



Contents lists available at ScienceDirect

Forensic Science International: Animals and Environments

journal homepage: www.sciencedirect.com/journal/forensic-science-international-animals-and-environments

Review Article

Current methods, future directions and considerations of DNA-based taxonomic identification in wildlife forensics

Kelly A. Meiklejohn^{a,*}, Mary K. Burnham-Curtis^b, Dyan J. Straughan^b, Jenny Giles^c, M. Katherine Moore^d^a North Carolina State University College of Veterinary Medicine, Dept. of Population Health and Pathobiology, 1060 William Moore Drive, Raleigh, NC 27607, USA^b U.S. Department of the Interior, Fish and Wildlife Service, OLE-National Fish and Wildlife Forensic Laboratory, 1490 East Main Street, Ashland, OR 97520, USA^c Fin Forensics, 1145 10th Ave E #322, Seattle, WA 98102, USA^d Forensic Laboratory, Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 331 Fort Johnson Road, Charleston, SC 29412 USA

ARTICLE INFO

Keywords:

Specimen identification
Species identification
Genetic markers
Taxonomic assignment
Molecular forensics
Wildlife trade
Wildlife crime
Mitochondrial DNA

ABSTRACT

Wildlife forensic analyses are frequently concerned with taxonomic identification, and very often employ amplification and Sanger sequencing of informative regions of the genome to achieve this. The materials submitted to wildlife forensic laboratories for taxonomic identification span a wide scope, from plant and animal parts in trade to assemblages of incidental biota at crime scenes. As these analyses take place within the context of legal proceedings, the wildlife forensic community is subject to unique requirements and considerations. These requirements and considerations are quite different from those of human forensic DNA, and have driven standardization in this field. While there has been extensive debate over appropriate DNA-based methods for taxonomic identification of a wide variety of biota in research settings, there has been little discussion on the issues associated with this approach in the high scrutiny environment of forensic science. This review outlines: key procedural and biological factors that may impact the accuracy of interpretation and reporting taxonomic identifications; resulting conventions employed by the wildlife forensics community; and implications for the use of emergent DNA sequencing technologies in taxonomic identification of wildlife in casework.

1. Introduction

One of the major aims of wildlife forensic analysis is to identify the taxonomic source (e.g., species, genus, etc.) of non-human biological material, either as items in trade, or direct or indirect evidence in other crimes. The trade in animals, plants, and their derivatives is expansive, and is one of the drivers of the current sixth mass extinction event [1]. Trade can be either licit (e.g., legally caught and marketed yellowfin tuna, carvings made of cow bone, boots made of rancher ostrich) or illicit (e.g., prohibited Goliath grouper, carvings made of elephant ivory, boots made of sea turtle). In some trades, such as timber and fisheries, the licit and illicit products are commonly co-mingled. In addition to the wide taxonomic scope of analyses required to detect violations in wildlife trade, wildlife forensic analyses may involve identification of direct evidence in wildlife crime (e.g., the alleged poisoning of a protected animal species) or of non-human biological trace evidence in human criminal investigations (e.g., pet hair, pollen, diatoms, insect

parts, etc.). Identifications of assemblages of trace evidence can provide probative information for investigative leads, either to link the crime scene to the victim and/or perpetrator, or to determine where a crime occurred based on the suite of species present.

Taxonomic classification is a man-made construct to categorize organisms based on observed similarities (or differences) in underlying measured traits. Though our understanding of taxonomy changes with the acquisition of new knowledge and can result in revisions to established names, it should be recognized that species categorizations are by and large "real" [2]. Morphological classification is the foundational method for taxonomic identification of biological materials. Morphological examination is a reliable, inexpensive method for identifying items encountered in forensic casework when sufficient diagnostic features are present, and applicable expertise and suitable comparative reference material are available (e.g., whole animals, intact bones, flight feathers, as in Trail [3]). When these conditions cannot be met and genetic capability is available, DNA is often used for taxonomic

* Corresponding author.

E-mail address: kameikle@ncsu.edu (K.A. Meiklejohn).<https://doi.org/10.1016/j.fsiae.2021.100030>

Received 4 June 2021; Received in revised form 14 September 2021; Accepted 17 September 2021

Available online 20 September 2021

2666-9374/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

identification. Wildlife forensic DNA analysts almost universally rely on sequencing of mitochondrial DNA (mtDNA) for specimen identification of animals (e.g., [4]), and a range of markers from the nuclear and chloroplast genomes of plants (e.g., [5,6]). The position and order of nucleotide bases in informative regions are the class characters used for diagnosis. Sequences from in-house and/or public databases are used as reference comparisons, and specimen identification is based on the degree of similarity between the unknown sequence and the reference. Such identifications are straightforward for well-separated groups with all closely related taxa characterized for a given genetic region, but are increasingly difficult in groups with incomplete taxon sampling and shallow coalescent depths [7].

This review paper aims to provide a perspective on current methods, future directions and considerations of DNA-based specimen identification in wildlife forensics. We take a deeper dive than previous reviews [8,9] and discuss: 1) commonly used regions for animals and plants; 2) sources of known sequences; 3) distance- and tree-based approaches for taxonomic identification; 4) biological considerations when using mitochondrial and chloroplast regions; 5) applications of next generation sequencing; and 6) minimum standards and best practices for DNA-based taxonomic identification. General requirements for laboratory practice involving DNA evidence in forensic casework, including those governed by laboratory accreditation (ISO 17025) and analyst certification, are not addressed. We wish to emphasize here that the analysis we refer to as “taxonomic identification” or “specimen identification” involves assigning unknown evidentiary material to an already-described taxon, and is not meant to include species discovery and description, which is a task best left to specialized taxonomists [10].

2. Body

2.1. Commonly used regions for taxonomic identification of animals and plants

Regions from the mitochondrial (mt) genome are the primary choice for forensic taxonomic identification of animal products because of the overwhelming availability of that molecule; the high copy number, cellular location and other molecular features of the mitochondrial genome means it is more likely to be recovered from forensic evidence which often contains DNA of low quality and quantity [11,12]. Given that mtDNA is haploid and maternally inherited, there is a lower effective population size among individuals of the same species [13]. Mitochondrial DNA also shows accelerated rates of evolution and differences in the rate of fixation of mutations through adaptive evolution [11,14,15]. These mutations can be harnessed to permit discrimination of closely related species. The ability to design universal primers that permit amplification across a broad range of taxa also makes mtDNA highly advantageous in wildlife forensics [8,16,17] as often only general information is known about the evidence sample prior to analysis (e.g., sample from a shark but without information about what family of shark). Further, given that the majority of initial molecular evolutionary studies targeted mtDNA [18,19], an abundance of foundational information and sequence data exist in the published literature and public sequence databases, respectively (e.g., [11,20,21]).

Although the mtDNA molecule evolves rapidly, the genes encoded in mtDNA do not mutate at equal rates [11]. This has an impact on the informativeness of individual mitochondrial regions; some regions have a lower mutation rate (e.g., cytochrome c oxidase subunit 1 (CO1) or 12S ribosomal RNA (12S)) and will be more useful for taxa that are well-separated or for higher level taxonomic classifications, while those that have a greater mutation rate (e.g., control region (CR), cytochrome b (cytB) or NADH dehydrogenase subunit 2 (ND2)) will be more useful for species level classifications of closely-related taxa [17]. The genetic diversity at a given region varies due to differences in the rate of mutation and/or the rate of fixation of adaptive mutations among taxa, and these factors impact the informativeness of different regions for

taxonomic identification. Further, the same mtDNA region can evolve at different rates in different taxa (e.g. [19,22–25]), meaning while a given mitochondrial region may be informative at the species level in one taxon, it may only be informative to the genus level in another. For example, *cytB* is used to discriminate among species of rhinoceros (Perissodactyla: Rhinocerotidae), but it is only effective at the genus level among deer in the genus *Odocoileus* (Artiodactyla: Cervidae), even though both groups are composed of closely related species of large mammals [26,27]. Thus, when selecting a mitochondrial region for use in taxonomic identification, it is important to have as complete a picture as possible of the level of sequence variation within and between species in the taxonomic group of interest.

The mitochondrial regions initially used for taxonomic identification in wildlife forensics are closely aligned with those used in early molecular evolutionary studies (e.g., [16,17]) based on restriction fragment length polymorphism (RFLP) and enzymatic techniques. Since DNA sequence analysis has become more common, there has been a desire to find a single region to serve as a ‘silver bullet’ for taxonomic identification. In 2003, a consortium of scientists promoted adoption of a single region for molecular taxonomic assignment of all animals, specifically a ~650-bp region of *CO1* [28]. While *CO1* offers species-level discrimination in many vertebrate and invertebrate taxa, closely related species can share the same *CO1* sequence (e.g., [29–32]). Other faster evolving mitochondrial regions, such as subunits of the NADH dehydrogenase and the CR are often used to delineate species [17,33,34] and characterize intraspecific relationships (i.e., subspecies, strains, etc.) (e.g., [35–37]). To highlight the breadth of mtDNA regions used in casework, wildlife forensic practitioners provided information about the commonly used regions for DNA-based taxonomic identification, noting that often multiple regions are used (Fig. 1).

In plants, the mitochondrial genome evolves slowly, meaning that mtDNA regions are typically not good candidates for DNA-based taxonomic identification [5,6]. A range of regions from the chloroplast

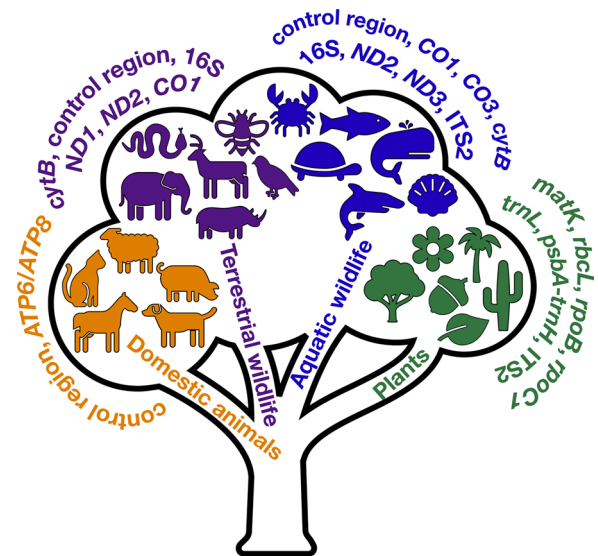


Fig. 1. Schematic highlighting the four main groups of taxa submitted to wildlife forensic laboratories for taxonomic identification, along with regions targeted for analysis (wrapped around the outside of the tree). Abbreviations are as follows: 1) Mitochondrial regions – 12S, 12S ribosomal RNA; 16S, 16S ribosomal RNA; *cytB*, cytochrome b; *CO1*, cytochrome c oxidase subunit 1; *CO3*, cytochrome c oxidase subunit 3; *ND1*, NADH dehydrogenase subunit 1; *ND2*, NADH dehydrogenase subunit 2; *ND3*, NADH dehydrogenase subunit 3; *ATP6*, ATP synthase membrane subunit 6; *ATP8*, ATP synthase membrane subunit 8. 2) Nuclear regions – *ITS2*, internal transcribed spacer subunit 2. 3) Chloroplast regions – *matK*, maturase K; *rbcL*, ribulose biphosphate carboxylase; *rpoB*, DNA-directed RNA polymerase subunit beta; *rpoC1*, plastid-encoded RNA polymerase subunit beta.

genome have long been used in evolutionary studies of land plants (e.g., [38,39]), given the higher substitution rate within the chloroplast genome. In 2009, the Consortium of the Barcode of Life completed a broad assessment of the utility of seven regions from the chloroplast genome across 397 species [5]. They identified a 2-region combination that provided the best compromise with respect to species level discrimination (72 %), amplification success, and sequence quality: ribulose biphosphate carboxylase (*rbcL*) and maturase K (*matK*) [5]. Supplementary regions from both the chloroplast genome (e.g., *trnL*, *psbA-trnH*, *rpoB*, *rpoC1*) and nuclear genome (primarily the internal transcribed spacer subunit 2 (ITS2)) are often required to ensure accurate specimen identification in some groups (Fig. 1).

2.2. Sources of known sequences for taxonomic identification

Accurate DNA-based classification of animal and plant material to species of origin is critically dependent on the availability of reference sequences for the taxonomic group of interest. These reference sequences must be from an informative gene region for the taxonomic level in question, be derived from accurately identified reference specimens, and achieve sufficient taxon sampling to rule out all other possibilities.

2.2.1. In-house reference sequence databases

Laboratories that routinely perform DNA-based classification of specific taxa often have developed an in-house reference database for this purpose, based on genetic sequences generated at that facility. The key benefit of an in-house database over a collection of genetic sequences derived from an external source is the ability to directly control and document the suitability of those sequences as reference material for forensic casework, including: 1) the suitability of originating reference specimens and region(s) sequenced; 2) completeness of taxon sampling; and 3) the integrity of the link between the electronic sequence data, the tissue sample and the originating specimen. Sequences to be deposited into an in-house database are obtained from reference material meeting a defined set of minimum quality, provenance and documentary criteria. Ideally, each reference sequence in the database would demonstrably link back to a specimen with sufficient species-diagnostic morphological characters. This specimen would be permanently retained in a collection to enable subsequent examination and taxonomic revision [40]. DNA or tissue samples obtained from collaborators are often used to augment the in-house database; though the originating specimens are not retained in the forensic facility's collection, they should have been identified by an appropriate expert, with morphological characters defined by published taxonomic studies and for which suitable metadata exists (e.g., collection locality, collection date, name of taxonomic expert performing the identification, etc.). To the fullest extent possible, reference specimens should be documented in such a manner that their taxonomic identity can be verified should questions arise [41], for example with photographs of species-diagnostic traits.

In practice, because of the difficulty of sourcing reference specimens, especially for rare taxa, in-house databases are often comprised of a core dataset generated in-house, augmented by additional sequence data meeting defined criteria (e.g., datasets from peer-reviewed phylogenetic or phylogeographic studies that include accessioned museum specimens). The core dataset may, for example, establish the discriminatory power of a given region for a desired level of inquiry in a specific taxonomic group, or may consist of sequences derived from accessioned museum specimens for taxa frequently misidentified in the field. For taxonomically complex groups, as many species as possible should be represented by genetic data derived from accessioned type specimens. The metadata associated with genetic sequences, tissue samples, and originating specimens are to be stored in a suitable relational database to ensure the integrity of the association among them (e.g., [42–44]). Forensic laboratories adhering to published standards implement multiple layers of checks and balances to ensure data and specimen

integrity, and for isolating possible explanations for unexpected data.

2.2.2. Public sequence databases

Publicly available DNA databases contain a wealth of information contributed by researchers around the world, and while their use requires caution, they are often suitable sources to augment in-house reference sequence databases, or to supply sequences of infrequently encountered species to laboratories unable to source material from a curated collection. There are numerous large public databases that contain DNA sequences, with the largest and most well-known being the International Nucleotide Sequence Databases (INSD), which is managed collaboratively by GenBank, the European Molecular Biology Laboratory (EMBL) and DNA Data Bank of Japan (DDBJ) [45]. Since the public release of the GenBank database in 1982 by the National Center for Biotechnology Information (NCBI), hundreds of millions of DNA sequences have been made publicly accessible, providing reference sequence data for mitochondrial, chloroplast and nuclear regions from hundreds of thousands of organisms [46]. More recently, whole genome sequences have been uploaded to GenBank, providing a rich array of material for sequence comparisons. Aside from GenBank, several other public sequence databases exist, some of which are taxon and region specific, for example, UNITE for eukaryote ITS sequences, and the Barcode of Life DataSystems (BOLD) for 'barcode' region sequences from plants, animals and fungi.

In addition to using public databases to source reference sequences, wildlife forensic geneticists use the native search interfaces of public sequence databases (e.g., BOLD, NCBI nucleotide search) at least occasionally for putative determination or exclusion of taxa at higher taxonomic levels, screening for contamination, or guiding marker choice with taxa not routinely processed by that laboratory. Identifications seldom rest solely on public database search results, as the output of such searches often do not contain enough information on the validity of the returned reference sequences. When using public databases, interpreting results relies on knowledge of relevant phylogenetic information, such as the composition of the genus or family in question, and any challenges at the species level. Additionally, species nomenclature undergoes constant review and revision, to not only resolve instances of synonymy (a single species that has two taxonomic names) but also to describe newly recognized species. Thus, the nomenclature for older submissions in some public databases may not be consistent with currently accepted nomenclature (e.g., the currently accepted *Ursus americanus* vs. the older, invalid name *Euarctos americanus* for the American black bear), or there may not be a consensus on the nomenclature for a species (e.g., *Cervus elaphus* vs. *Cervus canadensis* for red deer).

It is important to re-emphasize that analysts should be mindful that while sequences in a public database can be a valuable resource, the presence of incorrectly identified and erroneous sequence data has been well-documented (e.g., [28,47–53]). To remedy this to some extent, GenBank has developed a Reference Sequence (RefSeq) collection, containing public data that have been subjected to varying levels of validation, annotation and manual curation by staff [54]. However, the number of sequences in this collection is limited, such that it would not include sufficient representatives for many wildlife forensic applications. For example, the total number of *COI* sequences in GenBank is ~3,370,000 (as of March 2021), while the *COI* sequences in the RefSeq collection number just over 2,000. While an improvement in quality over other GenBank sequences, RefSeq samples are not required to meet the standard of validation for a forensic reference collection. A higher level of curation is performed on all sequences deposited in BOLD, including confirming that the sequence is not from a contaminant, is derived from a protein coding gene, and is a functional gene copy [55]. Importantly however, BOLD does source data from GenBank for inclusion in their publicly available dataset, thus misidentified and erroneous data may appear in BOLD search results. As data from more species are added to public databases, it will be easier to identify and remove

potentially misidentified sequences [56]. Fundamentally, using DNA sequences obtained from a public repository as reference data in a forensic setting requires substantial domain knowledge and interpretation by the analyst to reach an accurate identification at an appropriate taxonomic level.

2.3. Approaches for taxonomic identification

After sequencing the DNA from an unknown and gathering relevant reference sequences from in-house or public databases, two main approaches are typically undertaken to achieve taxonomic identification in wildlife forensics: distance-based and/or tree-based. It should be noted that identifications are often made using a combination of both approaches, unless a single approach has been validated for the taxon in question.

2.3.1. Distance-based approaches

Comparing the genetic distance, a measure of the number of nucleotides that differ between two sequences, is a common approach used for taxonomic identification. The Kimura-2-Parameter (K2P) model is often used to calculate genetic distances as it: 1) allows for different, undefined substitution rates between transitions and transversions; and 2) works effectively when nucleotide variation is limited [57]. Specimen identification based on calculated genetic distances often relies on either a 'gap' or 'threshold'. In a 'gap' based approach, the intraspecific genetic distance (*i.e.*, between individuals of a single species) must not overlap the interspecific genetic distance (*i.e.*, between individuals of differing species). The most common use of a 'gap' for taxonomic identification is for the *COI* barcode region, in which the general rule of thumb is that the intraspecific variation should be <3 % and interspecific variation should be >3 %, or 10X the mean of the intraspecific variation [58,59]. In the absence of complete taxonomic sampling or when the efficacy of a new region is being established, an 'experience threshold' is used for taxonomic assignment based on empirical observations of forensic practitioners across many taxa. A conservative 'threshold' of 1 % is often implemented in the scientific literature to minimize misidentifications that can occur when conspecifics are not included in the reference database (*e.g.*, [10]), but it must be re-emphasized here that there is no single threshold that will differentiate all taxa [7,60].

As several publications have examined the shortcomings of using either a 'gap' or 'threshold' approach for specimen identification (*e.g.*, [10,31,60,61]), only the main concerns raised in those studies most pertinent to wildlife forensics will be highlighted here. First and foremost, broad implementation of a 'gap' or 'threshold' for specimen identification is often not appropriate, and can lead to both false negatives and false positives [7]. In closely related species or sub-species which have a shallow coalescent depth, nucleotide variation at commonly analyzed regions may be limited. This scenario could result in a 'false negative', where multiple taxa would be lumped together given the interspecific distances are shallower than the proposed 'gap' or 'threshold' [10,29]. This was highlighted for a pair of sister species of forensically important Australian flesh flies (Diptera: Sarcophagidae) [62] and tuna species within the genus *Thunnus* [63], whereby the interspecific distances using the *COI* barcode region were below 3 %. Alternatively, targeted regions in fast evolving species may exhibit higher than expected intraspecific variation, causing morphologically indistinguishable individuals to be assigned to more than one species (known as splitting) [29]. For example, a high 'false positive' rate was reported for marine gastropods (cowries) when a 2 % 'threshold' was applied; 20 % of taxa were inaccurately split into more than one species [29]. This was rectified when the 'threshold' for specimen identification was increased. A few smaller, but still noteworthy considerations for distance-based specimen identification for wildlife forensics include: 1) the length of the sequence, as calculations based only on highly variable sections of the targeted region can inaccurately inflate the genetic distance (and vice versa) [8]; 2) calculating the genetic distance between

focal species does not provide any indication as to the relationship between other species (*i.e.*, species X can be equidistant to both species Y and Z, but the distance to X gives no insight into the distance between Y and Z); and 3) species that cannot be assigned confidently and accurately based on the genetic distances may still be recovered as reciprocally monophyletic in a phylogram.

Before implementing genetic distances for specimen identification in wildlife casework, the accuracy of this approach for the chosen taxonomic group must be performed. Firstly, verifying whether individuals from the extent of a species' geographical range were included when setting the 'gap' or 'threshold' is needed; if individuals from only one population were sampled, the expected variation would reflect the 'local gap' rather than the 'global gap', and if used broadly could result in misidentifications [10,31]. Secondly, the minimum and maximum genetic variation of the species in question should be established. Studies have shown that using the mean inaccurately inflates the interspecific variation, resulting in misleading specimen identification [64]. Finally, it is prudent to compare genetic distances reported in the scientific literature to those calculated from in-house reference sequences. It is important to emphasize that wildlife forensic scientists are not alpha taxonomists. It is not within the scope of their role to describe new species or set thresholds for how species are defined (either genetically or morphologically). Rather, they draw from previously published research by taxonomic experts in zoology and botany and apply that information in a forensic context.

2.3.2. Tree-based approaches

Construction of phylograms, which visually display an estimate of the evolutionary relationships among taxa, is also commonly used for taxonomic identification. Using discrete data, such as scored morphological characters or protein or DNA sequences, software algorithms create phylograms. Regardless of the software used or data type, aligned data are required as input. Commonly used multiple sequence alignment software packages, such as ClustalW [65] and MAAFT [66], are freely available either as standalone packages, within larger sequence analysis software packages (*e.g.*, Sequencher, Geneious, CLC Genomics Workbench, MEGA), or within open source web-based platforms (*e.g.*, Galaxy, EMBL-EBI). As the sensitivity and accuracy of the resulting alignment can be impacted by the scoring scheme (*i.e.*, matches, mismatches) and gap penalty settings (*i.e.*, the cost to insert a gap into the alignment), care needs to be taken when choosing software settings for use. After manual review of the resulting alignment, evolutionary relationships are reconstructed using one of four main algorithms, each differing in complexity and a priori assumptions about the data: 1) Distance (*e.g.*, neighbor joining, UPGMA), based on pairwise distances; 2) Maximum Parsimony, in which the simplest hypothesis is preferred; 3) Maximum Likelihood, based on the most likely tree topology when assuming a specific model of evolution; and 4) Bayesian, a statistical inference method based on Bayes' theorem. Confidence in the resulting topology is gleaned by nodal support (*e.g.*, bootstrap percentages, posterior probability out of 1.0) and also branch length (long branches indicate more nucleotide substitutions). Assigning an unknown to a particular taxon is done by examining the placement of that unknown in the phylogram. For example, if the unknown sequence is recovered within a clade (or group) composed only of individuals of a single species (known as monophyly), then the unknown specimen would be identified as that species (Ross et al. [7] refers to this as a "strict" tree-based approach). An example phylogram, generated using *COI* barcode region sequences from forensically relevant canids, is shown in Fig. 2.

While reconstructing a phylogram for specimen identification might seem like a straight-forward endeavor, there are many considerations that are especially prudent to forensic casework. Firstly, the regions targeted for species-level assignments should not be used to infer evolutionary higher-level relationships (*e.g.*, families or orders) without further validation, as their utility precisely lies in species-level resolution. For instance, the *COI* barcode region is not suitable for inferring

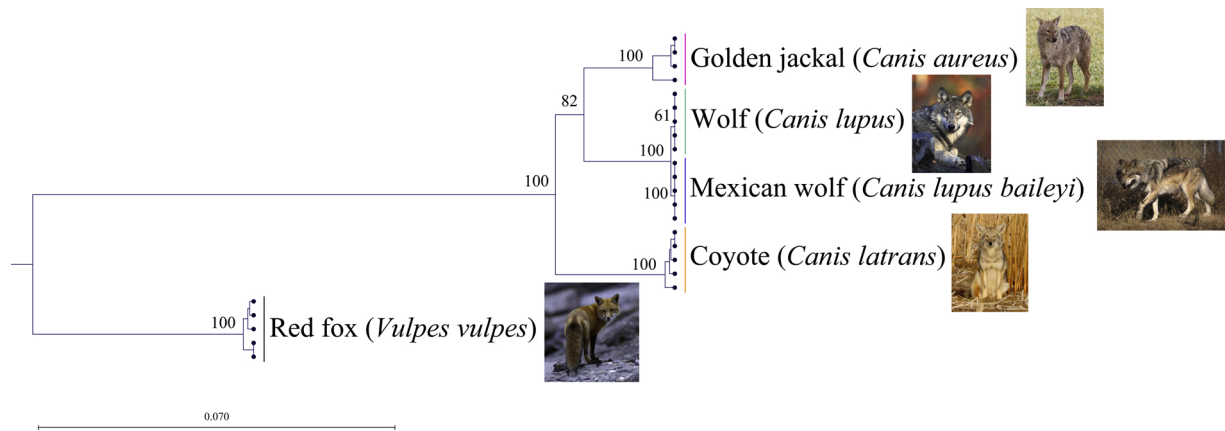


Fig. 2. Example phylogram for forensically relevant canid species. Sequences pertaining to the barcode region of *COI* were downloaded from GenBank and aligned in CLC Genomics Workbench (Qiagen). The UPGMA construction method and Kimura 80 nucleotide distance measure was used, and 1,000 replicates generated to determine bootstrap support. Red fox (*Vulpes vulpes*) was used as the outgroup. Multiple individuals of each species are recovered as a single clade with strong bootstrap support (100). Images sourced from the open source collection from the USFWS Digital Library and Wikipedia (Golden jackal).

inter-species relationships, and often higher-level nodes are recovered with poor support (*i.e.*, when that topology is recovered in <50 % of replicate trees; typically reported as a bootstrap support). Thus, if using a DNA region best suited for species-level relationships to instead infer higher-level relationships, analysts should proceed with caution or use data from other regions that are calibrated for higher taxonomic levels. Though forensic analysts are not called upon to determine if deep nodes in a tree accurately reflect evolutionary history, they do commonly “back up” and identify evidence sequences to genus or family when species determination is difficult due to shallow coalescent depth or incomplete taxonomic sampling. While species-level markers are generally safe for genus- and sometimes family-level diagnosis of an unknown nested within a monophyletic species clade, diagnosing higher-level taxonomic groupings of an unknown sequence becomes increasingly difficult with increasing genetic distance from the nearest known.

Secondly, every attempt should be made to ensure that reference sequences in the alignment not only include conspecifics and congeners from a range of biogeographical populations, but also where possible an appropriate outgroup. Collins and Cruickshank [60] emphasized that “comprehensive sampling and complete reference libraries [67,68], bring arguably the single biggest improvement to DNA barcode identification success [69].” Phylograms reconstructed with missing taxa and/or missing data can provide misleading results, and has been the topic of hundreds of published studies. For example, in a study targeting vertebrates, adding incomplete sequences (*i.e.*, those spanning only 10 % of the targeted region) of missing taxa greatly improved the resolution and associated node support in the resulting tree [70]. However, it is important to note that in some forensic scenarios complete taxon sampling is not feasible. Sampling of rare and endangered species is heavily regulated to protect them, and sampling remote populations can be challenging, especially if the wildlife in question is difficult to capture. In such scenarios, the resulting tree should be interpreted with caution.

2.4. Biological considerations when using mitochondrial and chloroplast regions for taxonomic identification

When using DNA-based approaches for taxonomic identification, it is important to take into account that biological processes such as heteroplasmy, polyploidy, introgression, hybridization, nuclear copies of mitochondrial genes (nUMTs) and incomplete lineage sorting can cause interpretation and reporting issues when not known and accounted for.

2.4.1. Heteroplasmy

Heteroplasmy is the presence of more than one mtDNA haplotype in an individual. Since mtDNA is a rapidly evolving molecule, new mutations arise frequently among the thousands of copies present in a cell as the result of imperfect replication and repair events [71,72]. Aside from random mutations, paternal mtDNA leakage during early development has been reported as a mechanism for heteroplasmy in some taxa (*e.g.*, *Drosophila* and cicadas [73]). Regardless of heteroplasmy origin, the means by which these new haplotypes become predominant in a maternal line are not well understood [71]. There are some taxa in which heteroplasmy is more common, for example bivalves [74], crickets [40], bees [73], lizards [75], and treefrogs [76]. For most animals, the actual prevalence of mtDNA recombination is not known, and it is thought to be maintained in populations through random genetic drift and mutation [77]. As heteroplasmy can produce functional haplotypes, identifying heteroplasmic sequences via the presence of premature stop codons or frameshift mutations is typically fruitless [73]. Rather, heteroplasmy is typically characterized by multiple base positions in a single Sanger sequence that have two alternate peaks. Cloning and/or bidirectional sequencing of a region may resolve the predominant sequence, or may just confirm that a length or point heteroplasmy is not a sequencing artifact. In the latter case, analysts should not attempt to assign one nucleotide, but rather use a degenerate base instead [73,78]. The inclusion of unverified heteroplasmic sequences in downstream data analysis could lead to incorrect specimen identification, as the number of unique species would be overestimated [79]. For example, when using *COI* for species identification in a genus of bees (*Hylaeus*), only 75 % of species known to have heteroplasmy were correctly recovered on a phylogenetic tree [73]. The inclusion of highly variable heteroplasmic sequences contributed to individuals of morphologically identifiable species being split into multiple clades on the phylogenetic tree. Thus, care must be taken when dealing with a forensic unknown, as heteroplasmic sequences can exacerbate difficulties of taxonomic identification within poorly separated taxa [73].

2.4.2. Polyploidy

Polyploidy, a form of reticulate evolution, refers to a biological condition in which an organism acquires additional copies (or chromosomes) of the genome [80,81]. While most common in plants – approximately 70 % of all angiosperms are estimated to be polyploids [82] – polyploidy has also been reported to a lesser extent in fungi [83], fish, insects, amphibians, reptiles, birds, molluscs, and mammals [80]. Polyploidy is classified into one of two types based on the mode of origin: 1) allopolyploids, where multiple copies originated from different species during hybridization; and 2) autopolyploids, where

multiple copies originated from genome duplication within a species during meiosis [80,84]. Regardless of the type, polyploidy generates substantial genetic and genomic variation, complicating taxonomic assignment; sequenced regions from polyploid individuals reflect species complexes which are difficult to accurately assign [85]. Studies have suggested that all copies of the target region should be sequenced from polyploid individuals (possible using either cloning or next generation sequencing) and included in subsequent analyses to ensure accurate assignment [86]. Considering this is not feasible for most wildlife forensic laboratories, a more appropriate solution would be to target regions derived through divergent evolution as they are less likely to be influenced by polyploidy (e.g., chloroplast regions in plants). Laboratories should be aware of the level of polyploidy documented in the scientific literature for the taxonomic group in question and choose regions appropriately, which in some cases might mean targeting multiple regions.

2.4.3. Introgression and hybridization

Hybridization can be defined as individuals from genetically distinct populations which interbreed to create offspring that possess genetic characteristics from each distinct group, while introgression is the gene flow between the hybridizing populations [87,88]. Introgression is most simply defined as the incorporation of genetic material from one species into the gene pool of a second divergent, but closely related species [89, 90]. The issues of hybridization and introgression are often discussed simultaneously with various species concepts, their definitions, and what constitutes a hybrid [88,91,92]. Wildlife forensic examiners, however, do not define what constitutes a species, and must use available information to identify the taxonomic source of an unknown item to the lowest taxonomic rank possible. It is estimated that approximately 10 % of animal species [93–95] and 25 % of plant species [94,95] hybridize to some extent. This number can be even higher in some groups. Ducks (Anatidae), for instance, have a rate of approximately 75 % [95], and hybridizations between wild and domestic species is well documented [96,97]. The complexities of hybridization and introgression are unique to each species; understanding the phylogenies of the species group in question, as well as a knowledge of anthropomorphic activities involving the species (e.g., hybridization with invasive species or intentional hybridizations [87,98,99]) is critical when ensuring correct taxonomic identification for forensic purposes.

The topic of identifying hybrids is beyond the scope of this paper, however there are considerations with regard to mtDNA region choice, types of hybrids (F1, F2, etc), geographic gene flow, and possible human mediated events that should be considered when using mtDNA for taxonomic identification. One example is wild canids in North America. Gray wolves *Canis lupus*, Eastern Gray wolves *C. lycaon* and coyotes *C. latrans* as well as domestic dogs *C. familiaris* are all known to naturally hybridize with each other, and multiple types of hybrids and back crossing events are possible [91,100,101]. Approximately 60 % of Eastern wolves in the Western Great Lakes region (Wisconsin, Michigan, Minnesota) of the United States have been documented to contain ‘coyote-like’ mitochondrial haplotypes as a result of a shared common ancestor, yet ongoing Eastern wolf x coyote hybridizations are known to occur in southern Ontario. Genetic studies on the relationships between these populations are numerous and somewhat contentious. The mitochondrial research has focused on the less conserved CR, and specific haplotypes have been defined as exclusive to specific populations. In this scenario, attempting to identify evidence items suspected to be a wolf using mtDNA regions without an understanding of the phylogeny of the organism, possible hybrids, or known introgression, can lead a forensic analyst to incorrectly assign the item in question to a coyote. Use of an additional nuclear region, combined with a clear understanding of the phylogeny and history of difficult hybridizing animals such as wolves, will avoid incorrect taxonomic identification.

Anthropomorphic mediated hybridization events can also lead to wide-spread introgression in natural populations. Throughout the 1990s

and early 2000’s, overharvesting of sturgeon from the Caspian Sea region (e.g., *Huso huso*, *Acipenser gueldenstaedtii/persicus* and *A. stellatus*) led to wide-spread replacement of caviar from other Acipenseriform species such as the North American white sturgeon *A. transmontanus* and American paddlefish *Polyodon spathula* [102]. Forensic scientists at the USFWS National Fish and Wildlife Forensics Laboratory tested over 6,000 caviars between the years 1998–2008, of which 85 % were declared as originating from Caspian Sea species. However, almost 30 % of caviars declared as Russian sturgeon *A. gueldenstaedtii* exhibited *cytB* sequences identical to Siberian sturgeon *A. baerii* [102]. This surprising result was not believed to occur as a result of intentional replacement, but rather that released Siberian sturgeon from aquaculture along the Danube River in Germany had successfully hybridized with wild Russian sturgeon, resulting in Siberian haplotype introgression into the wild Russian populations [102,103].

Knowledge of the geographic distribution of a species in question, as well as areas of overlapping distribution (sympatry) with hybridizing species and the amount of introgression is also an important consideration when identifying unknown samples. Members of the deer genus *Odocoileus* (*O. hemionus*, mule deer and black-tailed deer; *O. virginianus*, white-tailed deer) are wide spread across North America with a large range of sympatry between the species, and hybrids are known to exist [104–106]. Shared *cytB* sequences have been reported between white-tailed and mule deer, and it is currently unclear if that is due to introgression [107] or the paraphyletic sorting of shared mtDNA haplotypes [108]. Despite this, it should be noted that the degree of mtDNA divergence between white-tailed, black-tailed and mule deer is inconsistent with their current taxonomy; there is high sequence divergence between mule and black-tailed deer (both currently *O. hemionus*) in some geographic areas, and low sequence divergence between white-tailed and mule deer in others [105,109,110].

2.4.4. Nuclear copies of mitochondrial genes (nUMTs)

The utilization of mtDNA for specimen identification can be problematic since portions of the mitochondrial genome can become integrated into the nuclear genome, known as nuclear mitochondrial pseudogenes (nUMTs) [11,111–114]. Given there are fewer functional constraints when mtDNA segments are incorporated as non-coding regions of the nuclear genome, nUMTs typically accumulate mutations more quickly and can lead to the existence of distinct copies of some mitochondrial genes within an individual [111,115]. Using universal primers, true functional copies of mitochondrial genes and nUMTs can be amplified equally for downstream sequencing [79]. Subsequent taxonomic assignment based on nUMTs may provide misleading results, as the nUMT and the true mtDNA target sequence will have diverged more than anticipated for the taxon in question [79,111]. Most nuclear copies can, however be recognized by several identifiable characteristics, predominantly premature stop codons and indels, and eliminated prior to data analysis. Alternatively, nUMTs can be characterized in reference materials along with the mtDNA region from which they originated, and the data maintained in comparison databases.

2.4.5. Incomplete lineage sorting

It is commonplace in evolutionary biology to use a single gene region (or a combination of regions) to infer the relationships among a set of species using a phylogram that is typically referred to as a ‘gene’ tree. Incomplete lineage sorting (ILS) is a process which can cause incongruence between ‘gene’ and ‘species’ trees, complicating specimen identification [116]. ILS occurs when closely related species interbreed or have not completely diverged, resulting in the differing patterns of inheritance reflected by mitochondrial, chloroplast and nuclear DNA (or between different regions of a single DNA type). For instance, ancestral mitochondrial states may have been retained, while derived states are evident in the nuclear genome, or vice versa [117]. Phylogenetic relationships among species exhibiting ILS may be polyphyletic, such that taxa that do not share a recent common ancestor are clustered together

in a tree. This often occurs with species that have long generation times such as sturgeon [118]. One example from *Crotaphytus* lizards shows how evolutionary events rather than phylogenetic descent can be reflected in the mitochondrial genome, which could lead to erroneous taxonomic identification [119]. Caution should be taken when attempting to identify closely related species using a single highly conserved gene region.

2.5. Application of next generation sequencing to taxonomic identification

Next-generation sequencing (NGS) (also known as massively parallel sequencing and high throughput sequencing), has revolutionized molecular biology, making it easier, quicker and cheaper to generate large volumes of sequence data. Regardless of the specific NGS platform used (Illumina, PacBio, IonGeneStudio, MiniON, etc.), the underlying premise is the same: multiple DNA fragments are sequenced in parallel, and sophisticated bioinformatic workflows are used to process and interpret the resulting data [120]. Whilst some human-focused forensic laboratories have validated NGS for certain casework applications (e.g., Federal Bureau Investigation for mitochondrial control region sequencing [121]; genotyping by sequencing [122]), at the time of this review we are not aware of any ISO 17025 accredited forensic wildlife laboratories using NGS for taxonomic identification of wildlife in casework.

Several published studies have highlighted the utility and benefits that NGS could bring to wildlife casework. In addition to the higher throughput capabilities afforded by NGS, the increased sensitivity facilitates the recovery of full target sequences even from low quantity and quality DNA samples. Standard PCR and Sanger sequencing of hard matrices such as ivory, teeth, bone and timber is often difficult, as few intact DNA fragments remain. Using NGS, successful taxonomic assignment has been achieved for highly degraded ivory [123], mammoth tusks [124] and timber regulated by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) [125,126]. Food products encountered in forensic cases often pose a challenge for traditional analysis given that high temperatures along with mechanical and chemical treatment during processing can shear and degrade DNA. NGS has been applied to processed seafood products, in which taxonomic assignment is necessary to either confirm authenticity [127] or detect possible undisclosed allergens (e.g., crustaceans [128]). Commercial NGS assays have also been developed (ThermoFisher Scientific; A38452, A38454) to streamline taxonomic identification of meat and fish in food mixtures for authentication purposes [129]. The analysis of mixed samples can also be streamlined using NGS; individual reads are generated which can be assigned to separate taxa, whereas a mixed Sanger chromatogram cannot be reliably interpreted. Further, NGS can be used to detect and characterize taxa present in mixtures at very low levels (1%), which might confirm a CITES or mislabeling violation (e.g., [129–132]). For taxa in which multiple regions are needed to ensure reliable and accurate taxonomic identification, new techniques coupled with NGS can streamline the generation of data whilst conserving valuable DNA extract. For example, Shaili et al. [133] used “genome skimming” (a method in which all DNA templates in a sample are sequenced without PCR) on the portable MiniON sequencer to quickly generate full mitochondrial genome sequences for specimen identification of a CITES listed shark. To discriminate between wolves and coyotes, hybridization capture – a method in which biotinylated RNA “baits” complementary to the DNA of interest are used to isolate target DNA prior to NGS – was used to generate full mitochondrial genome sequences [140]. It is worth noting though, that cases from the research literature demonstrating what is technically possible in terms of detection of violations in the field may not currently be adequately validated for or compatible with the workflow of forensic casework, or may not be useful in an enforcement context [4].

Despite the highlighted advantages of using NGS for taxonomic

assignment, there are numerous considerations that are especially critical within a forensic context. First, while increased sensitivity is advantageous for processing challenging samples, it can also increase the likelihood that low level environmental and reagent contaminants are sequenced. This was highlighted by Dormontt et al. [134], who observed contamination in negative control samples, likely a result of sample carry over during DNA isolation. Likewise, contamination between evidence samples, which are often packaged together (Fig. 3), will be visible with NGS, meaning that validations must address acceptable levels of contamination. Each laboratory will need to conduct extensive validation studies to determine interpretation coverage thresholds and address error rates, which are inherent at some level with any NGS platform. Secondly, when universal primers are used with a mixed sample, bias in primer annealing may cause taxa with primer binding site mismatches to amplify poorly or not at all, leading to a false negative result [135,136]. Additionally, inhibitors that are commonly co-extracted with forensic samples can negatively impact the ligation of indices and adaptors during library preparation [135]. Without this ligation, downstream sequencing is not possible. Finally, there is a substantial cost investment to bring NGS into casework, which might be out of reach for smaller wildlife forensic laboratories. These costs extend beyond the instrumentation, to the reagents and consumables needed for validation and analyst training. Further, while NGS is more affordable than Sanger sequencing on a per nucleotide basis, this cost benefit is only capitalized when processing at high throughput. Given most wildlife forensic laboratories have low throughput, the cost of NGS still might be out of reach for at least the near term.

For laboratories considering implementing NGS in forensic casework, there are additional logistical challenges. Aside from completing the necessary validations, analysts would need to be trained on the laboratory workflows, which differ substantially from those currently used for Sanger sequencing. For simplicity, some of the time-consuming and analyst-sensitive library preparation steps (e.g., bead purification) can be completed by liquid handling robots. Further, the reagent and consumable set up for most NGS instruments is very straight-forward; the analyst is typically only required to pipette their sample into a reagent cartridge and subsequently load the cartridge into the instrument. A challenge unique to NGS concerns the storage and analysis of large amounts of sequence data. Given a single Illumina MiSeq v3 run generates up to 15 GB of data, laboratories would likely need a dedicated storage platform. Commercially developed and open source software programs are available to complete basic NGS analysis steps, such as



Fig. 3. It is common in wildlife forensic laboratories to receive evidence with many individuals packaged together, such as this assemblage of seahorses from multiple species submitted to the NOAA Forensic Laboratory. Cross-contamination between evidence items is seldom detected with PCR and Sanger sequencing, but could be apparent with NGS, which is more sensitive. Image kindly provided by M. Katherine Moore.

sample demultiplexing, quality filtering and primer trimming. To complement these, several bioinformatic pipelines have been developed to streamline taxonomic assignment, such as for metazoans [137], fish [138], CITES listed species (e.g., [125,132]), and taxa associated with environmental samples [139]. These pipelines provide a good starting point for the analysis of NGS data, and could be adapted to meet the specific needs (i.e., DNA region, taxa) of the laboratory.

2.6. Developed standards and guidelines for molecular taxonomic identification

To ensure forensic science is admissible in court, the methods used to analyze forensic evidence must be performed to recognized standards and guidelines. The Organization of Scientific Area Committees (OSAC) for Forensic Science was created in 2014 to address the lack of discipline-specific forensic science standards. This U.S.-based organization is composed of 550-plus members who are leading forensic practitioners from private and public forensic laboratories, industry, and academia. The two Biology subcommittees, Human and Wildlife Forensic Biology, have developed several general standards and guidelines for DNA analysis, interpretation, and reporting (e.g., ANSI/ASB 019, 048). Additional standards are directly applicable to DNA-based taxonomic identification in wildlife forensics, including training in DNA isolation and purification methods (ANSI/ASB 023), validating new primers for Sanger sequencing (ANSI/ASB 047), training in mtDNA analysis for taxonomic identification (ANSI/ASB 111) and report writing (ANSI/ASB 029). Further, forthcoming standards focusing on a) DNA sequencing using capillary electrophoresis, b) prevention, monitoring and mitigating DNA contamination, c) in-house sequence databases for taxonomic assignment of wildlife, and d) use of reference sequences from public databases for taxonomic identification, have been drafted. Published standards and documents pertinent to the interpretation and reporting of DNA-based taxonomic identification can be accessed online via <https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/wildlife-forensics-subcommittee> and <http://www.asbstandardsboard.org/published-documents/wildlife-forensics-published-documents/>.

3. Conclusions

One of the main requests of law enforcement to a wildlife forensic laboratory is identifying the species origin of an evidence item. As outlined in this review, amplification and Sanger sequencing of informative regions of the genome is the most commonly used approach in both animals and plants. Wildlife forensic laboratories often rely on the regions identified and analysis techniques developed by the research community, given that personnel and monetary resources needed to develop a tailored approach for a specific species in question are extremely limited. When completing DNA-based taxonomic identifications, wildlife forensic biologists have to employ both rigor and pragmatism: forensic science has little tolerance for error, the scope of species submitted for analysis are often much broader than those in most academic laboratories (e.g., a single laboratory could process samples from birds, fish, mammals, and timber), and the court system demands either a 'yes' or 'no' answer within a reasonable timeframe. Accurate taxonomic identification requires not only technical expertise in the laboratory, but also knowledge of evolutionary and coalescent theory, characteristics of mtDNA, and the phylogeny and biogeography of the species in question.

Disclaimer

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service or the National Oceanic and Atmospheric Administration. Mention of tradenames does not imply U.S. Government endorsement of

commercial products.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the members of the Wildlife Forensic Biology Subcommittee of the United States National Institute for Standards and Technology- The Organization of Scientific Area Committees for Forensic Science (NIST-OSAC), for their valuable feedback and contributions to this review article.

References

- [1] G. Ceballos, P.R. Ehrlich, The misunderstood sixth mass extinction, *Science* 360 (2018) 1080–1081.
- [2] J.C. Avise, D. Walker, Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome, *Proc. Natl. Acad. Sci. U. S. A.* 96 (3) (1999) 992–995.
- [3] P.W. Trail, Identifying Bald versus Golden Eagle bones: a primer for wildlife biologists and law enforcement officers, *J. Fish Wildl. Manag.* 8 (2017) 596–610.
- [4] C. Mori, S. Matsumura, Current issues for mammalian species identification in forensic science: a review, *Int. J. Legal Med.* 135 (2021) 3–12.
- [5] D.H. Janzen, P.M. Hollingsworth, L.L. Forrest, J.L. Spouge, M. Hajibabaei, S. Ratnasingham, M. Van der Bank, M.W. Chase, R.S. Cowan, D.L. Erickson, A. J. Fazekas, S.W. Graham, K.E. James, K. Kim, W.J. Kress, H. Schneider, J. Van AlphenStahl, S.C.H. Barrett, C. Van den Berg, D. Bogarin, K.S. Burgess, K. M. Cameron, M. Carine, J. Chacón, A. Clark, J.J. Clarkson, F. Conrad, D.S. Devey, C. Ford, T.A.J. Hedderson, M.L. Hollingsworth, B.C. Husband, L.J. Kelly, P. R. Kesanakurti, J.S. Kim, Y. Kim, R. Lahaye, H. Lee, D.G. Long, S. Madriñán, O. Maurin, I. Meusnier, S.G. Newmaster, C. Park, D.M. Percy, G. Petersen, J. E. Richardson, G.A. Salazar, V. Savolainen, O. Seberg, M.J. Wilkinson, D. Yi, D. P. Little, A DNA barcode for land plants, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 12794–12797.
- [6] P.M. Hollingsworth, S.W. Graham, D.P. Little, Choosing and using a plant DNA barcode, *PLoS One* 6 (2011) 1–13.
- [7] H.A. Ross, S. Murugan, W.L. Li, Testing the reliability of genetic methods of species identification via simulation, *Syst. Biol.* 57 (2008) 216–230.
- [8] R.N. Johnson, L. Wilson-Wilde, A. Linacre, Current and future directions of DNA in wildlife forensic science, *Forensic Sci. Int. Genet.* 10 (2014) 1–11.
- [9] M.K. Moore, K. Frazier, Humans are animals, too: critical commonalities and differences between human and wildlife forensic genetics, *J. Forensic Sci.* 64 (2019) 1603–1621.
- [10] R.A. Collins, R.H. Cruickshank, The seven deadly sins of DNA barcoding, *Mol. Ecol. Resour.* 13 (2013) 969–975.
- [11] C. Moritz, T.E. Dooling, W.M. Brown, Evolution of animal mitochondrial DNA: relevance for population biology and systematics, *Annu. Rev. Ecol. Syst.* 18 (1987) 269–292.
- [12] D.R. Foran, Relative degradation of nuclear and mitochondrial DNA: an experimental approach, *J. Forensic Sci.* 51 (2006) 766–770.
- [13] C.W. Birky Jr., T. Maruyama, P. Fuerst, An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results, *Genetics* 103 (1983) 513–527.
- [14] W.M. Brown, M. George, A.C. Wilson, Rapid evolution of animal mitochondrial DNA, *Proc. Natl. Acad. Sci. U. S. A.* 76 (1979) 1967.
- [15] J.E. James, G. Piganeau, A. Eyre-Walker, The rate of adaptive evolution in animal mitochondria, *Mol. Ecol.* 25 (2016) 67–78.
- [16] T.D. Kocher, W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca, A. C. Wilson, Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers, *Proc. Natl. Acad. Sci. U. S. A.* 86 (1989) 6196.
- [17] M.A. Cronin, D.A. Palmisciano, E.R. Vyse, D.G. Cameron, Mitochondrial DNA in wildlife forensic science: species identification of tissues, *Wildl. Soc. Bull.* 19 (1991) 94–105.
- [18] J.C. Avise, *Molecular Markers, Natural History and Evolution*, 2nd ed., Sinauer Associates, Sunderland (Massachusetts), 2004.
- [19] B. Nabholz, S. Glémin, N. Galtier, Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis, *Mol. Biol. Evol.* 25 (2008) 120–130.
- [20] S. Palumbi, A. Martin, S. Romano, W.O. McMillan, L. Stice, G. Grabowski, *The Simple Fool's Guide to PCR*, Dept. of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu, HI, 1991.
- [21] J.L. Boore, Animal mitochondrial genomes, *Nucleic Acids Res.* 27 (1999) 1767–1780.
- [22] J.C. Avise, B.W. Bowen, T. Lamb, A.B. Meylan, E. Bermingham, Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines, *Mol. Biol. Evol.* 9 (1992) 457–473.

- [23] G.C. Johns, J.C. Avise, A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene, *Mol. Biol. Evol.* 15 (1998) 1481–1490.
- [24] D. Huang, R. Meier, P.A. Todd, L.M. Chou, Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding, *J. Mol. Evol.* 66 (2008) 167–174.
- [25] S.A. Smith, M.J. Donoghue, Rates of molecular evolution are linked to life history in flowering plants, *Science* 322 (2008) 86–89.
- [26] H.-M. Hsieh, L.-H. Huang, L.-C. Tsai, Y.-C. Kuo, H.-H. Meng, A. Linacre, Lee J.C.-I, Species identification of rhinoceros horns using the cytochrome *b* gene, *Forensic Sci. Int.* 136 (2003) 1–11.
- [27] E.K. Latch, E.M. Kierepka, J.R. Heffelfinger, O.E.J. Rhodes, Hybrid swarm between divergent lineages of Mule Deer (*Odocoileus hemionus*), *Mol. Ecol.* 20 (2011) 5265–5279.
- [28] P.D. Bridge, P.J. Roberts, B.M. Spooner, G. Panchal, On the unreliability of published DNA sequences, *New Phytol.* 160 (2003) 43–48.
- [29] C.P. Meyer, G. Paulay, DNA barcoding: error rates based on comprehensive sampling, *PLoS Biol.* 3 (2005) e22.
- [30] R.D. Ward, T.S. Zemlak, B.H. Innes, P.R. Last, P.D.N. Hebert, DNA barcoding Australia's fish species, *Philos. Trans. Biol. Sci.* 360 (2005) 1847–1857.
- [31] R. Meier, K. Shiyang, G. Vaidya, P.K.L. Ng, DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success, *Syst. Biol.* 55 (2006) 715–728.
- [32] J. Waugh, DNA barcoding in animal species: progress, potential and pitfalls, *Bioessays* 29 (2007) 188–197.
- [33] S.P. Casey, H.J. Hall, H.F. Stanley, A.C.J. Vincent, The origin and evolution of seahorses (genus *Hippocampus*): a phylogenetic study using the cytochrome *b* gene of mitochondrial DNA, *Mol. Phylogenet. Evol.* 30 (2004) 261–272.
- [34] H.R.L. Lerner, D.P. Mindell, Phylogeny of eagles, Old World vultures, and other Accipitridae based on nuclear and mitochondrial DNA, *Mol. Phylogenet. Evol.* 37 (2005) 327–346.
- [35] S.A.M. Amer, Y. Kumazawa, Mitochondrial DNA sequences of the Afro-Arabian spiny-tailed lizards (genus *Uromastix*; family Agamidae): phylogenetic analyses and evolution of gene arrangements, *Biol. J. Linn. Soc.* 85 (2005) 247–260.
- [36] M. Venegas-Anaya, A.J. Crawford, A.H. Escobedo Galvan, O.I. Sanjurjo, L. D. Densmore 3rd, E. Bermingham, Mitochondrial DNA phylogeography of *Caiman crocodylus* in Mesoamerica and South America, *J. Exp. Zool.* 309 (2008) 614–627.
- [37] M.R. Buchalski, B.N. Sacks, D.A. Gille, M.C.T. Penedo, H.B. Ernest, S.A. Morrison, W.M. Boyce, Phylogeographic and population genetic structure of bighorn sheep (*Ovis canadensis*) in North American deserts, *J. Mammal.* 97 (2016) 823–838.
- [38] J.D. Palmer, Comparative organization of chloroplast genomes, *Annu. Rev. Genet.* 19 (1985) 325–354.
- [39] M.W. Chase, D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les, B.D. Mishler, M. R. Duvall, R.A. Price, H.G. Hills, Y.L. Qiu, K.A. Kron, J.H. Rettig, E. Conti, J. D. Palmer, J.R. Manhart, K.J. Sytsma, H.J. Michaels, H.J. Kress, K.G. Karol, W. D. Clark, M. Hedren, B.S. Gaut, R.K. Jansen, K.J. Kim, C.F. Wimpsey, J.F. Smith, G. R. Furnier, S.H. Strauss, Q.Y. Xiang, G.M. Plunkett, P.S. Soltis, S.M. Swensen, S. E. Williams, P.A. Gadek, C.J. Quinn, L.E. Eguarte, E. Golenberg, G.H. Learn, S. W. Graham, S.C.H. Barrett, S. Dayanandan, V.A. Albert, Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *RCBL*, *Ann. Mo. Bot. Gard.* 80 (1993) 528–580.
- [40] I.J. Harrison, P. Chakrabarty, J. Freyhof, J.F. Craig, Correct nomenclature and recommendations for preserving and cataloguing voucher material and genetic sequences, *J. Fish Biol.* 78 (2011) 1283–1290.
- [41] L. Prendini, R. Hanner, R. DeSalle, Obtaining, storing and archiving specimens and tissue samples for use in molecular studies, in: R. DeSalle, G. Giribet, W. Wheeler (Eds.), *Techniques in Molecular Systematics and Evolution*, Birkhäuser, Basel, 2002.
- [42] ISBER, 2012 best practices for repositories collection, storage, retrieval, and distribution of biological materials for research. International Society for Biological and Environmental Repositories, *Biopreserv. Biobank.* 10 (2012) 79–161.
- [43] L.D. Campbell, J.J. Astrin, Y. DeSouza, J. Giri, A.A. Patel, M. Rawley-Payne, A. Rush, N. Sieffert, The 2018 revision of the ISBER best practices: summary of changes and the editorial team's development process, *Biopreserv. Biobank.* 16 (2018) 3–6.
- [44] C.D. Phillips, J.L. Dunnum, R.C. Dowler, L.C. Bradley, H.J. Garner, K. A. MacDonald, B.K. Lim, M.A. Revezel, M.L. Campbell, H.L. Lutz, N.O. Garza, J. A. Cook, R.D. Bradley, S.T. Alvarez-Castañeda, J.E. Bradley, R.D. Bradley, L. N. Carraway, J.P. Carrera-E, C.J. Conroy, B.S. Coyner, J.R. Demboski, C.W. Dick, R.C. Dowler, K. Doyle, J.L. Dunnum, J.A. Esselstyn, E. Gutiérrez, J.D. Hanson, P. M. Holahan, T. Holmes, C.A. Iudica, R.N. Leite, T.E. Lee, B.K. Lim, J.L. Malaney, B.S. McLean, S.B. McLaren, N.D. Moncrief, L. Olson, N. Ordóñez-Garza, C. D. Phillips, M.A. Revezel, E.A. Rickart, D.S. Rogers, C.W. Thompson, N.S. Upham, P.M. Velazco, E. Heske, Curatorial guidelines and standards of the American Society of Mammalogists for collections of genetic resources, *J. Mammal.* 100 (2019) 1690–1694.
- [45] B. Soren, D. Antoine, H. Masahira, N. Haruki, S. Kazuo, M. Tara, P. Daphne, Nucleotide sequence database policies, *Science* 298 (2002) 1333.
- [46] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, *Nucleic Acids Res.* 41 (2013) D36–42.
- [47] V. Rytas, Taxonomic misidentification in public DNA databases, *New Phytol.* 160 (2003) 4–5.
- [48] K.E. Ashelford, N.A. Chuzhanova, J.C. Fry, A.J. Jones, A.J. Weightman, At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies, *Appl. Environ. Microbiol.* 71 (2005) 7724–7736.
- [49] R.H. Nilsson, M. Ryberg, E. Kristiansson, K. Abarenkov, K.-H. Larsson, U. Kõljalg, Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective, *PLoS One* 1 (2006) e59.
- [50] M.I. Bidartondo, T.D. Bruns, M. Blackwell, I. Edwards, A.F.S. Taylor, T. Horton, N. Zhang, U. Kõljalg, G. May, T.W. Kuyper, G. Gilbert, M. Wedin, D.W. Li, J. Y. Uchida, A. Jumpponen, A. Pringle, J. Deckert, R.M. Bateman, E.C. Villinga, H. J. Beker, C. Pashley, H.T. Lumbsch, S.O. Rogers, J. Xu, P. Johnston, R. A. Shoemaker, M. Liu, G. Marques, B. Summerell, S. Sokolski, A. Purvis, J. Borneman, J.E. Blair, K.H. Larsson, W. Chen, U. Thrane, P. Widden, J.N. Bruhn, V. Bianchinotti, D. Tuthill, T.J. Baroni, G. Barron, K. Hosaka, S.O. Suh, P. W. Crous, J.E. Janso, K. Jewell, M. Piepenbring, C. Schoch, G. Thorn, R. Sullivan, G.W. Griffith, S.G. Bradley, T. Aoki, D.H. Pfister, W.T. Yoder, Y.M. Ju, S. E. Carpenter, C. Hawkes, S.M. Berch, M. Trappe, W. Duan, S.A. Redhead, G. Bonito, M. Berbee, M. Binder, R.A. Taber, G. Coelho, G. Thorn, R. Abad, A. Ganley, T. Barraclough, R. Agerer, L. Nagy, B.A. Roy, T. Læssøe, E.W. Boehm, K. Ivors, N. Hallenberg, H.V. Tichy, G.M. Mueller, K. Voigt, J. Stalpers, E. Langer, A. Burt, M. Scholler, D. Krueger, S. Huhndorf, G. Pacioni, R. Pöder, T. Pennanen, S.N. Jeffers, M. Capelari, B. Arenz, K. Nakasone, J.P. Tewari, G.L. Andersen, R. H. Nilsson, W. Rossi, A.N. Miller, A.S. Methven, S. Schechter, P. Vance, D. Mahoney, Z.K. Atallah, S. Kang, A.W. Wilson, W. Schell, I. Anderson, I. Kohn, J.P. Rheeder, J. Mehl, M. Grief, G.N. Ngala, J. Ammirati, M. Kawasaki, Y. Gwo-Fang, T. Matsumoto, K.E. Conway, J.K. Harrison, S. Kroken, C. Decock, D. Smith, G. Koenig, I.A. Dickie, D. Luoma, T. May, M. Leonardi, L. Sigler, D.L. Taylor, L. Egerton-Warburton, J. Geml, C. Gibson, T. Sharpton, I. Alexander, J. Guarro, D. L. Hawksworth, J.C. Dianese, S.A. Trudell, P. Avis, B. Paulus, M. Padamsee, J. L. Mata, Z. Wang, P. Callac, N. Lima, M. White, J.M. Moncalvo, C. Barreau, S. T. Bates, M.A. Juncaí, B. Buyck, R.K. Rebeler, P. Dyer, M.R. Liles, D. Eastburn, S. Timonen, D. Estes, R. Carter, J.M. Herr Jr, J. Berube, G. Chandler, K. Kerekes, P. Bonello, G.H. Sung, J. Cruse-Sanders, R.G. Márquez, T.Z. DeSantis, E. Horak, M. Fitzsimons, H. Döring, R. Kjølner, S. Yao, J. Spatafora, N. Hynson, B. Dentinger, M. Ryberg, A.E. Arnold, P. Bridge, W.W.H. Ho, K. Hughes, E.G. Simmons, R. E. Baird, T.A.J. Volk, B.A. Perry, M. Wach, R.W. Kerrigan, F. Okafor, D.L. Lindner, J. W. Taylor, J. Campbell, J. Rajesh, D.R. Reynolds, D. Geiser, R.A. Humber, N. Hausmann, T. Szaro, J. Stajich, F.W. Spiegel, H.S. Vishniac, D.J. Lodge, A. Gathman, K.G. Peay, J. Stenlid, T. Henkel, C.H. Robinson, P.J. Pukkila, N. H. Nguyen, C. Villata, D. Dewsbury, P. Kennedy, S. Bergemann, M. Stadler, D. S. Yohalem, M.C. Aime, F. Kauff, A. Porras-Alfaro, R.M. Miller, C. Guéidan, A. Beck, J. Carroll, B. Andersen, S. Marek, J.A. Crouch, G. Turgeon, J. Kerrigan, M.E. Smith, J.B. Ristaino, K.T. Hodge, G. Kuldau, G.J. Samuels, T.M. Porter, S. E. Baker, H.J. Raja, H. Voglmayr, M. Gardes, R.W. Lichtwardt, D.P. Janos, J. D. Rogers, A.E. Glenn, P. Cannon, S.W. Woolfolk, R.P. Korf, H.C. Kistler, M. A. Castellano, S.L. Maldonado-Ramírez, H.E. Hallen, P.M. Kirk, E.L. Stewart, J. J. Farrar, T. Osmundson, R.S. Currah, M. Spiering, S. Branco, V. Vujanovic, Preserving accuracy in GenBank, *Science* 319 (2008) 1616.
- [51] S. Gontran, J. Kurt, B. Yves, B. Luc, D. Erena, B. Thierry, M. Marc de, D. Stijn, Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France, *ZooKeys* 365 (2013) 307–328.
- [52] F. Crocetta, P. Mariottini, D. Salvi, M. Oliverio, Does GenBank provide a reliable DNA barcode reference to identify small alien oysters invading the Mediterranean Sea? *J. Mar. Biol. Assoc. U.K.* 95 (2015) 111–122.
- [53] Y.G. Seah, A.F. Ariffin, T.N.A.M. Jaafar, Levels of COI divergence in family Leiognathidae using sequences available in GenBank and BOLD systems: a review on the accuracy of public databases, *Aquac. Conserv. Legis.* 10 (2017) 391–401.
- [54] K.D. Pruitt, T. Tatusova, G.R. Brown, D.R. Maglott, NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy, *Nucleic Acids Res.* 40 (2011) D130–D135.
- [55] S. Ratnasingham, P.D. Hebert, BOLD: the barcode of life data system, *Mol. Ecol. Notes* 7 (2007) 355–364.
- [56] A. Linacre, S.S. Tobe, An overview to the investigative approach to species testing in wildlife forensic science, *Investig. Genet.* 2 (2) (2011).
- [57] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* 16 (1980) 111–120.
- [58] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. deWaard, Biological identifications through DNA barcodes, *Proc. Biol. Sci.* 270 (2003) 313–321.
- [59] P.D.N. Hebert, M.Y. Stoeckle, T.S. Zemlak, C.M. Francis, Identification of birds through DNA barcodes, *PLoS Biol.* 2 (2004) 1657–1663.
- [60] R.A. Collins, R.H. Cruickshank, Known knowns, known unknowns, unknown unknowns and unknown knowns in DNA barcoding: a comment on Downton et al, *Syst. Biol.* 63 (2014) 1005–1009.
- [61] K.W. Will, B.D. Mishler, Q.D. Wheeler, The perils of DNA barcoding and the need for integrative taxonomy, *Syst. Biol.* 54 (2005) 844–851.
- [62] K.A. Meiklejohn, J.F. Wallman, S.L. Cameron, M. Downton, Comprehensive evaluation of DNA barcoding for the molecular species identification of forensically important Australian Sarcophagidae (Diptera), *Invertebr. Syst.* 26 (2012) 515–525.
- [63] J.H. Lowenstein, G. Amato, S.O. Kolokotronis, The real *maccoyii*: identifying tuna sushi with DNA barcodes—contrasting characteristic attributes and genetic distances, *PLoS One* 4 (2009) e7866.
- [64] R. Meier, G. Zhang, F. Ali, The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification, *Syst. Biol.* 57 (2008) 809–813.

- [65] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T. J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (2007) 2947–2948.
- [66] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Res.* 15 (2002) 3059–3066.
- [67] L.M. Boykin, K.F. Armstrong, L. Kubatko, P. De Barro, Species delimitation and global biosecurity, *Evol. Bioinform.* 8 (2012) 1–37.
- [68] M. Virgilio, K. Jordaens, F.C. Breman, T. Backeljau, M. De Meyer, Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case, *PLoS One* 7 (2012), e31581.
- [69] T. Ekrem, E. Willassen, E. Stur, A comprehensive DNA sequence library is essential for identification with DNA barcodes, *Mol. Phylogenet. Evol.* 43 (2007) 530–542.
- [70] J.J. Wiens, J. Tiu, Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling, *PLoS One* 7 (2012), e42925.
- [71] A. Rokas, E. Ladoukakis, E. Zouros, Animal mitochondrial DNA recombination revisited, *Trends Ecol. Evol.* 18 (2003) 411–417.
- [72] A. Linacre, Application of mitochondrial DNA technologies in wildlife investigations - species identification, *Forensic Sci. Rev.* 18 (2006) 1–8.
- [73] K.N. Magnacca, M.J.F. Brown, Mitochondrial heteroplasmy and DNA barcoding in Hawaiian Hylaeus (*Nesoprosopis*) bees (Hymenoptera: Colletidae), *BMC Evol. Biol.* 10 (2010) 174–189.
- [74] W.R. Hoeh, K.H. Blakley, W.M. Brown, Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA, *Science* 251 (1991) 1488–1490.
- [75] L.D. Densmore, J.W. Wright, W.M. Brown, Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*), *Genetics* 110 (1985) 689–707.
- [76] E. Bermingham, T. Lamb, J.C. Avise, Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates, *J. Hered.* 77 (1986) 249–252.
- [77] D.L. Hartl, A.G. Clark, *Principles of Population Genetics*, 2nd ed., Sinauer Associates, Sunderland, MA, 1989.
- [78] N.L. Vollmer, A. Viricel, L. Wilcox, M.K. Moore, P.E. Rosel, The occurrence of mtDNA heteroplasmy in multiple cetacean species, *Curr. Genet.* 57 (2011) 115–131.
- [79] H. Song, J.E. Buhay, M.F. Whiting, K.A. Crandall, Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 13486–13491.
- [80] S.P. Otto, J. Whitton, Polyploid incidence and evolution, *Annu. Rev. Genet.* 34 (2000) 401–437.
- [81] R. Donné, M.B. Nader, C. Desdouets, Cellular and molecular mechanisms controlling ploidy. Reference Module in Life Sciences, Elsevier, 2017.
- [82] L.A. Meyers, D.A. Levin, On the abundance of polyploids in flowering plants, *Evolution* 60 (2006) 1198–1206.
- [83] W. Albertin, P. Marullo, Polyploidy in fungi: evolution after whole-genome duplication, *Proc. Biol. Sci.* 279 (2012) 2497–2509.
- [84] K. St Onge, J.P. Foxe, J. Li, H. Li, K. Holm, P. Corcoran, S. Tanja, M. Lascoux, W. Stephen, Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd's Purse (*Capsella bursa-pastoris*), *Mol. Biol. Evol.* 29 (2012) 1721–1733.
- [85] M. Kyriakidou, H.H. Tai, N.L. Anglin, D. Ellis, M.V. Strömvik, Current strategies of polyploid plant genome sequence assembly, *Front. Plant Sci.* 9 (2018).
- [86] F.-W. Li, L.-Y. Kuo, C.J. Rothfels, A. Ebiyara, W.-L. Chiou, M.D. Windham, K. M. Pryer, rbcL and matK earn two thumbs up as the core DNA barcode for ferns, *PLoS One* 6 (10) (2011) e26597.
- [87] M.R. Judith, S. Daniel, Extinction by hybridization and introgression, *Annu. Rev. Ecol. Syst.* 27 (1996) 83–109.
- [88] F.W. Allendorf, R.F. Leary, P. Spruell, J.K. Wenburg, The problems with hybrids: setting conservation guidelines, *Trends Ecol. Evol.* 16 (2001) 613–622.
- [89] E. Anderson, G.L. Stebbins, Hybridization as an evolutionary stimulus, *Evolution* 8 (1954) 378–388.
- [90] G.L. Stebbins, The role of hybridization in evolution, *Proc. Am. Philos. Soc.* 103 (1959) 231–251.
- [91] S.M. Chambers, S.R. Fain, B. Fazio, M. Amaral, An account of the taxonomy of North American wolves from morphological and genetic analyses, *North Am. Fauna* (2012) 1–67.
- [92] A. Amorim, F. Pereira, C. Alves, O. García, Species assignment in forensics and the challenge of hybrids, *Forensic Sci. Int. Genet.* 48 (2020) 102333.
- [93] P.R. Grant, B.R. Grant, Hybridization of bird species, *Science* 256 (1992) 193–197.
- [94] J. Mallet, Hybridization as an invasion of the genome, *Trends Ecol. Evol.* 20 (2005) 229–237.
- [95] J. Mallet, Hybrid speciation, *Nature* 446 (2007) 279–283.
- [96] M.E. Heikkinen, M. Ruokonen, M. Alexander, J. Aspi, T. Pyhäjärvi, J.B. Searle, Relationship between wild greylag and European domestic geese based on mitochondrial DNA, *Anim. Genet.* 46 (2015) 485–497.
- [97] M.E. Heikkinen, M. Ruokonen, T.A. White, M.M. Alexander, I. Gündüz, K. M. Dohney, J. Aspi, J.B. Searle, T. Pyhäjärvi, Long-term reciprocal gene flow in wild and domestic geese reveals complex domestication history, *G3 Genes Genomes Genet.* 10 (2020) 3061–3070.
- [98] B.A. Gonzalez, A.M. Agapito, F. Novoa-Munoz, J. Vianna, W.E. Johnson, J. C. Marin, Utility of genetic variation in coat color genes to distinguish wild, domestic and hybrid South American camelids for forensic and judicial applications, *Forensic Sci. Int. Genet.* 45 (2020) 102226.
- [99] C.S. Quilodran, J.I. Montoya-Burgos, M. Currat, Harmonizing hybridization dissonance in conservation, *Commun. Biol.* 3 (2020) 391.
- [100] S.R. Fain, D.J. Straughan, B.F. Taylor, Genetic outcomes of wolf recovery in the western Great Lakes states, *Conserv. Genet.* 11 (2010) 1747–1765.
- [101] A. Trouwborst, Exploring the legal status of wolf-dog hybrids and other dubious animals: international and EU law and the wildlife conservation problem of hybridization with domestic and alien species, *Rev. Eur. Comp. Int. Environ. Law* 23 (2014) 111–124.
- [102] S.R. Fain, D.J. Straughan, B.C. Hamlin, R.M. Hoesch, J.P. LeMay, Forensic genetic identification of sturgeon caviars traveling in world trade, *Conserv. Genet.* 14 (2013) 855.
- [103] A. Ludwig, L. Lippold, L. Debus, R. Reinartz, First evidence of hybridization between endangered sterlets (*Acipenser ruthenus*) and exotic Siberian sturgeons (*Acipenser baerii*) in the Danube River, *Biol. Invasions* 11 (2008) 753–760.
- [104] S.M. Carr, G.A. Hughes, Direction of introgressive hybridization between species of North American deer (*Odocoileus*) as inferred from mitochondrial-cytochrome-b sequences, *J. Mammal.* 74 (1993) 331–342.
- [105] J.C. Cathey, J.W. Bickham, J.C. Patton, Introgressive hybridization and onconcordance evolutionary history of maternal and paternal lineages in North American deer, *Evolution* 52 (1998) 1224–1229.
- [106] G.E. Hornbeck, J.M. Mahoney, Introgressive hybridization of mule deer and white-tailed deer in Southwestern Alberta, *Wildl. Soc. Bull.* (1973–2006) 28 (2000) 1012–1015.
- [107] T. Russell, C. Cullingham, A. Kommadath, P. Stothard, A. Herbst, D. Coltman, Development of a novel Mule Deer genomic assembly and species-diagnostic SNP panel for assessing introgression in Mule Deer, White-Tailed Deer, and their interspecific hybrids, *G3 Genes Genomes Genet.* 9 (2019) 911–919.
- [108] M.A. Cronin, In my experience: mitochondrial DNA in wildlife taxonomy and conservation biology: cautionary notes, *Wildl. Soc. Bull.* 21 (1993) 339–348.
- [109] S.M. Carr, S.W. Ballinger, J.N. Derr, J.W. Bickham, Mitochondrial DNA analysis of hybridization between sympatric white-tailed deer and mule deer in West Texas, *Proc. Natl. Acad. Sci. U. S. A.* 83 (1986) 9576–9580.
- [110] M.A. Cronin, E.R. Vyse, D.G. Cameron, Genetic relationships between mule deer and white-tailed deer in Montana, *J. Wildl. Manage.* 52 (1988) 320–328.
- [111] D.X. Zhang, G.M. Hewitt, Nuclear integrations: Challenges for mitochondrial DNA markers, *Trends Ecol. Evol.* 11 (1996) 247–251.
- [112] J.L. Blanchard, M. Lynch, Organellar genes: why do they end up in the nucleus? *Trends Genet.* 16 (2000) 315–320.
- [113] G.D.D. Hurst, F.M. Jiggins, Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic Studies: the effects of inherited symbionts, *Proc. Biol. Sci.* 272 (2005) 1525–1534.
- [114] J.E. Buhay, “COI-like” sequences are becoming problematic in molecular systematic and DNA barcoding studies, *J. Crustacean Biol.* 29 (2009) 96–110.
- [115] T.R. Gregory, DNA barcoding does not compete with taxonomy, *Nature* 434 (2005) 1067.
- [116] B.C. Carstens, L.L. Knowles, Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers, *Syst. Biol.* 56 (2007) 400–411.
- [117] L. Choleva, Z. Musilova, A. Kohoutova-Sediva, J. Paces, P. Rab, K. Janko, Distinguishing between incomplete lineage sorting and genomic introgressions: complete fixation of allospecific mitochondrial DNA in a sexually reproducing fish (Cobitis; Teleostei), despite clonal reproduction of hybrids, *PLoS One* 9 (2014), e80641.
- [118] C. Anindo, A.D. Terry, The historical biogeography of sturgeons (Osteichthyes: Acipenseridae): a synthesis of phylogenetics, palaeontology and palaeogeography, *J. Biogeogr.* 25 (1998) 623–640.
- [119] J.A. McGuire, C.W. Linkem, M.S. Koo, D.W. Hutchison, A.K. Lappin, D.I. Orange, J. Lemos-Espinal, B.R. Riddle, J.R. Jaeger, Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards, *Evolution* 61 (2007) 2879–2897.
- [120] S. Shokralla, J.L. Spall, J.F. Gibson, M. Hajibabaei, Next-generation sequencing technologies for environmental DNA research, *Mol. Ecol.* 21 (2012) 1794–1805.
- [121] M.D. Brandhagen, R.S. Just, J.A. Irwin, Validation of NGS for mitochondrial DNA casework at the FBI laboratory, *Forensic Sci. Int. Genet.* 44 (2020) 102151.
- [122] B. McCord, S.B. Lee, A. Alonso, P.A. Barrio, P. Muller, S. Kocher, B. Berger, P. Martin, M. Bodner, S. Willuweit, W. Parson, L. Roewer, B. Budowle, Current state-of-art of STR sequencing in forensic genetics, *Electrophoresis* 39 (2018) 2655.
- [123] L.E. Cartier, M.S. Krzemnicki, M. Gysi, B. Lendvay, N.V. Morf, A case study of ivory species identification using a combination of morphological, gemmological and genetic methods, *J. Gemmol.* 37 (2020) 282–297.
- [124] T.M. Lan, Y. Lin, J. Njaramba-Ngatia, X.S. Guo, R.G. Li, H.M. Li, S. Kumar-Sahu, X. Wang, X.J. Yang, H.B. Guo, W.H. Xu, K. Kristiansen, H. Liu, Y.C. Xu, Improving species identification of ancient mammals based on next-generation sequencing data, *Genes* 10 (7) (2019) 509.
- [125] Y. Lammers, T. Peelen, R.A. Vos, B. Gravendeel, The HTS barcode checker pipeline, a tool for automated detection of illegally traded species from high-throughput sequencing data, *BMC Bioinform.* 15 (2014).
- [126] K.N. Finch, F.A. Jones, R.C. Cronn, Genomic resources for the neotropical tree genus *Cedrela* (Meliaceae) and its relatives, *BMC Genomics* 20 (2019).
- [127] D.C. Carvalho, R.M. Palhares, M.G. Drummond, M. Gadanho, Food metagenomics: next generation sequencing identifies species mixtures and mislabeling within highly processed cod products, *Food Control* 80 (2017) 183–186.

- [128] A. Giusti, A. Armani, C.G. Sotelo, Advances in the analysis of complex food matrices: species identification in surimi-based products using Next Generation Sequencing technologies, *PLoS One* 12 (2017) 1–18.
- [129] G. Cottenet, C. Blancpain, P.F. Chuah, C. Cavin, Evaluation and application of a next generation sequencing approach for meat species identification, *Food Control* 110 (2020) 107003.
- [130] M.L. Coghlan, J. Haile, J. Houston, D.C. Murray, N.E. White, P. Moolhuijzen, M. I. Bellgard, M. Bunce, Deep sequencing of plant and animal DNA contained within traditional chinese medicines reveals legality issues and health safety concerns, *PLoS Genet.* 8 (2012) 1–11.
- [131] A.O. Tillmar, B. Dell'Amico, J. Welander, G. Holmlund, A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures, *PLoS One* 8 (2013), e83761.
- [132] A.J. Arulandhu, M. Staats, R. Hagelaar, M.M. Voorhuijzen, T.W. Prins, I. Scholtens, A. Costessi, D. Duijsings, F. Rechenmann, F.B. Gaspar, M.T. Barreto Crespo, A. Holst-Jensen, M. Birck, M. Burns, E. Haynes, R. Hohegger, A. Klingl, L. Lundberg, C. Natale, H. Niekamp, E. Perri, A. Barbante, J.P. Rosec, R. Seyfarth, T. Sovova, C. van Moorlegem, S. van Ruth, T. Peelen, E. Kok, Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples, *GigaScience* 6 (2017) 1–18.
- [133] J. Shaili, S. Jitesh, C. Vito Adrian, R.F. Sam, A.E. Robert, M. Isabel, V. Asit, A. D. Elizabeth, 'Genome skimming' with the MinION hand-held sequencer identifies CITES-listed shark species in India's exports market, *Sci. Rep.* 9 (2019) 1–13.
- [134] E.E. Dormontt, D.I. Jardine, K.J. van Dijk, B.F. Dunker, R.R.M. Dixon, V. D. Hipkins, S. Tobe, A. Linacre, A.J. Lowe, Forensic validation of a SNP and INDEL panel for individualisation of timber from bigleaf maple (*Acer macrophyllum* Pursch), *Forensic Sci. Int. Genet.* 46 (2020) 102252.
- [135] Y.T. Lo, P.C. Shaw, Application of next-generation sequencing for the identification of herbal products, *Biotechnol. Adv.* 37 (2019) 107450.
- [136] E.K. Timpano, M.K.R. Scheible, K.A. Meiklejohn, Optimization of the second internal transcribed spacer (ITS2) for characterizing land plants from soil, *PLoS One* 15 (2020) 1–12.
- [137] M. Leray, S.L. Ho, L.J. Lin, R.J. Machida, MIDORI server: a webserver for taxonomic assignment of unknown metazoan mitochondrial-encoded sequences using a curated database, *Bioinformatics* 34 (2018) 3753–3754.
- [138] Y. Sato, M. Miya, T. Fukunaga, T. Sado, W. Iwasaki, MitoFish and MiFish Pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding, *Mol. Biol. Evol.* 35 (2018) 1553–1555.
- [139] E.E. Curd, Z. Gold, G.S. Kandlikar, J. Gomer, M. Ogden, T. O'Connell, L. Pipes, T. M. Schweizer, L. Rabichow, M. Lin, B. Shi, P.H. Barber, N. Kraft, R. Wayne, R. S. Meyer, D. Yu, Anacapa Toolkit: an environmental DNA toolkit for processing multilocus metabarcode datasets, *Methods Ecol. Evol.* 10 (2019) 1469.
- [140] M.K.R. Scheible, D.J. Straughan, M.K. Burnham-Curtis, K.A. Meiklejohn, Using hybridization capture to obtain mitochondrial genomes from forensically relevant canids: assessing sequence variation for species identification, *Forensic Sci. Int.: Anim. Environ.* 1 (2021) 100018.