flow-through and collection tube.
- (v) Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 µl Buffer AW2, and centrifuge for 3 min to dry the DNeasy membrane. Discard flow-through and collection tube.
- (vi) Place the DNeasy Mini spin column in a clean 1.5 ml or 2 ml microcentrifuge tube, and pipet 50-100 µl Buffer AE directly onto the DNeasy membrane. Incubate at room temperature (20°C) for 1 min, and then centrifuge for 1 min to elute.

(B) Alkaline lysis buffer-based HotSHOT DNA extraction

CRITICAL STEP FIELD This is a quick and inexpensive extraction method.

- (i) Submerge small amount of tissue (~10-25 mg) or small specimen in 50 µl of the alkaline lysis HotSHOT solution (see ³³ or <https://www.youtube.com/watch?v=Ej91tNYsdr8> for an overview of the method).
- (ii) Incubate at 95°C for 20-30 minutes (some tissue types may require longer incubation) and then cool to 4°C. **CRITICAL STEP FIELD** Miniaturized PCR machines usually use fans for cooling. Therefore they cannot cool samples down to 4°C. Alternatively, put the solution in a fridge or a cool place after the heating to cool down.
- (iii) Add an equal volume of neutralization reagent (40 mM Tris-HCl); mix well. The prep is ready for downstream PCR.

(C) Chelex® 100 Resin-based DNA extraction

CRITICAL STEP FIELD This is a quick and inexpensive extraction method.

- (i) Add 190 µl of 5% Chelex reagent (dissolved in DNase-free distilled water) and 10 µl of proteinase K (25 mg/mL) to each sample.
- (ii) **Optional:** Incubate at 55°C for 1 hr for samples that are not easy to lysate.
- (iii) Incubate at 95°C for 20 minutes.
- (iv) Remove supernatant (DNA extract) and transfer to a fresh 1.5ml Eppendorf tube. **!CAUTION** Make sure you do not transfer any of the beads, as these can inactivate Taq polymerases during the PCR.

PAUSE POINT Genomic DNA can be stored at 4° or -20°C until amplification and library preparation. If only limited or no cooling capacities are available, PCR products can be stored at RT for a few days without degradation.

PCR amplification and indexing

PCR Amplification of the target region (same for the 1-Step and 2-Step PCR setups) • **TIMING**

2.5 – 4h

5-6. In a 0.2 ml tube, set up a barcoding PCR reaction as follows:

Reagent	Volume (µl) per sample
---------	------------------------

Commented [HC21]: Can the extracted DNA be stored? If so please add a PAUSE POINT here and explain how long it could be stored for and under which conditions.

Commented [HC22]: Can it be stored, or should it be used straight away? If it can be stored please add a PAUSE POINT here and explain how long it could be stored for and under which conditions.

Commented [HC23]: How, by pipetting?

Commented [HC24]: Does this apply to the prep from Step Biii?

Commented [HC25]: Deleted this as I think it was incorrect – there seem to be different sections to follow for one or two step?

Formatted: Font: 11 pt, Font color: Custom Color(RGB(67,67,67))

Commented [HC26]: Please clarify whether this is for the one-step or two step PCR? It might be best to put in options A and B in this step to explain one step PCR option and 2 step PCR process?
I think you are implying that these steps are carried out in both one step and two step method – in which case you need to clarify that different primers need to be used here depending on whether this is for the one step PCR or for the first step of the two step PCR?

DreamTaq Hot Start master mix	6.25
Forward primer (10 μ M)	2
Reverse primer (10 μ M)	2
Template DNA from Step 4	1.25 (or up to 1 μ g)
Water, nuclease-free	Fill up to 12.5
Total	12.5

6.- Place in thermocycler and run the following program:

PCR Program	Temperature (°C)	Duration	No. of cycles
Initial denaturation	95	3 min	1
Denaturation	95	30 sec	34
Annealing	Variable, based on annealing temperature of the primer	30 sec	34
Optional: Ramp	72	ramp	34
Extension	72	30 sec	34
Final extension	72	5 min	1
Hold	4	∞	

CRITICAL STEP When closing the lid (either strip or single cap) on the PCR tube, be careful not to touch the inside of the lid while handling.

CRITICAL STEP The PCR settings may need to be adjusted depending on your amplicon length and choice of polymerase. For amplification with the DreamTaq™ Hot Start DNA Polymerase, elongation for 30 sec is sufficient for amplicons up to 1 kb. For longer amplicons, increase the elongation time by 1min/1kb. Other polymerases might need longer extension times! We recommend the use of the NEB Tm Calculator (<https://tmcalculator.neb.com/>) for calculating the annealing temperature of primers.

CRITICAL STEP FIELD In the optional **Ramp step** of the PCR program, the temperature is ramped from the primer annealing temperature to 72°C. This step was shown to improve the PCR yield. However, many field-friendly PCR machines do not have this option, in which case just continue straight from the **Annealing step** to the **Extension step** in the PCR program.

Commented [HC27]: Can you add a reference for this?

CRITICAL STEP FIELD Miniaturized PCR machines usually use fans to cool down the temperature during the PCR. Therefore they cannot cool samples down to 4°C. Make sure you put the PCR products in a fridge or a cool place after the PCR finished.

!CAUTION Leaving the PCR for a long time, such as overnight, on 4°C could damage the PCR machines.

7. Determine the concentration of the amplified barcodes either using the Agilent Bioanalyser, the TapeStation or the Qubit with the broad-range assay per the manufacturer's instructions. **CRITICAL STEP FIELD** Alternatively, DNA concentrations can be roughly estimated based on intensity of gel electrophoresis bands in comparison to a DNA ladder.

TROUBLESHOOTING

PAUSE POINT PCR products can be stored at 4°C until library preparation or up to a year at -20°C. If only limited or no cooling capacities are available, PCR products can be stored at RT for a few days without degradation.

8. For the 1-step PCR protocol continue straight to **Step 11**. For the 2-step PCR protocol, prepare the following mastermix **CRITICAL STEP** The final concentration of amplified barcodes (from Step 64) per 50 µl indexing reaction should be ~0.5 ng/µl (e.g., add 1 µl of a barcode solution with 25 ng/µl). Fill in the right amount in the table below (x). For long barcodes (>2,000 bp) it is advisable to extend the extension time of the PCR to assure complete extension. Template concentration and number of PCR cycles might also need to be increased.

Reagent	Volume (µl) per sample
PCR Index forward (10µM)	1
PCR Index reverse (10µM)	1
Amplified barcode from Step 64	x
DreamTaq Hot Start master mix	25
Nuclease-free water	fill up to 50
Total	50

9. Place in thermocycler and run the following program:

PCR Program	Temperature (°C)	Duration	No. of cycles
Initial denaturation	95	3 min	1
Denaturation	95	15 sec	12-15
Annealing	55	15 sec	12-15
Extension	72	30-60 sec	12-15

Commented [HC28]: What do you expect to see? What is the minimum (or maximum) concentration appropriate to proceed to library prep and sequencing?

Commented [HC29]: Please use the same terminology throughout – It isn't clear what this primer consists of, what is the sequence, what are the components (e.g. ONT tail and index?), which step does it come from – can you consider including a Table that lists all the primers needed, with their sequence, components, and which step they are used in (whether for 1 step or 2 step PCR) to make this clearer please?

Final extension	72	1 min	1
Hold	4	∞	

CRITICAL STEP Adjust the extension time according to the amplicon length and the polymerase's specific extension time (see **Step 6**).

CRITICAL STEP Temperatures here were determined for the DreamTaq Hot Start master mix. Adjust accordingly for different lengths of amplicons and the type of polymerase used.

10. Check the successful amplification by repeating **Step 7**.

CRITICAL Make sure that the addition of the indexes was successful by checking the amplicon size on the gel. The products from **Step 9** should be longer by twice the index length (see **Step 3** for the index length) than the products from **Step 6**. You might see two bands if the index addition was not complete.

TROUBLESHOOTING

PAUSE POINT PCR products can be stored at 4°C or -20°C until library preparation. If only limited or no cooling capacities are available, PCR products can be stored at RT for a few days without degradation.

Normalization and pooling of indexed barcode amplicons • TIMING 30-45 min

11. **Optional:** Determine the concentration of the PCR products as described in Step 7 and normalize samples to equimolar concentrations by diluting higher concentrated PCR products with nuclease-free water. If users wish to skip normalization or if no quantification methods are readily available, continue with **Step 12** but pool samples in equal volumes (note that this may result in unequal sequencing coverage depending on amplicon concentrations).
12. Pool aliquots of the indexed amplicons in equimolar concentrations (from **Step 11**) to prepare a pool of 1 µg of total indexed amplicons. Fill up with nuclease-free water to obtain a total of 53.5 µl if needed (otherwise, if the volume of your pool exceeds 53.5 µl proceed directly to Step 13).
CRITICAL STEP Start the protocol for MinION flow cells with ~ 1 µg of DNA or <100-200 fmol. For use with the Flongle flow cell ~500 ng is sufficient. See <https://www.promega.com/resources/tools/biomath/> to convert DNA from µg to pmol. Be careful to properly pool amplicons in equimolar ratios for best downstream sequencing results.

Cleanup and quantification of amplicons • TIMING ~30 min

13. To purify the amplified barcode regions using AMPure XP or homemade size-selection beads, firstly ~~take~~ the beads out of the fridge before use and ~~use~~ allow them to equilibrate to room temperature before use.

14. Mix the beads well so the liquid appears homogeneous and consistent in color. **CRITICAL STEP** For amplicons below 1,000 bp, we recommend using a ratio of 0.6-1.0x of bead volume to PCR product volume. For amplicons >1,000 bp, use a ratio of 0.6x-0.8x.

Commented [HC30]: How?

13. Take a new 1.5 ml tube, add the amplicon pool from **Step 12** first, then the beads [\(in the ratios explained above\)](#), and mix well by pipetting up and down (at least 10 times). **CRITICAL STEP** [For amplicons below 1,000 bp, we recommend using a ratio of 0.6-1.0x of bead volume to PCR product volume. For amplicons >1,000 bp, use a ratio of 0.6x-0.8x.](#)
14. Incubate for 5 min at room temperature on a [rotator mixer](#). **CRITICAL FIELD** Alternatively, carefully mix by inverting the tube by hand a couple of times.
15. Put the tubes in a magnetic rack and wait for solution to clear (3-5 min).
16. Keeping the tubes in the magnetic rack, [discard/remove](#) the cleared solution (supernatant) from the tubes without touching the beads [and discard](#).
17. Wash Step 1: keep the tubes in the magnetic rack and add 200µl 70% ethanol.
18. Let the tubes sit for 1 min; then [remove and discard](#) the ethanol.
19. Wash Step 2: keep the tubes in the magnetic rack and add 200µl 70% ethanol.
20. Let the tubes sit for 1 min; then remove and discard the ethanol.
21. Seal the tubes, remove from magnetic rack and [spin down briefly](#).
22. Return to the magnetic rack and wait for 1 min.
23. Remove any remaining ethanol, being careful not to touch the bead pellet.
24. Leave the tubes open until the beads are dry (usually 30 sec to 2 min). **CRITICAL STEP:** Do not over-dry, as this might decrease the yield.
25. Add 53.5 µl nuclease-free water directly to the bead pellet. Remove from magnetic rack and mix well by pipetting up and down at least 10 times.
26. Incubate for 3 min at room temperature.
27. Centrifuge briefly, then put back in the magnetic rack for 2-3 min or until the solution is clear.
28. [Remove 53.5 µl of the supernatant and transfer to a clean 0.2 ml PCR tube.](#)

Commented [HC31]: At what speed?

Commented [HC32]: How? By pipetting?

Commented [HC33]: How? By pipetting?

Commented [HC34]: Please give speed, duration and temp.

Commented [HC35]: Can the purified PCR product be stored – if so for how long and under which conditions – please add PAUSE POINT if necessary.

Library preparation and sequencing

Formatted: Indent: Left: 0 cm

Library Preparation (using the SQK-LSK110 sequencing kit) • TIMING ~1.5h

Commented [HC36]: We can't have this many levels of subheading in the Procedure so I have run them together below.

Library Preparation (using the SQK-LSK110 sequencing kit): End-prep • TIMING ~35 min

29. In a 0.2 ml PCR tube, mix the following:

	Standard Flow Cell	Flongle Flow Cell
Reagent	Volume (µl) per sample	Volume (µl) per sample
Ultra II End Prep Reaction Buffer	3.5	3.5
Ultra II End Prep Enzyme Mix	3	1.5
Pooled Barcoded Amplicons (from Step 28)	53.5	25
Total	60	30

CRITICAL STEP If prepping a sequencing library for a standard MinION flow cell, use the volumes in the Standard Flow Cell column. If prepping a sequencing library for a Flongle flow cell use the volumes in the Flongle Flow Cell column.

30. Incubate at room temperature ~~(20°C)~~ for 10 min.

CRITICAL STEP During this step, take out the AMX and BLUNT/TA ligase master mix and place on ice.

31. Incubate at 65°C for 10 min in the Thermocycler.

32. Place on ice to cool down for 30 sec.

33. Repeat **Steps 13 to 28** to perform a bead clean-up of the DNA sample. **CRITICAL STEP FIELD** ONT recommends that the end-prepped DNA sample be subjected to a bead clean-up (see **Steps 13 to 28**). This clean-up can be omitted for simplicity and to reduce library preparation time. However, it has been observed that omission of this clean-up can reduce subsequent adapter ligation efficiency, increase the prevalence of chimeric reads, and lead to an increase in pores being unavailable for sequencing. If omitting the clean-up step, proceed directly to **Step 34**.

Library Preparation (using the SQK-LSK110 sequencing kit): Adapter ligation and clean-up •

TIMING ~30min

34. In a 1.5 ml Eppendorf DNA LoBind tube, mix in the following order:

~~**CRITICAL** Perform this step on ice or ice packs if possible.~~

	Standard Flow Cell	Flongle Flow Cell
Reagent	Volume (µl) per sample	Volume (µl) per sample
DNA sample from Step 32	60	30
NEBNext Quick T4 DNA Ligase	10	5
Ligation Buffer (LNB)	25	12.5
Adapter Mix (AMX-F)	5	2.5
Total	100	50

CRITICAL STEP Perform this step on ice or ice packs if possible.

CRITICAL STEP If prepping a sequencing library for a standard MinION flow cell, use the volumes in the Standard Flow Cell column. If prepping a sequencing library for a flongle flow cell use the volumes in the Flongle Flow Cell column.

35. Mix gently then incubate at room temperature for 10-20 minutes

36. ~~To purify the library with a AMPure XP bead cleanup, firstly add~~ 40 µl of resuspended beads and mix by flicking tube.

37. Incubate at room temperature for 10 min.

38. Spin briefly and place in the magnetic rack until solutions clears (2-3 min).

39. Remove the supernatant, avoiding the pellet.

40. Wash Step 1: add 250 µl Short Fragment Buffer (SFB) at room temperature, remove from magnetic rack and resuspend by gently flicking.
41. Spin briefly and place in the magnetic rack until solution clears.
42. Remove and discard the supernatant, avoiding the pellet.
43. Wash Step 2: add 250 µl SFB at room temperature, remove from magnetic rack and resuspend by flicking.
44. Spin briefly and place back in the magnetic rack until solution clears.
45. Remove and discard the supernatant, avoiding the pellet.
46. Spin down and remove and discard all residual supernatant, allow to dry for ~30 seconds.
47. Remove tube from magnetic rack and add **15 µl (standard MinION flow cell)** or **7 µl (Flongle flow cell)** of Elution Buffer (EB) and resuspend beads by flicking.
48. Incubate at room temperature for 10 min.
49. Spin down briefly and pellet the beads on the magnetic rack until the eluate is clear and colorless.
50. Transfer 15 µl (standard MinION flow cell) or 7 µl (Flongle flow cell) of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube. The prepared library is used for loading into the flow cell in Step 1. **PAUSE POINT** Store the library on ice until ready to load.
51. **Optional:** Quantify 1 µl of eluted sample using a Qubit fluorometer
~~**PAUSE POINT** Store the library on ice until ready to load.~~

Commented [HC37]: Can you replace this with: Repeat Steps 13 to 28 to perform a bead clean-up of the DNA sample?

Commented [HC38]: Please add Step number at which this occurs.

Preparing solutions for sequencing. • **TIMING ~10min**

52. Thaw the Sequencing Buffer (SQB), Loading Beads (LB), Flush Tether (FLT) and one tube of Flush Buffer (FLB) at room temperature before placing the tubes on ice as soon as thawing is complete.
53. Mix the Sequencing Buffer (SQB) and Flush Buffer (FLB) tubes by mixing, spin down and return to ice.
54. Spin down the Flush Tether (FLT) tube, mix by pipetting, and return to ice.

Commented [HC39]: How? Pipetting?

Preparing and loading flow cells for sequencing • **TIMING ~10min**

55. This should be done following option (A) if using a standard MinION flow cell, or option (B) if using a Flongle flow cell
CRITICAL STEP ONT recommends loading 5–50 fmol of this the final prepared library onto R9.4.1 flow cells. For R10.3 flow cells, ONT recommends loading 25-75 fmol. **CAUTION:** Loading more than 50 fmol of DNA can have a detrimental effect on throughput. Dilute the library in Elution Buffer if required. **CRITICAL STEP:** If you are using the Flongle for sample prep development, ONT recommends loading 3-20 fmol instead. To calculate the right fmol we recommend to use the Promega online tool: <https://promega.com/resources/tools/biomath>.
CRITICAL STEP We recommend checking ONT's current protocol for updates and changes (<https://community.nanoporetech.com/protocols>). We also suggest checking ONT's YouTube channel for updated introductory videos (<https://www.youtube.com/channel/UC5yMIYjHSgFfZ37LYq-dzig>).

Commented [HC40]: What do you mean by this? How much do you load if you are just using a Flongle flow cell to follow through the Procedure during your experiment?

(A) Preparing and loading the Standard MinION flow cell

- (i) **Priming and loading:** Open the lid of the nanopore sequencing device and slide the flow cell's priming port cover clockwise so that the priming port is visible.
- (ii) Set a **P1000** pipette to 200 μL .
- (iii) Insert the tip into the priming port. After opening the priming port, check for a small bubble under the cover. Draw back a small volume to remove any bubble (20-30 μL).

CRITICAL STEP Care must be taken when drawing back buffer from the flow cell. The array of pores must be covered by buffer at all times. Removing more than 20-30 μL risks damaging the pores in the array.

CRITICAL STEP After opening the priming port, check for a small bubble under the cover. Draw back a small volume to remove any bubble (20-30 μL).

- (iv) Turn the wheel until the dial shows 220-230 μL , or until you can see a small volume of buffer entering the pipette tip.
- CRITICAL STEP** Visually check that there is continuous buffer from the priming port across the sensor array.

TROUBLESHOOTING

- (v) Prepare the flow cell priming mix: add 30 μL of thawed and mixed Flush Tether (FLT) [from Step 54](#) directly to the tube of thawed and mixed Flush Buffer (FLB) [from Step 52](#), and mix by pipetting up and down.
 - (vi) Load 800 μL of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for 5 minutes.
 - (vii) **Library dilution for sequencing:** Thoroughly mix the contents of the Loading Beads (LB) tube [from Step 52](#) by pipetting.
- CRITICAL STEP** The Loading Beads (LB) tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use.
- (viii) Thoroughly mix the contents of SQB (ONT Ligation Kit) and LB (ONT library loading bead kit) tubes [from Steps 52 & 53](#) by pipetting. Then mixmake the following reaction mix:

Reagent	Volume (μL) per sample
Sequencing Buffer SQB	37.5
Loading Beads II (LBII), mixed immediately before use	25.5
Library from Step 50	12
Total	75

Complete the flow cell priming:

Commented [HC41]: Can you be more specific?

Commented [HC42]: What do you do with this? Discard?

Commented [HC43]: What is this? You only refer to LB previously?

- (ix) Gently lift the SpotON sample port cover to make the SpotON sample port accessible.
- (x) Load 200 µl of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles again by reverse pipetting.
CRITICAL STEP Do not load the priming mix into the SpotON sample port.
- (xi) Mix the prepared library gently by pipetting up and down just prior to loading.
- (xii) Add 75 µl of sample to the flow cell via the SpotON sample port in a dropwise fashion. Ensure each drop flows into the port before adding the next.
CRITICAL STEP The library is loaded dropwise without putting the pipette tip firmly into the port. Take care to avoid introducing any air during pipetting.
TROUBLESHOOTING
- (xiii) Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, then close the priming port and the MinION lid.

PAUSE POINT

(B) Preparing and loading the Flongle flow cell

- (i) Peel back the seal tab until the sample port is exposed.
CRITICAL STEP The library is loaded by putting the pipette tip into the port. Take care to avoid introducing any air during pipetting.
- (ii) **Priming and loading:** Prepare the flow cell priming mix in a new tube: add 117 µl of mixed Flush Buffer (FLB) and 3 µl of mixed Flush Tether (FLT), and mix by pipetting up and down.
- (iii) Load 120 µl of the priming mix into the sample port, avoiding the introduction of air bubbles by reverse pipetting. Wait for 5 minutes.
- (iv) **Library dilution for sequencing:** Thoroughly mix the contents of the LB tube by pipetting.
CRITICAL STEP The Loading Beads (LB) tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use.
- (v) Thoroughly mix the contents of SQB (ONT Ligation Kit box) and LB (ONT library loading bead kit box) tubes by pipetting. Then [make the following reaction mix](#):

Reagent	Volume (µl) per sample
SQB	15
LB	10
Library from Step 50	5
Total	30

- (vi) Mix the prepared library gently by pipetting up and down just prior to loading.

Commented [HC44]: Do you do this again here even though priming mix has already been added in step vi above?

Commented [HC45]: Please add details of how long this can be paused for here and under which conditions?

Commented [HC46]: Please add the step numbers that reagents are taken from in this section of steps as I have done above for A.

- (vii) Add 30 µl of sample to the flow cell via the sample port by dialing down the pipetting volume.
- (viii) Gently reseal the sample port with the seal tab, making sure the sample port is sealed properly.
- (ix) Bring the top (wheel icon section) to its original position and close the MinION lid.

TROUBLESHOOTING

Getting ready for sequencing • TIMING 0.5-1h

- 56. To set up the MinKNOW software, firstly plug in your MinION device.
CRITICAL STEP From here on, the steps are the same for the standard MinION Mk1B flow cells and the Flongle flow cell using MinKNOW on a computer.
- 57. Open MinKNOW on the computer.
- 58. Select the sequencing device connected to the computer, then select the ‘Start Sequencing’ option on the Start homepage.
- 59. Enter information such as experiment name, sample ID and flow cell type, and select kit type.
- 60. **Optional:** Turn basecalling OFF if you want to basecall the data later on your laptop or a [mini-server](#), or if you plan to sequence for longer than a day.
- 61. [Select the following](#) ~~run~~ options: 24h or 72h depending on the maximum run time required, output file fasta5, and save to the local drive.

Sequencing • TIMING 1-3h (up to 3d depending on the required amount of data)

- 62. Start sequencing by pressing the START button.
The run time depends on the amount of coverage desired per individual DNA barcode (amplicon).
[Once enough reads have been produced the run can be stopped.](#) **CRITICAL STEP** We recommend aiming for >1,000x coverage per individual amplicon. This should give you a good representation for each DNA barcode even in the presence of slightly uneven pooling of the different amplicons.
- 63. TROUBLESHOOTING** **Optional:** To recover flow cell pores and re-use the flow cell after a run, a nuclease-flush protocol can be performed (see ONT flow cell wash kit).

DNA sequence processing workflow • TIMING 30min (up to 1d for large amounts of data if the basecalling still needs to be performed, depending on the mini-server/computer capacity)

CRITICAL Here, we outline a bioinformatics protocol, ~~which that~~ can be scaled up to analyze a large number of pooled amplicon samples. The bioinformatic processing can be carried out on UNIX platforms, on Windows with an Ubuntu terminal or on Mac platforms using the terminal (command line). In the example steps, we provide names for the files and file directories for clarity; however, these should be modified to match your file system when running. We also provide example parameters, which should be adapted to fit the individual DNA barcode preferences. The example commands and related data can be found [at https://github.com/ksahlin/NGSpeciesID](https://github.com/ksahlin/NGSpeciesID) (see EXAMPLE WORKFLOW section of the readme documentation).

Commented [HC47]: Can you be more specific?

64. *Basecalling and quality check.* In this step the raw current profiles are converted into fastq formatted sequences. We recommend using the live basecalling included in MinKNOW for shorter sequencing runs (1-3 hours) or when sequencing on the MinION Mk1C. Alternatively, for longer runs or if the data are being re-basecalled, use the following command:

```
guppy_basecaller --input_path minKNOW_input/ --save_path basecalled_fastqs/ -c dna_r9.4.1_450bps_fast.cfg --recursive --disable_pings
```

CRITICAL STEP This will carry out basecalling using Guppy for data sequenced on flow cells using the R9.4 pore (standard MinION flow cell or Flongle). For the R10.3 based standard MinION flow cell, use `-c dna_r10.3_450bps_fast.cfg`. Furthermore, in addition to Fast models, Guppy also offers High-accuracy models, which have a much longer runtime, but produce more accurate basecalls. Information on these models and the current model names can be found in the Guppy documentation. All the fastq files will be stored in the folder: *basecalled_fastqs*. We usually filter the reads after the basecalling, but if you want to filter at this point, use: `--min_qscore 7`, which roughly corresponds to a basecall accuracy of 85%. The tool can be parallelized using the command `--num_callers`. The option `--disable_pings` disables both the transmission of telemetry pings and automatic upload of crash reports, which require internet connection.

CRITICAL STEP ~~ONT uses different basecalling model names in different Medaka versions. Check the Medaka manual for the model name that corresponds to your version of Medaka.~~

65. Go to the folder with the fastq files generated by Guppy.

66. Concatenate all the read files into one large file [as follows](#):

```
cat *.fastq > sequencing_reads.fastq
```

67. Check raw read quality/stats with NanoPlot [as follows](#):

```
NanoPlot --fastq_rich sequencing_reads.fastq -o sequencing_run -p sequencing_run
```

This will create a html based report in the folder *sequencing_run* (option `-o`). The prefix of the output files can be specified using the option `-p`.

68. *Demultiplexing of the sequencing data.* Here we provide commands for demultiplexing using Guppy ([option A](#)) or minibar ([option B](#)). Guppy was developed for demultiplexing of ONT's barcoding kits. However, it now also supports the use of custom barcode sequences (https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14dec2018/barcoding-demultiplexing). Minibar also allows for demultiplexing of ONT and custom indexed libraries. It further offers the option to demultiplex based on primer sequence for multiplexed samples (where different amplicons are sequenced per sample). Example files can be found in: Supplementary Data 1 (a file containing 3,000 reads in fastq format) and Supplementary Data 2 (the index file used for demultiplexing with minibar), and on <https://github.com/ksahlin/NGSpeciesID>.

(A) Demultiplexing using Guppy

Commented [HC48]: How is this achieved? Automatically, or do you need to execute a command?

Commented [HC49]: Can you add a web address where this can be found?

i) [Use the following command:](#)

```
guppy_barcode -i sequencing_reads.fastq -s demultiplex_folder --trim_bar-  
codes --disable_pings
```

CRITICAL STEP The demultiplexing stringency can be adjusted with the `--min_score` flag. The default is 60. Increasing this threshold results in fewer demultiplexed reads, but more accurate read assignment. The option `-s` sets the output directory for the demultiplexed read files. The indexes will be automatically trimmed with the flag `--trim_bar-codes`. The demultiplexing can be parallelized using `--worker_threads`.

¶

(B) Demultiplexing using minibar

i) [Use the following command:](#)

```
python minibar.py indexes.txt sequencing_reads.fastq -T -F -e 3 -E 11
```

CRITICAL STEP This will carry out demultiplexing of sequencing reads ([see sequencing_reads.fastq](#) [corresponds to:](#) in Supplementary Data 1) using a set of pre-specified index and primer combinations ([see indexes.txt](#) [corresponds to:](#) in Supplementary Data 2). When using the option `-T`, minibar will trim the index and primer sequences on both ends. The option `-F` will tell minibar to output an individual fastq sequencing file for each index and primer combination. The stringency can be adjusted using the options `-e` and `-E`, which specify the edit distance allowed between indexes (`-e`) and primers (`-E`). Note that the edit distance is higher for primer regions, which is due to possible ambiguities in primer sequences. Lower edit distances will result in fewer reads, but more accurate read assignment.

69. *Read filtering, clustering, consensus generation and polishing.* Use the following command:

```
NGSpeciesID --ont --consensus --q 10 --sample_size 300 --m 800 --s 100 --medaka --pri-  
mer_file primers.txt --fastq barcode0.fastq --outfolder barcode0_consensus
```

CRITICAL STEP NGSspeciesID uses fastq or fasta files as read input. The output folder can be specified using `--outfolder`. We suggest filtering the sequencing reads for Phred quality score and length. This allows the user to remove reads with lower qualities, as well as many chimeric reads that result from polymerase jumping during the amplification or non-target reads. We recommend filtering for Phred score higher than 10 (using option `--q`). This can be lowered if needed (e.g. to increase the number of available reads) or increased if the data shows a high average quality (e.g. >13). To avoid chimeric and non-target reads, adjust the intended target length (`--m`) and maximum deviation from target length (`--s`) values for each amplicon. For example, if the target amplicon has a length of 800bp, we recommend removing reads < 600 bp and > 1,000 bp. As too many reads can add noise to the clustering and polishing, we recommend subsampling the read

Formatted: Font: Not Italic

Formatted: Numbered + Level: 1 + Numbering Style: i, ii, iii, ... + Start at: 1 + Alignment: Left + Aligned at: 3.81 cm + Indent at: 5.08 cm

Formatted: Font: 12 pt, Not Italic

Formatted: Indent: Left: 5.08 cm

data to 300 to 500 reads per sample using the option: `--sample_size` (see ²⁵). Use the `--ont` flag for MinION read data. The tool automatically generates consensus sequences for all read clusters that include more than 10% of all the reads. This value can be changed if the output is expected to include several consensus sequences (e.g. for species pools). This technically allows for the creation of individual consensus sequences for mixed samples. However, we have to caution that this might not work for distinguishing closely related species (i.e., within the same genus) due to the MinION's high error rate of raw sequence reads. The tool automatically carries out consensus sequence polishing using all the reads that make up the cluster from which the individual consensus sequence was created when the flags `--medaka` or `--racon` are used. The `--primer_file` flag defines the fasta file including the primer sequences to be removed from the consensus sequence (optional step). A primer example fasta file can be found in Supplementary Data 3 and on <https://github.com/ksahlin/NGSpeciesID>.

CRITICAL STEP ONT uses different basecalling model names in different Medaka versions. The default value in NGSpeciesID might differ from the current model names of Medaka. Check the Medaka manual for the model name that corresponds to your version of Medaka if you get a model name error, using the command: `medaka tools list models`

CRITICAL STEP We suggest using a bash script for rapid processing of multiple files. For example, create a text file called `consensus.sh`. In this file, save:

```
for file in *.fastq; do
    bn=`basename $file .fastq`
    NGSpeciesID --ont --consensus --q 10 --sample_size 500 --m 800 --s 100 --medaka --
    primer_file Supplementary_File4_primer.txt --fastq $file --outfolder ${bn}
done
```

An example file can be found on <https://github.com/ksahlin/NGSpeciesID>. Then execute the file in your UNIX or Mac terminal using the command:

```
bash consensus.sh
```

70. Compare the consensus sequences against a database (e.g. NCBI nucleotide or BOLD for Eukaryotes) using BLAST either offline (**option A**) or online (**option B**). If conducting BLAST offline, remember to download or curate a reference database locally on a laptop in fasta format.

(A) Offline comparison of the consensus sequences against a reference database using the command line.

- (i)** First, **convert** the database.fasta file ~~has to be converted~~ into a BLAST readable format **using the following command**. **CRITICAL STEP** This only needs to be done for each database file once.

```
makeblastdb -in database.fasta -dbtype nucl -out database
```

This will create a nucleotide database of the fasta file with the name database. You can create targeted databases or download the respective taxon groups from NCBI nucleotide or BOLD.

- (ii) BLAST the consensus sequences against the database.

```
blastn -db database -query barcode0_consensus.fasta -outfmt 6  
-out barcode0_consensus_blast.out
```

In this example, we compare the consensus sequence of barcode0 against the respective database. BLASTn can also be parallelized to save time when more than one thread is available (option *-num_threads*). Furthermore, it provides different output formats, which can be specified with *-outfmt* (6 is commonly used).

(B) Online comparison of the consensus sequences against a reference database

- (i) Open either online BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) or online BOLD (http://boldsystems.org/index.php/IDS_OpenIdEngine).
- (ii) Copy and paste the consensus sequence/s into the respective field and submit the search.

BLAST: BLAST provides several output statistics characterizing the quality of a match, as follows:

Bit score: A measure of the quality of the alignment between the reference and query. Generally, the higher the bit score, the better the quality of the hit.

Query Coverage: A measure for how much of the query (here our barcode consensus sequence) matches the reference. Coverage should be close to 100% for DNA barcodes.

E-value: A measure for the likelihood that a given sequence match is purely by chance (depends on the database size). The smaller the E-value, the better the match.

Percent Identity: A measure for sequence similarity between the query and the reference database match. The percent identity for species assignment can differ between taxa.

CRITICAL STEP Even though the percent identity is the most intuitive measure of quality of a database match, we recommend not relying on it as the sole measure for taxonomic assignment. Furthermore, note that depending on the taxon studied, the available sequence database can be very poorly covered and thus the best BLAST hit in the database used might not be the actual species or closest relative. In addition, some DNA barcodes are more suited than others for species delineation for certain taxa, and so even a 100% match might not be a reliable species assignment (particularly in very conserved markers). The E-value is very helpful as it takes into account the size of the database. We recommend looking at several BLAST hits to check whether other species match with equally good or poor quality and incorporating phylogenetic analyses, before deciding on a final taxonomic assignment.

BOLD: BOLD uses BLAST to identify single base indels and subsequently uses a Hidden Markov Model of the (within BOLD) translated protein sequence, to which it then aligns the

translated COI sequence. It outputs a similarity score (in % match) and a hierarchical placement of the sequence using Neighbor-joining tree reconstruction with closely related species.

CRITICAL STEP BOLD only works for the following DNA barcodes: COI (animals), ITS (fungi), and MATK and RBCL (plants).

CRITICAL STEP Online BOLD does not automatically check for the reverse complement of the submitted sequence. In case No Hits are found, try reverse complementing the sequence (using e.g. http://www.bioinformatics.org/sms/rev_comp.html) and submit again.

Commented [HC50]: Please expand these abbreviations at first use

TIMING

Commented [HC51]: Please amend step numbers following renumbering of the Procedure.

Primer and index design, • **TIMING 5min, Steps 1-3**

DNA extraction and purification from collected samples, • **TIMING 30min-1h, Step 4**

PCR Amplification of the target region, • **TIMING 2.5 – 4h, Steps 5 - 10**

Normalization and pooling of indexed barcode amplicons, • **TIMING 30 min, Step 11-12**

Cleanup and quantification of amplicons, • **TIMING ~30 min, Step 13-28**

Library Preparation (using the SQK-LSK110 sequencing kit), • **TIMING ~1.5-2h, Step 29 – 61**

Library Preparation End-prep, • **TIMING ~35 min, Steps 29 -33**

Adapter ligation and clean-up, • **TIMING ~30 min, Steps 34 - 51**

Preparing solutions for sequencing, • **TIMING ~10 min, Steps 52 - 54**

Preparing and loading flow cells for sequencing, • **TIMING ~10 min, Step 55**

Getting ready for Sequencing, • **TIMING 10min, Steps 56 - 61**

Sequencing, • **TIMING 1-3h** (up to 3d depending on the required amount of data), **Steps 62 - 63**

DNA sequence processing workflow, • **TIMING 30min (up to 1d (for large amounts of data if the basecalling still needs to be performed, depending on the server/computer capacity), Steps 64 - 70**

Troubleshooting

Troubleshooting guidance can be found in Table 1.

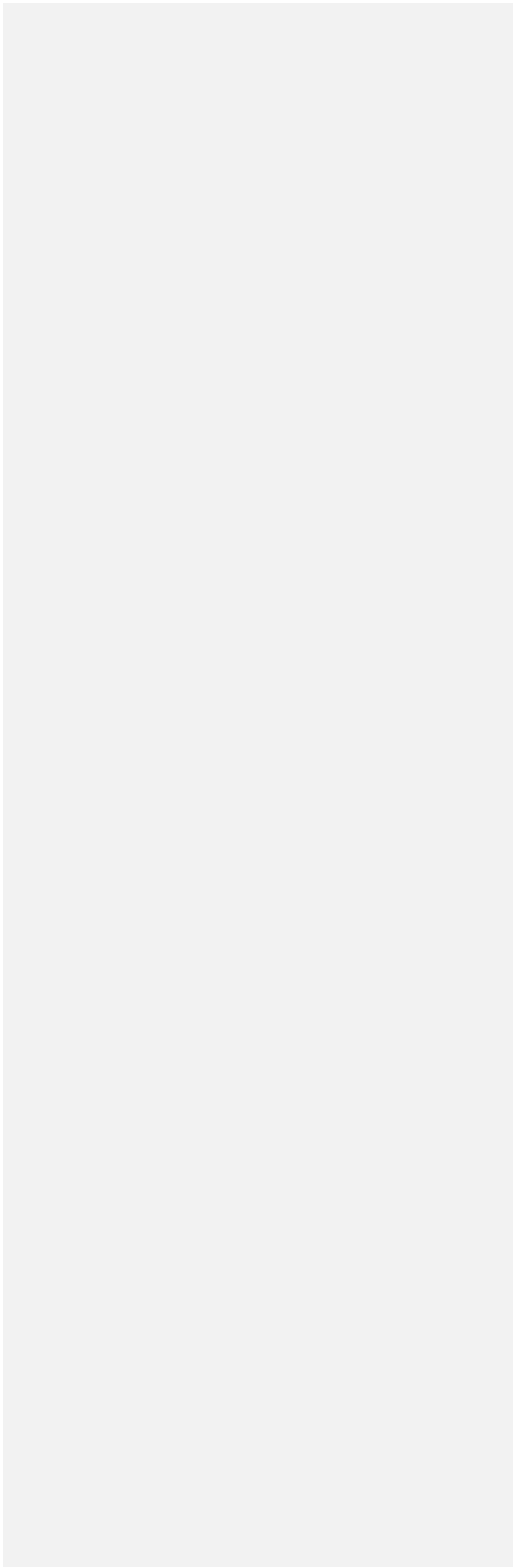
Commented [HC52]: Please amend step numbers following renumbering of the Procedure.



Table 1| Troubleshooting table

Step	Problem	Possible reason	Solution
------	---------	-----------------	----------

5- 7, 8-10	Low or no PCR product yield	Not enough template	Assess quantity of initial DNA extraction. Increase concentration of DNA template or perform DNA extraction of sample again
		Template is degraded	Use electrophoresis to check DNA quality
		Reaction mix components are compromised	Check expiration date of components Aliquot biological components of reaction mixture and avoid multiple freeze-thaw cycles
		Reaction is missing polymerase or other reaction components	Make sure each component was added to PCR reaction
		Over dried beads	Do not dry for more than 1 min
		There is residual ethanol and DNA does not elute well	Make sure to remove any remaining ethanol using pipette P20
		Low amplification efficiency	Try increasing the amount of primers used or use a more efficient polymerase such as the KAPA HiFi HotStart ReadyMix Consider performing an additional bead clean-up step, as described in Step

			13 -28.
--	--	--	---------



54  (iv)	A small bubble may remain even after you have removed more than 20-30µl		Very slowly add the flow cell priming mix and monitor the bubble. If it becomes dislodged, immediately stop and try drawing back enough volume to remove the bubble
54  (xii)	The library is not flowing into the SpotON port after each drop	The SpotON port is blocked with the beads	Carefully draw back the library and try again. More priming buffer may have to be added
61	MinKNOW fails to start script	Problem with MinKNOW installation	Reinstall MinKNOW from scratch and restart the script

Commented [HC53]: What happens if you can't remove it? Can you proceed with the bubble there?

Anticipated Results

Commented [HC54]: Can you add any references to figures or data that illustrate the results of the protocol?

The described protocol can be carried out under a variety of settings, including in the field, in standard laboratories, or in classroom environments. It was developed to require minimal equipment and funding to process samples on-site, which makes the protocol particularly useful for hands-on genomics teaching programs or for conducting genetic assessments of biological samples in areas with limited research funding or infrastructure. All the steps can be performed on battery power, which enables the user to run experiments even if constant electricity is lacking. The bioinformatic processing can be carried out on a laptop with minimal computational experience. The outlined pipeline generates polished consensus sequences for each DNA barcode amplified. It will output individual consensus files (in fasta format), which can be used for downstream analyses such as taxonomic assignments or phylogenetic analyses. Furthermore, this nanopore-based protocol is highly customizable for multiplexing any type of amplicon that the user is interested in, including long-range amplicons, or amplicon mixes generated using multiplex PCR.

Commented [HC55]: Can you provide an example of what the results look like when you have done this? What have you used the protocol to look into and what did you find? E.g. explain the experiments done and cite Supp. Data 1 in this section? Explain what you found out from the obtained dataset and what you were trying to investigate in this case? Once you have the raw read as in Supp. Data 1, can you explain and add examples of what you do with them, summarise what you have found? How did you identify the different species (was this the aim?)?

Data availability

Example data files can be found in the Supplementary information.

Supplementary data.

Supplementary Data 1 contains an example MinION reads dataset generated using the outlined protocol (3,000 reads). The file contains reads of three fish species: the Atlantic cod (*Gadus morhua*), the Haddock (*Melanogrammus aeglefinus*) and the Whiting (*Merlangius merlangus*), sequenced on a Flongle flow cell.

Supplementary Data 2 contains an example index file for demultiplexing with minibar (can be used to demultiplex the example read data in Supplementary Data 2).

Supplementary Data 3 contains an example file of the primer sequences in fasta format (can be used for primer removal using NGSspeciesID).

Commented [HC56]: Is this data new or previously published? If it is from a published study please add a reference. If it is previously unpublished, please state so.

Commented [HC57]: Should this say 1?

Code availability

The code used can be found on the NGSspeciesID github page: <https://github.com/ksahlin/NGSpeciesID>

References

1. Dirzo, R. *et al.* Defaunation in the Anthropocene. *Science* **345**, 401–406 (2014).
2. Barnosky, A. D. *et al.* Approaching a state shift in Earth's biosphere. *Nature* **486**, 52–58 (2012).
3. Diagne, C. *et al.* High and rising economic costs of biological invasions worldwide. *Nature* **592**, 571–576 (2021).
4. Seebens, H. *et al.* Global rise in emerging alien species results from increased accessibility of new source pools. *PNAS* **115**, E2264–E2273 (2018).
5. Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, S96–S99 (2003).
6. Ratnasingham, S. & Hebert, P. bold: The Barcode of Life Data System. *Mol. Ecol. Notes* **7**, 355–364 (2007).
7. Mizrahi, I. *Chapter 1: GenBank: The Nucleotide Sequence Database.*
8. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci Rep* **5**, 1–7 (2015).
9. Martinez, B. *et al.* Technology innovation: advancing capacities for the early detection of and rapid response to invasive species. *Biol Invasions* **22**, 75–100 (2020).
10. Krehenwinkel, H., Pomerantz, A. & Prost, S. Genetic Biomonitoring and Biodiversity Assessment Using Portable Sequencing Technologies: Current Uses and Future Directions. *Genes* **10**, 858 (2019).
11. Menegon, M. *et al.* On site DNA barcoding by nanopore sequencing. *PLoS One* **12**, (2017).
12. Pomerantz, A. *et al.* Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience* **7**, (2018).
13. Blanco, M. B. *et al.* Next-generation technologies applied to age-old challenges in Madagascar. *Conserv Genet* **21**, 785–793 (2020).
14. Chang, J. J. M., Ip, Y. C. A., Ng, C. S. L. & Huang, D. Takeaways from Mobile DNA Barcoding with BentoLab and MinION. *Genes* **11**, 1121 (2020).

15. Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y. & Tighe, S. W. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *J Biomol Tech* **28**, 2–7 (2017).
16. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
17. Faria, N. R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine* **8**, 97 (2016).
18. Watsa, M., Erkenswick, G. A., Pomerantz, A. & Prost, S. Portable sequencing as a teaching tool in conservation and biodiversity research. *PLOS Biology* **18**, e3000667 (2020).
19. Salazar, A. N. *et al.* An educational guide for nanopore sequencing in the classroom. *PLOS Computational Biology* **16**, e1007314 (2020).
20. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17**, 239 (2016).
21. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* **6**, (2017).
22. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol* **14**, (2018).
23. Krehenwinkel, H. *et al.* Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience* **8**, (2019).
24. Srivathsan, A. *et al.* Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology* **17**, 96 (2019).
25. Vasiljevic, N. *et al.* Developmental validation of Oxford Nanopore Technology MinION sequence data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Science International: Genetics* **53**, 102493 (2021).
26. Maestri, S. *et al.* A Rapid and Accurate MinION-Based Workflow for Tracking Species Biodiversity in the Field. *Genes* **10**, 468 (2019).
27. Seah, A., Lim, M. C. W., McAloose, D., Prost, S. & Seimon, T. A. MinION-Based DNA Barcoding of Preserved and Non-Invasively Collected Wildlife Samples. *Genes* **11**, 445 (2020).
28. Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. General concepts for PCR primer design. *Genome Research* **3**, S30–S37 (1993).
29. Singh, V. & Kumar, A. PCR Primer Design. *Molecular Biology Today* **2**, 27–32 (2001).
30. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115 (2012).
31. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
32. Labrador, K., Agmata, A., Palermo, J. D., Follante, J. & Pante, Ma. J. Authentication of processed Philippine sardine products using Hotshot DNA extraction and minibarcode amplification. *Food Control* **98**, 150–155 (2019).
33. Truett, G. e. *et al.* Preparation of PCR-Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques* **29**, 52–54 (2000).
34. Taberlet, P. *et al.* Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**, e14 (2007).
35. Knot, I. E., Zouganelis, G. D., Weedall, G. D., Wich, S. A. & Rae, R. DNA Barcoding of Nematodes Using the MinION. *Front. Ecol. Evol.* **8**, (2020).
36. Wurzbacher, C. *et al.* Introducing ribosomal tandem repeat barcoding for fungi. *Molecular Ecology Resources* **19**, 118–127 (2019).
37. Wilson, B. D., Eisenstein, M. & Soh, H. T. High-Fidelity Nanopore Sequencing of Ultra-Short DNA Targets. *Anal. Chem.* **91**, 6783–6789 (2019).

38. Cornelis, S., Gansemans, Y., Deleze, L., Deforce, D. & Van Nieuwerburgh, F. Forensic SNP Genotyping using Nanopore MinION Sequencing. *Sci Rep* **7**, 41759 (2017).
39. Bohmann, K. *et al.* Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*, early access, (2021).
40. Smyth, R. P. *et al.* Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**, 45–51 (2010).
41. Zizka, V. M. A., Elbrecht, V., Macher, J.-N. & Leese, F. Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Molecular Ecology Resources* **19**, 893–899 (2019).
42. Schnell, I. B., Bohmann, K. & Gilbert, M. T. P. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* **15**, 1289–1303 (2015).
43. Srivathsan, A. *et al.* MinION barcodes: biodiversity discovery and identification by everyone, for everyone. *BioRxiv* (2021) doi:10.1101/2021.03.09.434692.
44. Lange, V. *et al.* Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* **15**, 63 (2014).
45. Srivathsan, A. *et al.* A MinION™-based pipeline for fast and cost-effective DNA barcoding. *Molecular Ecology Resources* **18**, 1035–1049 (2018).
46. Sahlin, K., Lim, M. C. W. & Prost, S. NGSspeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecology and Evolution* **11**, 1392–1398 (2021).
47. Sahlin, K. & Medvedev, P. De Novo Clustering of Long-Read Transcriptome Data Using a Greedy, Quality-Value Based Algorithm. in *Research in Computational Molecular Biology* (ed. Cowen, L. J.) 227–242 (Springer International Publishing, 2019).
48. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
49. Daily, J. Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics* **17**, 81 (2016).
50. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
51. Byagathalli, G., Pomerantz, A., Sinha, S., Standeven, J. & Bhamla, M. S. A 3D-printed hand-powered centrifuge for molecular biology. *PLOS Biology* **17**, e3000251 (2019).

Acknowledgements

We thank Luca Comai for the adjustment of the barcode_generator tool to work for our MinION indexing protocol. We thank Oxford Nanopore Technologies for providing technical support, and for making the offline MinKNOW software available to us. We also thank Hitomi Asahara and the UC Berkeley DNA Sequencing Facility for technical support, and the reviewers for the valuable input. We thank Tropical Herping, Rainforest Expeditions, Field Projects International and the Inkaterra Guides Field Station for support during fieldwork.

Authors Contribution

AP, NV, AS, EH, SK, HK, SW, RO and SP optimized and further developed the laboratory protocols. AP and SP tested and validated the approach and equipment in the field. KS, ML and SP developed and validated the bioinformatics processing pipeline. RO and SP conceived the project. All authors wrote the manuscript and approved the contents of the manuscript.

Competing Financial Interests

Pomerantz became an employee of Oxford Nanopore Technologies Ltd. after the completion of the research described in the manuscript. None of the other authors declare any potential conflicts.

Figure Legends

Figure 1: Examples of equipment used to carry out field-deployed DNA amplicon sequencing. (A) Benchtop centrifuge connected to (B) external power source (RAVPower) to carry out DNA extraction steps. (C) Small thermocycler (miniPCR) to carry out polymerase chain reaction (PCR) amplification or heat block steps. (D) Miniaturized gel electrophoresis system (MiniOne) and mobile phone to visualize PCR amplification products. (E) Small magnetic rack to perform bead cleanups of PCR product and during ONT library preparation steps. (F) The portable MinION sequencer (Oxford Nanopore Technologies). (G) Example setup of portable genomics tools sequencing DNA amplicons in the Amazon rainforest.

Figure 2: Illustration of the steps involved in the laboratory part of the protocol. After isolating DNA from the samples, genetic regions of interest are amplified via polymerase chain reaction (PCR). Indexes can be added to the amplicons using a 1-Step or 2-Step PCR protocol in order to pool and sequence high numbers of amplicons from numerous samples. After pooling, library preparation is carried out and can be sequenced using the portable MinION Mk1B (connected to a laptop or mini-server) or Mk1C (standalone) platforms.

Figure 3: Illustration of the steps involved in the bioinformatic part of the protocol. After the sequencing data have been generated, DNA amplicons are basecalled, quality checked, and demultiplexed. The program NGSspeciesID can then be used to perform downstream processing in one step, including quality and size filtering, subsampling, read clustering, and final consensus generation for each of the DNA amplicons. The final polished consensus sequences can subsequently be compared against reference databases, such as the NCBI or BOLD.

Commented [HC58]: Please see my comments in the Anticipated Results section and try to give an example there of when you have used the protocol and what you found – can you explain what you were looking for in the amazon rainforest and what you found using this protocol? You use a different example in Supp. Data 1, I think?

Commented [HC59]: Please amend step numbers on Figure following renumbering of the Procedure.

Commented [HC60]: Please amend step numbers on Figure following renumbering of the Procedure.