Received: 1 July 2021 | Revised: 15 September 2021 | Accepted: 15 September 2021

DOI: 10.1002/edn3.256

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

ORIGINAL ARTICLE

Creating genetic reference datasets: Indirect sampling of target species using terrestrial leeches as sample "collectors"

Andreas Wilting[1] | Thanh V. Nguyen[1,2,3,4] | Jan Axtner[1] | An Nguyen[1,4] | Anke Schmidt[2] | Minh Le[3,5] | Anh H. Q. Nguyen[6] | Benjamin M. Rawson[6] | Andrew Tilker[1,4] | Joerns Fickel[2,7]

[1]Department of Ecological Dynamics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

[2]Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

[3]Central Institute for Natural Resources and Environmental Studies, Vietnam National University, Hanoi, Vietnam

[4]Re:wild, Austin, Texas, USA

[5]Department of Environmental Ecology, Faculty of Environmental Science, Hanoi University of Science, Vietnam National University, Hanoi, Vietnam

[6]WWF Viet-Nam, Hanoi, Vietnam

[7]Institute for Biochemistry and Biology, University of Potsdam, Potsdam-Golm, Germany

Correspondence
Andreas Wilting, Department of Ecological Dynamics, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Straße 17, 10315 Berlin, Germany
Email: wilting@izw-berlin.de

Funding information
Point Defiance Zoo and Aquarium; Safari Club International Foundation; Leibniz-Institut für Zoo- und Wildtierforschung; GreaterGood; Ocean Park Conservation Foundation, Hong Kong; Russell E. Train Education for Nature; National Geographic Society; Critical Ecosystem Partnership Fund; German Federal Ministry of Education and Research, Grant/Award Number: 01LN1301A; Mohamed bin Zayed Species Conservation Fund

Abstract

Sampling genetic material from rare and often secretive mammals can be difficult in challenging environments such as tropical rainforests. Large-scale sampling is important however for resolving species' taxonomic uncertainties, as well as to help provide genetic material for reference databases that can be used with DNA forensics to combat the illegal wildlife trade. Environmental DNA (eDNA) and invertebrate-derived DNA (iDNA) offer a promising way to overcome this sampling bottleneck, though to date e/iDNA studies have been primarily used to obtain information about species presence using barcoding or metabarcoding approaches. In this study, we collected 4,710 terrestrial leeches across six protected areas in the Annamite Mountains of Viet Nam and Laos and sequenced fragments (223–694 bp) of *mt-Cytb* and/or *mt-Dcr* of the mitogenomes from six species that are frequently found in the illegal wildlife trade. Although amplification success was generally low (<50%) because DNA was highly degraded, we showed that e/iDNA samples can be an important source of genetic material for difficult to sample species and show promise as a way to fill existing sampling gaps. Because e/iDNA sampling has the potential to generate much-needed genetic data to improve taxonomic assessments, this approach can be used to support conservation management by contributing to genetic databases that are necessary to determine the origin of wildlife products or to inform the reintroduction of confiscated animals.

KEYWORDS
DNA forensics, invertebrate-derived DNA (iDNA), phylogeography, Southeast Asia, wildlife trade

1 | INTRODUCTION

Understanding the genetic diversity of species and populations is key to taxonomic assignments, which in turn are necessary for modern adaptive species management (Isaac et al., 2004; Mace, 2004; Wilting et al., 2015) and biodiversity assessments following the UN Sustainable Development Goals (https://sdgs.un.org/goals). Furthermore, genetic information from populations across species' geographic range can assist forensic scientists in tracing the origin of confiscated animals and animal products in the wildlife trade (Blair et al., 2017; Eaton et al., 2010; Le et al., 2020; Wasser et al., 2008; Zhang et al., 2015). The lack of such genetic information often hampers law enforcement effectiveness (and potential convictions) as the origin of the traded animals or animal products cannot be determined. Therefore, data on genetic diversity of species and local differentiation of populations are not only important for evolutionary biologists, but are also of direct practical importance for conservation.

For most species, genetic data from across their distribution range are lacking, and extensive sampling efforts are needed to fill these gaps. However, efforts to create range-wide molecular reference databases are often resource intensive and challenging, especially for species that have large distributions, are rare or elusive, or occur in areas that are difficult to access. One potential solution to these difficulties is to use archival samples from museum collections to fill sampling gaps (Martins et al., 2018; Patel et al., 2017; Wilting et al., 2015; Yao et al., 2017, 2020). However, there are disadvantages to relying on museum samples, including the fact that museum collections often reflect opportunistic sampling by naturalists in the past, and lack precise locality information (Wehi et al., 2012). Furthermore, archival samples only reflect the species' genetic information from the time when samples were collected and may therefore not be currently representative of a species' genepool. Genetic changes within populations often occur through genetic drift, and these changes may increase if populations have undergone recent demographic changes through decline or expansion (Frankham et al., 2002). Many species that are of interest to conservationists are also not well represented in archival collections, and the few available samples rarely cover the species' entire range. In summary, archival collections may often only be of limited help in understanding population genetics for species across large geographic areas.

Analysis of environmental DNA (eDNA) provides a unique opportunity to avoid the aforementioned "sampling gap problem," because such samples can often be collected non-invasively across large areas (Deiner et al., 2017). Specifically, a subdiscipline of eDNA-based research, known as invertebrate-derived DNA (iDNA) research, targets the genetic material of prey or host species extracted from copro-, sarco-, or hematophagous invertebrates. Both eDNA and iDNA have been used in biodiversity studies more frequently in recent years, especially for terrestrial vertebrates that are otherwise difficult to sample (Calvignac-Spencer et al., 2013; Leempoel et al., 2020; Sales et al., 2020; Schnell et al., 2012, 2015). The most prominent method applied in such e/iDNA studies is the so-called metabarcoding approach, which can be designed to detect specific mammalian species, genera, or families (Schnell et al., 2012; Wilcox et al., 2016), or even entire mammalian communities (Ji et al., 2020; Weiskopf et al., 2018). However, the drawback of such metabarcoding approaches is that retrievable genetic information is limited to the short barcode sequence used for taxon assignment.

However, examples for the use of e/iDNA samples to infer the population genetic structure of a species are less common. Recent examples include the use of aquatic eDNA to study the genetic structure of fish (Weitemier et al., 2021), whales (Székely et al., 2021), and the use of iDNA obtained from copepods to study genetic diversity of whale sharks (Meekan et al., 2017). In terrestrial ecosystems, host DNA obtained from hematophagous leeches has been used to infer the population genetic structure of the rare and highly threatened Annamite striped rabbit *Nesolagus timminsi* (Nguyen et al., 2021). Although such e/iDNA studies are still uncommon, they have already shown great potential to become a widely used approach for the collection of genetic information in an effective and non-invasive manner across large areas.

In this study, we evaluated whether and how an iDNA biodiversity dataset derived from hematophagous leeches collected at multiple sites can be used to fill sampling gaps for a suite of terrestrial mammals. The leech samples were collected in the Annamite Mountains of Viet Nam and Laos. The Annamites ecoregion is particularly well suited for this study because (1) sampling of terrestrial hematophagous leeches as mammalian host DNA collectors is easy in the wet evergreen tropical forests of the region (Nguyen et al., 2021; Schnell et al., 2012), (2) samples from the Annamites are much rarer in international museum collections than samples from other regions in Southeast Asia (e.g., Sunda Shelf), (3) the Annamites contain several of the world's least known, most threatened, and evolutionarily unique mammal species, several of which were only recently discovered by science (e.g., saola *Pseudoryx nghetinhensis*, large-antlered muntjac *Muntiacus vuquangensis*, and Annamite striped rabbit *N. timminsi*; Sterling & Hurley, 2005), with evidence that some mammal populations found more widely may be genetically distinct from other populations in Southeast Asia, and (4) the Annamites are heavily impacted by the Southeast Asian snaring crisis (Belecky & Gray, 2020; Gray et al., 2018; Krishnasamy & Zavagli, 2020; Tilker et al., 2019), as evidenced by the fact that Annamite animals and their parts are often confiscated by wildlife protection authorities. Thus, focusing on the Annamite ecoregion will not only help to fill sampling gaps in species' distributions, but will also help contribute to the establishment of a geographically referenced molecular dataset for mammals in the Annamites region. Because such a reference database is currently missing, the exact origin of confiscated animals can often not be determined (Blair et al., 2017), which impedes investigative work on illegal wildlife trade routes and potentially even convictions. Here, we explored the suitability of iDNA as a way to address the "sampling gap problem" for six Annamites taxa (three carnivores and three ungulates), all of which are regularly traded (Kasper et al., 2020; Nguyen & Dinh, 2020). All six species occur across large parts of Southeast Asia or Indochina, and earlier genetic studies on those species have either failed to include samples from the Annamites or have done so to a minimal extent.

2 | MATERIAL AND METHODS

2.1 | Sample collection

We collected leeches during several field campaigns in the Annamite mountain range concurrent with systematic camera-trapping studies carried out between 2014 and 2019 (Nguyen et al., 2021; Tilker et al., 2020). In Viet Nam, surveys were conducted in Pu Mat National Park (PM used as abbreviation in the tables and figures), Phong Dien Nature Reserve (PD), Bach Ma National Park (BM), Hue Saola and Quang Nam Saola Nature Reserve (SNR), and Song Thanh Nature Reserve (ST). In Laos, we surveyed the eastern part of Xe Sap National Protected Area and adjacent forests near Ban Palé Village (XS) (Figure 1). Leeches were collected as bulk samples (i.e., several leeches were pooled per sample) around camera-trap stations. Each leech bulk sample was preserved on-site in nucleic acid preservation buffer (Camacho-Sanchez et al., 2013) and stored at −20°C after the end of each field campaign. We separated two types of leeches based on their phenotypic appearance, brown and tiger leeches (for detailed information on the collection procedure see Nguyen et al., 2021), because they are assumed to occupy different ecological niches and may have differences in vertebrate host preferences (Drinkwater et al., 2019).

2.2 | Laboratory work and bioinformatic analysis

2.2.1 | Species identification via metabarcoding

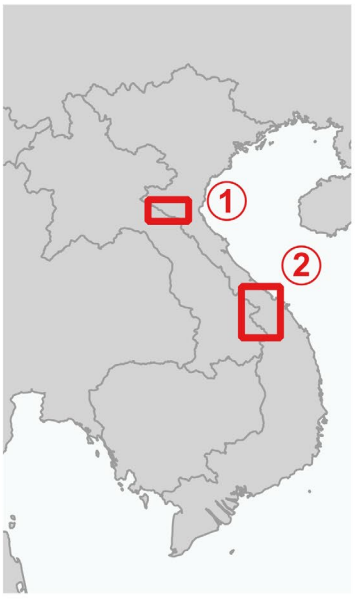
Each leech bulk sample was first screened for vertebrate host DNA using a metabarcoding approach, which included replicates and controls, following the general procedure outlined in Axtner et al. (2019), albeit with slight modifications. Instead of using all three suggested mitochondrial markers (Axtner et al., 2019), we performed six initial independent PCR screenings with the shortest marker only (16S rRNA/*mt-Rnr2*, 93 bp) and used the two longer markers (12S rRNA/*mt-RNR1*, 389 bp, and cytochrome b/mt-CYB, 302 bp) only to resolve taxonomic uncertainties (Tilker et al., 2020). PCR products were paired-end sequenced using Illumina's MiSeq Reagent Kit V3 with 2 × 300 cycles (Nguyen et al., 2021). We assigned the sequencing reads using PROTAX (Somervuo et al., 2016), accepting an assignment only when it was found in at least two independent PCRs (Axtner et al., 2019). We followed the bioinformatical workflow of Axtner et al. (2019) to analyze all the metabarcoding data.

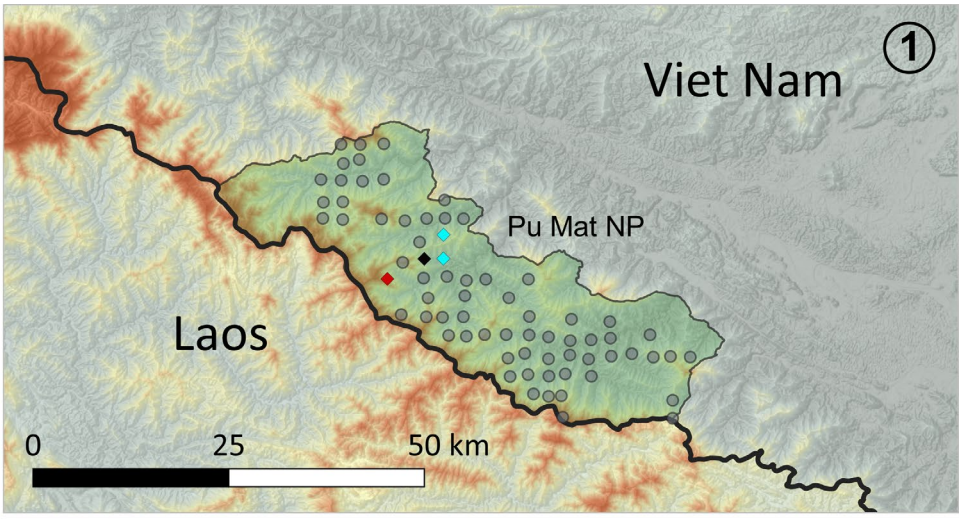
2.2.2 | Targeted PCRs for reference datasets

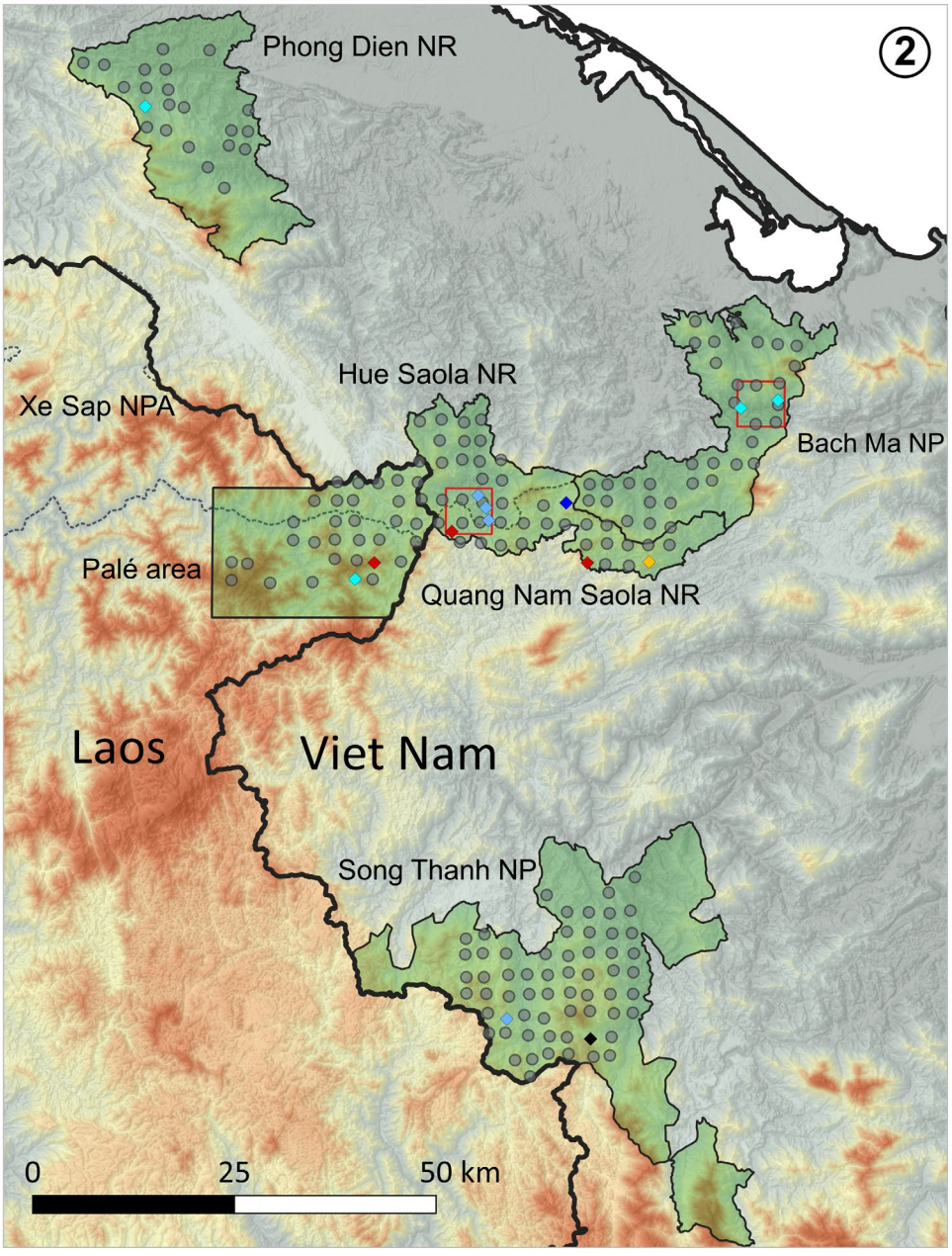
In the second step, we identified all leech bulk samples that contained one of six focal taxa (Table S1). For this study, we selected six taxa that were among the most frequently detected species in our metabarcoding approach and for which reference datasets using fragments of mitochondrial DNA were available. To be analyzed further, leech bulk samples containing mammalian iDNA had to meet the following criteria: (1) the bulk sample contained barcode sequences of few (ideally just one) mammal species to avoid amplification of non-target species and to avoid formation of chimera sequences (see Table S5), and (2) the bulk sample contained a high amount of target species iDNA as indicated by a high number of respective reads in the metabarcode sequencing. If more than eight bulks per taxon met these criteria, we selected eight bulks to facilitate downstream analyses from as many of the six sampling areas as possible. The iDNA of the selected bulks was then submitted to subsequent PCRs and Sanger sequencing to obtain sequenced fragments for the taxa of interest. For each focal species, we tested several primers (Table S2) to amplify a fragment of the cytochrome b gene (*mt-Cytb*) and the mitochondrial control region (*mt-Dcr*). To improve the amplification success rate due to the potential degradation of iDNA, marker fragments were designed to be short, ranging from 178 to 425 bp (Table S3). Fragments were amplified in 20 µl PCR volumes including 0.05 M of each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1× PCR buffer, 0.5 U AmpliTaq Gold™ (Invitrogen), and 2 µl of total DNA (range of total DNA concentration 1–10 ng/µl) extracted from the leech bulk sample. Cycling conditions were 5 min at 95°C for denaturation, followed by 35 cycles of 30 s at 95°C, primer-specific annealing conditions, and 45 s at 72°C. Amplification ended with a final step for 5 min at 72°C. All PCR setups included a no-template control. We used the ExoFastAP purification kit (Thermo Scientific) to purify PCR products from excess reagents and verified amplification visually by gel electrophoresis on 1.5% agarose gels. Sanger sequencing was carried out bidirectionally using the BigDye Terminator Cycle Sequencing Kit v.1.1 (Thermo Scientific) following the manufacturer's instruction. We used the BigDye XTerminator Purification Kit (Thermo Scientific) to purify our sequencing products. Terminated fragments were then separated on an ABI A3130xl automated genetic analyzer using the base-calling software SEQUENCE v.5.2 (both Applied Biosystems).

2.2.3 | Phylogenetic analyses

We used GENEIOUS v.8.0.5 (Biomatters Ltd.) to edit and align Sanger sequencing reads. PCR primer sequences were clipped, and sequencing reads were visually edited. For each focal species, we then mapped sequencing reads to a corresponding reference genome downloaded from NCBI GenBank (Table S4), also using GENEIOUS. Depending on availability, we downloaded additional sequences of the same species, the same genus or the same family to which we aligned the iDNA sequences (Table S4). We employed MEGA v.10.1.8 (Kumar et al., 2018) to analyze the genetic distance between the iDNA-retrieved sequences and the other sequences in the alignments. We build neighbor-joining (Saitou & Nei, 1987) phylogenies in MEGA to determine the position of our sample sequences in a tree together with the available reference sequences employing the Kimura 2-parameter substitution model. Due to gaps in the alignment and only partial overlaps of some sequences, we applied the







International border

Study area

Protected area border

0.5 - 1 km sampling

2.5 km sampling

Melogale spp.

Paguma larvata

Urva urva

Muntiacus vaginalis

Rusa unicolor

Sus scrofa

Elevation (m)

2,000

0

FIGURE 1 Location of the six study areas in the Annamites of Viet Nam and Laos. Gray circles are locations in which leech bulk samples were collected and colored-diamonds indicate the locations of leech bulk samples that amplified in this study for each of the six species. Note: Xe Sap National Protected Area and Palé area, as well as the Hue and Quang Nam Saola Nature Reserves, were treated each as one study area, as the leech sampling was conducted at the same time

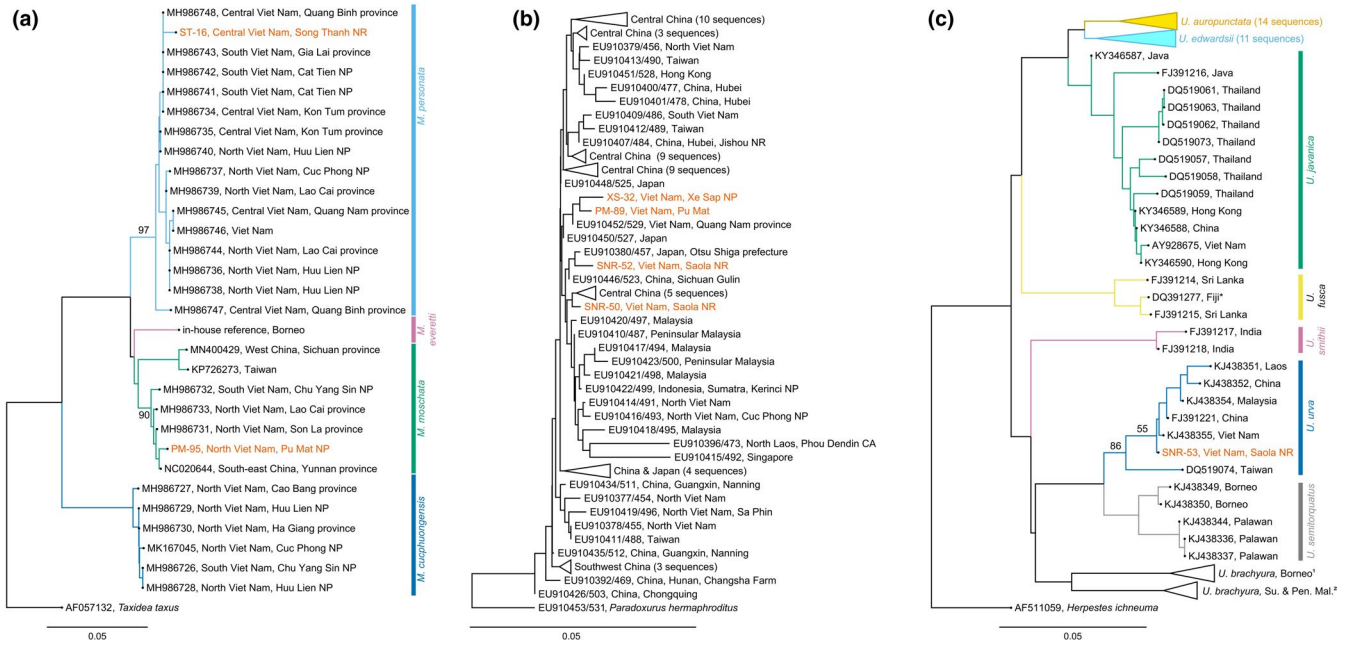


FIGURE 2 Neighbor-joining trees of carnivores, displaying (a) Ferret badgers *Melogale* spp. (609 bp mt-Cytb) (b) masked palm civet *Paguma larvata* (concatenated 173 bp *mt-Cytb* and 258 bp *mt-Dcr*) and (c) Crab-eating mongoose *Urva urva* (293 bp mt-Cytb) sequences from leech bulk samples (Table 1) in relation to reference sequences from GenBank (see Table S4). *Taxidea taxus* (a), *Paradoxurus hermaphroditus* (b), and *Herpestes ichneumon* (c) were used as outgroup. Leech bulk samples from this study are highlighted in dark-orange and branches from other regions were partly collapsed for better visualization. For better visualization, we only show bootstrap supports >50 for clades in which leech bulk samples were included. [1]Included 7 sequences; [2]included 14 sequences from Sumatra and Peninsular Malaysia

"pairwise deletions" parameter and 10,000 bootstrap replicates to assess node support. This analysis was performed to show the suitability of iDNA for this type of approach, and not to run in-depth phylogenetic analyses. Therefore, we only show the bootstrap supports above 50 for clades which contained one or more of the leech bulk samples from this study (Figures 2 and 3).

3 | RESULTS

Across the six sampling sites, we collected 4710 leeches (4403 brown leeches and 307 tiger leeches). The highest number of leeches per site was collected at Hue and Quang Nam Saola Nature Reserve (n = 2051; 1907 brown leeches, 144 tiger leeches), while the lowest number of leeches was collected at Phong Dien Nature Reserve (n = 54, all brown leeches). The 4710 leeches were pooled during the collection in 590 bulk samples (503 brown leech bulks, 87 tiger leech bulks). The mean number of leeches per bulk varied between the protected areas from 2.4 to 18.5 for brown leeches (total median = 5, mean ± SD: 8.5 ± 5.3) and from 1.8 to 4.4 for tiger leeches (total median = 3, mean ± SD: 2.7 ± 1.6 (Table 1). Applying the 16S rDNA metabarcoding approach, we obtained a total of 361 mammal assignments across all sampling sites representing 37 species, whereby iDNA extracted from brown leeches yielded 301 assignments from 34 mammal species and iDNA extracted from tiger leeches led to 60 mammal metabarcoding assignments from 23 species. The highest sampling efficiencies (lowest number of leeches per individual detection and lowest number of leeches per species detection) were observed at Phong Dien Nature Reserve (the site with least leeches collected) with 5.4 leeches per individual detection and 13.5 leeches per species detection. The lowest sampling efficiency for a mammal detection was observed at Pu Mat National Park with approx. 23 leeches per individual detection, and for species detection at Hue and Quang Nam Saola Nature Reserve (the two sites with most leeches collected) with 85.5 leeches needed per species detection (Table 1). Out of the 590 leech bulks, 218 (36.9%) contained iDNA of the six focal taxa (190 brown leech bulks, 28 tiger leech bulks). Of the focal taxa, Eurasian wild pig (*Sus scrofa*) iDNA was found most often, both in total (in 114 leech bulks) and at each of the six sampling sites. The focal species detected in the least number of bulks was the crab-eating mongoose (*Urva urva*), detected in 11 bulks at three sampling sites.

Sequence analyses for each of the six focal species are outlined below.

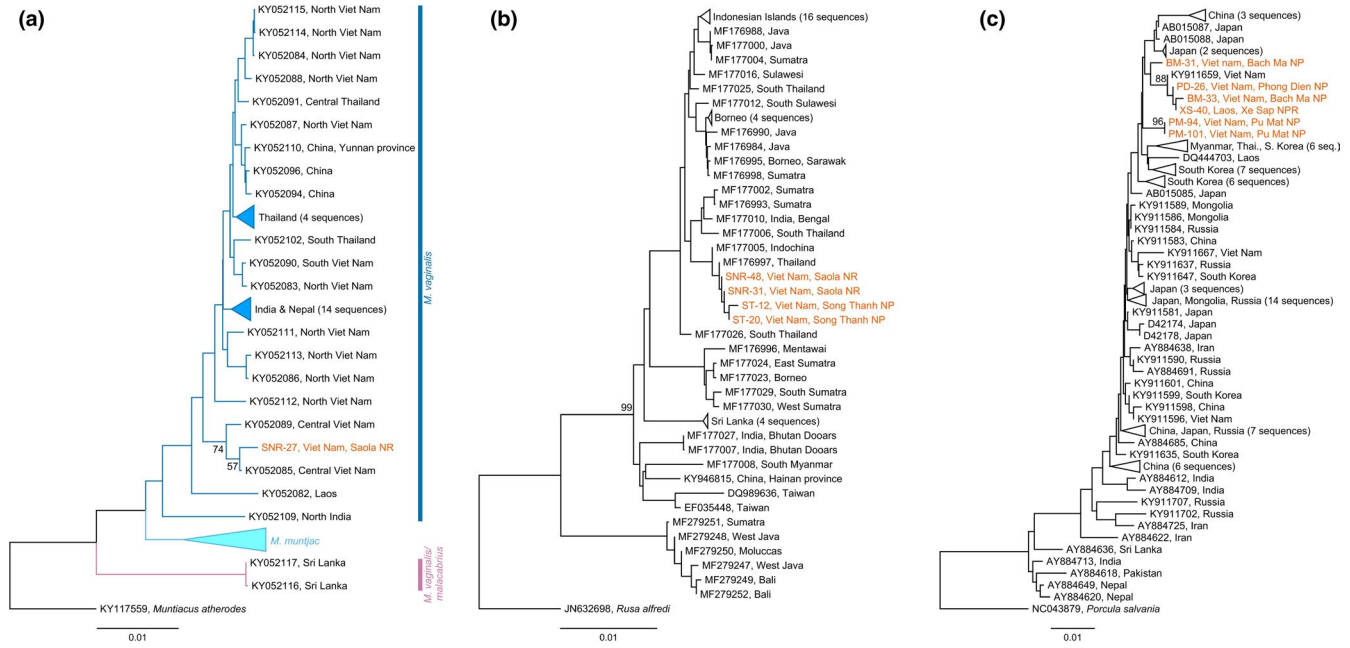


FIGURE 3 Neighbor-joining trees of ungulates, displaying (a) Red muntjacs *Muntiacus* spp. (full mitogenomes 16,412 bp for the reference samples and 424 bp *mt-Cytb* for the sample from this study) (b) Sambar *Rusa* (full mitogenomes 16,546 bp for the reference samples and up to 223 bp *mt-Cytb* for the samples from this study) and (c) Eurasian wild pig *Sus scrofa* (694bp *mt-Dcr*) sequences from leech bulk samples (Table 1) in relation to reference sequences from GenBank (see Table S4). *Muntiacus atherodes* (a), *Rusa alfredi* (b) and *Porcula salvania* (c) were used as outgroup. Leech bulk samples from this study are highlighted in dark-orange and branches from other regions were partly collapsed for better visualization. For better visualization, we only show bootstrap supports >50 for clades in which leech bulk samples were included

3.1 | Ferret badgers (*Melogale* spp.)

Based on the 16S rDNA metabarcoding assignments, ferret badger sequences were detected in 47 leech bulk iDNA samples collected in all six study areas (Table 1). As most mtDNA GenBank entries for *Melogale* were *mt-Cytb* sequences, we only attempted to amplify a *Melogale* ssp. *mt-Cytb* fragment from the corresponding leech bulks. Only two of the selected eight leech bulk samples yielded *mt-Cytb* sequences, one with 609 bp and the other one with 494 bp length (Table 2). We included all 30 available *Melogale mt-Cytb* sequences from GenBank (Li et al., 2019; Rozhnov et al., 2019; Zhou et al., 2020) as well as an in-house generated sequence of the Bornean ferret badger (*Melogale everetti*) in the alignment. These sequences covered four of the five known species of this genus. No genetic data were available for the fifth species, the Javan ferret badger (*Melogale orientalis*). However, all three species known to occur in Viet Nam (the Viet Nam ferret badger (*Melogale cucphongensis*), the small-toothed ferret badger (*Melogale moschata*), and the large-toothed ferret badger (*Melogale personata*) were included in our analyses (Table S4). The American badger (*Taxidea taxus*) was used as a taxonomic outgroup in the phylogenetic analyses. The NJ-tree reveals four clades that match the existing species (Figure 2). One of the leech bulk samples (ST-16; Figure 2a, upper clade) grouped together with the large-toothed ferret badger sequences, while the *Melogale* sequence obtained from bulk leech sample PM-95 resided within the clade of the small-toothed ferret badger (Figure 2a, second from bottom clade).

3.2 | Masked palm civet (*Paguma larvata*)

We detected masked palm civet iDNA in 17 leech bulk samples from five of the six study areas (all except Bach Ma National Park, although camera-trapping data showed that the species occurs in the park) (Table 1). Of the selected eight leech bulk samples, we could amplify short fragments of the *mt-Cytb* and *mt-Dcr* (Table 2) from four bulks coming from three different study areas (PM, SNR, and XS). For the NJ-tree reconstruction, we combined our iDNA sequences with reference sequences from a larger study on the genetic diversity of masked palm civets (Patou et al., 2009). The common palm civet (*Paradoxurus hermaphroditus*) was used as outgroup. Two of the four iDNA sequences (XS-32 and PM-89) grouped together with a GenBank reference sequence from the Central Annamites (Quang Nam province), while the other two grouped either with a sequence originating from Japan (SNR-52) or with sequences originating from China (SNR-50; Figure 2b).

3.3 | Crab-eating mongoose (*Urva urva*)

Although metabarcoding results had indicated the presence of crab-eating mongoose iDNA in 11 leech bulk samples, iDNA degradation had already progressed to the point that we were only able to amplify a short *mt-Cytb* fragment from one of the eight selected bulk samples. For the comparison, we used *mt-Cytb* sequences in a

neighbor-joining tree (Figure 2C) from other Asian mongoose species (Patou et al., 2009; Veron & Jennings, 2017; Veron et al., 2015). The Egyptian mongoose served as outgroup. The NJ-tree showed clear separation among the numerous Asian mongoose species. The iDNA *mt-Cytb* sequence from one leech bulk sample (SNR-53; Figure 2C, lower third) was placed within the *U. urva* clade, together with sequences from specimens from Viet Nam, Laos, China, Taiwan, and Malaysia.

3.4 | Northern red muntjac (*Muntiacus vaginalis*)

We detected northern red muntjac iDNA in 15 of our brown leech bulk samples from only two study areas (SNR and ST, although camera-trapping recorded the species in all six study areas). However, due to iDNA degeneration, only one of the primer systems generated a short 424 bp *mt-Cytb* fragment, and only for one out of the eight selected samples (SNR-27). As references for the NJ-tree reconstruction, we used aligned published mitogenome sequences from red muntjacs (*M. vaginalis* and *M. muntjac*) with known origins (Martins et al., 2017; Singh et al., 2019) to which we also aligned our leech-derived sequence. We used the Bornean yellow muntjac (*M. atherodes*) for the outgroup. The resulting NJ-tree (Figure 3a) shows the leech-derived sequence SNR-27 to cluster with other references from central Viet Nam.

3.5 | Sambar (*Rusa unicolor*)

Species assignments via metabarcoding analysis indicated the presence of sambar iDNA in 14 brown leech bulks coming from two of the six sampling sites (SNR and ST, the same sites with muntjac presence; Table 1). Here, iDNA degradation had not progressed as far as for iDNA samples from muntjac and crab-eating mongoose, and we were able to amplify short *mt-Cytb* fragments from four of the eight selected leech bulk samples. Of those four positive samples, one yielded a 161 bp long sequence (ST-12) and the other three a 223 bp long sequence (ST-20, SNR-31, and SNR-48; Table2). The sambar iDNA *mt-Cytb* sequences were aligned with GenBank mitogenome sequences of *R. unicolor* and *R. timorensis* (Javan deer) with known origins, whereby *R. timorensis* sequences were included because a recent study had revealed past hybridization events between both species (Martins et al., 2018). We used the Philippine spotted deer (*Rusa alfredi*) as outgroup. The NJ-tree displayed several clades that were not clearly geographically structured. The four sambar iDNA *mt-Cytb* sequences clustered close to each other within a clade that included predominantly Indochinese *R. unicolor* (Figure 3b, upper clade).

3.6 | Eurasian wild pig (*Sus scrofa*)

The species we detected most often in the 16S rDNA metabarcoding analysis, both in absolute numbers and in terms of bulks per sampling site, was the Eurasian wild pig (*S. scrofa*). The species was detected across all six sampling sites in 114 out of the 590 bulk samples (19.3%). Because the latest taxonomic revision for *S. scrofa* was based on sequences of the *mt-Dcr* (Choi et al., 2020), we amplified a fragment of this locus. We were able to amplify fragments from six *S. scrofa* iDNA containing bulks with lengths that ranged from 637bp ($n = 1$) and over 641bp ($n = 4$) to 694bp ($n = 1$). Of the 327 Eurasian wild pigs analyzed from Asia and Eastern Europe (Choi et al., 2020), we selected representative haplotypes for each region, summing to 118 sequences. We also added the published (Robins et al., 2006) *S. scrofa* sequence from a skull formerly suspected to have come from Heude's pig *Sus bucculentus* (Groves et al., 1997) as a representative from Laos. The pygmy hog (*Porcula salvania*) served as taxonomic outgroup, as it is the basal genus of the Suinae subfamily. The NJ-tree, which does not show a geographic structuring of the sequences, positioned all six iDNA *S. scrofa mt-Dcr* sequences together with other Eurasian wild pig samples from Viet Nam, Korea, China, and other Indochinese samples (Figure 3C, upper part).

4 | DISCUSSION

4.1 | e/iDNA as sources to fill geographic sampling gaps

In this study, we used iDNA extracted from hematophagous leeches to evaluate its potential for obtaining genetic information of mammalian host DNA to fill sampling gaps across large geographic areas. Recent e/iDNA studies have largely focused on the detection of single rare or invasive species (Barata et al., 2021; Dougherty et al., 2016) or been used to assess general biodiversity patterns (Calvignac-Spencer et al., 2013; Gogarten et al., 2020; Leempoel et al., 2020; Sales et al., 2020; Tilker et al., 2020; Weiskopf et al., 2018). Our results suggest that non-invasively obtained genetic material from e/iDNA can be used both for more in-depth studies on within-species genetic diversity (Nguyen et al., 2021) and for obtaining genetic information from populations that would otherwise be difficult to sample. Because invertebrates are often abundant and can be collected in large numbers, iDNA-based approaches have the potential to provide genetic material from species that are especially rare or elusive (Schnell et al., 2015). The advantage to first assessing mammalian diversity using metabarcoding approaches is that downstream selection of bulk samples for the analyses of the genetic diversity of the target species can be based on these results. However, our results also show that detectability of a species is largely influenced by the sampling effort, as some of our target species were not recorded in sites with low sampling effort (even though camera-trapping proved their occurrence in the protected areas). To improve e/iDNA detection rates for specific focal species, it may be helpful to first conduct pilot studies to identify areas of presence of the species of interest, ideally employing conventional methods such as camera-trapping; such preliminary studies may be particularly important for studies on rare species or those with

TABLE 1 Overview of the collected samples per study area and the results of the metabarcoding

	BM			PD			PM		
	tot.	B	T	tot.	B	T	tot.	B	T
No. collected leech bulk samples	129	118	11	22	22	0	82	70	12
No. of collected leeches	576	548	28	54	54	0	436	387	49
No. of recorded mammal species[a]	12	9	5	4	4	0	8	8	4
No. of mammal assignments[a]	38	33	5	10	10	0	19	14	5
Melogale spp.									
Positive bulk samples	3	3	0	1	1	0	6	5	1
Selected bulk samples	0			0			1	1	0
Paguma larvata									
Positive bulk samples	0			2	2	0	5	4	1
Selected bulk samples				1	1	0	1	0	1
Urva urva									
Positive bulk samples	0			0			0		
Selected bulk samples									
Muntiacus vaginalis									
Positive bulk samples	0			0			0		
Selected bulk samples									
Rusa unicolor									
Positive bulk samples	0			0			0		
Selected bulk samples									
Sus scrofa									
Positive bulk samples	14	13	1	4	4	0	8	7	1
Selected bulk samples	0			2	2	0	2	2	0

Abbreviations: B, brown leeches; BM, Bach Ma National Park; PD-, Phong Dien NR; PM, Pu Mat National Park; SNR, Saola Nature Reserves; ST, Song Thanh NP; T, tiger leeches; tot, total; XS, Xe Sap National Protected Area/Pale area.

[a]Using metabarcoding.

specific habitat requirements. Follow-up surveys can then employ intensive collection of e/iDNA to maximize detections and therefore the chance to obtain molecular data for the target species.

Regarding the invertebrate vector, our sampling was biased toward leech morphotype, since we collected approximately five times more bulk samples of brown leeches than of tiger leeches. Moreover, bulk samples from brown leeches contained more than twice as many leeches (median: brown = 5; tiger = 2). Interestingly, we did not detect any deer species (i.e., sambar or muntjac) in tiger leech samples. We believe the most likely explanation for this finding is that we had many more brown than tiger leech bulk samples, and not a consequence of morphotype feeding preferences; a recent study from Malaysian Borneo showed that deer and pigs were detected more frequently in tiger leech bulk samples than in brown leech bulks (Abrams et al., 2019). However, we also cannot exclude regional differences in host species preferences among leeches. Other studies have also found differences in iDNA detection rates and detected species between both leech types (Drinkwater et al., 2020). The question of leech morphotype feeding preferences is of relevance for biodiversity estimates or species distribution modelling, as varying detection probabilities can influence the results. However, for studies that simply aim to obtain genetic material of a target species to fill a sampling gap, we do not see the need to distinguish the two morphotypes during collection. This, in turn, would make field logistics and sampling even easier.

The molecular data obtained by e/iDNA sampling might be particularly valuable for efforts to counteract the global illegal wildlife trade. Depending on the source of DNA, a major advantage of e/iDNA data can be that the samples have precise location data, which is a prerequisite to establish a spatially explicit reference dataset to be used in wildlife DNA forensics (Wasser et al., 2008). This is at least true for some DNA sources such as soil, smaller water bodies, or relatively immobile invertebrates such as ticks or leeches. Because exact locality information is not known for most archival specimens (Boessenkool et al., 2010), and not available for samples that originated from a confiscated specimen, large reference databases with sequences from different populations of different geographic areas could be used by enforcement agencies to determine the source populations of traded wildlife, thus helping to identify wildlife trade routes and to establish the evidence base needed for follow-up legal actions. Furthermore, specific locality information for confiscated wildlife could be used to direct law enforcement to

SNR			ST			XS			Total		
tot.	B	T	tot.	B	T	tot.	B	T	tot.	B	T
206	165	41	90	84	6	61	44	17	590	503	87
2051	1907	144	704	693	11	889	814	75	4710	4403	307
24	19	15	16	16	1	24	14	10	37	34	23
155	122	33	59	57	2	80	65	15	361	301	60
20	16	4	4	4	0	13	9	4	47	38	9
2	1	1	3	3	0	2	1	1	8	6	2
5	4	1	1	1	0	4	4	0	17	16	1
3	3	0	1	1	0	2	2	0	8	7	1
8	5	3	1	1	0	2	1	1	11	7	4
5	3	2	1	1	0	2	1	1	8	5	3
10	10	0	5	5	0	0			15	15	0
6	6	0	2	2	0				8	8	0
9	9	0	5	5	0	0			14	14	0
4	4	0	4	4	0				8	8	0
48	41	7	26	24	2	14	11	3	114	100	14
2	2	0	1	1	0	1	1	0	8	8	0

poaching hotspots, thus potentially preventing further overexploitation of susceptible populations (Williams et al., 2020). Such genetic databases could also be used to inform the release of confiscated animals back into their population of origin, thus avoiding genetic outbreeding and introgression (Banes et al., 2016; Oklander et al., 2020), and in general helping to align release efforts with globally recommended IUCN guidelines (Maddison, 2019). Unfortunately, confiscated wildlife of unknown geographic origin is commonplace, especially in countries where wildlife trade in prevalent. For example, in April 2021, "Save Vietnam's Wildlife" confiscated 100 traded masked palm civets (Save Vietnam's Wildlife pers. comm.). It is unknown whether these animals originated from civet farms or were taken from the wild, but a reference database with local resolution for this species could support future release efforts.

Despite the advantages of obtaining genetic material via iDNA collection, our study also highlighted several challenges with the use of e/iDNA to fill sampling gaps.

1. Ingested DNA will be at various stages of degradation (see also Nguyen et al., 2021). If already highly degraded, iDNA samples will fail to amplify in PCRs. Of the 48 leech bulk samples which were selected in this study to be submitted for fragment-specific PCRs, less than half (20) yielded amplicons for at least parts of the target fragments (Table 2). Certainly, such a high failure rate limits the applicability of e/iDNA for exceptionally rare species, at least at reasonable costs. Here, we note that despite the large sampling effort in this study, only few detections of rare species such as the marbled cat (*Pardofelis marmorata*) and Owston's civet (*Chrotogale owstonii*) were obtained (see Tilker et al., 2020). In addition, due to DNA degradation, the amplified fragments might be too short to provide the needed taxonomic resolution. At the same time, this degradation might only pose a significant challenge for the rather conventional PCR-based methods applied in this project. Recent progress in ancient DNA studies show that with modern DNA enrichment techniques, it is possible to retrieve full mitogenomes from archival samples (Martins et al., 2017; Paijmans et al., 2020; Patel et al., 2016) and even full genomes from samples that are more than one million years old and thus highly degraded (van der Valk et al., 2021). We expect that such modern DNA enrichment techniques will become routine in the upcoming years. Coupled with the

TABLE 2 Leech samples that were amplified in the targeted PCRs and the number of sequenced basepairs (bp) for the respective mitochondrial markers

		Marker			
Bulk ID	Study area	*mt-Cytb* (bp)	*mt-Dcr* (bp)	Total length (bp)	GenBank accession numbers
Melogale					
PM-95	Pu Mat NP	609	—	609	MZ234298
ST-16	Song Thanh NP	494	—	494	MZ234299
Paguma larvata					
XS-32	Xe Sap NPA	173[a]	258[b]	431	MZ265766
PM-89	Pu Mat NP	173[a]	258[b]	431	MZ265765
SNR-50	Saola NR	173[a]	248[b]	421	MZ265764
SNR-52	Saola NR	173[a]	258[b]	431	MZ265767
Urva urva					
SNR-53	Saola NR	293	—	293	MZ234300
Muntiacus vaginalis					
SNR-27	Saola NR	424	—	424	MZ265768
Rusa unicolor					
ST-20	Song Thanh NP	223	—	223	MZ265769
ST-12	Song Thanh NP	161	—	161	MZ265772
SNR-31	Saola NR	223	—	223	MZ265771
SNR-48	Saola NR	223	—	223	MZ265770
Sus scrofa					
XS-40	Xe Sap NPA	—	641	641	MZ273045
PD-26	Phong Dien NR	—	694	694	MZ273046
PM-94	Pu Mat NP	—	641	641	MZ273047
PM-101	Pu Mat NP	—	641	641	MZ273048
BM-31	Bach Ma NP	—	641	641	MZ273049
BM-33	Bach Ma NP	—	637	637	MZ273050

Note: Study area and GenBank accession numbers of the new sequences are also shown.

[a]145 bp *mt-Cytb* followed by 28 bp *tRNA-Thr*.

[b]26 bp *tRNA-Pro* followed by 234 bp in *mt-Dcr*.

continuing decline of sequencing costs (Wetterstrand, 2021), we hope that such methods will also become available for e/iDNA studies. Recently, DNA enrichment using hybridization-capture approaches (Paijmans et al., 2016) have already been used successfully to assess the mammalian diversity in water samples (Seeber et al., 2019) and to screen leech samples for novel mammalian viruses (Alfano et al., 2021).

2. Sequences from e/iDNA cannot be linked to a real specimen at hand, so that no additional phenotypic data are available. The iDNA sequences from our six target species all grouped together with sequences that originated from Indochina, and no sequence was representative for a clearly distinct Annamite clade. Because the Annamite Mountains are known for their high rate of endemism (Baltzer et al., 2001), this result was unexpected, but avoided the problem of iDNA sequences not matching GenBank reference sequences. It is important to note that novel e/iDNA sequences without reference sequence alignment do not allow the possibility of drawing taxonomic conclusions. If similar sequences are not available in databases, such e/iDNA sequences would be need to be treated as an operational taxonomic unit until further analyses are conducted, ideally with genetic data from a real specimen for comparison. Therefore, e/iDNA studies cannot serve as a replacement for obtaining both DNA and morphological data from actual specimens. We do, however, think that e/iDNA studies can be used to identify unique populations—for example, if e/iDNA sequences do not match any existing reference population—as well as taxonomic uncertainties, which can then be prioritized for future in-depth phylogenetic and phylogeographic studies.

4.2 | Filling sampling gaps on the six target species

4.2.1 | Ferret badgers

Species assignment for ferret badgers (*Melogale* spp.) from camera-trap photographs or for animals found in trade is difficult due to the lack of external characters clearly differentiating the species (Robichaud, 2010; Schank et al., 2009). Species-level identification often requires a close inspection of skull morphology and dentition. Partly for this reason, little is known about the distribution of ferret badger species in Indochina (Schank et al., 2009). In 2005 and 2006, two animals were collected in northern Viet Nam and, based on their skull characteristics and a 423 bp long *mt-Cytb* fragment, they were described as a new species, the Viet Nam ferret badger *M. cucphuongensis* (Nadler et al., 2011). Subsequent sampling across Viet Nam has revealed that the new species appears to be relatively widespread and occurs sympatrically with both small-toothed (*M. moschata*) and large-toothed ferret badgers (*M. personata*) across large parts of the country (Rozhnov et al., 2019). In this study, we did not detect iDNA from *M. cucphuongensis*, but did detect iDNA from a small-toothed ferret badger in a leech bulk sample from the northern Annamites (PM-95; Pu Mat National Park, Figure 2a) and from a large-toothed ferret badger in a leech bulk sample from the central Annamites (ST-16; Song Thanh National Park, Figure 2a). Our simultaneous camera-trapping surveys showed that ferret badgers were among the most commonly recorded mammals in all six study areas (unpublished data). However, due to the lack of obvious external characteristics (Robichaud, 2010; Schank et al., 2009), ferret badgers captured by these camera-trap photographs could not be assigned to any of the three *Melogale* species occurring in the Annamites. Similarly, ferret badgers were among the most frequently recorded species in our metabarcoding approach, but the short 16S rDNA barcode had not been designed to distinguish between the small-toothed and large-toothed ferret badgers, while for the Viet Nam ferret badger a 16S rDNA reference sequence was missing. Therefore, a much broader sampling campaign—for example, of leech iDNA samples collected across the country, and supplemented by sequencing of species-diagnostic *mt-Cytb* regions—is needed to obtain an overview on the distribution and genetic diversity of the three ferret badger species in Viet Nam.

4.2.2 | Masked palm civet and crab-eating mongoose

For the masked palm civet (*Paguma larvata*), our results support earlier studies that found a low genetic diversity within the species compared with other Paradoxurinae species (Patou et al., 2009). Two of our iDNA samples clustered with a sequence originating from the Quang Nam province in central Vietnam, but the other two *P. larvata*-iDNA sequences grouped with sequences from Japan and China (Figure 2B). Such geographic assignment indicates a low level of genetic diversity within the species, likely due to a recent population expansion that has resulted in little geographic structuring. Nevertheless, the new *Paguma* iDNA sequences are the first step toward a more comprehensive dataset that can assist in determining the origin of masked palm civets confiscated by wildlife protection authorities in Viet Nam and elsewhere. The recent confiscation of 100 masked palm civets shows the urgency of such a reference database.

For the crab-eating mongoose (*U. urva*), our iDNA sequence from the Saola Nature Reserves (SNR-53; Figure 2c) clustered with all other samples of this species and was most closely related to a sample from Viet Nam. Only a much wider sampling of more populations can provide further insights on the genetic diversity within this species. As mongooses are regularly traded and found on wildlife markets, we hope that e/iDNA sampling and sequencing will contribute to a more complete Viet Nam wide reference database.

4.2.3 | Northern red muntjac and sambar

The sampling of Northern red muntjac (*Muntiacus vaginalis*) from Viet Nam was already quite intensive in an earlier study analyzing mitogenomes of red muntjacs from Southeast Asia, including the central Annamites (Martins et al., 2017). As expected, the single Northern red muntjac iDNA (SNR-27, Figure 3a) that amplified in this analysis clustered with that sample from the central Annamites (Martins et al., 2017).

Although sambar (*Rusa unicolor*) from the Annamites had not been included in the large phylogeographic study of *Rusa* (Martins et al., 2018), the four iDNA sequences obtained in this study formed one group and clustered with sequences originating from Indochina, India, and Sumatra (Figure 3B). Based on samples from both Indochina and the Sunda region, the aforementioned study showed the existence of several large *R. unicolor* mtDNA clades (Martins et al., 2018). However, as these clades showed no clear geographic structuring even when full mitogenomes were sequenced (Martins et al., 2018) (for example, one clade included samples from India, Thailand, and Sumatra), it might be difficult to determine the origin of traded sambar, at least with the mtDNA fragments used in this study.

4.2.4 | Eurasian wild pig

Our six iDNA sequences clustered closely with sequences from Southeast Asia, and mainly from Indochina (Viet Nam, Laos, Thailand, Myanmar, S. Korea; Figure 3c). We did not detect an Annamite-specific wild pig clade or species. Four iDNA samples from the central Annamites grouped with one GenBank reference sample from Viet Nam, while the two iDNA samples from Pu Mat National Park in northern Viet Nam seem somewhat distinct (Figure 3c). These results indicate some genetic differentiation between populations of the central and northern Annamites. Wild boars are often targeted by poachers, and their meat is widely found in wildlife markets

throughout Viet Nam (Drury, 2011; Sandalj et al., 2016). If subsequent studies substantiate the apparent genetic differentiation within Vietnamese populations mentioned above, a reference database could be established to determine the origin of Eurasian wild pig animals and animal products confiscated from the illegal wildlife trade. Such information would allow local authorities to identify areas that are under the greatest poaching threat and to direct law enforcement activities to these poaching hot spots.

5 | CONCLUSION

The use of e/iDNA has great value for conservation studies as a way to detect focal species. Here, we take the application of e/iDNA one step further by showing that this method can be used to obtain information over large areas that can fill geographic sampling gaps in genetic reference databases. We see enormous potential in the use of these genetic databases to inform conservation actions, especially with regards to DNA-based wildlife forensics. Two practical examples in which such a range-wide reference database could be useful include situations where it is necessary to determine the origin of confiscated animals and their parts to support law enforcement agencies, and to inform the release of confiscated animals back into their source populations. Although a relatively high sampling effort is needed to compensate for the degradation of e/iDNA, we believe that modern DNA enrichment techniques are likely to mitigate this obstacle. We see a promising future for the use of e/iDNA to fill sampling genomic gaps as a means to enhance conservation efforts.

ACKNOWLEDGMENTS

We thank the staff of the WWF-CarBi and Green Annamites projects for providing extensive logistical support. We thank the WWF Viet-Nam and WWF Laos for providing fieldwork and logistical assistant. We thank the Save Vietnam's Wildlife for providing a full-time working team leader and Vinh University for providing two team leaders. We also thank the management and the administration of Bach Ma NP, Hue Saola NR, Quang Nam Saola NR, Xe Sap NPA, Song Thanh NR, Phong Dien NR, and Pu Mat NP for providing permissions and personnel to conduct this research. A special thank goes to our field team leaders in Vietnam and Laos. The Central Institute for Natural Resources and Environmental Studies, Vietnam National University (VNU-CRES) supported the process to obtain Access and Benefit-sharing (ABS) certificates for the genetic samples. Funding for the surveys was provided by the German Federal Ministry of Education and Research (BMBF FKZ: 01LN1301A), Leibniz-IZW, Point Defiance Zoo and Aquarium, Safari Club International, Critical Ecosystem Partnership Fund, the National Geographic Society's Committee for Research and Exploration, Ocean Park Conservation Foundation Hong Kong, Mohamed bin Zayed Species Conservation Fund, and GreaterGood. Thanh V. Nguyen received financial support through a Russell E. Train Education for Nature program fellowship, the Leibniz-IZW, and the Vietnamese Ministry of Science and Technology Program 562. The CarBi project received support from the Internationale Klimaschutzinitiative (International Climate Protection Initiative) of the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMBU) and the Kreditanstalt für Wiederaufbau (KfW).

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interests.

AUTHOR CONTRIBUTIONS

AW conceptualized the idea. TVN, AN, and AT conducted fieldwork and collected data. ML, AHQN, and BMR assisted the fieldwork. TVN and AS conducted laboratory work. JA analyzed data. AW, JA, AT, and JF wrote the manuscript. All authors contributed revision of the article.

DATA AVAILABILITY STATEMENT

DNA sequences for Sanger sequencing have been deposited in GenBank (see Table 2 for accession numbers).

ORCID

Andreas Wilting https://orcid.org/0000-0001-5073-9186
Thanh V. Nguyen https://orcid.org/0000-0002-8057-7912
Jan Axtner https://orcid.org/0000-0003-1269-5586
An Nguyen https://orcid.org/0000-0003-0456-3866
Minh Le https://orcid.org/0000-0002-2953-2815
Benjamin M. Rawson https://orcid.org/0000-0003-4141-5985
Andrew Tilker https://orcid.org/0000-0003-3630-8691
Joerns Fickel https://orcid.org/0000-0002-0593-5820

REFERENCES

Abrams, J. F., Hörig, L. A., Brozovic, R., Axtner, J., Crampton-Platt, A., Mohamed, A., Wong, S. T., Sollmann, R., Yu, D. W., & Wilting, A. (2019). Shifting up a gear with iDNA: From mammal detection events to standardised surveys. *Journal of Applied Ecology*, *56*(7), 1637–1648. https://doi.org/10.1111/1365-2664.13411

Alfano, N., Dayaram, A., Axtner, J., Tsangaras, K., Kampmann, M.-L., Mohamed, A., Wong, S. T., Gilbert, M. T. P., Wilting, A., & Greenwood, A. D. (2021). Non-invasive surveys of mammalian viruses using environmental DNA. *Methods in Ecology and Evolution*, in press. https://doi.org/10.1111/2041-210X.13661

Axtner, J., Crampton-Platt, A., Hörig, L. A., Mohamed, A., Xu, C. C. Y., Yu, D. W., & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, *8*(4), giz029. https://doi.org/10.1093/gigascience/giz029

Baltzer, M. C., Nguyen, T. D., & Shore, R. G. (2001). *Towards a vision for biodiversity conservation in the forests of the lower Mekong ecoregion complex: Summary of the biological assessment for the Ecoregion Biodiversity Conservation Program in the forests of the lower Mekong ecoregion complex* (pp. 1–109). WWF INdochina, WWF US, WWF International.

Banes, G. L., Galdikas, B. M. F., & Vigilant, L. (2016). Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. *Scientific Reports*, *6*(1), 22026. https://doi.org/10.1038/srep22026

Barata, I. M., Griffiths, R. A., Fogell, D. J., & Buxton, A. S. (2021). Comparison of eDNA and visual surveys for rare and cryptic

bromeliad-dwelling frogs. *The Herpetological Journal*, *31*(1), 1–9. https://doi.org/10.33256/hj31.1.19

Belecky, M., & Gray, T. N. E. (2020). *Silence of the snares: Southeast Asia's snaring crisis* (pp. 1–88). WWF International. https://wwfeu.awsassets.panda.org/downloads/wwf_snaring_report_2020.pdf

Blair, M. E., Le, M. D., Sethi, G., Thach, H. M., Nguyen, V. T. H., Amato, G., Birchette, M., & Sterling, E. J. (2017). The importance of an interdisciplinary research approach to inform wildlife trade management in Southeast Asia. *BioScience*, *67*(11), 995–1003. https://doi.org/10.1093/biosci/bix113

Boessenkool, S., Star, B., Scofield, R. P., Seddon, P. J., & Waters, J. M. (2010). Lost in translation or deliberate falsification? Genetic analyses reveal erroneous museum data for historic penguin specimens. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 1057–1064. https://doi.org/10.1098/rspb.2009.1837

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., & Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, *22*(4), 915–924. https://doi.org/10.1111/mec.12183

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., & Leonard, J. A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, *13*(4), 663–673. https://doi.org/10.1111/1755-0998.12108

Choi, S. K., Kim, K. S., Ranyuk, M., Babaev, E., Voloshina, I., Bayarlkhagva, D., Chong, J.-R., Ishiguro, N., Yu, L. I., Min, M.-S., Lee, H., & Markov, N. (2020). Asia-wide phylogeography of wild boar (Sus scrofa) based on mitochondrial DNA and Y-chromosome: Revising the migration routes of wild boar in Asia. *PLoS One*, *15*(8), e0238049. https://doi.org/10.1371/journal.pone.0238049

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Dougherty, M. M., Larson, E. R., Renshaw, M. A., Gantz, C. A., Egan, S. P., Erickson, D. M., & Lodge, D. M. (2016). Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, *53*(3), 722–732. https://doi.org/10.1111/1365-2664.12621

Drinkwater, R., Jucker, T., Potter, J. H. T., Swinfield, T., Coomes, D. A., Slade, E. M., Gilbert, M. T. P., Lewis, O. T., Bernard, H., Struebig, M. J., Clare, E. L., & Rossiter, S. J. (2020). Leech blood-meal invertebrate-derived DNA reveals differences in Bornean mammal diversity across habitats. *Molecular Ecology*, *00*, 1–14. https://doi.org/10.1111/mec.15724

Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., Gilbert, M. T. P., & Rossiter, S. J. (2019). Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources*, *19*(1), 105–117. https://doi.org/10.1111/1755-0998.12943

Drury, R. (2011). Hungry for success: Urban consumer demand for wild animal products in Vietnam. *Conservation and Society*, *9*(3), 247–257. https://doi.org/10.4103/0972-4923.86995

Eaton, M. J., Meyers, G. L., Kolokotronis, S.-O., Leslie, M. S., Martin, A. P., & Amato, G. (2010). Barcoding bushmeat: Molecular identification of Central African and South American harvested vertebrates. *Conservation Genetics*, *11*(4), 1389–1404. https://doi.org/10.1007/s10592-009-9967-0

Frankham, R., Ballou, J., Briscoe, D., & McInnes, K. (2002). *Introduction to conservation genetics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808999

Gogarten, J. F., Hoffmann, C., Arandjelovic, M., Sachse, A., Merkel, K., Dieguez, P., Agbor, A., Angedakin, S., Brazzola, G., Jones, S., Langergraber, K. E., Lee, K., Marrocoli, S., Murai, M., Sommer, V., Kühl, H., Leendertz, F. H., & Calvignac-Spencer, S. (2020). Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, *2*(1), 63–76. https://doi.org/10.1002/edn3.46

Gray, T. N. E., Hughes, A. C., Laurance, W. F., Long, B., Lynam, A. J., O'Kelly, H., Ripple, W. J., Seng, T., Scotson, L., & Wilkinson, N. M. (2018). The wildlife snaring crisis: An insidious and pervasive threat to biodiversity in Southeast Asia. In *Biodiversity and conservation* (Vol. *27*). Springer Netherlands. https://doi.org/10.1007/s10531-017-1450-5

Groves, C. P., Schaller, G. B., Amato, G., & Khounboline, K. (1997). Rediscovery of the wild pig *Sus bucculentus*. *Nature*, *386*(6623), 335. https://doi.org/10.1038/386335a0

Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology & Evolution*, *19*(9), 464–469. https://doi.org/10.1016/j.tree.2004.06.004

Ji, Y., Baker, C. C. M., Li, Y., Popescu, V. D., Wang, Z., Wang, J., Wang, L., Wu, C., Hua, C., Yang, Z., Yang, C., Xu, C. C. Y., Wen, Q., Pierce, N. E., & Yu, D. W. (2020). Large-scale quantification of vertebrate biodiversity in Ailaoshan nature reserve from leech iDNA. *BioRxiv*. https://doi.org/10.1101/2020.02.10.941336

Kasper, K., Schweikhard, J., Lehmann, M., Ebert, C. L., Erbe, P., Wayakone, S., Nguyen, T. Q., Le, M. D., & Ziegler, T. (2020). The extent of the illegal trade with terrestrial vertebrates in markets and households in Khammouane Province, Lao PDR. *Nature Conservation*, *41*(August), 25–45. https://doi.org/10.3897/natureconservation.41.51888

Krishnasamy, K., & Zavagli, M. (2020). *Southeast Asia: At the heart of wildlife trade*. TRAFFIC, Southeast Asia Regional Office. https://www.traffic.org/site/assets/files/12648/sea-traps-february-2020.pdf

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Le, M. D., McCormack, T. E. M., Hoang, H. V., Duong, H. T., Nguyen, T. Q., Ziegler, T., Nguyen, H. D., & Ngo, H. T. (2020). Threats from wildlife trade: The importance of genetic data in safeguarding the endangered Four-eyed Turtle (*Sacalia quadriocellata*). *Nature Conservation*, *41*, 91–111. https://doi.org/10.3897/natureconservation.41.54661

Leempoel, K., Hebert, T., & Hadly, E. A. (2020). A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1918), 20192353. https://doi.org/10.1098/rspb.2019.2353

Li, S., Yu, G.-H., Liu, S., & Jin, C.-S. (2019). First record of the ferret-badger *Melogale cucphuongensis* Nadler et al, Nadler Nadler. *Zoological Research*, *40*(6), 575–579. https://doi.org/10.24272/j.issn.2095-8137.2019.067

Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1444), 711–719. https://doi.org/10.1098/rstb.2003.1454

Maddison, N. (Ed.). (2019). *Guidelines for the management of confiscated, live organisms*. IUCN, International Union for Conservation of Nature. https://doi.org/10.2305/IUCN.CH.2019.03.en

Martins, R. F., Fickel, J., Le, M., van Nguyen, T., Nguyen, H. M., Timmins, R., Gan, H. M., Rovie-Ryan, J. J., Lenz, D., Förster, D. W., & Wilting, A. (2017). Phylogeography of red muntjacs reveals three distinct mitochondrial lineages. *BMC Evolutionary Biology*, *17*(1), 34. https://doi.org/10.1186/s12862-017-0888-0

Martins, R. F., Schmidt, A., Lenz, D., Wilting, A., & Fickel, J. (2018). Human-mediated introduction of introgressed deer across Wallace's line: Historical biogeography of *Rusa unicolor* and *R. timorensis*. *Ecology and Evolution*, *8*(3), 1465–1479. https://doi.org/10.1002/ece3.3754

Meekan, M., Austin, C. M., Tan, M. H., Wei, N.-W.-V., Miller, A., Pierce, S. J., Rowat, D., Stevens, G., Davies, T. K., Ponzo, A., & Gan, H. M. (2017). iDNA at sea: Recovery of Whale Shark (*Rhincodon typus*)

mitochondrial DNA sequences from the Whale Shark Copepod (*Pandarus rhincodonicus*) confirms global population structure. *Frontiers in Marine Science*, *4*(December), 1–8. https://doi.org/10.3389/fmars.2017.00420

Nadler, T., Streicher, U., Stefen, C., Schwierz, E., & Roos, C. (2011). A new species of ferret-badger, Genus Melogale, from Vietnam. *Der Zoologische Garten*, *80*(5), 271–286. https://doi.org/10.1016/j.zoolgart.2011.08.004

Nguyen, D. H., & Dinh, T. M. (2020). Impacts of wildlife trade and sustainable development in Vietnam. *E3S Web of Conferences*, *157*, 03001. https://doi.org/10.1051/e3sconf/202015703001

Nguyen, T. V., Tilker, A., Nguyen, A., Hörig, L., Axtner, J., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., Wilting, A., & Fickel, J. (2021). Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit *Nesolagus timminsi*. *Environmental DNA*, *3*(4), 780–791. https://doi.org/10.1002/edn3.182

Oklander, L. I., Caputo, M., Solari, A., & Corach, D. (2020). Genetic assignment of illegally trafficked neotropical primates and implications for reintroduction programs. *Scientific Reports*, *10*(1), 3676. https://doi.org/10.1038/s41598-020-60569-3

Paijmans, J. L. A., Barlow, A., Henneberger, K., Fickel, J., Hofreiter, M., & Foerster, D. W. G. (2020). Ancestral mitogenome capture of the Southeast Asian banded linsang. *PLoS One*, *15*(6), e0234385. https://doi.org/10.1371/journal.pone.0234385

Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*(1), 42–55. https://doi.org/10.1111/1755-0998.12420

Patel, R. P., Förster, D. W., Kitchener, A. C., Rayan, M. D., Mohamed, S. W., Werner, L., Lenz, D., Pfestorf, H., Kramer-Schadt, S., Radchuk, V., Fickel, J., & Wilting, A. (2016). Two species of Southeast Asian cats in the genus Catopuma with diverging histories: An island endemic forest specialist and a widespread habitat generalist. *Royal Society Open Science*, *3*(10), 160350. https://doi.org/10.1098/rsos.160350

Patel, R. P., Wutke, S., Lenz, D., Mukherjee, S., Ramakrishnan, U., Veron, G., Fickel, J., Wilting, A., & Förster, D. W. (2017). Genetic structure and phylogeography of the Leopard Cat (*Prionailurus bengalensis*) inferred from mitochondrial genomes. *Journal of Heredity*, *108*(4), 349–360. https://doi.org/10.1093/jhered/esx017

Patou, M.-L., Chen, J., Cosson, L., Andersen, D. H., Cruaud, C., Couloux, A., Randi, E., Zhang, S., & Veron, G. (2009). Low genetic diversity in the masked palm civet *Paguma larvata* (Viverridae). *Journal of Zoology*, *278*(3), 218–230. https://doi.org/10.1111/j.1469-7998.2009.00570.x

Robichaud, W. G. (2010). A field record of Small-toothed Ferret Badger Melogale moschata in Central Laos, and other recent records of ferret badgers from the country. *Small Carnivore Conservation*, *42*, 32–34.

Robins, J. H., Ross, H. A., Allen, M. S., & Matisoo-Smith, E. (2006). Sus bucculentus revisited. *Nature*, *440*(7086), E7–E7. https://doi.org/10.1038/nature04770

Rozhnov, V. V., Korablev, M. P., & Abramov, A. V. (2019). Systematics and distribution of ferret badgers Melogale (Mammalia, Mustelidae) in Vietnam: First genetic data. *Doklady Biological Sciences*, *485*(1), 47–51. https://doi.org/10.1134/S001249661902008X

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Sales, N. G., Kaizer, M. D. C., Coscia, I., Perkins, J. C., Highlands, A., Boubli, J. P., Magnusson, W. E., Ferreira Da Silva, M. N., Benvenuto, C., & McDevitt, A. D. (2020). Assessing the potential of environmental DNA metabarcoding for monitoring Neotropical mammals: A case study in the Amazon and Atlantic Forest. *Brazil. Mammal Review*, *50*(3), 221–225. https://doi.org/10.1111/mam.12183

Sandalj, M., Treydte, A. C., & Ziegler, S. (2016). Is wild meat luxury? Quantifying wild meat demand and availability in Hue, Vietnam. *Biological Conservation*, *194*, 105–112. https://doi.org/10.1016/j.biocon.2015.12.018

Schank, C., Pollard, E. H. B., Sechrest, W., Timmins, R., Holden, J., & Walston, J. (2009). First confirmed records of Large-toothed Ferret Badger Melogale personata in Cambodia, with notes on country records of Melogale. *Small Carnivore Conservation*, *40*, 11–15.

Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Yu, D. W., Wilting, A., & Gilbert, M. T. P. (2015). IDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool – Prospects, pitfalls and avenues to be developed. *Frontiers in Zoology*, *12*(1), 24. https://doi.org/10.1186/s12983-015-0115-z

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F., & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, *22*(20), 1980. https://doi.org/10.1016/j.cub.2012.10.014

Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., & Greenwood, A. D. (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. *Molecular Ecology Resources*, *19*(6), 1486–1496. https://doi.org/10.1111/1755-0998.13069

Singh, B., Kumar, A., Uniyal, V. P., & Gupta, S. K. (2019). Complete mitochondrial genome of northern Indian red muntjac (*Muntiacus vaginalis*) and its phylogenetic analysis. *Molecular Biology Reports*, *46*(1), 1327–1333. https://doi.org/10.1007/s11033-018-4486-z

Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, *32*(19), 2920–2927. https://doi.org/10.1093/bioinformatics/btw346

Sterling, E., & Hurley, M. M. (2005). Conserving biodiversity in Vietnam: Applying biogeography to conservation research. *Proceeding of the California Academy of Sciences*, *56*(I9), 98–114.

Székely, D., Corfixen, N. L., Mørch, L. L., Knudsen, S. W., McCarthy, M. L., Teilmann, J., Heide-Jørgensen, M. P., & Olsen, M. T. (2021). Environmental DNA captures the genetic diversity of bowhead whales (*Balaena mysticetus*) in West Greenland. *Environmental DNA*, *3*(1), 248–260. https://doi.org/10.1002/edn3.176

Tilker, A., Abrams, J. F., Mohamed, A., Nguyen, A., Wong, S. T., Sollmann, R., Niedballa, J., Bhagwat, T., Gray, T. N. E., Rawson, B. M., Guegan, F., Kissing, J., Wegmann, M., & Wilting, A. (2019). Habitat degradation and indiscriminate hunting differentially impact faunal communities in the Southeast Asian tropical biodiversity hotspot. *Communications Biology*, *2*(1), 1–11. https://doi.org/10.1038/s42003-019-0640-y

Tilker, A., Abrams, J. F., Nguyen, A., Hörig, L., Axtner, J., Louvrier, J., Rawson, B. M. B. M., Nguyen, H. A. Q. H. A. Q., Guegan, F., Nguyen, T. V. T. V., Le, M., Sollmann, R., & Wilting, A. (2020). Identifying conservation priorities in a defaunated tropical biodiversity hotspot. *Diversity and Distributions*, *26*(4), 426–440. https://doi.org/10.1111/ddi.13029

van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., … Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, *591*(7849), 265–269. https://doi.org/10.1038/s41586-021-03224-9

Veron, G., & Jennings, A. P. (2017). Javan mongoose or small Indian mongoose–Who is where? *Mammalian Biology*, *87*, 62–70. https://doi.org/10.1016/j.mambio.2017.05.006

Veron, G., Patou, M.-L., Debruyne, R., Couloux, A., Fernandez, D. A. P., Wong, S. T., Fuchs, J., & Jennings, A. P. (2015). Systematics of the

Southeast Asian mongooses (Herpestidae, Carnivora): Solving the mystery of the elusive collared mongoose and Palawan mongoose. *Zoological Journal of the Linnean Society*, *173*(1), 236–248. https://doi.org/10.1111/zoj.12206

Wasser, S. K., Clark, W. J., Drori, O., Kisamo, E. S., Mailand, C., Mutayoba, B., & Stephens, M. (2008). Combating the illegal trade in African Elephant Ivory with DNA forensics. *Conservation Biology*, *22*(4), 1065–1071. https://doi.org/10.1111/j.1523-1739.2008.01012.x

Wehi, P. M., Whaanga, H., & Trewick, S. A. (2012). Artefacts, biology and bias in museum collection research. *Molecular Ecology*, *21*(13), 3103–3109. https://doi.org/10.1111/j.1365-294X.2012.05589.x

Weiskopf, S. R., McCarthy, K. P., Tessler, M., Rahman, H. A., McCarthy, J. L., Hersch, R., Faisal, M. M., & Siddall, M. E. (2018). Using terrestrial haematophagous leeches to enhance tropical biodiversity monitoring programmes in Bangladesh. *Journal of Applied Ecology*, *55*(4), 2071–2081. https://doi.org/10.1111/1365-2664.13111

Weitemier, K., Penaluna, B. E., Hauck, L. L., Longway, L. J., Garcia, T., & Cronn, R. (2021). Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. *Molecular Ecology*. https://doi.org/10.1111/mec.15811

Wetterstrand, K. A. (2021). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: https://www.genome.gov/sequencingcostsdata. Accessed September 24, 2021.

Wilcox, T. M., McKelvey, K. S., Young, M. K., Sepulveda, A. J., Shepard, B. B., Jane, S. F., Whiteley, A. R., Lowe, W. H., & Schwartz, M. K. (2016). Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char *Salvelinus fontinalis*. *Biological Conservation*, *194*, 209–216. https://doi.org/10.1016/j.biocon.2015.12.023

Williams, K. M., Barkdull, M., Fahmy, M., Hekkala, E., Siddall, M. E., & Kvist, S. (2020). Caught red handed: iDNA points to wild source for CITES-protected contraband leeches. *European Journal of Wildlife Research*, *66*, 80. https://doi.org/10.1007/s10344-020-01419-5

Wilting, A., Courtiol, A., Christiansen, P., Niedballa, J., Scharf, A. K., Orlando, L., Balkenhol, N., Hofer, H., Kramer-Schadt, S., Fickel, J., & Kitchener, A. C. (2015). Planning tiger recovery: Understanding intraspecific variation for effective conservation. *Science Advances*, *1*(5), e1400175. https://doi.org/10.1126/sciadv.1400175

Yao, L., Li, H., Martin, R. D., Moreau, C. S., & Malhi, R. S. (2017). Tracing the phylogeographic history of Southeast Asian long-tailed macaques through mitogenomes of museum specimens. *Molecular Phylogenetics and Evolution*, *116*, 227–238. https://doi.org/10.1016/j.ympev.2017.08.006

Yao, L., Witt, K., Li, H., Rice, J., Salinas, N. R., Martin, R. D., Huerta-Sánchez, E., & Malhi, R. S. (2020). Population genetics of wild *Macaca fascicularis* with low-coverage shotgun sequencing of museum specimens. *American Journal of Physical Anthropology*, *173*(1), 21–33. https://doi.org/10.1002/ajpa.24099

Zhang, H., Miller, M. P., Yang, F., Chan, H. K., Gaubert, P., Ades, G., & Fischer, G. A. (2015). Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Global Ecology and Conservation*, *4*(July), 414–422. https://doi.org/10.1016/j.gecco.2015.08.002

Zhou, F., Yao, Y., Li, D., Wu, J., Wen, A., Xie, M., Wang, Q., Zhu, G., Ni, Q., Zhang, M., & Xu, H. (2020). Complete mitochondrial genome sequence of a Chinese ferret-badger (*Melogale moschata*) and phylogenetic analysis. *Mitochondrial DNA Part B*, *5*(1), 1115–1116. https://doi.org/10.1080/23802359.2020.1713920

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wilting, A., Nguyen, T. V., Axtner, J., Nguyen, A., Schmidt, A., Le, M., Nguyen, A. H. Q., Rawson, B. M., Tilker, A., & Fickel, J. (2022). Creating genetic reference datasets: Indirect sampling of target species using terrestrial leeches as sample "collectors". *Environmental DNA*, 4, 311–325. https://doi.org/10.1002/edn3.256