RESEARCH ARTICLE

Rapid on-site universal vertebrate species identification via multi-barcode nanopore sequencing

Emily C. Patterson[1], Sandrella Morrison-Lanjouw[2,3], Mark A. Jobling[1]*,
Jon H. Wetton[1]*

1 Division of Genetics & Genome Biology, School of Biological & Biomedical Sciences, University of Leicester, United Kingdom, 2 University of Utrecht, Julius Center, Utrecht, The Netherlands, 3 Institute for Risk Assessment Sciences (IRAS), Utrecht, The Netherlands

* maj4@le.ac.uk (MAJ); jw418@le.ac.uk (JHW)

Abstract

The growing illegal wildlife trade (IWT) threatens biodiversity and is a conduit for zoonotic disease, yet its risk of detection is low. Once processed, trafficked species are difficult to identify morphologically, and currently require DNA-based approaches that are time-consuming, costly, and lab-based. There is thus a need for a rapid, cheap, on-site method for species identification. We describe VeRIF-ID (Vertebrate Rapid In-Field Identification via DNA), a method that employs simultaneous on-site nanopore sequencing of four different mitochondrial DNA barcodes. Primers were designed to produce short amplicons to aid analysis of damaged DNA, and to be effective over a broad taxonomic range of vertebrates from lamprey to chimpanzee. Validation demonstrated species-level identification in 91% of 83 tested species, and genus/tribe-level identification of the remaining species (which are also problematic with existing approaches). DNA extraction, PCR and library preparation steps were simplified and optimised so that sampling to species identification takes <3 h for a single sample. Species components are identifiable non-quantitatively in prepared mixtures of muscle tissue from up to five species, and laboratory tests of Traditional East Asian Medicine samples reveal DNA from species including critically endangered saiga antelope and black rhinoceros. In conjunction with a portable Bento Lab device the necessary equipment and reagents are easily portable, and we apply the method to analyse seized bushmeat and fish samples within an airport customs zone, identifying mammal and fish species in 15 samples within 6 h. The initial equipment costs for VeRIF-ID are ~$8K, and the cost per sample of ~$10–48 (depending on set-up), considerably cheaper than current conventional lab-based approaches. The method requires only basic hands-on skills. Ongoing trials with potential end-users will focus on establishing forensic reporting criteria prior to casework implementation. Future development of user-friendly bioinformatic interfaces will aim to fully democratise species identification.

Introduction

The trading of wildlife for food (bushmeat), traditional medicinal ingredients, and as pets, involves approximately a quarter of all terrestrial vertebrate species globally [1]. The illegal wildlife trade (IWT) is one of the most lucrative illicit industries, with an estimated worth of $7–23 billion per year, and a growth of between 5–7% annually [2]. Both the probability of detection and the penalties, when detected, are generally low [3]. The deleterious consequences include extinction, biodiversity loss, the introduction of invasive species, and the emergence of novel zoonotic diseases [4]. The trade poses a threat to all 17 United Nations Sustainable Development Goals (SDGs), and most clearly in SDG 16 (Peace, Justice and Strong Institutions) and SDG 8 (Decent Work and Economic Growth) [5]. Despite this, eliminating the IWT has only recently been recognised as a political priority [6,7].

The detection and seizure of illegal wildlife in the supply chain, for instance at border posts, ports, and airports, represents an essential measure for the enforcement of trade regulations [8]. Animal carcasses can be recognised through visual inspection, and their identification may be aided by the use of visual guides (for example [9]). However, processed wildlife products lacking discernible morphological characteristics (e.g., meats, fish fillets, animal parts and derivatives within products such as Traditional East Asian Medicines [TEAMs] [10]) are frequently encountered. Processing is often carried out to conceal the species of origin and can make products visually indistinguishable from those obtained from farmed or legally harvested species [11]. For example, 59% of 194 items of meat seized at Brussels Zaventem airport were unidentifiable morphologically [12].

Species substitutions are of particular concern in fisheries. Each year, illegal, unreported, and unregulated (IUU) fishing is thought to cost about $36.4 billion in losses [13]. Widespread food fraud throughout the fisheries logistics network has seen high-value species adulterated or substituted with low-value species, and imported seafood labelled as domestic [14]. Equally, illegally harvested protected species may be sold as low-value species [15], as seen, for example, in the UK where protected shark species have been a component of traditional fish and chips [16]. Further food fraud and adulteration cases include the presence of pork in seafood products [17], or horsemeat in beef products [18], causing serious ethical or religious issues for consumers.

Species identification in processed samples depends on DNA-based analysis, which is normally achieved through DNA barcoding [19], the analysis of sequences that display low intraspecific but high interspecific variation. DNA barcodes used for animals lie in the mitochondrial genome (mtDNA): a segment of ~600 bp in the cytochrome c oxidase subunit I gene (COI), is the principal barcode [19,20], while other commonly used barcodes include sequences in cytochrome b (CYTB) [21–23], 16S [24,25] and 12S ribosomal RNA (rRNA) genes [26,27]. In the barcoding process, DNA is extracted from the sample and the chosen barcode is amplified by PCR using conserved primers. Typically, such primers are not applied across multiple taxonomic groups, but are targeted at a particular group, for instance shark-specific primers for use in pet foods [28], or primers designed for the identification of tortoises [29], or crocodiles [30]. Amplicons are usually then Sanger-sequenced and the resulting

sequences compared to a relevant database (e.g., the NCBI Nucleotide database, or the Barcode of Life Data System (BOLD) [31]) to identify the best-matching species.

This approach has a number of disadvantages. Time-consuming and costly research and development is needed every time a new taxon-specific species identification system is developed. Testing is done in a laboratory with expensive equipment, specialist staff, and associated costs, so that each test costs > $300 [32]. The infrastructure requirements mean that samples seized at customs posts or in the field must be transported to a laboratory. This complicates the evidential chain, incurs additional time and expense, introduces biosecurity risks, and can add the administrative burden of satisfying CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) restrictions if export and import across national boundaries are necessary. There are also potential issues with identification via the barcoding process itself. Sequence databases are not free of errors [33,34], and mislabelling of sequences or the erroneous database inclusion of sequences from nuclear mitochondrial segments (numts; [35,36]) can lead to inaccurate identifications [37]. This is more likely to occur when only a single barcode, e.g., COI, is used in a test [34]. Use of single barcodes also risks misidentification due to the co-amplification (or preferential amplification) of numts in samples undergoing testing.

A number of mitigations can be envisaged for these problems. 'Universal' PCR primers that amplify barcode sequences across a very wide taxonomic range could provide an agnostic assay allowing a single test to be applied to any suspicious item whilst also reducing validation, training and consumables costs. Simultaneous amplification and sequencing of multiple barcodes could mitigate against problems with databases and numt amplification. Finally, on-site analysis could produce results more quickly, reinforce the chain of custody, and facilitate the application of penalties on those illegally trading in wildlife products. The need for portable tools to carry out rapid on-site species identification has been emphasised by recent reports and surveys [38,39].

The portability, low start-up costs and ready availability of the Oxford Nanopore Technologies (ONT) MinION sequencing device make it a good candidate for in-field species identification via DNA barcoding. Continual improvements in chemistries have been closing the quality-score gap with other sequencing methods [40]. Indeed, it satisfies the key cost and accuracy requirements of a field-based system for wildlife forensic investigations as determined by surveying wildlife forensic and law enforcement professionals [38]. The MinION has an initial cost of $2999, weighs just ~100 g, and is plugged directly into a laptop USB port for operation and sequence interpretation. This device has already been used to identify species, from diverse vertebrates in Tanzania [41], to reptiles and amphibians in the Ecuadorian Chocó rainforest [42], a shark fin specimen from a market in India [43], as well as closely related plant species [44], and non-invasively collected wildlife samples [45]. Sequencing of a 421-bp CYTB segment in vertebrates has shown that ONT MinION data in conjunction with appropriate bioinformatic tools can generate sufficiently accurate consensus DNA sequences for forensic genetic species identification [46]. Protocols have been developed for species identification that, depending on sample type, can be completed in 10 h by a skilled operator in a laboratory [47].

In this study, we design, validate, and test for on-site application a system that addresses many of the problems associated with traditional DNA barcode-based approaches. Our system, VeRIF-ID (Vertebrate Rapid In-Field Identification via DNA), uses nanopore sequencing of four short DNA barcodes amplified with 'universal' primers that are diagnostic over a very broad taxonomic range. We simplify and optimise DNA extraction, PCR and library preparation such that the process from sampling to species identification takes less than 3 h for a single sample and 6 h for a batch of 24. In conjunction with a portable Bento Lab device, the method is suitable for on-site application, as demonstrated by bushmeat species identification within a major airport's customs zone.

Methods

Primer design

Whole mitogenome sequences for a phylogenetically diverse range of mammal (n = 113), fish (n = 59), bird (n = 35), reptile (n = 27), and amphibian (n = 15) species were downloaded from GenBank (S1 Table in S1 File). Up to 20 sequences per

species were included. The FeatureExtract 1.2 Server (https://services.healthtech.dtu.dk/service.php?FeatureExtract-1.2) was used to extract COI, CYTB, and 16S and 12S rRNA gene sequences. A multiple sequence alignment of the resulting 843 sequences was carried out using Clustal Omega [48] on the European Bioinformatics server. JalView [49] was used to inspect the resulting alignment and barcode primers were designed such that their 3´ ends were in conserved blocks bracketing variable regions. For each barcode, the variable region between primer sites was extracted and used as input with MEGA (v10.1.8) to build a neighbour-joining (NJ) Kimura two-parameter (K2P) tree with 100 bootstrap replicates [50], and displayed using the interactive Tree Of Life tool [51]. The tree was examined to determine the discrimination power of the primer pair under consideration. Primers favouring shorter amplicons were preferred given the known failure of longer barcodes in more degraded DNA samples [52,53]. Primers were designed to give amplicons in the length range 150–353 bp (S2 Table in S1 File).

Sample collection

This research was approved by the University of Leicester's Animal Welfare and Ethical Review Board (ref. AWERB/2021/159). All methods were performed in accordance with the relevant guidelines and regulations, and the study is reported in accordance with ARRIVE guidelines (https://arriveguidelines.org). Consent to participate is not relevant. No experiments were carried out on live animals for this study, and no animals were sacrificed to provide samples. Samples were either from non-invasive sources, from sources intended for human consumption, or provided to us following sampling by qualified veterinarians during necessary procedures (S3 Table in S1 File).

Single-source samples encompassed a wide range of species and sample types, and are listed in S3 Table in S1 File. A further set of samples of muscle tissue from seven species (cattle, sheep, pig, chicken, turkey, salmon, and trout) were used to create lab-prepared mixtures. Finally, Traditional East Asian Medicine (TEAM) samples confiscated by the Metropolitan Police Service (London, UK) were analysed. Tissue, blood, and DNA extracts were stored at −20°C with all other sample types kept at room temperature prior to DNA extraction.

DNA extraction

Hair, shed reptile skin, hedgehog quill, frog spawn, claw, eggshell, and feather samples were washed twice in 500 μL 70% (v/v) ethanol and twice in 1 mL Milli-Q water. DNA was extracted from 1–2 mm³ of claw, tissue, shed skin, feather calamus, spawn, quill, 3 μL of blood, or 5–10 hairs. A previously developed alkaline extraction method [54] was used for DNA extractions. Samples were heated in 100 μL of lysis solution (200 mM KOH, 2 mM EDTA, 0.2% (v/v) Triton X-100) at 98°C for 10 mins and then neutralised with 300 μL of 100 mM Tris-HCL. This extraction method was used with 0.01 g of each of the seven tissue samples used to create 19 lab-prepared mixtures (S4 Table in S1 File). Bone and scute samples were wiped down successively in bleach, water (×2), and ethanol. After drying for 5–10 min, a hand-held drill (Ozito Cordless drill driver) with a 6-mm wood drill-bit was used to obtain ~200 mg of bone powder. About 200 mg of pill, paste, or oil were obtained from TEAM samples (S5 Table in S1 File) for DNA extraction. Pills were powdered using a mortar and pestle. Samples were incubated at 56°C for 2 h with 950 μL 0.5M EDTA, 50 μl 1% (w/v) n-lauroylsarcosine sodium salt, 14 μL of 1M DL-DTT, and 32 μL proteinase K (20 mg/ml) [55]. After centrifugation at 11,300 g for 3 min, three 125 μL fractions of the supernatant were concentrated using the MinElute PCR Purification kit (QIAGEN) with the DNA eluted in 30–60 μL of buffer.

The QIAamp DNA Investigator kit was used for DNA extractions from ~0.5 cm² of each TEAM plaster (S5 Table in S1 File) following the manufacturer's protocol for the isolation of total DNA from paper and similar materials. DNA was eluted in 30–60 μL of water.

PCR/Tetraplex amplification

HPLC-purified primers with an additional tandemly repeated 13-bp sequence at the 5´ end of each of the four barcode primer pairs (CYTB: [AGTGTCCTGCTAG]$_2$; COI: [CTGAGGTGATCAG]$_2$; 16S:[AACATGTGGTAAG]$_2$; 12S:

[TTATTCGCACACT]$_2$ [56]) were used during amplification (S2 Table in S1 File). This additional sequence was included to enable the addition of transposome introduced indices at the 5' end of primer sequences. Each reaction was performed in a 25-µL volume containing 12.5 µL of Type-it Master Mix (QIAGEN), 0.76 µL of forward and reverse CYTB primers (0.3 µM final concentration), 0.62 µL of forward and reverse COI and 16S primers (0.25 µM final concentration), and 0.2 µL of forward and reverse 12S primers (0.08 µM final concentration), plus 2–3 µL DNA extract. Cycling conditions consisted of an initial denaturation at 95°C for 5 min, then 35 cycles at 95°C for 30 s, 45°C for 90 s, 72°C for 30 s, and a final extension at 68°C for 10 min.

Nanopore flowcell runs

A total of 96 single-source samples were prepared for sequencing on a MinION R9.4.1 flow cell (FLO-MIN106D, ONT). Mixtures and a no-template control were sequenced on a second flow cell. A 14-µL volume of PCR products was purified using 1.8 × AppMag PCR Clean Up Beads (Appleton Woods Ltd), and the DNA eluted in 14 µL of water. The Rapid Barcoding kit 96 (SQK-RBK110.96, ONT) was used to prepare samples for sequencing where the manufacturer's protocol was followed with some exceptions. The name of this kit uses the term 'barcoding' to describe the inclusion of short sequences (known as barcodes, or indices) in library preparation that allow data from multiple samples to be deconvoluted into individual sequence datasets. We refer to this process hereafter as 'indexing' to avoid confusion with barcoding in the sense of species identification. Sample quantification and normalisation steps were omitted. Instead, a 9-µL volume of each sample was used during the fragmentation and nanopore index addition step ('tagmentation'), and a 9-µL volume of each tagmented sample was pooled together. Half of this volume was used during sequencing of the single-source samples, whereas the full volume was used for sequencing of the pooled mixed samples. Sequencing runs were carried out on a MinION sequencer and the software program MinKNOW (v21.06.0 for the single-source samples and v21.11.7 for the mixtures).

Data analysis

Guppy (v5.0.16) (https://nanoporetech.com/software/other/guppy/history?version=5-0-16) was used for basecalling with the high accuracy model, and demultiplexing with the trimming index option enabled. The resulting sequences were filtered using NanoFilt (v2.8.0) [57] to select for sequences with a quality score of ≥11 and length between 100 and 500 bp. Reads assigned to a given nanopore index were processed using Porechop (v0.2.4) [58] to search for each of the forward and reverse barcode primer sequences and assign reads to their respective species ID barcode. The NGSpeciesID pipeline (v0.1.1.1) [59] was used to generate consensus sequences from the resulting groups of reads. A cluster abundance ratio of 0.1% was applied for the mixed flow cell run. Consensus sequences were used as a BLAST query against a local NCBI Nucleotide database (constructed 02/02/2022), with the output file format specified via the option -outfmt "7 qseqid qlen evalue qcovs pident staxids stitle scomnames". The BLAST output file, sorted by E-value, was then sorted by query sequence ID and percent identity before filtering to remove duplicate lines if the same taxonomic identifier was present more than once for any given sequence, and then further filtered to retain only the top hit(s).

In the single-source samples sequencing run, consensus sequences were discarded if they had < 100 supporting reads, if the similarity for a given top match was < 90%, or the BLAST hits were not labelled to the species level, or were of non-coding RNA sequences. In the mixed samples sequencing run, consensus sequences were retained if they had > 20 supporting reads. This cut-off, which is about three times the highest number of supporting reads seen in the negative control, was selected to overcome any instances of contamination. A species was only considered present if forward and reverse consensus sequences were generated for at least two species ID barcodes in the mixtures run. In the single-source sequencing run, results for a species ID barcode were considered only when consensus sequences were generated in both the forward and reverse direction (though see S1 Text).

In selected cases, all reads assigned to a nanopore index were aligned to a mitochondrial reference sequence using minimap2 (v2.24) [60]. The SAM alignment file was transformed using SAMtools (v1.13) [61] and visualised using IGV (v2.3.82) [62].

On-site species identification

The methods used for on-site species identification at Twycross Zoo and Brussels Zaventem Airport varied depending on the sample source and site. Details are provided in S6 Table in S1 File.

Results

We set out to design a rapid, robust and cost-effective system for vertebrate species identification that could be applied on-site with its only external requirement being a power supply. Below we describe the development and validation of this system, VeRIF-ID (Vertebrate Rapid In-Field Identification via DNA), and also give proof-of-principle examples of applications on-site, including deployment in a customs reception site within a major European airport.

Primer design and in silico assessment of barcode performance

We first designed a new set of 'universal' primers for the amplification of four commonly used barcodes (COI, CYTB, 12S and 16S rRNA) for species identification in vertebrates. Design of previously published systems was based on a limited number of sequences and species, and given the growing availability of sequence data we began afresh. We used 843 sequences from a phylogenetically diverse range of 249 mammal, bird, fish, reptile, and amphibian species for primer design, aiming for short amplicons to maximise success in analysing degraded DNA. This resulted in novel 16S rRNA and COI primers that are quite distinct from other published primers; however, for CYTB and 12S rRNA, similarities to existing primers are notable (S1 Fig), indicating that the inclusion of additional sequences made little difference here. The discrimination power of the four new primer pairs, which generate amplicon lengths between 150 and 353 bp, was assessed by building NJ K2P trees and seeking instances where a species-level identification was not possible (S2–S5 Figs). Across all barcodes, an incorrect species-level identification was seen in only two cases; for a tolai hare (*Lepus tolai*) and a saltwater crocodile (*Crocodylus porosus*). These results may be explained by introgression previously noted in hares [63] or by sequence mislabelling in GenBank [64] and/or hybridisation in the crocodile. Species-level identification was achieved across the remaining 99.2% (247/249) species by examining all four barcodes.

Development of a simple and rapid workflow for VeRIF-ID

The adoption of portable on-site species identification is dependent on the development of a fast, affordable, accurate and user-friendly system [38]. We examined each stage of the DNA barcoding process with these considerations in mind. Although commercial DNA extraction kits have been favoured in MinION-based barcoding studies [42,65–67] because they yield extracts of high quality and concentration, these methods require several items of equipment, involve multiple manipulation steps, and have a high cost per sample (~$4). Instead, we assessed previously developed DNA extraction methods that use commonly available laboratory chemicals [54,68]. We settled on a two-step method [54] based on an alkaline (KOH) lysis buffer. Incubating samples for 10 min at 98°C in this buffer, followed by neutralisation, produced DNA extracts of suitable quality for downstream barcoding steps without tube-to-tube transfers. This ~10-min extraction method worked well across a wide range of sample types (blood, muscle tissue, skin, feathers, hair, and claws [69]) and reduced the extraction time from ~30 mins for blood, ~1–3 h for tissue and hair (using commercial kit-based extractions) or 5–12 h for feathers [70], and over 24 h for claws [71]. A second extraction method was necessary for bone and Traditional East Asian Medicines (TEAM) samples, and a ~2.5-h protocol was chosen based on a previously developed method [55] versus >12 h using a DNA extraction kit. These simpler extraction methods considerably reduced the cost of extraction per

sample from ~$7 using a commercial kit to <$0.10 using a KOH-based DNA extraction method, or ~$5 using the EDTA-rich DNA extraction method for bone and TEAM samples.

To maximise efficiency, we set up a PCR tetraplex to co-amplify the four barcodes. The redundancy afforded by multiple barcodes favoured this approach despite the expected potential loss of barcode sequence data due to competition between targets. Heating steps accounted for ~45–47% (~47/100 or 54.5/120 min) of the time taken for PCR during on-site tests.

Oxford Nanopore Technologies' rapid indexing kit (see Methods for an explanation for the name we use here) was selected to prepare samples for sequencing. The rapid kit was chosen despite manufacturer recommendations to use it with longer (500 bp – 5 kb) fragments of DNA (https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-sequencing-v14-amplicon-sequencing-sqk-rbk114-24-or-sqk/v/raa_9198_v114_revh_29nov2023), and its relatively low throughput compared to the ligation kit. The choice is for speed and simplicity: the method comprises just two main steps, takes only 10 minutes, and requires no third-party reagents. The short turnaround time is largely due to the use of a transposase complex that fragments DNA and simultaneously attaches nanopore indices to both newly-created 5′ strand termini (https://store.nanoporetech.com/rapid-barcoding-sequencing-kit-24-v14.html) during the first main step ('tagmentation'). This simplifies preparation for sequencing by significantly reducing the number of pipetting and tube transfer steps (by 82% and 72% respectively compared to the native indexing kit (SQK-NBD114.24)) without sacrificing sequencing accuracy. The use of the rapid indexing kit does introduce some complexities in bioinformatic processing of generated sequence data which are discussed in S1 Text (see also S6 Fig).

Samples were sequenced on a flow cell connected to a laptop with an offline version of the operating software Min-KNOW and a local copy of the NCBI Nucleotide database, eliminating the need for an internet connection. The workflow summarising these steps as well as the equipment and reagents required, which together fit into a standard luggage item, is shown in Fig 1.

Testing VeRIF-ID on single-source samples

We first tested the developed method on a nanopore flow cell to identify the species of origin for 96 single-source samples, originating from 45 mammals, 19 birds, 12 fish, six reptiles, one amphibian, and two invertebrates (S3 Table in S1 File). A total of 12.1 million reads were obtained (for details of Nanopore run statistics see S7 Table in S1 File), of which 88% were assigned to a nanopore index. An average of 58K reads per sample (range, 10K - 164K) were taken forward. Because of the mode of action of the rapid indexing kit used for library preparation, it was not possible to obtain a single consensus sequence per species ID barcode. Instead, forward and reverse consensus sequences were obtained for each species ID barcode (see S1 Text). Despite the transposase-introduced sequence resulting in longer than expected consensus sequences, and the use of a fragmentation-based kit, the alignment lengths between the query barcode consensus sequence and top database sequence match were generally in agreement with expected barcode insert sizes (S1 Text).

Of the 83 unique vertebrate species (S3 Table in S1 File), 75 could be assigned to species level, seven to genus, and one to tribe level (Fig 2). The eight less precisely assigned species include recently diverged domesticates, and are detailed in S8 Table in S1 File. For a further seven species (African pygmy hedgehog [*Atelerix albiventris*], Argentine anchovy [*Engraulis anchoita*], Malagasy giant rat [*Hypogeomys antimena*], great grey owl [*Strix nebulosa*], red crested turaco [*Tauraco erythrolophus*], tope shark [*Galeorhinus galeus*], and pink-backed pelican [*Pelecanus rufescens*]) there were missing or partial reference data for some barcode sequences (S8 Table in S1 File). This meant that alternative candidate species appeared as the top match(es) in the absence of data from the true species. Unexpected results were obtained in a further three instances, due to a range of issues (S8 Table in S1 File). Distinguishing nyala [*Tragelaphus angasii*] from bushbuck [*T. scriptus*] is affected by NCBI reference sequence mislabelling [72]. Unexpected species assignment was also seen for the herring (*Clupea harengus*) and Arctic cod (*Boreogadus saida*) samples, which were
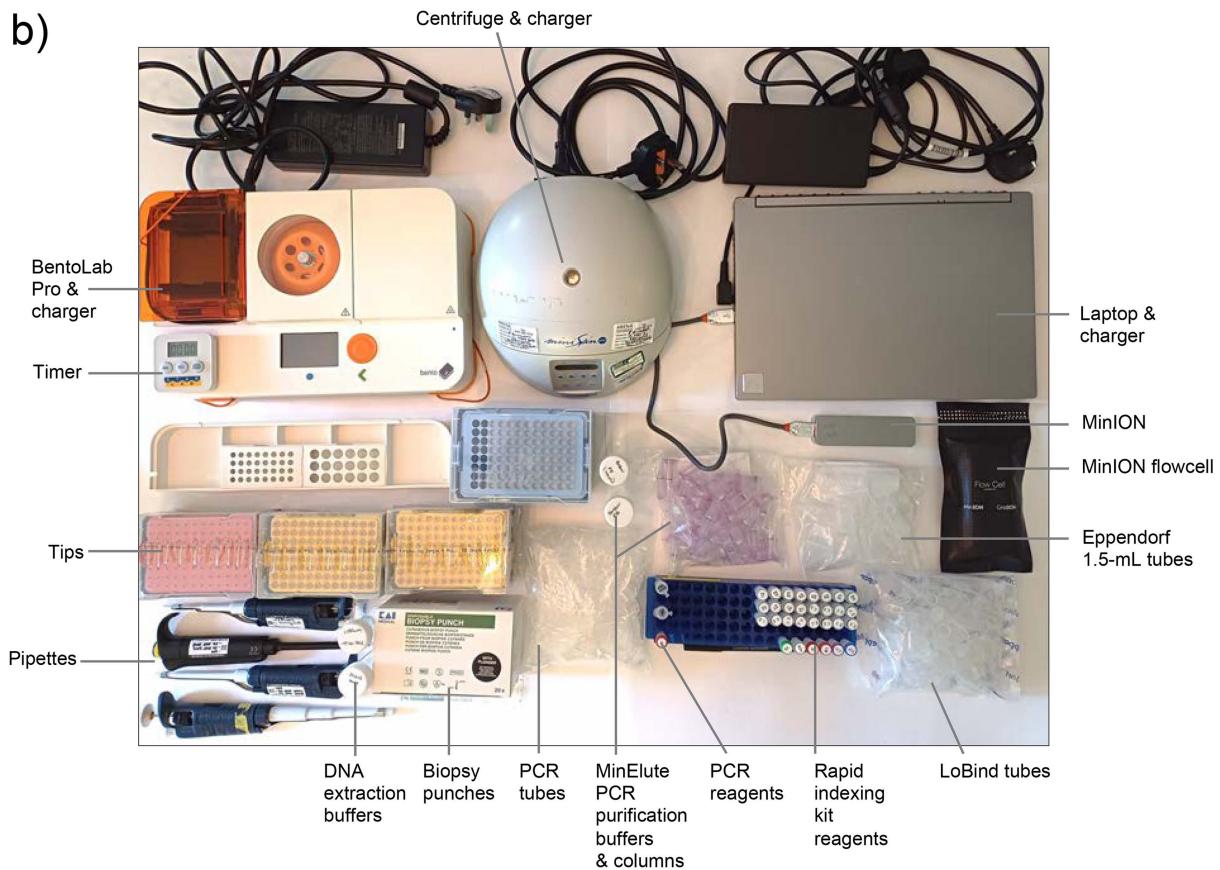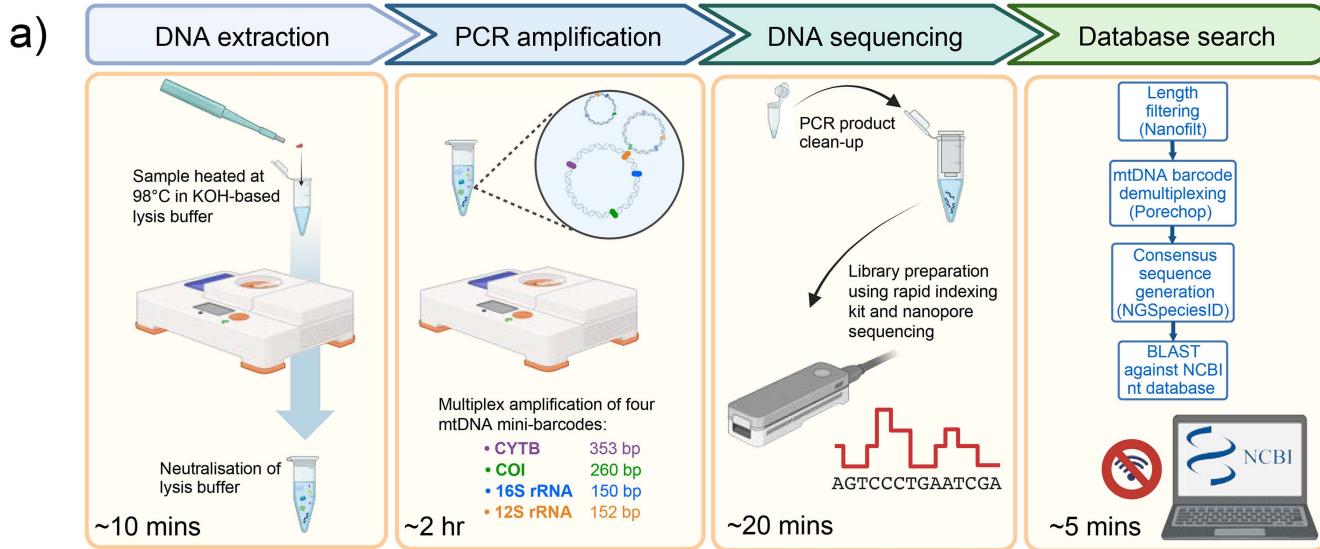
Fig 1. VeRIF-ID workflow and equipment. a) Schematic workflow for species identification. b) Photograph showing the necessary equipment and reagents, as deployed in Brussels Zaventem airport. See also S1 Text.

https://doi.org/10.1371/journal.pone.0336383.g001
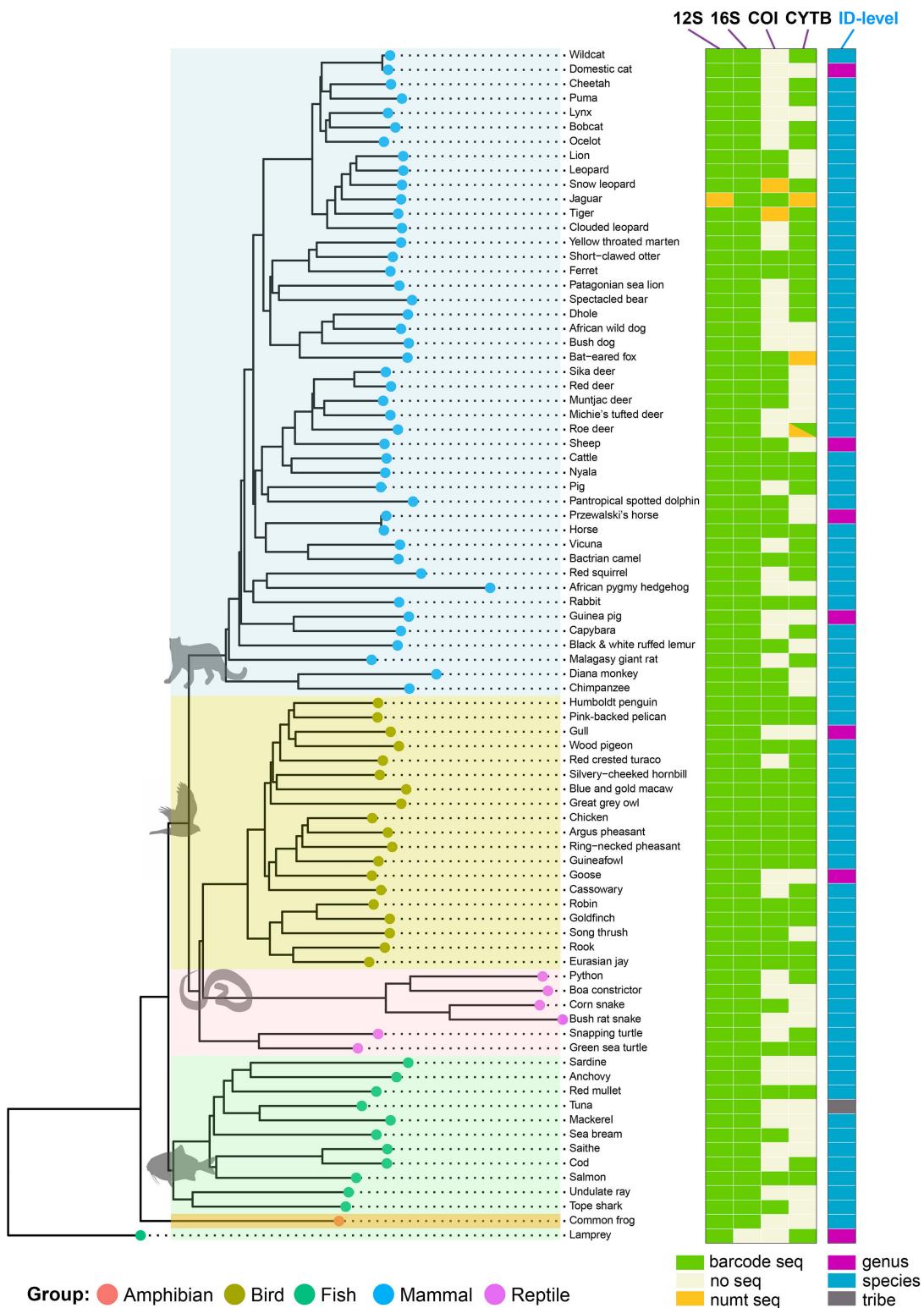
Fig 2. Barcode recovery across 83 vertebrate species following a nanopore flow cell run. To the left is a NJ K2P tree, created using publicly available cytochrome b sequences from GenBank for the species (or closely related species) displayed. Recovery of barcode and numt sequences for each of the four tested barcodes is shown schematically to the right, as well as the taxonomic level of identification achieved. Sequences were labelled

as numts if their top match was to nuclear DNA and/or the top mitochondrial match was in disagreement with the other barcode sequence matches. Two additional invertebrate species were also analysed with some success [69] but are not shown here.

identified as saithe (*Pollachius virens*) and Atlantic cod (*Gadus morhua*) respectively. These results appeared genuine based on a reference-based approach where reads were aligned to the expected and assigned species. This surprising outcome is consistent with the finding that 55% of shop-bought fish samples tested in the UK were mislabelled [73].

Species identification in mixtures

Because tested samples often contain mixtures of material from more than one species, we carried out a qualitative assessment of species identification in known mixtures of vertebrate muscle tissues, as proof of principle. Nineteen lab-prepared mixed samples (S4 Table in S1 File) and one negative amplification control were sequenced on a flow cell. A total of 2.5 million reads were obtained, and of these, 1.4 million remained after filtering, and 988K (70%) were assigned to their respective species ID barcode. It was possible to successfully identify all expected species in mixtures with up to five species of mammals and birds. It is important to note that here we consider only the presence or absence of a species, rather than a quantitative assessment. As shown schematically in S7 Fig, fewer reads than expected are observed for some species in particular mixtures, despite the addition of equal mass of tissue from each species to a mixture. This is possibly due to species differences in mtDNA copy number per unit mass, and to primer bias, and is particularly marked for fish species (S7 Fig).

To assess the performance of VeRIF-ID in mixtures of interest to law enforcement, we analysed a range of seven Traditional East Asian Medicine (TEAM) products (S5 Table in S1 File), from which 203K reads were obtained. About half (100K) remained after filtering, and an average of 9.4K reads per sample remained after barcode demultiplexing. For two samples, no results were obtained, but the species identified within the remaining samples rarely matched those listed on the packaging (when available [S5 Table in S1 File]). Two species of particular interest, the black rhinoceros (*Diceros bicornis*) and saiga antelope (*Saiga tatarica*), were identified within a wax-covered ball intended to tranquillise the nervous system and treat a range of ailments including fever, annoyance, panic, and facial nerve disorder. At the time of manufacture of the TEAM product, both species were listed as critically endangered (www.iucnredlist.org; the saiga antelope's status was revised in 2023 to near threatened). Packaging in this instance suggested the presence of saiga antelope (as "Cornu Antelopis"), but not black rhinoceros.

Use of VeRIF-ID for on-site testing

The VeRIF-ID system has been designed to be used on site. We first tested its suitability at Twycross Zoo, UK. We were provided with three samples (a hair, feather, and blood sample) from three known species during a 'blind' test. The species of origin was successfully determined in all cases (respectively, Coppery titi, *Callicebus cupreus*; Demoiselle crane, *Grus virgo*; Francois's langur, *Trachypithecus francoisi*) in just over 5 h (S9 Table in S1 File).

We next deployed the VeRIF-ID system within a customs zone at Brussels Zaventem Airport, an important gateway for unregulated import of wildlife derivatives, through which around 3.8 tonnes of bushmeat are estimated to enter Europe each month [12]. Bushmeat samples, and fish contaminated with insects (therefore disallowed for import), were seized by customs officials during a passenger baggage compliance action on 24/01/2024. Without disturbance to normal customs activities, the species of origin of these 15 samples was determined in approximately 6 h (S10 and S11 Tables in S1 File). Samples seized included common livestock (pig [*Sus scrofa*] and cattle [*Bos taurus*]), but also two species of duiker, a small to medium sized antelope (Fig 3). Maxwell's duiker (*Philantomba maxwellii*) is considered 'of least concern' in the International Union for Conservation of Nature (IUCN) classification (www.iucnredlist.org), while bay duiker (*Cephalophus dorsalis*) is listed as 'near threatened'.
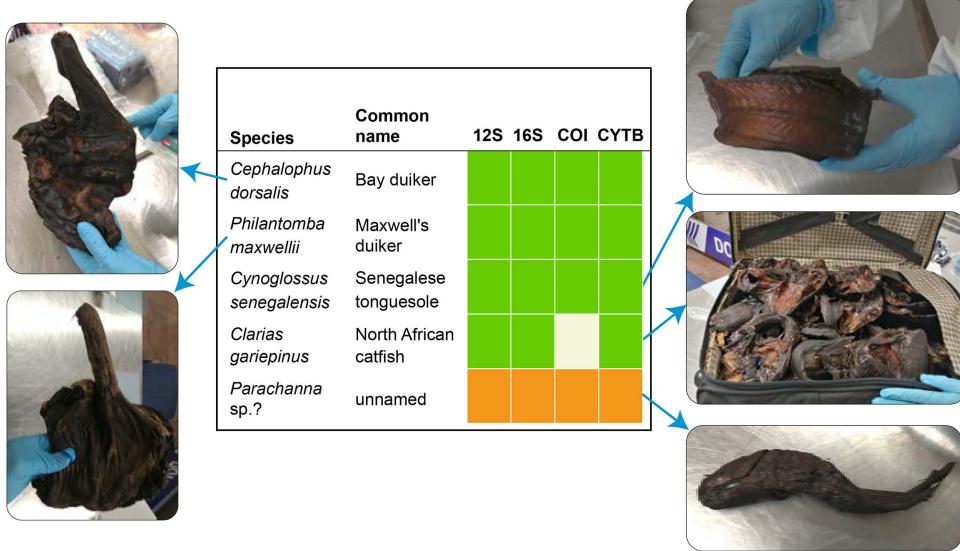
Fig 3. Non-livestock bushmeat and fish species analysed on-site at Brussels Zaventem Airport. Photographs show that bushmeat and fish are dried, and/or smoked. The table gives species and common names (where known) for the five non-livestock samples. Results for the four barcodes are shown to the right (green: amplified and identifying; orange: amplified but no hit in NCBI Nucleotide database (possible undescribed species); white: no amplification). See S10 Table in S1 File for further information.

https://doi.org/10.1371/journal.pone.0336383.g003

Discussion

We have described the development of VeRIF-ID (Vertebrate Rapid In-Field Identification via DNA). Our aims were to maximise taxonomic range and the robustness of species identification [74], to reduce time, cost and complexity, and to ensure portability and on-site capability. We chose to use short barcode amplicons to maximise the recovery of data from challenging samples, but clearly there is a balance to be struck between this and the greater certainty of identification offered by longer recovered sequences.

In our test of single-source samples across a wide range of vertebrate classes from lamprey to chimpanzee (S3 Table in S1 File), 91% (75/83) could be assigned to species level, 8% (7/83) to genus, and 1% to tribe level (Fig 2; S8 Table in S1 File). This breadth makes VeRIF-ID a quasi-universal test – a valuable property because it saves the time and expense of specific test development. It is also worth noting that it enables an immediate response, as changes in regulation and market trends alter the species targeted by the IWT [75]. Despite our vertebrate-focused design, we also generated taxonomically useful information for some invertebrate species (S5 and S10 Tables in S1 File; [69]). As a considerable proportion of illegally traded species consists of invertebrates and plants [76], it would be valuable to multiplex 'universal' invertebrate and plant barcode primers in a similar manner, and this should be feasible.

The use of multiple barcodes increases the robustness of identification with respect to barcode drop-out, incomplete or erroneous reference sequence data [33,34], and the co-amplification or preferential amplification of numts (Fig 2). The 12S and 16S rRNA barcodes were successfully amplified across 100% (83/83) and 99% (82/83) of the species tested respectively, with the 16S sequence missing from only the lamprey, a basal jawless vertebrate. The inclusion of more discriminating COI and CYTB barcodes allowed us to achieve a species-level identification in 20% (15/75) of species where this would not have been possible with the rRNA barcodes alone. In part, this was due to the lack of reference 12S and 16S rRNA sequence data in some species.

Any obtained barcode sequence must be searched against a relevant database of known sequences to effect an identification. Generally, the Barcode of Life Database (BOLD) is preferred since uploaded information must be of a certain

quality, and derived from a specimen for which taxonomic assignment can be reviewed [31]. However, BOLD is currently centred around the barcode COI, so not well suited to multi-barcode query. Here we used NCBI's Nucleotide database, which incorporates GenBank. The latter is a public DNA sequence repository containing non-quality-controlled, unvalidated data in which mistakes are not uncommon [77]. Notably, it has been reported that much of the BOLD data is unauthenticated and itself transferred from GenBank [78]. More generally, any database is susceptible to the misidentification of specimens, a problem for which there is no easy solution [79]. Identification errors are more likely when available data are confined to a single barcode sequence from a single source. By querying multiple different barcode sequences, often derived from different laboratories, the effect of database errors is reduced.

Amplifying numts in addition to, or instead of, target mitogenome sequences can cause problems in species identification. Examples of the mitigation provided by multiple barcodes against numt amplification can be seen in Fig 2. Where numts are amplified and sequenced, they comprise at most two of the four barcodes (prominently in the Felids [35]), while the remaining barcode sequences reflect the endogenous mitogenome.

The multi-species mixture testing in our study was carried out as a proof-of-concept to show that the method could identify individual species within such mixtures. We do not envisage VeRIF-ID being used for quantitative analysis in practice, but our analysis of lab-prepared mixtures of muscle tissues from up to five different vertebrates shows that the method can generally detect the presence or absence of species reliably. Primer bias leads to dropout of certain barcodes for some species, particularly for fish in the presence of non-fish species. The method is thus susceptible to false negatives and has limited sensitivity, but nonetheless could still detect two critically endangered species in a TEAM sample (S5 Table in S1 File). Such identifications could be followed up with species- or genus-specific primers for confirmation and quantification.

Minimising the time required for species identification is of great practical importance. The first requirement is to ensure that the method can be deployed close to a site of sample seizure. The equipment and reagents required for VeRIF-ID are easily portable by a single operator, and the onboard database obviates the need for an internet connection. A power supply is necessary, and a refrigerator is desirable to prolong the life of flow-cells and reagents. Once analysis is commenced, a rapid testing time is important. While we have reduced this considerably, in practice a duration of an hour or less would make testing while a suspect was detained, or a shipment in transit was held, more useful. For VeRIF-ID, the longest step in the process is PCR. Further optimisation of denaturation, annealing and extension time could be explored, and overall time could also be shortened through the use of more efficient thermocyclers than the Bento Lab Pro, employed here. Such devices include the NextGenPCR™ thermocycler, which has been reported to reduce the time for PCR to around 25 min [80]. However, a compromise is required between speed, expense, and portability – the thermocycler costs about $22K, and weighs 17 kg.

To demonstrate proof of principle, we trialled VeRIF-ID in the customs zone at Brussels Zaventem airport, successfully analysing a variety of seized fish and bushmeat samples from passengers on flights from Addis Ababa, Kigali via Entebbe, and Luanda via Kinshasa. The method allowed 15 samples to be sampled and analysed in around 6 hours, without disruption to normal airport activities (Fig 3 and S9 and S10 Tables in S1 File). Given the very high traffic of bushmeat via European airports [12,81], longer term deployment of equipment combined with an efficient workflow is needed to allow higher numbers of samples to be processed during a day. The species identified in Brussels included the bay duiker, which is a reservoir for Ebola virus [82] and has been linked to Ebola spillovers into human populations in West and Central Africa [83]. On-site species identification, potentially followed by pathogen testing in known reservoir species, could provide valuable data for the tracing of pathogens along supply chains and contribute to pandemic prevention preparedness.

Democratising species identification testing demands cheap methods. The total cost of all the required equipment for VeRIF-ID (a MinION sequencer, laptop, Bento Lab, centrifuge and a set of pipettes), is relatively low, at approximately $8K. Considering the test itself, while high-throughput laboratory-based DNA barcoding costs as low as $0.10 have been

reported [84], in practice forensic testing costs tend to be much higher. Excluding consumables transportation charges, the overall per sample cost of VeRIF-ID is ~$43–49 if using a flow cell once (before goods/sales tax, and assuming analysis of 24 samples using a flowcell). This is about 160% cheaper than prices typically charged for species identifications [32]; the on-site nature of the testing also eliminates the need for courier charges for sample shipping. Most of our method's cost is in the sequencing step, which may be reduced by halving reagent volumes during the library preparation step, and either reusing flow cells or through the use of cheaper, single-use Flongle flow cells; see S1 Text). Depending on the chosen set-up, per-sample costs can be reduced, for instance if a flow cell is re-used once, the cost per sample decreases to ~$27–33. Synthesising a set of primer sequences with the indices incorporated within their 5′ tails would allow samples to be pooled immediately after PCR. Only one PCR clean-up would be required, reducing the consumables required, the complexity of the process, and the overall time taken. The slightly cheaper rapid sequencing kit (SQK-RAD114) could then be used.

The simpler a DNA barcoding technique is, the more widely it can be deployed, since relatively little operator expertise is required, and training is simplified. In terms of hands-on abilities, VeRIF-ID is simple, but demands basic lab skills of pipetting; using lyophilised PCR mixes could simplify this. However, flow cell loading is the most crucial and demanding practical step. Challenges in flow cell loading have been recognised by ONT, which aims to revise the process with a simpler, more user-tolerant design. Overall, the most challenging aspect of the method we present here is the need for some familiarity with bioinformatics; developing a simple user interface is therefore an obvious future goal. The development of a user-friendly software with a graphical user interface for use across all commonly used operating systems would greatly increase the accessibility of the workflow. Such software has previously been developed to generate barcodes in real-time [85].

Supporting information

S1 Fig. Barcode primer positions in mtDNA, and relationships to previously designed primers. a) Schematic diagram of typical vertebrate mtDNA, with approximate positions of the four barcode sequences. Primer sequences are given in S2 Table in S1 File. b) Schematic representation of the relationship between the primers in this study and previously published primers. Numbers indicate coordinates, based on the human mtDNA reference sequence (NC_012920). c) Detailed comparison of primer sequences for barcodes COI, CYTB, and 12S rRNA. 16S rRNA primers are not shown since they have no overlaps with previously described primers.
(PDF)

S2 Fig. Phylogenetic tree based on cytochrome b. a) A NJ-K2P tree with 100 bootstrap replicates was calculated using MEGA (Hall, 2013) and displayed using the interactive Tree Of Life (Letunic and Bork, 2021) for the chosen 307-bp segment of cytochrome b. Clades have been collapsed when multiple sequences of the same species (or closely related species) were present. Instances where species differentiation was not possible are displayed in panels b) and c). Inconsistent species identifications are highlighted by red underlining.
(PDF)

S3 Fig. Phylogenetic tree based on cytochrome c oxidase subunit I. a) A NJ-K2P tree with 100 bootstrap replicates was calculated using MEGA (Hall, 2013) and displayed using the interactive Tree Of Life (Letunic and Bork, 2021) for the chosen 214-bp segment of cytochrome c oxidase subunit I. Clades have been collapsed when multiple sequences of the same species (or closely related species) were present. Instances where species differentiation was not possible are displayed in panels b) to f). Inconsistent species identifications are highlighted by red underlining.
(PDF)

S4 Fig. Phylogenetic tree based on 12S ribosomal RNA. a) A NJ-K2P tree with 100 bootstrap replicates was calculated using MEGA (Hall, 2013) and displayed using the interactive Tree Of Life (Letunic and Bork, 2021) for the chosen 111-bp segment of 12S rRNA. Clades have been collapsed when multiple sequences of the same species (or closely related species) were present. Instances where species differentiation was not possible are displayed in panels b) to h). Inconsistent species identifications are highlighted by red underlining.
(PDF)

S5 Fig. Phylogenetic tree based on 16S ribosomal RNA. a) A NJ-K2P tree with 100 bootstrap replicates was calculated using MEGA (Hall, 2013) and displayed using the interactive Tree Of Life (Letunic and Bork, 2021) for the chosen 103-bp segment of 16S rRNA. Clades have been collapsed when multiple sequences of the same species (or closely related species) were present. Instances where species differentiation was not possible are displayed in panels b) to h). Inconsistent species identifications are highlighted by red underlining.
(PDF)

S6 Fig. Action of the rapid indexing kit and generation of sequences for database query. a) PCR products are generated by primers with 26-nt 5′ extensions to increase the target size for transposome complex binding. b) Transposome complex introduces a double-strand break at a random point within each duplex and adds a sample-identifying index to each of the two internal free 5′ ends. c) Following sequencing adapter addition, sequencing proceeds from each adapter through the nanopore in the 5′-3′ direction. d) In bioinformatic processing, a read is only retained if it contains a primer sequence at the distal end (primer sequences are then trimmed – not shown), and reads are used as input with NGSpeciesID for consensus sequence generation. This leads to reduced confidence at the proximal end of each strand consensus. Top- and bottom-strand consensuses are used separately for database query and only results with consensus sequences based on at least 100 independent reads and generated in both directions are considered (though see main text).
(PDF)

S7 Fig. Heatmap displaying the relative percentage of reads assigned to a given species within 19 lab-prepared mixtures. For each mixture, the first column indicates the relative proportion of reads expected for each species present in the mixture for that indicated barcode in the minibarcode tetraplex. The second column for each mixture indicates the relative proportion of reads assigned to the barcode of each of the detected species based on the highest number of supporting reads of that species' barcode consensus sequence. In sequence processing consensus sequences with >20 supporting reads were retained. A species was only considered present if forward and reverse consensus sequences were generated for at least two species ID barcodes.
(PDF)

S1 File. S1–S11 Tables. Supplementary tables with additional information on methods, samples analysed, and data.
(XLSX)

S1 Text. Supplementary text: Effect of the rapid-indexing kit on sequence data, VeRIF-ID costs and protocols.
(DOCX)

Acknowledgments

Author contributions

Conceptualization: Mark A. Jobling, Jon H. Wetton.

Formal analysis: Emily C. Patterson, Jon H. Wetton.

Funding acquisition: Emily C. Patterson.

Investigation: Emily C. Patterson, Sandrella Morrison-Lanjouw.

Supervision: Mark A. Jobling, Jon H. Wetton.

Visualization: Mark A. Jobling, Emily C. Patterson, Jon H. Wetton.

Writing – original draft: Mark A. Jobling, Emily C. Patterson, Jon H. Wetton.

Writing – review & editing: Mark A. Jobling, Emily C. Patterson, Jon H. Wetton, Sandrella Morrison-Lanjouw.

References

1. Scheffers BR, Oliveira BF, Lamb I, Edwards DP. Global wildlife trade across the tree of life. Science. 2019;366(6461):71–6. https://doi.org/10.1126/science.aav5327 PMID: 31604304

2. Smart U, Cihlar JC, Budowle B. International Wildlife Trafficking: a perspective on the challenges and potential forensic genetics solutions. Forensic Sci Int Genet. 2021;54:102551. https://doi.org/10.1016/j.fsigen.2021.102551 PMID: 34134047

3. Chaber A, Allebone-Webb S, Lignereux Y, Cunningham AA, Marcus Rowcliffe J. The scale of illegal meat importation from Africa to Europe via Paris. Conserv Lett. 2010;3(5):317–21. https://doi.org/10.1111/j.1755-263x.2010.00121.x

4. Cardoso P, Amponsah-Mensah K, Barreiros JP, Bouhuys J, Cheung H, Davies A, et al. Scientists' warning to humanity on illegal or unsustainable wildlife trade. Biol Conserv. 2021;263:109341. https://doi.org/10.1016/j.biocon.2021.109341

5. Transnational Alliance to Combat Illicit Trade. Mapping the impact of illicit trade on the Sustainable Development Goals; 2019. Available from: http://www.tracit.org/uploads/1/0/2/2/102238034/tracit_sdg_july2019_highres.pdf

6. UNODC. World Wildlife Crime Report - Trafficking in protected Species. United Nations Office on Drugs and Crime; 2020.

7. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Revision of the EU action plan against wildlife trafficking; 2022 [cited 2024 Mar]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A581%3AFIN&qid=1667989438184

8. 't Sas-Rolfes M, Challender DWS, Hinsley A, Veríssimo D, Milner-Gulland EJ. Illegal wildlife trade: scale, processes, and governance. Annu Rev Environ Resour. 2019;44(1):201–28. https://doi.org/10.1146/annurev-environ-101718-033253

9. UNEP-WCMC. Checklist of CITES species - CITES Identification Manual; 2020 [cited 2024 Mar]. Available from: https://checklist.cites.org/#/en

10. Iyengar A. Forensic DNA analysis for animal protection and biodiversity conservation: a review. J Nat Conserv. 2014;22(3):195–205. https://doi.org/10.1016/j.jnc.2013.12.001

11. ICCWC. Wildlife and Forest Crime. Analytic toolkit. 2nd ed. International Consortium on Combating Wildlife Crime; 2022 [cited 2024 Mar]. Available from: https://www.unodc.org/documents/Wildlife/Toolkit_e.pdf

12. Chaber A-L, Moloney GK, Renault V, Morrison-Lanjouw S, Garigliany M, Flandroy L, et al. Examining the international bushmeat traffic in Belgium: a threat to conservation and public health. One Health. 2023;17:100605. https://doi.org/10.1016/j.onehlt.2023.100605 PMID: 37577053

13. Hughes AC. Wildlife trade. Curr Biol. 2021;31(19):R1218–24. https://doi.org/10.1016/j.cub.2021.08.056 PMID: 34637735

14. Lawrence S, Elliott C, Huisman W, Dean M, van Ruth S. The 11 sins of seafood: assessing a decade of food fraud reports in the global supply chain. Compr Rev Food Sci Food Saf. 2022;21(4):3746–69. https://doi.org/10.1111/1541-4337.12998 PMID: 35808861

15. Reilly A. Overview of food fraud in the fisheries sector. Rome, Italy: Food and Agriculture Organization of the United Nations; 2018. 32 p. http://www.fao.org/3/i8791en/I8791EN.pdf

16. Hobbs CAD, Potts RWA, Bjerregaard Walsh M, Usher J, Griffiths AM. Using DNA barcoding to investigate patterns of species utilisation in UK shark products reveals threatened species on sale. Sci Rep. 2019;9(1):1028. https://doi.org/10.1038/s41598-018-38270-3 PMID: 30705397

17. Ho JKI, Puniamoorthy J, Srivathsan A, Meier R. MinION sequencing of seafood in Singapore reveals creatively labelled flatfishes, confused roe, pig DNA in squid balls, and phantom crustaceans. Food Control. 2020;112:107144. https://doi.org/10.1016/j.foodcont.2020.107144

18. O'Mahony PJ. Finding horse meat in beef products--a global problem. QJM. 2013;106(6):595–7. https://doi.org/10.1093/qjmed/hct087 PMID: 23625529

19. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270(1512):313–21. https://doi.org/10.1098/rspb.2002.2218 PMID: 12614582

20. Verzeletti A, Cortellini V, Cisana S, Pretto S, Cerri N, De Ferrari F. Genetic approach for species identification: forensic application of DNA barcoding. Forensic Sci Int: Genet Suppl Ser. 2015;5:e576–8. https://doi.org/10.1016/j.fsigss.2015.09.228

21. Ewart KM, Frankham GJ, McEwing R, Webster LMI, Ciavaglia SA, Linacre AMT, et al. An internationally standardized species identification test for use on suspected seized rhinoceros horn in the illegal wildlife trade. Forensic Sci Int Genet. 2018;32:33–9. https://doi.org/10.1016/j.fsigen.2017.10.003 PMID: 29035720

22. Lee JC-I, Hsieh H-M, Huang L-H, Kuo Y-C, Wu J-H, Chin S-C, et al. Ivory identification by DNA profiling of cytochrome b gene. Int J Legal Med. 2009;123(2):117–21. https://doi.org/10.1007/s00414-008-0264-0 PMID: 18612647

23. Lee JC-I, Tsai L-C, Liao S-P, Linacre A, Hsieh H-M. Species identification using the cytochrome b gene of commercial turtle shells. Forensic Sci Int Genet. 2009;3(2):67–73. https://doi.org/10.1016/j.fsigen.2008.10.005 PMID: 19215874

24. Mitani T, Akane A, Tokiyasu T, Yoshimura S, Okii Y, Yoshida M. Identification of animal species using the partial sequences in the mitochondrial 16S rRNA gene. Leg Med (Tokyo). 2009;11 Suppl 1:S449-50. https://doi.org/10.1016/j.legalmed.2009.02.002 PMID: 19289300

25. Vences M, Thomas M, van der Meijden A, Chiari Y, Vieites DR. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Front Zool. 2005;2(1):5. https://doi.org/10.1186/1742-9994-2-5 PMID: 15771783

26. Hu Z, Wang Z, Zhang S, Bian Y, Li C. Species identification through pyrosequencing 12S rRNA gene. Forensic Sci Int: Genet Suppl Ser. 2015;5:e561–3. https://doi.org/10.1016/j.fsigss.2015.09.222

27. Balitzki-Korte B, Anslinger K, Bartsch C, Rolf B. Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene. Int J Legal Med. 2005;119(5):291–4. https://doi.org/10.1007/s00414-005-0537-9 PMID: 15875208

28. French I, Wainwright BJ. DNA barcoding identifies endangered sharks in pet food sold in Singapore. Front Mar Sci. 2022;9. https://doi.org/10.3389/fmars.2022.836941

29. Hsieh H-M, Huang L-H, Tsai L-C, Liu C-L, Kuo Y-C, Hsiao C-T, et al. Species identification of Kachuga tecta using the cytochrome b gene. J Forensic Sci. 2006;51(1):52–6. https://doi.org/10.1111/j.1556-4029.2005.00004.x PMID: 16423223

30. Naga Jogayya K, Meganathan PR, Dubey B, Haque I. Mitochondrial 16S ribosomal RNA gene for forensic identification of crocodile species. J Forensic Leg Med. 2013;20(4):334–8. https://doi.org/10.1016/j.jflm.2012.09.018 PMID: 23622485

31. Ratnasingham S, Hebert PDN. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes. 2007;7(3):355–64. https://doi.org/10.1111/j.1471-8286.2007.01678.x PMID: 18784790

32. PAW Forensic Working Group. Current forensic tests available for the investigation of wildlife crime in the UK; 2024. Available from: https://www.tracenetwork.org/wp-content/uploads/2024/01/UK-wildlife-forensic-costs-and-contacts_Jan2024.pdf

33. Locatelli NS, McIntyre PB, Therkildsen NO, Baetscher DS. GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. Proc Natl Acad Sci U S A. 2020;117(51):32211–2. https://doi.org/10.1073/pnas.2007421117 PMID: 33234565

34. Meiklejohn KA, Damaso N, Robertson JM. Assessment of BOLD and GenBank - their accuracy and reliability for the identification of biological materials. PLoS One. 2019;14(6):e0217084. https://doi.org/10.1371/journal.pone.0217084 PMID: 31216285

35. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol. 1994;39(2):174–90. https://doi.org/10.1007/BF00163806 PMID: 7932781

36. Calabrese FM, Balacco DL, Preste R, Diroma MA, Forino R, Ventura M, et al. NumtS colonization in mammalian genomes. Sci Rep. 2017;7(1):16357. https://doi.org/10.1038/s41598-017-16750-2 PMID: 29180746

37. Marshall C, Parson W. Interpreting NUMTs in forensic genetics: seeing the forest for the trees. Forensic Sci Int Genet. 2021;53:102497. https://doi.org/10.1016/j.fsigen.2021.102497 PMID: 33740708

38. Masters A, Ogden R, Wetton JH, Dawnay N. Defining end user requirements for a field-based molecular detection system for wildlife forensic investigations. Forensic Sci Int. 2019;301:231–9. https://doi.org/10.1016/j.forsciint.2019.05.041 PMID: 31181408

39. Esmail N, Wintle B, Athanas A, Beale C, Bending Z, Dai R, et al. Emerging illegal wildlife trade issues: a global horizon scan. Conserv Lett. 2020;13:e12715. https://doi.org/10.1111/conl.12715

40. Cuber P, Chooneea D, Geeves C, Salatino S, Creedy TJ, Griffin C, et al. Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. Ecol Genet Genomics. 2023;28:100181. https://doi.org/10.1016/j.egg.2023.100181

41. 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;12(10):e0184741. https://doi.org/10.1371/journal.pone.0184741 PMID: 28977016

42. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. Gigascience. 2018;7(4):giy033. https://doi.org/10.1093/gigascience/giy033 PMID: 29617771

43. Johri S, Solanki J, Cantu VA, Fellows SR, Edwards RA, Moreno I, et al. "Genome skimming" with the MinION hand-held sequencer identifies CITES-listed shark species in India's exports market. Sci Rep. 2019;9(1):4476. https://doi.org/10.1038/s41598-019-40940-9 PMID: 30872700

44. Parker J, Helmstetter AJ, Devey D, Wilkinson T, Papadopulos AST. Field-based species identification of closely-related plants using real-time nanopore sequencing. Sci Rep. 2017;7(1):8345. https://doi.org/10.1038/s41598-017-08461-5 PMID: 28827531

45. Seah A, Lim MCW, McAloose D, Prost S, Seimon T. MinION-based DNA barcoding of preserved and non-invasively collected wildlife samples. Genes. 2020;11(445). https://doi.org/10.1101/2020.01.29.925081

46. Vasiljevic N, Lim M, Humble E, Seah A, Kratzer A, Morf NV, et al. Developmental validation of Oxford Nanopore Technology MinION sequence data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification. Forensic Sci Int Genet. 2021;53:102493. https://doi.org/10.1016/j.fsigen.2021.102493 PMID: 33770699

47. Pomerantz A, Sahlin K, Vasiljevic N, Seah A, Lim M, Humble E, et al. Rapid in situ identification of biological specimens via DNA amplicon sequencing using miniaturized laboratory equipment. Nat Protoc. 2022;17(6):1415–43. https://doi.org/10.1038/s41596-022-00682-x PMID: 35411044

48. Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. In: Russell DJ, editor. Multiple sequence alignment methods. Totowa (NJ): Humana Press; 2014. p. 105–16.

49. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

50. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;30(5):1229–35. https://doi.org/10.1093/molbev/mst012 PMID: 23486614

51. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6. https://doi.org/10.1093/nar/gkab301 PMID: 33885785

52. Meusnier I, Singer GAC, Landry J-F, Hickey DA, Hebert PDN, Hajibabaei M. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics. 2008;9:214. https://doi.org/10.1186/1471-2164-9-214 PMID: 18474098

53. Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes. 2006;6(4):959–64. https://doi.org/10.1111/j.1471-8286.2006.01470.x

54. Tagliavia M, Nicosia A, Salamone M, Biondo G, Bennici CD, Mazzola S, et al. Development of a fast DNA extraction method for sea food and marine species identification. Food Chem. 2016;203:375–8. https://doi.org/10.1016/j.foodchem.2016.02.095 PMID: 26948627

55. Hasap L, Chotigeat W, Pradutkanchana J, Vongvatcharanon U, Kitpipit T, Thanakiatkrai P. A novel, 4-h DNA extraction method for STR typing of casework bone samples. Int J Legal Med. 2020;134(2):461–71. https://doi.org/10.1007/s00414-019-02232-9 PMID: 31897668

56. Srivathsan A, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, et al. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. BMC Biol. 2019;17(1):96. https://doi.org/10.1186/s12915-019-0706-9 PMID: 31783752

57. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9. https://doi.org/10.1093/bioinformatics/bty149 PMID: 29547981

58. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3(10):e000132. https://doi.org/10.1099/mgen.0.000132 PMID: 29177090

59. Sahlin K, Lim MCW, Prost S. NGSpeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. Ecol Evol. 2021;11(3):1392–8. https://doi.org/10.1002/ece3.7146 PMID: 33598139

60. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242

61. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330 PMID: 21653522

62. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6. https://doi.org/10.1038/nbt.1754 PMID: 21221095

63. Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC. Recurrent introgression of mitochondrial DNA among hares (Lepus spp.) revealed by species-tree inference and coalescent simulations. Syst Biol. 2012;61(3):367–81. https://doi.org/10.1093/sysbio/syr114 PMID: 22201159

64. Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN. BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries? PLoS One. 2020;15(4):e0231814. https://doi.org/10.1371/journal.pone.0231814 PMID: 32298363

65. Maestri S, Cosentino E, Paterno M, Freitag H, Garces JM, Marcolungo L, et al. A Rapid and accurate MinION-based workflow for tracking species biodiversity in the field. Genes (Basel). 2019;10(6):468. https://doi.org/10.3390/genes10060468 PMID: 31226847

66. Knot IE, Zouganelis GD, Weedall GD, Wich SA, Rae R. DNA barcoding of nematodes using the MinION. Front Ecol Evol. 2020;8. https://doi.org/10.3389/fevo.2020.00100

67. Toxqui Rodríguez M del S, Vanhollebeke J, Derycke S. Evaluation of DNA metabarcoding using Oxford Nanopore sequencing for authentication of mixed seafood products. Food Control. 2023;145:109388. https://doi.org/10.1016/j.foodcont.2022.109388

68. Kitpipit T, Chotigeat W, Linacre A, Thanakiatkrai P. Forensic animal DNA analysis using economical two-step direct PCR. Forensic Sci Med Pathol. 2014;10(1):29–38. https://doi.org/10.1007/s12024-013-9521-8 PMID: 24435950

69. Patterson EC. Combatting illegal trade in animal products by rapid on-site DNA sequencing, PhD Thesis. UK: University of Leicester; 2024. https://doi.org/10.25392/leicester.data.26056789.v1

70. Bello N, Francino O, Sánchez A. Isolation of genomic DNA from feathers. J Vet Diagn Invest. 2001;13(2):162–4. https://doi.org/10.1177/104063870101300212 PMID: 11289214

71. Hedmark E, Ellegren H. Microsatellite genotyping of DNA isolated from claws left on tanned carnivore hides. Int J Legal Med. 2005;119(6):370–3. https://doi.org/10.1007/s00414-005-0521-4 PMID: 15690184

72. Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, et al. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C R Biol. 2012;335(1):32–50. https://doi.org/10.1016/j.crvi.2011.11.002 PMID: 22226162

73. Leahy S. Revealed: seafood fraud happening on a vast global scale. The Guardian; 2021 [cited 2023 Mar]. Available from: https://www.theguardian.com/environment/2021/mar/15/revealed-seafood-happening-on-a-vast-global-scale

74. Webster LMI, Prigge T-L, Frankham GJ. A guide for the validation of DNA based species identification in forensic casework. FSIAE. 2024;5:100080. https://doi.org/10.1016/j.fsiae.2023.100080

75. UNODC. World wildlife crime report 2024: trafficking in protected species; 2024. https://www.unodc.org/cofrb/uploads/documents/ECOS/World_Wildlife_Crime_Report_2024.pdf

76. Fukushima CS, Mammola S, Cardoso P. Global wildlife trade permeates the Tree of Life. Biol Conserv. 2020;247:108503. https://doi.org/10.1016/j.biocon.2020.108503 PMID: 32454527

77. Ahlers N, Creecy J, Frankham G, Johnson RN, Kotze A, Linacre A, et al. 'ForCyt' DNA database of wildlife species. Forensic Sci Int: Genet Suppl Ser. 2017;6:e466–8. https://doi.org/10.1016/j.fsigss.2017.09.195

78. Dawnay N, Ogden R, Wetton JH, Thorpe RS, McEwing R. Genetic data from 28 STR loci for forensic individual identification and parentage analyses in 6 bird of prey species. Forensic Sci Int Genet. 2009;3(2):e63-9. https://doi.org/10.1016/j.fsigen.2008.07.001 PMID: 19215871

79. Collins RA, Cruickshank RH. The seven deadly sins of DNA barcoding. Mol Ecol Resour. 2013;13(6):969–75. https://doi.org/10.1111/1755-0998.12046 PMID: 23280099

80. Vasilita C, Feng V, Hansen AK, Hartop E, Srivathsan A, Struijk R, et al. Express barcoding with NextGenPCR and MinION for species-level sorting of ecological samples. Mol Ecol Resour. 2024;24(3):e13922. https://doi.org/10.1111/1755-0998.13922 PMID: 38240168

81. Morrison-Lanjouw S, Spijker R, Mughini-Gras L, Coutinho RA, Chaber AL, Leeflang M. A systematic review of the intercontinental movement of unregulated African meat imports into and through European border checkpoints. One Health. 2023;17:100599. https://doi.org/10.1016/j.onehlt.2023.100599 PMID: 37545542

82. EFSA Panel on Animal Health & Welfare, More S, Botner A, Butterworth A, Calistri P, Depner K, et al. Assessment of listing and categorisation of animal diseases within the framework of the Animal Health Law (Regulation (EU) No 2016/429): Ebola virus disease. EFSA J. 2017;15(7):e04890. https://doi.org/10.2903/j.efsa.2017.4890 PMID: 32625555

83. Judson SD, Fischer R, Judson A, Munster VJ. Ecological contexts of index cases and spillover events of different ebolaviruses. PLoS Pathog. 2016;12(8):e1005780. https://doi.org/10.1371/journal.ppat.1005780 PMID: 27494600

84. Hebert PDN, Floyd R, Jafarpour S, Prosser SWJ. Barcode 100K specimens: in a single nanopore run. Mol Ecol Resour. 2025;25(1):e14028. https://doi.org/10.1111/1755-0998.14028 PMID: 39387679

85. Srivathsan A, Feng V, Suárez D, Emerson B, Meier R. ONTbarcoder 2.0: rapid species discovery and identification with real-time barcoding facilitated by Oxford Nanopore R10.4. Cladistics. 2024;40(2):192–203. https://doi.org/10.1111/cla.12566 PMID: 38041646