Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 34 (1): 54–61, 2023
OPEN ACCESS
Available online at:
http://www.italian-journal-of-mammalogy.it
doi:10.4404/hystrix-00628-2023

Research Article

Shortcomings of DNA barcodes: a perspective from the mammal fauna of Switzerland

Manuel Ruedi[1,*], Julie Manzinalli[1], Adrian Dietrich[2], Lorenzo Vinciguerra[2,3]

[1]*Muséum d'histoire naturelle de la Ville de Genève, BP 6434, 1211 Geneva 6, Switzerland*
[2]*UNA AG, Schwarzenburgstrasse 11, 3007 Bern Switzerland*
[3]*Naturmuseum Rorschach Strasse 263, 9016 St. Gallen, Switzerland*

Keywords:
DNA barcoding
COI
Cytochrome b
16S rRNA
mitochondrial
vouchers

Article history:
Received: 5 April 2023
Accepted: 6 July 2023

Acknowledgements
We would like to dedicate this study to late Prof. Peter Vogel, who contributed immensely to the knowledge of the Swiss mammal fauna. Sampling was significantly enlarged thanks to the help of Juerg Paul Mueller (BNM), Nicola Zambelli (MSNL). Thanks are due to Lucie Cauwet, Janik Pralong and Nagwa Othman (MHNG) for technical assistance. Thomas Briner (Naturmuseum Solothurn), Jacques Gilliéron (Geneva), Faune Concept and Maddalena and associate who provided several specimens or DNA sequences. We also thank two anonymous reviewers for suggestions to improve the quality of the initial manuscript.

Abstract

The use of DNA barcodes is a popular approach to identify unknown environmental samples based on reference DNA databases. However, as shown previously, the success of reaching accurate species-level identification will not only depend on the taxonomic group under interest, but also on the availability of reliable and comprehensive reference databases. We applied this methodology for a sampling of vouchered specimens of wild mammals of Switzerland and demonstrated that reliable species identification with standard DNA barcodes in this reputedly well-known taxon may be challenging. The overall success of unambiguous species-level identifications with help of three commonly used mitochondrial markers barely reached 70 % in unsupervised queries submitted to BOLD or GenBank. Pitfalls were mostly due to misidentified or mislabelled sequences available in public databases, to the presence of highly divergent cryptic lineages or to missing reference sequences for the rarest species (i.e., those that are in greatest need for conservation attention). Divergent cryptic lineages (over 5 % sequence divergence) found in Swiss mammals were either due to highly distinct intraspecific haplotypes or to the existence of an overlooked cryptic biological species (*Muscardinus speciosus*). To circumvent part of the observed pitfalls, we provide a curated and complete reference database for the wild mammals of Switzerland based on carefully identified, vouchered specimens. We finally acknowledge that the identification of naturally hybridizing or introgressed species remains a significant challenge that is often overlooked in massive, unsupervised DNA barcode analyses.

Introduction

The use of short DNA sequences to identify unknown biological material against reference databases is a very popular approach that has been expanding quickly in the literature with the ease of obtaining high-throughput sequencing data from various sources of DNA (Kocher et al., 2017a). Such a relatively straightforward method based on DNA barcoding has become a very useful tool in many fields of research, including in forensics (Galimberti et al., 2013; Gouda et al., 2020), in environmental DNA metabarcoding (Zeale et al., 2011; Yao et al., 2022) or to document trophic interactions among species (Razgour et al., 2011; Andriollo et al., 2021).

Historically, several markers of mitochondrial origin, including fragments of the cytochrome *b* (CytB), the NADH dehydrogenase subunit 1 (ND1) or ribosomal RNA (12S or 16S) genes, to name a few, have been used as comparative sequences for help in identifying animal taxa (Zeale et al., 2011; Giguet-Covex et al., 2014; Kocher et al., 2017b). In general, these sequences are publicly available from global repositories such as GenBank (Benson et al., 2018) from where unknown homologues can be queried with simple BLAST searches (Boratyn et al., 2013) to inform taxonomic identification. Since 2003, Paul Hebert and colleagues standardized this animal DNA barcoding approach (Hebert et al., 2003b,a) by restricting the marker of choice to a small fragment (ca. 650 bp in length) of the cytochrome *c* oxidase I gene (COI) and by developing a new dedicated repository, the Barcode Of Life Data System (BOLD, Ratnasingham and Hebert, 2007). In addition of compiling and storing well-curated COI sequences, this online interface also aimed to associate those records with available specimen information such as geographic origin or pictures of vouchered material. This platform further implemented a novel taxon identification engine (ID engine) facilitating species identification with refined, automated queries. These queries are based on a RESL algorithm which essentially groups similar sequences into molecular operational taxonomic units assigned to unique Barcode Index Numbers (BINs). This iterative process of allocating similar lineages into unique BINs thus does not require prior species-level identification (Ratnasingham and Hebert, 2013), and can be automated via scripts in order to identify thousands of unknown sequences.

However, the success of reaching positive and accurate species-level identification strongly depends on the animal group under focus (Kvist, 2013) and on the quality of the reference database used to compare unknown biological material (Kocher et al., 2017b; Meiklejohn et al., 2019). Indeed, in order to reach unambiguous species-level identification, the group of interest must show a clear genetic discontinuity between intraspecific and interspecific differentiation of DNA barcodes, i.e. a so called barcoding gap. The existence of such genetic discontinuities varies among animal taxa, from extensive in some lepidopterans or mammals (Borisenko et al., 2008), to fuzzy in annelids or other invertebrates (Kvist, 2013, 2016; Ivanov et al., 2018). Previous studies have shown that other critical factors for successful DNA barcode species-level identifications include the reliability (Meiklejohn et al., 2019; Pentinsaari et al., 2020) and completeness of reference databases (Kvist, 2013; Kocher et al., 2017b; Yao et al., 2022). Speciose groups such a dipterans or hymenopterans may lack adequate taxonomic coverages, or the labelled sequences available in public repositories for comparisons might represent erroneous species-level iden-

*Corresponding author
Email address: manuel.ruedi@ville-ge.ch (Manuel Ruedi)

tification, compromising the entire DNA barcode approach (Meier et al., 2006; Meiklejohn et al., 2019). Limitations due to the use of a single mitochondrial marker to recognize biological species (Ballard and Whitlock, 2004) are also critical when hybridization, introgression or incomplete lineage sorting are common in taxa that have diverged recently (Ermakov et al., 2015; Mutanen et al., 2016).

During cursorial examination of results from queried DNA sequences against available DNA barcodes of small European mammals, we realized that many species in this supposedly well-known group were represented by multiple problematic reference sequences, or conversely lacked any close genetic match. We therefore decided to build a new reference database of DNA barcodes covering all 99 species of wild mammals living in Switzerland (Graf and Fischer, 2021), and examine the outcome of species-level identifications obtained with available sequences deposited in the two main public repositories of DNA sequences, BOLD and GenBank. The main aims of this report were thus (1) to build a new complete database of reference sequences essentially based on carefully identified, vouchered specimens, (2) to examine the reliability of species-level identification obtained with these sequences and unsupervised barcode approaches, (3) to propose recommendations to minimise pitfalls to achieve better species-level identifications.

Material and Methods

Ethics statement

This work was exclusively based on existing tissue or DNA extracts available in different museum collections and thus required no ethical approval.

Sampling design

One key initial step to achieve reliable species-level identification throughout the DNA barcoding approach is to rely on sequences linked to voucher specimens properly identified by taxonomic experts. Hence, most of the samples used in this survey were issued from the frozen tissue collection associated to specimens housed in the collections of the Natural History Museum of Geneva (MHNG, n=201), the Naturmuseum of St. Gallen (NMSG, n=22), the Bündner Naturmuseum of Chur (BNM, n=4), the Museo cantonale di Storia naturale of Lugano (MCSN, n = 2), the Musée cantonal des Sciences naturelles of Lausanne (IZEA, n =1) and the Estacion Biologica de Doñana of Sevilla (EBD, n=1). Each sampled mammal was critically evaluated with respect to diagnostic morphological characters (Dietz and von Helversen, 2004; Marchesi et al., 2008; Dietz and Kiefer, 2015), collection locality and (for a small subset of specimens) previous DNA analyses of the same individual. Because larger mammal species of the Swiss fauna have already been subjected to broad phylogeographic surveys with multiple molecular markers (e.g. Davison et al., 2001; Rodriguez et al., 2010; Zachos et al., 2010; Nussberger et al., 2013), we focussed our taxonomic sampling on the lesser known and morphologically more challenging small mammals from the orders Eulipotyphla, Chiroptera and Rodentia, which represent over 70 % of the local mammalian diversity. Information about the origin and museum catalogue number of all specimens examined can be found in the Table S1.

We tried to capture potential intraspecific genetic variation by sampling multiple individuals from distinct biogeographic compartments within Switzerland, e.g., north and south of the Alpine range, as illustrated on the map (Fig. 1). For the larger mammals and for few other species that are present in neighbouring countries but not yet formally recorded in Switzerland, i.e. *Erinaceus roumanicus* (Bolfíková and Hulva, 2012), *Sicista betulina* (Andersen et al., 2022) and *Rhinolophus euryale* (Ruedi, 2021), we relied preferably on sequences associated with museum vouchers or on referenced specimens that were sequenced for their entire mitogenomes. These mitogenomes were issued mainly from large-scale initiatives to document molecular variation of local vertebrates, e.g., in Denmark (Margaryan et al., 2021) or in France (Hassanin et al., 2009). Each of these mitogenomes were carefully checked for taxonomic consistency, as they were not freed of potential labelling errors. For instance, we discarded the mitogenome MT584130 deposited in the GenBank and associated to the label "*Eptesicus serotinus* DM334", which in fact proved to be identical to the mitogenome MN122907 of a common noctule (*Nyctalus noctula*) sequenced by the same research team (Margaryan et al., 2021). Accession numbers corresponding to the 332 downloaded sequences used for comparisons are available in the Table S1.

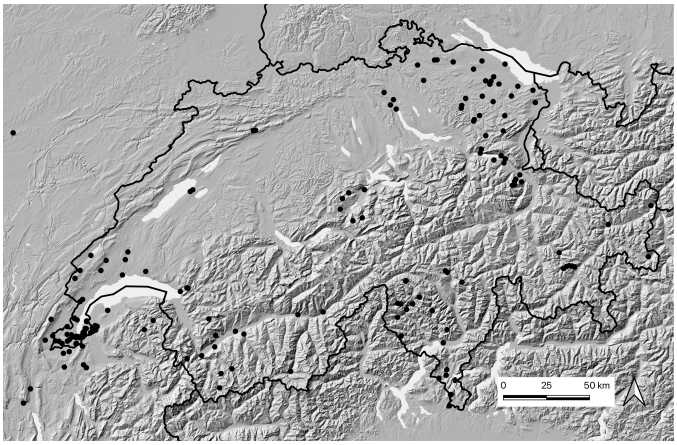

Figure 1 – Geographic origin (black dots) of the tissue samples newly sequenced in this genetic survey of wild mammals of Switzerland. Precise location, taxonomy and voucher reference numbers from which sequences were obtained are detailed in the Table SI.

DNA extraction and sequencing

A small fragment (20-50 mg) of frozen or ethanol preserved tissue sample was used to extract whole DNA. In total DNA was extracted from 193 samples representing 78 species (Tab. S1), using the DNeasy® Blood and Tissue Kit (Qiagen, Germany), following the manufacturer's instructions. The extracted DNA was eluted in 200 µL of TE buffer and stored at −20 °C. DNA extracts were used for PCR amplification to obtain sequences of three mitochondrial markers. Two pairs of primers were used to amplify the standard fragment for COI barcode gene: one preferred primer pair was UTyr (5'-ACCYCTGTCYTTAGATTTACAGTC- 3') and C1L705 (5'-ACTTCDGGGTGNCCRAARAATCA-3') that was designed by Hassanin et al. (2012) to amplify a very large array of mammalian species. An important mismatch at the 3' end of the forward priming site, however, was found in all assayed *Rhinolophus* species and prevented the initiation of amplification reaction with these primers. To overcome this problem, we used for this taxon the more classical barcode primer pair VR1d (5' -TAGACTTCTGGGTGGCCRAARAAYCA- 3') and VF1d (5' -TCTCAACCAACCACAAAGACATTGG- 3') (Ivanova et al., 2007) instead. Because some small European mammals were extensively or exclusively characterized so far with other mitochondrial markers (e.g. Yannic et al., 2008; Beysard et al., 2012), we also sequenced for comparisons the CytB or part of the 16SrRNA genes in a subset of samples with the following primer pairs: Molcit-F (5' -AATGACATGAAAAATCACCGTTGT- 3') and CYTB-H (5' -CTTTTCTGGTTTACAAGACCAG- 3') (Ibáñez et al., 2006) were used to amplify the entire CytB gene (about 1200 bp), whereas part of the 16S (ca. 560 bp) was amplified with the primer pair 16SAr-L (5' -CGCCTGTTTATCAAAAACAT- 3') and 16SBr-H (5' -CCGGTCTGAACTCAGATCACGT- 3') (Palumbi et al., 1991).

All amplifications were achieved in a 25 µL reaction volume with a commercial kit (Qiagen, Switzerland). PCR cocktail included 2.5 µL 10x reaction buffer, 1 μM $MgCl_2$, 0.8 µmM dNTP, 0.2 µM of each primer, 5 µL QSolution, 1 U Taq polymerase and 4 µL of extracted DNA. To minimize the amplification of unspecific DNA fragments, typical thermal cycling profiles included a touch-down strategy whereby the initial annealing temperature was set 5 °C above the one recommended for a specific primer pair for 10 cycles (with a decreasing temperature of 0.5 °C at each cycle), before continuing with 26 cycles at

a lower annealing temperature (i.e. 55°C for the COI or 16S, and 45 °C for the CytB). Amplicons were purified enzymatically with an ExoSAP kit (Thermo Fisher, Switzerland) and sent for sequencing (using the same primers) to a commercial company (Macrogen Europe Inc., Holland).

Sequence alignments and haplotype analyses

Chromatograms of both strands were visualised, assembled, and edited using Sequencher v.4.1 (Gene Codes Corp., USA). Sequences were examined individually for the presence of double peaks or stop codons in the coding genes to check for possible non-targeted amplicons of nuclear origin (or numts; see Dubey et al., 2009). Multiple identical sequences from the same species were then collapsed into unique haplotypes. The package MEGA v.7 (Kumar et al., 2016) was used to align sequences with the Clustal W algorithm (Thompson et al., 2003). For each gene alignment, the same package was used to calculate pairwise genetic distances with the commonly used Kimura-2-parameters model (K2P) and to generate neighbour-joining trees (NJ). As our intent was not to interpret phylogenetic relationships, this simple tree-building method was ideal to quickly visualise the clustering pattern of newly generated haplotypes across the different species.

Each distinct COI haplotype was submitted to a standard, unsupervised DNA barcoding analysis using the ID engine provided in BOLD v.4 and queried against available species-level barcodes. We investigated these suggested species-level identifications using either the default reference database, i.e. the "Species Level Barcode Records" (>4.8 million COI sequences available, representing >244 K nominal species), or the more stringent "Public Record Barcode Database" (>2.6 COI sequences available, representing >156 K nominal species). We also queried these unique COI haplotypes and those of the CytB and 16S with BLAST searches (Boratyn et al., 2013) using the default parameters, against all available sequences in GenBank (as of January 27, 2023).

Following similar rationales suggested by Meiklejohn et al. (2019), we considered four different outcomes for results of these queries of species-level assignations, only the first one being regarded as accurate, the remaining ones being problematic: (i) When the top hits suggested a single species corresponding to the morphology-based identification of the voucher, we considered the query as "**correct ID**". (ii) When the top hit suggested more than one species name with highly similar barcodes (100-98% similarity), including the morphology-based one, we considered the query as "**ambiguous ID**". (iii) When the top hits suggested a single species which did not correspond to the a priori ID, we considered the query as "**erroneous ID**". (iv) When no matching sequence was found in the database within 98% sequence similarity, the query was considered as "**failed ID**".

Queries of COI haplotypes could be easily categorized in this scheme based on the downloadable file returned by the ID engine in BOLD, except for the second category (ambiguous ID). In this case, it was necessary to examine one by one the suggested top 20 matches available in the detailed output, to detect the potential presence of different species with highly similar barcodes (within 2 % sequence divergence). Similarly, species-level information from BLAST searches issued from GenBank for all three markers were extracted from the output text files. Taxonomic discrepancies purely due to nomenclatural issues such as accepted synonyms, or the use of different generic names for the same species, were not considered as errors. For instance, the European water vole *Arvicola amphibius* appeared in the GenBank under various synonymous taxa (*Arvicola terrestris* or *A. sherman*), while the European edible dormouse *Glis glis* was also stored as *Myoxus glis*. These taxonomic variations were not considered as errors. However, accepted taxonomic splits such as *Microtus lavernedii* (split from *M. agrestis*; see Kryštufek, 2018) or *Myotis crypticus* (split from *M. nattereri*; see Juste et al., 2018) were considered as errors if the wrong species name was recovered in the queries.

Results

Patterns of mtDNA divergence

Most sequenced amplicons produced clean and easily assembled sequences in all three assayed genes. Exceptions included partially degraded tissue samples which could only be amplified for the shorter marker (ca. 550 bp of the 16S), while longer target sequences (COI and CytB) could only be recovered in individuals associated with better preserved tissue material. For the CytB gene, and only for this marker, we found, however, major difficulties to obtain clean sequences in most assayed specimens of wood mice (genus *Apodemus*), including those associated with fresh tissue samples. Indeed, chromatograms often showed multiple double peaks at specific nucleotide positions, in both the forward and reverse sequences. These chromatograms were thus interpreted as resulting from the co-amplification of the targeted CytB gene and a nuclear paralog (numt). Several stop codons were inferred in the consensus sequences resulting from these amplicons, confirming that a non-coding pseudogene co-amplified with the genuine CytB, as already evidenced by Dubey et al. (2009) in *A. sylvaticus.* We found that this pseudogene amplified (or co-amplified and thus produced chimeric sequences) in all three assayed species of *Apodemus*. A BLAST search of this known CytB pseudogene (e.g., GenBank # AF159395) against the full genome of a wood mouse deposited by the Wellcome Sanger Institute (sequence number OX359319) indeed confirmed that it matched at 98 % similarity with several nuclear fragments of the chromosome 22. The COI and 16S sequences issued from the same individuals showed no evidence of pseudogene in that nuclear genome. The CytB nuclear pseudogene diverged by 8-12 % K2P distance from genuine mitochondrial haplotypes found in *A. sylvaticus*, *A. flavicollis* or *A. alpicola* (Table 1). A full understanding of the evolution of this CytB pseudogene is beyond the scope of the present survey and will need further scrutiny. Interestingly, a known pseudogene of the COI in wild cats (e.g., GenBank # KF297791) did not amplify with the primer pair used here (Hassanin et al., 2012), but would differ by at least 7.5 % compared to the genuine, mitochondrial version of the COI barcode (Table 1).

Table 1 – Unusually low (<5%) interspecific K2P distances observed among distinct species of wild mammals found in Switzerland. We indicate the corresponding divergences observed at the 16S for illustrative purpose only, as this fragment is known to be much more conserved compared to the other two mitochondrial markers. In the bottom of the table, we mentioned K2P divergence values of a pseudogene found in *Apodemus sylvaticus* (GenBank # AF159395) or in *Felis silvestris* (GenBank # KF297791) against genuine, mitochondrial versions of the corresponding marker in different species. Interspecific distances not available for those divergent taxa in alternative markers are marked as "n.a".

Species	COI	CytB	16S
Myotis crypticus vs. *M. nattereri*	0.0%	0.1%	n.a.
Myotis myotis vs. *M. blythii*	0.0%	0.5%	n.a.
Eptesicus serotinus vs. *E. nilssonii*	0.6%	1.3%	0.2%
Canis auratus vs. *Canis lupus*	3.7%	7.3%	1.3%
Cervus elaphus vs. *C. nippon*	3.0%	6.6%	0.9%
Sorex araneus vs. *S. antinorii*	2.6%	2.1%	1.7%
Sorex araneus vs. *S. coronatus*	3.8%	4.5%	1.5%
Arvicola amphibius vs. *A. italicus*	5.7%	4.1%	1.9%
Pseudo-*Apodemus sylvaticus* vs. *A. sylvaticus*	n.a.	12.1%	n.a.
Pseudo-*Apodemus sylvaticus* vs. *A. flavicollis*	n.a.	11.0%	n.a.
Pseudo-*Apodemus sylvaticus* vs. *A. alpicola*	n.a.	8.3%	n.a.
Pseudo-*Felis silvestris* vs. *F. silvestris*	7.5%	n.a.	n.a.

After excluding these chimeric or non-targeted amplicons, we obtained high-quality sequences for a total of 196 individuals from 78 mammal species to build the following three datasets. For the COI dataset, 136 new sequences (varying in length from 567 to 710 bp) representing 104 unique haplotypes of 68 species were retained for further analyses; for the CytB dataset we obtained 108 CytB sequences (length 1094 to 1140 bp) representing 95 unique haplotypes of 68 species; and

for the 16S dataset 141 sequences (length 517 to 540 bp) representing 109 haplotypes of 73 species. The mean ± SD number of individuals examined per species was 1.95±1.70 sample for the COI, 1.59±1.03 for the CytB and 1.93±1.25 for the 16S marker (range 1-20 per species). All newly obtained sequences were deposited in the GenBank under accession numbers OQ706633 - OQ706768 (COI), OQ706633 - OQ706768 (CytB) and OQ707963 - OQ708103 (16S), as detailed in the Table S1.

We also downloaded 335 sequences issued from 26 species from GenBank to obtain comprehensive reference datasets for all 99 species recorded in Switzerland (plus 5 other species found in its immediate periphery, see Graf and Fischer, 2021), as detailed in the Table S1. These downloaded sequences were of course not submitted to identification queries but were carefully checked for taxonomic consistency.

For both COI and CytB haplotypes (the 16S gene being much more conservative, and hence not directly comparable to the other mitochondrial markers), intraspecific pairwise comparisons followed the usual pattern of variation reported within mammal species (Baker and Bradley, 2006), with most K2P distances being smaller than 2 %. Interspecific K2P distance generally exceeded 5 % divergence, suggesting the existence of a barcode gap in this group (matrices of pairwise K2P distances for all genes can be found in the Table S2). However, there were notable exceptions to this usual pattern. Six species of rodents, five species of bats and two eulipotyphlans showed higher levels of intraspecific divergence in one or both markers (Table 2), with *Muscardinus avellanarius* showing up to 11.6 % K2P sequence divergence between haplotypes. The bats *Hypsugo savii* and *Myotis crypticus* also exceeded 9 % divergence between conspecific individuals sampled within Switzerland. Conversely, the following eight pairs of species had unusually similar haplotypes (<5 %): for the bats *Myotis myotis / M. blythii* and *M. crypticus / M. nattereri* some haplotypes were identical (0% divergence for the COI and 0.1-0.5 % for the CytB) in both taxa, whereas for *Eptesicus serotinus / E. nilssonii*, they diverged from each other minimally at 0.6-1.3 % (Table 1). The three closely related shrews *Sorex araneus, S. antinorii* and *S. coronatus* were slightly more divergent from each other (2-4.5 %) (Table 1). Overall, these extremely low interspecific values therefore largely overlapped with those found in most intraspecific comparisons (Table 2).

Table 2 – Unusually high (>2 %) intraspecific K2P distance observed among haplotypes of mammals from Switzerland estimated for the COI and CytB mitochondrial markers (with corresponding values for the 16S gene). Markers envisioned here and other conventions are the same as in the legend of Table 1. Intraspecific distances not available in alternative markers are marked as "n.a.".

Species	COI	CytB	16S
Muscardinus avellanarius	11.6 %	11.3 %	8.9 %
Myotis crypticus	9.0 %	9.1 %	4.5 %
Hypsugo savii	8.9 %	9.7 %	4.3 %
Dryomys nitedula	6.9 %	6.9 %	5.2 %
Pipistrellus kuhlii	4.5 %	5.6 %	3.4 %
Erinaceus europaeus	4.4 %	6.2 %	0.7 %
Apodemus sylvaticus	3.9 %	4.3 %	0.9 %
Arvicola amphibius	3.8 %	4.1 %	1.5 %
Eliomys quercinus	3.0 %	4.9 %	2.3 %
Plecotus auritus	2.6 %	4.5 %	1.7 %
Sorex minutus	2.3 %	2.3 %	1.5 %
Apodemus alpicola	n.a.	2.3 %	0.6 %
Myotis daubentonii	1.5 %	2.1 %	0.2 %

Species-level identification success

The resulting 104, 95 and 109 distinct haplotypes for the COI, CytB and 16S markers, respectively, were submitted to various queries and results were tabulated according to the four categories envisioned here (Fig. 2). Regarding the COI dataset, 61 and 69 % of queries submitted to BOLD ID engine (using the default parameter or choosing the public database, respectively) returned the correct species-level identification, 22 and 11 % returned ambiguous identifications (i.e. with more than one species, including the correct one), 11 and 4 % returned an erroneous identification, while 7 and 16 % resulted in no matching taxon (failed ID). Prevailing erroneous identifications obtained with the default parameters in the ID engine were found in bats (6 haplotypes of 5 species), rodents (4 haplotypes of 2 species) and an erinaceid (1 haplotype of 1 species), while haplotypes with no matching records (failed ID) concerned the mole *Talpa caeca* (1 haplotype), the voles *Arvicola italicus* (2 haplotypes) and *Microtus multiplex* (3 haplotypes), and the dormouse *Dryomys nitedula* (1 haplotype). The more stringent criteria applied to the public records dataset included in BOLD resulted in less ambiguous or erroneous ID, but more queries failed (16 % of 104 haplotypes queried, Fig. 2), including all haplotypes of *Microtus subterraneus* and of *Eliomys quercinus*, and 1 each of the hedgehog *Erinaceus europaeus*, the bat *Hypsugo savii*, the shrew *Sorex alpinus* and the wood mouse *Apodemus sylvaticus*. The same 104 distinct COI haplotypes submitted to BLAST searches in GenBank returned 67 % correct ID, 20 % ambiguous ID, 1 % erroneous ID and 12 % failed ID (i.e. no matching record within 2 % sequence similarity). This last category involved the same species listed above for the ID engine results. Notice that not all the different COI variants found in one species resulted in the same ID category (see Table S3 for details). For instance, of the four distinct *A. flavicollis* COI sequences queried, 3 were erroneously classified as *A. sylvaticus* with the default ID engine in BOLD (at 99.7-100 % sequence similarity) and 1 as either *A. flavicollis* or *A. sylvaticus* with the public dataset, whereas all four were ambiguously identified as *A. flavicollis* or *D. nitedula* (sic) in GenBank searches. The later species appears among the top hits obviously due to a labelling error associated with the COI sequence # MZ661159 which is identical to a series of homologous sequences of *A. flavicollis* from various origins available in GenBank (results not shown).

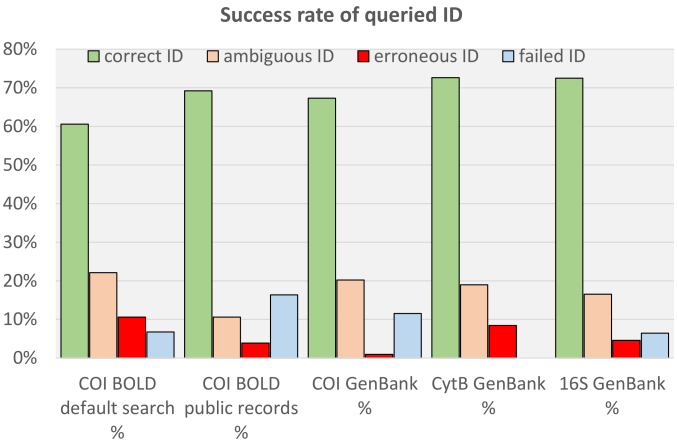

Figure 2 – Success rate of species-level identifications from queries submitted to the identification engine available in BOLD or following a BLAST search in GenBank. Results were categorized as "correct ID" when species corresponded to the morphological ID, as "ambiguous ID" when other species were also suggested within 98% sequence similarity, as "erroneous ID" when the species was discordant with the morphological ID, or as "failed ID" when no record was suggested within 98% sequence similarity. The queried datasets consisted in 104, 95 and 109 distinct haplotypes for the COI, CytB and 16S markers, respectively.

Results from the other two mitochondrial markers (CytB and 16S) are not directly comparable to those obtained for the COI, as haplotypes were not necessarily sequenced from the same individual and available number of sequences for comparisons were much reduced. However, the global pattern of identification success at species-level resolution was similar to that of the COI marker, with a little over 70 % correct ID, nearly 20 % ambiguous ID, 5 to 8 % erroneous ID and 0 to 6 % of failed ID obtained with these two alternative mitochondrial markers (Fig. 2).

Discussion

The molecular identification of unknown samples throughout the DNA barcoding approach based on the COI or other mitochondrial markers has got huge momentum during the last decade, both thanks to the exponential growth of reference sequences available online and to the ease of getting these markers from many sources of biological material (Kocher et al., 2017a). This approach has been particularly useful to identify massive amounts of sequences produced during environmental DNA surveys (Srivathsan et al., 2021; Gostel and Kress, 2022), including while elucidating the diet of interacting species (Andriollo et al., 2021), or for the remote monitoring of elusive species (Galimberti et al., 2015; Gillet et al., 2015; Jamwal et al., 2021). The present survey focusing on the small mammals of Switzerland, however, resulted in surprisingly high levels of taxonomic uncertainty during the standard, automated species-level identifications that are usually performed during such DNA barcoding or metabarcoding approaches. Using newly generated sequences of the popular COI marker in these mammals, queries conducted with default parameters in the ID engine implemented in BOLD returned up to 33 % problematic species-level identifications (Fig. 2). These ambiguous, erroneous or failed identifications were inferred from comparison with sequences generated from carefully identified voucher specimens (Tab. S1). This is worrying given that mammals represent an a priori well-known and well-documented group (Galimberti et al., 2015) and is likely to be much more severe in other, lesser-known taxa. The main reason which led to these inferred taxonomic discordances among Swiss mammals is reviewed hereafter and will serve for better practice to perform more reliable and informed species-level identification based on curated reference databases, as already suggested by earlier studies on mammals (e.g. Galimberti et al., 2015).

Labelling or identification errors

The most common problem compromising unsupervised species-level identifications encountered in our survey was the numerous incorrectly labelled sequences present in publicly available databases (Kvist, 2013; Mutanen et al., 2016). These problematic records include wrong identification of the sequenced specimen or simply labelling errors (Bidartondo et al., 2008; Pentinsaari et al., 2020) and explained 19 out of 78 discordant species ID of the queried mammal haplotypes. This common source of error compromised for instance all identification of haplotypes issued from two bat species easily identified by external characters (the common noctule *Nyctalus noctula* and the serotine bat *Eptesicus serotinus*) because one published complete mitochondrial genome was clearly issued from a mislabelled (or wrongly identified) individual (GenBank #MT584130). Hence all current queries generated wrong or ambiguous identifications (even at 100% match), regardless of the marker or the method used for comparisons. Another example of such errors generating discordant results was the published CytB sequence labelled as *Rattus rattus* (#UAM97703) but which matched 100% with any other *R. norvegicus* present in GenBank (including our queried new haplotype), and thus rendered species-level identification ambiguous. The use of more stringent criteria during queries, e.g., the use of the “Public Record Barcode Database" (Kvist, 2013) available in BOLD or iBOLD instead of the default one (Fig. 1) or of a lower threshold (e.g., 1% maximal dissimilarity) before accepting species-level identification (Galimberti et al., 2012) would not solve this source of errors. Such pervasive labelling errors are unfortunately widespread in GenBank and BOLD repositories, despite efforts to better curate these databases (Kvist, 2013). Unless such errors can be easily flagged by expert taxonomists (Hosner et al., 2022), one way to avoid them would be to analyse not only the best match, but all returned sequences within a given similarity level of genetic distance and check their taxonomic coherence. BOLD already offers this possibility, and already warns when more than one species appears in the same BIN, but this is usually overlooked when hundreds or thousands of unknown sequences are queried with automated pipelines. This coherence check would also allow building external reference libraries that are purged from obvious errors. In the supplementary material (Tab. S4), we provide such a curated reference library for all 104 wild mammals living in or near Switzerland and for three popular mitochondrial markers (COI, CytB and 16S).

Incoherence due to outdated taxonomies

Another common source of major taxonomic discordances observed in the queried datasets was due to outdated taxonomies used in the public repositories. When a species name is associated to a changing genus, like for the red vole *Clethrionomys glareolus* which was sometimes known as *Myodes glareolus*, the coexistence of both binomials in repositories does not generate ambiguities in DNA barcoding approaches, as both designations will point to a single taxon. However, when taxonomic revisions involve the split of a species complex into several independent species (e.g. Juste et al., 2018; Kryštufek, 2018), then sequences deposited prior to the change will likely not be updated and will generate ambiguous or wrong species-level ID in subsequent queries. These taxonomic inconsistencies were experienced here for the species pairs *Myotis nattereri/crypticus*, *Microtus agrestis/lavernedii, Arvicola amphibius/italicus* and *Sorex araneus/antinorii/coronatus*, all of which produced ambiguous or wrong species-level identifications in both GenBank and BOLD queries. Updating these repositories is a major challenge, as deposited sequences generally lack the appropriate metadata to allow objective taxonomic reassignment or were not based on vouchered specimens that could be re-identified. We suggest that for those revised species complexes, and unless reference databases are based on up-to-date taxonomies, the use of aggregates should be associated to them (e.g., *Sorex araneus* aggr.) to warn for potential inconsistencies.

Such errors due to outdated taxonomies in reference libraries may have far reaching consequences. For instance, during the survey of potential zoonotic reservoirs of coronaviruses in Chinese bats, several hundred potential horsts (chiropterans) were biopsied and identified through their mitochondrial DNA barcodes (Latinne et al., 2020). The authors concluded that several of those barcoded bats were potential reservoirs, including six species (*Myotis daubentonii*, *M. myotis, Miniopterus schreibersii*, *Miniopterus fuscus*, *Pipistrellus pipistrellus*, and *Plecotus auritus*) that do not live in China (Wilson and Mittermeier, 2019). The reason for this confusion is that these taxa are represented in the GenBank by sequences issued from Asian specimens (and indeed very similar to the queried Chinese sequences) but labelled with old names that are currently considered as different species. Such taxonomic confusions thus may seriously compromise our understanding of the evolution of important zoonotic diseases.

Introgressed or hybridizing taxa

It is largely accepted that recently evolved species such as domesticated forms of wild-living ancestors (e.g., dog/wolf, domestic/wild cat, ferret/polecat) cannot be properly identified with classical mitochondrial markers either because they lack diagnostic mutations, or because they share haplotypes through repeated episodes of hybridization (Mallet, 2005; Berthier et al., 2006; Seixas et al., 2018). In such cases, specific markers based on nuclear genome must be developed (e.g. Nussberger et al., 2013; Marques et al., 2017), but they lack the versatility of being used in broader surveys of entire groups such as all local mammals. Genome skimming (Bohmann et al., 2020) provides a promising avenue to alleviate this lack of resolution, but still is not available in most applications. Although relatively rare among mammals (Mallet, 2005), biological species hybridizing under natural conditions do occur in the wild. If overlooked these situations may generate ambiguous or erroneous species-level identifications throughout the mitochondrial DNA barcoding approach (Salokannel et al., 2021). During the present survey of mammalian species, such errors or ambiguities were not observed because we sampled representative individuals away for known hybrid zones. But several species pairs of wild mammals are known to hybridize regularly in restricted areas of Switzerland, including *Sorex araneus/antinorii* (Lugon-Moulin et al., 1996), *Lepus timidus/europaeus* (Zachos et al., 2010) in the Alps, or *Microtus agrestis/lavernedii* in the Jura mountains (Beysard et al., 2012). Al-

though hybrids are relatively rare in these areas, surveys based on environmental mitochondrial DNA might also lead to errors if they are ignored.

Perhaps an even more pervasive phenomenon is the occurrence of wild mammals which inherited alien mitochondrion throughout ancient, massive introgression (Berthier et al., 2006; Artyushin et al., 2009; Juste et al., 2013; Afonso et al., 2017; Çoraman et al., 2019), which indeed explained most identification errors we observed in the species pairs *Myotis myotis/M. blythii*, *Myotis nattereri/M. crypticus* and *Eptesicus serotinus/E. nilsonii*. In these pairs of sister species, both taxa share extremely similar (<2 % sequence divergence; see Table 2) or even identical mitogenomes over extensive areas in Europe. Unless nuclear markers are developed to differentiate them specifically, they again should best be regarded as species aggregates (Galan et al., 2018), even if 100 % match is returned in automated queries. This case was observed for instance for two haplotypes issued from morphologically identified *M. myotis*, which matched 100 % with COI sequences of *M. blythii* previously deposited in GenBank or BOLD. The same limitation of the DNA barcoding approach would apply to other pairs of sister taxa found elsewhere in Europe (Juste et al., 2013; Çoraman et al., 2020; Zolotareva et al., 2020) so that caution should be exerted since the exact extent of introgression are still unknown.

Highly divergent haplotypes or cryptic species?

Although individual genetic variation is often minimal within most mammal species (Bradley and Baker, 2001), geographic barriers to gene flow or historical movements between previously isolated subpopulations may favour the persistence or coexistence of highly differentiated lineages in situation of close geographic proximity (Hewitt, 2004). Such highly divergent mitochondrial lineages might therefore simply reflect unusual intraspecific variation or could flag the existence of additional taxonomic diversity (Meier et al., 2006; Dasmahapatra et al., 2010), but clearly evidence drawn from other type of characters (e.g., from nuclear genes, chromosomes or morphology) must be considered before concluding to either possibility. The two major haplogroups differing by up to 4.5 % divergence at the CytB (Table 1) evidenced in Kuhl's pipistrelles (*P. kuhlii*) illustrate this situation. An earlier study showed that such unusually high divergence of mitochondrial markers in *P. kuhlii* should not be interpreted in terms of cryptic taxonomic diversity, as individuals carrying these lineages were analysed simultaneously with multiple nuclear markers (microsatellites) in several populations of Western Switzerland and clearly showed extensive interbreeding, as if drawn from a single panmictic population (Andriollo et al., 2015).

Other surveyed mammals from Switzerland showed comparatively high levels of mitochondrial differentiation (Table 1). For instance, the COI haplotype of a European hedgehog *E. europaeus* sampled in the Geneva region differed by at least 4.4 % from any other COI sequence deposited so far in GenBank or BOLD (Table 1). Previous phylogeographic studies based on other mitochondrial markers showed that two divergent mtDNA lineages (E1 and E2) in this species occur in Europe and meet along a line running from Italy northwards through Switzerland, Germany, and Scandinavia (Seddon et al., 2001), while a third lineage (E3) was endemic to Sicily. All lineages, however, lacked corresponding genetic discontinuities at nuclear genes, confirming that they represented intraspecific mitotypes (Seddon et al., 2001). As only representatives of one of these main lineages were available so far for the COI gene, our queries returned no matching sequence within 2 % divergence (i.e. were classified as a failed ID). Similar results were obtained with several other mammals, including *Apodemus sylvaticus*, *Eliomys quercinus*, *Dryomys nitedula* or *Hypsugo savii* (Table 1), all of which were already known to exhibit major intraspecific variation among disjunctive or sympatric populations within Western Europe (Ibáñez et al., 2006; Perez et al., 2013; Herman et al., 2016; Mohammadi et al., 2021). These CytB-based lineages are currently underrepresented in COI databases, resulting in some of the "failed" species-level identifications of our queries (Fig. 2).

Divergent lineages from one of the investigated Gliridae, the hazel dormouse (*Muscardinus avellanarius*), however, underline a different biological situation, as they most likely represent two cryptic species. Mitochondrial markers sequenced in three individuals from various places in Switzerland had very high maximal levels of divergence when compared to other European samples (over 11 % for the COI or CytB, and over 8 % for the 16S; Table 1). A previous, broad-scale phylogeographic study (Mouton et al., 2012) showed that northern and eastern European populations were characterized by a L2 lineage that was very distinct from the L1 lineage prevalent in all assayed Western European and Italian populations. The individuals from western Switzerland and nearby France (field number M2030, M2278, T-1451) analysed here indeed pertained to the L1 lineage, while those from eastern Switzerland (V7366 and MZ661234) had similar haplotypes to those found in Denmark or the UK, i.e., representing the L2 lineage (Tab. S1 and S2). Mouton et al. (2017) later demonstrated that individuals carrying these two main lineages across Europe shared no allele in common at 10 highly variable microsatellite loci, nor at two nuclear introns, demonstrating ancient and complete lack of gene flow between those two groups. As animals corresponding to both lineages also differ in some external and dental characters (Miller, 1912; Von Witte, 1962), we suggest that there is ample independent evidence to consider each group/lineage as two independent biological species. We thus propose to apply the following species name for them: the nominal form originally described from Scandinavia (i.e. corresponding to lineage L2) should retain the oldest name *Muscardinus avellanarius* (Linnaeus, 1758), while animals from Western Europe should be called *Muscardinus speciosus* a taxon originally described from southern Italy (Dehne, 1855). Although *M. pulcher* (Barrett-Hamilton, 1898) has also been proposed to name Italian specimens (e.g. Miller, 1912) *speciosus* has priority as this is the most ancient available name unambiguously attributable to populations carrying the L1 lineage. Recognizing both lineages as distinct species has important bearing for conservation because hazel dormice are declining and have been classified as locally threatened in several countries (e.g. in Switzerland; Capt, 2022). If translocations are proposed to reinforce local populations, this might lead to problematic biological interactions if both *M. avellanarius* and *M. speciosus* are inadvertently involved in such conservation programmes.

Other experimental errors – pseudogenes

A final example which explained part of the abnormally divergent haplotypes found in intraspecific comparisons of the mammals investigated here was the existence of paralogs of mitochondrial genes (Ermakov et al., 2015). We inadvertently amplified and sequenced a pseudogene of the CytB gene in several *Apodemus* mice which turned to be identical to a GenBank sequence (AF159395) already flagged by Dubey et al. (2009) as a nuclear copy of the CytB gene. This nuclear copy differed by up to 12 % from genuine CytB gene sequences (Table 1) and was present in several copies within the chromosome 22 genome of a completely sequenced *A. sylvaticus* individual (see Results section). As the amplification of non-target fragments is primer-dependent, the use of different sets of primers, or a combination of different markers (as done here) will help detecting inconsistencies. Careful examination of the obtained raw sequences for the existence of anomalies such as stop codons, insertion/deletions or double peaks should also be performed to avoid further confusions (Dubey et al., 2009; Shi et al., 2017). Completely avoiding the amplification of nuclear copies of a targeted mitochondrial gene is not easy but several strategies, including the use of long-range PCR (Emser et al., 2021), have been proposed to minimize the problem. If overlooked, such inadvertently amplified and sequenced paralogs may easily confound species-level identification with the DNA barcoding approach (Dubey et al., 2009).

Incomplete reference libraries

Incomplete reference database is often mentioned as a reason for failure to identify unknown samples with the DNA barcode approach (Galimberti et al., 2015; Bengtsson-Palme et al., 2018). We were sur-

prised that the following four mammalian species of the Swiss fauna lacked any reference sequence for the COI or 16S markers: *Microtus multiplex*, *Dryomys nitedula*, *Talpa caeca* and *Sorex alpinus* (the only COI sequence of *S. alpinus* - EF636551 - being a quite divergent pseudogene). According to a recent review (Graf and Fischer, 2021), these species are rare and/or endangered in Switzerland and need more distributional data to inform specific conservation plans. If those plans would include surveys based on environmental DNA, they would likely have been overlooked due to lack of reference sequence. It is thus of paramount importance to conduct such broad, DNA-based surveys with carefully built libraries (Bengtsson-Palme et al., 2018), where lists of all potential species under focus must be checked for availability. If no tissue is available for new DNA sequencing, another possibility is to rely on alternative markers, as exemplified here with the CytB gene for which matching sequences for all these four species were already publicly available in the GenBank (Fig. 2). In future studies, this problem will likely be less severe, at least for the most popular or tractable taxa such as mammals, as more and more sequences are submitted to GenBank or BOLD (Mutanen et al., 2016).

References

Afonso, E., Goydadin A.C., Giraudoux P., Farny G., 2017. Investigating hybridization between the two sibling bat species *Myotis myotis* and *M. blythii* from guano in a natural mixed maternity colony. Plos One 12: e0170534.

Andersen L.W., Jacobsen M.W., Frydenberg J., Møller J.D., Jensen T.S., 2022. Phylogeography using mitogenomes: A rare Dipodidae, Sicista betulina, in North-western Europe. Ecol. Evol. 12: e8865.

Andriollo T., Naciri Y., Ruedi M., 2015. Two mitochondrial barcodes for one biological species: The case of European Kuhl's pipistrelles (*Chiroptera*). Plos One 10: e0134881.

Andriollo T., Michaux J.R., Ruedi M., 2021. Food for everyone: Differential feeding habits of cryptic bat species inferred from DNA metabarcoding. Mol. Ecol. 30: 4584–4600.

Artyushin I.V., Bannikova A.A., Lebedev V.S., Kruskop S.V., 2009. Mitochondrial DNA relationships among North Palaearctic Eptesicus (*Vespertilionidae, Chiroptera*) and past hybridization between Common Serotine and Northern Bat. Zootaxa 2262: 40–52.

Baker R.J., Bradley R.D., 2006. Speciation in mammals and the genetic species concept. J. Mammal. 87: 643–662.

Ballard J.W.O., Whitlock M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13: 729–744.

Bengtsson-Palme J., Richardson R.T., Meola M., Wurzbacher C., Tremblay É.D., Thorell K., Kanger K., Eriksson K.M., Bilodeau G.J., Johnson R.M., Hartmann M., Nilsson R.H., 2018. Metaxa2 Database Builder: enabling taxonomic identification from metagenomic or metabarcoding data using any genetic marker. Bioinformatics 34: 4027–4033.

Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Ostell J., Pruitt K.D., Sayers E.W., 2018. GenBank. Nucl. Acids Rec. 46: D41–D47.

Berthier P., Excoffier L., Ruedi M., 2006. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. Proc. R. Soc. Lond. [Biol] 273: 3101–3109.

Beysard M., Perrin N., Jaarola M., Heckel G., Vogel P., 2012. Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). J. Evolution. Biol. 25: 400–408.

Bidartondo M.I., Bruns T.D., Blackwell M., Edwards I., Taylor A.F., Horton T., Zhang N., Kõljalg U., May G., Kuyper T.W., Bever J.D., 2008. Preserving accuracy in GenBank. Science 319: 1616a.

Bohmann K., Mirarab S., Bafna V., Gilbert M.T.P., 2020. Beyond DNA barcoding: The unrealized potential of genome skim data in sample identification. Mol. Ecol. 29: 2521–2534.

Bolfíková B., Hulva P., 2012. Microevolution of sympatry: landscape genetics of hedgehogs *Erinaceus europaeus* and *E. roumanicus* in Central Europe. Heredity 108: 248–255.

Boratyn G.M., Camacho C., Cooper P.S., Coulouris G., Fong A., Ma N., Madden T.L., Matten W.T., Mcginnis S.D., Merezhuk Y., Raytselis Y., Sayers E.W., Tao T., Ye J., Zaretskaya I., 2013. BLAST: a more efficient report with usability improvements. Nucl. Acids Rec. 41: W29–W33.

Borisenko A., Lim B., Ivanova N., Hanner R., Hebert P., 2008. DNA barcoding in surveys of small mammal communities: A field study in Suriname. Mol. Ecol. Resour. 8: 471–479.

Bradley R., Baker R., 2001. A test of the genetic species concept: cytochrome b sequences and mammals. J. Mammal. 82: 960–973.

Capt S., 2022. Liste rouge des mammifères (hors chauves-souris). Espèces menacées en Suisse. Office fédéral de l'environnement (OFEV); Info Fauna (CSCF), Bern and Neuchâtel, Switzerland. [in French]

Çoraman E., Dundarova H., Dietz C., Mayer F., 2020. Patterns of mtDNA introgression suggest population replacement in Palaearctic whiskered bat species. R. Soc. open sci. 7: 191805.

Çoraman E., Dietz C., Hempel E., Ghazaryan A., Levin E., Presetnik P., Zagmajster M., Mayer F., 2019. Reticulate evolutionary history of a Western Palaearctic bat complex explained by multiple mtDNA introgressions in secondary contacts. J. Biogeogr. 46: 343–354.

Dasmahapatra K.K., Elias M., Hill R.I., Hoffman J.I., Mallet J., 2010. Mitochondrial DNA barcoding detects some species that are real, and some that are not. Mol. Ecol. Resour. 10: 264–273.

Davison A., Birks J.D.S., Brookes R.C., Messenger J.E., Griffiths H.I., 2001. Mitochondrial phylogeography and population history of pine martens *Martes martes* compared with polecats *Mustela putorius*. Mol. Ecol. 10: 2479–2488.

Dehne A., 1855. *Myoxus speciosus* Dene. Prächtiger Haselschläfer. Allg. Dt. Naturhist. Zeitg. 7: 180. [in German]

Dietz C., von Helversen O., 2004. Illustrated identification key to the bats of Europe. www.uni-tuebingen.de/tierphys/Kontakt/mitarbeiter_seiten/dietz.htm

Dietz C., Kiefer A., 2015. Les chauves-souris d'Europe. Connaître, déterminer, protéger. Delachaux et Niestlé, Paris, France. [in French]

Dubey S., Michaux J., Brunner H., Hutterer R., Vogel P., 2009. False phylogenies on wood mice due to cryptic cytochrome-b pseudogene. Mol. Phylogenet. Evol. 50: 633–641.

Emser S.V., Schaschl H., Millesi E., Steinborn R., 2021. Extension of mitogenome enrichment based on single long-range PCR: mtDNAs and putative mitochondrial-derived peptides of five rodent hibernators. Frontiers in Genetics: 2434.

Ermakov O.A., Simonov E., Surin V.L., Titov S.V., Brandler O.V., Ivanova N.V., Borisenko A.V., 2015. Implications of hybridization, NUMTs, and overlooked diversity for DNA barcoding of Eurasian ground squirrels. Plos One 10.

Galan M., Pons J.B., Tournayre O., Pierre E., Leuchtmann M., Pontier D., Charbonnel N., 2018. Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. Mol. Ecol. Resour. 18: 474–489.

Galimberti A., Sandionigi A., Bruno A., Bellati A., Casiraghi M., 2015. DNA barcoding in mammals: what's new and where next? Hystrix 26: 12.

Galimberti A., De Mattia F., Losa A., Bruni I., Federici S., Casiraghi M., Martellos S., Labra M., 2013. DNA barcoding as a new tool for food traceability. Food Res. Int. 50: 55–63.

Galimberti A., Spada M., Russo D., Mucedda M., Agnelli P., Crottini A., Ferri E., Martinoli A., Casiraghi M., 2012. Integrated Operational Taxonomic Units (IOTUs) in echolocating bats: A bridge between molecular and traditional taxonomy. Plos One 7: e40122.

Giguet-Covex C., Pansu J., Arnaud F., Rey P.J., Griggo C., Gielly L., Domaizon I., Coissac E., David F., Choler P., Poulenard J., Taberlet P., 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat. Commun. 5 : 3211.

Gillet F., Tiouchichine M.-L., Galan M., Blan F., Némoz M., Aulagnier S., Michaux J.R., 2015. A new method to identify the endangered Pyrenean desman (*Galemys pyrenaicus*) and to study its diet, using next generation sequencing from faeces. Mamm. Biol. 80: 505–509.

Gostel M.R., Kress W.J., 2022. The expanding role of DNA barcodes: Indispensable tools for ecology, evolution, and conservation. Diversity 14: 213.

Gouda S., Kerry R.G., Das A., Chauhan N.S., 2020. Wildlife forensics: A boon for species identification and conservation implications. Forensic Science International 317: 110530.

Graf R.F., Fischer C. (Eds.), 2021. Atlas des mammifères de Suisse et du Liechtenstein. Haupt Verlag, Berne. [in French]

Hassanin A., Ropiquet A., Couloux A., Cruaud C., 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J. Mol. Evol. 68: 293–310.

Hassanin A., Delsuc F., Ropiquet A., Hammer C., Van Vuuren B.J., Matthee C., Ruiz-Garcia M., Catzeflis F., Areskoug V., Nguyen T.T., Couloux A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C. R. Biologies 335: 32–50.

Hebert P.D.N., Ratnasingham S., Dewaard J.R., 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. [Biol] 270: S96–S99.

Hebert P.D.N., Cywinska A., Ball S.L., Dewaard J.R., 2003b. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. [Biol] 270: 313–321.

Herman J.S., Jóhannesdóttir F., Jones E.P., Mcdevitt A.D., Michaux J.R., White T.A., Wójcik J.M., Searle J.B., 2016. Post-glacial colonization of Europe by the wood mouse, *Apodemus sylvaticus*: evidence of a northern refugium and dispersal with humans. Biol. J. Linn. Soc. 120: 313–332.

Hewitt G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. Phil. Trans. Royal Soc London. Series B: Biol. Sci. 359: 183–195.

Hosner P.A., Zhao M., Kimball R.T., Braun E.L., Burleigh J.G., 2022. Updating splits, lumps, and shuffles: Reconciling GenBank names with standardized avian taxonomies. Ornithology 139.

Ibáñez C., García-Mudarra J.L., Ruedi M., Stadelmann B., Juste J., 2006. The Iberian contribution to cryptic diversity in European bats. Acta Chiropterol. 8: 277–297.

Ivanov V., Lee K.M., Mutanen M., 2018. Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. Mol. Ecol. 27: 1681–1695.

Ivanova N., Zemlak T., Hanner R., Hebert P., 2007. Universal primer cocktails for fish DNA barcoding. Mol. Ecol. Notes 7: 544–548.

Jamwal P.S., Bruno A., Galimberti A., Magnani D., Krupa H., Casiraghi M., Loy A., 2021. First assessment of eDNA-based detection approach to monitor the presence of Eurasian otter in Southern Italy. Hystrix 32: 176–181.

Juste J., Benda P., Garcia-Mudarra J.L., Ibáñez C., 2013. Phylogeny and systematics of Old World serotine bats (genus *Eptesicus*, Vespertilionidae, Chiroptera): an integrative approach. Zool. Scr. 42: 441–457.

Juste J., Ruedi M., Puechmaille S.J., Salicini I., Ibáñez C., 2018. Two new bat species within the *Myotis nattereri* species complex (Vespertilionidae, Chiroptera) from the Western Palaearctic. Acta Chiropterol. 20: 285–301.

Kocher A., De Thoisy B., Catzeflis F., Valière S., Bañuls A.-L., Murienne J., 2017a. iDNA screening: Disease vectors as vertebrate samplers. Mol. Ecol. 26: 6478–6486.

Kocher A., De Thoisy B., Catzeflis F., Huguin M., Valière S., Zinger L., Bañuls A.-L., Murienne J. 2017b. Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. Methods Ecol. Evol. 8: 1276–1283.

Kryštufek, B. 2018. Mediterranean field vole *Microtus lavernedii* (p. 333). In: Wilson D. E., Lacher T. E., Mittermeier R. A. (Eds). Handbook of the mammals of the world: Rodents II, Lynx Edicions, Barcelona, Spain, Vol 7.

Kumar S., Stecher G., Tamura K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874.

Kvist S., 2013. Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Mol. Phylogenet. Evol. 69: 39–45.

Kvist S., 2016. Does a global DNA barcoding gap exist in Annelida? Mitochondrial DNA A 27: 2241–2252.

Latinne A., Hu B., Olival K.J., Zhu G., Zhang L., Li H., Chmura A.A., Field H.E., Zambrana-Torrelio C., Epstein J.H., Li B., Zhang W., Wang L.-F., Shi Z.-L., Daszak P., 2020. Origin and cross-species transmission of bat coronaviruses in China. Nat. Commun. 11: 4235.

Lugon-Moulin N., Wyttenbach A., Brünner H., Goudet J., Hausser J., 1996. Study of gene flow through a hybrid zone in the common shrew (*Sorex araneus*) using microsatellites. Hereditas 125: 159–168.

Mallet J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20: 229–237.

Marchesi P., Blant M., Capt S. (Eds.), 2008. Mammifères de Suisse - Clés de détermination. CSCF and SSBF, Neuchâtel, Switzerland. [in French]

Margaryan A., Noer C.L., Richter S.R., Restrup M.E., Bülow-Hansen J.L., Leerhøi F., Langkjær E.M.R., Gopalakrishnan S., Carøe C., Gilbert M.T.P., Bohmann K., 2021. Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark. Environ. DNA 3: 472–480.

Marques J.P., Ferreira M.S., Farelo L., Callahan C.M., Hackländer K., Jenny H., Montgomery W.I., Reid N., Good J.M., Alves P.C., Melo-Ferreira J., 2017. Mountain hare transcriptome and diagnostic markers as resources to monitor hybridization with European hares. Sci. Data 4: 1–11.

Meier R., Shiyang K., Vaidya G., Ng P.K.L., 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst. Biol. 55: 715–728.

Meiklejohn K.A., Damaso N., Robertson J.M., 2019. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. Plos One 14: e0217084.

Miller G.S., 1912. Catalogue of the Mammals of Western Europe (Europe exclusive of Russia) in the collection of the British Museum. Trustees of the British Museum (Natural History), London, UK.

Mohammadi Z., Kami H.G., Ghorbani F., Khajeh A., Olsson U., 2021. Cryptic lineage diversity within Forest Dormice (Mammalia: *Dryomys nitedula*) revealed by deep genetic divergence among different subspecies on the Iranian Plateau and in adjacent areas. Mamm. Biol. 101: 21–34.

Mouton A., Grill A., Sara M., Kryštufek B., Randi E., Amori G., Juškaitis R., Aloise G., Mortelliti A., Panchetti F., 2012. Evidence of a complex phylogeographic structure in the common dormouse, *Muscardinus avellanarius* (Rodentia: Gliridae). Biol. J. Linn. Soc. 105: 648–664.

Mouton A., Mortelliti A., Grill A., Sara M., Kryštufek B., Juškaitis R., Latinne A., Amori G., Randi E., Büchner S., Schulz B., Ehlers S., Lang J., Adamik P., Verbeylen G., Dorenbosch M., Trout R., Elmeros M., Aloise G., Mazzoti S., Matur F., Poitevin F., Michaux J.R., 2017. Evolutionary history and species delimitations: a case study of the hazel dormouse, *Muscardinus avellanarius*. Conserv. Genet. 18: 181–196.

Mutanen M., Kivelä S.M., Vos R.A., Doorenweerd C., Ratnasingham S., Hausmann A., Huemer P., Dincă V., Van Nieukerken E.J., Lopez-Vaamonde C., Vila R., Aarvik L., Decaëns T., Efetov K.A., Hebert P.D.N., Johnsen A., Karsholt O., Pentinsaari M., Rougerie R., Segerer A., Tarmann G., Zahiri R., Godfray H.C.J., 2016. Species-level para- and polyphyly in DNA barcode gene trees: Strong operational bias in European Lepidoptera. Syst. Biol. 65: 1024–1040.

Nussberger B., Greminger M.P., Grossen C., Keller L.F., Wandeler P., 2013. Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. Mol. Ecol. Resour. 13: 447–460.

Palumbi S.R., Kewalo A.M., Romano S., Mcmillan W.O., Stice L., Grabowski G., 1991. The simple fool's guide to PCR. (Universtiy of Hawaii, Honolulu, USA), 45 pp.

Pentinsaari M., Ratnasingham S., Miller S.E., Hebert P.D.N., 2020. BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? Plos One 15: e0231814.

Perez G.C.L., Libois R., Nieberding C.M., 2013. Phylogeography of the garden dormouse Eliomys quercinus in the western Palearctic region. J. Mammal. 94: 202–217.

Ratnasingham S., Hebert P.D.N., 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol. Ecol. Notes 7: 355–364.

Ratnasingham S., Hebert P.D.N., 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. Plos One 8: e66213.

Razgour O., Clare E.L., Zeale M.R.K., Hanmer J., Schnell I.B., Rasmussen M., Gilbert T.P., Jones G., 2011. High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. Ecol. Evol. 1: 556–570.

Rodriguez F., Perez T., Hammer S.E., Albornoz J., Dominguez A., 2010. Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus *Rupicapra*). BMC Evol. Biol. 10.

Ruedi M., 2021. Espèces additionnelles - Rhinolophe euryale Rhinolophus euryale (Blasius, 1853) (p. 438). In: Graf R. F., Fischer C. (Eds). Atlas des mammifères de Suisse et du Liechtenstein Haupt Verlag Berne Société Suisse de Biologie de la Faune SSBF. [in French]

Salokannel J., Lee K.M., Rinne A., Mutanen M., 2021. ddRAD sequencing sheds light on low interspecific and high intraspecific mtDNA divergences in two groups of caddisflies. Insect Syst. Div. 5: 4.

Seddon J.M., Santucci F., Reeve N.J., Hewitt G.M., 2001. DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*: Pleitocene refugia, postglacial expansion and colonisation routes. Mol. Ecol. 10: 2187-2198.

Seixas F.A., Boursot P., Melo-Ferreira J., 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biol. 19: 91.

Shi H., Xing Y., Mao X., 2017. The little brown bat nuclear genome contains an entire mitochondrial genome: Real or artifact? Gene 629: 64–67.

Srivathsan A., Lee L., Katoh K., Hartop E., Kutty S.N., Wong J., Yeo D., Meier R., 2021. ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. BMC Biol. 19: 217.

Thompson J.D., Gibson T.J., Higgins D.G., 2003. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics : 2.3. 1–2.3. 22.

Von Witte G.R., 1962. Zur Systematik der Haselmaus *Muscardinus avellanarius* L. Bonner zool. Beitr. 13: 312-320. [in German]

Wilson D.E., Mittermeier R.A. (Eds.), 2019. Handbook of the Mammals of the World: Vol. 9 Bats. Lynx Edicions, Barcelona, Spain.

Yannic G., Basset P., Hausser J., 2008. A new perspective on the evolutionary history of western European *Sorex araneus* group revealed by paternal and maternal molecular markers. Mol. Phylogenet. Evol. 47: 237–250.

Yao M., Zhang S., Lu Q., Chen X., Zhang S.-Y., Kong Y., Zhao J., 2022. Fishing for fish environmental DNA: Ecological applications, methodological considerations, surveying designs, and ways forward. Mol. Ecol. 31: 5132–5164.

Zachos F.E., Ben Slimen H., Hackländer K., Giacometti M., Suchentrunk F., 2010. Regional genetic in situ differentiation despite phylogenetic heterogeneity in Alpine mountain hares. J. Zool. 282: 47–53.

Zeale M.R., Butlin R.K., Barker G.L., Lees D.C., Jones G., 2011. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. Mol. Ecol. Resour. 11: 236–244.

Zolotareva K.I., Belokon M.M., Belokon Y.S., Rutovskaya M.V., Hlyap L.A., Starykov V.P., Politov D.V., Lebedev V.S., Bannikova A.A., 2020. Genetic diversity and structure of the hedgehogs *Erinaceus europaeus* and *Erinaceus roumanicus*: Evidence for ongoing hybridization in Eastern Europe. Biol. J. Linn. Soc. 132: 174–195.

Associate Editor: R. Caniglia

Supplemental information

Additional Supplemental Information may be found in the online version of this article:

Table S1 Sampling information of all mammals newly sequenced or downloaded from the GenBank and retained as reference COI, CytB and 16S datasets.

Table S2 Matrices of pairwise K2P distances among all newly sequenced haplotypes of COI, CytB and 16S. Genus and species name abbreviation correspond to the initial three letters of the binomial names (e.g., "Eptser" stands for *Eptesicus serotinus*) followed by field number, haplotype and gene abbreviation (e.g., "Eptser_M2638_H2_COI"). For convenience, we partitioned matrices into two groups of species (a) those belonging to the orders Chiroptera and Lipotyphla, and (b) to all other orders (Carnivora, Rodentia and Ungulata).

Table S3 Listing of the 304 newly sequenced haplotypes of COI, CytB and 16S queried in BOLD and GenBank, with the corresponding species-level identification success obtained in those queries.

Table S4 Reference sequences of all species of wild mammals recorded in Switzerland and neighbouring regions, for the COI, CytB and 16S database, respectively. The content of column G can easily be transformed in fasta format to get complete alignments.