nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-022-00682-x

Check for updates

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

Aaron Pomerantz[1,2,13], Kristoffer Sahlin[3,13], Nina Vasiljevic[4], Adeline Seah[5], Marisa Lim[5], Emily Humble[6], Susan Kennedy[7,8], Henrik Krehenwinkel[8], Sven Winter[9], Rob Ogden[6] and Stefan Prost[10,11,12]✉

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 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 h, 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 barcodes[5], 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 National Center for Biotechnology Information (NCBI) GenBank database[7]) and the ability to simultaneously analyze high numbers of samples[8] have further increased the utility of DNA sequencing as a valuable tool for biodiversity characterization, wildlife forensics and biosurveillance for invasive alien species.

[1]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. [2]Marine Biological Laboratory, Woods Hole, MA, USA. [3]Department of Mathematics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden. [4]Institute of Forensic Medicine Zurich, University of Zurich, Zurich, Switzerland. [5]Wildlife Conservation Society, Zoological Health Program, Bronx Zoo, Bronx, NY, USA. [6]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, UK. [7]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan. [8]Department of Biogeography, University of Trier, Trier, Germany. [9]Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany. [10]Natural History Museum, Vienna, Central Research Laboratories, Vienna, Austria. [11]LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany. [12]South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa. [13]These authors contributed equally: Aaron Pomerantz, Kristoffer Sahlin. [15]These authors jointly supervised this work: Rob Ogden, Stefan Prost. ✉e-mail: stefanprost.research@protonmail.com

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 major 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 (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 economies[9]. 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 (ONT)), have gained in popularity, as they are relatively inexpensive compared with traditional, bulky molecular laboratory equipment and allow for in situ processing of genetic material[10] (Fig. 1). Portable genomics laboratories have been deployed around the world in a diverse set of ecosystems and settings, including the rainforests of Tanzania[11], Ecuador[12] and Madagascar[13], out at sea[14] and in the Antarctic[15], as well as being used to monitor disease outbreaks such as Ebola in West Africa[16] or Zika in Brazil[17], 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 among 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 (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 sequence[20]. 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 with 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 accuracy[14]. 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 refs. [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 reliable[25].

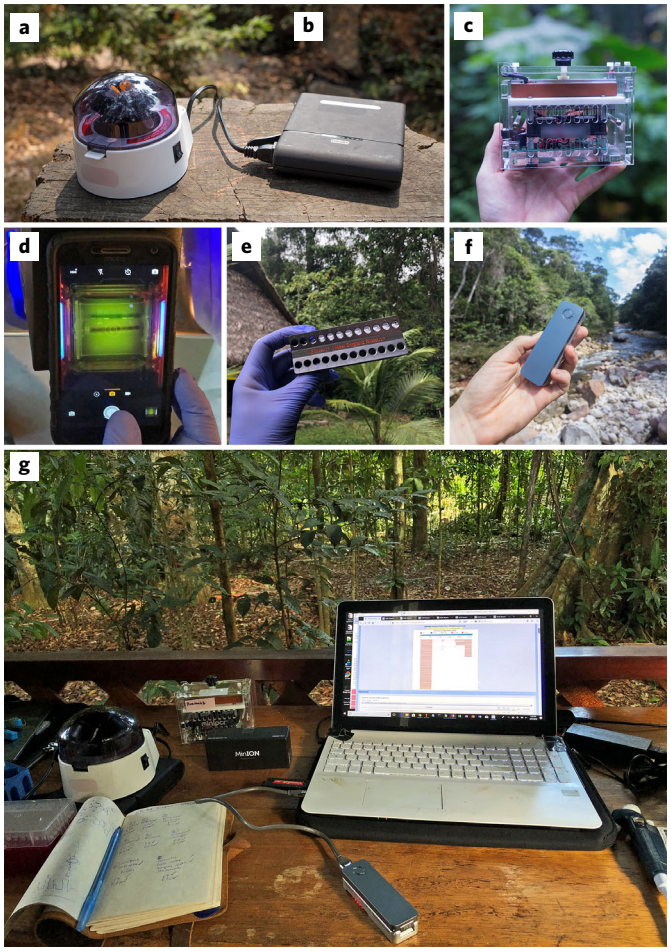

Fig. 1 | Examples of equipment used to carry out field-deployed DNA amplicon sequencing. a,**b**, Benchtop centrifuge (**a**) connected to external power source (RAVPower) (**b**) to carry out DNA extraction steps. **c**, Small thermocycler (miniPCR) to carry out PCR amplification or heat block steps. **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 (ONT). **g**, Example setup of portable genomics tools sequencing DNA amplicons in the Amazon rainforest. This example showcases sequence data being generated in real time of long ribosomal DNA amplicons from arthropods and plants, which enables taxonomic analysis in the field (see ref. [10] for further details).

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 and wildlife forensic efforts, on-site invasive species detection, laboratory-based diagnostics and educational genomics programs, as users can tailor the protocol to work with 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 of 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 with 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.

Alternative methods

The presented protocol is based on insights gained from previously published research papers, such as refs. [12,18,23,24,26,27]. Throughout the protocol, we have attempted to highlight alternative methods that users may wish to assess and compare for their applications. For example, 1-Step PCR amplification strategies have been carried out in studies such as refs. [23,28], while 2-Step PCR amplification strategies were conducted in refs. [12,26,27]. Different strategies to demultiplex custom-designed barcodes have been implemented, for instance with minibar[23,25,29,30] and miniBarcoder[14,24]. Finally, different polishing and reference-free consensus-generating tools have been developed for nanopore-based amplicons, including NGSpeciesID[31], ONTrack[26], Decona (https://github.com/Saskia-Oosterbroek/decona) and ONTbarcoder[28].

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 (Fig. 1f, g). The approach can be used in typical molecular laboratories or nonconventional 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 '**FIELD**' 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 Figs. 2, 3 respectively.

Experimental design

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)[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 2–8 °C for several days with portable refrigeration systems (e.g., the Crēdo Cube, Pelican BioThermal). While DNA primers and most non-enzymatic reagents are fairly stable and can be transported at room temperature (RT, 20 °C) for hours or sometimes even days, many enzymes, such as ligases, require subzero or near-subzero storage. Transport and storage of these can be achieved using portable refrigeration systems and/or transport using ice. Suboptimal transport or storage conditions can lead to a decrease in activity or even loss of functions in the enzymes (see, e.g., ref. [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 materials[15]. 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) or Q5 Hot Start High-Fidelity 2X Master Mix (NEB) for carrying out PCR amplifications. We note that these polymerase master mixes can be relatively costly and that a variety of commercial polymerases and PCR master mixes are available, and recommend the user 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 reagents[11]. However, these still require more testing for use in in situ DNA barcoding applications. ONT offers a field sequencing kit (SQK-LRK001) for cold-chain-free library preparation, but this option is typically used for sequencing genomic DNA as input

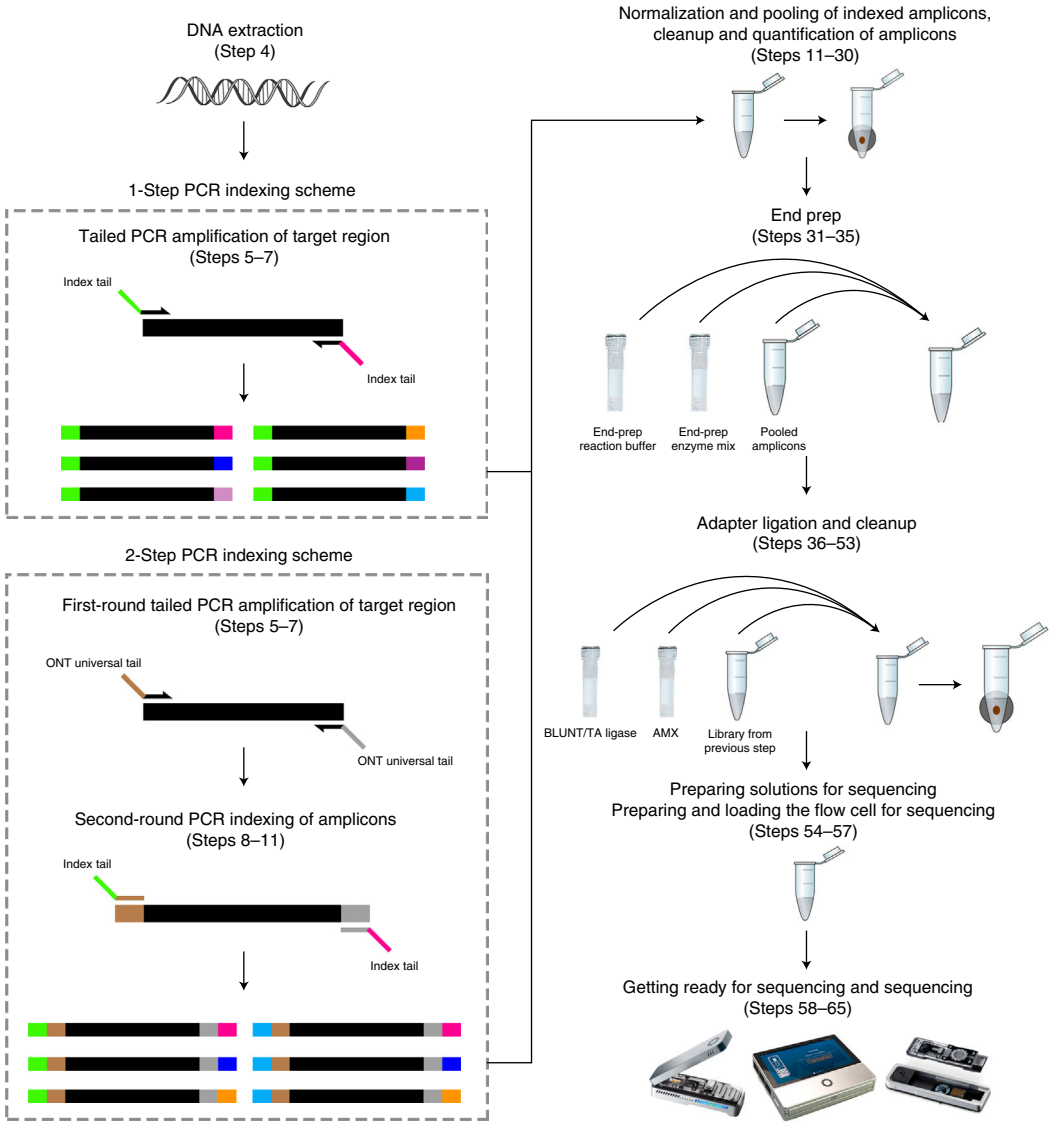

Fig. 2 | Illustration of the steps involved in the laboratory part of the protocol. After isolating DNA from the samples, genetic regions of interest are amplified via PCR. Indexes can be added to the amplicons using a 1-Step or 2-Step PCR protocol to pool and sequence high numbers of amplicons from 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.

material, rather than for DNA barcoding, as the kit relies on transposase activity, which cleaves template molecules.

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 labeled, preserved and deposited at publicly accessible research collections, so that they can serve as verifiable and permanent records. The presented protocol has been used for a variety of samples, from freshly sampled blood and tissue to dried scat, hair, feather and formalin-fixed paraffin-embedded tissue samples, from a variety of taxa including plants, invertebrates (such as insects) and vertebrates (such as reptiles and mammals).

Primer design, multiplexing indexes and PCR strategies (Steps 1–3)

Before starting the laboratory work, the appropriate set of primers must be designed or selected from the scientific literature. These are used to amplify DNA barcodes (genetic regions of interest), which can encompass coding or noncoding regions, and are usually chosen to include a region with high

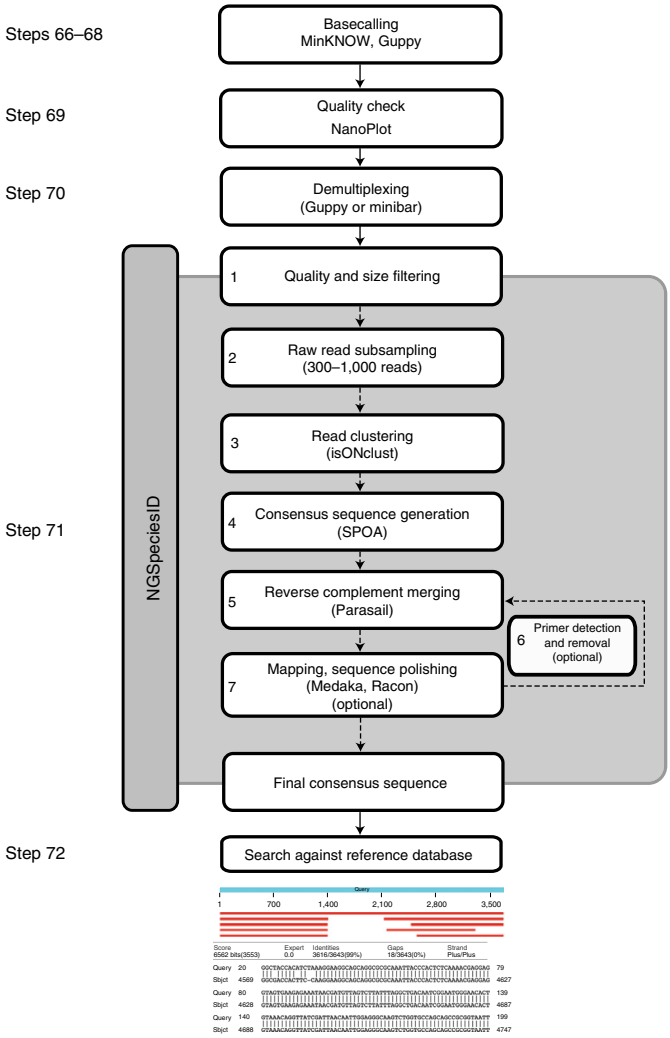

Fig. 3 | Illustration of the steps involved in the bioinformatic part of the protocol. After the sequencing data have been generated, DNA amplicons are basecalled, quality checked and demultiplexed. The program NGSpeciesID can then be used to perform downstream processing in one step, including quality 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.

discriminatory power for the taxon group investigated[5,32]. 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%[33,34]. These parameters should be well matched for both primers in a pair for the best chance of efficiency and specificity[33]. In general, annealing temperature differences between the primers, and in the case of multiplexing (the simultaneous PCR amplification of several amplicons within the same reaction) 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 Primer3 (ref. [35]) 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-BLAST[36]. For more details on primer design, especially in applications that require multiplex PCR or universal primers for a large variety of organisms, please consult refs. [33,34], for example.

To pool amplicons from numerous samples in the same sequencing run (called multiplexing), unique sample-specific indexes must be added to the amplicons. This can be done using either a 1-Step or a 2-Step PCR protocol (Fig. 2), which require slightly different primer setups. In the case of

1-Step PCR, unique indexes are added to the primer sequence directly (which have the following configuration: 5′-index-primer-3′), whereas in the 2-Step PCR protocol universal tails are added first to the DNA barcodes during the amplification of the target region (using amplification primers that have the following configuration: 5′-universal tail-primer-3′) followed by a separate indexing PCR reaction to add the sample-specific indexes (indexing primer configuration: 5′-index-universal tail-3′). The 1-Step protocol (see, e.g., refs. [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 step[37]. Chimeric amplicons primarily form due to incomplete PCR products serving as primers on template DNA[38]. They are particularly a problem in community metabarcoding experiments, where multiple taxa are amplified in a community sample[37,39]. Creation of chimera between many different taxa can highly 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 this protocol, chimeric reads are filtered out during the cluster formation step (Step 69). Alternatively, the 2-Step PCR protocol may be the preferred option for studies that aim to use multiple primer sets (e.g., for multilocus PCR amplification[40] or the application of primer cocktails, which is common for universal cytochrome c oxidase subunit 1 (COI) amplification[5]). As mentioned above, 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 also allows the user to share indexes between projects, taxons 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). However, while adding indexes to the primers allows for multiplexing of many samples, it might reduce the amplification efficiency during the PCR[41]. This can lead to lower demultiplexing rates. With decreasing error rates of the newly developed flow cells and chemistries, index lengths should be reduced to increase the amplification efficiency (see, e.g., ref. [28]). 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'). If researchers plan to also sequence the amplicons on an Illumina HTS platform, Illumina's universal tails (see ref. [42]) 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 run[41]. 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. We also recommend adding one very divergent sample to the processing to monitor for cross-contamination during the laboratory work, or to test the right settings for the demultiplexing of samples during the bioinformatic processing (Step 68).

DNA extraction (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 HotSHOT extraction[43]. 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), which produced sufficient yields and qualities for subsequent amplification of target regions (see reagents list and refs. [12,23,27]). These include standard kits such as the Qiagen QIAmp DNeasy Blood & Tissue Kit or the Bio-Rad Chelex 100 Resin extraction. The latter is more cost-efficient (~0.17 USD per sample[27]) 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 protocol[44] 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 noninvasively sampled DNA sources (e.g., from scat, hair and

feathers), which may contain PCR inhibitors or be degraded from environmental exposure[27]. 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 DNA isolation, we have not experienced reduction in sequencing efficiency or consensus read quality for these extractions[20]. We recommend running DNA extraction blanks (where nuclease-free water is added to the extraction instead of a sample as a negative control) when working with potentially degraded samples.

Amplification, indexing and pooling (Steps 5–30)

After isolating DNA from the samples, DNA barcodes are amplified via PCR. PCR amplification can be efficiently performed using inexpensive mobile thermocyclers. For instance, researchers have reported success using miniPCR (https://www.minipcr.com), MiniOne (https://theminione.com) and BentoLab (https://www.bento.bio) for thermocycling steps under field conditions (see, e.g., refs. [12,14,26,45]). 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 (Fig. 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., refs. [23,46]). 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 PCR[47] or amplicon ligation[48] have been suggested to overcome this lower read length limit.

To multiplex several samples in the same sequencing run, indexes are added (see 'Primer design, multiplexing indexes and PCR strategies' section above and Steps 1–3 in the Procedure). After amplification the amplicons should be pooled in an equimolar ratio to generate a single sequencing library encompassing even coverage for all samples.

Assessment of DNA, either after amplification or during the sequencing library preparation, can be carried out using miniaturized gel-electrophoresis systems, such as blueGel (miniPCR), MiniOne system (MinOne) or BentoLab (Bento Bioworks), or quantification instruments such as Qubit (Thermo Fisher Scientific) or TapeStation (Agilent) if these are available. To reduce time and resources running gels, users can check a small subset of reactions by gel electrophoresis to ensure that there is no widespread amplification failure or presence of amplicons in negative controls.

Sequencing library preparation (Steps 31–53)

The presented sequencing library preparation protocol is based on ONT's library preparation manual for the Ligation Sequencing Kit, experiences gained from working in the field (see refs. [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) step. The library preparation can be carried out in ~1.5 h in remote conditions, and requires pipettes, a small centrifuge, a miniature thermocycler and a small magnetic rack for bead cleanups (Fig. 1). We have included optional steps that require additional equipment in case the protocol is performed in a fully equipped laboratory.

Sequencing (Steps 54–65)

The MinION Mk1B sequencer can be run using external battery power on either a laptop (with sufficient space and speed; see https://nanoporetech.com/community/lab-it-requirements) or standalone devices such as ONT's Mk1C. Alternatively, DIY setups using mini-servers such as 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 the case of high-throughput multiplexing (e.g., hundreds to thousands of samples), up to 72 h. 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 ~10,000 sequenced reads per minute for ~500-bp-long DNA barcodes on the standard

MinION flowcell with pores R9.4 and R10.3. The Flongle has fewer pores and therefore produces fewer reads. Note that the rate of data generated and total output is dependent on the number of active pores, which decrease during the sequencing run time. Previous assessment of raw read coverage using data generated from R9.4 flow cells showed that 300–500× coverage per amplicon produced highly reliable consensus sequences[23,25]. Due to improvements in raw read accuracy, recent studies have demonstrated that as little as 10–50× coverage can be sufficient to produce consensus sequences with 100% accuracy[23,25,28]. This will likely improve further with decreasing error rates of the newer flow cell generations, chemistries and basecalling algorithms (https://nanoporetech.com/accuracy). ONT provides updated videos on how to load the different flow cell types and other useful tutorials on their YouTube channel (https://www.youtube.com/channel/UC5yMlYjHSgFfZ37LYq-dzig).

Bioinformatics (Steps 66–72)

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 1–5 h, or when the sequencing is carried out using ONT's Mk1C device. Alternatively, 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 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 substantially faster than the 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: minibar[23] (https://github.com/calacademy-research/minibar) and Guppy (ONT). These tools were specifically developed for nanopore sequencing data. In addition to demultiplexing of dual or single index libraries, minibar also offers the option to demultiplex samples on the basis of primer sequences, in case different DNA barcodes were amplified in single reactions (multiplex PCR). There are other options available (see, e.g., ref. [49]). 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. If a divergent taxon is added during the laboratory processing, different settings for the allowed numbers of mismatches in the indexes can be tested. If too many mismatches are allowed, reads from this control sample will be present in other demultiplexed samples. This can be tested by blasting the raw read data against a reference or reference database (see Step 70).

Quality assessment and further downstream processing. NanoPlot (https://github.com/wdecoster/NanoPlot) outputs summary statistics and plots showing the quality of the sequenced reads. For downstream processing of raw reads, we use NGSpeciesID, a tool developed for the generation of highly accurate consensus sequences from third-generation long-read sequencing technologies[31]. We note that there are a variety of bioinformatic workflows that have been reported for generating accurate consensus amplicon sequences using MinION[12,24,26–28,46,49]. Pipelines such as ONTrack[26] or Consension[46] 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 ref. [12]). NGSpeciesID has been benchmarked against current pipelines, and all have been found to be comparable for generating highly accurate consensus sequences[31]. NGSpeciesID offers many options (including read filtering on quality and size, subsampling and 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., ref. [28]. Furthermore, it is the only consensus caller for ONT data that has formally been validated for use in non-human forensic applications[25]. 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 below). The final polished consensus sequences can subsequently be compared against different databases, such as the BOLD[6] or the NCBI GenBank

nucleotide database[7]. 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 BLAST[7] and provide recommendations and guidelines on how to interpret the results.

NGSpeciesID (Step 71)

NGSpeciesID 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 on the basis of expected sequence similarity using read-specific error rates with the isONclust algorithm[50]. A draft consensus is formed using spoa[51] 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 notably lower error rates than the individual reads. NGSpeciesID then detects and merges any consensus sequences classified as reverse complements to each other using pairwise alignment with parasail[52]. Two consensus sequences are merged if they have a sequence identity above a user-defined parameter given to NGSpeciesID (default 10%). Sequence identity is calculated as 1 − mismatches/alignment length, where mismatches can be either indels or substitutions. Finally, all draft consensus sequences passing this step are polished using the original reads and Medaka (https://github.com/nanoporetech/medaka) or Racon (https://github.com/rvaser/spoa). To do so, all reads on which the consensus sequence is based are first mapped to the respective consensus sequence using minimap2[53], and then processed either with Medaka or Racon. The polished sequences are the final output of NGSpeciesID. Optionally, users can specify to automatically trim priming sequences from the consensus, in which case NGSpeciesID will also carry out an additional round of reverse complement detection and polishing.

Materials

Reagents

DNA extraction

- QIAmp DNeasy Blood and Tissue Kit (Qiagen, cat. no. 69504) or HotSHOT method (see ref. [44]), or Chelex 100 Resin (Bio-Rad, cat. no. 142-1253), or QuickExtract (Lucigen, cat. no. QE09050), or the 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 types[27]. Alternatively, buffer-based extractions, such as the HotSHOT method[44], are easy and quick to carry out and very cheap alternatives (see ref. [43]). 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 either Q5 Hot Start High-Fidelity 2X Master Mix (NEB, cat. no. M0494S) or assemble your own PCR mix (e.g., using Taq DNA polymerase, buffer, $MgCl_2$, dNTPs) ▲ **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 premixed 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 including ONT's universal tail sequences if performing 2-Step PCR, or with custom indexes attached if using 1-Step PCR (ordered from suppliers such as Integrated DNA Technologies or Sigma Aldrich)

- Custom-made indexing oligos (ordered from suppliers such as Integrated DNA Technologies or Sigma Aldrich) or (optional) PCR Barcoding Expansion 1-96 (ONT, cat. no. EXP-PBC096). To avoid confusion with the scientific term 'DNA barcoding' referring to standardized DNA fragments that are used for species identification, this kit (in the name of which the term barcoding refers to the identification of reads from multiplexing of many samples) 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 indexing 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) ▲ **CRITICAL** **FIELD** If no scale is available, preweighed 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)
- SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, cat. no. S33102) or GelGreen stain (MiniPCR, cat. no. RG-1550-01)

ONT library preparation

- ONT Sequencing Kit (ONT, 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, or (optional) alternatively, users can purchase individual NEBNext products: NEBNext 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., Thermo Fisher, cat. no. AM9937)
- (Optional) D1000 ScreenTape and Reagents (Agilent, cat. no. 5067-5582, 5067-5583)
- (Optional) Flow Cell Wash Kit (ONT, 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) ▲ **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 that spin at 2,000*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 isolations[54] ▲ **CRITICAL** **FIELD** Use external battery packs to be able to run miniature PCR machines and microcentrifuge 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) or (optional) TapeStation 2200 (Agilent) ▲ **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) 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 (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) or BentoLab (Bento Bioworks)
- Cell phone with camera
- Standard kitchen microwave ▲ **CRITICAL** **FIELD** Alternatively, agarose gel solutions can be heated up in a water bath on a camping stove.
- Laboratory bottles (e.g., VWR, cat. no. VWRI215-1593-E)

MinION sequencing

- MinION (ONT, cat. no. MinION Mk1B or MinION MK1C) ▲ **CRITICAL** **FIELD** Make sure to obtain the 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 a MinION Mk1C (a standalone device for running MinION sequencing that can be used in place of MinION Mk1B plus a laptop) or a custom-built mini-server such as Nvidia Jetson Xavier NX (https://developer.nvidia.com/buy-jetson?product=all&location=US)

MinION flow cell options

- Standard MinION flow cell R9.4 (ONT, cat. no. FLO-MIN106D) or R10.x (ONT, cat. no. FLO-MIN111) or a Flongle adapter plus flow cells (ONT, 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 with 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 (ONT) ▲ **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 (ONT); can be used instead of live basecalling with MinKNOW

Quality control

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

Read de-multiplexing

- Guppy (ONT) 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/nuccore) 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

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.

1× TBE buffer

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

Agarose gels for gel electrophoresis

To make a 2% agarose gel, use 0.4 g of agarose and add it to 20 ml 1× TBE 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

Primer and index design: adapting amplification primers for the 1-Step and 2-Step PCR protocols ● Timing 5 min

1 Design project-specific primers using primer design tools such as Primer3 (ref. [35], and see also 'Experimental design' for more information on primer design parameters) or select appropriate primers from the literature.

2 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 or option B to use ONT's indexing kits. We recommend a custom indexing approach, as it is cost-effective and dual index combinations allow the user to pool >96 samples on a single flow cell.

(A) **Designing custom indexes for multiplexing ● Timing 5 min**

(i) We recommend using barcode_generator (https://github.com/lcomai/barcode_generator) to generate custom indexes, with the following command:
python3 barcode_generator_3.4.py none 24 48 12 40 50
This tool allows the user to specify a range of settings for the generation of indexes, as described in the following paragraph and covered in detail in the tool's manual (https://github.com/lcomai/barcode_generator). 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 ref. [28]). 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 of 40–50, similar to ONT's indexes). The tool automatically removes indexes that show homopolymers >4 bp, which could otherwise cause issues due to the MinION's error profile. This number can be increased or decreased as needed. Each forward and reverse index can be used in different combinations to increase the amount of experimental multiplexing. In the example command above, 24 unique forward and 24 unique reverse indexes are generated (specified as 48 in the example command), which yields 576 sample combinations. Barcode generator will automatically attach ONT's universal tails to the indexes (forward TTTCTGTTGGTGCTGATATTGC and reverse ACTTGCCTGTCG CTCTATCTTC) as follows, which is necessary for the 2-Step PCR protocol to allow the addition of index sequences to DNA barcode amplicons in Step 8.

2-Step PCR protocol

Example of forward index primer with the forward index (lower case) appended to ONT forward universal tail (upper case):

5′-ctgtagacaaatcaaggcctccagTTTCTGTTGGTGCTGATATTGC-3′

Example of reverse index primer with the reverse index (lower case) appended to ONT reverse universal tail (upper case):

5′-tgtcgttagtagctcgttctacctACTTGCCTGTCGCTCTATCTTC-3′

1-Step PCR protocol

However, in the 1-Step PCR protocol, the indexes are added directly to the PCR primers; thus, for the 1-Step PCR protocol, replace ONT's universal tail sequence with the DNA barcode primer sequence from Step 1 and order these as index primers (as shown below) for use in Step 5.

Example forward index (lower case) and primer combination:

5′-ctgtagacaaatcaaggcctccag-[project-specific forward primer sequence]-3′

Example reverse index (lower case) and primer combination:

5′-tgtcgttagtagctcgttctacct-[project-specific reverse primer sequence]-3′

For the 1-Step protocol, continue with Step 4, while for the 2-Step protocol, continue with Step 3 to also order the DNA barcoding primers for use in Step 5 (as well as ordering indexing primers designed in this step for use in Step 8).

▲ **CRITICAL STEP** Be sure to take into account 5′–3′ strand orientation when appending custom indexes on forward and reverse primer sequences.

(B) **ONT's indexing kit (PCR Barcoding Expansion Kit)** ● **Timing 5 min**

(i) Alternatively, for the 2-Step PCR protocol, you can select indexes from ONT's indexing kit. These fit to ONT's universal tails, which are added to the project-specific DNA barcode primers as described in Step 3. The selected indexes are added to the amplicon via PCR amplification in Step 8.

3 For the 2-Step PCR, add ONT's universal tail (capitalized nucleotides) to your primer sequences (from Step 1) and order these DNA barcoding oligos for use in Step 5. Example forward primer with ONT's forward universal tail (capital letters) added:

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

Example reverse primer with ONT's reverse universal tail (capital letters) added:

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

▲ **CRITICAL STEP** For the 2-Step protocol, you should have two different primer types: DNA barcode primers with the ONT universal tail attached (e.g., 5′-TTTCTGTTGGTGCTGATATTGC-[project-specific forward primer sequence]-3′ designed in Step 3) and indexing primers (e.g., 5′-ctgtagacaaatcaaggcctccagTTTCTGTTGGTGCTGATATTGC-3′; designed in Step 2A or taken from ONT's indexing kit (Step 2B)) that consist of the index sequence and the ONT universal tail sequence. The DNA barcode primers are used in Step 5 and the indexing primers in Step 8.

DNA extraction and purification from collected samples ● Timing 30 min to 1 h

▲ **CRITICAL** In general, all steps are performed at RT if not specifically stated otherwise.

4 DNA can be extracted using a variety of methods. Use option A for DNeasy Blood & Tissue Kit, option B for alkaline lysis buffer or option C for Chelex Resin. Samples should be cut into small pieces (for instance, 10–25 mg for most tissues and plant samples) to ensure rapid lysis and high yields.

▲ **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.

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

(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 h.

(ii) Mix well for 15 s by inverting the tube. 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 at 2,000*g* for 1 min. Discard flow-through and collection tube.

▲ **CRITICAL STEP** **FIELD** We have used microcentrifuges for this protocol that spin at 2,000*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 pipette 50–100 µl Buffer AE directly onto the DNeasy membrane. Incubate at RT 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 a small amount of tissue (~10–25 mg) or small specimen in 50 µl of the alkaline lysis HotSHOT solution (see ref. [44] or https://www.youtube.com/watch?v=Ej91tNYsdr8 for an overview of the method).

(ii) Incubate at 95 °C for 20–30 min (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 it 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 h for samples that are not easy to lysate.

(iii) Incubate at 95 °C for 20 min.

(iv) Remove supernatant (DNA extract) with a pipette, and transfer to a fresh 1.5 ml 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 (from Step 4A(vi), 4B(iii) and 4C(iv)) can be stored at 4 °C or −20 °C until amplification and library preparation. If only limited or no cooling capacities are available, PCR products are relatively stable at RT.

PCR amplification of the target region (1-Step and 2-Step PCR setups) ● Timing 2.5–4 h

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

Reagent	Volume (µl) per sample
DreamTaq Hot Start master mix	6.25
Forward index primer (10 µM) from Step 2A for 1-Step PCR, or forward DNA barcoding primer (10 µM) from Step 3 for 2-Step PCR	2
Reverse index primer (10 µM) from Step 2A for 1-Step PCR, or reverse DNA barcoding primer (10 µM) from Step 3 for 2-Step PCR	2
Template DNA from Step 4	1.25 (or up to 1 µg)
Water, nuclease-free	Fill up to 12.5
Total	**12.5**

▲ **CRITICAL STEP** Make sure you are using the appropriate primer setup, depending on whether you use the 1-Step or 2-Step PCR setup (see Steps 2–3). The index primers used here should consist of 5′-index-[project-specific forward primer sequence]-3′ for forward and reverse indexing respectively (see Step 2) for the 1-Step PCR protocol, and the DNA barcoding primers used here should consist of 5′-ONT universal tail-[project-specific forward primer sequence]-3′ for forward and reverse primers, respectively (see Step 3) for the 2-Step PCR protocol.

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 s	34
Annealing	Variable, based on annealing temperature of the primer	30 s	
Optional: Ramp	72	Ramp	
Extension	72	30 s	
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 s is sufficient for amplicons up to 1 kb. For longer amplicons, increase the elongation time by 1 min/1 kb. 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 (unpublished observation, N.V., E.H. and R.O.). 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.

▲ **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 has finished.

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

7 (Optional) 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 with a DNA ladder.

■ **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 are relatively stable at RT.

? TROUBLESHOOTING

8 For the 1-Step PCR protocol, continue straight to Step 11. For the 2-Step PCR protocol, prepare the following mastermix:

Reagent	Volume (µl) per sample
Forward index primer (10 µM) from Step 2	1
Reverse index primer (10 µM) from Step 2	1
Amplified barcode from Step 6	*x*
DreamTaq Hot Start master mix	25
Nuclease-free water	Fill up to 50
Total	**50**

▲ **CRITICAL STEP** The final concentration of amplified barcodes (from Step 6) 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 above (*x*). For long barcodes (>2,000 bp) it is advisable to extend the extension time of the PCR to ensure complete extension. Template concentration and number of PCR cycles might also need to be increased. The index primers used here should consist of 5′-index-ONT universal tail-3′ for forward and reverse indexing, respectively (see Step 2).

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 s	12–15
Annealing	55	15 s	
Extension	72	30–60 s	
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.

■ **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.

? TROUBLESHOOTING

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.

Cleanup and quantification of amplicons ● Timing ~30 min

13 To purify the pooled amplicons using AMPure XP or homemade size-selection beads, first take the beads out of the fridge before use and allow them to equilibrate to RT before use.

14 Mix the beads well by vortexing, or gently flicking or inverting the tubes until the liquid appears homogeneous and consistent in color.

15 Take a new 1.5 ml tube, add the amplicon pool from Step 12 first, then the beads from Step 14, and mix well by pipetting up and down (at least ten times).
▲ CRITICAL STEP For amplicons below 1,000 bp, we recommend using a ratio of 0.6–1.0 of bead volume to PCR product volume. For amplicons >1,000 bp, use a ratio of 0.6–0.8.

16 Incubate for 5 min at RT, gently rotating on a rotator mixer.
▲ CRITICAL STEP **FIELD** Alternatively, carefully mix by inverting the tube by hand a couple of times.

17 Put the tubes in a magnetic rack, and wait for solution to clear (3–5 min).

18 Keeping the tubes in the magnetic rack, remove the cleared solution (supernatant) from the tubes using a pipette without touching the beads, and discard.

19 Wash step 1: keep the tubes in the magnetic rack, and add 200 µl 70% ethanol.

20 Let the tubes sit for 1 min, then remove and discard the ethanol using a pipette.

21 Wash step 2: keep the tubes in the magnetic rack, and add 200 µl 70% ethanol.

22 Let the tubes sit for 1 min, then remove and discard the ethanol using a pipette.

23 Seal the tubes, remove from magnetic rack and spin down briefly.

24 Return to the magnetic rack, open the caps and wait for 1 min.

25 Remove any remaining ethanol, being careful not to touch the bead pellet.

26 Leave the tubes open until the beads are dry (usually 30 s to 2 min).
▲ CRITICAL STEP Do not over-dry, as this might decrease the yield.

27 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 ten times.

28 Incubate for 3 min at RT.

29 Centrifuge briefly, then put back in the magnetic rack for 2–3 min or until the solution is clear.

30 Remove 53.5 µl of the supernatant, and transfer to a clean 0.2 ml PCR tube.
■ PAUSE POINT The pooled and cleaned PCR products can be stored at 4 °C or −20 °C until library preparation. If only limited or no cooling capacities are available, they can be stored at RT for a few days without degradation.

Library preparation (using the SQK-LSK110 sequencing kit); end prep ● Timing ~35 min

31 In a 0.2 ml PCR tube, mix the following:

	Volume (µl) per sample	
Reagent	**Standard flow cell**	**Flongle flow cell**
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.

32 Incubate at RT for 10 min.
▲ CRITICAL STEP During this step, take out the AMX and BLUNT/TA ligase master mix and place on ice.

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

34 Let cool down for 30 s.

35 Repeat Steps 13–28 to perform a bead cleanup of the DNA sample.
▲ CRITICAL STEP **FIELD** ONT recommends that the end-prepped DNA sample be subjected to a bead cleanup (see Steps 13–28). This cleanup can be omitted for simplicity and to reduce library preparation time. However, it has been observed that omission of this cleanup 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 cleanup step, proceed directly to Step 34.

Library preparation (using the SQK-LSK110 sequencing kit); adapter ligation and cleanup ● Timing ~30 min

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

	Volume (µl) per sample	
Reagent	**Standard flow cell**	**Flongle flow cell**
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.

37 Mix gently, then incubate at RT for 10–20 min.

38 To purify the library with a bead cleanup, add 40 µl of resuspended beads and mix by flicking tube.

39 Incubate at RT for 10 min.

40 Spin briefly and place in the magnetic rack until solution clears (2–3 min).

41 Remove the supernatant by pipetting and discard, avoiding the pellet.

42 Wash Step 1: add 250 µl Short Fragment Buffer (SFB) at RT, remove from magnetic rack and resuspend by gently flicking.

43 Spin briefly and place in the magnetic rack until solution clears.

44 Remove and discard the supernatant by pipetting, avoiding the pellet.

45 Wash Step 2: add 250 µl SFB at RT, remove from magnetic rack and resuspend by flicking.

46 Spin briefly and place back in the magnetic rack until solution clears.

47 Remove and discard the supernatant by pipetting, avoiding the pellet.

48 Spin down and remove and discard all residual supernatant; allow to dry for ~30 s.

49 Remove tube from magnetic rack, and add 15 µl (standard MinION flow cell) or 7 µl (Flongle flow cell) of Elution Buffer and resuspend beads by flicking.

50 Incubate at RT for 10 min.

51 Spin down briefly and pellet the beads on the magnetic rack until the eluate is clear and colorless.

52 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 in Step 55A(viii) or Step 55B(v).

■ **PAUSE POINT** Store the library on ice until ready to load.

53 (Optional) Quantify 1 µl of eluted sample using a Qubit fluorometer

Preparing solutions for sequencing ● Timing ~10 min

54 Thaw the Sequencing Buffer II (SBII), Loading Beads II (LBII), Flush Tether (FLT) and one tube of Flush Buffer (FLB) at RT before placing the tubes on ice as soon as thawing is complete.

55 Mix the SBII and FLB tubes by pipetting the liquid up and down, spin down and return to ice.

56 Spin down the FLT tube, mix by pipetting and return to ice.

Preparing and loading flow cells for sequencing ● Timing ~10 min

57 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 the final prepared library onto R9.4.1 flow cells. For R10.3 flow cells, ONT recommends loading 25–75 fmol. If you are using the Flongle, ONT recommends loading 3–20 fmol instead.

! CAUTION Loading more than the maximum fmol recommended can have detrimental effects on throughput. Dilute the library in Elution Buffer if required. To calculate the right fmol, we recommend using 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/UC5yMlYjHSgFfZ37LYq-dzig).

(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 bubble (20–30 µl). To do so, turn the wheel until the dial shows 220–230 µl, or until you can see a small volume of buffer entering the pipette tip. Discard the liquid afterwards.

▲ 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 >20–30 µl risks damaging the pores in the array.

? TROUBLESHOOTING

(iv) Prepare the flow cell priming mix: add 30 µl of thawed and mixed FLT from Step 54 directly to the tube of thawed and mixed FLB from Step 52, and mix by pipetting up and down.

(v) Load 800 µl of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for 5 min.

(vi) *Library dilution for sequencing:* thoroughly mix the contents of the LBII tube from Step 52 by pipetting.

▲ CRITICAL STEP The LBII tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use.

(vii) Thoroughly mix the contents of SBII (ONT Ligation Kit) and LBII (ONT library loading bead kit) tubes from Steps 52 and 53 by pipetting. Then make the following reaction mix:

Reagent	Volume (µl) per sample
SBII	37.5
LBII, mixed immediately before use	25.5
DNA Library from Step 50	12
Total	**75**

Complete the flow cell priming:

(viii) Gently lift the SpotON sample port cover to make the SpotON sample port accessible.

(ix) Load an additional 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.

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

(xi) 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

(xii) Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, then close the priming port and the MinION lid.

(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 FLB from Step 53 and 3 µl of mixed FLT from Step 54, 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 min.

(iv) ***Library dilution for sequencing:*** thoroughly mix the contents of the LBII tube by pipetting.
▲ **CRITICAL STEP** The LBII tube from Step 52 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 SBII (ONT Ligation Kit box) and LBII (ONT library loading bead kit box) tubes from Steps 52 and 53 by pipetting. Then make the following reaction mix:

Reagent	Volume (µl) per sample
SBII	15
LBII	10
Library from Step 50	5
Total	**30**

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

(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.

Getting ready for sequencing ● Timing 0.5–1 h

58 To set up the MinKNOW software, first 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.

59 Open MinKNOW on the computer.

60 Select the sequencing device connected to the computer, then select the 'Start Sequencing' option on the Start homepage.

61 Enter information such as experiment name, sample ID and flow cell type, and select kit type.

62 (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.

63 Select the following run options: 24 h or 72 h depending on the maximum run time required, output file fasta5, and save to the local drive.

Sequencing ● Timing 1–3 h (up to 3 d depending on the required amount of data)

64 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.
(Optional) We recommend sequencing until ~1,000× coverage per individual amplicon has been achieved as a conservative estimate, although fewer reads are required to generate accurate consensus sequences. This should provide a good representation for each DNA barcode, even in the presence of slightly uneven pooling of the different amplicons.
? TROUBLESHOOTING

65 (Optional) To recover flow cell pores and reuse the flow cell after a run, a nuclease-flush protocol can be performed (see ONT Flow Cell Wash Kit).

DNA sequence processing workflow ● Timing 30 min (up to 1 d 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 that can be scaled up to analyze a large number of pooled amplicon samples (tens to thousands of 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 (see 'Example workflow' section of the readme documentation).

66 *Basecalling*. 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 h) 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 (using the command `medaka_concensus -h`). 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.

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

68 Concatenate all the read files into one large file as follows:

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

69 *Quality check*. Check raw read quality/stats with NanoPlot as follows:

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

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

70 *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 on the basis of 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**

(i) Use the following command for indexing with ONT's indexes:

```
guppy_barcoder -i sequencing_reads.fastq -s demultiplex_folder
--trim_barcodes --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_barcodes`. The demultiplexing can be parallelized using `--worker_threads`. Refer to ONT's Community Guppy help page (https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revaa_14dec2018) for commands to demultiplex custom indexes.

(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` in Supplementary Data 1) using a set of prespecified index and primer combinations (see `indexes.txt` 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.

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

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

▲ **CRITICAL STEP** NGSpeciesID 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 quality, as well as many chimeric reads that result from polymerase jumping during the amplification or non-target reads. We recommend filtering for Phred score >10 (using option `--q`). This can be lowered if needed (e.g., to increase the number of available reads) or increased if the data show a high average quality (e.g., >13). To avoid chimeric and nontarget 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 800 bp, 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 data to 300–500 reads per sample using the option: `--sample_size` (see refs. [23,25]). Use the `--ont` flag for MinION read data. The tool automatically generates consensus sequences for all read clusters that include >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) owing 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
```

72 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 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). The specific columns can also be specified manually (see ref. [55]).

(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 translated protein sequence (within BOLD), 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 for animals, internal transcribed spacer for fungi, and maturase K and ribulose-bisphosphate carboxylase for plants.

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

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
7, 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
			Aliquot DNA extraction to avoid multiple freeze–thaw cycles
		Reaction mix components are compromised	Check expiration date of components
			Aliquot biological components of reaction mixture, and avoid multiple freeze–thaw cycles
			Check with a positive control
		Reaction is missing polymerase or other reaction components	Make sure each component was added to PCR reaction
		Overdried beads	Do not dry for >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 or DNA template used or use a more efficient polymerase such as the KAPA HiFi HotStart ReadyMix
			Consider performing an additional bead cleanup step, as described in Steps 13–28
57A(iii)	A small bubble may remain even after you have removed >20–30 µl		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
57A(xi)	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
64	MinKNOW fails to start script	Problem with MinKNOW installation	Reinstall MinKNOW from scratch, and restart the script

Timing

Steps 1–3, primer and index design: 5 min
Step 4, DNA extraction and purification from collected samples: 30 min to 1 h
Steps 5–10, PCR amplification of the target region: 2.5–4 h
Steps 11–12, normalization and pooling of indexed barcode amplicons: 30–45 min
Steps 13–30, cleanup and quantification of amplicons: ~30 min
Steps 31–63, library preparation (using the SQK–LSK110 sequencing kit): ~1.5–2 h
Steps 31–35, end prep: ~35 min
Steps 36–53, adapter ligation and cleanup: ~30 min
Steps 54–56, preparing solutions for sequencing: ~10 min
Step 57, preparing and loading flow cells for sequencing: ~10 min
Steps 58–63, getting ready for sequencing: 0.5–1 h

Steps 64–65, sequencing: 1–3 h (up to 3 d, depending on the required amount of data)
Steps 66–72, DNA sequence processing workflow: 30 min (up to 1 d for large amounts of data if the basecalling still needs to be performed, depending on the mini-server/computer capacity)

Anticipated results

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, such as those carried out by ref. [13] and ref. [18] in Madagascar and Peru, respectively, rapid identification of pests at border inspection sites, monitoring for pathogens[56] or conducting genetic assessments of biological samples in areas with limited research funding or infrastructure. For example, studies employing the protocol have found that the consensus amplicons generated have high sequence accuracy, which has enabled phylogenetic analyses of plants and arthropods, and forensic identification of wild mammals in a straightforward, cost-effective, portable approach[10,25]. 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.

When demultiplexing the example file (Supplementary Data 1) using Minibar (Step 68B), you should get 389 reads for sample_c1, 366 reads for sample_h1 and 402 reads for sample_w1. You can increase these numbers by increasing option *-e*. When generating consensus sequences for these read files using NGSpeciesID (Step 69), you should obtain one single consensus sequence per file (in the folder medaka_cl_id_x when using Medaka for polishing, or racon_cl_id_x if you are using Racon, respectively; the x stands for the number of the respective cluster). As the 300 sequences are chosen randomly in Step 69, you might find two consensus sequences per barcode in very rare cases. Usually, these blast to the right species, but with slightly lower similarity (see discussion in ref. [25]). When comparing these with the NCBI Nucleotide database (nr/nt) using online Blast (Step 70B), you will find the consensus sequence of sample_c1 to blast to the Atlantic cod (*Gardus morhua*) with a 99.70% similarity to the first hit for Medaka and 99.11% for Racon, the consensus sequence of sample_h1 to haddock (*Melanogrammus aeglefinus*) with a 99.70% similarity for Medaka and 97.19% for Racon, and sample_w1 to the whiting (*Merlangius merlangus*) with a 99.41% similarity for Medaka and 98.51% for Racon. These values can differ slightly depending on the randomly chosen reads in Step 69.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Example data files can be found in the Supplementary Information.

Code availability

The code used can be found on the NGSpeciesID GitHub page: https://github.com/ksahlin/NGSpeciesID. The code in this protocol has been peer reviewed.

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. *Proc. Natl Acad. Sci. USA* **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. *Proc. R. Soc. Lond. B* **270**, S96–S99 (2003).
6. Ratnasingham, S. & Hebert, P. bold: The Barcode of Life Data System. *Mol. Ecol. Notes* **7**, 355–364 (2007).
7. Mizrachi, I. *Chapter 1: GenBank: The Nucleotide Sequence Database* (NCBI, 2013); https://www.ncbi.nlm.nih.gov/books/NBK470040/
8. Shokralla, S. et al. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* **5**, 9687 (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**, e0184741 (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**, giy033 (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 Med.* **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 Biol.* **18**, e3000667 (2020).
19. Salazar, A. N. et al. An educational guide for nanopore sequencing in the classroom. *PLoS Comput. Biol.* **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. *F1000Research* **6**, 1–32 (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**, e1006583 (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**, giz006 (2019).
24. Srivathsan, A. et al. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biol.* **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 Sci. Int. Genet.* **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. Srivathsan, A. et al. MinION barcodes: biodiversity discovery and identification by everyone, for everyone. Preprint at *BioRxiv* https://doi.org/10.1101/2021.03.09.434692 (2021).
29. Atkins, P. A. P., Gamo, M. E. S. & Voytas, D. F. Analyzing plant gene targeting outcomes and conversion tracts with nanopore sequencing. *Int. J. Mol. Sci.* **22**, 9723 (2021).
30. Simmons, D. R. et al. The Collection of Zoosporic Eufungi at the University of Michigan (CZEUM): introducing a new repository of barcoded *Chytridiomyceta* and *Blastocladiomycota* cultures. *IMA Fungus* **11**, 20 (2020).
31. Sahlin, K., Lim, M. C. W. & Prost, S. NGSpeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecol. Evol.* **11**, 1392–1398 (2021).
32. Taberlet, P. et al. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14 (2007).
33. Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. General concepts for PCR primer design. *Genome Res.* **3**, S30–S37 (1993).
34. Singh, V. & Kumar, A. PCR primer design. *Mol. Biol. Today* **2**, 27–32 (2001).
35. Untergasser, A. et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
36. Ye, J. et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 134 (2012).
37. Bohmann, K. et al. Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Mol. Ecol. Res.* **00**, 1–16 (2021).
38. Smyth, R. P. et al. Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* **469**, 45–51 (2010).

39. Zizka, V. M. A., Elbrecht, V., Macher, J.-N. & Leese, F. Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Mol. Ecol. Resour.* **19**, 893–899 (2019).
40. Arulandhu, A. J. et al. Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *Gigascience* **6**, gix080 (2017).
41. Schnell, I. B., Bohmann, K. & Gilbert, M. T. P. Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* **15**, 1289–1303 (2015).
42. Lange, V. et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* **15**, 63 (2014).
43. 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).
44. Truett, G. E. et al. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). *BioTechniques* **29**, 52–54 (2000).
45. Knot, I. E., Zouganelis, G. D., Weedall, G. D., Wich, S. A. & Rae, R. DNA barcoding of nematodes using the MinION. *Front. Ecol. Evol.* https://doi.org/10.3389/fevo.2020.00100 (2020).
46. Wurzbacher, C. et al. Introducing ribosomal tandem repeat barcoding for fungi. *Mol. Ecol. Resour.* **19**, 118–127 (2019).
47. Wilson, B. D., Eisenstein, M. & Soh, H. T. High-fidelity nanopore sequencing of ultra-short DNA targets. *Anal. Chem.* **91**, 6783–6789 (2019).
48. Cornelis, S., Gansemans, Y., Deleye, L., Deforce, D. & Van Nieuwerburgh, F. Forensic SNP Genotyping using Nanopore MinION Sequencing. *Sci. Rep.* **7**, 41759 (2017).
49. Srivathsan, A. et al. A MinION™-based pipeline for fast and cost-effective DNA barcoding. *Mol. Ecol. Resour.* **18**, 1035–1049 (2018).
50. Sahlin, K. & Medvedev, P. in *Research in Computational Molecular Biology* (ed. Cowen, L. J.) 227–242 (Springer, 2019).
51. 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).
52. Daily, J. Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinform.* **17**, 81 (2016).
53. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
54. Byagathvalli, G., Pomerantz, A., Sinha, S., Standeven, J. & Bhamla, M. S. A 3D-printed hand-powered centrifuge for molecular biology. *PLoS Biol.* **17**, e3000251 (2019).
55. Madden, T. *Appendices. BLAST Command Line Applications User Manual [Internet]* (Bethesda (MD): National Center for Biotechnology Information (US, 2021); https://www.ncbi.nlm.nih.gov/books/NBK279684/
56. Watsa, M., Wildlife Disease Surveillance Focus Group. Rigorous wildlife disease surveillance. *Science* **369**, 145–147 (2020).

Acknowledgements

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

Author contributions

A.P., N.V., A.S., E.H., S.K., H.K., S.W., R.O. and S.P. optimized and further developed the laboratory protocols. A.P. and S.P. tested and validated the approach and equipment in the field. K.S., M.L. and S.P. developed and validated the bioinformatics processing pipeline. R.O. and S.P. conceived the project. All authors wrote the manuscript and approved the contents of the manuscript.

Competing interests

A.P. became an employee of ONT PLC after the completion of the research described in the paper. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-022-00682-x.

Correspondence and requests for materials should be addressed to Stefan Prost.

Peer review information *Nature Protocols* thanks Simon Creer, Sujeevan Ratnasingham and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 February 2020; Accepted: 4 January 2022;
Published online: 11 April 2022

Related links

Key references using this protocol

Pomerantz, A. et al. *Gigascience* **7**, giy033 (2018): https://doi.org/10.1093/gigascience/giy033
Sahlin, K. et al. *Ecol. Evol.* **11**, 1392–1398 (2021): https://doi.org/10.1002/ece3.7146
Vasiljevic, N. et al. *Forensic Sci. Int. Genet.* **53**, 102493 (2021): https://doi.org/10.1016/j.fsigen.2021.102493

nature research

Corresponding author(s): Stefan Prost

Last updated by author(s): 2020/02/20

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We only provide example codes, and did not carry out any analyses for this protocol paper. We added all the relevant links to the software tools in the protocols paper. All software tools are open source.
Data analysis	We only provide example codes, and did not carry out any analyses for this protocol paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The example data is provided in the Supplementary Material

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We describe a method to carry out in-situ DNA barcoding to monitor biodiversity. DNA from samples is extracted, amplified, and sequenced on the MinION sequencing device
Research sample	The protocol can be used for all sample types. We submitted 1,000 reads of fish data with the protocol as an example for the bioinformatic processing. We chose these, because we have a different study on them. They are commercially available food species.
Sampling strategy	NA
Data collection	NA
Timing and spatial scale	NA
Data exclusions	NA
Reproducibility	The protocol is based on experiences gained from many different published research studies.
Randomization	NA
Blinding	NA

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	NA
Wild animals	Scientific authenticated fish samples (Haddock, Whitening, Cod) from the DTU Aqua university. These are not protected species.
Field-collected samples	NA
Ethics oversight	No ethical approval or guidance was needed.

Note that full information on the approval of the study protocol must also be provided in the manuscript.