Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale

Henrik Krehenwinkel[1,4], Aaron Pomerantz[2], James B. Henderson[3,4], Susan R. Kennedy[1], Jun Ying Lim[1,2], Varun Swamy[5], Juan Diego Shoobridge[6], Nipam H. Patel[2,7], Rosemary G. Gillespie[1], Stefan Prost[2,8]

[1] Department of Environmental Science, Policy and Management, University of California, Berkeley, USA
[2] Department of Integrative Biology, University of California, Berkeley, USA
[3] Institute for Biodiversity Science and Sustainability, California Academy of Sciences, San Francisco, USA
[4] Center for Comparative Genomics, California Academy of Sciences, San Francisco, USA
[5] San Diego Zoo Institute for Conservation Research, Escondido, USA
[6] Applied Botany Laboratory, Research and development Laboratories, Cayetano Heredia University, Lima, Perú
[7] Department of Molecular and Cell Biology, University of California, Berkeley, USA
[8] Research Institute of Wildlife Ecology, Department of Integrative Biology and Evolution, University of Veterinary Medicine, Vienna, Austria

Corresponding authors: Henrik Krehenwinkel (krehenwinkel@berkeley.edu) and Stefan Prost (stefan.prost@berkeley.edu)

Keywords

Biodiversity, ribosomal, eukaryotes, long DNA barcodes, Oxford Nanopore Technologies, MinION

Abstract

Background
In light of the current biodiversity crisis, DNA barcoding is developing into an essential tool to quantify state shifts in global ecosystems. Current barcoding protocols often rely on short amplicon sequences, which yield accurate identification of biological entities in a community, but provide limited phylogenetic resolution across broad taxonomic scales. However, the phylogenetic structure of communities is an essential component of biodiversity. Consequently, a barcoding approach is required that unites robust taxonomic assignment power and high phylogenetic utility. A possible solution is offered by sequencing long ribosomal DNA (rDNA) amplicons on the MinION platform (Oxford Nanopore Technologies).

Results
Using a dataset of various animal and plant species, with a focus on arthropods, we assemble a pipeline for long rDNA barcode analysis and introduce a new software (MiniBar) to demultiplex dual indexed nanopore reads. We find excellent phylogenetic and taxonomic resolution offered by long rDNA sequences across broad taxonomic scales. We highlight the simplicity of our approach by field barcoding with a miniaturized, mobile laboratory in a remote rainforest. We also test the utility of long rDNA amplicons for analysis of community diversity through metabarcoding and find that they recover highly skewed diversity estimates.

Conclusions

Sequencing dual indexed, long rDNA amplicons on the MinION platform is a straightforward, cost effective, portable and universal approach for eukaryote DNA barcoding. Long rDNA amplicons scale up DNA barcoding by enabling the accurate recovery of taxonomic and phylogenetic diversity. However, bulk community analyses using long-read approaches may introduce biases and will require further exploration.

Introduction

The world is changing at an unprecedented rate, threatening the integrity of biological communities [1, 2]. To understand the impacts of change, whether a system is close to a regime shift, and how to mitigate the impacts of a given environmental stressor, it is important to consider the biological community as a whole. In recognition of this need, there has been a shift in emphasis from studies that focus on single indicator taxa, to comparative studies across multiple taxa and metrics that consider the properties of entire communities [3]. Such efforts require accurate information on the identity of the different biological entities within a community, as well as the phylogenetic diversity that they represent.

Comparative ecological studies across multiple taxa have been greatly simplified by molecular barcoding [4], where species identifications are based on short PCR amplicon "barcode" sequences. Different barcode marker genes have been established across the tree of life [5, 6], with mitochondrial cytochrome oxidase subunit I (COI) commonly used for animal barcoding [4]. The availability of large sequence reference databases and universal primers, together with its uniparental inheritance and fast evolutionary rate, make COI a useful marker to distinguish even recently diverged taxa. In recent years, DNA barcoding has greatly profited from the emergence of next generation sequencing (NGS) technology. Current NGS platforms enable the parallel generation of barcodes for hundreds of specimens at a fraction of the cost of Sanger sequencing [7]. Furthermore, NGS technology has enabled metabarcoding, the sequencing of bulk community samples, which allows scoring the diversity of entire ecosystems [8].

However, despite their undeniable advantages, barcoding approaches using short, mitochondrial markers have several drawbacks. The phylogenetic resolution offered by short barcodes is very limited, as they contain only a restricted number of informative sites. This problem is exacerbated by the fast evolutionary rate of mitochondrial DNA, which leads to a quick saturation with mutations, increasing the probability of homoplasy between divergent lineages. The accurate estimation of phylogenetic diversity across wide taxonomic scales, however, is an important component of biodiversity research [9]. Moreover, mitochondrial DNA is not always the best marker to reflect species differentiation, as different factors are known to inflate mitochondrial differentiation in relation to the nuclear genomic background. For example, male biased gene flow [10] or infections with reproductive parasites [11] (e.g. *Wolbachia*) can lead to highly divergent mitochondrial lineages in the absence of nuclear differentiation. In contrast, introgressive hybridization can cause the complete replacement of mitochondrial genomes (see e.g. [12, 13]), resulting in shared mitochondrial variation between species.

Considering this background, it would be desirable to complement mitochondrial DNA based barcoding with additional information from the nuclear genome. An ideal nuclear barcoding marker should possess sufficient variation to distinguish young species pairs, but also provide support for phylogenetic hypotheses between divergent lineages. Moreover, the marker should be present across a wide range of taxa and amplification should be possible using universal primers. A marker that fulfils all the above requirements is the nuclear ribosomal DNA (rDNA).

As an essential component of the ribosomal machinery, rDNA is a common feature across the tree of life from microbes to higher eukaryotes [14]. All eukaryotes share homologous transcription units of the 18S, 5,8S and 28S-rDNA genes, which include two internal transcribed spacers (ITS1 and ITS2) [15]. Due to varying evolutionary constraints acting on different parts of the rDNA, it consists of regions of extreme evolutionary conservation, which are interrupted by highly variable sequence stretches [16]. While some rDNA gene regions are entirely conserved across all eukaryotes, the two ITS sequences are distinguished by such rapid evolutionary change that they separate even lineages within species [5, 17]. rDNA markers thus offer taxonomic and phylogenetic resolution at a very broad taxonomic scale. As an essential component of the translation machinery, nuclear rDNA is required in large quantities in each cell. It is thus present in multiple copies across the genome [15] and is readily accessible for PCR amplification. Due to the above advantages, rDNA already is a popular and widely used marker for molecular taxonomy and phylogenetics in many groups of organisms [5, 6, 15, 17, 18].

Spanning about 8 kb, the ribosomal cluster is fairly large, and current barcoding protocols, e.g. using Sanger sequencing or Illumina amplicon sequencing, can only target short sequence stretches of 150 – 1,000 bp. Such short stretches of 28S and 18S are often too conserved to identify young species pairs [19]. The ITS regions, on the other hand, are too variable to design truly universal primers, leading to a considerable amount of taxon dropout during PCR. Moreover, ITS sequences can show considerable length variation between taxa, and are often too long for short amplicon-based barcoding [20]. Consequently, it would be ideal to amplify and sequence a large part of the ribosomal cluster in one fragment. A solution to sequence the resulting long amplicons is offered by recent developments in third generation sequencing platforms, which now enable researchers to generate ultra-long reads, of up to 800 kb [21]. Recently, amplicons of several kilobases of the rDNA cluster were sequenced using Pacific Bioscience (PacBio) technology, to explore fungal community composition [22, 23]. But while PacBio sequencing is well suited for long amplicon sequencing, it is currently not readily available to every laboratory due to the high cost and limited distribution of sequencing machines.

A cost-efficient and readily available alternative is provided by nanopore sequencing technology. The MinION sequencer (Oxford Nanopore Technologies) is small in size, lightweight, allows for sequencing of several Gb's of DNA with average read lengths over 10 kb on a single flow cell [24] and is available starting at $1,000. Despite a raw read error rate of about 12-22 % [21], highly accurate consensus sequences can be called from nanopore data [25, 26]. The MinION is well suited for amplicon sequencing, and a simple dual indexing strategy can be used to demultiplex amplicon samples [27]. This technology offers tremendous potential for long-read barcoding applications, as recently shown in an analysis in fungi [26]. However, current analyses are still exploratory or limited in taxonomic focus and streamlined analysis pipelines to establish the method across the eukaryote tree of life are still missing.

Considering this background, we explore the feasibility of nanopore sequencing of long rDNA amplicons as a simple, cost efficient and universal eukaryote DNA barcoding approach. We compile a workflow from PCR amplification, to library preparation, to demultiplexing and consensus calling (see Fig. 1 for an overview). We explore the error profile of nanopore consensus sequences and introduce MiniBar, a new software to demultiplex dual indexed nanopore amplicon sequences. We test the utility of the ribosomal cluster for molecular taxonomy and phylogenetics across divergent plant and animal taxa. A particular focus of our analysis are arthropods, the most diverse group in the animal kingdom [28], which are highly threatened by current mass extinctions [29]. Using a dataset of spiders, we compare the

taxonomic resolution of the ribosomal cluster with that offered by molecular barcoding using mitochondrial COI, the currently preferred barcode marker for arthropods. Oxford Nanopore Technologies' MinION is a portable sequencer, and Nanopore based DNA barcoding has been applied in remote sites outside of conventional labs (see eg. [25, 30, 31]). Such field-based applications confront researchers with additional complexities and challenges. To highlight the simplicity of our approach, we tested it under field conditions and generated long rDNA barcode sequences using a miniaturized mobile laboratory in a Peruvian rainforest.

We also tested the efficacy of long-read rDNA sequencing for metabarcoding of bulk community samples. A study of bacterial communities [32] suggests Nanopore long-read sequencing as a powerful tool for community characterization, but also found pronounced biases in the recovered taxon abundance. Currently, little is known about the utility of long-read sequencing for animal community analysis. Metabarcoding protocols for community samples need to be carefully optimized, as they can suffer from pronounced taxonomic biases, e.g. due to primer binding or polymerase efficiency [33]. Well established Illumina based short read metabarcoding protocols can account for these biases and allow for a very high qualitative and even quantitative recovery of taxa in communities [34]. However, additional, yet unexplored, biases may affect long-read metabarcoding. We thus also test the utility of long-read rDNA barcoding to recover taxonomic diversity from arthropod mock communities. We compare the qualitative (species richness) and quantitative (species abundance) recovery of taxa by long-read sequencing with that based on short read Illumina amplicon sequencing of the 18SrDNA.

Overall, we demonstrate that long rDNA amplification and sequencing on the MinION platform is a straightforward, cost effective, and universal approach for eukaryote DNA barcoding. It combines robust taxonomic assignment power with high phylogenetic resolution and will enable future analyses of taxonomic and phylogenetic diversity across wide taxonomic scales.

Materials and Methods

DNA extraction, PCR and library preparation
We analyzed 114 specimens of eukaryotes including 17 insect and 42 spider species, two annelid and nine plant species (Supplementary Table 1). Some feeder insects and the annelids were purchased at a pet store. The remaining specimens were collected in oak forest on the University of California Berkeley's campus or in native rainforests of the Hawaiian Archipelago (under the Hawaii DLNR permit: FHM14-349). We particularly focused our arthropod sampling on spiders, which are ubiquitous and essential predators in all terrestrial ecosystems. Recent phylogenomic work [35] provided us with a solid baseline to test the efficiency of rDNA amplicons for phylogenetic and taxonomic purposes. We included a taxonomically diverse collection of 16 spider families from the Araneoidea, the RTA clade and a haplogyne outgroup species. Within spiders, we additionally focused on the genus *Tetragnatha*, which has undergone a striking adaptive radiation on Hawaii.

DNA was extracted from each sample using the Qiagen Archivepure kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocols. The DNA integrity was checked on an agarose gel. Only samples with high DNA integrity were used for the following PCRs. All DNA extracts were quantified using a Qubit fluorometer using the high sensitivity dsDNA assay (Thermo Fisher, Waltham, MA, USA) and diluted to concentrations of 20 ng/µl. We designed a primer pair of each 27 bases to amplify a ~4,000 bp fragment of the ribosomal DNA, including partial 18S and 28S as well as full ITS1, 5.8S and ITS2 sequences (18S_F4 GGCTACCACATCYAARGAAGGCAGCAG and 28S_R8

TCGGCAGGTGAGTYGTTRCACAYTCCT). The primers were designed using alignments of partial 18S and 28S sequences of ~1,000 species of eukaryotes, with a focus on animals. The primers targeted highly conserved regions across all analyzed taxa. Degenerate sites were incorporated to account for variation. We aimed for high annealing temperatures (65-70°C) to impose stringent amplification. These were calculated using the NEB Tm Calculator (https://tmcalculator.neb.com/#!/main).

To index every PCR amplicon separately, we used a dual indexing strategy with each primer carrying a unique 15 bp index sequence at its 5'-tail. Index sequences were designed using Barcode Generator (http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator) with a minimum distance of 10 bases between each index. A total of 15 forward and 16 reverse indexes were designed. Every sample was amplified separately using the Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswitch, MA, USA) in 15 µl reactions, at 68°C annealing temperature, with 35 PCR cycles and using 50 ng of template DNA per PCR. All PCR products were checked and quantified on an agarose gel and then pooled. The final pool was cleaned from residual primers by 0.75 X AMpure Beads XP (Beckman Coulter, Brea, CA, USA). DNA library preparation was carried out according to the 1D PCR barcoding amplicons SQK- LSK108 protocol (Oxford Nanopore Technologies, Oxford, UK). Barcoded DNA products were pooled with 5 µl of DNA CS (a positive control provided by ONT) and an end-repair was performed (NEB-Next Ultra II End-prep reaction buffer and enzyme mix), then purified using AMPure XP beads. Adapter ligation and tethering was carried out with 20 µl Adapter Mix and 50 µl of NEB Blunt/TA ligation Master Mix. The adapter-ligated DNA library was then purified with AMPure beads XP, followed by the addition of Adapter Bead binding buffer, and finally eluted in 15 µl of Elution Buffer. Each R9 flow cell was primed with 1000 µl of a mixture of Fuel Mix and nuclease-free water. Twelve µl of the amplicon library were diluted in 75 µL of running buffer with 35 µL RBF, 25.5 uL LLB, and 2.5 µL nuclease-free water and then added to the flow cell via the SpotON sample port. The "NC_48Hr_sequencing_FLO-MIN107_SQK-LSK108_plus_Basecaller.py" protocol was initiated using the MinION control software, MinKNOW.

Field trial in the Amazon rainforest

A field trial using the protocol described above was conducted in Tambopata, Peru, at the Refugio Amazonas lodge (-12.874797, -69.409669) using two butterflies, a grasshopper, one mosquito, unidentified insect eggs and two plant specimens. Collection permits in Peru were issued by the Servicio Nacional Forestal y de Fauna Silvestre, 403-2016-SERFOR-DGGSPFFS, 019-2017-SERFOR-DGGSPFFS. DNA extractions, PCR and library preparation were performed in the field using a highly miniaturized laboratory consisting of portable equipment. Equipment used for sequencing under remote tropical conditions is described in further detail in Pomerantz, et al. [25]. DNA extractions were carried out with the Quick-DNA Miniprep Plus Kit (Zymo Research, Irvine, CA, USA) according to manufacturer's protocol. PCRs were performed using the Q5 Hot Start High-Fidelity 2X Master Mix and the same primers as described above. A battery operated portable miniPCR device (Amplyus, Cambridge, MA, USA) was used to run PCRs. The sequencing on the MinION was carried out as described above.

Bioinformatics

Raw data processing and consensus calling

The fastq files generated by the ONT software MinKNOW were de-multiplexed using MiniBar (see description below), with index edit distances of 2, 3, and 4 and a primer edit distance of 11. Next, the reads were filtered for quality (>13) and size (>3kb) using Nanofilt [36](

https://github.com/wdecoster/nanofilt). Individual consensus sequences were created using Allele Wrangler (https://github.com/transplantation-immunology/allele-wrangler/) for demultiplexed fastq files with a minimum coverage of 30. Error correction was performed using RACON [37] (https://github.com/isovic/racon). To do so, we first mapped all the reads back to the consensus using minimap (https://github.com/lh3/minimap2). We performed two cycles of running minimap and RACON. Final consensus sequences were compared against the NCBI database using BLASTn to check if the taxonomic assignment was correct.

We performed multiple tests to validate and optimize the consensus accuracy of long-read barcode sequences. To comparatively assess the accuracy, we used consensus sequences of short 18S and 28SrDNA amplicons, which were previously generated using Illumina amplicon sequencing for the 47 analyzed Hawaiian *Tetragnatha* specimens (Kennedy unpublished data). These sequences were aligned with the respective stretches of our nanopore consensus sequences using ClustalW in MEGA [38]. All alignments were then visually inspected and edited manually, where necessary. Pairwise distances between Illumina and nanopore consensus were calculated in MEGA.

To measure consensus accuracy over the whole ribosomal amplicon, we utilized genome skimming data [39] for six Hawaiian *Peperomia* plant species (Lim et al unpublished data). 150 bp paired-end TruSeq gDNA shotgun libraries for the six *Peperomia* samples were sequenced on a single HiSeq v4000 lane (Illumina, San Diego, CA, USA). The resulting paired-end reads were trimmed and filtered using Trimmomatic v0.36 [40] and mapped to their respective nanopore consensus sequences using bowtie2 [41] under default parameter values and allowing for minimum and maximum fragment size of 200 and 700 bases respectively. Mapping coverage of Illumina reads to nanopore consensus sequences ranged between 150 - 600 X with a mean of ~ 300 X across all six samples. We called Illumina read based consensus sequences for each *Peperomia* species using bcftools [42], and aligned them with the previously generated nanopore consensus sequences. Pairwise genetic distances were then calculated in MEGA as described above. We performed two independent distance calculations: 1) excluding indels, i.e. only using nucleotide substitutions to estimate genetic distances, and 2) including indels as additional characters.

Our demultiplexing software allows flexible edit distances to identify forward and reverse indexes from Nanopore reads. Due to the high raw read error rate, too large edit distances could lead to carryover between samples during demultiplexing. This carryover could possibly affect the accuracy of the called consensus sequence. On the other hand, too stringent edit distances may result in very large read dropout. Assuming an average error rate of 12-22 %, 3 bp of our 15 bp indexes should maximize sequence recovery. We thus tested index edit distances of 2, 3, and 4 bp in MiniBar for the six *Peperomia* specimens for which we had generated Illumina based consensus sequences. We counted the number of recovered reads and estimated the accuracy of the resulting consensus sequence based on the according edit distances as described above.

A recent study [25] showed that accurate consensus sequences from nanopore data can be generated using only 30x coverage. We tested 18 different assembly coverages from 10 to 800 sequences for a *Peperomia* species, to explore optimal assembly coverage. We randomly subsampled the quality filtered and demultiplexed fastq file for the according specimen each 10 times for each tested assembly coverage. Consensus sequences were then assembled and genetic distances to the Illumina consensus calculated as described above.

Phylogenetic and taxonomic analysis

We carried out phylogenetic analyses on two hierarchical levels. First, we built a phylogeny for all higher eukaryote taxa in our dataset, which included plants, animals and fungi. Second, we took a closer look into the phylogeny of spiders. The resulting quality checked consensus sequences of all taxa were aligned using ClustalW in MEGA. The alignments were visually inspected and manually edited. Due to the deep divergence in the eukaryote data set, the highly variable ITS sequences could not be aligned and were excluded. For the analyses of spiders, we retained both ITS sequences and aligned the whole rDNA amplicon. Appropriate models of sequence evolution for each gene fragment of the rDNA cluster were identified using PartitionFinder [43]. Phylogenies were built using MrBayes [44], with 4 heated chains, a chain length of 1,100,000, subsampling every 200 generations and a burnin length of 100,000.

Focusing on the endemic Hawaiian *Tetragnatha* species, we also tested the utility of the ribosomal cluster for taxonomic identification, as we also had COI barcodes available for these species. Our dataset contained ribosomal DNA sequences for 47 specimens in 16 species. We calculated pairwise genetic distances between and within all species for the whole ribosomal cluster and for each separate gene region of the rDNA cluster using MEGA. As the 18S and 5.8S did not yield any species level resolution within Hawaiian *Tetragnatha*, they were not analyzed separately. To compare the taxonomic resolution of the ribosomal cluster with that of the commonly used mitochondrial COI, we calculated inter- and intraspecific distances for an alignment of 418 bp of the COI barcode region for the same spider specimens (Kennedy et al. unpublished data). We performed a Mantel test using the R package ade4 [45] to test for a significant correlation between COI and ribosomal DNA based distances. A comparison of intraspecific and interspecific distances for mitochondrial COI and ribosomal DNA also allowed us to test for the presence of a barcode gap.

Nanopore based arthropod metabarcoding

To test for the possibility of estimating arthropod community composition from Nanopore sequencing, we prepared four mock communities of different amounts of DNA extracts from 9 species of arthropods from different orders (see Supplementary Table 2). The samples were amplified using the Q5 High Fidelity Mastermix as described above at 68 °C annealing temperature and 35 PCR cycles. We additionally tested two variations of PCR conditions. We either reduced the annealing temperature to 63 °C or reduced the PCR cycle number to 25.

In order to compare our results with those from an optimized Illumina short read protocol, we amplified all samples for a ~300 bp fragment of the 18SrDNA using the primer pair 18S2F/18S4R [46]. Amplification and library preparation were performed as described in [47] using the Qiagen Multiplex PCR kit. The 18S amplicon pools were sequenced on an Illumina MiSeq using V3 chemistry and 2 x 300 bp reads. Sequence quality filtering, read merging and primer trimming were performed as described in [34].

A library of 18S sequences for all included arthropod species (from [34]) was used as a reference database to identify the recovered sequences using BLASTn [48], with a minimum e-value of 10^{-4} and a minimum overlap of 95 %. Despite the high raw error rate of nanopore reads, taxonomic status of sequences could be assigned using BLAST, as our pools contained members of highly divergent orders. We compared the qualitative (number of species) and quantitative (abundance of species) recovery of taxa from the communities by nanopore long-read and Illumina short read data. To estimate the recovery of taxon abundances, we calculated a fold change between input DNA amount and recovered reads for each taxon and mock community. A fold change of zero corresponded to a 1:1 association of taxon abundance and read count, while positive or negative values indicated higher or lower read counts than the taxon's actual abundance.

MiniBar

We created a de-multiplexing software, called MiniBar. It allows customization of search parameters to account for the high read error rates and has built-in awareness of the dual barcode and primer pairs flanking the sequences. MiniBar takes as input a tab-delimited barcode file and a sequence file in either fasta or fastq format. The barcode file contains, at a minimum, sample name, forward barcode, forward primer, reverse barcode, and reverse primer for each of the samples potentially in the sequence file. The software searches for barcodes and for a primer, each permitting a user defined number of errors, an error being a mismatch or indel. Error count to determine a match can either be a percentage of each of their lengths or can be separately specified for barcode and primer as a maximum edit distance [49]. Output options permit saving each sample in its own file or all samples in a single file, with the sample names in the fasta or fastq headers. The found barcode primer pairs can be trimmed from the sequence or can remain in the sequence distinguished by case or color. MiniBar, written in Python 2.7, can also run in Python 3 and has the single dependency of the Edlib library module for edit distance m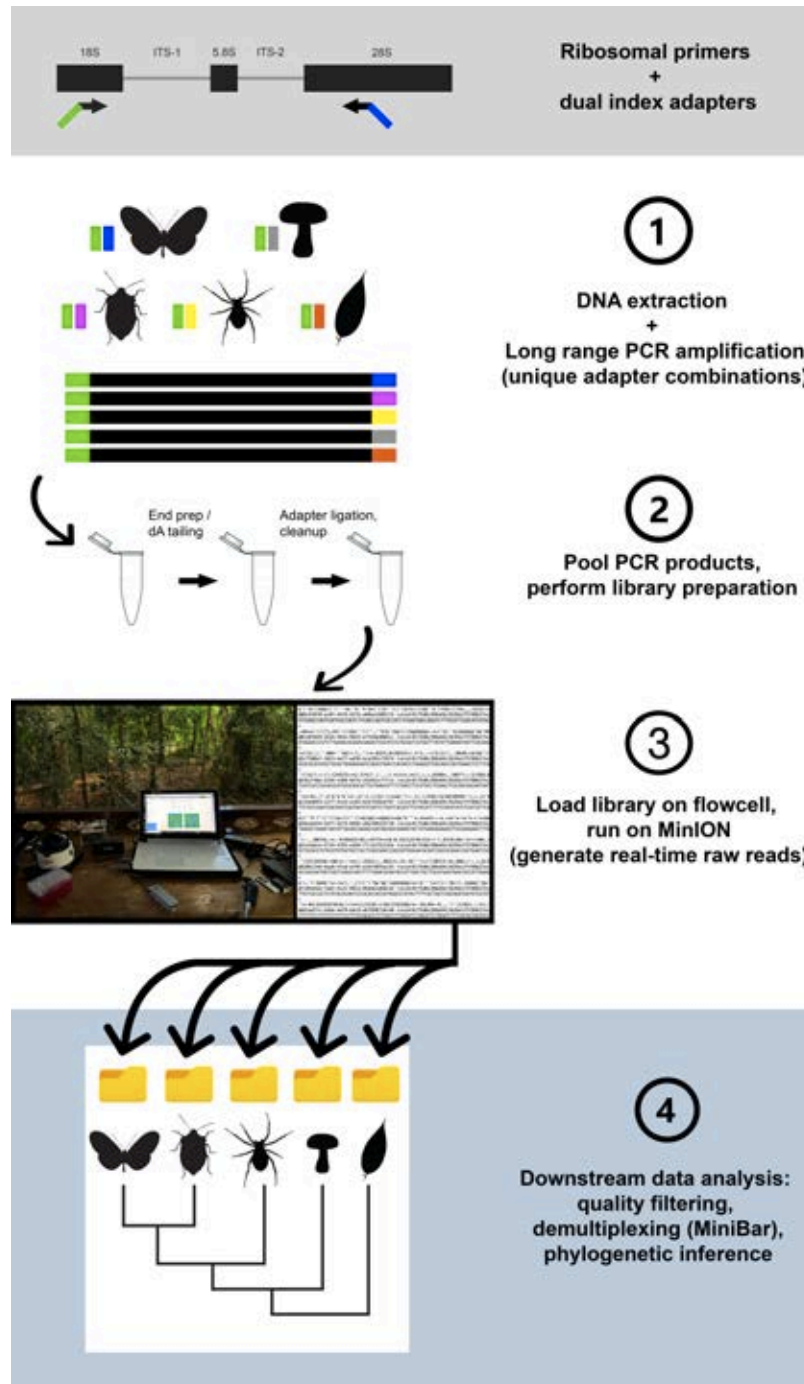easured approximate search [50]. MiniBar can be found at https://github.com/calacademy-research/minibar along with test data.

Figure 1. Workflow for the design, amplification, and sequencing of the ribosomal DNA cluster.

Results

Sequencing, specimen recovery and consensus quality

After quality filtering and trimming, our nanopore run yielded 245,433 reads. We tested edit distances of two, three and four bases in MiniBar to demultiplex samples. Increasing edit distances led to a significant increase in read numbers assigned to index combinations

(Pairwise Wilcoxon Test, FDR-corrected P-value < 0.05). On average, we found 355 reads per specimen for an edit distance of two, 647 for a distance of three and 1,051 for a distance of four. However, at an edit distance of four, we found a considerable increase of wrongly assigned samples. Using Illumina shotgun sequencing-derived consensus sequences of rDNA from six *Peperomia* plants, we tested the accuracy of the nanopore consensus assemblies based on the three edit distances (Fig. 2). While a distance of four yielded the highest number of assigned reads (1,785 on average), it also led to slightly more inaccurate consensus assemblies, with an average distance of 2.072 % to Illumina based consensus sequences. We found a significant increase of consensus accuracy (Pairwise Wilcoxon Test, FDR corrected P < 0.05) for edit distances of two (0.165 % average distance) and three (0.187 % average distance). Despite significant differences in assigned reads (1,091 vs. 637 reads on average), there was not a significant difference in consensus accuracy of edit distances of two versus three bases (Pairwise Wilcoxon Test, FDR corrected P > 0.05).
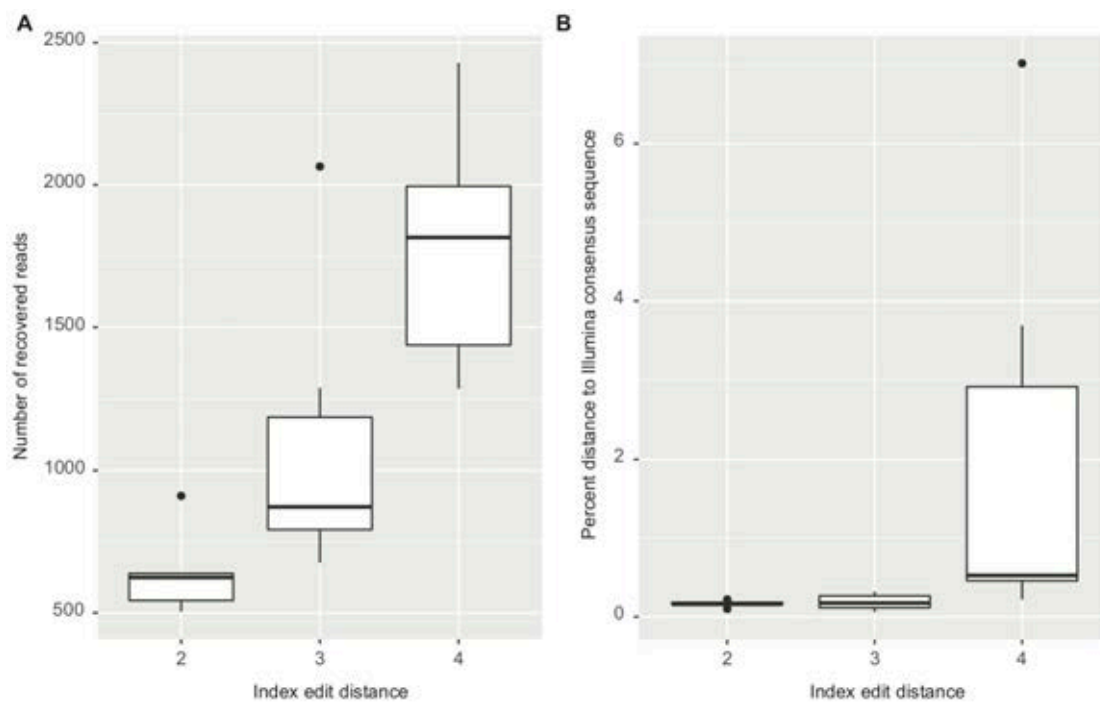
Figure 2: Comparison of recovered sequences and consensus accuracy for different index edit distances in Minibar. A) Number of recovered reads for six *Peperomia* species at index edit distances of two, three and four. B) Pairwise sequence divergence between Illumina and Nanopore based consensus sequences of the same six *Peperomia* specimens at the same index edit distances.

We chose a minimum coverage of 30 (see below) and an edit distance of two (which showed the smallest final consensus error rate) for all subsequent analyses. BLAST analyses suggested a correct taxonomic assignment for the majority of these consensus sequences. However, we found some notable exceptions. For two insect specimens, we amplified mite rDNA sequences. One of these specimens was *Drosophila hydei*, with the mite taxon being a well known phoretic associated with arthropods. A different mite taxon was assembled from an unidentified termite species. A species of isopod and a neuropteran yielded fungal sequences after assembly. The larva of a butterfly and a feeder mealworm (*Zophobas morio*) generated consensus sequences for plants.

A comparison of our consensus sequences for 47 Hawaiian specimens of the spider genus *Tetragnatha* with short Illumina amplicon sequencing-derived 18S and 28S rDNA sequences suggests a very high consensus accuracy. Except for a single specimen, with a single substitution error, all nanopore based consensus sequences were completely identical to the Illumina based consensus. However, the corresponding 18S and 28S fragments did not contain long stretches of homopolymer sequences, where nanopore raw read errors are known to accumulate [51]. Despite containing several homopolymers, the nanopore derived *Peperomia* consensus sequences were highly accurate (Supplementary Fig. 1). Including gaps in the alignment, an average distance of 0.165 % to Illumina based consensus sequences was found. Errors were clustered in Indel regions. After excluding gaps, the average distance dropped to 0.102 %.

We found only a small effect of sequence coverage on consensus assembly accuracy (Supplementary Fig. 2). Even at 10-fold coverage, a low average distance of 0.257% to Illumina consensus sequences was observed. However, at 20-fold coverage, the average distance significantly decreased to 0.128 % (Pairwise Wilcoxon Test, FDR corrected P < 0.05). A slight, but not significant, decrease of distance was observed with increasing coverage, with optimal consensus accuracy at 300-fold coverage (0.031 % distance). At coverages larger than 300, the consensus accuracy slightly decreased (average distance of 0.103 % at 800 X coverage).

The length of the rDNA amplicon was quite variable between taxa. Compared to animals, plant specimens showed a significantly shorter amplicon (Pairwise Wilcoxon Test, FDR corrected P < 0.05). The length difference was found for the actual gene sequences (18S, 5.8S, 28S: 3,063 vs 2,781 bp on average; Supplementary Fig. 3A) as well as including the ITS sequences (3,741 vs. 3,241 bp on average, Supplementary Fig. 3B). Within arthropods, we found significant length differences between arachnid and insect sequences. On average, insects carried a significantly longer rDNA sequence than arachnids (Supplementary Fig. 4; Pairwise Wilcoxon Test, FDR corrected P < 0.05). This holds true for the gene sequences (3,154 vs. 3,047 bp for 18S, 5.8S, 28S on average), as well as the whole amplicon, including ITS sequences (4,192 vs. 3,644 bp on average). While most spiders showed very stable length distributions for the rDNA amplicon length (average length ± standard deviation across all Araneae: 3,629 bp ± 81), several insect orders had rDNA sequences of more variable length (Coleoptera: 4,488 bp ± 352; Lepidoptera: 4363 bp ± 603).

In contrast to the variable length of the rDNA cluster, we found a very stable GC content across the whole taxonomic spectrum (46.75 ± 2.67 % across all taxa). GC content of plants and animals was highly similar (Supplementary Fig 3c) (plants: 46.01 ± 1.66 %; animals: 46.82 ± 2.74 %). Highly similar GC content was also found between insects (46.67 ± 3.73 %) and arachnids (46.93 ± 2.47 %) (Supplementary Fig 4c).
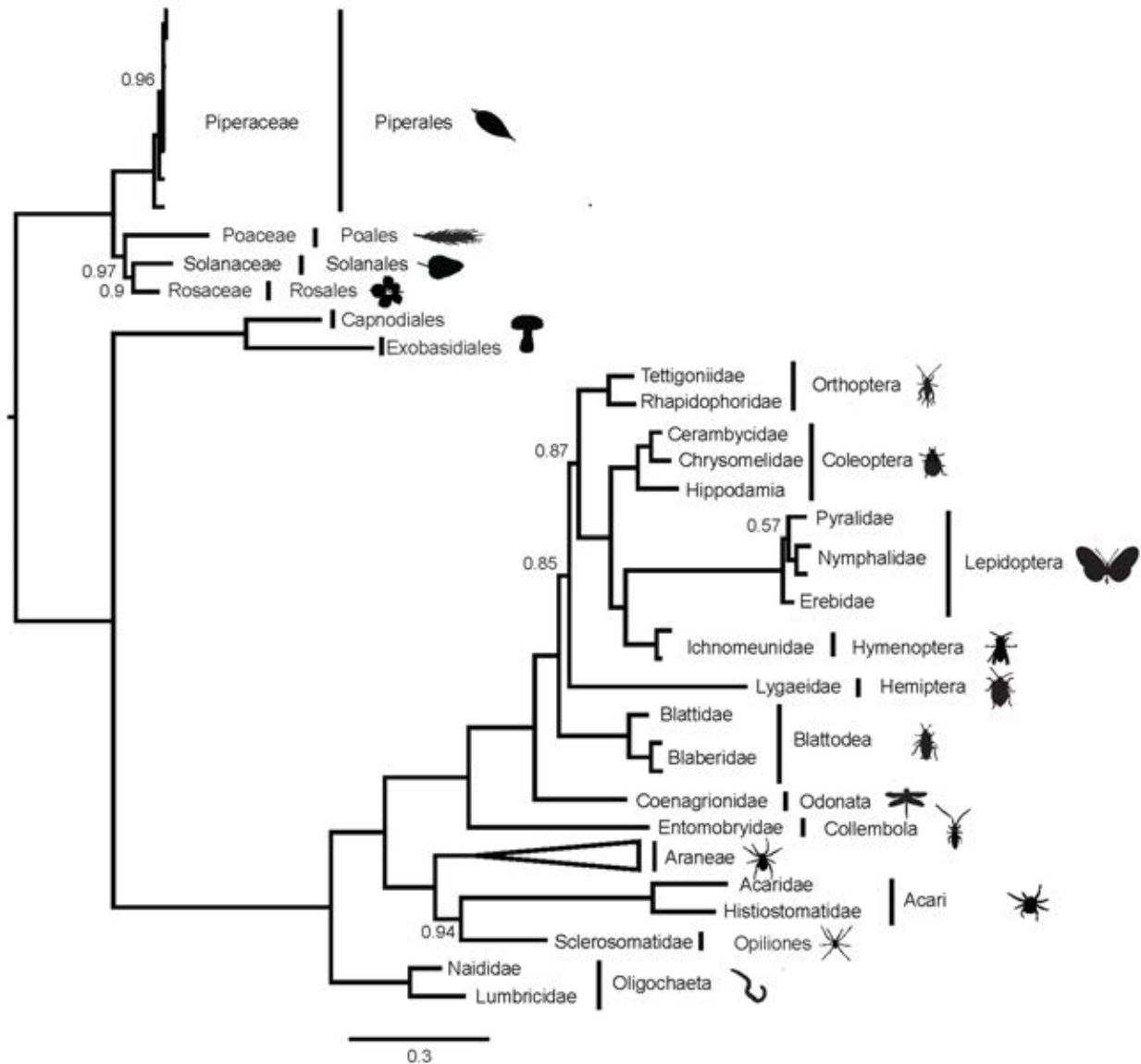
Phylogenetic reconstruction

Figure 3 Bayesian consensus phylogeny based on a 3,656 bp alignment of 18S, 5.8S and 28S sequences of 117 animal, fungal and plant taxa. The phylogeny is rooted using plants as outgroup. Branches are annotated with family and order level taxonomy. The Araneae clade of 83 specimens is collapsed. Only posterior probability values below 1 are displayed.
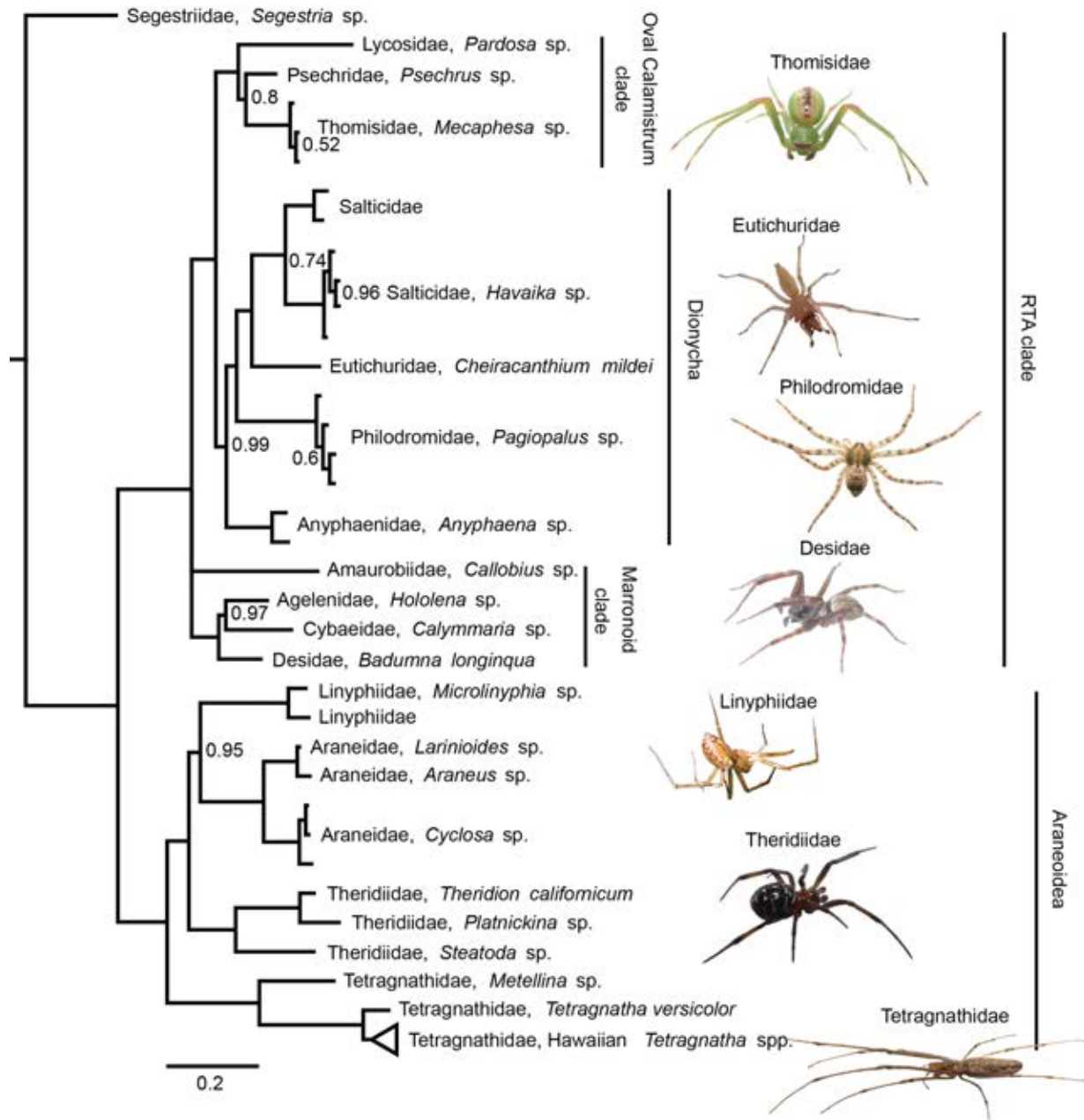
Figure 4. Bayesian consensus phylogeny of 83 spiders in 16 families, based on a 4,214 bp alignment of 18S, ITS1, 5.8S, ITS2 and 28S. The phylogeny is rooted using the basal haplogyne *Segestria* sp. The clade containing Hawaiian members of the genus *Tetragnatha* is collapsed (the uncolapsed clade is shown in Fig. 5). Only posterior probability values below 1 are displayed.
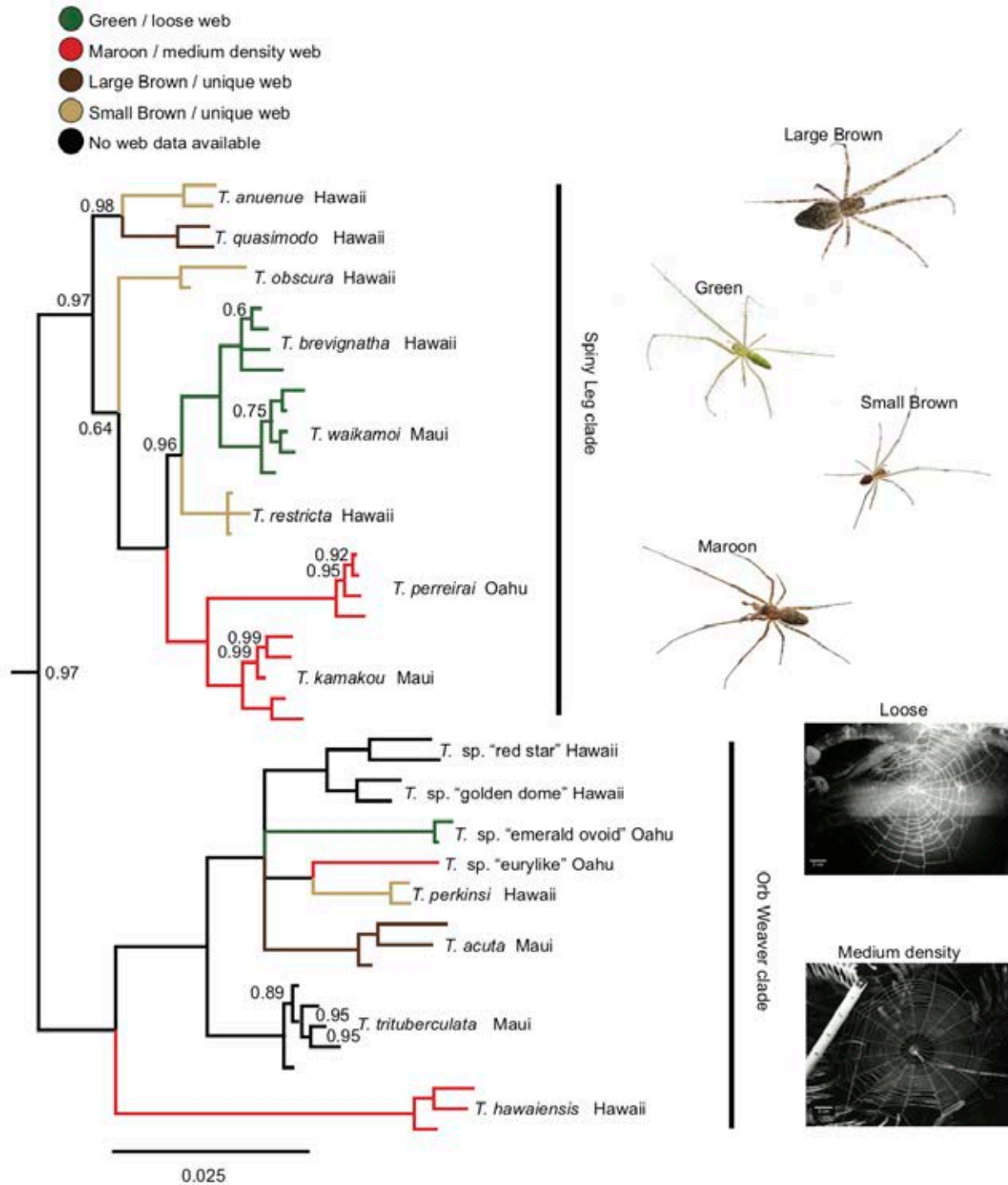
Figure 5. Section of the same phylogeny as Fig. 4, with expansion of the clade of 16 Hawaiian *Tetragnatha* species. Different "Spiny Leg" ecomorphs and web architectures are indicated by branch coloration. Only posterior probability values below 1 are displayed.

We generated an alignment of 3,656 bp for 117 concatenated 18S, 5.8S and 28S sequences of plants, fungi, annelids and arthropods. Our phylogeny was well supported (most posterior support values equal one; Fig. 3). A basal split separated plants from fungi and animals. Within plants, the genus *Peperomia* was recovered as monophyletic. Fungi formed the sister group of animals. Within animals, annelids formed a separate clade from arthropods. Arthropods separated into arachnids and hexapods. Each separate arthropod order formed well supported groups. The hexapod phylogeny generally resembled that found in latest phylogenomic work [52]. The Collembola species *Salina* sp. formed the base to the insect tree, followed by the odonate *Argia* sp. A higher branch led to Blattodea, Hemiptera and Orthoptera. However, the support values for the relationships between these three orders were comparatively low (~ 0.85). Finally, holometabolan insects (Hymenoptera, Coleoptera and Lepidoptera) were recovered as monophyletic. The two Acari species, together with Opiliones, formed the sister clade to the monophyletic Araneae clade.

Next, we generated a separate alignment of rDNA sequences for 83 spiders, including both ITS regions (totaling 4,214 bp). The spider phylogeny was also strongly supported (Fig. 4). Overall, our phylogenetic tree topology agreed with the most recent phylogenetic work of [53] and [35]. With the haplogyne *Segestria* sp. (family Segestriidae) forming the root, we recovered the so-called RTA clade (represented in our dataset by families Agelenidae, Amaurobiidae, Anyphaenidae, Cybaeidae, Desidae, Eutichuridae, Lycosidae, Philodromidae, Psechridae, Salticidae and Thomisidae) and the Araneoidea (Araneidae, Linyphiidae, Tetragnathidae, Theridiidae) as two well supported monophyla. Within these clades, all families and genera formed well supported monophyletic groups. Similar to recent studies, we found the Marronoid clade as basal to the rest of the RTA clade; more derived clades were the Oval Calamistrum and the Dionycha clade. Inter-family relationships also closely matched those found in recent work: Lycosidae was basal to the clade formed by Psechridae and Thomisidae; Salticidae was closest to Eutichuridae and Philodromidae, with Anyphaenidae falling basal within Dionycha. Within Araneoidea, our results differed slightly from recent studies in that we recovered Tetragnathidae, rather than Theridiidae, as basal.

We recovered Hawaiian *Tetragnatha* as a well supported monophyletic clade within the Tetragnathidae. We found two main clades of Hawaiian *Tetragnatha* (Fig. 5), both of which have been supported by earlier work [54-57]: the orb weaving clade and the "Spiny Leg clade" of actively hunting species. All *Tetragnatha* species formed monophyletic groups, and the relationships among different species were mostly well supported. Within the Spiny Leg clade, species fell into one of four ecotypes, each of which is associated with a particular substrate type: "large brown" (*T. quasimodo*) with tree bark, "small brown" (*T. anuenue*, *T. obscura* and *T. restricta*) with twigs, "green" (*T. brevignatha* and *T. waikamoi*) with green leaves, and "maroon" (*T. perreirai* and *T. kamakou*) with lichen. While green and maroon ecotypes clustered phylogenetically, small brown species appeared in three separate clades on the tree. Within the orb weaving clade, *T. hawaiensis*, a generalist species which occurs on all of the Hawaiian Islands, fell basal. The characteristic web structures of some of these species have been documented [35, 58]. We found a pattern of apparent convergence in web structure for some species. *T.* sp. "emerald ovoid" spins a loose web with widely spaced rows of capture silk. *T. hawaiensis* and *T.* sp. "eurylike," which are distant relatives within the Hawaiian *Tetragnatha* clade, both spin webs of medium silk density, i.e. with more rows of capture silk per unit area than *T.* sp. "emerald ovoid." *T. perkinsi* and *T. acuta* each spin a web structure that is not comparable in its silk density or size to any other known *Tetragnatha* species in this group [58], and are thus classified as "unique".

Taxonomic resolution

Our inferred genetic distances for rDNA sequences within and between Hawaiian *Tetragnatha* species were significantly correlated to those found for COI sequences of the same taxa (R^2 = 0.70, *P* < 0.001) (Fig. 6a). A Mantel test also suggested highly significant correlation of mitochondrial COI and nuclear rDNA based distances (Mantel test, 9999 replicates, *P* < 0.001). Hence, the rDNA cluster supported a very similar pattern of genetic differentiation to COI. However, the faster evolutionary rate of COI was reflected in lower distances for the whole rDNA than for COI. Interspecific distances were significantly higher than intraspecific ones for COI and rDNA (Fig 6b,c). No overlap of intra and interspecific distances was evident for COI, suggesting the presence of a barcode gap. A small overlap of intra and interspecific distances was evident for the rDNA (Supplementary Table 3). Like the combined rDNA cluster, genetic distances for different parts of the rDNA cluster all showed significant correlation with COI based distances, when analyzed separately (R^2 28S = 0.57, R^2 ITS1 = 0.68, R^2 ITS2 = 0.56, *P* < 0.001) (Supplementary Fig. 5). While the 28SrDNA showed considerably lower distances than COI, those for ITS1 and ITS2 were more comparable to COI (Supplementary Fig. 5b-d). Yet, interspecific and intraspecific distances for COI were significantly different from those for any part of the rDNA cluster (Pairwise Wilcoxon Test, FDR corrected *P* < 0.05).
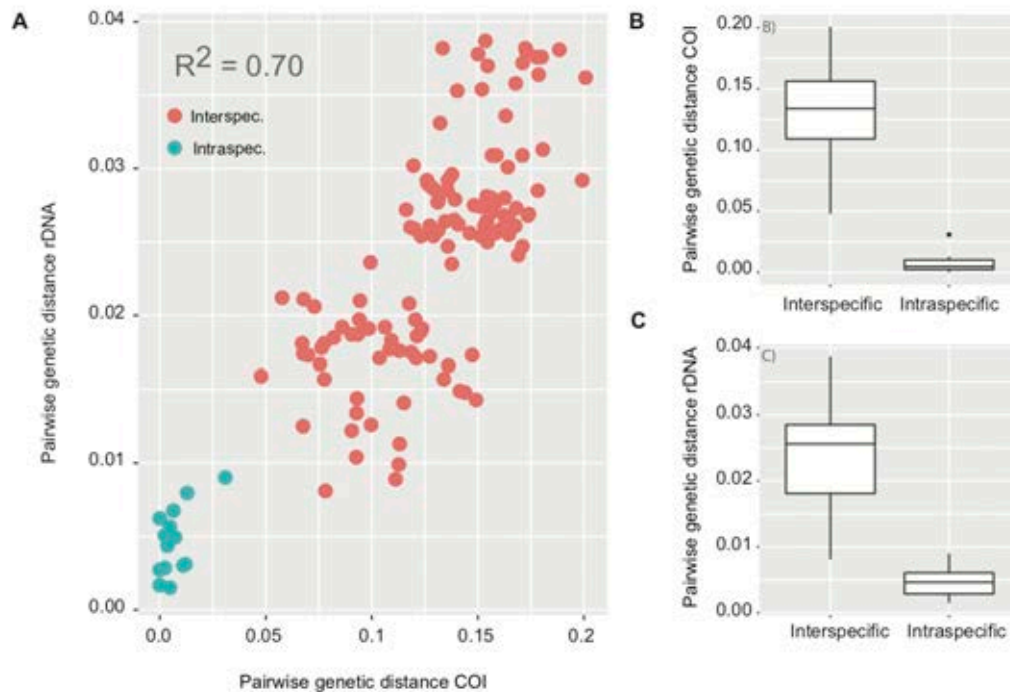

Figure 6 Inter and intraspecific genetic distances for the nuclear rDNA and mitochondrial COI for Hawaiian *Tetragnatha* spiders. A) Correlation of pairwise genetic distance between (red) and within (green) 16 Hawaiian *Tetragnatha* species based on COI and the full rDNA amplicon. B) Interspecific and intraspecific genetic distances for the same spider species based on mitochondrial COI and C) the whole rDNA amplicon.

Field trial in the Amazon rainforest

On March 26, 2018 we set out to test this method and a portable laboratory (as described in Pomerantz, et al. [25]) during an expedition to the Peruvian Amazon at the Refugio Amazonas Lodge (Supplementary Fig. 6). This field site is a "Terra firme" forest in the sector of "Condenado", approximately two and a half hours by boat up river from the native community of

Infierno on the buffer zone of the Tambopata National Reserve. We collected plant and insect material, extracted DNA, amplified the rDNA cluster, and sequenced material on the MinION platform using the MinKNOW offline software (provided by ONT). The first run generated 17,149 reads and the second one 20,167 reads. We generated consensus sequences for five out of the seven analyzed specimens. One plant sample and the grasshopper could not be assembled due to too low read coverage. Moreover, BLAST analysis of the reads assigned to the grasshopper suggested that we had sequenced a mite, instead of the grasshopper DNA. The unidentified insect eggs resulted in a butterfly consensus sequence, possibly a Pierid species.

Nanopore based arthropod metabarcoding
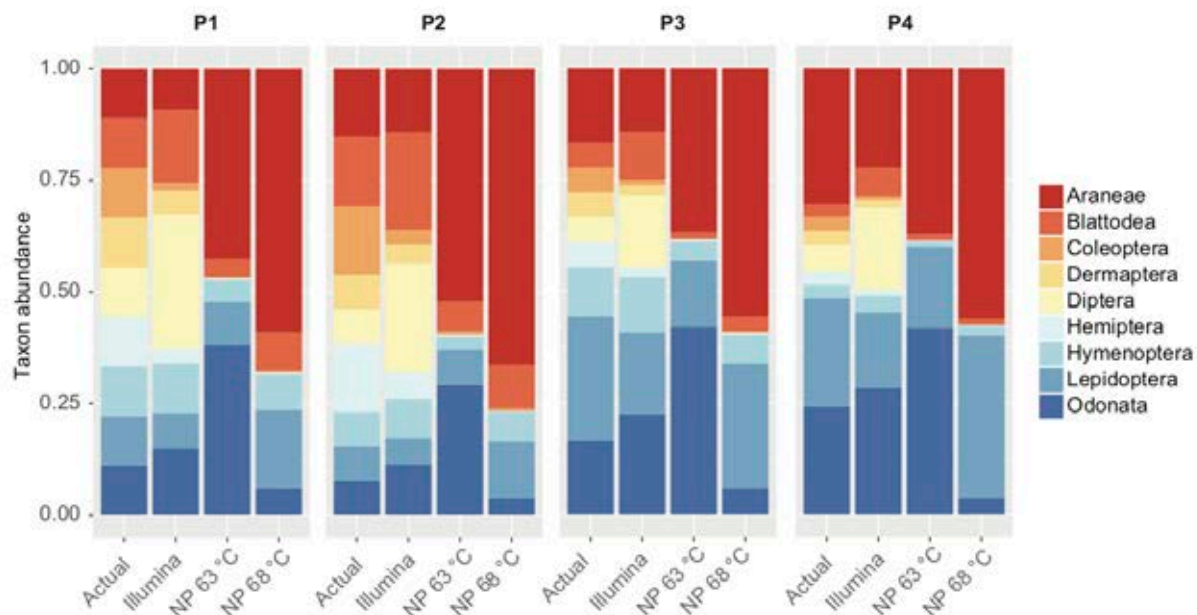
Figure 7: Relative abundances for nine arthropod species in our four mock communities (actual), compared to an Illumina amplicon sequencing protocol, and nanopore protocols at 63 °C and 68 °C annealing temperature

On average, we recovered 2,645 reads for each Illumina sequenced mock community and 1,149 for each nanopore mock community. The optimized Illumina amplicon sequencing based 18SrDNA protocol resulted in a very good taxon recovery. All nine taxa were recovered from all four mock communities (Fig. 7). Moreover, the Illumina based protocol allowed for accurate predictions of taxon abundances. The average fold change between input DNA and recovered read count was closely distributed around zero (Supplementary Fig 6). In contrast, the long-read nanopore protocol showed very biased qualitative and quantitative taxon recovery (Fig. 7). On average, only 83.33 % of taxa were recovered per nanopore sequenced mock community. Moreover, the fold change of input DNA and recovered read count were highly biased between taxa. Some taxa were considerably over or underrepresented among the read population. This led to a significantly higher variation of fold change between input DNA and read count compared to the Illumina amplicon based protocol (Levene's test $P < 0.05$; Supplementary Fig. 7). A reduction of PCR annealing temperature did result in a considerable increase of Odonata sequences, but overall did not have a strong effect on qualitative (77.78 % of taxa recovered) or quantitative taxon recovery (Fig. 7). The variation of fold change between different PCR annealing temperatures was not significantly different (Levene's test, $P > 0.05$). A reduction of

PCR cycle number by 10 also did not yield any significant effect on qualitative (88.89 % of taxa recovered) or quantitative taxon recovery (Supplementary Fig. 7).

Discussion

Phylogenetic and taxonomic utility of long rDNA amplicons
Developments in long-read sequencing hold great promise for molecular taxonomy and phylogenetics across very broad taxonomic scales. We recovered phylogenetic relationships across the eukaryote tree of life, which were mostly consistent with the current state of research (e.g. [52]). Separate orders of arthropods all formed well supported monophyletic groups. Our spider phylogeny was highly congruent with recent work based on whole transcriptomes [35] and multi-amplicon data [53]. Moreover, using the rDNA cluster allowed us to resolve young phylogenetic divergences: the relationships within the recent adaptive radiation of the genus *Tetragnatha* in Hawaii confirmed previous research [59, 60].

Besides their high phylogenetic utility, long rDNA amplicons showed excellent support for taxonomic hypotheses. All morphologically identified species of Hawaiian *Tetragnatha* were recovered as monophyletic groups. The divergence patterns and taxonomic classifications of spiders based on rDNA were strongly correlated to those based on mitochondrial COI, the most commonly used animal barcode marker [4]. rDNA may thus be ideal to complement mitochondrial barcoding. A universal and variable nuclear marker as a supplement to COI barcoding will be particularly useful in cases of mito-nuclear discordance due to male biased gene flow [10, 61], hybridization [12] or infections with reproductive parasites [11].

Their high phylogenetic utility across very broad taxonomic categories also provide long rDNA amplicons with a distinct advantage over short read barcoding protocols, which are not well suited to support broad scale phylogenetic hypotheses [62]. The inclusion of long amplicons would make it possible to scale up barcoding from simple taxon assignment to community wide phylogenetic inferences [9]. Recently, the amplification of whole mitochondrial genomes was suggested for animal barcoding [63]. This would increase taxonomic and phylogenetic resolution and thus alleviate some disadvantages of short COI amplicons. However, it is challenging to develop truly universal primers to target mitochondrial genomes across a wide range of taxonomic groups [64]. Moreover, mitochondrial genomes will not allow cases of mito-nuclear discordance to be identified. A straightforward way to achieve highly resolved phylogenies may be the combination of long rDNA amplicon sequencing with multiplex PCRs of short mitochondrial amplicons, to amplify multiple mitochondrial DNA fragments [65].

Simple, accurate, universal and cost efficient long-read DNA barcoding
Despite the high raw read error of nanopore data, consensus sequences were highly accurate, and library preparation and sequencing for our protocol are simple and cost efficient. Using a single pair of universal primers, long rDNA amplicons can be amplified across diverse eukaryote taxa. A simple dual indexing approach during PCR allows large numbers of samples to be pooled before library preparation [27]. Only a single PCR is required per specimen, while subsequent cleanup and library preparation can be performed on pooled samples. The simplicity of our approach is additionally highlighted by its effectiveness even under field conditions in a remote rainforest site. Nanopore sequencing technology is affordable and universally available to any laboratory. Our ONT MinION generated about 250,000 reads per run. Aiming for about 1,000 reads per amplified specimen, 250 long rDNA barcodes could be generated in single MinION run. Input DNA amounts for different specimens will have to be carefully balanced to maximize the recovery. The total reagent costs, including PCR, library

preparation and sequencing, then amount to less than $4 for each long barcode sequence generated.

Pitfalls of nanopore based long-read barcoding

While our protocol was generally straightforward and reliable, we found several drawbacks, which require further considerations and optimization. First, it needs to be noted that long rDNA amplification will not be possible with highly degraded DNA molecules, e.g. from historical specimens [66]. Moreover, amplification success of long range PCRs proved less consistent than that for amplification of short amplicons. We observed a complete failure of some PCRs when too high template DNA concentrations were loaded. The long range polymerase may be more sensitive to PCR inhibitors present in some arthropod DNA extractions [67]. PCR conditions will have to be carefully optimized for reliable and consistent amplification. We also found that highly universal eukaryote primers may result in undesired amplification, for example plants from beetle and butterfly larval guts, phoretic mites, or fungal sequences. However, as long as the DNA of the target taxon is still dominating the resulting amplicon mixture, this undesired amplification will not affect consensus calling. It may be advisable to check the taxonomic composition of amplicon samples before assembly, e.g. by blasting against a reference library. To avoid unspecific amplification, PCR primers could also be redesigned to exclude certain lineages from amplification. It should also be noted that our approach results in only a single consensus sequence for each processed specimen. As a diploid marker, the rDNA cluster can contain heterozygous positions in some specimens, in particular within the ITS regions. This information is currently lost, and a different assembly approach may be necessary to recover heterozygosity as well. Furthermore, index length and edit distance are also important considerations. We used indexes of 15 bp and with a minimum distance of 10 bp to index both sides of our amplicons. Index edit distance of only 4 bp between samples already led to considerable cross-specimen index bleeding. It may thus be better to increase the length and edit distances of indexes. Indexes of 20 or 30 bp could be easily attached to the 5'-tails of PCR primers without strongly affecting PCR efficiency.

Nanopore based arthropod metabarcoding

It is well known that Illumina amplicon sequencing of short 18SrDNA fragments can yield very accurate qualitative and quantitative taxon recovery in metabarcoding experiments [48], a finding that is confirmed by our results. In contrast, little is known on the performance of long-read nanopore sequencing for community diversity assessments [32]. Our long barcode based approach resulted in the dropout of several taxa and highly skewed relative taxon abundances. Skewed abundances were already found in microbial community analysis using nanopore [32]. In the simplest case, primer mismatches may be responsible for biased amplification [32, 68]. However, the targeted priming sites in our study were extremely conserved. Also, a change of PCR cycle number and annealing temperature did not have a strong effect on taxon abundances, as would be expected in the case of PCR priming bias [69]. Another possibility is the preferential amplification of template molecules with a certain GC content by the DNA polymerase [33]. However, we found the GC content of the rDNA cluster to be very stable across taxa. Yet another potential explanation for the differential recovery of taxa in community samples is taxonomic bias in DNA degradation [70], but we do not expect DNA degradation to have played a role in our experiment because we used only high quality DNA extractions (verified by gel electrophoresis) from fresh specimens. The most plausible explanation appears to be that variable rDNA lengths are found between different taxa. It is well known that shorter sequences are amplified preferentially in a PCR, especially after it reaches the plateau stage [71]. Such dominance of shorter amplicons could explain the observed biases very well. In fact, the most abundant taxon in our pools was a spider, which also had the shortest amplicon length. The dominant amplification of shorter sequences may also explain the amplification of

plant DNA from a butterfly and a flour beetle larva, as plants showed considerably shorter rDNA amplicons than insects. We found a very high variation of rDNA amplicon length within many taxonomic groups, this could be a considerable problem for long read metabarcoding applications. More research into the causes and possible mitigation of these biases will be required before long-read sequencing can be routinely utilized for metabarcoding applications.

Conclusion

Sequencing long dual indexed rDNA amplicons on Oxford Nanopore Technologies' MinION is a simple, cost effective, accurate and universal approach for eukaryote DNA barcoding. Long rDNA amplicons offer high phylogenetic and taxonomic resolution across broad taxonomic scales from kingdom down to species. They also prove to be an excellent complement to mitochondrial COI based barcoding in arthropods. However, despite the long-read advantages in the analysis of separate specimens, we found considerable biases associated with sequencing bulk community samples. The observed taxonomic bias is possibly a result of taxon-specific length variation of the rDNA cluster and preferential amplification of species with shorter rDNA. Further research into the sources of the observed bias is required before long rDNA amplicon sequencing can be utilized as a reliable resource for the analysis of bulk samples.

Author contributions

HK and SP designed the study. HK, AP, SRK, JYL, VS and JDS collected the specimens. Laboratory work was carried out by HK, AP and SRK and the data were subsequently analyzed by HK, AP, JBH, SRK and SP. The paper was writen by HK, AP, JBH, SRK, JYL, VS, JDS, NHP, RGG, SP.

Acknowledgements

References

1. Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., and Kinzig, A. (2000). Global biodiversity scenarios for the year 2100. Science *287*, 1770-1774.
2. Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., and Sexton, J.O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. Science *344*, 1246752.
3. Rominger, A., Goodman, K., Lim, J., Armstrong, E., Becking, L., Bennett, G., Brewer, M., Cotoras, D., Ewing, C., and Harte, J. (2016). Community assembly on isolated islands: macroecology meets evolution. Global ecology and biogeography *25*, 769-780.
4. Hebert, P.D., Ratnasingham, S., and de Waard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London B: Biological Sciences *270*, S96-S99.
5. Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., and Crous, P.W. (2012). Nuclear ribosomal internal

transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences *109*, 6241-6246.

6. China Plant BOL Group, Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L., and Chen, S.-L. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences *108*, 19641-19646.

7. Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B., and Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Scientific reports *5*, 9687.

8. Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., and Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution *3*, 613-623.

9. Graham, C.H., and Fine, P.V. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. Ecology Letters *11*, 1265-1277.

10. Krehenwinkel, H., Graze, M., Rödder, D., Tanaka, K., Baba, Y.G., Muster, C., and Uhl, G. (2016). A phylogeographical survey of a highly dispersive spider reveals eastern Asia as a major glacial refugium for Palaearctic fauna. Journal of Biogeography *43*, 1583-1594.

11. Hurst, G.D., and Jiggins, F.M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society of London B: Biological Sciences *272*, 1525-1534.

12. Bernatchez, L., Glémet, H., Wilson, C.C., and Danzmann, R.G. (1995). Introgression and fixation of Arctic char (Salvelinus alpinus) mitochondrial genome in an allopatric population of brook trout (Salvelinus fontinalis). Canadian Journal of Fisheries and Aquatic Sciences *52*, 179-185.

13. Melo−Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N., and Alves, P. (2005). Invasion from the cold past: extensive introgression of mountain hare (Lepus timidus) mitochondrial DNA into three other hare species in northern Iberia. Molecular Ecology *14*, 2459-2464.

14. Soltis, P.S., and Soltis, D.E. (1998). Molecular evolution of 18S rDNA in angiosperms: implications for character weighting in phylogenetic analysis. In Molecular systematics of plants II. (Springer), pp. 188-210.

15. Hillis, D.M., and Dixon, M.T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. The Quarterly review of biology *66*, 411-453.

16. Black IV, W.C., Klompen, J., and Keirans, J.E. (1997). Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene. Molecular Phylogenetics and Evolution *7*, 129-144.

17. Powers, T.O., Todd, T., Burnell, A., Murray, P., Fleming, C., Szalanski, A.L., Adams, B., and Harris, T. (1997). The rDNA internal transcribed spacer region as a taxonomic marker for nematodes. Journal of Nematology *29*, 441.

18. Sonnenberg, R., Nolte, A.W., and Tautz, D. (2007). An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. Frontiers in zoology *4*, 6.

19. Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., and Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences *109*, 16208-16212.

20. von der Schulenburg, J.H.G., Hancock, J.M., Pagnamenta, A., Sloggett, J.J., Majerus, M.E., and Hurst, G.D. (2001). Extreme length and length variation in the first ribosomal internal transcribed spacer of ladybird beetles (Coleoptera: Coccinellidae). Molecular Biology and Evolution *18*, 648-660.

21.   Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., and Fiddes, I.T. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature biotechnology *36*, 338.

22.   Heeger, F., Bourne, E.C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., Overmann, J., Mazzoni, C.J., and Monaghan, M.T. (2018). Long-read DNA metabarcoding of ribosomal rRNA in the analysis of fungi from aquatic environments. bioRxiv, 283127.

23.   Tedersoo, L., Tooming–Klunderud, A., and Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. New Phytologist *217*, 1370-1385.

24.   Giordano, F., Aigrain, L., Quail, M.A., Coupland, P., Bonfield, J.K., Davies, R.M., Tischler, G., Jackson, D.K., Keane, T.M., and Li, J. (2017). De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. Scientific reports *7*, 3935.

25.   Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A., Barrio-Amorós, C.L., Salazar-Valenzuela, D., and Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience *7*, giy033.

26.   Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S., Kristiansson, E., Kagami, M., and Nilsson, R.H. (2018). Introducing ribosomal tandem repeat barcoding for fungi. bioRxiv, 310540.

27.   Srivathsan, A., Baloğlu, B., Wang, W., Tan, W.X., Bertrand, D., Ng, A.H., Boey, E.J., Koh, J.J., Nagarajan, N., and Meier, R. (2018). A Min ION™–based pipeline for fast and cost–effective DNA barcoding. Molecular ecology resources.

28.   Giribet, G., and Edgecombe, G.D. (2012). Reevaluating the arthropod tree of life. Annual review of entomology *57*, 167-186.

29.   Hochkirch, A. (2016). The insect crisis we can't ignore. Nature News *539*, 141.

30.   Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., and Mikhail, A. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature *530*, 228.

31.   Edwards, A., Debbonaire, A.R., Sattler, B., Mur, L.A., and Hodson, A.J. (2016). Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N. bioRxiv, 073965.

32.   Benítez-Páez, A., Portune, K.J., and Sanz, Y. (2016). Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION™ portable nanopore sequencer. GigaScience *5*, 4.

33.   Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E., and Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular ecology resources.

34.   Krehenwinkel, H., Wolf, M., Lim, J.Y., Rominger, A.J., Simison, W.B., and Gillespie, R.G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports *7*, 17668.

35.   Fernández, R., Kallal, R.J., Dimitrov, D., Ballesteros, J.A., Arnedo, M.A., Giribet, G., and Hormiga, G. (2018). Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. Current Biology *28*, 1489-1497. e1485.

36.   De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long read sequencing data. bioRxiv, 237180.

37.   Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research *27*, 737-746.

38.   Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution *30*, 2725-2729.

39.   Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: Next–generation sequencing for plant systematics. American Journal of Botany *99*, 349-364.
40.   Bolger, A., and Giorgi, F. Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. URL http://www. usadellab. org/cms/index. php.
41.   Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754-1760.
42.   Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.
43.   Lanfear, R., Calcott, B., Ho, S.Y., and Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution *29*, 1695-1701.
44.   Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754-755.
45.   Dray, S., and Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software *22*, 1-20.
46.   Machida, R.J., and Knowlton, N. (2012). PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. PLoS one *7*, e46180.
47.   Krehenwinkel, H., Kennedy, S., Pekár, S., and Gillespie, R.G. (2017). A cost–efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large–scale gut content analysis by Illumina sequencing. Methods in Ecology and Evolution *8*, 126-134.
48.   Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.
49.   Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady, Volume 10. pp. 707-710.
50.   Šošić, M., and Šikić, M. (2017). Edlib: a C/C++ library for fast, exact sequence alignment using edit distance. Bioinformatics *33*, 1394-1395.
51.   Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. bioRxiv, 015552.
52.   Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., and Beutel, R.G. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science *346*, 763-767.
53.   Wheeler, W.C., Coddington, J.A., Crowley, L.M., Dimitrov, D., Goloboff, P.A., Griswold, C.E., Hormiga, G., Prendini, L., Ramírez, M.J., and Sierwald, P. (2017). The spider tree of life: phylogeny of Araneae based on target–gene analyses from an extensive taxon sampling. Cladistics *33*, 574-616.
54.   Gillespie, R.G. (1991). Hawaiian spiders of the genus Tetragnatha: I. Spiny leg clade. Journal of Arachnology, 174-209.
55.   Gillespie, R.G. (1999). Comparison of rates of speciation in web-building and non-web-building groups within a Hawaiian spider radiation. Journal of Arachnology, 79-85.
56.   Gillespie, R.G. (2016). Island time and the interplay between ecology and evolution in species diversification. Evolutionary applications *9*, 53-73.
57.   Gillespie, R.G., Croom, H.B., and Hasty, G.L. (1997). Phylogenetic relationships and adaptive shifts among major clades of Tetragnatha spiders (Araneae: Tetragnathidae) in Hawai'i.
58.   Blackledge, T.A., Binford, G.J., and Gillespie, R.G. (2003). Resource use within a community of Hawaiian spiders (Araneae: Tetragnathidae). In Annales Zoologici Fennici. (JSTOR), pp. 293-303.

59. Blackledge, T.A., and Gillespie, R.G. (2004). Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. Proceedings of the National Academy of Sciences of the United States of America *101*, 16228-16233.

60. Gillespie, R. (2004). Community assembly through adaptive radiation in Hawaiian spiders. Science *303*, 356-359.

61. Wilmer, J.W., Hall, L., Barratt, E., and Moritz, C. (1999). Genetic Structure and Male-Mediated Gene Flow in the Ghost Bat (Macroderma gigas). Evolution, 1582-1591.

62. Kjer, K.M., Zhou, X., Frandsen, P.B., Thomas, J.A., and Blahnik, R.J. (2014). Moving toward species-level phylogeny using ribosomal DNA and COI barcodes: an example from the diverse caddisfly genus Chimarra (Trichoptera: Philopotamidae). Arthropod Systematics & Phylogeny *72*, 345-354.

63. Deiner, K., Renshaw, M.A., Li, Y., Olds, B.P., Lodge, D.M., and Pfrender, M.E. (2017). Long–range PCR allows sequencing of mitochondrial genomes from environmental DNA. Methods in Ecology and Evolution *8*, 1888-1898.

64. Briscoe, A.G., Goodacre, S., Masta, S.E., Taylor, M.I., Arnedo, M.A., Penney, D., Kenny, J., and Creer, S. (2013). Can long-range PCR be used to amplify genetically divergent mitochondrial genomes for comparative phylogenetics? A case study within spiders (Arthropoda: Araneae). PLoS one *8*, e62404.

65. Krehenwinkel, H., Kennedy, S., Rueda, M., and Gillespie, R. (in press). Low cost molecular systematics of entire arthropod communities: Primer sets for rapid multi locus analyses by multiplex PCRs and Illumina amplicon sequencing. . Methods in Ecology and Evolution.

66. Krehenwinkel, H., and Pekar, S. (2015). An analysis of factors affecting genotyping success from museum specimens reveals an increase of genetic and morphological variation during a historical range expansion of a European spider. PLoS one *10*, e0136337.

67. Margam, V.M., Gachomo, E.W., Shukle, J.H., Ariyo, O.O., Seufferheld, M.J., and Kotchoni, S.O. (2010). A simplified arthropod genomic-DNA extraction protocol for polymerase chain reaction (PCR)-based specimen identification through barcoding. Molecular biology reports *37*, 3631-3635.

68. Sipos, R., Székely, A.J., Palatinszky, M., Révész, S., Márialigeti, K., and Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiology Ecology *60*, 341-350.

69. Suzuki, M.T., and Giovannoni, S.J. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and environmental microbiology *62*, 625-630.

70. Krehenwinkel, H., Fong, M., Kennedy, S., Huang, E.G., Noriyuki, S., Cayetano, L., and Gillespie, R. (2018). The effect of DNA degradation bias in passive sampling devices on metabarcoding studies of arthropod communities and their associated microbiota. PLoS one *13*, e0189188.

71. Wattier, R., Engel, C., Saumitou–Laprade, P., and Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in Gracilaria gracilis (Rhodophyta). Molecular Ecology *7*, 1569-1573.

RESEARCH

Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building

Aaron Pomerantz [1,*], Nicolás Peñafiel[2], Alejandro Arteaga[3,4,5], Lucas Bustamante[5], Frank Pichardo[5], Luis A. Coloma[6], César L. Barrio-Amorós[7], David Salazar-Valenzuela[2] and Stefan Prost[1,8,*]

[1]Department of Integrative Biology, University of California, Berkeley, CA, USA, [2]Centro de Investigación de la Biodiversidad y Cambio Climático (BioCamb) e Ingeniería en Biodiversidad y Recursos Genéticos, Facultad de Ciencias de Medio Ambiente, Universidad Tecnológica Indoamérica, Machala y Sabanilla, Quito, Ecuador, [3]Richard Gilder Graduate School, American Museum of Natural History, New York, USA, [4]Department of Herpetology, American Museum of Natural History, New York, USA, [5]Tropical Herping, Quito, Ecuador, [6]Centro Jambatu de Investigación y Conservación de Anfibios, Fundación Otonga, Quito, Ecuador, [7]Doc Frog Expeditions, Uvita, Costa Rica and [8]Program for Conservation Genomics, Department of Biology, Stanford University, Stanford, CA, USA

*Correspondence address. Aaron Pomerantz. Department of Integrative Biology, University of California, Berkeley, CA, USA. Tel: +1 310 9464911, E-mail: pomerantz_aaron@berkeley.edu http://orcid.org/0000-0002-6412-9001; Stefan Prost, E-mail: stefanprost.research@protonmail.com

Abstract

Background: Advancements in portable scientific instruments provide promising avenues to expedite field work in order to understand the diverse array of organisms that inhabit our planet. Here, we tested the feasibility for *in situ* molecular analyses of endemic fauna using a portable laboratory fitting within a single backpack in one of the world's most imperiled biodiversity hotspots, the Ecuadorian Chocó rainforest. We used portable equipment, including the MinION nanopore sequencer (Oxford Nanopore Technologies) and the miniPCR (miniPCR), to perform DNA extraction, polymerase chain reaction amplification, and real-time DNA barcoding of reptile specimens in the field. **Findings:** We demonstrate that nanopore sequencing can be implemented in a remote tropical forest to quickly and accurately identify species using DNA barcoding, as we generated consensus sequences for species resolution with an accuracy of >99% in less than 24 hours after collecting specimens. The flexibility of our mobile laboratory further allowed us to generate sequence information at the Universidad Tecnológica Indoamérica in Quito for rare, endangered, and undescribed species. This includes the recently rediscovered Jambato toad, which was thought to be extinct for 28 years. Sequences generated on the MinION required as few as 30 reads to achieve high accuracy relative to Sanger sequencing, and with further multiplexing of samples, nanopore sequencing can become a cost-effective approach for rapid and portable DNA barcoding. **Conclusions:** Overall, we establish how mobile laboratories and nanopore sequencing can help to accelerate species identification in remote areas to aid in conservation efforts and be applied to research facilities in developing countries. This opens up possibilities for biodiversity

studies by promoting local research capacity building, teaching nonspecialists and students about the environment, tackling wildlife crime, and promoting conservation via research-focused ecotourism.

Data Description

Background

Biodiversity is defined as the variety of life found on Earth, including variation in genes, species, and ecosystems. While about 1.9 million species have been described to date, there are an estimated 5–30 million species in total on the planet, with most of the diversity contained within tropical rainforests [1–3]. For instance, Ecuador, despite its small size of 283,561 km^2 (roughly 1.5% of South America), is one of the most biologically diverse countries in the world [4, 5]. Biodiversity is fundamentally important to natural and agro-ecosystems; it provides humans with an array of foods and materials, contributes to medical discoveries, furnishes the economy, and supports ecological services that make life on our planet possible [6]. Today, species are going extinct at an accelerated rate because of environmental changes caused by human activities, including habitat loss, spread of nonnative species, pollution, and climate change [7, 8]. All of these threats have put a serious strain on the diversity of species on Earth.

In the past decade, an ever-growing body of readily accessible knowledge, coupled with new tools in molecular genetics and bioinformatics, have resulted in species being described with greater accuracy, in greater detail, and with additional information on morphological differences. As a result of this increase in quality and content, desirable as it is, the actual process of species description has become slower, while the rate at which species are being lost to extinction has become faster. For many groups of animals, species delimitation can be challenging using solely morphological characteristics [9, 10]; this can be improved by incorporating molecular data [11, 12]. This is relevant for the conservation of threatened animals because programs and laws can be implemented more effectively when the existence of a species or population is formally described. DNA barcoding, which is a diagnostic technique that uses short conserved DNA sequences, has become a popular tool for a variety of studies, including species identification and molecular phylogenetic inference [13–15]. Ongoing initiatives, such as Barcode of Life [16], seek to identify species and create large-scale reference databases via diagnostic DNA sequences using a standardized approach to accelerate taxonomic progress.

While projects that use standard molecular markers have grown in popularity in the last decade, a fundamental challenge remains in transporting biological material to a site where DNA sequencing can be performed. Furthermore, complex and overwhelming regulations can impede biological research in biodiverse countries and can make it challenging to export material out of the country of origin [17, 18]. Additionally, many research institutions in developing parts of the world do not have access to conventional sequencing technologies within the country, further limiting identification options. This is the case for Ecuador, where most laboratories ship their samples internationally to be sequenced, often creating a delay of weeks to months between tissue collection and the availability of the sequence data. Performing genetic analyses on site or at a nearby facility within the country can help to avoid project delays and decrease the risk of sample quality decline associated with ex-

tensive transport. It is now possible to take portable lab equipment to remote regions, perform *in situ* experiments, and obtain genetic information relevant for biological studies and conservation policies in real time.

Portable Sequencing

The MinION (Oxford Nanopore Technologies) is a recently developed nanopore-based DNA sequencing platform. This technology has several advantages over traditional sequencing technologies, including long-read output, low initial startup costs relative to other commercial sequencers, portability, and rapid real-time analysis (reviewed by [19, 20]). Due to its small size (10 × 3.2 × 2 cm), light weight (90 grams), and ease of power and data transfer (a single USB connection to a standard laptop computer), the MinION has emerged as a valuable tool for portable sequencing projects. This device has been used at remote sites outside of conventional labs, including West Africa to monitor the 2014–2015 Ebola outbreak [21] and Brazil for Zika virus outbreak surveillance [22, 23]. It has also been used in the Antarctic to sequence microbial communities [24, 25], in Tanzania to sequence frog DNA [26], and in Snowdonia National Park, Wales, for shotgun genomic sequencing of closely related plant species [27]. The MinION has even been run aboard the International Space Station to evaluate performance off-Earth [28]. Indeed, nanopore sequencing appears to hold promise for a variety of molecular experiments in the field.

Scientists have mused over the possibility of a portable method for DNA barcoding for more than a decade [29, 15]. In this study, our goal was to determine if the steps involved in barcoding, including real-time sequencing with the MinION, could be carried out entirely during a field expedition. We specifically targeted DNA barcodes with existing reference databases because they are the standard approach in molecular biodiversity studies and allowed us to rapidly produce genetic data for the identification of several animal taxa by multiplexing. Our field site was situated in a remote tropical rainforest and did not offer the commodities of a sophisticated laboratory environment, including consistent power sources and Internet access. We assessed the feasibility for *in situ* genetic sequencing of reptiles and amphibians for rapid species identification using a portable laboratory fitting within a single backpack at one of the world's most imperiled biodiversity hotspots, the Ecuadorian Chocó rainforest. We demonstrate that portable DNA amplicon sequencing with the MinION allows rapid, accurate, and efficient determination at the species level under remote tropical environmental conditions, as well as quick turnaround time for DNA barcodes of undescribed and threatened species at a research facility within the country.

Analyses

Site, sampling, digital photos, and tissue collection

We performed all field-based research in the Canandé Reserve (0.52993 N, 79.03541 W, 594 m), a 2,000-hectareprotected area owned by the Jocotoco Foundation [30] in Esmeraldas province, northwestern Ecuador. The reserve is located in the Chocó

ecoregion and is approximately 6 hours by car, depending on road conditions, from the city of Quito. The majority of organisms sampled in this study were located by space-constrained visual examination of ground-level substrates [1]. The remaining individuals were detected by turning over logs, rocks, and other surface objects. All specimens included in the genetic analyses were morphologically identified based on [2] and [3]. The sample (a tadpole, CJ 7191) of *Atelopus ignescens* was provided by the Museum of Centro Jambatu, Ecuador, and was preserved in ethanol 95%. We took vouchers for all samples collected and processed them in the field. These were deposited at the Museo de Zoología of the Universidad Tecnológica Indoamérica (MZUTI 5375 *Bothriechis schlegelii*, MZUTI 5383 *Lepidoblepharis* aff. *grandis*. (Gecko 1), MZUTI 5384 *Lepidoblepharis* aff. *buchwaldi* (Gecko 2).

Portable laboratory equipment and setup

The main items for portable laboratory equipment included the following: two MinION devices, a USB 3.0 cable, three SpotON flow cells (R9.5, Oxford Nanopore Technologies (ONT)), one miniPCR thermocycler (miniPCR), and a benchtop centrifuge (USA Scientific), as well as standard laboratory pipettes and sample racks (Fig. 1, Supplementary Fig. S3). The MinKNOW offline software (ONT) required for operation of the MinION was installed and ran on a Windows Vaio Sony laptop with an external SSD drive (VisionTek, 240GB). All heat block and temperature cycling steps were performed using the miniPCR machine, which is a portable thermo-cycler weighing 0.45 kg. The miniPCR was programmed via an application on the laptop and powered by an external battery (PowerAdd). The total amount of equipment could fit into one carry-on backpack; a full list of laboratory hardware is provided in Supplementary Table S1. Reagents for sequencing required frozen transport from the United States; this was achieved by use of packaging with cold packs in a Styrofoam box and later transfer to a plastic cool box with further cold packs upon arrival at Quito, Ecuador. MinION flow cells require storage at +2–8°C and were therefore transferred in a food storage container with chilled cold packs. At the field site, reagents and supplies were stored inside a local refrigerator and freezer.

Molecular techniques

Genomic DNA was extracted from fresh blood or tissue samples stored in 95% ethanol using either the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol and eluted in 100 μL ddH$_2$O or a modified salt precipitation method based on the Puregene DNA purification kit (Gentra Systems) that involved cellular lysis with SDS and proteinase K, protein precipitation using guanidine isothiocyanate, and DNA precipitation by isopropanol. Tools for manipulating and lysing tissues were sterilized with a flame in between processing samples. We amplified the following mitochondrial DNA fragments: 16S gene using primers 16Sar-L and 16Sbr-H-R from [4], CytB gene using primers L14910 and H16064 developed by [5], and the gene coding for subunit 4 of the NADH dehydrogenase with primers ND4 developed by [6]. All polymerase chain reaction (PCR) primers contained universal tailed sequences for the ONT barcoding kit (Supplementary Table S2). We used the ONT PCR Barcoding Kit that allows up to 12 different libraries (barcodes 1–12) to be combined and loaded onto a single flow cell at the same time. PCR reactions contained approximately 1 μLof PCR product, 2.5 μL10X PCR buffer, 1 μL25mM MgCl$_2$, 200 μM dNTP mix, 0.2 μM of each primer, and 0.25 Platinum Taq DNA Polymerase (Thermo Fisher Scientific) in a 25 μL total volume.

All samples for the first PCR run were amplified on the same miniPCR under the following settings: initial denaturation 94°C for 2 minutes, 35 cycles of denaturation at 94°C for 45 seconds, annealing at 56°C for 60 seconds, extension at 72°C for 60 seconds, and a final extension at 72°C for 120 seconds. Then a second round of PCR was carried out, including 2 μLof ONT PCR Barcode, 2 μL of first-round PCR product, 41 μL H$_2$O, and 50 μLPCR reaction mix (0.5 μLTaq DNA polymerase, 1 μL dNTP mix, 2 μL MgCl2, 41 μL H$_2$O). The second round of PCR barcode conditions were modified based on ONT protocol for the Platinum Taq Polymerase used in this study as follows: initial denaturation at 95°C for 3 minutes, 15 cycles of denaturation at 95°C for 15 seconds, annealing at 62°C for 15 seconds, extension at 72°C for 60 seconds, and final extension at 72°C for 120 seconds. For verification of samples sequenced in the field, PCR products were subsequently cleaned with Exonuclase I and Alkaline Phosphatase (Illustra ExoProStar by GE Healthcare) at the Universidad Tecnológica Indoamérica (UTI) in Quito and sent to Macrogen Inc. (Korea) for Sanger sequencing. All PCR products were sequenced on an ABI3730XL sequencer in both forward and reverse directions with the same primers that were used for amplification.

MinION sequencing

DNA library preparation was carried out according to the 1D PCR barcoding amplicons SQK-LSK108 protocol (ONT). Barcode DNA products were pooled with 5 μLof DNA CS (a positive control provided by ONT) and an end-repair was performed (NEB-Next Ultra II End-prep reaction buffer and enzyme mix, New England Biolabs), then purified using AMPure XP beads. Adapter ligation and tethering was then carried out with 20 μL ofAdapter Mix (ONT) and 50 μLof NEB Blunt/TA ligation Master Mix (New England Biolabs). The adapter ligated DNA library was then purified with AMPure beads, followed by the addition of Adapter Bead binding buffer (ONT), and finally eluted in 15 μLof Elution Buffer (ONT). Each R9 flow cell was primed with 1,000 μLof a mixture of Fuel Mix (ONT) and nuclease-free water. Then 12 μL of the amplicon library was diluted in 75 μL of running buffer with 35 μL RBF, 25.5 uL LLB, and 2.5 μL nuclease-free water and added to the flow cell via the SpotON sample port. The NC_48Hr_sequencing_FLO-MIN107_SQK-LSK108_plus_Basecaller.py protocol was initiated using the MinION control software, MinKNOW (offline version provided by ONT).

Bioinformatics

The commands used can be found in the Supplementary Materials and Methods section.

To retrieve the nucleotide sequences from raw signal data generated by the MinKNOW software, we used Albacore 1.2.5 [31] for base calling and demultiplexing of the ONT barcodes (Albacore, RRID:SCR_015897). The FAST5 files were then converted to fastq files using Nanopolish [7, 32]. We then filtered the raw reads for quality (score of >13) and read length (> 200bp) using Nanofilt [33], and generated consensus sequences using both reference-based mapping and *de novo* assembly. For the reference-based mapping, we used BWA 0.7.15 (BWA, RRID:SCR_010910) [8, 34] to align the reads to the reference, samtools 1.3 (SAMTOOLS, RRID:SCR_002105) [9] to process the mapping file, and ANGSD [10] to call the consensus sequence. The *de novo* assembly of each amplicon was carried out using Canu (Canu, RRID:SCR_015880) [11, 35], with parameters fitting for our application. Given that we used short amplicons for the assembly, we set the minimum read length to 200 bpand

Figure 1: Process of nanopore sequencing in the Ecuadorian Chocó rainforest. (A) Sampling endemic fauna; eyelash pitviper next to MinION. (B) Extraction of blood or tissue samples. (C) DNA extraction using the DNeasy kit and benchtop centrifuge, and PCR amplification with the MiniPCR. (D) Oxford nanopore library preparation of DNA barcodes. (E) Bioinformatic processing of nanopore data in the field. (F) Primary equipment used in portable sequencing, left to right: MiniPCR sitting atop Poweradd external battery, MinION plugged into a Windows laptop displaying Geneious Pro software of raw nanopore data.

the minimum overlap to 50 bp. We subsequently extracted the consensus sequences using tgStoreDump. After the consensus calling (for both methods), we mapped the reads back to the consensus sequence (using BWA mem and samtools as described above) and polished the sequencing using Nanopolish (Nanopolish, RRID:SCR_016157) [7]. Adapters were removed using Cutadapt (cutadapt, RRID:SCR_011841) [12]. The consensi were then aligned to the Sanger sequences of the same amplicons to investigate the quality of the consensus sequences generated from MinION reads using SeaView (SeaView, RRID:SCR_015059) [13] and AliView (AliView, RRID:SCR_002780) [14]. Sanger

sequencing reads were edited and assembled using Geneious R10 software (Geneious, RRID:SCR_010519) [15]; mapping files were inspected by eye using Tablet [17].

We further tested the impact of coverage on the consensus accuracy by randomly subsampling three sets of 30, 100, 300, and 1,000 reads, respectively, for the eyelash palm pitviper and gecko 1. Subsampling was performed with famas [36]. These sets were assembled *de novo* and processed using the same approach that was used for the full datasets (see above).

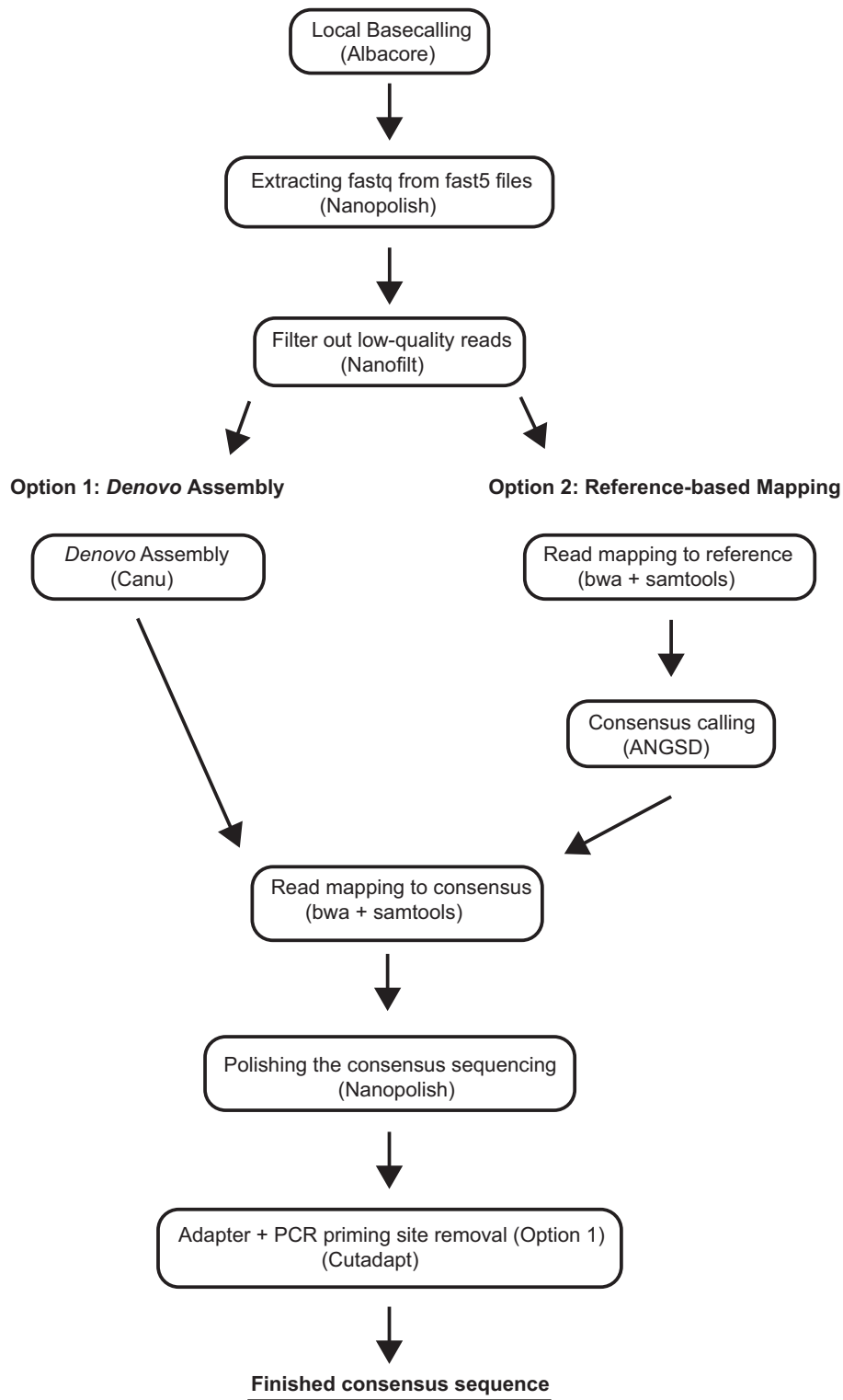We then created species alignments for all barcodes (using sequences obtained from GenBank; accession numbers can be

Figure 2: Bioinformatics workflow summarizing the steps performed during nanopore sequencing analysis with either a *de novo* approach (left) or reference-based mapping approach (right) in order to generate a consensus sequences.

found in the phylogenetic tree reconstructions in the Supplementary Material). We inferred the best substitution model using jModelTest (jModelTest, RRID:SCR_015244) [18] and reconstructed their phylogenetic trees using the maximum likelihood approach implemented in Mega 5 [19] with 1,000 bootstrap replicates (for bioinformatics workflow see Fig.2). The output tree files, including the accession numbers, are provided in the Supplementary Material.
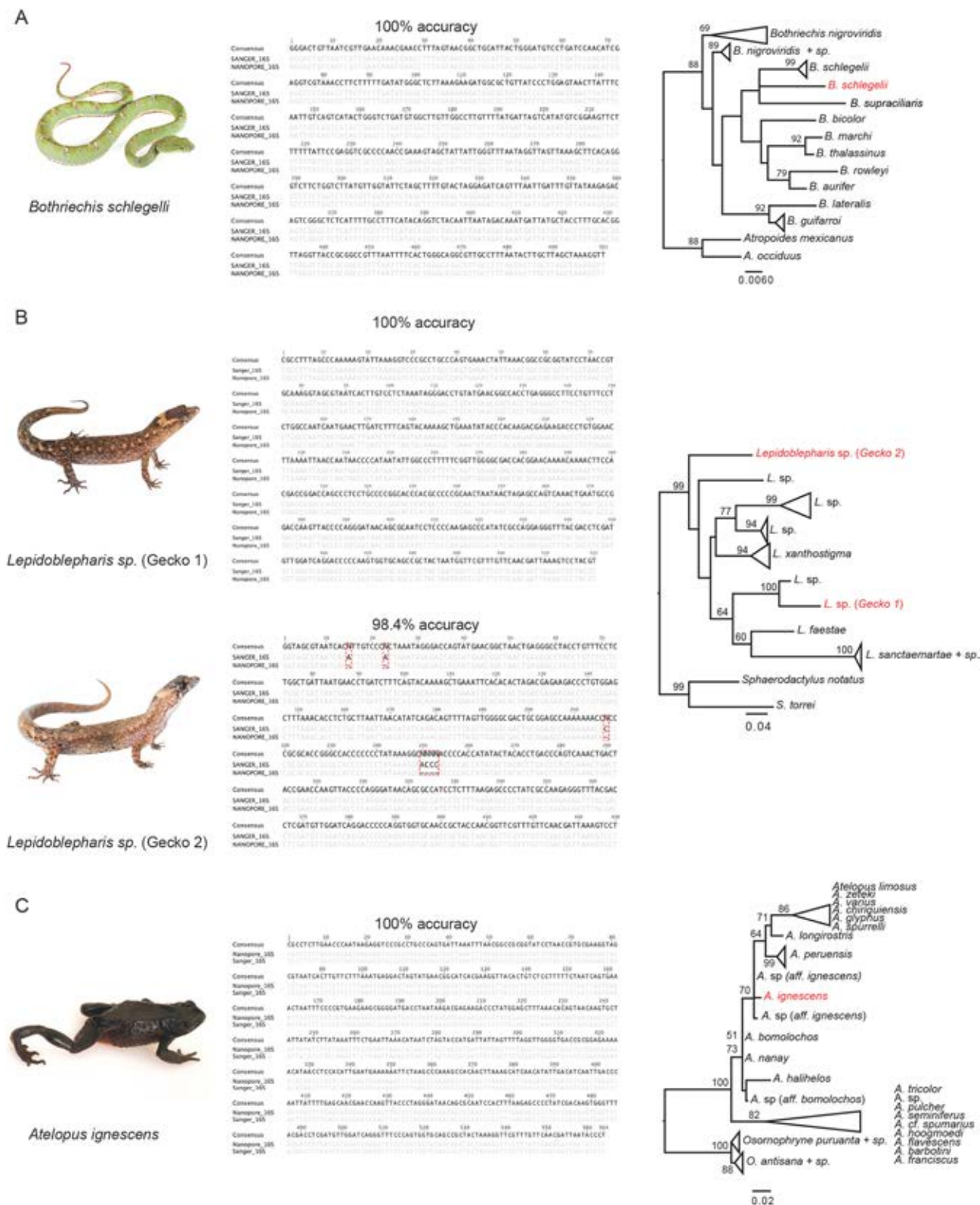
Figure 3: Species investigated, nucleotide alignments of nanopore and Sanger sequences comparing consensus accuracy, and maximum likelihood trees of 16S sequences for (A) eyelash pitviper, *Bothriechis schlegelii*, (B) two species of dwarf gecko, *Lepidoblepharis* sp., and (C) the Jambato toad, *Atelopus ignescens*. Red labels in the phylogenetic trees indicate the sequences generated by the MinION.
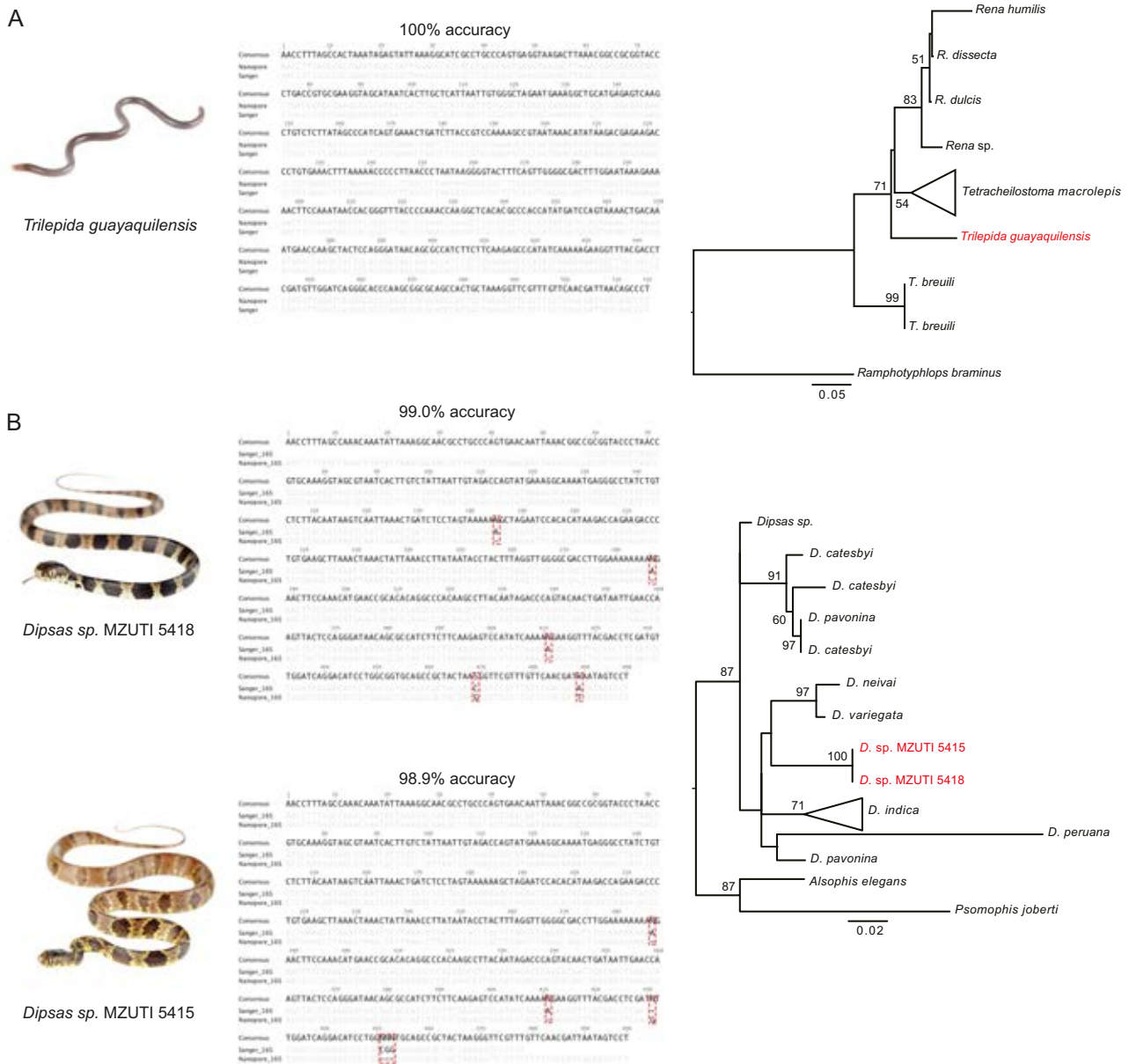
Figure 4: Species investigated, nucleotide alignments of nanopore and Sanger sequences comparing consensus accuracy, and maximum likelihood trees of 16S sequences for (A) Guayaquil blind snake, *Trilepida guayaquilensis*, and (B) two species of *Dipsas* snakes. Red labels in the phylogenetic trees indicate the sequences generated by the MinION.

Results

On 11 July 2017, we arrived at the field site at approximately 1500 hours and collected reptile and amphibian samples from 2000 to 2300 hours. Next, back at the field station, we extracted DNA and performed PCR amplification for 16S, CytB, and ND4 genes. On 12 July, the PCR barcodes were pooled, the library was prepared, and sequencing was initiated at approximately 1600 hours on a flow cell using the offline MinKNOW software, generating 16,663 reads after approximately 2 hours. The MinKNOW software was then paused in order to assess the reads generated. Within 24 hours of collecting reptiles and amphibians in the Ecuadorian Chocó, we successfully generated consensus sequences for 16S and ND4 genes of an eyelash palm pitviper (*Bothriechis schlegelii*) and 16S for the dwarf gecko (*Lepidoblepharis* sp.; gecko 1). The CytB gene was not successfully sequenced, which

was later confirmed at UTI's lab by lack of PCR product on a gel (Supplementary Table S3, Supplementary Fig. S4). The field-generated sequence data were analyzed that evening on a laptop using a number of open source and custom-developed bioinformatic workflows (see Materials and Methods section). Phylogenetic trees generated using the nanopore sequences and the previously generated reference database yielded accurate species identification (Figs. 3 and 4).

Upon return to UTI's lab in Quito, we created one additional DNA barcode library with new samples. With our remaining flow cell, we were interested in quickly generating genetic information for additional specimens that were collected during our field expedition (gecko 2), undescribed snake species collected the week before our expedition (Genera: *Dipsas* and *Sibon*), an endangered species that would have been difficult to export out of

the country (Jambato toad), a rare species lacking molecular data (Guayaquil blind snake), and combinations of barcoded samples through multiplexing (for the eyelash palm pitviper and gecko 1).

Initially, this second sequencing run appeared to perform well. However, after using Albacore to demultiplex the reads, we determined the adapter ligation enzyme likely degraded because the output primarily consisted of adapter sequences (Supplementary Fig. S1, Supplementary Table S1). Nevertheless, we were able to generate consensus sequences for 16S of the Jambato toad, the two *Dipsas* species, the dwarf gecko, and the Guayaquil blind snake (Figs. 3 and 4).

The pore count of the flow cells appeared to be unaffected by travel conditions, as indicated by the multiplexer (MUX) scan, an ONT program that performs a quality check by assessing flow cell active pore count. The first run in the field had an initial MUX scan of 478, 357, 177, and 31, for a total of 1,043 active pores; after approximately 2 hours of sequencing, the flow cell generated 16,484 reads. The second flow cell ran at UTI had a MUX scan of 508, 448, 277, and 84, for total of 1,317 active pores; the run produced 21,636 reads within 2 hours. This is notable since this run was performed 8 days after arriving in Ecuador and the flow cell had been stored at suboptimal conditions on site and during travel. The presence or absence of PCR product and size was later determined by gel electrophoresis and quantified using a Quantus Fluorometer (Promega) at UTI. Amplification for 16S and ND4 was successful for all samples. However, amplification of CytB was unsuccessful, perhaps due to suboptimal PCR settings, as samples were run concurrently due to the limitation and time constraint of having only one miniPCR machine available (Supplementary Fig. S4). While the ONT protocol calls for equimolar ratios of pooled PCR product, we did not have an accurate way of quantifying DNA in the field and, as such, had an overrepresentation of 16S sequences, likely due to PCR bias. On future field expeditions, an inexpensive device such as the bluegel DNA electrophoresis (produced by miniPCR) can be used to assess DNA and PCR products.

Sequencing and bioinformatics

Eyelash Palm Pitviper (Bothriechis schlegelii)
The eyelash palm pitviper (*B. schlegelii*) is an iconic venomous pitviper species found in mesic forests of central and northwestern South America [3]. One individual was captured on the evening of the 11July 2017 and sequenced on the MinION the following evening. We obtained 3,696 reads for the 16S fragment, 65 reads for CytB, and 94 for ND4. The 16S reads showed an average length of 655 bpincluding the sequencing adapters. The best contig created by Canu was based on 55 reads, to which 3,695 reads mapped for the polishing step. The consensus sequence was 501 bpand showed a 100% nucleotide match to the respective Sanger sequence. For this species, we did not find any differences between the *de novo* and the reference-based mapping consensus sequences (generated by mapping against a reference from the same species). The individual clusters with all other *B. schlegelii* and *B. supraciliaris* (considered by some authors to be conspecific with *B. schlegelii*) sequences in the phylogenetic tree (Fig. 3A). While the CytB *de novo* assembly did not succeed (no two reads assembled together), the best supported contig for ND4 (864bp) was based on 50 sequences and achieved an accuracy of 99.4% after polishing (using 95 reads that mapped to the *de novo* consensus).

Dwarf Geckos (Genus: Lepidoblepharis)
Dwarf geckos (genus: *Lepidoblepharis*) are small-bodied leaf litter geckos found in Central and South America. Dwarf geckos can be difficult to identify in the field, and it is suspected that there are several cryptic species within this genus in Ecuador. We captured two individuals on the evening of the 11 July 2017. Because the two geckos differed in the shape and size of the dorsal scales (Fig.3B) and were difficult to confidently identify by morphological characters, we decided to investigate them further with DNA barcoding.

Gecko 1 (Lepidoblepharis aff. grandis) Gecko 1 was included in the first sequencing run in the field. We obtained 4,834 reads for the 16S fragment, 63 reads for CytB, and 76 for ND4. The consensus sequence (522bp) for this individual showed a 100% nucleotide match to the respective Sanger sequence. We then performed reference-based mapping using *L. xanthostigma* (GenBank accession: KP845170) as a reference; the resulting consensus had 99.4% accuracy. We found three insertions compared to the Sanger and *de novo* consensus sequences (position 302: G and 350–351: AA). Next, we attempted assemblies for CytB and ND4. While the assembly for the CytB reads failed, we were able to assemble the ND4 reads. However, the polished consensus sequence showed a relatively high error rate compared to the Sanger sequence (92.1% accuracy). We then blasted all ND4 reads against NCBI. For ND4 we found 8 sequences to blast to ND4 from squamates, 4 to 16S (3 to a viper and 1 to a gecko), 3 to the positive control, 10 very short hits (negligible hits), and 46 to find no blast hit. Interestingly, while only 8 reads were hits for ND4 from squamates, 72 reads mapped to the consensus of the *de novo* assembly. The higher error rate can thus be explained by the fact that contaminant reads were used to assemble and correct consensus. The *de novo* assembled consensus showed an accuracy of 91.7% compared to 92.1% for the polished sequence.

Gecko 2 (Lepidoblepharis aff. buchwaldi) Gecko 2 was included in the second sequencing run at UTI. We generated 325 reads (for more information, see discussion on the possible issue with the adapter ligation enzyme). After filtering for read quality and assembly, we found the best contig to be supported by 30 reads. Of the 325 barcoded reads, we found 308 to map to the consensus. After running Nanopolish, we found it to match 98.4% to the Sanger sequence. All of the observed differences were indels (mostly 1 bp, but also one 4 bp indel; positions: 15, 23, 217, and 250–253, respectively; Fig.3B). Positions 15 and 23 show an A in the reference, which is not found in the nanopore consensus (filtered or unfiltered, and polished or not polished). Position 217 is a C in the Sanger reference. None of the consensi for the nanopore data showed the C. This error can potentially be explained as it lies within a 6 bp cytosine homopolymer. Interestingly, we saw only a 1 bpmismatch instead of the 4 bpindel at position 250–253 in the filtered, but not polished, nanopore consensus sequence. After polishing, all sequences (filtered or unfiltered) showed the 4 bpindel. Next, we applied reference-based mapping (same protocol and reference as for gecko 1). The resulting consensus sequence showed an accuracy of 97.9%. Phylogenetic tree reconstruction showed that gecko 1 and gecko 2 are clearly two distinct species (see Fig.3B).

Jambato toad (Atelopus ignescens)
Laboratory processing and sequencing for *Atelopus ignescens* was carried out in the lab at UTI using a preserved tadpole sample. We obtained 503 reads for this species. The best supported *de novo* assembled contig was based on 56 reads. We then mapped

the reads back to this contig for the polishing step, which resulted in 491 mapped reads. However, while the total coverage was 434x for the segment, the average coverage was only 212x. The discrepancy can be explained by a high percentage of reads that exclusively consisted of adapter sequences (probably caused by inefficient adapter ligation; see Discussion section; Supplementary Fig. S1). The resulting sequence fits 100% to the respective Sanger sequence (Fig. 3C). Next, we used the reference-based approach to construct a consensus sequence, using *Atelopus hoogmoedi* (GenBank accession: EU672974) as a reference; the consensus achieved an accuracy of 100% after polishing. The phylogenetic tree reconstruction clusters our sequence with samples described as *A*. sp. *aff. ignescens*.

Guayaquil blind snake (*Trilepida guayaquilensis*)

The Guayaquil blind snake (*Trilepida guayaquilensis*) belongs to the family of slender blind snakes (Leptotyphlopidae). This family is found in North and South America, Asia, and Africa. They are fossorial snakes adapted to life underground. The Guayaquil blind snake was only known from one individual described in 1970 and is endemic to Ecuador [20]. For a second specimen collected by Jose Vieira on 3 March 2016 at Pacoche, province of Manabi, Ecuador (S1.0677 W80.88169 323m), we obtained 756 sequences. However, many of those reads were adapter sequences. The Canu *de novo* assembled sequence was generated from 16 reads. We then mapped 740 reads back to this consensus. After polishing, the consensus sequence matched 100% of the Sanger-generated sequence (Fig.4A; 516 bpconsensus length). We further investigated the accuracy of reference-based mapping for this species. We used *Trilepida macrolepis* (GenBank accession: GQ469225) as a reference, which is suspected to be a close relative of *T. guayaquilensis*. However, the resulting consensus sequence had a lower accuracy (97.7%) compared to the *de novo* assembled consensus (100%). Our sequence is sister to the clade comprising *Trilepida macrolepis* and all *Rena* species in the phylogenetic tree.

Dipsas snakes (Genus: *Dipsas*)

Dipsas are nonvenomous New World colubrid snakes that are found in Central and South America. Here we included two specimens collected one week prior to our expedition.

Dipsas oreas (MZUTI 5418) We generated 779 reads for *Dipsas oreas* (MZUTI5418). The best supported contig of the Canu *de novo* assembly (498 bpconsensus length) was based on 59 reads and matched the corresponding Sanger sequence to 99% after polishing (Fig.4B). Three of 5 mismatches were indels in poly-A stretches (position: 185, 287, 411). The remaining two mismatches are a C to G at position 469 and a T to A at position 489 for the nanopore compared to the Sanger sequence. Interestingly, the reference-based consensus sequence (using *Dipsas* sp., GenBank accession: KX283341 as a reference) matched the Sanger sequence to 99.4% after polishing. We generated 816 reads for the CytB barcode. However, *de novo* assembly was not successful as only three reads blasted to CytB. However, the lengths of the hits were insignificant. Two sequences blasted to 16S, one blasted to a Dipsadine snake and one blastedto *Atelopus*. One read belonged to the positive control, and 53 showed insignificantly short hits.

Dipsas oreas (MZUTI 5415) We generated 487 reads for *Dipsas* (MZUTI 5415). Sequences with a quality score of >13 were retained, resulting in 193 sequences. The best supported contig of the Canu *de novo* assembly was based on 59 reads (498 bpconsen-

sus length). After polishing, the consensus sequence matched the corresponding Sanger sequence to 98.9% (Fig.4B). The first two mismatches are typical nanopore errors, namely, indels in poly-A stretches (positions: 287, 411). The nanopore sequence shows an insertion of a single G compared to the Sanger sequence as position 431. The last mismatch is a three base-pair deletion compared to the Sanger sequence (positions: 451–453). The reference-based consensus (using *Dipsas* sp., GenBank accession: KX283341 as a reference) achieved a 98.4% match after polishing. We generated 1,077 reads for the CytB barcode. Again, *de novo* assembly was not successful, as only four reads actually belonged to CytB. Four sequences belonged to the positive control, 7 to 16S (four blasted to Colubridae and three to squamates), 1 to a Viperidae microsatellite, and 51 gave insignificantly short hits. The two *Dipsas* specimens clustered together in the phylogeny. They are sister to the clade comprising *D. neivai* and *D. variegata*. However, this part of the phylogeny shows low support (bootstraps <50).

Sibon sp. (Genus: *Sibon*)

Sibon snakes are found in northern South America, Central America, and Mexico [21]. We generated 339 reads for the 16S barcode of this species. However, we were not able to create a consensus sequence for this barcode, as almost all the reads were adapter sequences (all but 11 reads). Furthermore, we generated 1,425 reads for the CytB barcode but were not able to create a consensus sequence.

Subsampling

We further investigated the read depth needed to call accurate consensus sequences using our approach. We used the eyelash palm pitviper and gecko 1 to test subsampling schemes, since we obtained thousands of reads for these samples. We randomly subsampled to 30, 100, 300, and 1,000 reads (in three replicates; see Supplementary Table S4). For the eyelash palm pitviper, we achieved accuracies ranging from 99.4% to 99.8% using only 30 reads, 99.6% to 100% using 100 reads, 99.8% for 300 reads, and 99.8% to 100% for 1,000 reads. For gecko 1, we achieved even better accuracy overall, with 30 reads ranging from 99.4% to 99.8%, 100 reads from 99.8% to 100%, all 300 reads sets achieved an accuracy of 100%,and for 1,000 reads all but one set (99.8%) achieved 100% accuracy.

Multiplexing

We further sequenced multiplexed barcodes (16S and ND4) for the eyelash palm pitviper and gecko 1. However, we did not obtain reads for this sample from sequencing run 2, most likely due to the adapter ligation issues. We thus generated artificial multiplexes for the eyelash palm pitviper, pooling random sets of 1,000 16S reads with all 96 ND4 reads to investigate the performance of the *de novo* assembly using multiplexed samples. We assembled the reads *de novo* and processed them using the same approach as discussed above. In all three cases, we found the first two contigs of the canu run to be 16S and ND4 contigs. After polishing, the 16S consensus sequences achieved a 99.8% accuracy (all three assemblies showed a deletion in a stretch of four Ts compared to the Sanger sequence) and the ND4 sequences a 99.4% accuracy. All errors, but one (which shows a T compared to the C in the Sanger sequence), in ND4 are deletions in homopolymer stretches.

Discussion

Performance in the field

Our objective was to use a portable laboratory in a rainforest to quickly identify endemic species with DNA barcoding. Our protocols resulted in successful DNA extraction, PCR amplification, nanopore sequencing, and barcode assembly, with a turnaround time of less than 24 hours. We observed that the MinION sequencing platform performed well in the field after extended travel, indicating the potential for nanopore-based sequencing on future field expeditions. Although we demonstrate that the successful molecular identification of organisms in a remote tropical environment is possible, challenges with molecular work in the field remain. Although our field site had inconsistent electrical power, we were able to use a conventional small centrifuge for several steps of DNA extraction and to power a refrigerator for storage of flow cells and some of the reagents, although temperatures were likely suboptimal. Lack of electrical supply can impede adequate storage of temperature-sensitive reagents for extended periods of time. Our experiments were performed during a relatively short field trial, with 10 days being the longest time period that reagents were kept at inconsistent freezing temperatures. It is uncertain how well nanopore kit reagents or flow cell integrity would endure over longer periods without consistent cooling temperatures, and we suspect the adapter ligation enzyme was compromised during our second nanopore run, as demultiplexing led to a majority of barcode adapters in each folder (Supplementary Table S3). While the MinION sequencer fits in the palm of a hand and needs only a USB outlet to function, bioinformatic analyses can be hampered under remote field conditions, because Internet access, large amounts of data storage, and long periods of time are often required for such analytical tasks. In our study, using short DNA fragments with a relatively small number of samples for barcoding allowed us to perform all bioinformatic analyses in the field, but larger data outputs may require additional storage and more computational resources.

Implications for conservation and biodiversity assessments

Tropical rainforests, such as the Ecuadorian Chocó, are often rich in biodiversity, as well as species of conservation concern. The Chocó biogeographical region is one of the world's 25 biodiversity hotspots [37], and several studies have identified the Chocó region of western Colombia and Ecuador as a global conservation priority [37–39]. We therefore chose this region for proof-of-principle *in situ* molecular work to highlight the importance of expediting field work in order to produce genetic information of endemic fauna. Our rapidly obtained DNA barcodes allowed us to accurately identify organisms while in the field. When samples are not required to be exported out of the country to carry out molecular experiments, real-time sequencing information can contribute to more efficient production of biodiversity reports that advise conservation policy, especially in areas of high conservation risk.

Of particular note in this study was the critically endangered harlequin Jambato toad, *Atelopus ignescens*. Although not a denizen of the Chocó rainforests, this Andean toad is a good example to demonstrate how nanopore sequencing can aid in the conservation of critically endangered species. *Atelopus ignescens* was previously presumed extinct (it is currently still listed as "extinct" on the International Union for Conservation of Nature (IUCN) [40]) and was only recently rediscovered [41]. The last confirmed record of *A. ignescens* dates back to 1988, and

this species was presumed to be extinct before one population was rediscovered in 2016, 28 years later. *Atelopus* is a species-rich genus of neotropical toads that contains 96 species, most of which are possibly extinct or endangered. In Ecuador there are 11 species of *Atelopus* that are critically endangered (tagged as possibly extinct [42]). Extinctions of *Atelopus* (and other anurans) are beyond control and are increasingly exacerbated by a combination of factors, including habitat loss, climate change, and pathogens [43–45]. For the many endangered species that are protected by international laws and treaties, sample transport requires permits that can often be difficult to obtain, even when research is expressly aimed at conservation, resulting in project delays that can further compromise sample quality. By working within the country, under permits issued by Ministerio del Ambiente de Ecuador to local institutions, we were able to generate sequence data for the endangered harlequin Jambato toad *A. ignescens* within 24 hours of receiving the tissue, whereas obtaining permits to ship samples internationally in the same time frame would have not been possible. Rapidly identifying the phylogenetic affinity of populations of *Atelopus* toads could speed up conservation efforts for these animals. Namely, a better understanding of the systematics of the group facilitated by real-time sequencing could help establish species limits, geographic distributions, *in situ* conservation actions, and *ex situ* breeding programs.

Species identifications

It is important to note that we do not intend for rapidly obtained portable sequence information to substitute for standard species description processes. Instead, we aim to demonstrate that obtaining real-time genetic information can have beneficial applications for biologists in the field, such as raising the interesting possibility of promptly identifying new candidate species, information that can be used to adjust fieldwork strategies or sampling efforts. As we have shown, the latter could be especially important with organisms and habitats that face pressing threat. Rapidly obtaining genetic sequence information in the field can also be useful for a range of other applications, including identifying cryptic species, hybrid zones, immature stages, and species complexes.

Furthermore, we acknowledge that in most cases multiple loci are needed to reliably infer species position in a phylogenetic tree. DNA barcoding has been shown to hold promise for identification purposes in taxonomically well-sampled clades but may have limitations or pitfalls in delineating closely related species or in taxonomically understudied groups [46, 47]. However, our aim in this study was to demonstrate that portable sequencing can be used in the field and that the final sequences have an accuracy needed to achieve reliable identification of a specimen. While a recent study has demonstrated a field-based shotgun genome approach with the MinION to identify closely related plant species [27], DNA barcoding already offers a robust reference database for many taxa thanks in part to global barcoding initiatives (the current Barcode of Life Data System contains 4,013,927 specimens and 398,087 Barcode Index Numbers [48] as of September 2017).

Finally, while highlighting the value of real-time portable DNA barcoding in this study, we do not wish to downplay the significance of taxonomic experts who have invaluable specialist knowledge about specific groups of organisms. Even with the advent of molecular diagnostic techniques to describe and discover species, placing organisms within a phylogenetic context based on a solid taxonomic foundation is necessary. An integra-

tive approach that uses molecular data and morphological taxonomy can lead to greater insight of biological and ecological questions [49]. As noted by Bik [49], "There is much to gain and little to lose by deeply integrating morphological taxonomy with high-throughput sequencing and computational workflows."

Bioinformatic challenges

While we were able to show that nanopore sequencing results in high-quality DNA barcode sequences, some challenges during the read processing remain. To our knowledge, no software solution specifically designed to assemble DNA barcodes from long read technologies is available. Here, we created our own pipeline (Supplementary Fig. S2). This required changing the settings for Canu [50], a whole genome *de novo* assembler (see Materials and Methods in the Supplementary Information and discussion below). However, software geared toward the specifics of assembling DNA barcodes from long read data would be beneficial to make the bioinformatics analysis easier and more widely applicable.

We were also interested in investigating the minimum coverage needed to create reliable consensus sequences. Therefore, we used different subsampling schemes. Overall, a coverage of 30 reads achieved an accuracy of 99.4–99.8%. With 100x read coverage, almost all assemblies were 100% accurate, indicating that an excessive number of reads is not needed to produce high-quality consensus sequences. Furthermore, we applied Nanopolish to all consensus sequences. This tool has been shown to be very effective at correcting typical nanopore errors, such as homopolymer errors [51, 52]. As can be seen in the section "post-nanopolish assembly identity" in [52], accuracy of the resulting consensus increases significantly after polishing. While we did not measure the improvement in accuracy in our study, we did notice a high accuracy after polishing. However, as can be seen in Fig.4B, nanopolish is not always able to accurately correct homopolymer stretches.

We further tested reference-based mapping vs. *de novo* assembly because a reference-based mapping approach may introduce bias, making it possible to miss indels. Overall, we saw that consensus sequences generated using reference-based mapping had slightly lower accuracy. However, in two cases (the eyelash palm pitviper and the Jambato toad), an accuracy of 100% was achieved with reference-based mapping. Interestingly, in the case of *Dipsas* sp. (MZUTI 5418), reference-based mapping resulted in a slightly better accuracy than the *de novo* approach (99.4% compared to 99%). However, in general, we recommend the use of a *de novo* assembly approach as this method can be applied even if no reference sequence is available and generally produced more accurate sequences. An alternative approach would be to generate consensus sequences by aligning the individual reads for each barcode to one another, which would not be affected by a reference bias. This method is implemented in the freely available software tool Allele Wrangler [53]. However, at the time of submission, this tool picked the first read as the pseudo reference, which can lead to errors in the consensus if this read is of particularly low quality or an incorrect (contaminant) sequence. Future developments might establish this method as an alternative to *de novo* assembly algorithms, which are typically written for larger genomes (e.g., the minimum genome size in Canu is 1,000 bp) and can have issues with assemblies where the consensus sequence is roughly the size of the input reads (personal communications, Adam Phillippy).

Each of our two runs showed a very high number of reads not assigned to any barcode sequence after demultiplexing with Al-

bacore 1.2.5 (7,780 and 14,272 for the first and second sequencing run, respectively). In order to investigate whether these reads belong to the target DNA barcodes but did not get assigned to sequencing barcodes or if they constitute other sequences, we generated two references (one for each sequencing run) comprising all consensi found within each individual sequencing run. We then mapped all reads not assigned to barcodes back to the reference. We were able to map 2,874 and 4,997 reads to the reference for the first and the second sequencing run, respectively, which shows that a high number of reads might be usable if more efficient demultiplexing algorithms become available. Here, we used Albacore 1.2.5, an ONT software tool, to demultiplex the sequencing barcodes. This tool is under constant development and thus might offer more efficient demultiplexing in later versions. Alternatively, third-party software tools like npBarcode [54] or Porechop [55] can be used.

Cost-effectiveness and local resource development

Next-generation sequencing technologies are constantly evolving, along with their associated costs. Most major next-generation sequencing platforms require considerable initial investment in the sequencers themselves, costing hundreds of thousands of dollars, which is why they are often consolidated to sequencing centers at the institutional level [56]. In this study, we used the ONT starter pack, which currently costs $1,000and includes two flow cells and a library preparation kit (six library preparations), as well as the ONT 12 barcoding kit, which is currently $250 for six library preparations (for a full list of equipment and additional reagents, see Supplementary Table S1). Using this setup, each barcode amplicon sequence generated costs of approximately $45 (this includes cost for the starter pack, etc.; a detailed cost account can be found in the Supplementary Material). At this cost, further multiplexing of samples on each flow cell is necessary to achieve a cost-effectiveness for DNA sequencing relative to other commercial options. However, it will likely not be long until much higher multiplexing (>500 samples) becomes achievable on the MinION platform, which would pave the way for MinION-based DNA barcode costs to be reduced to less than $1, similar to advancements achieved with Illumina and PacBio-based pipelines (see [57–59]). On the contrary, Sanger sequencing from UTI in Ecuador shipped internationally for processing costs of approximately $10 per sample, independent of the through-put. Thus, the Oxford Nanopore MinION has the potential to be a cost-effective sequencing option for resource-limited labs, especially in developing countries without access to standard sequencing devices.

The small size and low power requirements of the MinION will likely continue to enable its evolution as a field-deployable DNA sequencing device, opening up new avenues for biological research in areas where the typical laboratory infrastructure for genetic sequencing is unavailable. With some training, in-the-field molecular analyses could also potentially be performed by students (see [60]) or assistants, providing an opportunity for local teaching and research capacity building, as well as community involvement via research-focused ecotourism or citizen-science projects.

Future outlook

Technological developments in lab equipment and reagent chemistry are increasingly enabling the incorporation of genetic analyses into field projects. Several portable technologies have been used to perform molecular experiments in the field, par-

ticularly for disease diagnostics [61, 62]. Advances in lyophilized and room-temperature reagents are also promising for field applications, such as EZ PCR Master Mix [63] and loop-mediated isothermal amplification [64, 65]. A hand-powered centrifuge [66] could also act as a substitute for a standard benchtop centrifuge during DNA extraction steps. Automatic devices, such as VolTRAX (a compact microfluidic device designed to automate nanopore library preparation; ONT) and improved library construction methods may offer faster and high-throughput methods for preparing nanopore libraries in the future. As the ONT MinION evolves, it could greatly advance field researchers' capacity to obtain genetic data from wild organisms while in the field. These technologies currently depend on reagents that require freezing but can be used at field sites with solar or portable freezer options. Faster and more automated sample processing, as well as cost reductions, are needed for adoption in low-income settings.

Beyond short PCR-based amplicons aimed at species identification, other exciting potential applications of nanopore sequencing in the field include sequencing of entire mitochondria from gDNA samples [67] or via long-range PCR, shotgun genome sequencing [27], analysis of environmental DNA [68, 25], sequencing of direct RNA [69, 70] or cDNA to rapidly profile transcriptomes ([71], and pathogen diagnostics and monitoring (such as chytrid fungus [72]). Rapid portable sequencing can also be applied to wildlife crime in order to perform species identification of animals affected by illegal trafficking, as well as serve to aid in early detection of invasive species that threaten local biodiversity and agriculture and emerging infectious diseases.

Potential implications

While we live in a period of amazing technological change, biodiversity and ecosystem health are decreasing worldwide. Portable sequencing will not be a silver bullet for conservation biology but it can be a powerful tool to more efficiently obtain information about the diversity of life on our planet. This is particularly important for many biodiversity hotspots, such as tropical rainforests like the Ecuadorian Chocó, which are often under high risk of habitat loss. Here, we show that portable DNA barcoding with the MinION sequencer allows rapid, accurate, and efficient determination at the species level under remote and tropical environmental conditions. We also demonstrate that portable sequencing can allow nimble use of rapidly generating data for endangered, rare, and undescribed species at nearby facilities within the country. As portable technologies develop further, this method has the potential to broaden the utility of biological field analyses, including real-time species identification, cryptic species discovery, biodiversity conservation reports, pathogen detection, and environmental studies.

Availability of supporting data

Raw sequencing data are available in the SRA via bioproject number PRJNA438544. Other supporting data are available in the *GigaScience* GigaDB repository [73].

Additional file

Supplementary Figure S1: Large portion of adapter sequences contained in demultiplexed barcodes, indicating possible adapter ligation degradation for the second nanopore run at UTI.

Supplementary Figure S2: Bioinformatics workflow summarizing the steps performed during nanopore sequencing analysis with either a *de novo* approach (left) or reference-based mapping approach (right), in order to generate a consensus sequences.

Supplementary Figure S3: Additional images highlighting the portable lab equipment and setup used for nanopore sequencing in Ecuador. (A) The handheld MinION DNA sequencer (Oxford Nanopore Technologies). (B) miniPCR Thermocycler (miniPCR). (C) Mobile setup for DNA extraction and PCR amplification. (D) Loading the ONT flow cell in the field. (E) Running the MinION using offline MinKNOW software. (F) Local collaborator loading the MinION at a nearby research facility, highlighting the opportunity for capacity building and community involvement.

Supplementary Figure S4: Gel of PCR product that was produced in the field using the miniPCR and imaged at UTI in Quito. Note that 16S and ND4 from samples amplified but CytB did not.

Supplementary Table S1: List of equipment, consumables, and reagents used for portable nanopore sequencing in Ecuador.

Supplementary Table S2: Primers used in this study. Bold nucleotides indicate the universal tailed ONT sequences for the barcoding kit.

Supplemental Table S3: Summary of de-multiplexed reads.

Supplementary Table S3: Subsampling report: impact of coverage on the consensus accuracy by randomly subsampling.

Abbreviations

ONT: Oxford Nanopore Technologies; PCR: polymerase chain reaction; UTI: Universidad Tecnológica Indoamérica.

Competing interests

The authors declare that they have no competing interests.

Author contributions

Acknowledgements

References

1. Erwin TL. Tropical forests their richness in coleoptera and other arthropod species. The Coleop Bull 1982;**36**:74–75.
2. Stork NE. How many species are there? Biodivers Conserv 1993;**2**(3):215–32.
3. Scheffers BR, Joppa LN, Pimm SL, et al. What we know and don't know about Earth's missing biodiversity. Trends in Ecology & Evolution 2012;**27**(9):501–10.
4. Sierra R, Campos F, Chamberlin J. Assessing biodiversity conservation priorities: ecosystem risk and representativeness in continental Ecuador. Landscape and Urban Planning 2002;**59**(2):95–110.
5. Cuesta F, Peralvo M, Merino-Viteri A, et al. Priority areas for biodiversity conservation in mainland Ecuador. Neotropical Biodiversity 2017;**3**(1):93–106.
6. Gascon C, Brooks TM, Contreras-MacBeath T, et al. The importance and benefits of species. Current Biology 2015;**25**(10):R431–8.
7. Pimm SL, Jenkins CN, Abell R, et al. The biodiversity of species and their rates of extinction, distribution, and protection. Science 2014;**344**(6187):1246752.
8. Ceballos G, Ehrlich PR, Dirzo R. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci U S A 2017;**114**:E6089–96.
9. Coloma LA, Duellman WE, Almendariz A, et al. Five new (extinct?) species of atelopus (Anura: Bufonidae) from Andean Colombia, Ecuador, and Peru. Zootaxa 2010;**2574**:1–54.
10. Lötters S, Van der Meijden A, Coloma LA, et al. Assessing the molecular phylogeny of a near extinct group of vertebrates: the neotropical harlequin frogs (Bufonidae; Atelopus). Systematics and Biodiversity 2011;**9**(1):45–57.
11. Barley AJ, White J, Diesmos AC, et al. The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. Evolution 2013;**67**(12):3556–72.
12. Carstens BC, Pelletier TA, Reid NM, et al. How to fail at species delimitation. Mol Ecol 2013;**22**(17):4369–83.
13. Hebert PD, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 2003;**270**(1512):313–21.
14. Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. Syst Biol 2005;**54**(5):852–9.
15. Savolainen V, Cowan RS, Vogler AP, et al. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 2005;**360**(1462):1805–11.
16. www.barcodeoflife.org.
17. Gilbert N. Biodiversity law could stymie research. Nature 2010;**463**(7281):598.
18. Fernández F. The greatest impediment to the study of biodiversity in Colombia. Caldasia 2011;**33**:3–5.
19. Laver T, Harrison J, O'Neill PA, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular Detection and Quantification 2015;**3**:1–8.
20. Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 2014;**14**(6):1097–102.
21. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature 2016;**530**(7589):228–32.
22. Faria NR, Sabino EC, Nunes MR, et al. Mobile real-time surveillance of Zika virus in Brazil. Genome Med 2016;**8**(1):97.
23. Faria NR, Quick J, Claro IM, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 2017;**546**(7658):406–10.
24. Edwards A, Debbonaire AR, Sattler B, et al. Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78°N. bioRxiv 2016 073965; https://doi.org/10.1101/073965.
25. Johnson SS, Zaikova E, Goerlitz DS, et al. Real-time DNA sequencing in the Antarctic dry valleys using the Oxford Nanopore sequencer. J Biomol Tech 2017;**28**:2–7.
26. Menegon M, Cantaloni C, Rodriguez-Prieto A, Centomo C, Abdelfattah A, Rossato M, et al. On site DNA barcoding by nanopore sequencing. PLoS One 2017;(10): e0184741 https://doi.org/10.1371/journal.pone.0184741.
27. Parker J, Helmstetter AJ, Devey D, et al. Field-based species identification of closely-related plants using real-time nanopore sequencing. Sci Rep 2017;**7**(1):8345.
28. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA sequencing and genome assembly on the international space station. Sci Rep 2017;**7**(1):18022.
29. Janzen DH, Hajibabaei M, Burns JM, et al. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 2005;**360**(1462):1835–45.
30. http://www.fjocotoco.org/canandeacute1.html.
31. Albacore GitHub Repository , https://github.com/dvera/albacore.
32. Nanopolish GitHub Repository, https://github.com/jts/nanopolish.
33. Nanofilt GitHub Repository Nanofilt GitHub Repository, https://github.com/wdecoster/nanofilt.
34. BWA GitHub Repository https://github.com/lh3/bwa/releases.
35. Canu readthedocs documentation, https://canu.readthedocs.io.
36. Famas GitHub Repository, https://github.com/andreas-wilm/famas.
37. Myers N, Mittermeier RA, Mittermeier CG, et al. Biodiversity hotspots for conservation priorities. Nature 2000;**403**(6772):853–8.
38. Dinerstein E. A Conservation Assessment of the Terrestrial Ecoregions of Latin America and the Caribbean. Washington, D.C.: World Bank. 1995;**xvii**:129.
39. Olson DM, Dinerstein E. The Global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. Conservation Biology 1998;**12**(3):502–15.
40. Ron S, Coloma LA, Lötters S, Duellman W, et al. Bustamante, wilmar bolívar, enrique la marca. 2004 *Atelopus ignescens*. The IUCN Red List of Threatened Species 2004: e.T54518A11157432.
41. Coloma LA. El Jambato negro del páramo, Atelopus ignescens, resucitó. 2016 IMCiencia. otonga.org/?p=438
42. Tapia EE, Coloma LA, Pazmiño-Otamendi G, et al. Rediscovery of the nearly extinct longnose harlequin frog *Atelopus longirostris* (Bufonidae) in Junín, Imbabura, Ecuador. Neotropical Biodiversity 2017;**3**(1):157–67.
43. Kiesecker JM. Global stressors and the global decline of amphibians: tipping the stress immunocompetency axis. Ecol Res 2011;**26**(5):897–908.

44. Pounds JA, Coloma LA. Beware the lone killer. Nature Reports Climate Change 2008;(0804):57–59.

45. Lips KR, Diffendorfer J, Mendelson JR, III, et al. Riding the wave: reconciling the roles of disease and climate change in amphibian declines. PLoS Biol 2008;**6**(3):e72.

46. Elias M, Hill RI, Willmott KR, et al. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proceedings of the Royal Society B: Biological Sciences 2007;**274**(1627):2881–9.

47. Meyer CP, Paulay G. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 2005;**3**(12):e422.

48. iBOL Barcode Library http://ibol.org/resources/barcode-library/.

49. Bik HM. Let's rise up to unite taxonomy and technology. PLoS Biol 2017;**15**(8):e2002231.

50. Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;**27**(5):722–36.

51. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods 2015;**12**(8):733–5.

52. Wick RR, Judd LM, Gorrie CL, et al. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 2017;**3**(10):e000132.

53. Allel Wrangler GitHub Repository https://github.com/transplantation-immunology/allele-wrangler/.

54. Nguyen SH, Duarte TPS, Coin LJM, et al. Real-time demultiplexing nanopore barcoded sequencing data with npBarcode. Bioinformatics 2017;**33**(24):3988–90.

55. Porechop GitHub Repository https://github.com/rrwick/Porechop.

56. Glenn TC. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 2011;**11**(5):759–69.

57. Meier R, Wong W, Srivathsan A, et al. $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. Cladistics 2016;**32**(1):100–10.

58. Liu S, Yang C, Zhou C, et al. Filling reference gaps via assembling DNA barcodes using high-throughput sequencing - moving toward barcoding the world. GigaScience 2017;6(12):1–8.

59. Hebert PDN, Brakmann TWA, Prosser SWJ, et al. A sequel to sanger: amplicon sequencing that scales. bioRxiv 191619; doi: https://doi.org/10.1101/191619.

60. Zeng Y, Martin CH. Oxford nanopore sequencing in a research-based undergraduate course. bioRxiv 2017;227439.

61. Canier L, Khim N, Kim S, et al. An innovative tool for moving malaria PCR detection of parasite reservoir into the field. Malar J 2013;**12**(1):405.

62. Marx V. PCR heads into the field. Nat Methods 2015;**12**(5):393–7.

63. Guevara EE, Frankel DC, Ranaivonasy J, et al. A simple, economical protocol for DNA extraction and amplification where there is no lab. Conservation Genetics Resources 2017;1–7. doi:10.1007/s12686-017-0758-5

64. Centeno-Cuadros A, Abbasi I, Nathan R. Sex determination in the wild: a field application of loop-mediated isothermal amplification successfully determines sex across three raptor species. Mol Ecol Resour 2017;**17**(2):153–60.

65. Howson ELA, Armson B, Madi M, et al. Evaluation of two lyophilized molecular assays to rapidly detect foot-and-mouth disease virus directly from clinical samples in field settings. Transbound Emerg Dis 2017;**64**(3):861–71.

66. Bhamla MS, Benson B, Chai C, et al. Hand-powered ultralow-cost paper centrifuge. Nat Biomed Eng 2017;**1**(1):0009.

67. Chandler J, Camberis M, Bouchery T, et al. Annotated mitochondrial genome with nanopore R9 signal for *Nippostrongylus brasiliensis* [version 1; referees: 1 approved, 2 approved with reservation]. F1000Res 2017;**6**:56.

68. Rees HC, Maddison BC, Middleditch DJ, et al. REVIEW: the detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J Appl Ecol 2014;**51**(5):1450–9.

69. Ayub M, Hardwick SW, Luisi BF, et al. Nanopore-based identification of individual nucleotides for direct RNA sequencing. Nano Lett 2013;**13**(12):6144–50.

70. Garalde DR, Snell EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Meth 2018;**15**(3):201–6.

71. Hargreaves AD, Mulley JF. Assessing the utility of the Oxford nanopore MinION for snake venom gland cDNA sequencing. Peer J 2015;**3**:e1441.

72. Weldon C, du Preez LH, Hyatt AD, et al. Origin of the amphibian chytrid fungus. Emerg Infect Dis 2004;**10**:2100–5.

73. Pomerantz A, Peñafiel N, Arteaga A, et al. Supporting data for "Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building." GigaScience Database 2018, http://dx.doi.org/10.5524/100426.

CrossMark

METHODS AND RESOURCES ARTICLE

A simple, economical protocol for DNA extraction and amplification where there is no lab

Elaine E. Guevara[1,2] · David C. Frankel[1] · Jeannin Ranaivonasy[3] ·
Alison F. Richard[2] · Joelisoa Ratsirarson[3] · Richard R. Lawler[4] · Brenda J. Bradley[1]

Abstract Genetic analyses are well suited to address many research questions in the study of wild populations, yet species of interest often have distributions that are geographically distant from molecular laboratories, necessitating potentially lengthy transport of biological specimens. Performing basic genetic analyses on site would avoid the project delays and risks of sample quality decline associated with transport, as well as allow original specimens to remain in the country of origin. Further, diagnostic genetic assays performed in the field could provide real-time information allowing for more nimble adjustments to research plans and use of resources. To this end, we developed protocols for reliably performing front-end genetics bench work in the field, without the requirements of electricity or permanent shelter. We validated these protocols on buccal swabs collected during routine capturing of sifaka lemurs (*Propithecus verreauxi*) at Bezà Mahafaly Special Reserve in Southwest Madagascar and faecal samples collected from captive sifakas (*P. coquereli*) at the Duke Lemur Center. Our basic protocol pipeline involves a chelating resin based DNA extraction followed by whole genome amplification or polymerase chain reaction using reagents stored at ambient temperature and portable, compact equipment powered by a lightweight solar panel. We achieved a high success rate (>80%) in downstream procedures, demonstrating the promise of such protocols for performing basic genetic analyses in a broad range of field situations.

Keywords Population genetics · Genotyping · Sex-typing · Field methods · Capacity building · Madagascar

✉ Elaine E. Guevara
elaine.guevara@yale.edu

[1] Department of Anthropology, Center for the Advanced Study of Human Paleobiology, The George Washington University, Science & Engineering Hall, Suite 6000, 800 22nd St. NW, Washington, DC 20052, USA

[2] Department of Anthropology, Yale University, 10 Sachem St., New Haven, CT 06511, USA

[3] Département des Eaux et Forêts, University of Antananarivo, BP 175 ESSA Forêts, 101 Antananarivo, Madagascar

[4] Department of Sociology and Anthropology, James Madison University, MSC 7501 Sheldon Hall, 71 Alumnae Dr., Harrisonburg, VA 22807, USA

Introduction

Genetic analyses are a valuable tool in the study of wild populations and, as a result, have been widely adopted to address questions related to conservation (Taberlet et al. 1997; Goossens et al. 2005; Kohn et al. 2006; Bergl and Vigilant 2008; Allendorf and Luikart 2009; Quéméré et al. 2010; Gray et al. 2014; Caragiulo et al. 2015), demographic and population history (Thalmann et al. 2011; Ruegg et al. 2013), social and reproductive organization (Griffin et al. 2003; Castro et al. 2004; Bradley et al. 2005; Alberts et al. 2006), pathogen load (Pallen 2014; Wedrowicz et al. 2016), and phenotype status of ecologically relevant traits like color vision phenotypes (Jacobs et al. 2017) or immune system profile (Schwensow et al. 2008). However, most genetic assays are performed in specialized molecular laboratories that are often distant from the species' natural geographic distributions. Thus, the biological samples from which genetic material is derived frequently need to be transported internationally. There are several disadvantages to this situation. First, without reliable refrigeration

or freezer systems, it can be challenging to adequately store specimens to avoid degradation at field sites or during transport. Further, for the many endangered species that are protected by international laws and treaties, transport requires permits that can often be difficult to obtain, even when research is expressly aimed at conservation, resulting in project delays that can further compromise sample quality. The removal of samples from the country of origin also generally precludes the involvement of local collaborators and students in the genetic components of research projects. Finally, prolonged periods between sample collection and dissemination of results can hinder conservation efforts, such as surveys of elusive species. In these and similar cases, real-time and in situ information provided by diagnostic genetic assays (e.g., species identity, sex-typing) could be highly beneficial, obviating travel between the lab and the field and potentially allowing for the active direction of expeditions while "on the ground."

To address these issues, we developed and tested protocols for basic genetics laboratory work under field conditions, without access to temperature control, an electrical grid, or permanent shelter.

Specifically, we used portable equipment that was powered by a lightweight, fold-up solar panel, along with reagents kept at ambient temperatures, to test a DNA extraction protocol. Extractions were then followed by either whole genome amplification or diagnostic sex-typing polymerase chain reaction (PCR) experiments. We tested these protocols at a remote field site [Bezà Mahafaly Special Reserve (BMSR)] in Madagascar and at a naturalistic captive site [Duke Lemur Center (DLC)] in North Carolina using buccal swab and faecal samples collected from sifaka lemurs (genus *Propithecus*). We obtained a high level of success, suggesting these protocols can yield rapid results and facilitate international sharing of samples without the need for establishing costly on-site laboratories or freezer storage.

## Materials and methods

Field sites and sample collection

Remote field site: BMSR encompasses 4600 hectares of dry forest in Southwest Madagascar and is home to four species of lemur. The field site is 5 h or more from Toliara, the nearest major city, by car, depending on the conditions of the roads, which vary seasonally. The BMSR sifaka (*Propithecus verreauxi*) population has been the subject of continuous research since 1984 (Richard et al. 2002; Sussman et al. 2012), but the reserve is off the electrical grid and there is no freezer storage. This long-term project includes the annual capturing and marking

of unmarked animals. Animals are demobilized using a tranquilizer dart launched from a Dan-Inject blowgun. Typically, unmarked yearlings and immigrants are captured, and a range of qualitative and quantitative measurements are taken from each anaesthetized animal. The animals are also given collars bearing an ID number prior to their release (Richard et al. 2002). We collected buccal cells using foam-tipped swabs from individuals while anaesthetized during the August capture season. August corresponds to the austral winter and the dry season in this region of Madagascar. Sunshine was plentiful, with around 9 h of direct sunlight reaching the solar panel per day. One DNA extraction per buccal swab was performed within an hour of collection.

Naturalistic captive site: We also tested our protocol on faecal samples, which we collected from captive animals at the DLC in North Carolina, a research site that partially mimics naturalistic habitat and conditions (Zehr et al. 2014). Faecal samples were collected from a closely related species of sifaka lemur (*Propithecus coquereli*) housed at the DLC. Samples were collected in the morning shortly after the sifakas awake and become active. Fresh faeces were collected immediately upon deposit from the floors of enclosures, which included both interior concrete floors (cleaned daily), and from the forest floor of natural habitat enclosures. Faecal pellets were gently swabbed with a foam-tipped swab dipped in sterile DNase-free water. Swabs were then returned to the sterile wrapper and placed in a small plastic bag stored at ambient indoor temperature (~25 °C) until extraction.

Equipment and set-up

See Fig. 1 for a schematic of the equipment set-up. All heat block and temperature cycling steps were performed using a miniPCR™ machine, which is a portable thermocycler weighing 0.45 kg. The miniPCR was programmed via an application on an android tablet (Google Nexus, weight: 0.44 kg). The miniPCR and tablet were powered by a solar battery (Apollo Pro 23000mAh battery, weight: 0.64 kg, made by PowerAdd). We implemented gel electrophoresis for the sex-typing assay on-site using a portable gel box with built-in power source and blue light illuminator (blueGel™, weight: 0.34 kg, made by miniPCR). The gel box was powered by a 27,000 mAh battery with an AC outlet (weight: 1.08 kg, made by ChargeTech). Batteries were charged from two fold-up solar panels (a 40 W panel, weight: 1.08 kg, made by PowerAdd, and a 60 W panel, weight: 1.63 kg, made by AllPowers). Standard laboratory pipettes and sample racks were used. The total weight of this equipment was around 10 kg and the total cost, excluding pipettes, was just over US$2000.
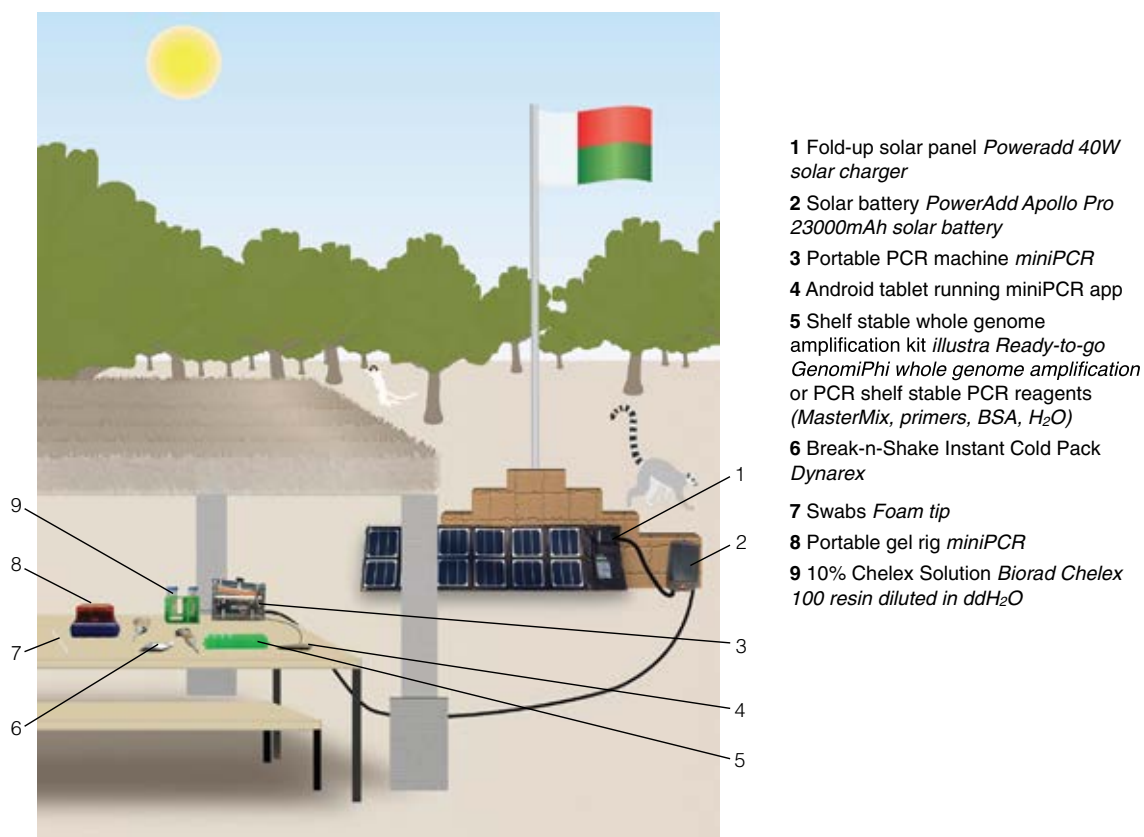
1 Fold-up solar panel *Poweradd 40W solar charger*

2 Solar battery *PowerAdd Apollo Pro 23000mAh solar battery*

3 Portable PCR machine *miniPCR*

4 Android tablet running miniPCR app

5 Shelf stable whole genome amplification kit *illustra Ready-to-go GenomiPhi whole genome amplification* or PCR shelf stable PCR reagents *(MasterMix, primers, BSA, H_2O)*

6 Break-n-Shake Instant Cold Pack *Dynarex*

7 Swabs *Foam tip*

8 Portable gel rig *miniPCR*

9 10% Chelex Solution *Biorad Chelex 100 resin diluted in ddH$_2$O*

Fig. 1 Depiction of equipment and field setting

DNA extraction

Swabs were swirled for a few seconds in 1.5 ml tubes containing approximately 300 µl of Chelex® 100 chelating resin (BioRad) in solution and then discarded. For buccal swabs, a 10% Chelex solution was prepared, and for faecal swabs, a 20% Chelex solution was used to counteract potential inhibition from compounds in the faeces. Chelex solution was then split between at least two 0.2 ml PCR tubes that were subsequently incubated at 95 °C for 10 min in the miniPCR unit. Samples were set to cool for 10 min, after which time, the supernatant was used as template in downstream amplification steps (Fig. 2; see Supplemental File 1 for protocol sheet). Quantification of DNA extractions was possible using a NanoDrop 2000 (Thermo-Fisher Scientific) spectrophotometer at the captive site.

Amplification

Option A: whole genome amplification and downstream PCR

Four whole genome amplifications (WGAs) per individual were performed immediately following DNA extraction
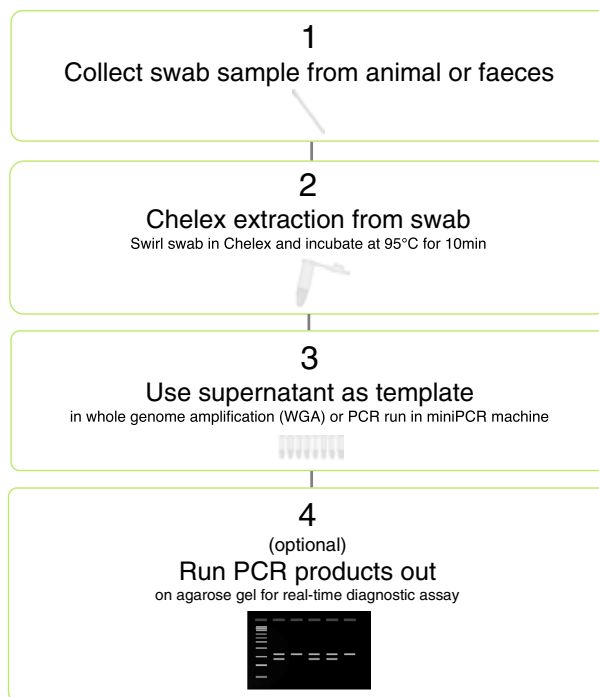
Fig. 2 Flow chart of the basic protocol

from buccal swabs using the illustra Ready-to-go Genomi-Phi™ v3 whole genome amplification kit (GE Life Sciences). All reagents in this kit are shelf stable and the protocol consists of three basic incubation steps: a denaturation step, an amplification step, and a deactivation step. Denaturation step: In a strip of 0.2 ml PCR tubes, 10 ul denaturation buffer was added, followed by 8 µl sterile DNase-free water, and 2 µl of template DNA. Samples were incubated for 3 min at 95 °C after which they were cooled on instant ice packs (Dynarex). In the meantime, a cooling program was begun on the miniPCR that was set up for incubation at the lowest temperature setting (20 °C), which results in continuous blowing of the miniPCR fans. Kit-provided strip tubes containing reagent cakes were placed in the miniPCR unit with the lid left open, and the 20 µl samples were then transferred to these tubes. Amplification step: all samples were incubated at 30 °C for 90 min in the miniPCR. Deactivation step: samples were incubated at 65 °C for 10 min in the miniPCR and then removed for storage at ambient temperature. All steps in this protocol and the preceding extraction from the buccal swabs were performed at a basic picnic table.

WGAs were kept at ambient temperature during storage in the field (~30 °C during the day) and transported back to the United States within 5–15 days, after which they were stored at −80 °C. Success of WGAs was assessed via visualization on a 2% agarose gel pre-stained with GelRed (Biotium). WGAs were then used as template in downstream PCRs targeting seven microsatellite loci that had already been characterized in this sifaka population (Lawler et al. 2001). Microsatellite PCRs using fluorescently labelled primers were performed in duplicate for each sample. Amplicons were run out on an agarose gel to confirm amplification of products of the target length and fragment analysis was performed using capillary electrophoresis with an ABI 3730xl 96-Capillary Genetic Analyzer at the DNA Analysis Facility at Yale University. GeneMarker v2.6.3 (SoftGenetics) was used to visualize the electropherograms and call genotypes. Homozygote genotypes were confirmed in a minimum of four independent PCRs. Allele frequency analysis was performed using CERVUS 3.0 (Kalinowski et al. 2007).

Option B: direct PCR

PCR reactions targeting the UTX/UTY sex-typing loci (Villesen and Fredsted 2006) were performed in the miniPCR using template from the faecal swabs. Most reactions were performed directly after extraction. This assay uses three primers to target an X-Y homologous portion of the ubiquitously transcribed tetratricopeptide repeat protein gene (Villesen and Fredsted 2006). The two target regions differ in length by 41 bp (X = 127 bp and Y = 86 bp) so that they may be visually distinguished via gel electrophoresis. Males will present as heterozygotes (two distinct bands) while females present as homozygotes (one band) (Fig. 3).

Reactions included 4 µl template DNA, 12.5 µl of EZ PCR Master Mix (miniPCR), 2 µl Ambion Ultrapure non-acetylated Bovine Serum Albumin (40 µg), 1 µl UTY/UTX primer (forward primer for both UTX and UTY regions), 4 µl UTY reverse primer, and 0.25 µl UTX reserve primer (all at a concentration of 10 µM), and 1.25 µl sterile DNase-free water. The Master Mix, primers, and BSA were stored frozen prior for use in these PCRs, but all of these reagents are stable at ambient temperature for a month or longer (based on manufacturer specifications and direct experience). Cycling conditions were as follows: 94 °C for 30 s, 59 °C for 40 s, 72 °C for 60 s for 35 cycles preceded by an activation step at 95 °C for 15 m and followed by a final extension step at 72 °C for 7 min. PCR products were visualized on a 2% agarose gel pre-stained with GelGreen (Biotium) in a portable gel box with built-in power source and blue light illuminator (blueGel, weight: 0.75 lbs, made by miniPCR). Gels were pre-made and stored wrapped in tinfoil inside a sealed plastic bag at room temperature for up to 2 weeks.

Results and discussion

During the August 2015 capture season at BMSR, we collected buccal swabs from 26 anaesthetized sifakas (21 yearlings and five adults) from which we extracted DNA and performed WGAs. The WGAs performed in field conditions showed a high rate of successful amplification (72% overall), as determined by gel electrophoresis visualization. Downstream microsatellite PCRs performed using successful WGAs as template also showed a high rate of successful amplification. We were able to call genotypes successfully from fragment analyses of the amplicons 89% of the time. (individual locus success rates were 80–97%; Table 1). These outcomes are similar to those we typically observe following our traditional microsatellite genotyping workflow based on silica column extractions from exported

Fig. 3 Gel showing successful amplification of sex-typing PCR

Table 1 Protocol success rates

Procedure	Test N	Success rate	
		Test	Typical
Trial 1: Buccal WGA			
Microsatellite genotyping success rate			
WGAs	104	72%	87%
Locus			
PV1	40	85%	86%
PV4	40	78%	77%
PV6	40	97%	87%
PV8	40	95%	91%
PV14	40	87%	91%
PV15	40	93%	87%
PV16	40	87%	89%
Total	280	89%	87%
Trial 2: Faecal PCR			
Sex-typing PCRs	42	94%	91%
Median DNA concentration (ng/μl)		53[a]	15
Median DNA molecular weight (ng)		3975[a]	2100

Trial 1: WGA success rate was determined as visible DNA following gel electrophoresis and microsatellite genotyping success rate determined as callable genotypes from electropherograms resulting from fragment analysis. Trial 2: Sex-typing PCR success rate was determined by visualization via gel electrophoresis. Typical success rates determined by these same procedures from "typical" sifaka samples in our hands (from gDNAs derived from tissue samples exported to the US and processed using traditional methods in the laboratory) are also shown

[a]Total DNA, likely includes high levels of bacterial DNA

tissue samples (Table 1). Genotypes for these individuals were not obtained using our traditional workflow because original samples remain in Madagascar, precluding a direct comparison. However, two adult males in our sample had been captured during previous field seasons and genotype data generated from our traditional protocol was available. In these cases, the genotypes derived for these individuals from both protocols matched, except in the case of one allele at one locus in which the estimated length was shifted by 2 bp, likely reflecting an error in allele calling due to stutter. Our sample size was too low to estimate deviation from Hardy–Weinberg, but the rates of observed versus expected heterozygosity did not markedly differ from what we typically observe.

We tried a variation on the incubation step of the WGA protocol to potentially conserve electricity in which we left samples at ambient temperature, which was around 30 °C, for 90 min rather than incubating them inside the miniPCR. We found that the WGAs incubated in the miniPCR amplified much more successfully (92% success rate) than those incubated at ambient temperature (43%), as determined by gel electrophoresis. In contrast,

we found no relationship between the number of days that WGAs were stored at ambient temperature and amplification success.

Fresh faeces were collected from 11 sifakas at the DLC that were used in the faecal swab and downstream sex-typing PCR protocol. The median concentration of faecal swab extractions was 53 ng/μl, as measured on the NanoDrop spectrophotometer, which measures total DNA concentration and, in the case of faecal-derived DNAs like these, could include relatively high concentrations of bacterial DNA. The median volume of supernatant was 75 μl. 94% of the UTX/UTY PCRs performed using supernatant directly after extraction from freshly collected samples showed amplification, as determined by gel electrophoresis visualization. PCRs using template DNA from extractions or samples stored at ambient temperature overnight or longer did not amplify as reliably; however, successful amplification was achieved from faecal swabs that were up to 13 days old. Sex genotypes based on multiple amplifications matched the known sex of the animals. One of the challenges we faced was optimizing extractions from non-invasively collected samples. While at BMSR, we opportunistically collected ten faecal samples that were used in our extraction-WGA protocol. In place of the swab, we dunked and swirled the faecal pellet in the tube containing Chelex solution with tweezers. However, few of the WGAs performed from these extractions showed amplification, as determined by gel electrophoresis, after returning to the lab. In our subsequent attempts at the DLC, we found that swabbing the exterior of the pellet to be critical as this likely disproportionately targeted host cells and avoided inhibiting compounds. We also increased the concentration of the Chelex solution used in faecal swab extractions from 10 to 20% to more effectively remove compounds in the faecal samples that could potentially inhibit downstream PCR.

The equipment proved functional, resilient to our field conditions, and easily transportable. Solar energy was sufficient to perform the planned number of analyses, and rates of DNA extraction and amplification were high. The PowerAdd and ChargeTech batteries functioned fully when left in direct sun, however additional types of lithium ion batteries proved unable to withstand similar temperatures during charging.

Our success was likely in part due to favourable field conditions; specifically, these included plentiful sunshine and an arid climate without extreme heat, which benefits sample preservation. In contrast, overcast climate or denser forest conditions could present challenges for obtaining sufficient power. Nevertheless, our protocol could likely be adjusted to suit such conditions with alternative equipment. For example, we also successfully ran the miniPCR from a car battery.

These experiments were also performed during a relatively short field season, with 15 days being the longest time period that reagents and WGAs were kept at ambient temperature. It is uncertain how well WGA integrity would hold up over longer periods without freezing. We tried storing WGAs on FTA cards, which could potentially extend the longevity of nucleic acid at ambient temperatures; however, attempts to extract WGA template from the cards failed. Similarly, pre-cast agarose gels are unlikely to hold up well during longer field excursions and in those cases it would be preferable to heat agarose solutions to cast fresh gels in the field. Alternatively, a number of suppliers offer gel cassettes that might be used.

Our protocols resulted in a success rate comparable or superior to the rate we typically observe in equivalent analyses performed in a dedicated molecular lab using samples transported internationally from BMSR. Recent technological developments in lab equipment, solar energy, and reagent chemistry will increasingly enable the incorporation of genetic analyses into field projects. A number of portable and semi-portable technologies have been used in molecular analyses, particularly for disease diagnostics (Marx 2015), and some of these technologies have allowed for genetic analyses to be performed in field laboratories (Bunting et al. 2014). Further, emerging portable genomics devices like the Oxford Nanopore MinION could greatly advance field researchers' capacity to obtain vast amounts of genetic data from wild organisms while in the field. These technologies currently depend on reagents that require freezing, but could potentially be used at field sites with solar and/or portable freezer options.

Our protocol offers a method that is highly portable, can be performed immediately upon sample collection in the field, and has the potential to be powered by green energy. These protocols also have an advantage in avoiding research delays related to permit applications. For example, U.S. Fish and Wildlife Services and the Center for Disease Control do not currently require permits for the import of synthetic nucleic acids. Further, the relative low cost and ease of use of the components of this "portable lab" could provide a means for researchers to incorporate genetic components into projects that previously did not include these analyses, without having to invest in a molecular lab. With some training, these analyses could also potentially be performed by local students or assistants, providing an opportunity for capacity building and community involvement.

References

Alberts SC, Buchan JC, Altmann J (2006) Sexual selection in wild baboons: from mating opportunities to paternity success. Anim Behav 72:1177–1196

Allendorf FW, Luikart G (2009) Conservation and the genetics of populations. Wiley-Blackwell, Oxford

Bergl RA, Vigilant L (2008) Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). Mol Ecol 16:501–516

Bradley BJ, Robbins MM, Williamson EA, Steklis HD, Steklis NG, Eckhardt N, Boesch C, Vigilant L (2005) Mountain gorilla tug-of-war: Silverbacks have limited control over reproduction in multimale groups. Proc Natl Acad Sci USA 102:9418–9423

Bunting S, Burnett E, Hunter RB, Field R, Hunter KL (2014) Incorporating molecular genetics into remote expedition fieldwork. Trop Conserv Sci 7:270–281

Caragiulo A, Pickles RS, Smith JA, Smith O, Goodrich J, Amato G (2015) Tiger (*Panthera tigris*) scent DNA: a valuable conservation tool for individual identification and population monitoring. Conserv Genet Resour 7:681–683

Castro I, Mason KM, Armstrong DP, Lambert DM (2004) Effect of extra-pair paternity on effective population size in a reintroduced population of the endangered hihi, and potential for behavioural management. Conserv Genet 5:381–393

Goossens B, Chikhi L, Jalil MF, Ancrenaz M, Lackman-Ancrenaz I, Mohamed M, Andau P, Bruford MW (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orangutan (*Pongo pygmaeus*) populations from Sabah, Malaysia. Mol Ecol 14:441–456

Gray TN, Vidya TN, Potdar S, Bharti DK, Sovanna P (2014) Population size estimation of an Asian elephant population in eastern Cambodia through non-invasive mark-recapture sampling. Conserv Genet 15:803–810

Griffin AS, Pemberton JM, Brotherton PN, McIlrath G, Gaynor D, Kansky R, O'Riain J, Clutton-Brock TH (2003) A genetic analysis of breeding success in the cooperative meerkat (*Suricata suricatta*). Behav Ecol 14:472–480

Jacobs RL, MacFie TS, Spriggs AN, Baden AL, Morelli TL, Irwin MT, Lawler RR, Pastorini J, Mayor M, Lei R, Culligan R, Hawkins MTR, Kappeler PM, Wright PC, Louis EE, Mundy NI, Bradley BJ (2017) Novel opsin gene variation in large-bodied, diurnal lemurs. Biol Lett 13:20170050

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. Trends Ecol Evol 21:629–637

Lawler RR, Richard AF, Riley MA (2001) Characterization and screening of microsatellite loci in a wild lemur population (*Propithecus verreauxi verreauxi*). Am J Primatol 55:253–259

Marx V (2015) PCR heads into the field. Nat Methods 12:393–397

Pallen MJ (2014) Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections. Parasitology 141:1856–1862

Quéméré E, Crouau-Roy BR, Rabarivola C, Louis EE Jr, Chikhi L (2010) Landscape genetics of an endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. Mol Ecol 19:1606–1621

Richard AF, Dewar RE, Schwartz M, Ratsirarson J (2002) Life in the slow lane? Demography and life histories of male and female sifaka (*Propithecus verreauxi verreauxi*). J Zool 256:421–436

Ruegg K, Rosenbaum HC, Anderson EC, Engel M, Rothschild A, Baker CS, Palumbi SR (2013) Long-term population size of the North Atlantic humpback whale within the context of worldwide population structure. Conserv Genet 14:103–114

Schwensow N, Eberle M, Sommer S (2008) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. Proc R Soc Lond B 275:555–564

Sussman RW, Richard AF, Ratsirarson J, Sauther ML, Brockman DK, Gould L, Lawler R, Cuozzo FP (2012) Bezà Mahafaly Special Reserve: long-term research on lemurs in southwestern Madagascar. In: Kappeler PM, Watts DP (eds) Long-term field studies of primates. Springer, Berlin Heidelberg, pp 45–66

Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Mol Ecol 6:869–876

Thalmann O, Wegmann D, Spitzner M, Arandjelovic M, Guschanski K, Leuenberger C, Bergl RA, Vigilant L (2011) Historical sampling reveals dramatic demographic changes in western gorilla populations. BMC Evol Biol 11:85. doi:10.1186/1471-2148-11-85

Villesen P, Fredsted T (2006) Fast and non-invasive PCR sexing of primates: apes, old world monkeys, new world monkeys and Strepsirrhines. BMC Ecol 6:8. doi:10.1186/1472-6785-6-8

Wedrowicz F, Saxton T, Mosse J, Wright W, Hogan FE (2016) A non-invasive tool for assessing pathogen prevalence in koala (*Phascolarctos cinereus*) populations: detection of Chlamydia pecorum and koala retrovirus (KoRV) DNA in genetic material sourced from scats. Conserv Genet Resour 8:511–521

Zehr SM, Roach RG, Haring D, Taylor J, Cameron FH, Yoder AD (2014) Life history profiles for 27 strepsirrhine primate taxa generated using captive data from the Duke Lemur Center. Sci Data 1:140019. doi:10.1038/sdata.2014.19.

Rapid identification of health care–associated infections with an integrated fluorescence anisotropy system

Ki Soo Park,[1,2]* Chen-Han Huang,[1,2]* Kyungheon Lee,[1,2] Yeong-Eun Yoo,[3] Cesar M. Castro,[1,4] Ralph Weissleder,[1,2,5]† Hakho Lee[1,2]†

Health care–associated infections (HAIs) and drug-resistant pathogens have become a major health care issue with millions of reported cases every year. Advanced diagnostics would allow clinicians to more quickly determine the most effective treatment, reduce the nonspecific use of broad-spectrum antimicrobials, and facilitate enrollment in new antibiotic treatments. We present a new integrated system, polarization anisotropy diagnostics (PAD), for rapid detection of HAI pathogens. The PAD uses changes of fluorescence anisotropy when detection probes recognize target bacterial nucleic acids. The technology is inherently robust against environmental noise and economically scalable for parallel measurements. The assay is fast (2 hours) and performed on-site in a single-tube format. When applied to clinical samples obtained from interventional procedures, the PAD determined the overall bacterial burden, differentiated HAI bacterial species, and identified drug resistance and virulence status. The PAD system holds promise as a powerful tool for near-patient, rapid HAI testing.

INTRODUCTION

Health care–associated infections (HAIs) and the emergence of drug-resistant pathogens are major health care issues. On any given day, 1 of 25 hospitalized patients becomes infected and as many as 1 of 9 succumb to death (1). HAIs incur a significant socioeconomic burden arising from prolonged hospital stays, lasting disability, and demand for new antimicrobials. In the United States, it is estimated that more than 600,000 patients develop HAIs every year (1), and HAI-related costs amount to $100 billion to $150 billion per year (2). Rapid, sensitive detection of pathogenic bacteria is a key to initiating timely treatment with proper antibiotics, preventing disease spread, and identifying infection sources in hospitals, homes, and other field settings (3–5).

Although bacterial culture is the clinical gold standard, it has drawbacks including long process times (up to several days), personnel cost, and the need for specialized equipment and species-specific protocols. As an alternative, nucleic acid (NA) testing has been increasingly adopted in clinical laboratories. On the basis of polymerase chain reaction (PCR) amplification of bacterial nucleotides, NA tests allow for comprehensive pathogen identification. The target sequence library is also rapidly expanding to empower these tests, aided by advances in bacterial whole-genome sequencing (6). Technical challenges, however, still remain in translating NA tests into routine clinical workflows. First, system operation can be complex, requiring trained operators. Although fully automated systems are available, they tend to be bulky and expensive. Second, assay costs are often higher than those for conventional screening. For example, in sequence-specific detection (for example, TaqMan, Molecular Beacon, and LightCycler), specialized probes for each NA target should be designed. Finally, NA tests are susceptible to false positives

due to accidental contamination by amplified products. To minimize the potential for cross-contamination, it is necessary to have the pre- and post-PCR areas in different spaces or PCR workstations. These constraints limit the penetration of NA tests to centralized hospital laboratories. For prompt, effective HAI control, assay platforms that can bring NA testing to the patient level (for example, community clinics and doctor's offices) are needed.

Here, we report a new detection system designed for rapid, cost-effective HAI diagnostics. Termed polarization anisotropy diagnostics (PAD), the system measures changes in fluorescence anisotropy when detection probes recognize target bacterial NAs (7). The detection is ratiometric, independent of fluorescence intensity, which makes the assay robust against environmental factors. We advanced the PAD to enable multilevel, on-site HAI diagnostics. Specifically, we have (i) developed a compact device with a disposable cartridge for sample preparation and multiwell detection, (ii) optimized the assay to perform NA amplification and detection without washing steps, (iii) embedded contamination control in the assay protocol, and (iv) created a library of sequence-specific probes to assess bacterial burden, pathogen types, antibiotic resistance, and virulence. As proof of concept, we applied PAD to detect clinically relevant HAI pathogens (8, 9): Gram-negative *Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa* and Gram-positive *Staphylococcus aureus*. The PAD demonstrated detection sensitivity down to the single bacterium level and determined drug resistance and virulence status. With clinical samples, the PAD achieved an accuracy comparable to that of bacterial culture; however, the PAD had a much shorter turnaround time (~2 hours) and allowed for on-site operation.

RESULTS

Polarization anisotropy diagnostics

Figure 1A shows the assay scheme. Following bacterial lysis, target NAs [for example, regions in 16S ribosomal RNA (rRNA) or mRNA] are

[1]Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. [2]Department of Radiology, Harvard Medical School, Boston, MA 02114, USA. [3]Department of Nanomanufacturing Technology, Korea Institute of Machinery and Materials, Daejeon 305-343, Korea. [4]Department of Medicine, Harvard Medical School, Boston, MA 02114, USA. [5]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.
*These authors contributed equally to this work.
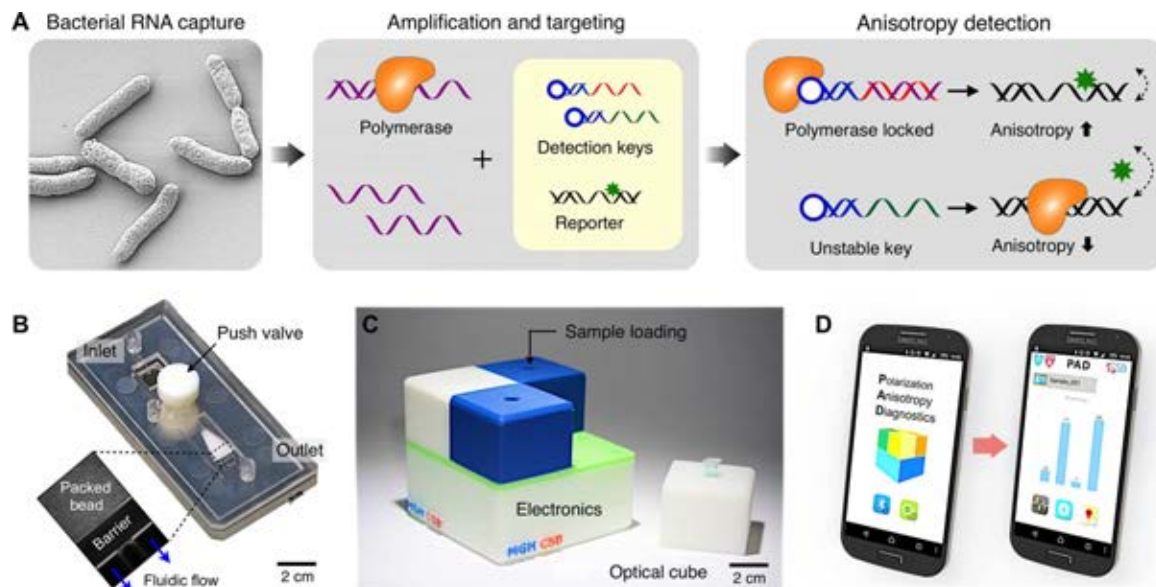†Corresponding author. Email: rweissleder@mgh.harvard.edu (R.W.); hlee@mgh.harvard.edu (H.L.)

Fig. 1. PAD system. (**A**) Assay procedure. Bacteria are lysed, and total RNA is extracted. Following the RT-PCR amplification, samples containing amplicons and DNA polymerase are incubated with an all-in-one master mix that has the detection key and the reporter. The resulting fluorescence anisotropy of the sample is then measured. (**B**) Photograph of a disposable RNA extraction cartridge made in plastic. The device has an RNA extraction chamber packed with glass beads (inset). (**C**) Photograph of a portable system for fluorescence anisotropy detection. Four separate optical cubes can be plugged into an electronic base station. (**D**) PAD measurement is controlled through a custom-designed application in a smartphone. The PAD device and the smartphone communicate via Bluetooth.

amplified via asymmetric reverse transcription PCR (RT-PCR). Next, an all-in-one PAD mix is added. The mix consists of two reagents: (i) a detection key that is derived from an aptamer (specific to DNA polymerase) adjoined with a complementary sequence to target NAs (7) and (ii) a reporter DNA with a fluorophore. The detection key, stabilized through hybridization with the target NA, locks into DNA polymerase and deactivates its enzymatic activity. The reporter DNA then retains its structure and assumes high fluorescence anisotropy (r) due to slow diffusional motion. Conversely, the anisotropy is low in the absence of the NA targets, because unlocked DNA polymerase cleaves the reporter's fluorophore during the extension reaction. The assay is fast (2 hours for completion) and is performed without any washing steps.

PAD system

To enable on-site HAI detection, we implemented a compact PAD system (Fig. 1, B and C). The system had two major parts: a disposable sample-processing cartridge and a compact reader for fluorescence anisotropy detection. The cartridge was used to extract bacterial NAs (see fig. S1 for the device structure). It had a fluidic chamber packed with glass beads (diameter, 30 μm) to create an on-chip filter (fig. S1A) (10): negatively charged NAs could be adsorbed to the positively charged bead surface under high-salt condition and then be eluted by changing the salt concentration. We implemented the device in poly(methylmethacrylate) via injection molding (see Materials and Methods for details). When compared to a commercial column filter, the fluidic cartridge showed comparable performance, extracting RNA with good quality (fig. S2). The RNA integrity numbers (RINs), ranging from 1 (poor) to 10 (best), were 9.6 ± 0.3 (fluidic cartridge) and 9.5 ± 0.2 (column filter). By using the on-chip cartridge, however, we could remove centrifugal washing steps.

The detection system was designed for portable, parallel PAD assays. It had a modular structure consisting of a base unit for signal processing and four plug-in optical cubes (Fig. 1C and fig. S3). Each optical cube could be customized to accommodate differently sized sample containers and fluorescence optics. The packaged PAD system had a small form factor ($8 \times 8 \times 8$ cm^3) and weighed ~400 g. The system wirelessly communicated with a computer or a smartphone through Bluetooth connection, which further improved system portability and simplified the assay setup. We designed a smartphone application for system operation as well as for data logging with geographic information (Fig. 1D and fig. S4).

PAD optics

Figure 2A shows the schematic of an optical cube in the PAD system. We used a light-emitting diode (LED) as an affordable light source. The illumination light passed through a linear polarizer and focused onto a sample to excite fluorophore (Fig. 2A, inset). The intensity of emission light was then measured by a pair of photodiodes. We measured both the parallel (I_x) and perpendicular (I_y) components relative to the polarized excitation light and calculated the fluorescence anisotropy (r) as $r = (I_x - I_y) \cdot (I_x + 2I_y)^{-1}$. The optics was encased into an opaque plastic body to minimize the interference from adjacent modules (fig. S3B).

For robust signal detection, we adopted the lock-in measurement technique. Exploiting the fact that r is intensity-independent, we modulated the intensity of the excitation light at the carrier frequency of 1 kHz. The emission signals were then frequency-locked through the homodyne signal path (Fig. 2B). The scheme significantly improved the signal-to-noise ratio (>28 dB; Fig. 2C) and allowed for reliable system operation under ambient light. Data acquisition was directed by a microcontroller unit (MCU; Arduino). A multichannel digital-to-analog converter (DAC) was used to deliver a tunable, modulating signal to the driver in each optical cube. The signal from a photodetector was amplified by a current amplifier, passed through an analog lock-in circuit,
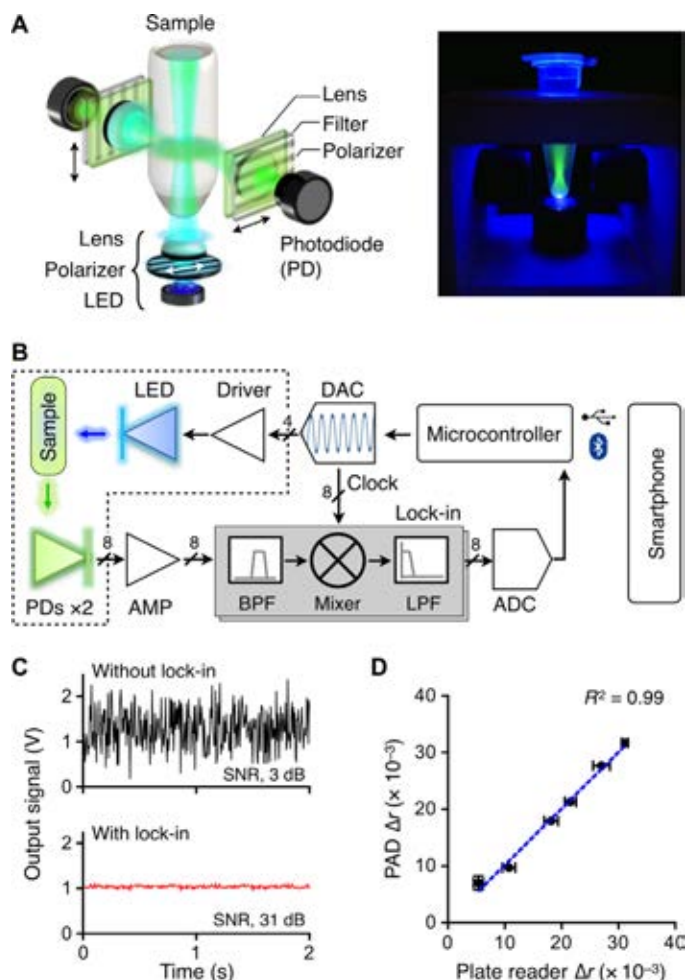
Fig. 2. Optical detection design. (**A**) Schematic of an optical cube. The optical excitation module has an LED, a linear polarizer, and a focusing lens. The emission light is measured by a pair of detector sets, each consisting of a lens, a polarization filter, and a photodiode (PD). (**B**) Circuit diagram. The on-board computer controls the entire system and communicates with a smartphone. To enhance the signal-to-noise ratio (SNR), the system uses the optical lock-in detection scheme. The intensity of the excitation light is amplitude-modulated, and the resulting emission intensities are mixed with the carrier frequency. The dotted box indicates an optical cube. BPF, band-pass filter; LPF, low-pass filter; AMP, amplifier. (**C**) The lock-in method significantly improved the signal-to-noise ratio (630 times, 28 dB). (**D**) The accuracy of the PAD was benchmarked against a commercial plate reader ($R^2 = 99\%$). Experiments were performed in triplicate, and the data were displayed as means ± SD. Horizontal and vertical error bars were from the plate reader and the PAD measurements, respectively.

and digitized by a multichannel analog-to-digital converter (ADC). The MCU was programmed to serially poll the optical cubes, control all peripheral components, and communicate with external devices via Bluetooth (fig. S5).

We benchmarked the PAD system against a commercial plate reader (Sapphire 2, Tecan). To prepare samples of different fluorescence anisotropy, we varied the ratio between free fluorescein-labeled DNA (FAM-DNA; high fluorescence anisotropy) and template-bound FAM-DNA (low anisotropy; see Materials and Methods). The observed r values
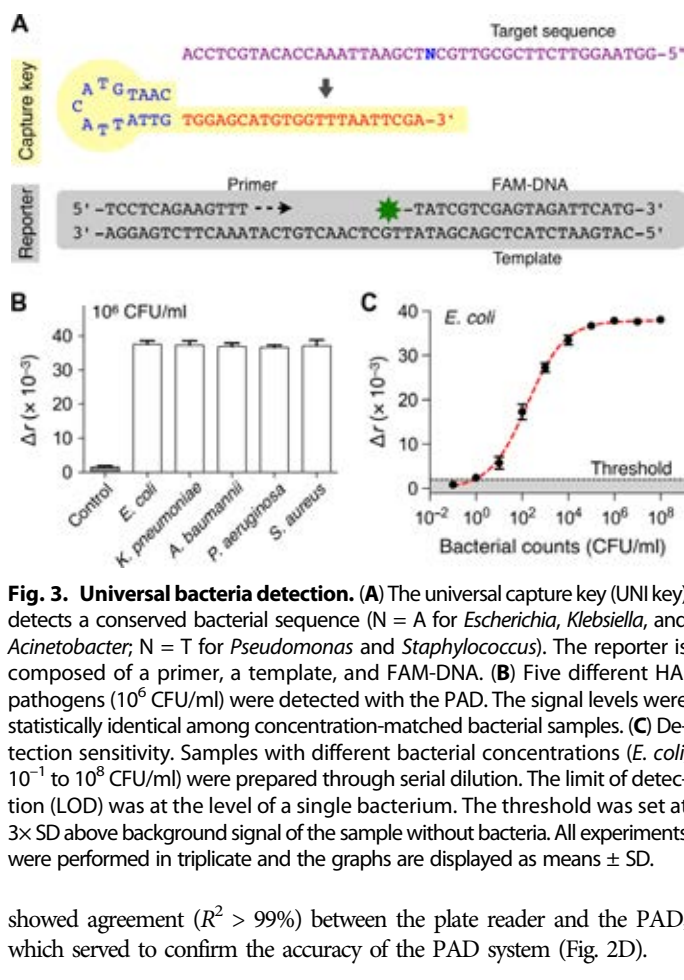
Fig. 3. Universal bacteria detection. (**A**) The universal capture key (UNI key) detects a conserved bacterial sequence (N = A for *Escherichia*, *Klebsiella*, and *Acinetobacter*; N = T for *Pseudomonas* and *Staphylococcus*). The reporter is composed of a primer, a template, and FAM-DNA. (**B**) Five different HAI pathogens (10^6 CFU/ml) were detected with the PAD. The signal levels were statistically identical among concentration-matched bacterial samples. (**C**) Detection sensitivity. Samples with different bacterial concentrations (*E. coli*, 10^{-1} to 10^8 CFU/ml) were prepared through serial dilution. The limit of detection (LOD) was at the level of a single bacterium. The threshold was set at 3× SD above background signal of the sample without bacteria. All experiments were performed in triplicate and the graphs are displayed as means ± SD.

showed agreement ($R^2 > 99\%$) between the plate reader and the PAD, which served to confirm the accuracy of the PAD system (Fig. 2D).

Universal key to measure bacterial load

We first used the PAD to detect the overall bacterial burden. We designed a single, universal key (UNI key) that targets a conserved region of 16*S* rRNA in different bacterial species (Fig. 3A and Table 1). We also prepared a common reporter probe that was composed of FAM-DNA, its primer, and its template (Fig. 3A). The PAD output was defined as $\Delta r = r - r_0$, where r_0 is the fluorescence anisotropy of control samples containing DNA polymerase and the reporter only.

We applied the designed assay (UNI-PAD) to detect representative HAI pathogens (*E. coli*, *K. pneumoniae*, *A. baumannii*, *P. aeruginosa*, and *S. aureus*). Across different bacterial species, we observed consistent Δr values in concentration-matched samples [Fig. 3B; $P = 0.8857$, one-way analysis of variance (ANOVA)]; this result supported the use of the UNI-PAD in estimating total bacterial load. We next performed titration experiments with serially diluted bacterial samples (Fig. 3C; see Materials and Methods for details). The PAD assay achieved the dynamic range spanning $>10^4$ colony-forming units (CFU); the limit of detection was down to single-digit CFU.

Assay optimization for point-of-care operation

We optimized the PAD system for its point-of-care (POC) applications. One major issue in POC NA testing is to control false positives caused by sample contamination with PCR products (that is, carryover contamination) (*11*, *12*). To minimize such effects, we adopted

Table 1. Sequence of detection keys designed for the PAD (see table S1 for other targets). PVL, Panton-Valentine leukocidin.

Category	Target	DNA sequence (5′ to 3′)
UNI	Universal	CAATGTACAGTATTGTGGAGCATGTGGTTTAATTCGA
HAI	*Escherichia*	CAATGTACAGTATTGTGGAGGAAGGGAGTAAAGTTAAT
	Klebsiella	CAATGTACAGTATTGTGGAGGCAAGGCGATAAGGT
	Acinetobacter	CAATGTACAGTATTGTTCTAGTTAATACCTAGGGATAGTG
	Pseudomonas	CAATGTACAGTATTGTGGAGGAAGGGCAGTAAGTTAA
	Staphylococcus	CAATGTACAGTATTGTGGGAAGAACATATGTGTAAGTAAC
ARV	*nuc*	CAATGTACAGTATTGTTGCTTCAGGACCATATTTCTCTAC
	femB	CAATGTACAGTATTGTCCAGAAGCAAGGTTTAGAATTG
	mecA	CAATGTACAGTATTGTCTGCTATCCACCCTCAAACAG
	PVL	CAATGTACAGTATTGTCGGTAGGTTATTCTTATGGTGGAG
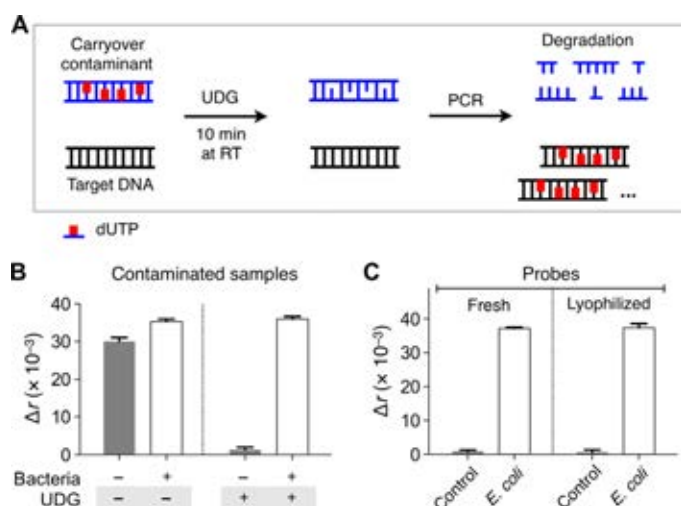
Fig. 4. Assay optimization for POC operation. (**A**) Schematic illustration of UDG-mediated control on carryover contamination. Uracil-containing carryover contaminant was specifically broken down by UDG, which allows for the amplification of the true target DNA only. (**B**) Uracil-containing contaminants (10^7 copies) were added to all samples. Contaminated samples produced high signal even in the absence of target bacteria. When samples were treated with UDG, the false-positive signal was eliminated. (**C**) The PAD reagents were lyophilized to facilitate their transport and extend their shelf life. After 4 weeks of storage in ambient condition, the reagents were used for bacterial detection. No difference was observed between fresh and lyophilized reagents. Bacterial samples in (B) and (C) contained *E. coli* (10^6 CFU/ml). All experiments were performed in triplicate, and the data are displayed as means ± SD. RT, room temperature.

the uracil-DNA glycosylase (UDG) method (*13*, *14*). By substituting deoxythymidine triphosphate (dTTP) with deoxyuridine triphosphate (dUTP) during PCR, we rendered all amplicons to have a uracil-containing DNA backbone. Applying UDG cleaved these amplicons, which selectively destroyed carryover contaminants from PCRs while keeping bona fide DNA templates (Fig. 4A). We confirmed that the method was compatible with the PAD assay. The signal level remained the same when dUTP replaced dTTP in DNA targets (fig. S6). We next

tested the efficacy of this contamination control. As a carryover contaminant, dUTP-containing PCR products (10^7 amplicons) were spiked into samples. Without UDG treatment, the negative control with no bacterial targets showed false positives (Fig. 4B). This signal was eliminated with the addition of UDG, and only samples containing true bacterial targets yielded high Δr (Fig. 4B).

To maximize system portability, we further combined the PAD with a miniaturized thermocycler (miniPCR, Amplyus) (*15*). The performance of the POC system matched that of conventional benchtop equipment (fig. S7). We also lyophilized all chemical reagents (for example, detection keys and reporters) to facilitate their transport and storage (see Materials and Methods for details). The reagents retained their activity after >2 weeks of storage in ambient conditions; we observed statistically identical Δr values ($P > 0.64$, two-tailed t test) with fresh and stored agents (Fig. 4C and fig. S8).

Differential keys for pathogen classification

To differentiate HAI-causing pathogens, we designed a set of detection keys (HAI keys) in which each key targets the hypervariable region of 16S rRNA in different bacterial species (Table 1; see table S1 for extended targets). The sequence homology among genus types was kept <50% to minimize nonspecific binding. When tested, the HAI keys assumed high specificity. For example, the *Escherichia* key (Fig. 5A) showed high signal (Δr) only with its intended target, whereas off-target signals were negligible even in high biological background (10^6 CFU of other bacterial species; Fig. 5B). Similarly, other HAI keys displayed excellent specificity with minimal crosstalk (Fig. 5C). Electrophoretic band-shift analysis confirmed that the detection keys, in the presence of complementary target amplicon, bound to DNA polymerase and inhibited its catalytic activity (fig. S9).

We next prepared probes for antibiotic resistance and virulence (ARV keys) for further bacterial phenotyping. These keys targeted bacterial genes that make pathogens antibiotic-resistant or highly virulent. As a model system, we profiled samples for *mecA*, PVL, *nuc*, and *femB* genes. *mecA* is the determining factor conferring MRSA, a common multidrug-resistant HAI pathogen (*16*). PVL, *nuc*, and *femB* are virulence factors that contribute to the pathogenicity of *S. aureus*. We used two representative MRSA strains that have the following known genotypes: health care–associated MRSA (*mecA*$^+$, PVL$^-$) and community-acquired MRSA
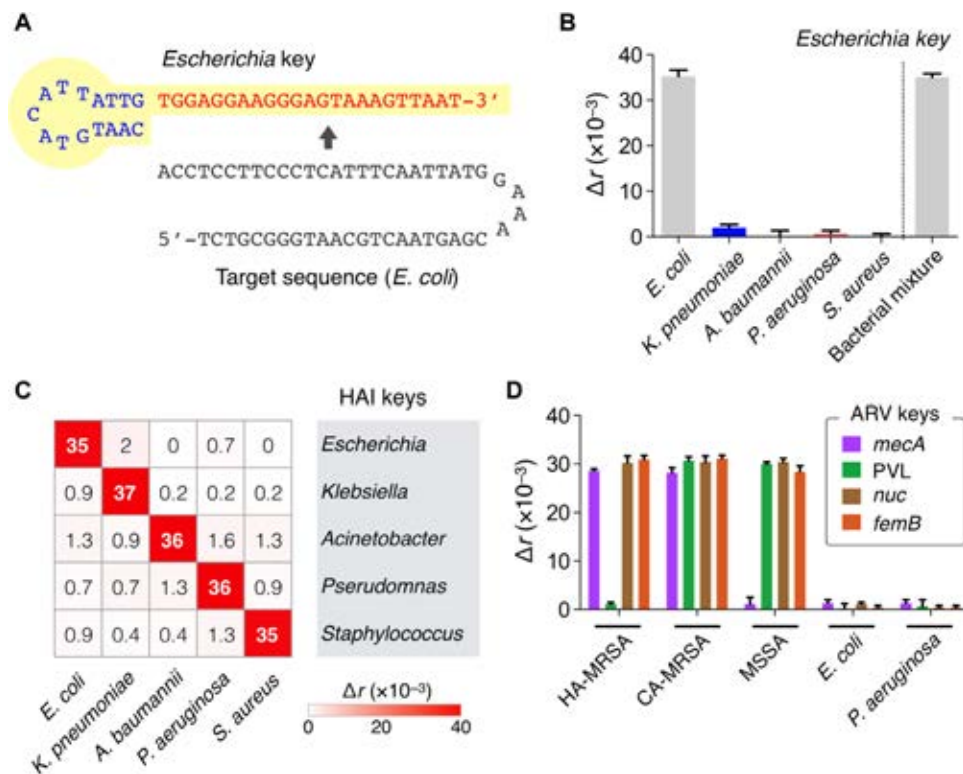
Fig. 5. Bacteria typing with PAD. (A) A set of detection keys specific for HAI pathogens was designed (HAI keys). An *Escherichia* key is shown as an example. **(B)** The specificity of HAI keys was tested. The signal was high only in the presence of the target species even in the mixture of other bacterial species. An example of *E. coli* (10^6 CFU/ml) detection is shown. **(C)** Heat map of Δr values obtained for HAI detection. Bacterial concentration was 10^6 CFU/ml. **(D)** Detection keys for antibiotic resistance and virulence (ARV keys) were designed for further typing. Two types of methicillin-resistant *S. aureus* (MRSA; 10^6 CFU/ml), health care–associated MRSA (HA-MRSA) and community-acquired MRSA (CA-MRSA), were identified by targeting the specific regions in *mecA* and PVL genes. Three pathogens [methicillin-sensitive *S. aureus* (MSSA), *E. coli*, and *P. aeruginosa*; 10^6 CFU/ml] were included as controls. All experiments were performed in triplicate. The heat map displays mean values and the bar graphs display means ± SD.

($mecA^+$, PVL^+). Control samples were MSSA ($mecA^-$, PVL^-, nuc^-, $femB^-$), *E. coli* ($mecA^-$, PVL^-, nuc^-, $femB^-$), and *P. aeruginosa* ($mecA^-$, PVL^-, nuc^-, $femB^-$) (table S2) (*17–20*). The ARV-PAD correctly genotyped bacteria, agreeing with a quantitative real-time PCR (qPCR) (Fig. 5D and fig. S10).

Clinical application

Finally, we applied the PAD for clinical HAI diagnostics. The assay started with bacterial lysis, followed by the NA collection using the plastic cartridge. Following the NA amplification, the samples were analyzed in parallel for total bacterial burden (UNI key), HAI pathogens (HAI keys), and antibiotic resistance and virulence status (ARV keys) (fig. S11). The total assay time was ~2 hours, and the required sample volume was ~40 μl.

We acquired patient samples and aliquoted them for the PAD test (2 hours) and conventional culture (3 to 5 days) in a clinical microbiology laboratory. Test results for all detection keys (UNI, HAI, and ARV) are shown in Fig. 6A and fig. S12. Samples negative with UNI-PAD were also negative with HAI-PAD, suggesting the potential use of universal detection for sample triaging. Among six UNI-PAD–positive samples, the HAI-PAD detected HAI pathogens in five samples, and the differentiation results matched the bacterial culture readouts (Fig. 6B). One patient (no. 6) was positive with bacterial load but was negative with HAI keys; the patient was later found to be infected with *Providencia*

rettgeri (nontargeted in the current HAI-PAD). For the sample positive for *S. aureus* (patient no. 2), the ARV-PAD showed $mecA^-$ status (that is, MSSA); this result matched the MRSA-negative pathology report and qPCR (Fig. 6 and fig. S13).

DISCUSSION

HAIs have become a ubiquitous, recalcitrant, and costly problem in modern health care. They increase the emergence of antibiotic resistance, cause significant morbidity and mortality, and prolong hospital stays. One of the key mandates to control this endemic is to equip local hospitals and community centers with more effective surveillance systems. The PAD platform presented here could enable rapid HAI detection in those settings. The detection system is compact and user-friendly with minimal operation complexity. The assay is comprehensive, assessing for overall bacterial burden, pathogen types, antibiotic resistance, and virulence factors.

Several features make the PAD ideal for field operation. First, the sensing scheme (fluorescence anisotropy) is inherently robust against environmental noise. We further incorporated the optical lock-in technique to significantly enhance the sensitivity. Second, the assay flow involves minimal complexity and hands-on time. A disposable plastic chip is used
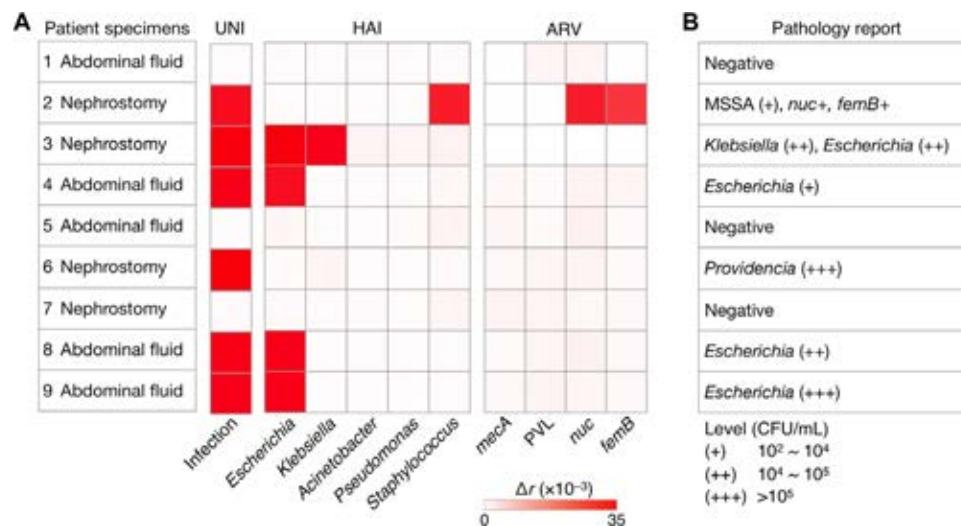
Fig. 6. Clinical application of PAD for HAI detection. (**A**) Nine samples from different patients were processed by the PAD for bacterial load (UNI), presence of the HAI species (HAI), and resistance/virulence status (ARV). (**B**) The clinical samples were also tested by a clinical pathology laboratory (culture and qPCR). The PAD and pathology reports agreed with each other.

to collect NAs, and the remaining processes are performed in a single tube without washing steps. The assay protocol is also refined to automatically dissolve carryover contaminants, thereby minimizing false positives. Third, the platform is highly affordable. The PAD device has simple electronics and can readily be expanded for parallel detection. Making an injection-molded cartridge brings advantages including high performance reproducibility, less cross-contamination between samples, and lower cost. Finally, the PAD is scalable for comprehensive screening. Decoupling detection and signaling probes enables such assays to be cost-effective, because a common fluorescence reporter can be used for all detection targets. We have already designed detection probes for >35 targets (table S1); the incremental assay cost for additional targets is ~$0.01.

In a pilot clinical test, PAD accuracy was comparable to that of bacterial culture. In contrast to the culture, the PAD assay was fast (~2 hours), multiplexed, and cost-effective (<$2 per assay). However, we note the following limitations in the current study. First, no clinical samples were found to contain drug-resistant strains. Further studies with larger cohorts are needed to verify PAD's capacity for drug resistance screening. Second, the PAD may present ambiguous results when target NAs overlap. For example, the current ARV keys would fail to discern community-acquired MRSA ($mecA^+$, PVL^+) from the mixture of health care–associated MRSA ($mecA^+$, PVL^-) and MSSA ($mecA^-$, PVL^+). Such incomplete classification is an inherent issue with NA-based tests. We expect that this issue would be overcome as more bacteria genomic data are accrued from whole-genome sequencing efforts.

The current prototype system could be further improved. First, we envision a self-contained, closed system in which sample preparation, thermocycling, and detection functions are all housed in a single device. Such a system would effectively eliminate erroneous results from sample contamination, user interference, or both and proffer "sample-in and answer-out" tests. Second, the current assay time could be shortened, particularly for DNA amplification. One promising direction is to adopt a photonic thermocycling system, which can complete the entire PCR in <5 min (21). Isothermal amplification is also an alternative; some isothermal reactions can be completed in 20 min, which could bring down

the total assay time to <1 hour (22). Using isothermal amplification would simplify the hardware requirement and make a battery-powered device feasible. In the future, we plan to expand the test library for broader pathogen and antibiotic resistance screening. The modular nature of PAD probes should facilitate incorporating new targets such as host-response factors (for example, interleukin-4, platelet-derived growth factor B chain, monocyte chemoattractant protein-1, and C-X-C motif chemokine 10) (23), as well as other viral, fungal, and parasitic markers.

MATERIALS AND METHODS

Fabrication of the plastic cartridge
The device was made in plastic via injection molding. A metal block of the mold was first machined to have surface structures such as channels, chambers, and ports negatively shaped to those on the top and bottom parts of the plastic cartridge. To confine microbeads in the RNA capture chamber, a weir-shaped physical barrier was designed at the outlet side of the chamber. The top and bottom parts of the device were injection-molded in a foundry (Korea Institute of Machinery and Materials). More than two devices were produced per minute. The top and bottom parts were glued together and the three-way valve was inserted. Fluidic connection was made by inserting polyethylene tubes in the fluidic ports.

Device preparation
The fluidic cartridge was filled with glass beads (diameter, ~30 μm; Polysciences). Beads were suspended in 75% ethanol (Sigma) and introduced through the inlet. The beads were retained in the RNA capture chamber due to the weir-style physical barrier in the outlet side of the chamber (fig. S1A). Following bead capture, excess ethanol was collected and removed. The entire device was then flushed with cycles of RNaseZap (Life Technologies), ribonuclease (RNase)–free water (Life Technologies), and ethanol, and dried. All fluidic flow was generated by manually operating syringes.

Detection system

The illumination source in the optical cube consisted of an LED (λ = 470 nm; Thorlabs), a dichroic film polarizer (polarization efficiency, >99%; Thorlabs), and a convex lens [focal length (f) = 8 mm]. The detection part had a convex lens (f = 8 mm), a long-pass filter (EL0500, Thorlabs), a dichroic film polarizer, and a photodiode (S1223, Hamamatsu). A 16-bit DAC (LTC1597, Linear Technology) was used to deliver the modulated control signal to a custom-designed LED driver (LF356, Texas Instrument). The signal from the photodiode was amplified by a custom-designed current amplifier (AD549, Analog Devices). For the lock-in detection, the amplified signal was passed through a band-pass filter (center frequency, 1 kHz; bandwidth, 100 Hz), mixed with a carrier signal, and passed through a low-pass filter (time constant, 1 ms). The conditioned signal was digitized by a 16-bit ADC (LTC1867, Linear Technology). A microcontroller (Arduino MEGA 2560) was programmed to control the light sources for multiplexing, to perform real-time data recording, and to communicate with an external device (for example, computers and smartphones) via a USB 2.0 or a Bluetooth interface. The entire system was powered by a 9-V battery housed inside the base station. A typical power consumption during a single optical measurement (one cube) was ~400 mW, and each test took ~30 s.

Smartphone application

We created a custom-designed Android application to facilitate system operation and data recording. Control software was designed using MIT App Inventor 2. The application connected the smartphone to the PAD system and sent the triggering signal for the fluorescence anisotropy detection. The measured data were sent back to the phone and combined with a time stamp and Global Positioning System coordinates.

Signal detection

Fluorescence anisotropy values were measured with the excitation and emission wavelengths of 470 and 525 nm, respectively. Fluorescence anisotropy (r) was calculated using the following equation: $r = (I_x - I_y) \cdot (I_x + 2I_y)^{-1}$, where I_x and I_y are emission intensities when the emission polarizers are in parallel with and perpendicular to the excitation polarizer, respectively. The LOD was estimated by setting the threshold at $3 \times$ SD above the background signal of samples without bacteria. For comparison with the benchtop equipment (Sapphire 2, Tecan), we measured $\Delta r = r - r_{FAM}$, where r_{FAM} is the fluorescence anisotropy only in the presence of FAM-DNA (62.5 nM) and r is the fluorescence anisotropy in the presence of FAM-DNA (62.5 nM) along with its template. We varied the template concentrations (10, 20, 30, 40, 50, and 60 nM) to produce different amounts of hybridized FAM-DNA.

Probe design

DNA oligonucleotides were synthesized by Integrated DNA Technologies. The list of DNA sequences is summarized in Table 1 and tables S1 and S3. For the universal and species-specific detection of pathogenic bacteria, individual 16S rRNA sequences of different bacterial genera [from the National Center for Biotechnology Information (NCBI) nucleotide database] were aligned using MegAlign software (DNASTAR), and both conserved and variable regions were selected as target sequences (24). To detect ARV factors, the specific regions of *nuc*, *femB*, *mecA*, and PVL (from the NCBI nucleotide database) were selected as target sequences (17–20). The detection keys were designed to have a hairpin structure joined by a single-stranded capture sequence (~20 to 25 nucleotides in length) (7, 25).

Lyophilization of master mix

An all-in-one PAD mix (20 μl) was concocted by mixing a detection key (400 nM), a reporter (150 nM), a primer (150 nM), and FAM-labeled DNA (125 nM) in 20 mM tris-HCl (pH 8.3), 20 mM KCl, 5 mM $(NH_4)_2SO_4$, and 6 mM $MgCl_2$. The mixture was first frozen in liquid nitrogen and then dried in VirTis Freezemobile 25EL Freeze Dryer (SP Scientific). The lyophilized reagents were stored at room temperature and reconstituted before use by adding 20 μl of UltraPure DNase/RNase-free distilled water (Life Technologies).

Experiment with cultured bacteria

All bacteria were purchased from the American Type Culture Collection (ATCC). Bacterial cultures were grown to mid-log phase in vendor-recommended medium: *E. coli* (#25922) in LB medium (BD Biosciences); *P. aeruginosa* (#142), *K. pneumoniae* (#43816), and MRSA (#BAA-1720 and #BAA-1707) in tryptic soy broth (BD Biosciences); *A. baumannii* (#15149) in nutrient broth (BD Biosciences); and *S. aureus* (#25923) in *Staphylococcus* broth (BD Biosciences). Bacteria were collected via centrifugation (6000g, 10 min), and pellets were resuspended with the preheated TRIzol (Life Technologies). The resuspended cells were transferred to 2-ml Safe-Lock tubes (Eppendorf) containing sterilized disruptor beads (0.1 mm; Scientific Industries) and lysed using a vortex mixer. After centrifugation, the supernatant was transferred to a new tube.

RNA extraction

Bacterial lysate mixed with an equal volume of ethanol was flown through the RNA extraction chamber, in which RNA was captured by packed glass beads. Subsequent flushing with Direct-zol RNA Pre-Wash (Zymo Research) and RNA Wash Buffer (Zymo Research) was done to remove traces of impurities and chaotropic salts. Finally, RNA was eluted in RNase-free water. For comparison of the fluidic cartridge with a commercial column (Zymo-Spin column, Zymo Research), bacterial lysate was divided into two aliquots. One sample was processed by the commercial column and the other by the fluidic cartridge. We checked the quality of the extracted RNA through an electrophoretic assay (2100 Bioanalyzer, Agilent). RNA molecular weight ladder provided in the kit (RNA 6000 Nano Chip, Agilent) was used as reference, and electrophoretic runs were performed per the manufacturer's protocol. The analysis assigned RINs to samples, ranging from 1 to 10, where 1 indicates highly degraded RNA and 10 completely intact RNA.

PAD assay

The single-stranded complementary DNA (cDNA) was synthesized using random priming with Promega's Reverse Transcription System as per the manufacturer's protocol. The asymmetric PCR amplification was then carried out in a total reaction volume of 25 μl containing 2.5 μl of cDNA, 0.8 μM excess primer, 0.08 μM limiting primer (table S3), 1× PCR buffer [20 mM tris-HCl, 20 mM KCl, 5 mM $(NH_4)_2SO_4$, and 2 or 3 mM $MgCl_2$], 0.2 mM deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), and deoxycytidine triphosphate (dCTP), 0.4 mM dUTP (Thermo Scientific), 2 U of Antarctic Thermolabile UDG (New England Biolabs), and 2.5 U of Maxima Hot Start Taq DNA polymerase (Thermo Scientific). For the asymmetric PCR on a miniaturized thermocycler (miniPCR, Amplyus), we used the following cycling conditions: 25°C for 10 min and 94°C for 4 min; 35 cycles of 30 s at 94°C, 30 s at 56°C, and 30 s at 72°C; and an extension step of 10 min at 72°C. With a benchtop thermocycler (MasterCycler, Eppendorf), the reaction conditions were 25°C for 10 min and 94°C for 4 min;

35 cycles of 5 s at 94°C, 15 s at 56°C, and 15 s at 72°C; and a final 10 min at 72°C. The PCR solution (20 μl) was mixed with an all-in-one PAD mix composed of a detection key (200 nM) and a reporter that was pre-formed with a template (75 nM), a primer (75 nM), and FAM-labeled DNA (62.5 nM) at room temperature for 15 min, making a total volume of 40 μl in 20 mM tris-HCl (pH 8.3), 20 mM KCl, 5 mM $(NH_4)_2SO_4$, and 4 mM $MgCl_2$.

UDG-mediated control of carryover contamination

To mimic the carryover contamination, we spiked dUTP-containing amplification products (carryover contaminants) into new reaction samples. The copy number of carryover contaminants was ~10^7; a single aerosol after PCR typically contains as many as 10^6 amplification products (26). The dUTP-containing amplicons were prepared following the same procedures outlined above except that equal amounts of limiting and excess primers were used. The obtained amplicons were purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) and quantified by measuring the absorbance at 260 nm with NanoDrop 1000 (Thermo Scientific). The copy number was estimated on the basis of the conversion factor (26 kD per amplicon).

Clinical samples

This study was approved by the Partners Institutional Review Board. Excess and discarded samples were collected from nine subjects with clinical suspicion for infected bodily fluid or abscess and referred for drainage. Specimens were collected using routine image-guided approaches by Massachusetts General Hospital Interventional Radiology physicians and analyzed blindly using the PAD assay. Specimens (500 μl) were mixed with 1.5 ml of TRIzol LS (Life Technologies) which is more concentrated than TRIzol, and the same RNA extraction procedure was applied as in pure bacterial cultures.

Electrophoretic band-shift experiment

Solution containing 100 or 200 nM detection key, PCR products, and 12.5 U of Taq DNA polymerase (New England Biolabs) in 20 mM tris-HCl (pH 8.3), 20 mM KCl, 5 mM $(NH_4)_2SO_4$, and 4 mM $MgCl_2$ was incubated at room temperature for 20 min. The solution was mixed with 6× loading buffer and subjected to electrophoresis on a 20% poly-acrylamide gel (Life Technologies). The analysis was carried out in 1× TBE (100 mM tris, 90 mM borate, 1 mM EDTA) at 150 V for 160 min at 4°C. After GelRed (Biotium) staining, gels were scanned using an ultraviolet transilluminator. DNA polymerase was neither fluorescent nor stained by the dye (27).

Quantitative real-time PCR

The cDNA derived from in vitro cultured bacteria or clinical samples was mixed with 1× PowerUp SYBR Green Master Mix (Life Technologies) and 0.4 μM specific primers used in the PAD assay. Thermal cycling was then carried out on the 7500 Fast Real-Time PCR system (Life Technologies) with the following conditions: UDG activation (50°C, 2 min), initiation (95°C, 2 min); 40 cycles of denaturation (95°C, 5 s); annealing (56°C, 15 s); extension (72°C, 30 s). The 7500 Fast software automatically calculates the C_t value, which represents the first PCR cycle at which the fluorescence signal exceeds the signal of a given uniform threshold. No-template control (NTC) remained undetected, not crossing the established threshold for 40 cycles, and was arbitrarily given a C_t value of 41. The ΔC_t was generated by subtracting the C_t value of the specimen from the C_t value of NTC (28).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/5/e1600300/DC1

table S1. Target sequences recognized by established detection keys.
table S2. Summary of a set of ARV factors targeted in this study.
table S3. DNA sequences used in this study.
fig. S1. Schematic of the plastic cartridge for RNA extraction.
fig. S2. Comparison between the fluidic cartridge and a commercial column.
fig. S3. The optical detection system in the PAD.
fig. S4. Snapshots of a PAD application.
fig. S5. The circuit diagram of the detection system.
fig. S6. The effect of dUTP on the PAD assay.
fig. S7. Portable PAD system.
fig. S8. Lyophilized probes.
fig. S9. Electrophoretic band-shift assay.
fig. S10. Detection of ARV factors.
fig. S11. Overall assay procedure for clinical samples.
fig. S12. Universal and species-specific detection of HAI pathogens in clinical samples by the PAD system.
fig. S13. Detection of ARV factors in clinical samples with PAD (top) and qPCR (bottom).
References (29, 30)

REFERENCES AND NOTES

1. S. S. Magill, J. R. Edwards, W. Bamberg, Z. G. Beldavs, G. Dumyati, M. A. Kainer, R. Lynfield, M. Maloney, L. McAllister-Hollod, J. Nadle, S. M. Ray, D. L. Thompson, L. E. Wilson, S. K. Fridkin; Emerging Infections Program Healthcare-Associated Infections and Antimicrobial Use Prevalence Survey Team, Multistate point-prevalence survey of health care–associated infections. N. Engl. J. Med. 370, 1198–1208 (2014).

2. A. Marchetti, R. Rossiter, Economic burden of healthcare-associated infection in US acute care hospitals: Societal perspective. J. Med. Econ. 16, 1399–1404 (2013).

3. B. Allegranzi, S. Bagheri Nejad, C. Combescure, W. Graafmans, H. Attar, L. Donaldson, D. Pittet, Burden of endemic health-care-associated infection in developing countries: Systematic review and meta-analysis. Lancet 377, 228–241 (2011).

4. R. A. Polin, S. Denson, M. T. Brady; Committee on Fetus and Newborn; Committee on Infectious Diseases, Epidemiology and diagnosis of health care–associated infections in the NICU. Pediatrics 129, e1104–e1109 (2012).

5. M. Klompas, D. S. Yokoe, R. A. Weinstein, Automated surveillance of health care–associated infections. Clin. Infect. Dis. 48, 1268–1275 (2009).

6. E. A. Mothershed, A. M. Whitney, Nucleic acid-based methods for the detection of bacterial pathogens: Present and future considerations for the clinical laboratory. Clin. Chim. Acta 363, 206–220 (2006).

7. K. S. Park, C. Y. Lee, H. G. Park, Target DNA induced switches of DNA polymerase activity. Chem. Commun. 51, 9942–9945 (2015).

8. P. D. Mauldin, C. D. Salgado, I. S. Hansen, D. T. Durup, J. A. Bosso, Attributable hospital cost and length of stay associated with health care-associated infections caused by antibiotic-resistant gram-negative bacteria. Antimicrob. Agents Chemother. 54, 109–115 (2010).

9. D. M. Sievert, P. Ricks, J. R. Edwards, A. Schneider, J. Patel, A. Srinivasan, A. Kallen, B. Limbago, S. Fridkin; National Healthcare Safety Network (NHSN) Team and Participating NHSN Facilities, Antimicrobial-resistant pathogens associated with healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. Infect. Control Hosp. Epidemiol. 34, 1–14 (2013).

10. H. Shao, J. Chung, K. Lee, L. Balaj, C. Min, B. S. Carter, F. H. Hochberg, X. O. Breakefield, H. Lee, R. Weissleder, Chip-based analysis of exosomal mRNA mediating drug resistance in glioblastoma. Nat. Commun. 6, 6999 (2015).

11. M. C. Longo, M. S. Berninger, J. L. Hartley, Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene 93, 125–128 (1990).

12. A. Borst, A. T. A. Box, A. C. Fluit, False-positive results and contamination in nucleic acid amplification assays: Suggestions for a prevent and destroy strategy. Eur. J. Clin. Microbiol. Infect. Dis. 23, 289–299 (2004).

13. E. W. Taggart, K. C. Carroll, C. L. Byington, G. A. Crist, D. R. Hillyard, Use of heat labile UNG in an RT-PCR assay for enterovirus detection. J. Virol. Methods 105, 57–65 (2002).

14. K. Hsieh, P. L. Mage, A. T. Csordas, M. Eisenstein, H. T. Soh, Simultaneous elimination of carryover contamination and detection of DNA with uracil-DNA-glycosylase-supplemented loop-mediated isothermal amplification (UDG-LAMP). Chem. Commun. 50, 3747–3749 (2014).

15. V. Marx, PCR heads into the field. Nat. Methods 12, 393–397 (2015).

16. P. Cornejo-Juárez, D. Vilar-Compte, C. Pérez-Jiménez, S. A. Ñamendys-Silva, S. Sandoval-Hernández, P. Volkow-Fernández, The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *Int. J. Infect. Dis.* **31**, 31–34 (2015).

17. B. Pichon, R. Hill, F. Laurent, A. R. Larsen, R. L. Skov, M. Holmes, G. F. Edwards, C. Teale, A. M. Kearns, Development of a real-time quadruplex PCR assay for simultaneous detection of *nuc*, Panton–Valentine leucocidin (PVL), *mecA* and homologue *mecA*$_{LGA251}$. *J. Antimicrob. Chemother.* **67**, 2338–2341 (2012).

18. J.-A. McClure, J. M. Conly, V. Lau, S. Elsayed, T. Louie, W. Hutchins, K. Zhang, Novel multiplex PCR assay for detection of the staphylococcal virulence marker Panton-Valentine leukocidin genes and simultaneous discrimination of methicillin-susceptible from -resistant Staphylococci. *J. Clin. Microbiol.* **44**, 1141–1144 (2006).

19. K. Zhang, J.-A. McClure, S. Elsayed, T. Louie, J. M. Conly, Novel multiplex PCR assay for simultaneous identification of community-associated methicillin-resistant *Staphylococcus aureus* strains USA300 and USA400 and detection of *mecA* and Panton-Valentine leukocidin genes, with discrimination of *Staphylococcus aureus* from coagulase-negative Staphylococci. *J. Clin. Microbiol.* **46**, 1118–1122 (2008).

20. K. L. Thong, M. Y. Lai, C. S. J. Teh, K. H. Chua, Simultaneous detection of methicillin-resistant *Staphylococcus aureus, Acinetobacter baumannii, Escherichia coli, Klebsiella pneumoniae and Pseudomonas aeruginosa* by multiplex PCR. *Trop. Biomed.* **28**, 21–31 (2011).

21. J. H. Son, S. Hong, A. J. Haack, L. Gustafson, M. Song, O. Hoxha, L. P. Lee, Rapid optical cavity PCR. *Adv. Healthcare Mater.* **5**, 167–174 (2016).

22. O. Piepenburg, C. H. Williams, D. L. Stemple, N. A. Armes, DNA detection using recombination proteins. *PLOS Biol.* **4**, e204 (2006).

23. S. Buch, Y. Sui, N. Dhillon, R. Potula, C. Zien, D. Pinson, S. Li, S. Dhillon, B. Nicolay, A. Sidelnik, C. Li, T. Villinger, K. Bisarriya, O. Narayan, Investigations on four host response factors whose expression is enhanced in X4 SHIV encephalitis. *J. Neuroimmunol.* **157**, 71–80 (2004).

24. H. J. Chung, C. M. Castro, H. Im, H. Lee, R. Weissleder, A magneto-DNA nanoparticle system for rapid detection and phenotyping of bacteria. *Nat. Nanotechnol.* **8**, 369–375 (2013).

25. K. S. Park, R. C. Charles, E. T. Ryan, R. Weissleder, H. Lee, Fluorescence polarization based nucleic acid testing for rapid and cost-effective diagnosis of infectious disease. *Chemistry* **21**, 16359–16363 (2015).

26. J. Aslanzadeh, Preventing PCR amplification carryover contamination in a clinical laboratory. *Ann. Clin. Lab. Sci.* **34**, 389–396 (2004).

27. W. U. Dittmer, A. Reuter, F. C. Simmel, A DNA-based machine that can cyclically bind and release thrombin. *Angew. Chem. Int. Ed.* **43**, 3550–3553 (2004).

28. J. A. Jordan, M. B. Durso, Real-time polymerase chain reaction for detecting bacterial DNA directly from blood of neonates being evaluated for sepsis. *J. Mol. Diagn.* **7**, 575–581 (2005).

29. M. E. Olson, T. K. Nygaard, L. Ackermann, R. L. Watkins, O. W. Zurek, K. B. Pallister, S. Griffith, M. R. Kiedrowski, C. E. Flack, J. S. Kavanaugh, B. N. Kreiswirth, A. R. Horswill, J. M. Voyich, *Staphylococcus aureus* nuclease is an SaeRS-dependent virulence factor. *Infect. Immun.* **81**, 1316–1324 (2013).

30. G. Lina, Y. Piémont, F. Godail-Gamot, M. Bes, M.-O. Peter, V. Gauduchon, F. Vandenesch, J. Etienne, Involvement of Panton-Valentine leukocidin–producing *Staphylococcus aureus* in primary skin infections and pneumonia. *Clin. Infect. Dis.* **29**, 1128–1132 (1999).

Science Advances

Rapid identification of health care–associated infections with an integrated fluorescence anisotropy system

Ki Soo Park, Chen-Han Huang, Kyungheon Lee, Yeong-Eun Yoo, Cesar M. Castro, Ralph Weissleder and Hakho Lee

ARTICLE TOOLS	http://advances.sciencemag.org/content/2/5/e1600300
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2016/05/03/2.5.e1600300.DC1
REFERENCES	This article cites 30 articles, 6 of which you can access for free http://advances.sciencemag.org/content/2/5/e1600300#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Performance of MiniPCR™ mini8, a portable thermal cycler

Han-Sol Kwon, Hyun-Chul Park, Kyungmyung Lee, Sanghyun An, Yu-Li Oh, Eu-Ree Ahn, Ju Yeon Jung and Si-Keun Lim★

Forensic DNA Division, National Forensic Service, Wonju 220-170, Korea

휴대용 DNA증폭기 MiniPCR™ mini8 Thermal Cycler의 성능 검토

권한솔 · 박현철 · 이경명 · 안상현 · 오유리 · 안으리 · 정주연 · 임시근★

국립과학수사연구원 법유전자과

Abstract: A small and inexpensive thermal cycler (PCR machine), known as the MiniPCR™ Mini8 Thermal Cycler (Amplyus, Cambridge, MA, USA), was developed. In this study, the performance of this PCR machine was compared with the GeneAmp® PCR system 9700 (Applied Biosystems) using four autosomal short tandem repeat (STR) kits, a Y-chromosome STR kit, and a mitochondrial DNA HV1/HV2 sequence analysis. The sensitivity and stochastic effects of the STR multiplex kits and the quality of the DNA sequence analysis were similar between the two PCR machines. The MiniPCR™ Mini8 Thermal Cycler could be used for analyses at forensic DNA laboratories and crime scenes. The cost of the PCR is so economical that school laboratories and individuals could use the machines.

요　약: 최근 손안에 들어올 정도로 크기가 작아 범죄현장 등에서 사용이 가능하며, 다른 일반적인 장비들에 비해 가격이 1/10이하로 저렴하여 누구나 사용할 수 있는 MiniPCR™ mini8 Thermal Cycler (Amplyus, Cambridge, MA, USA)가 개발되었다. 본 연구에서는 DNA감식에 일반적으로 사용되고 네 가지 종류의 상염색체 STR 다중증폭 키트들과 한 종류의 Y 염색체 STR 증폭키트, 그리고 미토콘드리아 DNA HV1/HV2의 염기서열 분석법을 사용하여 MiniPCR™ mini8 Thermal Cycler의 성능을 Applied Biosystems사의 GeneAmp® PCR system 9700와 비교하였다. STR 다중증폭키트 키트들의 민감도와 증폭 불균형 정도를 비교한 결과 두 PCR 장비에서 큰 차이가 없었으며, 미토콘드리아 DNA HV1/HV2의 염기서열 분석 결과도 동등하였다. MiniPCR™ mini8 Thermal Cycler는 DNA 감식 실험실은 물론이고, 가격이 저렴해 학교와 개인이 간편하게 사용할 수 있으며, 휴대가 간편해 차량이나 야외에서 활용 될 수 있을 것으로 기대된다.

Key words: MiniPCR™ mini8 Thermal Cycler, GeneAmp® PCR system 9700, performance, STR multiplex kits, mtDNA HV1/HV2

★ Corresponding author
Phone : +82-(0)33-902-5721　Fax : +82-(0)33-902-5941
E-mail : neobios@korea.kr

1. 서 론

1983년 Kary Mullis에 의해 개발된 중합효소연쇄반응(PCR; Polymerase chain reaction)은 DNA의 일부분을 증폭시키는 기술로서 분자생물학 발전에 큰 영향을 끼쳐왔다[1]. PCR은 사용 목적에 따라 다양한 분석방법이 개발되어 왔다.[2-4] DNA감식에 의한 개인식별과 신원확인은 법과학 분야에 혁명적 발전을 가져왔으며, 지문분석과 함께 막강한 도구로 자리 잡고 있다. 현재 STR좌위의 분석은 개인식별에 가장 유용하게 사용되는 DNA감식 수단이며,[5-7] PCR 증폭기는 성공적인 DNA감식을 결정하는 가장 중요한 장비 중 하나이다.

최근 미국의 스타트업 기업인 Amplyus사(Cambridge, MA, USA)에서 손 안에 들어올 정도로 크기가 작고,

Fig. 1. MiniPCR™ mini8 Thermal Cycler (source: http://www.minipcr.com)

스마트폰 어플을 이용해 PCR 증폭조건을 설정하고 모니터링 할 수 있으며, 기존의 일반적인 PCR장비에 비해 가격이 1/10정도로 저렴한 MiniPCR™ mini8 Thermal Cycler (이하 MiniPCR)를 개발하였다(*Fig.* 1).

본 연구에서는 MiniPCR의 성능을 현재 많은 DNA감식 실험실에서 사용 중인 GeneAmp PCR systems 9700 thermal cycler (Life Technologies, 이하 9700)와 비교해 보았다. 이를 위해 개인식별과 신원확인에 주로 사용되고 있는 상 염색체 STR 다중증폭키트인 AmpF/STR Identifiler amplification kit (이하 ID), AmpF/STR Identifiler Plus amplification kit (이하 ID plus), GlobalFiler™ PCR amplification kit(Applied Biosystems, Foster City, CA, USA, 이하 GF), PowerPlex® Fusion system (Promega, Madison, WI, USA, 이하 Fusion)과 Y 염색체 STR 다중증폭키트인 PowerPlex®Y23 system (Promega, Madison, WI, USA, 이하 Y23)를 사용하여 DNA 농도에 따른 증폭 민감도 및 대립유전자 증폭 불균형(Stochastic effect)정도를 분석하였다. 또한 분해된 시료 등과 모계 혈통 확인에 주로 사용되는 미토콘드리아 DNA의 HV1과 HV2의 염기서열을 분석하여 두 PCR 장비의 성능을 비교하였다.

2. 재료 및 방법

2.1. DNA 시료

본 실험에서는 DNA감식의 표준 DNA인 2800M DNA (Promega, Madison, WI, USA)를 사용하였다. DNA는 연속 2배 희석하여 1,000, 500, 250, 125, 62.5, 31.25 pg/μL의 농도로 준비하였다.

Table 1. Constitution of PCR mixtures and PCR protocols for STR Multiplex kits

		ID[a]	ID plus[b]	GF[c]	Fusion[d]	Y23[e]
Reagents (μL)	Master mix	6.3	8	7.5	5	5
	Primer Set	1.8	4	2.5	5	2.5
	Taq gold	0.3	-	-	-	-
	D.W.	4.6	6	13	13	15.5
	Target DNA	2	2	2	2	2
	Total Volume	15	20	25	25	25
PCR Protocol	Initial step	95 °C (11 min)	95° C (11 min)	95 °C (1 min)	96 °C (1 min)	9 5°C (2 min)
	Denaturation	94 °C (1 min)	94 °C (20 sec)	94 °C (10 sec)	94 °C (10 sec)	94 °C (10 sec)
	Annealing	59 °C (1 min)			59 (1 min)	61 °C (1 min)
	Extension	72 °C (1 min)	59 °C (3 min)	59° C (90 sec)	72 (30 sec)	72 °C (30 sec)
	Final	60 °C (60 min)	60 °C (10 min)	60 °C (10 min)	60 °C (10 min)	60 °C (20 min)

[a]AmpF/STR Identifiler amplification, [b]AmpF/STR Identifiler Plus amplification, [c]GlobalFiler™ PCR, [d]PowerPlex® Fusion system amplification, [e]PowerPlex®Y23 system

2.2. 증폭 및 모세관 전기영동

각 STR 다중증폭 키트들의 반응액 조성과 증폭 조건은 제조자의 지시에 따랐다(*Table* 1). 각 PCR 장비에서 증폭된 PCR 산물 1 μL에 GeneScan 600 LIZ size standard (Thermo Fisher Scientific, Waltham, MA, USA) 0.4 μL 그리고 Hi-Di Formamide (Thermo Fisher Scientific, Waltham, MA, USA) 9.6 μL를 혼합한 후, 3500 xl genetic analyzer (Thermo Fisher Scientific,

Waltham, MA, USA)에서 전기영동을 수행하였으며, GeneMapper ID-X 소프트웨어 v1.4 (Thermo Fisher Scientific, Waltham, MA, USA)와 엑셀 프로그램 (Microsoft)을 사용하여 결과를 분석 하였다.

2.3. 미토콘드리아 DNA HV1/HV2 염기서열 분석

미토콘드리아 DNA의 과변이부위(Hypervariable region)인 HV1과 HV2의 증폭과 염기서열 분석을 위

Fig. 2. PCR product (A) and plot of PowerPlex Fusion system(B) amplified by MiniPCR and 9700.

해 Gold ST★R 10X buffer (PROMEGA, Madison, WI, USA), AmpliTaq DNA polymerase 1.25 unit에 프라이머(F15910/R16410 및 F15/R484, 10 mM each)를 넣고 반응 시켰다.[8] PCR산물의 확인을 위해 Agilent DNA kits(Agilent Technologies, Waldbronn, Germany) 및 Agilent 2100 Bioanalyzer를 사용하였다.

3. 결   과

3.1. 민감도

MiniPCR과 9700을 이용해 STR 다중증폭키트를 각 농도 별 표준 DNA시료에서 증폭한 후, 피크의 높이 (peak height)와 대립유전자의 검출율을 비교하였다. Agilent 2100 Bioanaylzer를 이용해 확인한 PCR 증폭 산물과 Genetic analyzer와 GeneMapper를 이용해 분석한 STR plot은 *Fig.* 2와 같다. 실험에 사용한 네 가지 종류의 상 염색체 STR 다중증폭키트들은 모두 31.25 pg/μL 이상의 DNA에서 MiniPCR과 9700사이에 피크 높이에 큰 차이 없이 성공적으로 증폭되었다

(*Fig.* 3). 그러나 Y23 키트의 경우에는 MiniPCR에서 9700에 비해 상대적으로 낮은 피크 높이를 보여주었다. 또한 각 STR 다중 키트 별로 대립유전자 검출율을 비교한 결과 MiniPCR과 9700 사이에 큰 차이가 없었다(*Fig.* 4).

3.2. 대립유전자 증폭 불균형 분석

DNA의 농도가 낮을 경우에 나타나는 이형접합 대립유전자의 증폭 불균형 현상(stochastic effect)은 PCR 다중증폭키트는 물론이고 PCR 증폭기의 성능을 비교할 수 있는 기준이 된다. MiniPCR과 9700을 사용해 증폭한 다섯 가지의 STR 다중증폭키트 모두에서 이형접합 대립유전자 증폭 산물의 피크 및 좌위의 균형에 큰 차이가 없었다(*Fig* 5).

3.3. 미토콘드리아 DNA HV1/HV2 염기서열 분석

법과학 분야에서 미토콘드리아 DNA 분석은 오래되거나 변질되어 핵 DNA의 분석이 불가능할 경우나 신원확인을 위해 모계 혈통 분석이 필요한 경우에 사
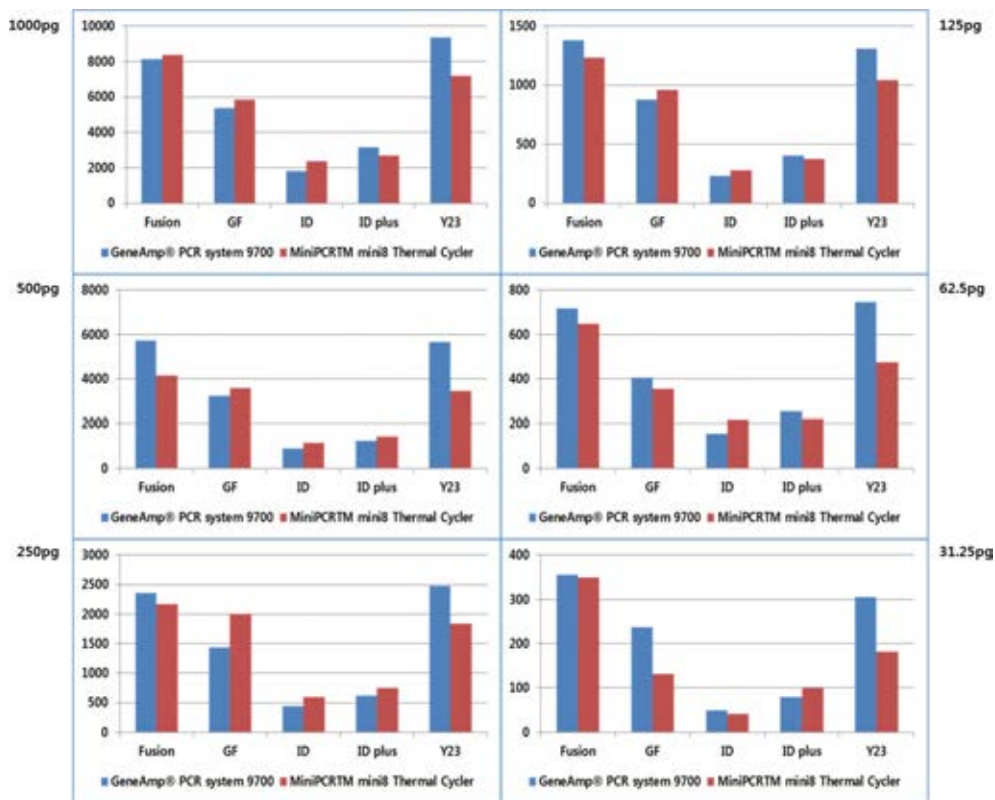
Fig. 3. Comparison of peak heights with regard to six STR multiplex kits using MiniPCR and 9700.
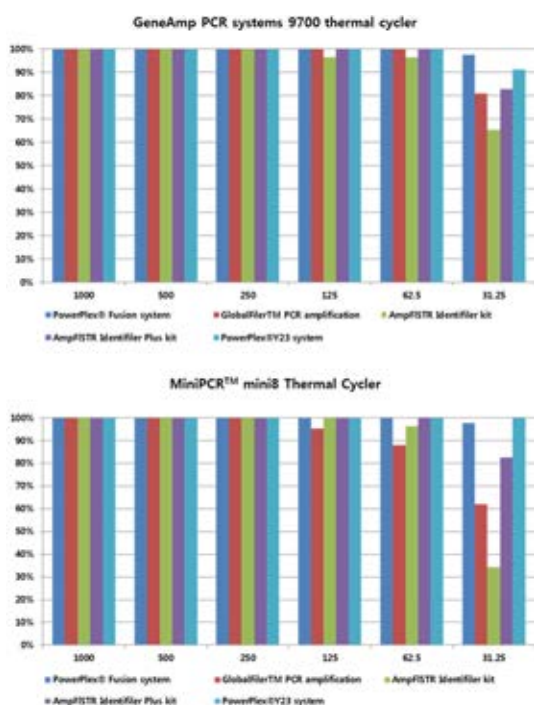
Fig. 4. Comparison of detected allele ratio with regard to five autosomal STR multiplex kits using MiniPCR and 9700.
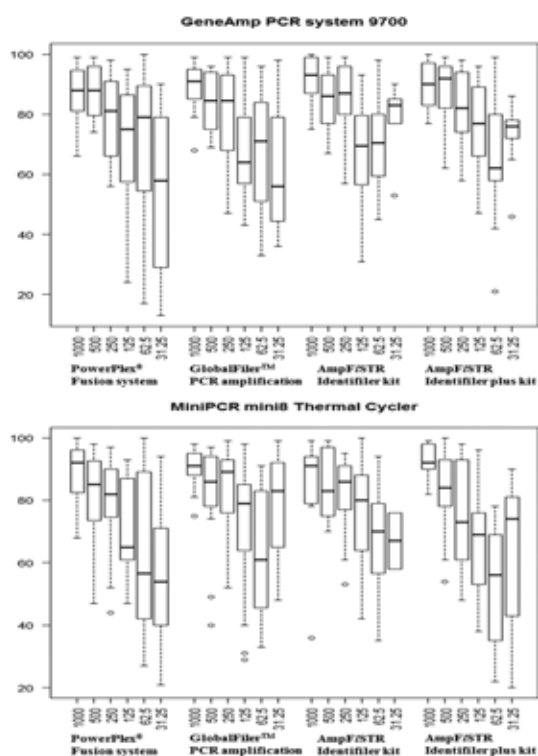

Fig. 5. Comparison of peak balance of heterozygous alleles from STR multiplex kits amplified by MiniPCR and 9700.
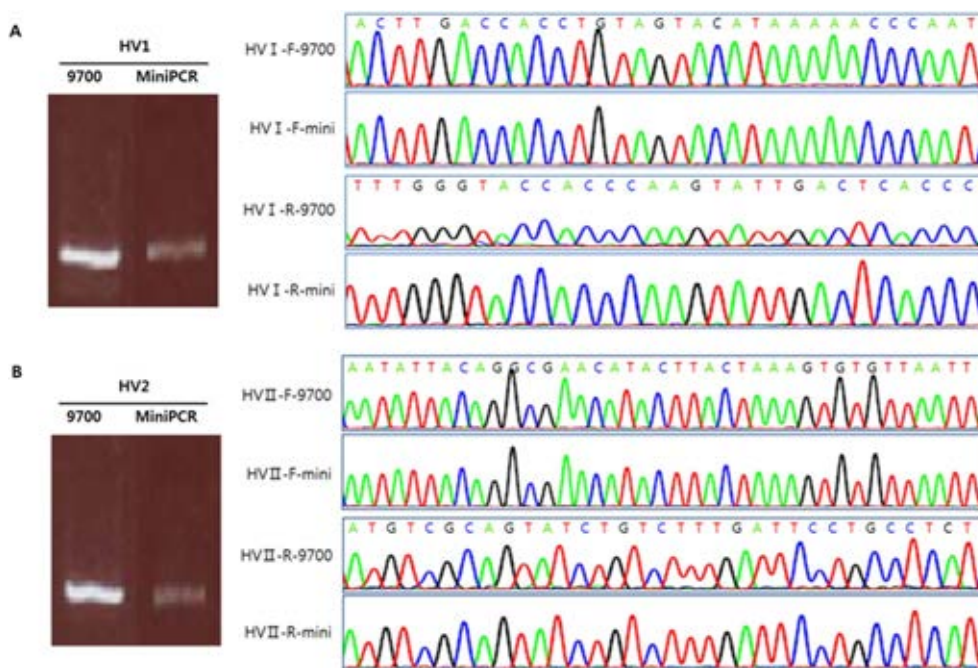

Fig. 6. PCR products and DNA sequence of mtDNA HV1(A), and HV2(B) amplified by MiniPCR and 9700.

용된다. 미토콘드리아 DNA의 HV1 (16024-16365)과
HV2 (73-340)은 사람마다 염기서열에 차이가 있는데,
염기서열 분석을 위해 먼저 PCR을 통해 두 부분을
증폭해야 한다. MiniPCR과 9700을 이용해 HV1과
HV2를 증폭한 후 염기서열을 분석해 비교한 결과, 모
두 동일한 염기서열을 얻을 수 있었다(*Fig.* 6).

4. 고 찰

　　PCR은 DNA 감식은 물론이고 거의 모든 생명과학
분야에서 필수적인 분석 방법이 되었다. 지금까지 수
많은 PCR 증폭기가 개발되어 실험실에서 사용되고
있는데, 대부분 가격이 비싸고 부피가 커서 실험실이
아닌 곳에서는 사용할 수 없었다. 본 연구에서는 작고
저렴한 MiniPCR 장비의 성능을 검토하기 위해 DNA
감식 실험실에서 개인식별과 신원확인을 하기 위해
일반적으로 수행되고 있는 STR 다중증폭과 DNA
염기서열 분석을 수행하여 기존의 일반적인 PCR 증
폭기와 비교하였다. 본 연구에 사용된 STR 다중증
폭키트들은 최소 15 개에서 23 개의 좌위를 동시에
증폭할 수 있는데, MiniPCR과 9700에서 거의 유사
한 결과를 얻을 수 있었다. 또한 두 장비를 사용해
미토콘드리아 DNA HV1/HV2의 염기서열 분석을
위한 PCR을 수행하여 그 결과를 비교한 결과도 큰
차이가 없었다. 본 연구에서 수행한 MiniPCR의 성
능은 가격과 작은 부피를 감안하면 현재 사용되고
있는 고가의 PCR 장비를 충분히 대체할 수 있으며,
나아가 범죄현장 등 야외에서도 유용하게 사용할 수
있을 것으로 사료 된다.

References

1. Bartlett, J. M. S. and Stirling, D. "A Short History of the Polymerase Chain Reaction". *PCR Protocols*. Methods in Molecular Biology, **226** (2nd ed.). pp. 3-6. doi: 10.1385/1-59259-384-4:3. ISBN 1-59259-384-4 (2003).
2. Newton, C. R. Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F., *Nucleic Acids Research*, **17**(7), 2503-2516 (1989).
3. Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L., *Gene*, **164**(1), 49-53 (1995).
4. Hayden, M. J., Nguyen, T. M., Waterman, A., and Chalmers, K. J., *BMC Genomics,* **9**, 80, doi: 10.1186/1471-2164-9-80 (2008).
5. Bartlett, J. M. *J. Forensic Sci.,* **51**, 253-265 (2006).
6. Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., and Chackraborty, R., *Am J. Hum. Genet*., **55**, 175-189 (1994).
7. Tautz, D., *Nucleic Acids. Res.*, **17**, 6463-6471 (1989).
8. Lee, E. J., Choung, C. M., Moon, S. H., Chung, J. Y., Jin, G. N., Kim, M. I., Lee, J. Y., Kim, S. H., and Lee, Y. H., *Korean J. Forensic Sci.*, **15**(1), 87-90 (2014).

npj | Microgravity

BRIEF COMMUNICATION   OPEN

Successful amplification of DNA aboard the International Space Station

Anna-Sophia Boguraev[1], Holly C. Christensen[2,3], Ashley R. Bonneau[4], John A. Pezza[5], Nicole M. Nichols[5], Antonio J. Giraldez[4], Michelle M. Gray[6], Brandon M. Wagner[6], Jordan T. Aken[6], Kevin D. Foley[6], D. Scott Copeland[6], Sebastian Kraves[7] and Ezequiel Alvarez Saavedra [ID][7]

As the range and duration of human ventures into space increase, it becomes imperative that we understand the effects of the cosmic environment on astronaut health. Molecular technologies now widely used in research and medicine will need to become available in space to ensure appropriate care of astronauts. The polymerase chain reaction (PCR) is the gold standard for DNA analysis, yet its potential for use on-orbit remains under-explored. We describe DNA amplification aboard the International Space Station (ISS) through the use of a miniaturized miniPCR system. Target sequences in plasmid, zebrafish genomic DNA, and bisulfite-treated DNA were successfully amplified under a variety of conditions. Methylation-specific primers differentially amplified bisulfite-treated samples as would be expected under standard laboratory conditions. Our findings establish proof of concept for targeted detection of DNA sequences during spaceflight and lay a foundation for future uses ranging from environmental monitoring to on-orbit diagnostics.

INTRODUCTION

On Earth, human health is continually improved by the use of molecular diagnostic techniques, however, these technologies remain largely untested in space conditions. It is well established that exposure to microgravity can result in profound impacts to the human body. For example, the human immune system's function is impacted in space, and may result in a decreased response to extracellular pathogens and an altered autoimmune response.[1,2] In order to preserve the health of astronauts during long-duration space missions, it will be important to rapidly detect molecular changes such as alterations in gene expression and epigenetic modifications. The polymerase chain reaction (PCR), a method to amplify DNA is routinely used on Earth to detect changes in DNA and gene expression and to diagnose infections among other applications.[3] In this investigation, we sought to determine the conditions under which PCR can be carried out in microgravity to enable DNA analysis and to establish the basis for a PCR-based assay to monitor crewmember health during long-term missions. This student-led investigation resulted from the winning proposal of the first Genes in Space Competition.[4]

RESULTS AND DISCUSSION

To test hardware performance and its capacity to amplify DNA under a variety of conditions in space, we conducted four runs in the miniPCR thermal cycler. The first "dry run" experiment, without biological samples, confirmed that the temperature profiles during thermal cycling were comparable to those obtained on Earth (Supplementary Fig. 1). The second and third experiments were used to determine the baseline conditions for

efficient DNA amplification, and the fourth experiment attempted to detect changes in DNA methylation patterns in genomic DNA. For each experiment duplicate reactions were prepared, stored under similar conditions, and amplified on Earth in parallel with those in the International Space Station (ISS).

The first experiment with a biological sample attempted to amplify plasmid DNA, as it provides a high-quality, low-complexity template that yields robust amplification on Earth. We conducted extensive stability studies of complete reactions on Earth and determined that reactions were viable after prolonged storage periods of at least 3 months (Fig. 1a), more than the time expected to lapse between sample preparation on Earth and operations at the ISS. "Complete" reactions, including template DNA, primers, polymerase, deoxynucleotides and reaction buffer were prepared on Earth and launched frozen on an ISS National Lab mission to the ISS[4] (Fig. 1b). Aboard the ISS, the miniPCR device was connected to the onboard computer in the maintenance work area (MWA) for the duration of the experiment and then stowed away for later use (Fig. 1c). To determine whether optimal conditions for PCR amplification differed in the ISS environment, we varied several experimental parameters. Changes in fluid dynamics such as reduced convection, and changes in surface tension in the ISS[5,6] might impact heat transfer throughout samples and liquid distribution in the reaction vessels, potentially impacting amplification efficiency due to incomplete DNA denaturation, poor primer annealing, or reduced polymerase efficiency. To determine whether the altered fluid dynamics would impact DNA amplification, we prepared reactions in total volumes of 12.5, 25, and 50 µl, which span the range of volumes typically used in laboratories. We also varied the amount of template DNA

[1]Yale University, New Haven, CT, USA; [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA; [3]Whitehead Institute, Cambridge, MA, USA; [4]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; [5]New England Biolabs, Inc., Ipswich, MA, USA; [6]Boeing, Houston, TX, USA and [7]miniPCR, Cambridge, MA, USA
Correspondence: Ezequiel Alvarez Saavedra (zeke@minipcr.com)

npj | nature partner journals

npj

2
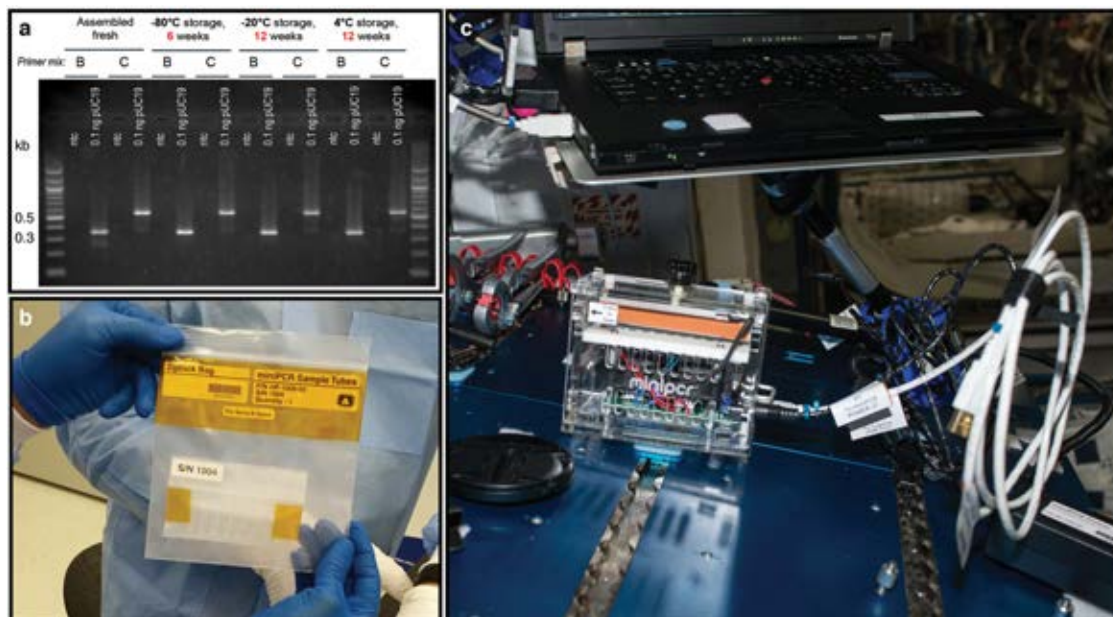
Successful amplification of DNA in space
A-S Boguraev et al.

Fig. 1 a Long-term stability studies. Complete reactions using Hot Start *Taq* (NEB) were prepared at the same time and stored at −80, −20, or 4 °C for either 6 or 12 weeks as indicated. Samples were then thawed or removed from the fridge, amplified using PCR and run on a 1.5% agarose gel. ntc: no template control. First and last lanes contain 0.25 μg 100 bp ladder (NEB). b Prepared eight-tube PCR strips were stored inside two zip bags, sealed with Kapton® tape and frozen. The samples remained frozen through all transportation steps until operations on the ISS. c Maintenance work area (MWA) on the ISS showing the miniPCR device connected to a laptop computer for programming and monitoring of experiments. Astronaut Tim Peake programmed and operated the miniPCR device
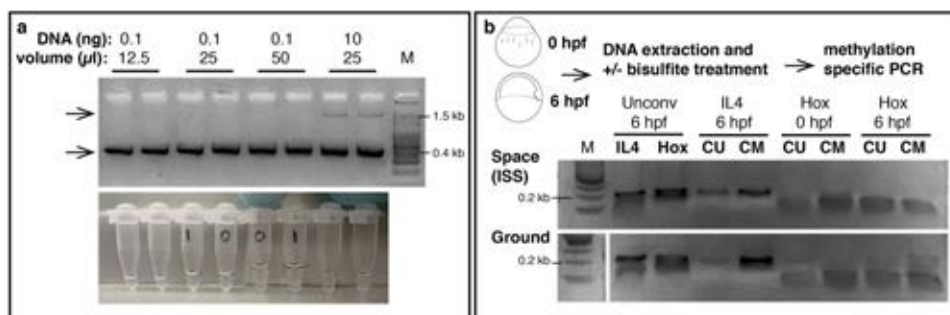

Fig. 2 Samples were returned to Earth frozen and run on a 2% agarose gel. a Top: All samples with plasmid DNA were successfully amplified on ISS regardless of sample volume or amount of template DNA. Bottom arrow indicates expected amplicon size. Top arrow indicates input DNA visible in samples with 10 nanograms starting DNA. Bottom: Samples derive from the same experiment and were processed in parallel to run on the gel. Photograph of tubes corresponding to samples in the gel, showing the different reaction volumes. b Top: Experimental overview: *Danio rerio* embryos were harvested at 0 h and 6 h post fertilization (hpf) and genomic DNA was purified. DNA for each timepoint was divided and either left untreated or treated with sodium bisulfite. PCR was performed on untreated (unconv) DNA in lanes 2 and 3 or bisulfite-treated DNA in lanes 4 through 8. CM: primers specific to sequences that were methylated. CU: primers specific to sequences that were not methylated. DNA in lanes 2 and 3 was amplified with CU primers. Lane 1 contains 0.25 μg 100 bp ladder (NEB). Samples derive from the same experiment and were processed in parallel to run on the gel

between 0.1 and 10 ng, and utilized two different DNA polymerases, *Taq* polymerase and Q5® polymerase. We found that DNA amplification in the ISS was successful under all experimental conditions tested (Fig. 2a), indicating that PCR can be conducted aboard the ISS without modification of standard protocols. Since the Genes in Space-1 studies were performed, two other independent efforts, NASA WetLab-2 and Water Monitoring Suite, have also amplified DNA aboard the ISS (unpublished).

It is becoming increasingly clear that altered epigenetic profiles can lead to a variety of disorders, and that spaceflight produces measurable changes to the epigenome.[7,8] An assay to evaluate epigenetic changes such as DNA methylation during spaceflight would aid our understanding of the effects of spaceflight on the epigenome and would allow monitoring of crewmembers' health.

For example, such an assay could aid in the detection of immune system alterations that may to lead to increased susceptibility to autoimmune disease, allergies, and other diseases in space.[9,10] The standard method used to evaluate the methylation profile of DNA is bisulfite conversion.[11] During bisulfite treatment unmethylated cytosines are converted to uracil through deamination, while methylated cytosines are protected from the conversion. These resulting alterations can be revealed using primers specific to converted or non-converted sequences.[12] We sought to determine whether such epigenetic changes could be detected using PCR technology on the ISS through the use of methylation-specific primers. We used zebrafish embryos as an experimental system because its methylation patterns have been studied in detail at a genome-wide level.[13] We selected to study the promoters of *Hoxb3a*, a developmental gene, and *Interleukin 4* (*IL4*), an immune

Successful amplification of DNA in space
A-S Boguraev et al.

npj

3

gene, as the methylation patterns for these genes have been shown to change in embryos between 0 and 6 h post-fertilization. DNA for each time point was prepared, divided and either left untreated or treated with sodium bisulfite. Complete amplification reactions to detect converted and non-converted DNA were prepared on Earth and sent frozen to ISS. PCR performed on untreated DNA (Fig. 2b, lanes 2 and 3) resulted in robust amplification of *IL4* and *Hox* in space samples and ground controls. At 6 hpf *IL4* space-amplified samples with primers specific to sequences that were methylated (Fig. 2b, top gel, lane 5) amplified more robustly than samples with primers specific to sequences that were not methylated (Fig. 2b, top gel, lane 4). These results mirror the ones from samples amplified on Earth (Fig. 2b, bottom gel, lanes 4 and 5), suggesting that methylation-specific PCR can be conducted in space without modifications. Reactions targeting the Hox promoter that contained bisulfite-treated DNA resulted in poor amplification in both space samples and ground controls (Fig. 2b, lanes 6–9), likely due to the increased fragility of bisulfite-treated DNA.[14]

Taken together, our experiments reveal that PCR can be robustly carried out in a microgravity environment without adaptations to the hardware, consumables, or reagents, opening the door to DNA analysis in space. Additionally, we found that differences in DNA methylation are detectable in this microgravity environment. Furthering our ability to evaluate molecular changes in space is essential to both our understanding of the effects of spaceflight on the human body and the maintenance of astronaut health to enable long-term missions. Recent reports of DNA sequencing aboard the ISS using samples prepared on Earth,[15] in combination with the successful on-orbit DNA amplification reported in this article, suggest that a complete sample-to-sequence DNA analysis workflow in space will be plausible in the near future.

MATERIALS AND METHODS

Equipment and consumables used for PCR

The mini8 miniPCR is manufactured by miniPCR (www.minipcr.com). The 0.2 ml PCR tubes were purchased from Eppendorf (Cat. No. 0030124359) and consist of 8-tube strips with hinged lids.

Long-term stability studies

"Complete" reactions containing 25 μl Q5® Hot Start High-Fidelity 2X Master Mix (NEB M0494), 5 μl primer mix (5 μM each), 19 μl water and 1 μl 0.1 ng/μl pUC19 (see below) were stored at the indicated temperatures for 6 or 12 weeks. pUC19 was replaced for water in no template controls. Samples were then thawed and immediately amplified. Two sets of primers, B and C, were tested (primer sequences in Supplementary Methods). Primer set B was selected for experiments conducted in ISS and ground controls. pUC19 was obtained in purified form from New England Biolabs (N3041). pUC19 is isolated from *E. coli* ER2272 (dam+ dcm+ EcoK M−) by a standard plasmid purification procedure. Amplification conditions for long-term stability studies were as following: 98 °C/30 s [98 °C/15 s, 64 °C/15 s, 72 °C/60 s]x30, 72 °C/5 min.

Space plasmid amplification

For space samples, complete reactions were prepared as described above and placed in 0.2 ml PCR tubes in 12.5, 25, and 50 μl aliquots as indicated. Samples were kept frozen until operations aboard the ISS. Amplification conditions for space samples and ground controls were as following: 95 °C/ 30 s [95 °C/15 s, 50 °C/15 s, 68 °C/60 s]x30, 68 °C/5 min.

DNA purification from Danio rerio embryos

Zebrafish lines were maintained in accordance with AAALAC research guidelines, under a protocol approved by Yale University IACUC. Embryos were obtained from natural crosses of TU-AB and TLF strains, with mating pairs selected from random pools of wild-type females and males.

DNA was extracted from embryos using previously described nuclei isolation protocols.[13] In brief, embryos were first chemically dechorionated (Pronase 1 mg/ml) and allowed to develop till the desired stage. Batches of embryos were disrupted in 1 ml lysis buffer (10 mM Tris-Cl [pH 8.0], 10 mM NaCl, 0.5% NP-40) using a 20 G needle. Nuclei were collected by centrifugation at 3500×g for 5 min at 4 °C. Unless otherwise noted, 500 and 100 embryos were used for 0 hpf and 6 hpf, respectively. Nuclei were resuspended in 1 ml nuclei lysis buffer (50 mM Tris-Cl [pH 8.0], 10 mM EDTA, 1% SDS) with 1 μl of RNase Cocktail (100 mg/ml, QIAGEN) and 10 μl Proteinase K Solution (Invitrogen). Following a 10-minute incubation at 55 ° C, equal volume phenol:chloroform (Invitrogen) was added and samples centrifuged at 4 °C for 30 min at 14,000 rpm. The aqueous phase was collected and transferred to a new microcentrifuge tube. To further reduce the large maternal messenger RNA contribution, 1 μl of RNase Cocktail was added and samples incubated for at least 3 h at 37 °C. A second phenol:chloroform extraction was performed as detailed above and DNA precipitated by adding 1.2 volume of isopropanol and 1/10 volume 3 M sodium acetate at −80 °C for one hour. Samples were centrifuged at 4 °C for 30 min at 14,000 rpm and the supernatant discarded. Afterwards, the DNA pellet was twice washed in 70% ethanol and spun at 14,000 rpm for 15 min at 4 °C. The DNA pellet air dried for 10 min at room temperature and was resuspended in nuclease-free water.

Bisulfite treatment

Sodium bisulfite treatment was performed using the EZ DNA methylation lightning reagents (Zymo) according to the manufacturers recommendations. Briefly, 500 ng of DNA was denatured and converted by incubating at 98 °C for 8 min followed by a 54 °C, 60 min step using the conversion reagent. Desulphonation and cleanup of converted DNA was performed using a DNA-binding spin column. Converted DNA was eluted in 10 μl of nuclease-free water.

Methylation-specific PCR of Danio rerio genomic DNA

"Complete" reactions containing 12.5 μl Q5® dU Bypass Master Mix (2x) (NEB M0598), 5 μl primer mix (5 μM each), 6.5 μl water and 1 μl genomic DNA. Samples were kept frozen until operations aboard the ISS. Amplification conditions for space samples and ground controls were as following: 94 °C/60 s [94 °C/10 s, 63 °C/30 s, 72 °C/30 s]x35, 72 °C/2 min.

Sample preparation and ISS operations

The Genes in Space-1 samples flew to the ISS on a Space X Dragon vehicle mated to a Falcon rocket. The samples were prepared on the ground and then delivered frozen to the Kennedy Space Center (KSC) as "Late Load." Samples were kept frozen during transit using −20 °C phase change cold packs (Cryopak) and were delivered to KSC approximately 3 days before launch where they were kept in the POLAR freezer on the Dragon vehicle with a set point of −35 °C. At ISS samples were placed in the MELFI freezer at −95 °C and stored until the miniPCR run would take place.

During operations, ISS crew removed the sample from the freezer and allowed 5 min for it to thaw before placing in the miniPCR. Parameters for thermal cycling were programmed by crew through the onboard computer and uploaded to the miniPCR via a USB cable. After upload the operations continued unattended. When the run was complete, the miniPCR data from the run was saved and all power was removed. After 30 min to allow for cool down, the sample was removed from the miniPCR and placed back into an ISS freezer (GLACIER), where it stayed until loaded back into a Space X Dragon for return. Samples were returned in an unpowered coldbag that was held at +4 °C via coldbricks. Upon splash-down, the vehicle was unloaded and the samples were put into a freezer for return to the Johnson Space Center, then turned over to the Boeing team and packaged in Fedex cold storage shipper (ice packs) for return to miniPCR for final analysis.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

npj

Successful amplification of DNA in space
A-S Boguraev et al.

4

England Biolabs and Math for America for their support of the Genes in Space competition that led to this study. This study was funded by miniPCR and Boeing.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Microgravity* website (https://doi.org/10.1038/s41526-017-0033-9).

Competing interests: E.A.S. and S.K. are employed by miniPCR, manufacturers of the device used for DNA amplification. N. N. and J.P. are employed by New England Biolabs, manufacturer of the amplification reagents described in this manuscript. The remaining authors declare no competing financial interests.

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

Change history: A correction to this article has been published and is linked from the HTML version of this article.

REFERENCES

1. Mermel, L. A. Infection prevention and control during prolonged human space travel. *Clin. Infect. Dis.* **56**, 123–130 (2013).
2. Sonnenfeld, G. & Shearer, W. T. Immune function during space flight. *Nutrition* **18**, 899–903 (2002).
3. Powledge, T. M. The polymerase chain reaction. *Adv. Physiol. Educ.* **28**, 44–50 (2004).
4. Genes in Space-1. https://www.nasa.gov/mission_pages/station/research/experiments/1913.html. (Accessed: 2nd July 2017)
5. Straub, J. The role of surface tension for two-phase heat and mass transfer in the absence of gravity. *Exp. Therm. Fluid Sci.* **9**, 253–273 (1994).
6. Napolitano, L. G. Materials: marangoni convection in space microgravity environments. *Science* **225**, 197–198 (1984).
7. Ou, X. et al. Spaceflight-induced genetic and epigenetic changes in the rice (*Oryza sativa L.*) genome are independent of each other. *Genome* **53**, 524–532 (2010).
8. Tauber, S., Yi, B., Chouk r, A. & Ullrich, O. *Effect of Spaceflight and Spaceflight Analogue Culture on Human and Microbial Cells*, Vol. 7, 121–129 (Springer, New York, 2016).
9. Yi, B., Crucian, B., Tauber, S., Ullrich, O. & Chouk r, A. *Effect of Spaceflight and Spaceflight Analogue Culture on Human and Microbial Cells,* Vol. 57, 61–79 (Springer, New York, 2016).
10. Taylor, G. R., Konstantinova, I., Sonnenfeld, G. & Jennings, R. in *Adv. Space Biol. Med*. **6**, 1–32 (1997).
11. Zuo, T., Tycko, B., Liu, T. -M., Lin, J. -J. L. & Huang, T. H. -M. Methods in DNA methylation profiling. *Epigenomics* **1**, 331–345 (2009).
12. Herman, J. G. Approaches, Methods, and Applications. in *DNA Methylation,* (ed. Manel Esteller) 65–72 (CRC Press, 2004). https://doi.org/10.1201/9780203487013.ch5.
13. Potok, M. E., Nix, D. A., Parnell, T. J. & Cairns, B. R. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**, 759–772 (2013).
14. Mill, J. & Petronis, A. Profiling DNA Methylation from Small Amounts of Genomic DNA Starting Material: Efficient Sodium Bisulfite Conversion and Subsequent Whole-Genome Amplification in *DNA Methylation Methods in Molecular Biology* (ed. Tost J, Vol. 507) 371–391 (Humana Press, 2009).
15. McIntyre, A. B. R. et al. Nanopore sequencing in microgravity. *Npj Microgravity* **2**, 131 (2016).