

## RESEARCH

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

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

<sup>1</sup>Department of Integrative Biology, University of California, Berkeley, CA, USA, <sup>2</sup>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, <sup>3</sup>Richard Gilder Graduate School, American Museum of Natural History, New York, USA, <sup>4</sup>Department of Herpetology, American Museum of Natural History, New York, USA, <sup>5</sup>Tropical Herping, Quito, Ecuador, <sup>6</sup>Centro Jambatu de Investigación y Conservación de Anfibios, Fundación Otonga, Quito, Ecuador, <sup>7</sup>Doc Frog Expeditions, Uvita, Costa Rica and <sup>8</sup>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](mailto:pomerantz.aaron@berkeley.edu)  <http://orcid.org/0000-0002-6412-9001>; Stefan Prost, E-mail: [stefanprost.research@protonmail.com](mailto: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

Received: 17 December 2017; Revised: 13 February 2018; Accepted: 22 March 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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

**Keywords:** nanopore sequencing; portable; DNA barcoding; biodiversity; field-based; real-time

## 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<sup>2</sup> (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-hectare-protected 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  $\mu$ L ddH<sub>2</sub>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  $\mu$ L of PCR product, 2.5  $\mu$ L 10X PCR buffer, 1  $\mu$ L 25mM MgCl<sub>2</sub>, 200  $\mu$ M dNTP mix, 0.2  $\mu$ M of each primer, and 0.25 Platinum Taq DNA Polymerase (Thermo Fisher Scientific) in a 25  $\mu$ L 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  $\mu$ L of ONT PCR Barcode, 2  $\mu$ L of first-round PCR product, 41  $\mu$ L H<sub>2</sub>O, and 50  $\mu$ L PCR reaction mix (0.5  $\mu$ L Taq DNA polymerase, 1  $\mu$ L dNTP mix, 2  $\mu$ L MgCl<sub>2</sub>, 41  $\mu$ L H<sub>2</sub>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 Exonuclease 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  $\mu$ L 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, New England Biolabs), then purified using AMPure XP beads. Adapter ligation and tethering was then carried out with 20  $\mu$ L of Adapter Mix (ONT) and 50  $\mu$ L of 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  $\mu$ L of Elution Buffer (ONT). Each R9 flow cell was primed with 1,000  $\mu$ L of a mixture of Fuel Mix (ONT) and nuclease-free water. Then 12  $\mu$ L of the amplicon library was diluted in 75  $\mu$ L of running buffer with 35  $\mu$ L RBF, 25.5  $\mu$ L LLB, and 2.5  $\mu$ L 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](https://doi.org/10.1093/bioinformatics/btt111)). 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](https://doi.org/10.1093/bioinformatics/btt111)) [8, 34] to align the reads to the reference, samtools 1.3 (SAMTOOLS, [RRID:SCR.002105](https://doi.org/10.1093/bioinformatics/btt111)) [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](https://doi.org/10.1093/bioinformatics/btt111)) [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 bp and



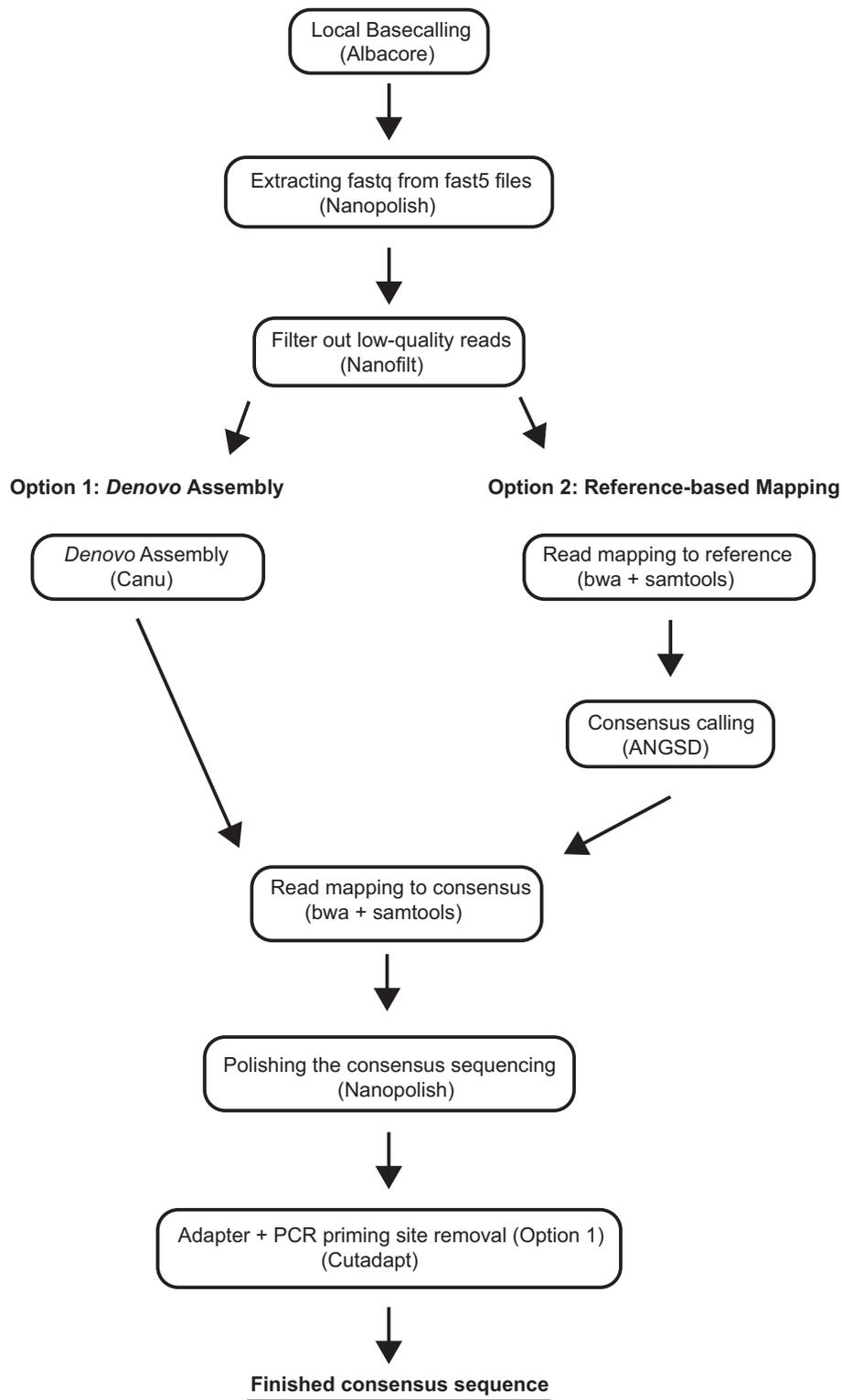
**Figure 1:** Process of nanopore sequencing in the Ecuadorian Chocó rainforest. (A) Sampling endemic fauna; eyelash pitviper next to MinION. (B) Extraction of blood or tissue samples. (C) DNA extraction using the DNeasy kit and benchtop centrifuge, and PCR amplification with the MiniPCR. (D) Oxford 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](https://doi.org/10.1093/bioinformatics/btt103)) [7]. Adapters were removed using *Cutadapt* (*cutadapt*, [RRID:SCR\\_011841](https://doi.org/10.1093/bioinformatics/btt103)) [12]. The consensus 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](https://doi.org/10.1093/bioinformatics/btt103)) [13] and *AliView* (*AliView*, [RRID:SCR\\_002780](https://doi.org/10.1093/bioinformatics/btt103)) [14]. Sanger

sequencing reads were edited and assembled using *Geneious R10* software (*Geneious*, [RRID:SCR\\_010519](https://doi.org/10.1093/bioinformatics/btt103)) [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](https://doi.org/10.1186/1471-2107-14-315)) [18] and reconstructed their phylogenetic trees using the maximum likelihood approach implemented in Mega 5 [19] with 1,000 bootstrap repli-

cates (for bioinformatics workflow see Fig.2). The output tree files, including the accession numbers, are provided in the Supplementary Material.

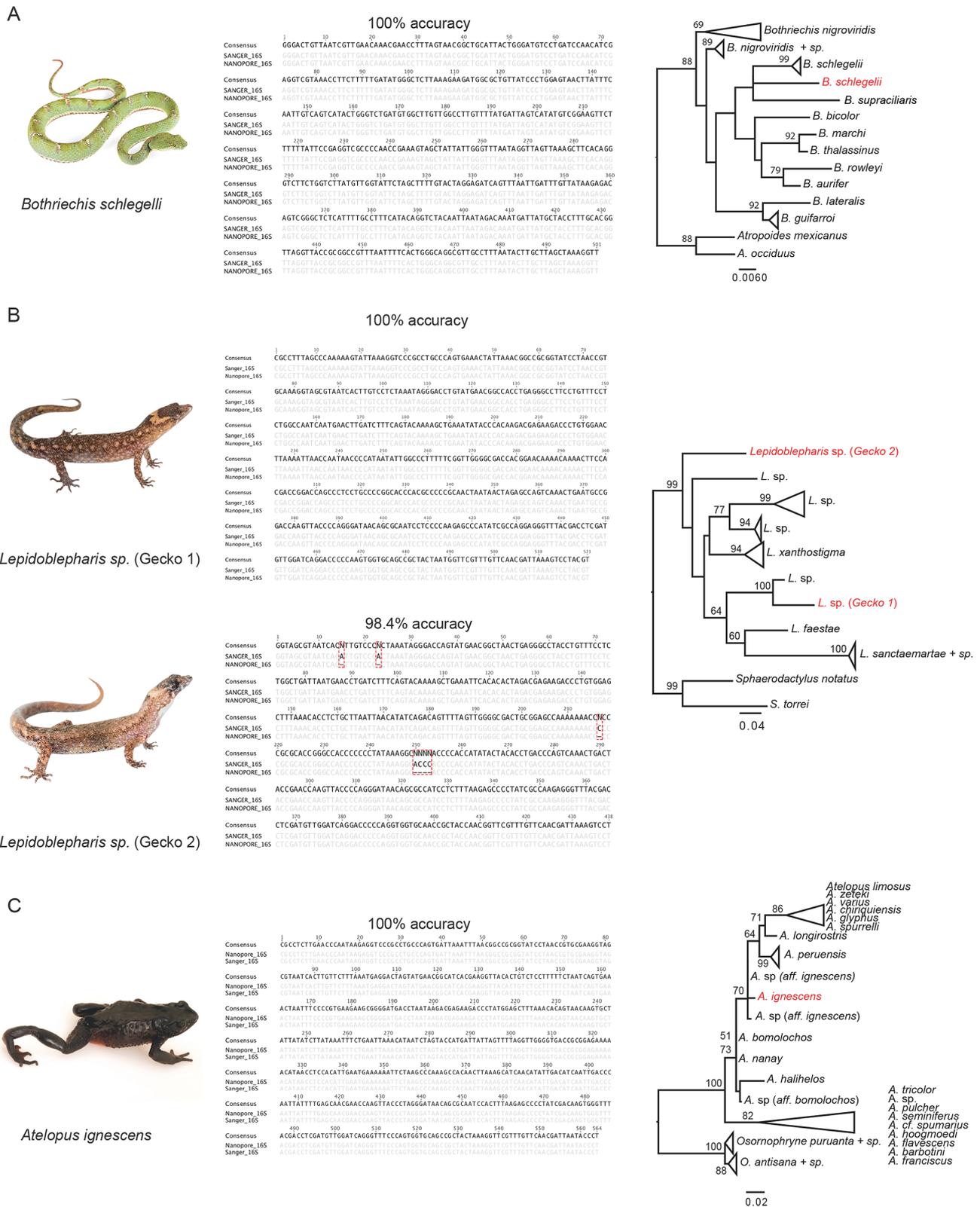
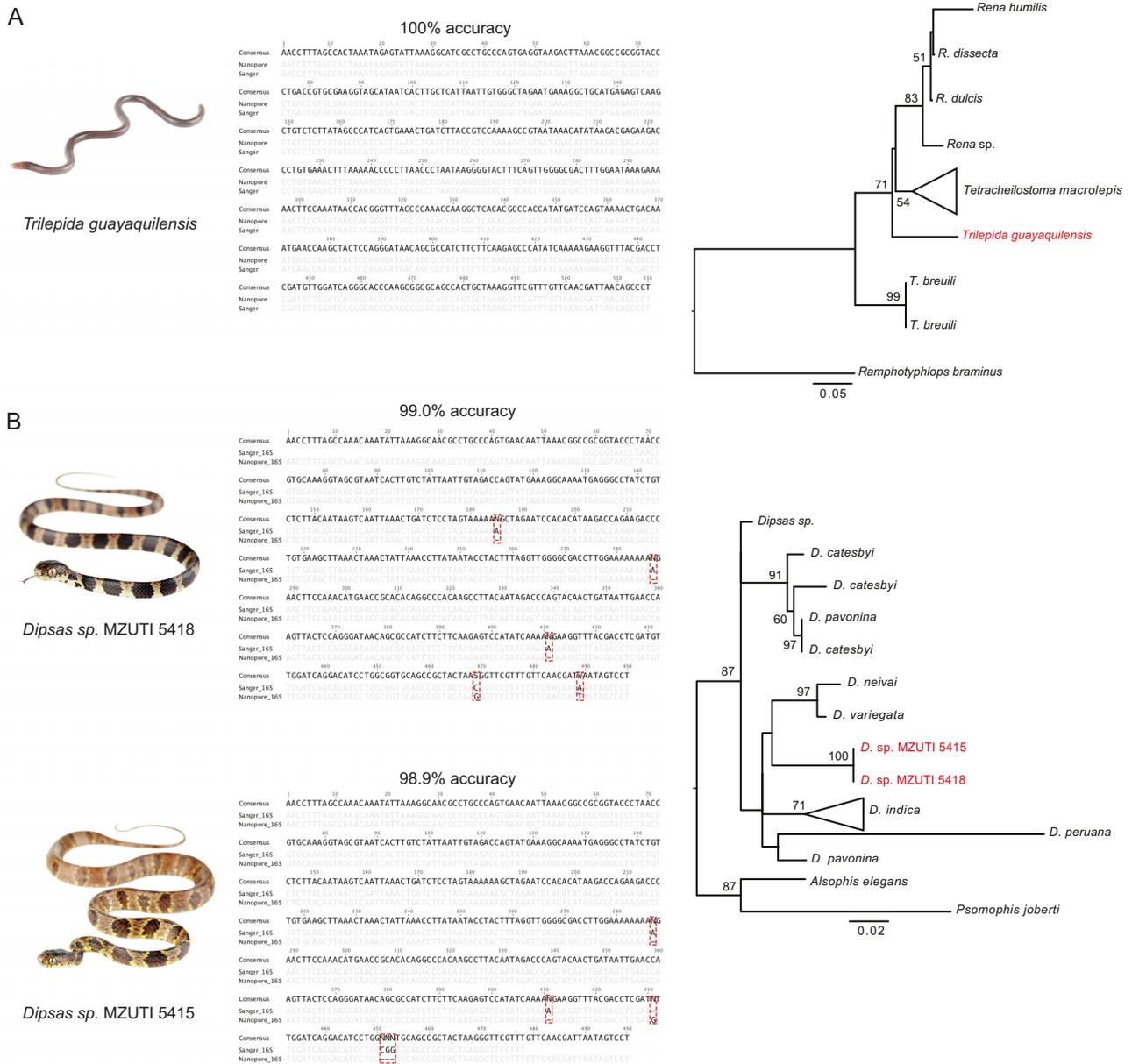


Figure 3: Species investigated, nucleotide alignments of nanopore and Sanger sequences comparing consensus accuracy, and maximum likelihood trees of 16S sequences for (A) eyelash pitviper, *Bothriechis schlegelii*, (B) two species of dwarf gecko, *Lepidoblepharis sp.*, and (C) the Jambato toad, *Atelopus 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 north-western South America [3]. One individual was captured on the evening of the 11 July 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 bp including 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 bp and 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 consensus 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 bp mismatch instead of the 4 bp indel at position 250–253 in the filtered, but not polished, nanopore consensus sequence. After polishing, all sequences (filtered or unfiltered) showed the 4 bp indel. 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 bp consensus 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 bp consensus 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 blasted to *Atelopus*. One read belonged to the positive control, and 53 showed insignificantly short hits.

*Dipsas oreas* (MZUTI 5415) We generated 487 reads for *Dipsas oreas* (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 bp consen-

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 \$1,000 and 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.

Supplementary 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

A.P. and S.P. designed the project. A.P., N.P., A.A., L.B., F.P., C.B., D.S.V., and S.P. carried out specimen collection. A.P. and N.P. performed the laboratory work. A.A., L.B., F.P., L.C., C.B., and D.S.V. performed morphological species identification and SP computational analyses. A.P., N.P., A.A., L.B., F.P., L.C., C.B., D.S.V., and S.P. wrote the paper.

### Acknowledgements

Fieldwork for this project was made possible with the support of Tropical Herping and Universidad Tecnológica Indoamérica. For granting access to the Canandé reserve, we are grateful to Martín Schaefer of Fundación Jocotoco. Funding for equipment and reagents was supported by the National Geographic Society/Waite grant (W412-15). Travel funding for Stefan Prost was provided by the Program of Conservation Genomics at Stanford University. Laboratory work was carried out at Universidad Tecnológica Indoamérica in Quito. Research and collection permits were issued by the Ministerio del Ambiente de Ecuador (MAE-DNB-CM-2015-0017). The Jambato toad tissue was provided by the Museum of Centro Jambatu under the Ministerio del Ambiente de Ecuador project "Conservation of Ecuadorian amphibian diversity and sustainable use of its genetic resources." We thank Oxford Nanopore Technologies for providing technical support, making the offline MinKNOW software available, and for providing two free flow cells. We also thank Hitomi Asahara and the UC Berkeley DNA Sequencing Facility for lending the bench-

top centrifuge used in this study; Jared Simpson, Sergey Koren, and Adam Phillippy for very helpful advice and discussion on the bioinformatic pipeline; and Ellie E. Armstrong for valuable input on the manuscript.

## References

- Erwin TL. Tropical forests their richness in coleoptera and other arthropod species. *The Coleop Bull* 1982;36:74–75.
- Stork NE. How many species are there? *Biodivers Conserv* 1993;2(3):215–32.
- 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.
- 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.
- Cuesta F, Peralvo M, Merino-Viteri A, et al. Priority areas for biodiversity conservation in mainland Ecuador. *Neotropical Biodiversity* 2017;3(1):93–106.
- Gascon C, Brooks TM, Contreras-MacBeath T, et al. The importance and benefits of species. *Current Biology* 2015;25(10):R431–8.
- 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.
- 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.
- 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.
- 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.
- 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.
- Carstens BC, Pelletier TA, Reid NM, et al. How to fail at species delimitation. *Mol Ecol* 2013;22(17):4369–83.
- 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.
- Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. *Syst Biol* 2005;54(5):852–9.
- 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.
- [www.barcodeoflife.org](http://www.barcodeoflife.org).
- Gilbert N. Biodiversity law could stymie research. *Nature* 2010;463(7281):598.
- Fernández F. The greatest impediment to the study of biodiversity in Colombia. *Caldasia* 2011;33:3–5.
- 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.
- Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 2014;14(6):1097–102.
- Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;530(7589):228–32.
- Faria NR, Sabino EC, Nunes MR, et al. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med* 2016;8(1):97.
- 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.
- Edwards A, Debonnaire 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>.
- 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.
- 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>.
- 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.
- 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.
- 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.
- <http://www.fjocotoco.org/canandeacute1.html>.
- Albacore GitHub Repository, <https://github.com/dvera/albacore>.
- Nanopolish GitHub Repository, <https://github.com/jts/nanopolish>.
- Nanofilt GitHub Repository Nanofilt GitHub Repository, <http://github.com/wdecoester/nanofilt>.
- BWA GitHub Repository <https://github.com/lh3/bwa/releases>.
- Canu readthedocs documentation, <https://canu.readthedocs.io>.
- Famas GitHub Repository, <https://github.com/andreas-wilm/famas>.
- Myers N, Mittermeier RA, Mittermeier CG, et al. Biodiversity hotspots for conservation priorities. *Nature* 2000;403(6772):853–8.
- Dinerstein E. A Conservation Assessment of the Terrestrial Ecoregions of Latin America and the Caribbean. Washington, D.C.: World Bank. 1995;xvii:129.
- 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.
- Ron S, Coloma LA, Lötters S, Duellman W, et al. Bustamante, wilmar bolivar, enrique la marca. 2004 *Atelopus ignescens*. The IUCN Red List of Threatened Species 2004: e.T54518A11157432.
- Coloma LA. El Jambato negro del páramo, *Atelopus ignescens*, resucitó. 2016 *IMCiencia*. [otonga.org/?p=438](http://otonga.org/?p=438)
- 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.
- 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>.