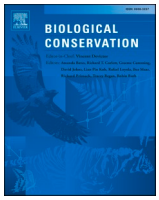
Biological Conservation 284 (2023) 110168

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon



Monitoring terrestrial wildlife by combining hybridization capture and metabarcoding data from waterhole environmental DNA

Juan Li [a,b], Peter Seeber [a], Jan Axtner [c], Rachel Crouthers [d], Milou Groenenberg [d], Arnulf Koehncke [e], Alexandre Courtiol [f], Pin Chanratana [g], Alex D. Greenwood [a,h,*]

[a] Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany
[b] Department of Biologie, Chemie, and Pharmazie, Freie Universität Berlin, Berlin 14195, Germany
[c] Department of Ecological Dynamics, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany
[d] World Wild Fund for Nature, Phnom Penh 120102, Cambodia
[e] WWF Germany, Berlin 10117, Germany
[f] Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, 10315 Berlin, Germany
[g] Ministry of Environment, Phnom Penh 120101, Cambodia
[h] School of Veterinary Medicine, Freie Universität Berlin, Berlin 14163, Germany

ARTICLE INFO

Keywords:
eDNA
Hybridization capture
Metabarcoding
Terrestrial animals

ABSTRACT

In conservation science, observation-based methods are generally applied to wildlife monitoring. While useful, such approaches are often restricted to well-characterized and conspicuous species. Environmental DNA (eDNA) can complement observation-dependent surveys, as sample collection is generally less labor-intensive. Furthermore, eDNA can be used to simultaneously detect multiple taxa in various habitats. Most eDNA applications rely on short PCR product-based meta-barcoding approaches. However, such approaches may be less robust when applied to genetically uncharacterized species. Hybridization capture techniques while less sensitive than metabarcoding, can identifying divergent sequences, especially those obtained from highly degraded DNA. To assess eDNA based methods for surveillance in a conservation context, we collected samples during the dry season from randomly selected waterholes from a protected area in eastern Cambodia. We applied both hybridization capture enrichment and metabarcoding targeting mammal mitogenomes to water (N = 46) and sediment (N = 10) samples. Seventeen species were detected, including 11 mammals, three amphibians, two reptiles, and one bird. Six species overlapped between the two applied methods. Seven species were hybridization capture-specific detections, and four were metabarcoding-specific. Metabarcoding was more sensitive to abundant or large body-size species while hybridization capture provided more mitogenomic information. While both methods have some advantages over observational approaches, combining them may improve the sensitivity, number of species detected and amount of genetic information obtained from eDNA. We demonstrate that eDNA from tropical forest waterholes can be used to determine the presence of wildlife that may be difficult to detect using other observational approaches.

1. Introduction

Wildlife monitoring in Cambodia, and elsewhere, is generally observation-based. To date, Line transects (O'Kelly et al., 2012; Gray, 2013; Nuttall et al., 2021), camera traps (Gray et al., 2014a; Rostro-García et al., 2018; Pin et al., 2020), and direct counts (Wright et al., 2013; Loveridge et al., 2019) have been implemented to gain great insight into wildlife population size and distribution. To further improve the inclusivity of wildlife monitoring, audible detection surveys (Tak et al., 2022) and DNA-based approaches (Pollard et al., 2008; Gray et al., 2014b) have been applied in Cambodia. However, conservation management strategies would benefit from methods that identify multiple species from indirect non-invasive sources.

Cambodia has over 50 protected areas, covering nearly 7.5 million hectares (UNEP-WCMC, UN Environment Programme - World Conservation Monitoring Centre) that serve as natural habitats for various wildlife species (Wharton, 1957; Hughes, 2017). Severe ecological disruptions are driving dramatic biodiversity declines in Indochina with

* Corresponding author at: Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany.
E-mail address: greenwood@izw-berlin.de (A.D. Greenwood).

https://doi.org/10.1016/j.biocon.2023.110168
Received 28 March 2023; Received in revised form 1 June 2023; Accepted 11 June 2023
Available online 7 July 2023
0006-3207/© 2023 Elsevier Ltd. All rights reserved.

numerous wildlife species being driven to the brink of extinction or beyond, for example kouprey (*Bos sauveli*), Javan rhinoceroses (*Rhinoceros sondaicus*), and tigers (*Panthera tigris*) (Duckworth, 1998, Poole and Duckworth, 2005, Forrest et al., 2011, Rostro-García et al., 2018, Rostro-García et al., 2023). Increased wildlife trade has shifted local strategies from subsistence and selective hunting to large-scale illegal and often indiscriminate hunting at the human-wildlife interface, which continues to threaten biodiversity (Loucks et al., 2009; Gray et al., 2017; Ibbett et al., 2021).

Eastern Cambodia supports vast areas of deciduous dipterocarp forests that experience five to six months of seasonal drought annually (Pin et al., 2020). During the dry season, waterholes in the forests become vital water resources for numerous wildlife species (Gray et al., 2015; Pin et al., 2020), which in turn can result in increased concentrations of shed DNA into the water through body fluid, mucus, feces or tissues, i.e. environmental DNA (eDNA). DNA typically binds to suspended or sedimented particulate matter, thus water and sediment represent genomic archives of terrestrial animals. Over the last decade, eDNA methods have been applied to detect the presence of various rare, cryptic, or low-density species (Jerde et al., 2011; Bohmann et al., 2014; Parsons et al., 2018). Initially used for aquatic species detection, eDNA has recently been successfully applied to terrestrial wildlife monitoring (Harper et al., 2019; Seeber et al., 2019; Lyet et al., 2021; Mena et al., 2021). In many cases, multiple species can be detected simultaneously from a single sample. In Cambodia, non-invasive DNA methods have thus far been limited to a singular focal species (elephants) (Pollard et al., 2008; Maltby and Bourchier, 2011; Gray et al., 2014b).

Species detection is frequently the focus of eDNA approaches. Metabarcoding, which targets short-sequence regions, is currently the most common technique in eDNA monitoring. PCR-based metabarcoding can achieve relatively high-resolution of species identification (Seeber and Epp, 2022). However, inevitable biases inherent in PCR approaches (e.g. primer bias) may confound the final species detection and assignment (Seeber and Epp, 2022). Moreover, PCR primers developed based on known taxa may fail to identify molecularly uncharacterized species (Bourret et al., 2020). Hybridization capture enrichment of eDNA is an alternative to PCR-based methods. It relies on DNA or RNA oligonucleotide probes (termed baits) that are complementary to genomic regions of interest, potentially including numerous loci or full genomes (Gasc et al., 2016). Targeting various loci or entire genomes can yield important information on genetic diversity and provide more accurate species identification (Seeber et al., 2019, Jensen et al., 2021).

Environmental DNA metabarcoding yields comparable results to observation-based approaches in various habitats (Ushio et al., 2017; Lyet et al., 2021; Mena et al., 2021; Farrell et al., 2022). In some cases, eDNA approaches may perform better, identifying overlooked species, especially when hybridization capture methods are applied (Harper et al., 2019; Giebner et al., 2020). By applying baits designed from full mitogenomes on eDNA, previous studies retrieved complete mitogenomes of target species with high coverage (Seeber et al., 2019, Jensen et al., 2021). However, few studies have directly compared the effectiveness of eDNA hybridization capture with metabarcoding approaches using terrestrial vertebrate DNA from environmental samples. In the current study, we extracted eDNA from water and sediment collected from waterholes in eastern Cambodia in the dry season. Our aims were to 1) assess the range of species which could be identified, 2) compare hybridization capture and metabarcoding performance using species detection as the key indicator, and 3) to determine the effect of sample type on method performance.

2. Materials and methods

2.1. Sample collection and DNA isolation

Sampling of waterholes was conducted within the Srepok Wildlife Sanctuary (SWS, 3729 km^2), previously known as Mondulkiri Protected Forest, which is situated in Cambodia's Eastern Plains Landscape (Fig. 1). Previous surveys have highlighted the biodiversity significance of this area which supports various threatened mammals, birds, and reptiles (Phan et al., 2010, Gray et al., 2012a, Gray et al., 2014a, Groenenberg et al., 2020). Environmental samples in the current study have been collected from the waterholes that were a subset of the concurrent camera trap study published by Pin et al., 2020. Sampling was conducted from January to April 2016, which corresponded to the local dry season. Waterhole visitation time included three sampling points: Visit 1 (January 25–February 5, 2016), Visit 2 (February 25–March 5, 2016) and Visit 3 (March 25–April 5, 2016). At each sampling site, 50 mL water was collected from the water surface. For a subset of the waterholes, sediment samples were taken from the top sediment layer using sterile 50-mL tubes. The number of waterholes surveyed decreased throughout the dry season due to seasonal dry-outs. Consequently, fewer samples were collected at the end of the dry season. For hybridization capture, 28 samples were from first visit, 19 from second visit, and 9 from the third visit. For metabarcoding, 31 samples were from first visit, 17 from second visit, and 8 from third visit (Table 1). Forty samples (30 water and 10 sediments) were analyzed by both hybridization capture and metabarcoding.

DNA from water samples was extracted using a NucleoSpin Soil kit (Macherey-Nagel, Düren, Germany) combined with an eDNA isolation kit to optimize DNA yields. Briefly, each sample was centrifuged at 4000 $\times g$ for 45 min for separate processing of supernatants and pellets. The supernatants were processed as per the instructions of the kit manufacturer, followed by an additional inhibitor removal step from the eluted DNA (NucleoSpin Inhibitor Removal, Macherey-Nagel). DNA from the pellet was extracted using a NucleoSpin Soil kit and eluted with 100 microliter (μL) Clean-up DNA elution buffer. DNA from sediment samples was isolated using a NucleoSpin Soil kit (Macherey-Nagel), according to the manufacturer's instructions. With each extraction batch (N = 7, each), one negative control (DNA-free water) was processed and subjected to subsequent library preparation, capture, and metabarcoding. DNA integrity and concentration were assessed using an Agilent 2200 Tapestation (Agilent, Santa Clara, CA, USA) with genomic chips, and concentrations were additionally measured using a Qubit™ dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). Low DNA-binding reaction tubes (Sarstedt, Nümbrecht, Germany) were used throughout extractions and all downstream steps to minimize DNA loss.

2.2. Hybridization capture

2.2.1. Capture evaluation and database preparation

Custom-designed RNA oligonucleotide baits (Daicel Arbor Biosciences, Ann Arbor, MI, USA) were used for hybridization capture. The bait panel was produced based on the mitochondrial genome sequences of 38 wild mammal species (Seeber et al., 2019), belonging to 26 families in 15 orders. Four species in four genera overlapped with wildlife present in Cambodia: Javan mongoose (*Herpestes javanicus*), Eurasian otter (*Lutra lutra*), leopard (*Panthera pardus*), and large flying fox (*Pteropus vampyrus*). The baits were 80 base pairs (bp) long, and threefold tiling was used to account for eDNA degradation. A total of 19,496 unique baits were generated.

Performance of the bait set was simulated in silico using the simulation package CapSim (Cao et al., 2018). All available mitochondrial genomes of wildlife species in Cambodia were downloaded from NCBI refseq with Entrez Direct (https://www.ncbi.nlm.nih.gov/books/NBK179288/) to build a customized mitogenome reference database. The Cambodian mitochondrial genome reference database included 62 mammal species, 45 reptile species, 12 amphibian species, and 168 avian species (287 species in total). We mapped bait sequences to the custom database with permissive parameters using Bowtie2 (Langdon, 2015), allowing multiple position alignments to probe all potentially aligned regions (parameters: –local –very-sensitive-local –mp 8 –rfg 10,8 –rfg 10,8 -k 10000). Sequences with at least one probe aligning to them

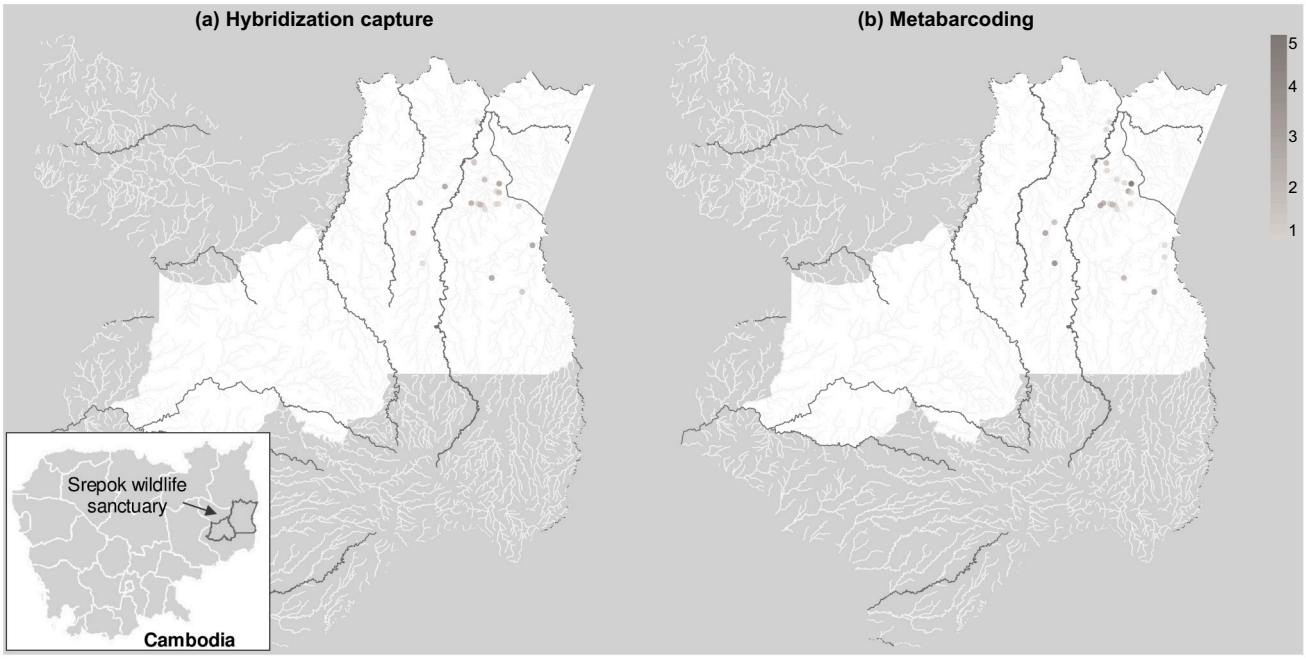


Fig. 1. Locations are shown for species detected from eDNA by (a) hybridization capture and (b) metabarcoding. The numbers of detected species at each location are shown as a gradient red. The grey boundary depicts the Srepok wildlife sanctuary in Cambodia where field work was conducted. Rivers are illustrated as blue lines and streams are shown in light blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Detected species.

eDNA method		Hybridization capture							Metabarcoding						
Visitation time		Visit 1		Visit 2		Visit 3		Total counts	Visit 1		Visit 2		Visit 3		Total Counts
Species category	Species (IUCN status)	Water	Sed	Water	Sed	Water	Sed		Water	Sed	Water	Sed	Water	Sed	
Mammals	*Bos javanicus* (EN)	2						2	2	1	1				4
	Bos taurus	2				1		3	1						1
	Elephas maximus (EN)			4		1		5	2		2				4
	Muntiacus vaginalis	1	1	1		1		4	1	1	1		1		4
	Rucervus eldii (EN)	1		1				2	1						1
	Sus scrofa	9	2	5	1	2	1	20	15	3	5		1	1	25
	Viverricula indica	1		1				2							
	Canis species								6		5		2		13
	Rattus exulans								2						2
	Rattus norvegicus								2						2
	Rattus tanezumi								1						1
Reptiles	*Cuora amboinensis* (EN)	2		1			1	4							
	Varanus salvator	1						1							
Amphibians	*Hoplobatrachus rugulosus*			1				1							
	Kaloula pulchra	1		1				2							
	Microhyla ornata					1		1							
Birds	*Gallus gallus*	1						1							
Positive samples/Total processed samples		13/24	2/4	7/15	1/4	4/7	1/2	28/56	19/27	3/4	8/13	1/4	3/6	1/2	35/56
Positive rate of species detection		0.54	0.50	0.47	0.25	0.57	0.50	0.50	0.70	0.75	0.62	0.25	0.50	0.50	0.63

Visitation time: the waterhole visit time. Visit 1 (2016.1.25 to 2016.2.5), Visit 2 (2016.2.25 to 2016.3.5), Visit 3 (2016.3.25 to 2016.4.5).
Water: water samples; Sed: Sediment samples.
Total counts: the total detection counts of respective species.
Positive samples: the number of samples that had positive detections for any species.

were retained, sorted, and indexed accordingly using Samtools (Li et al., 2009), which was utilized as reference for in silico sequencing with CapSim (parameters: fmedian = 400, smedian = 320, illen = 250, ilmode = pe).

2.2.2. Library preparation, hybridization capture, and sequencing

An established Illumina sequencing library preparation protocol (Meyer and Kircher, 2010) was used with modifications (Seeber et al., 2019). Physical fragmentation was conducted using an ultrasonicator (Covaris M220; Covaris, Woburn, MA, USA) to a fragment size of approximately 400 bp. Shearing efficiency and fragment distribution

were assessed on an Agilent 2200 TapeStation with D1000 ScreenTapes. End repair, adapter ligation, and fill-in reactions were performed with 42.5 μL sheared eDNA using the respective NEBNext kits (New England Biolabs, New England Biolabs, Ipswitch, MA, USA) according to the manufacturer's instructions. To pool samples for capture and high-throughput sequencing, samples carrying P5/P7 adaptors were indexed with unique combinations of tags, using five amplification cycles and three technical replicates.

We pooled four libraries each at equal molarity to optimize the utilization and performance of the RNA baits. Capture reactions were performed on magnetic beads following the manufacturer's instructions (Daicel Arbor Biosciences). Incubation was 48 h at 60 °C, followed by 15-cycle post-capture PCR (in duplicate) of on-bead capture products using P5/P7 bridge primers and KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland). The enriched products were cleaned using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany), visualized, and quantified using an Agilent 2200 Tapestation and Qubit measurement. For sequencing, 16 capture products (i.e., 56 enriched libraries), plus two capture products from negative controls (8 libraries), were randomly separated into two batches and were pooled at equimolarity. Each pool was diluted to 8 pM and mixed with a PhiX DNA control spike-in of 1 % for two separate runs. The two pooled libraries were sequenced on an Illumina Miseq platform (Miseq v2, 500 cycles, 2 × 250 bp paired-end reads; Illumina, San Diego, CA, USA).

2.2.3. Hybridization capture bioinformatic analyses

Adaptors were removed from demultiplexed raw reads using CUTADAPT v1.15 (Martin, 2011), followed by quality trimming with TRIMMOMATIC v0.38 (Bolger et al., 2014). A single window of 10 bp with quality score above 20 and minimal length of 50 bp was set to remove low quality reads that were shorter than 50 bp. The remaining reads were merged using FLASH v1.2.11 (Magoc and Salzberg, 2011) with a minimum overlap of 20 bp. Simulation data were processed in the same way as sequencing data. Given presumable degradation of the eDNA, unmerged reads were also retained to perform consecutive mapping against our customized mitogenome reference database.

The selected sequences were mapped against the custom reference database using BWA (Li and Durbin, 2009). To minimize misclassifications owing to low-complexity regions, we identified low-complexity sequences using dustmasker (-level 45) (Morgulis et al., 2006). Low complexity sequences were replaced with N's using an in-house bash script. Merged and unmerged reads were individually mapped against the masked mitogenome database with default settings, which were concatenated to a single bam file for each sample using Samtools (Li et al., 2009). Mapped reads were extracted, sorted, deduplicated, and indexed using Samtools and GATK -4.2.0.0 (McKenna et al., 2010). Total coverage of each mitochondrial genome was calculated with bedtools (Quinlan, 2014). Alignments that covered at least 2 % of the whole mitogenome of each reference with more than three reads and similarity above 95 %, were kept. Retained reads were used for further blast searches against the complete NCBI nucleotide reference database to filter out ambiguous matches. A final cutoff of above 95 % identity was applied. Mitogenome mapping results were visualized using R package circlize (Gu et al., 2014). A local BLASTx search was performed using unaligned reads against the NCBI protein reference database using Diamond (e-value 1-e3) (Buchfink et al., 2015). Blast results were filtered with the thresholds of 99 % query coverage and 98 % query identity, which were then parsed to each classification level using perl script tax_trace.pl (https://github.com/theo-allnutt-bioinformatics/scripts/blob/master/tax_trace.pl).

2.3. Metabarcoding

2.3.1. Library preparations and sequencing

The eDNA metabarcoding libraries were prepared following a two-round PCR workflow (Axtner et al., 2019). As in previous studies (Tilker et al., 2019; Nguyen et al., 2021), we modified the protocol and used only one mammal-specific marker, a 93 bp-fragment of the mitochondrial encoded 16S rRNA (Taylor, 1996). We used six PCR replicates instead of four replicates. The primers for the marker gene targeted for the first round. Illumina sequencing adaptors were synthesized with different mirrored tag pairs for the second round. This allowed for processing unique twin tag pairs, which facilitated identification of cross-contamination and tag-jumping.

The PCRs were run in reactions of 20 μL (AmpliTaq Gold™ 360, Invitrogen) with 2 μL template. Cycling conditions were 5 min at 95 °C, 38 cycles (14 cycles for the second round) of 30 s at 95 °C, 30 s at 55 °C, and 45 s at 72 °C. A final extension was run for 5 min at 72 °C. Two μL of the first PCR was used as template DNA for the second PCR under the same cycling conditions. Bank vole (*Myodes glareolus*) DNA was used as positive control. The qualification of final PCR products was assessed by agarose gel electrophoresis. Additional measurements were conducted on positive and negative controls using the Agilent 2200 TapeStation system. To determine the final DNA concentrations, all PCR samples were purified with magnetic AMPure® beads (Beckman Coulter) at a ratio of 0.7 according to the manufacturer's instructions and measured with a Quant-iT™ PicoGreen® dsDNA assay kit. Finally, all samples with a target DNA band, plus positive and negative controls, were pooled to equimolarity at a final concentration of 4 nM. Following the Illumina MiSeq Reagent Kit v3 600 cycle instructions, the pool was diluted to 11 pM mixed with 25 % of PhiX control, and sequenced on the Illumina MiSeq platform.

2.3.2. Metabarcoding bioinformatic analyses

Raw reads were de-multiplexed from the basecall files. We merged the resulting paired reads using a described bioinformatic pipeline (Axtner et al., 2019). We removed unpaired reads and dereplicated identical reads for the final FASTA files with their frequency of occurrence in each sample added to the file header.

For species identification, we used a probabilistic taxonomic assignment method PROTAX (Somervuo et al., 2016). We updated a species list of Tilker et al. (2019) to include 173 mammal and bird species from across the broader Indochinese ecoregion. We searched INSDC databases for available references of the listed species and updated the reference database from Axtner et al. (2019). We added reference sequences for 12 species and added 48 reference sequences to the 16S rRNA reference database. In total, the reference database covered 5872 reference sequences of 3962 tetrapod species. We trained PROTAX models and weighted them based on our Indochinese species list by assigning a prior probability of 90 % to these species and a 10 % probability to all others (Somervuo et al., 2016) using the pipeline of Axtner et al. (2019). For each sequence PROTAX provided an assignment to the level of class, order, family, genus, and species along with a log-probability value for the assignment. Pairwise sequence similarities of queries and references were calculated using LAST (Kielbasa et al., 2011). We aggregated the results for each sample, PCR replicate, and each taxonomic level. We summed up the total number of reads per sample assigned to certain taxa, by calculating the mean probability of the assignments and by calculating the mean similarity score to the nearest reference for the assignments. Species assignment was allocated when the same assignment was made in at least two PCR replicates (out of a total of six) and five reads per sample. Lastly, we performed a local blast against the complete NCBI nucleotide reference database (e-value 1-e5) to identify the demultiplexed reads in order to compare results with the PROTAX assignments. Blast results were parsed following the same procedure described above for hybridization capture data.

2.4. Statistical analyses

To estimate the capacity of hybridization capture to retrieve target sequences, we fitted a linear model (LM) using the R function lm(). We used the logged number of reads mapped to mitogenome references in

each sample as the response variable, and DNA integrity number (DIN) measured by the TapeStation 2200, DNA concentration, sample types and the logged number of processed reads as fixed effects. We performed anova tests using the function anova() in R to evaluate the significance of each effect.

To assess the performance of hybridization capture and metabarcoding on species detection, we fitted a generalized linear mixed-effect model (GLMM) using R package spaMM (Rousset and Ferdy, 2014). Species detection was used as the response variable, and method, species category, sample type, and sampling interval days as fixed effects. Individual sample ID, species, and location of each sampling site were included as random effects. The response variable (species detection) was coded with a one or zero depending on whether or not a species was detected for any given sampling date and location by either of the methods. All species detected at least once across all dates, locations and methods, were considered throughout the analysis. We fitted the model using a binomial family with a logit link and PQL/L method. We estimated marginal R^2 (proportion of total variance explained by fixed effects) and conditional R^2 (proportion of total variance explained by fixed and random effects) values, and the proportion of variance contributed by random effects using function pseudoR2() in spaMM package. We compared the final full model with reduced models, which did not contain a respective fixed or random effect of interest to evaluate the significance of each effect, using a log-likelihood ratio test for which the distribution of the test statistics under the null hypothesis was generated using 1000 parametric bootstrap replicates. Modelling assumptions were supported by diagnostic tests offered by the R package DARHMa (Hartig, 2022).

2.5. Data access

Sequencing data generated in the current study are available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA943158 and PRJNA933918.

3. Results

3.1. Evaluation of hybridization capture bait performance

In total, 72.11 % of the baits matched references in our Cambodian mitogenome reference database. Mammal species showed the highest mitogenome coverage, ranging from 36.95 % to 100 %, followed by 3.19 % to 20.4 % mitogenome coverage for reptiles, 5.28 % to 13.4 % for amphibians, and 4.29 % to 17.64 % for birds (Fig. S1). Simulation results from Capsim, taking off-target capture into consideration, showed a marked increase in mitogenome coverage, i.e., 93.67 % to 100 % in mammals, 17.95 % to 79.71 % in reptiles, 27.63 % to 61.75 % in amphibians, and 15.91 % to 80.20 % in birds (Fig. S1). Among mammal species, baits were distributed evenly across the full mitogenomes, but were more concentrated at the 12S ribosomal RNA (rRNA) and 16S rRNA regions, tilling to 99-fold coverage at maximum (Fig. 2). Baits aligning to non-mammal mitogenome references favored the variable regions 7–9 (V7–9) of 16S ribosomal RNA with 98-fold coverage at maximum (Fig. 2).

3.2. Hybridization capture

A total of 37,123,056 of 38,555,232 reads from 46 water and 10 sediments passed the pre-processing steps. Samples produced variable numbers of reads ranging from 156,031–1,847,627 reads due to the inconsistent proportion of potential target sequences in the libraries pooled in each capture reaction. The demultiplexing using corresponding index combinations confirmed that negative controls did not generate sequences. After mitogenome mapping, 222,889 reads from 28 samples (24 water and 4 sediment samples) aligned to the Cambodian mitochondrial genome reference database, corresponding to a minimum 313 reads (0.05 %) and maximum 58,034 reads (3.66 %) per sample. Statistical analysis showed that sample types, and total DNA concentration and integrity had no significant effect on the number of reads aligned to the database. However, sequencing throughput significantly affected number of reads aligned (Fig. 3a and Table 2). The fitted linear model was assessed as shown in Fig. S2a.

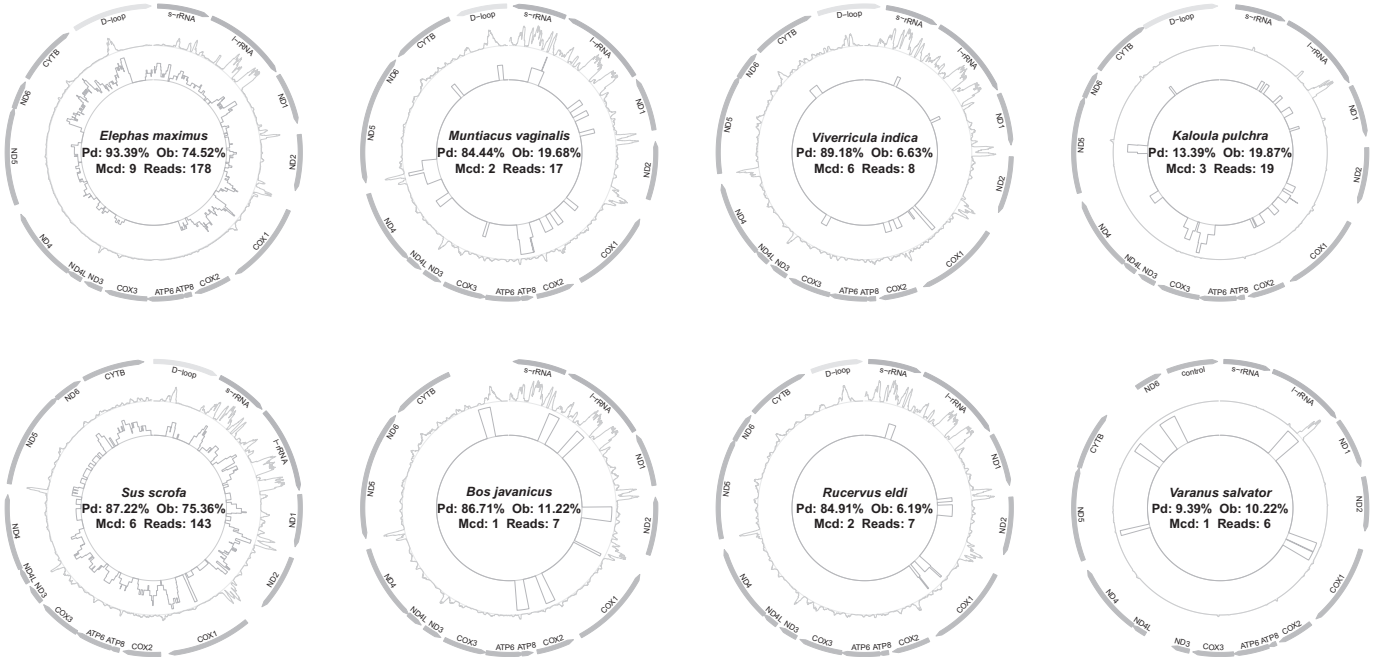


Fig. 2. Clockwise overviews of mitogenome coverage of respective target species are shown. The outer circles represent the annotated mitogenome reference, the middle circle depicts individual base pair coverage of corresponding species based on simulations, and the innermost circles depict base pair coverage from our hybridization capture results. Reads are the number of reads mapped to reference; Mcd: Maximal coverage depth; Pd: Predicted coverage; Ob: Observed coverage.

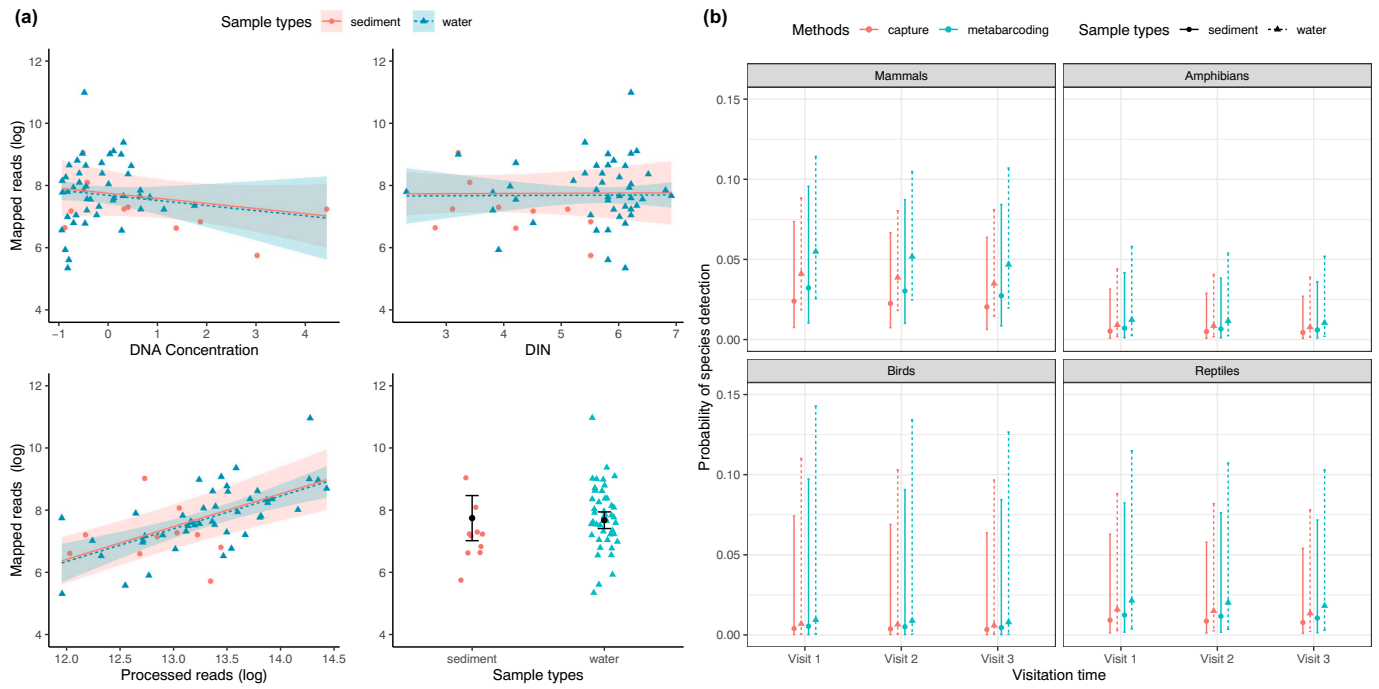


Fig. 3. Plots from the linear models (LM) and generalized logistic mixed models (GLMM) are shown. Panel (a) shows the observations and LM based predictions of the logged number of reads mapped to the reference database versus DNA concentration, DIN, processed reads, (logged) and sample type. Panel (b) shows the GLMM estimated species detection probability of the two eDNA-based methods. Mapped reads indicates the number of reads that were mapped to the reference database; DIN: the DNA integrity number generated from the Tapestation system; Processed reads: the 23 total number of reads that past pre-processing step from raw data; Visitation time: Visit 1 (2016.1.25 to 2016.2.5), Visit 2 (2016.2.25 to 2016.3.5), Visit 3 (2016.3.25 to 2016.4.5); Capture: hybridization capture.

Table 2
Results from linear model that predicts the effects of sample properties and sequencing on mapped reads.

Fixed effects	Mapped reads (log)			
	Estimates	Std. error	T value	P value
(Intercept)	−6.291	2.761	−2.278	0.027
DIN	0.007	0.127	0.052	0.958
DNA Concentration	−0.164	0.138	−1.192	0.239
Processed reads (log)	1.052	0.197	5.334	**<0.001*****
Sample types[Sediments]	0.068	0.419	0.162	0.872
R^2 / R^2 adjusted	0.395 / 0.347			

Mapped reads: the number of reads mapped to the Cambodian mitochondrial genome reference database.
DIN: DNA Integrity Number that was generated from Tapestation system for the assessment of DNA quality.
Processed reads: the total number of reads that have past pre-processing step.
Statistical significance: P value<0.05*, P<0.01**, P<0.001***.

Species presence was detected in 28 of the 56 samples used. This included 15 samples collected from the Visit 1, 8 samples collected from the Visit 2, and 5 sample collected from the Visit 3 (Fig. 1, Fig. 4a and Table 1). In total, we detected thirteen species. Seven mammals detected included: domesticated cattle (*Bos taurus*), banteng (*Bos javanicus)*, Asian elephant (*Elephas maximus*), red muntjac (*Muntiacus vaginalis*), Eld's deer (*Rucervus eldii*), wild boar (*Sus scrofa*) and small Indian civet (*Viverricula indica*). Four amphibians detected were the East Asian bullfrog (*Hoplobatrachus rugulosus*), the banded bullfrog (*Kaloula pulchra*) and the ornate narrow-mouthed toad (*Microhyla ornate*). Reptiles detected included the water monitor (*Varanus salvator),* and the Amboina box turtle (*Cuora amboinensis*). One bird species was detected (red junglefowl *Gallus gallus*). Of all the species identified, twelve species were detected from the water samples. Only two mammal species were detected from sediment samples (Fig. 4a and Table 1). Wild boar (*S. scrofa*) was the most-frequently detected species, detected in sixteen water and four sediment samples, followed by Asian elephant (*E. maximus*, five water samples) and red muntjac (*M. vaginalis*, one sediment sample and three water samples). Banteng (*B. javanicus*), Eld's deer (*R. eldii*), and small Indian civet (*V. indica*), and the banded bullfrog (*K. pulchra*) were each only detected in two water samples (Table 1.). Species only detected on one occasion from different water samples included; the Asian bullfrog (*H. rugulosus*), ornate narrow-mouthed toad (*M. ornate*), the water monitor (*V. salvator*), and red junglefowl (*G. gallus*) (Table 1).

Mammal species showed higher mitogenome coverage than non-mammal species (Fig. 2 and Fig. 4a). Asian elephant (*E. maximus*) and wild boar (*S. scrofa*) showed the highest coverage, with 178 reads aligned to the elephant (*E. maximus*), covering 74.52 % of its mitogenome (no coverage for COX3, ATP6, and repeats in D-loop) with 9-fold maximal base coverage. Wild boar (*S. scrofa*) had 143 reads aligned, covering 75.36 % of the mitogenome with 6-fold maximal base coverage. Seventeen reads were aligned to red muntjac (*M. vaginalis*) with 19.68 % mitogenome coverage; 7 reads aligned to banteng (*B. javanicus*) with 11.22 % coverage; 8 reads aligned to small Indian civet (*V. indica*) with 6.63 % coverage; 7 reads aligned to Eld's deer (*R. eldii*) with 6.19 % coverage. Among amphibians, banded bullfrog (*K. pulchra*) had 19.87 % coverage by 19 reads and the ornate narrow-

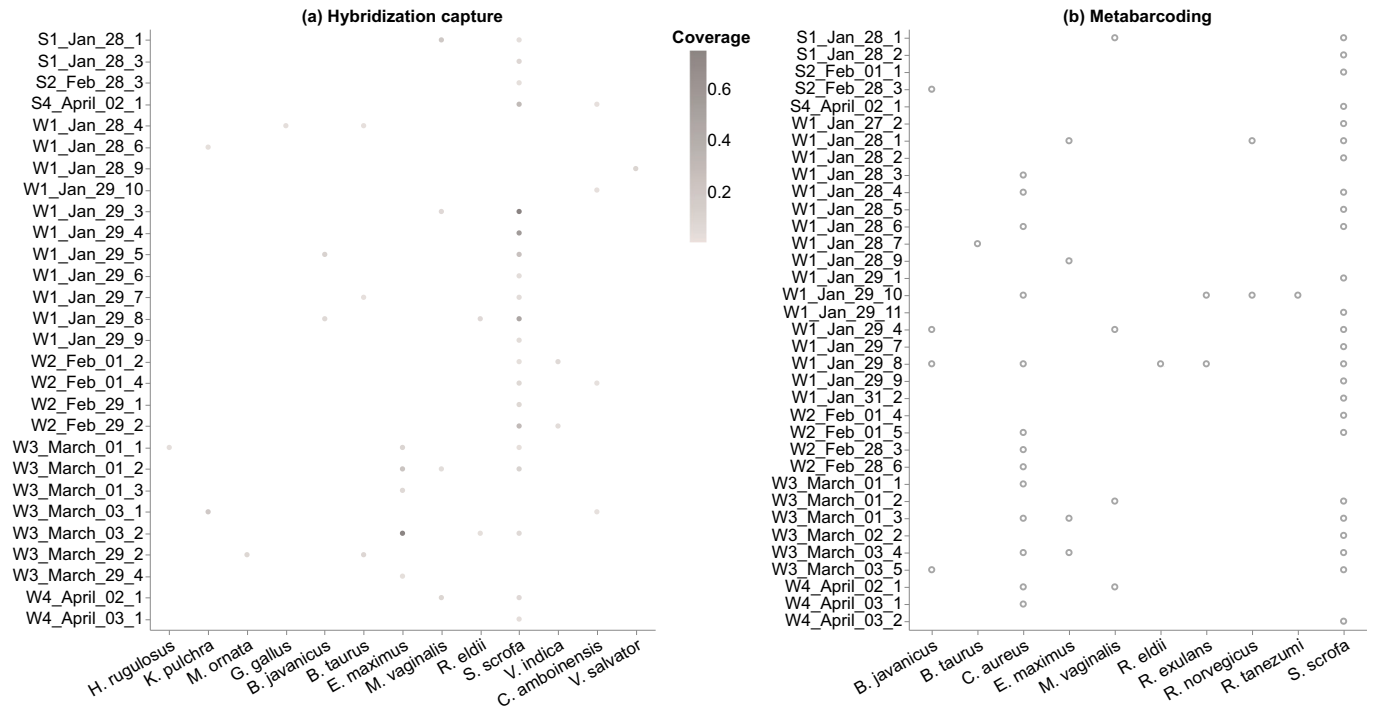


Fig. 4. The dot plots of detected species in respective samples are shown in (a) by hybridization capture and (b) by metabarcoding. The mitogenome coverages (Plot a) of each species reference from capture-based detection were illustrated in a gradient gray from zero to one, corresponding to 0 to 100. Samples were named based on sample types and respective sampling date; the prefix S represents sediment and W represents water, followed by sampling months and respective days.

mouthed toad (*M. ornate*) 8.35 % coverage by 36 reads. The water monitor (*V. salvator*), a reptile, had 10.22 % mitogenome coverage represented by six reads. The sole avian species captured, red junglefowl (*G. gallus*), produced 3 reads covering 3.10 % of the mitogenome.

Blastx results showed that 57.98 % ± 14.79 % of reads (approximately 94,796–2,651,121 reads) of each sample were not assigned to any species. Results indicated that 39.48 % ± 14.89 % (75,604–1,432,006) were assigned to bacteria; 0.30 % ± 0.36 % (11,612,186 reads) were Archaea; 0.31 % ± 0.33 % (15,718,729 reads) were viruses; and 1.59 % ± 0.67 % (214,498,687 reads) were other Eukaryota: Viridiplantae, Fungi, and Metazoa (Fig. S3 a). Blast searches using chordate candidate reads against the NCBI reference nucleotide database showed that 290 reads from 35 samples were assigned to Chordata (at similarity >98 %), comprising the four classes Actinopteri, Amphibia, Lepidosauria, Mammalia (assigned to *Homo sapiens*; Fig. S3 b).

3.3. Metabarcoding

A total of 15,657,307 sequences were generated by metabarcoding from 56 samples (46 water and 10 sediment samples) plus 3 positive in six replicates, at an average of 48,163 ± 15,210 reads (329 ± 82 representative reads) per sample. No reads were generated from negative controls after demultiplexing. PROTAX assigned sequences to eight vertebrate families and seven genera (Fig. S4 and 4). We detected species presence in 22 samples collected from the Visit 1, 9 samples collected from the Visit 2 and 4 samples collected from the Visit 3 (Fig. 1, Fig. 4b and Table 1). In total, ten mammal species were confirmed in 5 sediment samples and 30 water samples. These included: wild boar (*S. scrofa*), Asian elephant (*E. maximus*), red muntjac (*M. vaginalis*), Eld's deer (*R. eldii*), banteng (*B. javanicus*), domesticated cattle (*B. taurus*), little rat (*Rattus exulans*), brown rat (*Rattus norvegicus*), Asian house rat (*Rattus tanezumi*), and *Canis* sp. Wild boar (*S. scrofa*), *Canis sp*, Asian elephant (*E. maximus*), banteng (*B. javanicus*), and red muntjac (*M. vaginalis*) were the most frequently detected species, with 25, 13, 4, 4, and 4 detections, respectively (Table 1). Wild boar (*S. scrofa*), Asian elephant (*E. maximus*), red muntjac (*M. vaginalis*), Eld's deer (*R. eldii*), banteng (*B. javanicus*), and domesticated cattle (*B. taurus*) overlapped with the species detected by hybridization capture.

Blast results assigned the metabarcoding data to 5 orders, 11 classes, 16 families, and 19 genera with above 98 % similarity. Nine families overlapped with the families from PROTAX, three families were exclusively detected by PROTAX (avian family Corvidae, reptile family Geoemydidae, and rodent family Sciuridae), and seven families were exclusively detected by Blast searches (four freshwater fish families: Anabantidae, Channidae, Cyprinidae, Danionidae, and three amphibian (frog) families: Dicroglossidae, Microhylidae, Rhacophoridae) as shown in Fig. S4. Seven genera overlapped with the genera from PROTAX, three genera *Urocissa*, *Leopoldamys*, and *Cuora* were only detected by PROTAX, and 12 genera were exclusively detected by Blast searches (*Channa*, *Esomus*, *Bandicota*, *Kaloula*, *Polypedates*, *Rucervus*, *Barbodes*, *Anabas*, *Microhyla*, *Hoplobatrachus*, *Glyphoglossus*, and *Rasbora*; Fig. S5).

3.4. Method comparison

More samples were positive in species presence by metabarcoding compared with hybridization capture (63 % vs 50 %, shown in Table 1). However, hybridization capture detected a broader spectrum of species (Fig. 3 and Fig. 4). Hybridization capture exclusively detected amphibians (East Asian bullfrog *H. rugulosus* and banded bullfrog *K. pulchra*), reptiles (water monitor *V. salvator* and Amboina box turtle *C. amboinensis*), and one bird (red junglefowl *G. gallus*) (Fig. 4). Whereas, metabarcoding exclusively detected two rodent (little rat *R. exulans* and Asian house rat *R. tanezumi*), and one carnivore species (*Canis* sp.). Hybridization capture detected domesticated cattle in two additional locations compared to metabarcoding. Moreover, Asian elephant (*E. maximus*) and Eld's deer (*R. eldii*) were recorded in one more location than metabarcoding. In contrast, metabarcoding detected wild boar in

Table 3
Results from generalized logistic mixed model estimated the probability of species detection.

	Species detections				
	Estimates	Std. error	Z value	Chi2_LR	P value
Fixed effects					
(Intercept)	−3.40828	0.45426	−7.503		
Methods [metabarcoding]	0.31605	0.248143	1.274	1.6473	0.1993
Species category[Amphibians]	−1.714775	0.874554	−1.961	5.3763	0.1462
Species category [Birds]	−2.023221	1.51938	−1.332		
Species category[Reptiles]	−1.072033	0.954097	−1.124		
Sample type [Sediment]	−0.643074	0.503967	−1.276	1.6625	0.1973
Days	−0.003606	0.007491	−0.481	0.2328	0.6295
Random effects					
Sample	0.95	0.74		110.62	**<0.001*****
Species	1.06	1.03		11.207	**<0.001*****
Locations	2.36E-10	1.54E-05		0	1
ICC	0.38				
Marginal R^2 / Conditional R^2	0.116 / 0.451				

Days: the interval days since the first day when the field sampling started.
Sample: ID used for each sample throughout the lab process.
ICC: intraclass correlation coefficient (estimate of the proportion of variance explained by random effects).

five more locations and banteng in two more locations than hybridization capture (Fig. 4).

GLMM estimated a slightly higher detection probability by metabarcoding, but no significant differences were observed for species detections between the two methods (Fig. 3b and Table 3). The fitted GLMM model was assessed as shown in Fig. S2b.Water samples had a slightly higher detection probability compared with sediment samples. Mammal species had the highest detection probability with both methods, followed by reptiles, amphibians, and birds (Fig. 3). Detection probability was not significantly correlated with sampling date in the dry season (Table 3).

4. Discussion

4.1. Species detection

Over the last few decades, increased efforts to survey biodiversity across Cambodia has provided a greater understanding of wildlife dynamics in the region (Pollard et al., 2008; Gray et al. 2012a; O'Kelly et al., 2012; Gray et al., 2014a; Gray et al., 2014b; Rostro-García et al., 2018; Loveridge et al., 2019; Pin et al., 2020). Nevertheless, longitudinal population trend data are still limited to a small set of species and sites (Groenenberg et al., 2020, Nuttall et al., 2021, Rostro-García et al., 2023) and monitoring data gaps remain for an array of different taxa across the country. Understanding population dynamics are challenging especially for elusive species occurring in low densities (Gray et al. 2012a; Nuttall et al., 2021). A recent assessment of the globally endangered Eld's deer (*R. eldii*) emphasized the absence of robust data for this large charismatic globally endangered species (Ladd et al., 2022). Therefore, wildlife conservation and management will require technological developments to complement observation-based techniques and fulfill the needs of an increased scale and frequency of biodiversity monitoring. In the current study, by applying metabarcoding and hybridization capture to water and sediment samples, we detected five classes of Chordata: Actinopteri, Amphibia, Aves, Lepidosauria, and Mammalia. With the exception of the Actinopteri class, other fauna could be assigned to the species level. Overall, seventeen species were identified, four of which were listed as endangered in the IUCN (International Union for Conservation of Nature) Red List: banteng (*B. javanicus*), Asian elephant (*E. maximus*), Eld's deer (*R. eldii*), and Amboina box turtle (*C. amboinensis*).

Ungulates were the most frequently detected species in the present study, by both hybridization capture and metabarcoding. Wild boar (*S. scrofa*) was the most frequently observed ungulate species followed by red muntjac (*M. vaginalis*), banteng (*B. javanicus*), and Eld's deer (*R. eldii*) in decreasing frequency. The relative detection frequency of the four ungulates was consistent with 2016 camera trap results (Pin et al., 2020). Overall, fewer species were observed by eDNA methods in the present study compared to the camera trap records. Waterhole samples were collected on multiple occasions, however both water and sediment samples were in low-volume and collected at a single time point. In contrast, the camera traps were set at the waterhole edge with the highest diversity of wildlife footprints. Photographical recordings were operated 24 h per day until the corresponding waterholes receded completely throughout the entire survey period (Pin et al., 2020). Nevertheless, frequency of species detected by eDNA methods and camera traps was consistent thus indicating that eDNA can reliably reflect ungulate species presence, even for low-density species such as Eld's deer (*R. eldii*). Despite a recent increase in conservation and enforcement efforts in Cambodia, the assemblage of globally threatened ungulate populations, even previously abundant species (e.g. red muntjac *M. vaginalis*), have suffered dramatic declines (Gray et al., 2012b, Groenenberg et al., 2020). As species densities decrease, observation-based monitoring will likely become less effective (Gray et al. 2012a, Groenenberg et al., 2020) requiring other methods such as DNA based approaches to complement them. Results from this study emphasize that eDNA from waterholes can provide invaluable information on presence and distribution of rare species occurring in low densities.

Asian elephant (*E. maximus*) populations have dramatically declined across several countries in Southeast Asia due to a range of anthropogenic factors (Menon and Tiwari, 2019). Previous non-invasive DNA surveys suggested that eastern Cambodia likely supports >300 individuals (Pollard et al., 2008; Gray et al., 2014b). This eDNA study detected Asian elephants (*E. maximus*) in five samples by capture and four by metabarcoding. While direct comparison of DNA methods with camera trapping is difficult with regard to comparability of different spatial and temporal sampling patterns, detection frequency recorded in this study exceeded those of the concurrent camera trapping records (recorded in two waterholes) (Pin et al., 2020). Elephants access water bodies at a high frequency, which contributes to the high likelihood of retrieving their DNA from waterhole samples (Dejean et al., 2011; Laramie et al., 2015). These results indicate that eDNA methods may be particularly useful for animals which shed abundant DNA into their environment and frequently visit waterholes but which may still be difficult to monitor directly. A set of genetic markers developed for use with elephant fecal samples achieved relatively precise population estimates and can discriminate individual animals at comparable rates to

observational surveys (Pollard et al., 2008; Maltby and Bourchier, 2011; Gray et al., 2014b). Considering the large amount of shed elephant DNA, inclusion of eDNA facilitate population structure determination, diversity and sex ratios.

Transect and camera traps surveys implemented in eastern Cambodia have provided an invaluable insight into the population status of a different species (Gray et al., 2014a, Rostro-García et al., 2018, Groenenberg et al., 2020, Nuttall et al., 2021). However, these surveys are often designed and biased towards larger species (MacKenzie et al., 2002) and detectability can be biased depending on survey locations (Gray, 2013; Rostro-García et al., 2018). In general, reptiles, amphibians and other small mammal species are often difficult to detect or distinguish morphologically in transect or camera trap surveys (Raemy and Ursenbacher, 2018; Matthias et al., 2021). In the current study, eDNA methods detected ten animals with small body size, constituting four mammals, three amphibians, two reptiles, and one bird species: seven by hybridization capture and three by metabarcoding. Eight out of the ten animals were exclusively detected by eDNA methods (three rodents, three amphibian and two reptiles) compared with the respective camera trapping, indicating eDNA detection has the potential to identify many more species and fill in monitoring and surveillance gaps.

Camera trap surveys conducted during the same timeframe and locations by Pin et al., 2020 recorded nine bird species, four of which were photographed at more than fifteen waterholes. In contrast, eDNA results from this study only identified one bird species (red junglefowl *G. gallus*). As of the time of writing, three of the nine bird species (red junglefowl *G. gallus*, green peafowl *Pavo muticus*, and sarus crane *Antigone Antigone*) identified in the camera traps study have full-mitogenome records and one (lesser adjutant *Leptoptilos javanicus*) has available COX1 sequences. The lack of mitogenome records in the database may explain some of the discrepancy. Moreover, all the hybridization capture baits and metabarcoding primers were designed to target mammals, which make the two eDNA methods less optimized for non-mammal species. Further optimization of the hybridization capture bait set and metabarcoding primers could improve detection of non-mammalian taxa, which would allow eDNA methods to be applied to target various taxa and across different landscapes.

4.2. Methodology

Metabarcoding methods are more economic and discriminating than shotgun sequencing, and therefore are preferred in eDNA studies (Bista et al., 2018; Bell et al., 2021; Rubiola et al., 2022; Seeber and Epp, 2022). However, specific target amplification may exclude substantial genetic information and miss species due to primer sequence mismatch (Bourret et al., 2020; Seeber and Epp, 2022). Hybridization capture enrichment before shotgun sequencing, while less sensitive than PCR-based methods such as metabarcoding, can yield considerable genomic information (Jensen et al., 2021). In the current study, metabarcoding and hybridization capture detected 27.6 % of the species recorded in the concurrent camera trap survey conducted by Pin et al., 2020. However, nine additional species were exclusively detected by eDNA methods: one ungulate (domesticated cattle *B. taurus*), three rodents, two reptiles, and three amphibians. Among them, only hybridization capture enrichment detected non-mammalian species, albeit at low rates in the capture simulation by Capsim, indicating high tolerance for sequence divergence. The small Indian civet (*V. indica*) photographed at six waterhole locations was also detected by hybridization capture at two waterholes, but not by the metabarcoding approach. None of the sequences identified by hybridization capture were located within the 16S rRNA region that the metabarcoding primers targeted. A single short target is an inherent drawback of PCR-based methods in eDNA detection and may be a cause of false negative detections (metabarcoding method not detecting small Indian civet). In contrast, hybridization capture targets a wider range of sequence regions, thus providing more opportunities for detection and species discrimination. For eDNA metabarcoding, detection is also limited by DNA degradation especially under extreme conditions found in tropical regions. However, when large amounts of DNA are shed (e.g. by wild boar and Asian elephants), metabarcoding may be more sensitive than hybridization capture. This may be attributed to the exponential amplification step by metabarcoding, which increased the detectability of animals shedding more DNA in a given environment. On the other hand, PCR amplification can potentially be initiated by a single target DNA sequence that may be missed in hybridization capture. Golden jackal (*Canis aureus*) was photographed at six waterholes, although using eDNA methods, results could only identify *Canis* sp. at the genus level in thirteen samples by metabarcoding. The inability to define the species by metabarcoding is likely due to limitations of the 16S reference database. Three *Rattus* species were detected by metabarcoding but not by hybridization capture or camera traps. None of the three rodents identified has full mitogenome records in the public databases. Lack of reference sequences may explain discrepancies between metabarcoding and hybridization capture, and camera traps. Overall, both methods are sensitive enough to detect species presence even when little DNA is present. However, the absence of reference sequences substantially influence the ability to reliably identify all fauna to a species level species regardless of method applied. Therefore, improving reference databases will be fundamental to future eDNA applications.

Noninvasive genetic sampling has been widely used for population genetic studies (Taberlet et al., 1993; Eggert et al., 2003; Pollard et al., 2008). In the current study, hybridization capture enrichment retrieved 74.52 % coverage of the Asian elephant (*E. maximus*) mitogenome, 75.36 % coverage of the wild boar (*S. scrofa*) mitogenome, and over 10 % coverage of the mitogenomes for banteng (*B. javanicus*), red muntjac (*M. vaginalis*), water monitor (*V. salvator*), and banded bullfrog (*K. pulchra*). This indicates eDNA hybridization capture can yield multilocus or even whole mitogenomic data for terrestrial animals. We chose to sample waterholes during the dry season when animals tend to congregate. Despite lower base coverage compared with aquatic animals (Sigsgaard et al., 2017; Jensen et al., 2021), the retrieved mitogenomic DNA reached 9-fold base coverage for Asian elephant (*E. maximus*) and 6-fold for wild boar (*S. scrofa*). We used comparatively small sample volumes (50 mL) to previous studies that isolated terrestrial mammal eDNA from 500 mL water (Ushio et al., 2017; Broadhurst et al., 2021) or 2 L (Harper et al., 2019; Sales et al., 2020). Thus, to further improve on the results of the current study in terms of total read numbers and genomic coverage, larger volumes of water, while more difficult to transport unless filtering can be performed on site, might yield better coverage. Dependent upon the survey objective, target capture could use a more specific bait panel for selected target taxa that remove redundant and highly conserved loci. Finally, sequencing to higher depth would likely improve coverage for the hybridization capture-based approach. Further methodological refinements could expand the species detected, increase the amount of genomic information obtainable and improve the implementation of such techniques in the field.

4.3. Sample type effects

DNA can bind to sediments and persist in the environment for long periods. By contrast, water likely reflects short-term species presence, either before degradation occurs or DNA settles in the sediments (Turner et al., 2015). Thus, it was surprising that significant differences between water and sediment samples were not observed in terms of species detection frequencies, and more species were detected from water samples. Aqueous eDNA is generally distributed heterogeneously in water bodies at low concentrations (Takahara et al., 2012; Pilliod et al., 2013). Multiple samplings per location are preferable for obtaining representative water samples. In general, sediment DNA extraction is performed on a small amount of sediment that is spatially restricted and which may not be as representative of water body diversity as surface water sampled from multiple sites.

The decreasing number of waterholes from the beginning to the end of dry season did not significantly change the species detection probability. DNA concentration and integrity did not affect species detection for either of the approaches used (hybridization capture or metabarcoding). This indicates that both eDNA methods are relatively robust even when DNA is likely degraded. Significant differences were not observed in species detection at different time points in the dry season. In principle, increasing scarcity of water resources should lead to a higher number of species utilizing the limited waterholes. However, anecdotal reports suggested that several of the waterholes dried out earlier than previous years. Climate and water availability could therefore influence species distribution and range. Consequently, eDNA monitoring of waterholes alone may be less feasible for some species. Expansion of waterholes (rather than repeat visits) and inclusion of other waterbodies, such as perennial rivers may lead to an increase in number of species identified. Water-based eDNA can be also be complemented by obtaining eDNA from other sources such as invertebrates which could provide more comprehensive terrestrial mammal monitoring options (Drinkwater et al., 2021). The benefit of being able to detect small bodied animals and ease of sampling may outweigh some of the limitations imposed by the heterogeneity of environment, animal behavior, and distributions.

5. Conclusion

We demonstrated that eDNA can be used for detecting a diverse array of species from waterhole samples in Cambodian tropical forests. Both hybridization capture and metabarcoding methods identified species that were not detected and or reported by observational approaches applied previously. Hybridization capture enrichment before sequencing is similarly robust to metabarcoding with respect to species detection. However, hybridization capture enrichment can detect a wider range of taxa and provides far more genomic information, whereas metabarcoding is more sensitive. With respect to laboratory, labor and time expenditure, metabarcoding (ca. 50 euro/sample) costs are approximately half as much as hybridization capture costs and it takes about one third of the wet lab time. If specific target species are known in advance, metabarcoding may be the preferable approach. If faunal assemblages are unclear, hybridization capture-based approaches may be preferable if only individual methods can be employed. In many conservation projects spatial information, species identification and in some cases genomic data are vital. Incorporating both methods, particularly in ecosystems that are not amenable to observational approaches should be considered for broad-scale monitoring. While the target substrate of this study was water and waterhole sediments, other alternative eDNA or invertebrate DNA sources could complement or replace water in other environmental or conservation contexts (Axtner et al., 2019; Nguyen et al., 2021). The same molecular approaches could be applied in these contexts and would likely exhibit similar benefits and restrictions observed in the current study for water while expanding the available resources for wildlife detection and monitoring.

CRediT authorship contribution statement

ADG, RC and AK developed the project. JL and PS developed the laboratory workflow for hybridization capture. JA developed the laboratory workflow for metabarcoding and implemented the bioinformatics analysis. JL implemented the entire experiments, bioinformatics analysis of hybridization capture data, statistical analysis and data visualization. AC provided statistical expertise. RC, and PC collected samples and oversaw project co-ordination and implementation. RC, MG, AK assisted with the project arrangements and funding support. RC and MG provided the expertise in Cambodian wildlife conservation. JL, ADG, and PS wrote the article. All authors discussed the project, data, results and have read and approved the final manuscript.

Declaration of competing interest

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

Data availability

Sequencing data generated in the current study are available in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA943158 and PRJNA933918.

Acknowledgements

The authors thank the Mondulkiri Provincial Department of Environment, the Ministry of Environment, and the Forestry Administration of the Royal Government of Cambodia for enabling and supporting this study. We are extremely grateful to all government and community researchers that collected waterhole samples throughout the survey period. This work was financially supported by the WWF. We also thank Anke Schmidt and Karin Hönig for laboratory technical assistance, Dorina Meneghini for bioinformatics support and Carin Hoffmann for administrative support. JL acknowledges the support of the China Scholarship Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2023.110168.

References

Axtner, J., Crampton-Platt, A., Horig, 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.

Bell, K.L., Petit, R.A., Cutler, A., Dobbs, E.K., Macpherson, J.M., Read, T.D., Burgess, K.S., Brosi, B.J., 2021. Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. Ecol. Evol. 11, 16082–16098.

Bista, I., Carvalho, G.R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S.L., Christmas, M., Creer, S., 2018. Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. Mol. Ecol. Resour. 18, 1020–1034.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. 29, 358–367.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bourret, V., Albert, V., April, J., Cote, G., Morissette, O., 2020. Past, present and future contributions of evolutionary biology to wildlife forensics, management and conservation. Evol. Appl. 13, 1420–1434.

Broadhurst, H.A., Gregory, L.M., Bleakley, E.K., Perkins, J.C., Lavin, J.V., Bolton, P., Browett, S.S., Howe, C.V., Singleton, N., Tansley, D., Sales, N.G., McDevitt, A.D., 2021. Mapping differences in mammalian distributions and diversity using environmental DNA from rivers. Sci. Total Environ. 801, 149724.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60.

Cao, M.D., Ganesamoorthy, D., Zhou, C., Coin, L.J.M., 2018. Simulating the dynamics of targeted capture sequencing with CapSim. Bioinformatics 34, 873–874.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., Miaud, C., 2011. Persistence of environmental DNA in freshwater ecosystems. PLoS One 6, e23398.

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., 2021. Leech blood-meal invertebrate-derived DNA reveals differences in Bornean mammal diversity across habitats. Mol. Ecol. 30, 3299–3312.

Duckworth, J.W.H.S., 1998. Tracking Tigers: A Review of the Status of Tiger, Asian Elephant, Gaur, and Banteng in Vietnam, Lao, Cambodia, and Yunnan (China), With Recommendations for Future Conservation Action. WWF Indochina, Hanoi, Vietnam, 282pp.

Eggert, L.S., Eggert, J.A., Woodruff, D.S., 2003. Estimating population sizes for elusive animals: the forest elephants of Kakum National Park, Ghana. Mol. Ecol. 12, 1389–1402.

Farrell, M.J., Govender, D., Hajibabaei, M., M. van der Bank, Davies, T.J., 2022. Environmental DNA as a management tool for tracking artificial waterhole use in savanna ecosystems. Biol. Conserv. 274.

Forrest, J.L., Bomhard, B., Budiman, A., Coad, L., Cox, N., Dinerstein, E., Hammer, D., Huang, C., Huy, K., Kraft, R., Lysenko, I., Magrath, W., 2011. Single-species conservation in a multiple-use landscape: current protection of the tiger range. Anim. Conserv. 14, 283–294.
Gasc, C., Peyretaillade, E., Peyret, P., 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. Nucleic Acids Res. 44, 4504–4518.
Giebner, H., Langen, K., Bourlat, S.J., Kukowka, S., Mayer, C., Astrin, J.J., Misof, B., Fonseca, V.G., 2020. Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. Mol. Ecol. Resour. 20, 1333–1345.
Gray, T.N.E., 2013. Activity patterns and home ranges of Indochinese leopard Panthera pardus delacouri in the Eastern Plains landscape, Cambodia. Nat. Hist. Bull. Siam Soc. 59, 10.
Gray, T.N., Phan, C., Pin, C., Prum, S., 2012a. Establishing a monitoring baseline for threatened large ungulates in eastern Cambodia. Wildl. Biol. 18, 406–413.
Gray, T.N.E., Prum, S., Pin, C., Phan, C., 2012b. Distance sampling reveals Cambodia's Eastern Plains landscape supports the largest global population of the endangered banteng Bos javanicus. Oryx 46, 563–566.
Gray, T.N.E., Pin, C., Phan, C., Crouthers, R., Kamler, J.F., Prum, S., 2014a. Camera-trap Records of Small Carnivores from Eastern Cambodia, 1999–2013.
Gray, T.N.E., Vidya, T.N.C., Potdar, S., Bharti, D.K., Sovanna, P., 2014b. Population size estimation of an Asian elephant population in eastern Cambodia through non-invasive mark-recapture sampling. Conserv. Genet. 15, 803–810.
Gray, T.N.E., McShea, W.J., Koehncke, A., Sovanna, P., Wright, M., 2015. Artificial deepening of seasonal waterholes in eastern Cambodia: impact on water retention and use by large ungulates and waterbirds. J. Threat. Taxa 7, 7189–7195.
Gray, T.N.E., Lynam, A.J., Seng, T., Laurance, W.F., Long, B., Scotson, L., Ripple, W.J., 2017. Wildlife-snaring crisis in Asian forests. Science 355, 255–256.
Groenenberg, M.C., Yoganand, R., K., 2020. Population status of ungulates in the Eastern Plains landscape. In: Srepok Wildlife Sanctuary and Phnom Prich Wildlife Sanctuary, Cambodia. WWF-Cambodia, Phnom Penh, Cambodia.
Gu, Z., Gu, L., Eils, R., Schlesner, M., Brors, B., 2014. Circlize implements and enhances circular visualization in R. Bioinformatics 30, 2811–2812.
Harper, L.R., Handley, L.L., Carpenter, A.I., Ghazali, M., Di Muri, C., Macgregor, C.J., Logan, T.W., Law, A., Breithaupt, T., Read, D.S., McDevitt, A.D., Hanfling, B., 2019. Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. Biol. Conserv. 238, 108225.
Hartig, F., 2022. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R Package Version 0.4.6.
Hughes, A.C., 2017. Understanding the drivers of southeast Asian biodiversity loss. Ecosphere 8.
Ibbett, H., Keane, A., Dobson, A.D.M., Griffin, O., Travers, H., Milner-Gulland, E.J., 2021. Estimating hunting prevalence and reliance on wild meat in Cambodia's Eastern Plains. Oryx 55, 878–888.
Jensen, M.R., Sigsgaard, E.E., Liu, S., Manica, A., Bach, S.S., Hansen, M.M., Moller, P.R., Thomsen, P.F., 2021. Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. Mol. Ecol. Resour. 21, 690–702.
Jerde, C.L., Mahon, A.R., Chadderton, W.L., Lodge, D.M., 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv. Lett. 4, 150–157.
Kielbasa, S.M., Wan, R., Sato, K., Horton, P., Frith, M.C., 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487–493.
Ladd, R., Crouthers, R., Brook, S., Eames, J.C., 2022. Reviewing the status and demise of the endangered Eld's deer and identifying priority sites and conservation actions in Cambodia. Mammalia 86, 407–421.
Langdon, W.B., 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min. 8, 1.
Laramie, M.B., Pilliod, D.S., Goldberg, C.S., 2015. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. Biol. Conserv. 183, 29–37.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., S. Genome Project Data Processing, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Loucks, C., Mascia, M.B., Maxwell, A., Huy, K., Duong, K., Chea, N., Long, B., Cox, N., Seng, T., 2009. Wildlife decline in Cambodia, 1953-2005: exploring the legacy of armed conflict. Conserv. Lett. 2, 82–92.
Loveridge, R., Ryan, G.E., Sum, P., Grey-Read, O., Mahood, S.P., Mould, A., Harrison, S., Crouthers, R., Ko, S., Clements, T., Eames, J.C., Pruvot, M., 2019. Poisoning causing the decline in South-East Asia's largest vulture population. Bird Conserv. Int. 29, 41–54.
Lyet, A., Pellissier, L., Valentini, A., Dejean, T., Hehmeyer, A., Naidoo, R., 2021. eDNA sampled from stream networks correlates with camera trap detection rates of terrestrial mammals. Sci. Rep. 11, 11362.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255.
Magoc, T., Salzberg, S.L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963.
Maltby, M., Bourchier, G., 2011. Current Status of Asian Elephants in Cambodia.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.
Matthias, L., Allison, M.J., Maslovat, C.Y., Hobbs, J., Helbing, C.C., 2021. Improving ecological surveys for the detection of cryptic, fossorial snakes using eDNA on and under artificial cover objects. Ecol. Indic. 131.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Mena, J.L., Yagui, H., Tejeda, V., Bonifaz, E., Bellemain, E., Valentini, A., Tobler, M.W., Sanchez-Vendizu, P., Lyet, A., 2021. Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. Ecol. Appl. 31, e02335.
Menon, V., Tiwari, S.K., 2019. Population status of Asian elephants Elephas maximus and key threats. Int. Zoo Yearb. 53, 17–30.
Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. In: Cold Spring Harb Protoc 2010 pdb prot5448.
Morgulis, A., Gertz, E.M., Schäffer, A.A., Agarwala, R., 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences, 13, 1028–1040.
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. Environ. DNA 3, 780–791.
Nuttall, M.N., Griffin, O., Fewster, R.M., McGowan, P.J.K., Abernethy, K., O'Kelly, H., Nut, M., Sot, V., Bunnefeld, N., 2021. Long-term monitoring of wildlife populations for protected area management in Southeast Asia. Conserv. Sci. Pract. 4, e614.
O'Kelly, H.J., Evans, T.D., Stokes, E.J., Clements, T.J., Dara, A., Gately, M., Menghor, N., Pollard, E.H., Soriyun, M., Walston, J., 2012. Identifying conservation successes, failures and future opportunities; assessing recovery potential of wild ungulates and tigers in Eastern Cambodia. PLoS One 7, e40482.
Parsons, K.M., Everett, M., Dahlheim, M., Park, L., 2018. Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. R. Soc. Open Sci. 5, 180537.
Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P., 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. Can. J. Fish. Aquat. Sci. 70, 1123–1130.
Pin, C., Ngoprasert, D., Gray, T.N.E., Savini, T., Crouthers, R., Gale, G.A., 2020. Utilization of waterholes by globally threatened species in deciduous dipterocarp forest of the Eastern Plains Landscape of Cambodia. Oryx 54, 572–582.
Pollard, E., Eggert, L., Chanvibol, C., Hedges, S., 2008. The Status and Conservation of Asian Elephants in the Seima Biodiversity Conservation Area, Cambodia. WCS Cambodia Program, Phnom Penh.
Poole, C.M., Duckworth, J.W., 2005. A documented 20th century record of Javan Rhinoceros Rhinoceros sondaicus from Cambodia. Mammalia 69, 443–444.
Quinlan, A.R., 2014. BEDTools: the Swiss-Army tool for genome feature analysis. Curr. Protoc. Bioinformatics 47 (11 12), 11–34.
Raemy, M., Ursenbacher, S., 2018. Detection of the European pond turtle (Emys orbicularis) by environmental DNA: is eDNA adequate for reptiles? Amphibia-Reptilia 39, 135–143.
Rostro-García, S., Kamler, J.F., Crouthers, R., Sopheak, K., Prum, S., In, V., Pin, C., Caragiulo, A., Macdonald, D.W., 2018. An adaptable but threatened big cat: density, diet and prey selection of the Indochinese leopard (Panthera pardus delacouri) in eastern Cambodia. R. Soc. Open Sci. 5, 171187.
Rostro-García, S., Kamler, J.F., Sollmann, R., Balme, G., Augustine, B.C., Kéry, M., Crouthers, R., Gray, T.N.E., Groenenberg, M., Prum, S., Macdonald, D.W., 2023. Population dynamics of the last leopard population of eastern Indochina in the context of improved law enforcement. Biol. Conserv. 283.
Rousset, F., Ferdy, J.B., 2014. Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography 37, 781–790.
Rubiola, S., Macori, G., Civera, T., Fanning, S., Mitchell, M., Chiesa, F., 2022. Comparison between full-length 16S rRNA metabarcoding and whole metagenome sequencing suggests the use of either is suitable for large-scale microbiome studies. Foodborne Pathog. Dis. 19, 495–504.
Sales, N.G., McKenzie, M.B., Drake, J., Harper, L.R., Browett, S.S., Coscia, I., Wangensteen, O.S., Baillie, C., Bryce, E., Dawson, D.A., Ochu, E., Haenfling, B., Handley, L.L., Mariani, S., Lambin, X., Sutherland, C., McDevitt, A.D., 2020. Fishing for mammals: landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. J. Appl. Ecol. 57, 707–716.
Seeber, P.A., Epp, L.S., 2022. Environmental DNA and metagenomics of terrestrial mammals as keystone taxa of recent and past ecosystems. Mammal Rev. 52, 538–553.
Seeber, P.A., McEwen, G.K., Lober, U., Forster, D.W., East, M.L., Melzheimer, J., Greenwood, A.D., 2019. Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. Mol. Ecol. Resour. 19, 1486–1496.
Sigsgaard, E.E., Nielsen, I.B., Bach, S.S., Lorenzen, E.D., Robinson, D.P., Knudsen, S.W., Pedersen, M.W., Jaidah, M., Orlando, L., Willerslev, E., Moller, P.R., Thomsen, P.F., 2017. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nat. Ecol. Evol. 1.
Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., Ovaskainen, O., 2016. Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics 32, 2920–2927.
Taberlet, P., Mattock, H., Duboispaganon, C., Bouvet, J., 1993. Sexing free-ranging brown bears ursus-arctos using hairs found in the field. Mol. Ecol. 2, 399–403.
Tak, C., Crouthers, R., Sukumal, N., Chhin, S., Savini, T., 2022. Importance of Srepok Wildlife Sanctuary, Cambodia, for the endangered green peafowl: implications of co-occurrence near human use areas. Raffles Bull. Zool. 70, 249–256.
Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., Kawabata, Z., 2012. Estimation of fish biomass using environmental DNA. PLoS One 7, e35868.

Taylor, P.G., 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Mol. Biol. Evol. 13, 283–285.
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. Commun. Biol. 2, 396.
Turner, C.R., Uy, K.L., Everhart, R.C., 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. Biol. Conserv. 183, 93–102.
Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., Miya, M., 2017. Environmental DNA enables detection of terrestrial mammals from forest pond water. Mol. Ecol. Resour. 17, e63–e75.
Wharton, C.H., 1957. An Ecological Study of the Kouprey (Novibos sauveliUrbain). Monographs of the Institute of Science and Technology, Monograph 5, Manila, Philippines, vol 111, pp. 91–93.
Wright, H.L., Ko, S., Norin, N., Phearun, S., 2013. White-shouldered Ibis Pseudibis davisoni population size and the impending threat of habitat conversion. Forktail 29, 162–165.
Phan, C.P., Prum, S., Gray, T.N.E., 2010. Recent camera trap records of globally threatened species from the Eastern Plains Landscape, Mondulkiri. Cambodian. J. Nat. Hist. 2010 (2), 89–93.