Environmental DNA for tracking waterhole visitation in savanna ecosystems

Maxwell J. Farrell[1*], Danny Govender[2,3],
Mehrdad Hajibabaei[4], Michelle van der Bank[4,5], T. Jonathan Davies[5,6]

[1]Ecology & Evolutionary Biology, University of Toronto
[2]SANParks Scientific Services
[3]University of Pretoria
[4]University of Guelph
[5]African Centre for DNA Barcoding, University of Johannesburg, South Africa
[6]Botany, Forest & Conservation Sciences, University of British Columbia
[*]To whom correspondence should be addressed; E-mail: maxwell.farrell@utoronto.ca

November 3, 2020

Abstract

The analysis of waterborne environmental DNA (eDNA) is effective for detecting invasive species and conducting large-scale biodiversity assessments, making it a potentially powerful tool for documenting diversity at sites where large numbers of species aggregate. We explore the utility of eDNA from waterholes for describing local mammal communities, quantifying patterns of species co-occurrences, and monitoring of rare or threatened species. In savanna ecosystems water can be a scarce resource during dry seasons and in periods of drought, promoting the aggregation of medium to large mammals. To explore the reliability of eDNA as a biodiversity indicator in these arid and semi-arid environments, we compare eDNA metabarcoding and camera traps for documenting waterhole use by mammals in the Kruger National Park, South Africa. We find that eDNA metabarcoding can recover the majority of mammal species detected in camera traps, but the DNA signatures of mammal visitation are temporally limited. Detection rates varied across sites, sampling time, species, and choice of reference library, with the best performance for water-dependent large bodied mammals visiting within two days of sampling, and matched to a curated system-specific reference library. Our results demonstrate that eDNA-based approaches can be used to track mammals of conservation concern, and reflect patterns of recent waterhole use and co-occurrence across water-dependent species, but also highlight limitations including the lack of long-term eDNA persistence in small and highly utilized waterholes and variability in detection rates among species. Sequencing of eDNA is a valuable tool for next-generation biodiversity sampling and has many exciting applications, but it is not sufficient to capture long-term waterhole visitation patterns or reliably detect rare and small-bodied species.

Introduction

The advent of next generation sequencing, coupled with comprehensive genomic reference libraries and techniques for isolating DNA from environmental samples (eDNA) is revolutionizing traditional biodiversity surveys (Bohmann et al., 2014). eDNA based approaches have been successful for targeted monitoring of invasive, rare, and threatened species (Rees et al., 2014). Screening of species via conserved marker genes, known as "metabarcoding" (Taberlet et al., 2012), is used for monitoring organisms from freshwater, marine, and terrestrial environments (Deiner et al., 2017), and allows for the detection of multiple species with a single sample. eDNA metabarcoding may be useful for describing local communities, quantifying patterns of species co-occurrence, and monitoring of endangered species, but its efficacy can be highly context dependent (Ushio et al., 2017; Seeber et al., 2019; Harrison et al., 2019; Sales et al., 2020; Leempoel et al., 2020). Here, we evaluate the performance of eDNA sampled from waterholes relative to species visitations documented by camera traps, and consider its application as a biodiversity survey tool to both track individual mammal species presence as well as their co-occurrences in a savanna ecosystem of high conservation value.

In savannas the availability of surface water for drinking and thermoregulation is a major structuring force for communities of medium and large mammals (Redfern et al., 2005). These essential resources become seasonally restricted in arid environments, and lack of water close to available forage can lead to mass die-offs and shifting distributions in periods of severe drought (Smit et al., 2020). As vital resources for multiple species, waterholes are sites of aggregation for mammals, which can modify predator-prey interactions (Valeix et al., 2009; Amoroso et al., 2020), parasite transmission (Turner et al., 2016; Franz et al., 2018), and drive large-scale shifts in community composition (Redfern et al., 2005; Smit et al., 2007). Waterholes are thus prime targets for biodiversity monitoring and eDNA sampling. Sequencing of eDNA present in waterholes and ponds has been successfully used for surveying terrestrial mammals in sub-Saharan Africa (Seeber et al., 2019) and Japan (Ushio et al., 2017), multiple vertebrates in Australian arid zones (Furlan et al., 2020) including the endangered Gouldian finch (*Erythrura gouldiae*) (Day et al., 2019), and both semi-aquatic and terrestrial mammals in the United Kingdom (Harper et al., 2019). Despite the promise of this approach, there remain many questions about its limitations, including the physical properties of eDNA that influence transport and decay (Harrison et al., 2019), best practices for limiting species identification errors due to molecular and bioinformatic workflows (Furlan et al., 2020), choice of molecular reference library for metabarcoding approaches (Porter and Hajibabaei, 2018b), and how eDNA biomonitoring compares to other emerging methods for biodiversity surveys (Schnell et al., 2015; Sales et al., 2020; Leempoel et al., 2020; Gogarten et al., 2020).

To explore the utility of waterhole eDNA metabarcoding as a tool for biodiversity surveys, we conduct a comparative study of eDNA metabarcoding and camera trapping for documenting mammal diversity at waterholes in a mammal-rich African savanna. Compared to eDNA approaches, camera traps have a longer history of use by wildlife researchers, but have seen increased popularity in recent years due to improved picture quality, decreased costs of modern digital cameras, and progress in citizen science and machine learning for species identification (Norouzzadeh et al., 2018). Camera trap surveys are now being conducted at larger scales and with dramatically increased detection rates (Swanson et al., 2015). Most often targeting mammals, when camera traps are employed in conjunction with a proper survey design and statistical model, they can gather a range of ecological data including relative abundances, activity patterns, diet, identification of individuals, population density estimates, and distribution patterns through informing habitat occupancy models (Burton et al., 2015). Although camera trapping and eDNA based approaches both have recognised strengths and weaknesses (Harper et al., 2019; Leempoel et al., 2020; Sales et al., 2020; Gogarten et al., 2020), they have not yet been formally contrasted in savanna ecosystems, which have been the focus of perhaps one of the most intensive co-ordinated global camera trap monitoring efforts (Swanson et al., 2015).

We conduct our study in the the Kruger National Park, South Africa (KNP), which harbours excep-

tionally high mammal and bird diversity, and is a key conservation site for a number of threatened species including wild free-ranging white rhinoceros (*Ceratotherium simum*) (Ferreira et al., 2017), a species of global conservation concern. The KNP is among the oldest game reserves in southern Africa, and contains a diversity of geologies, topographies, and vegetation types, and is home to 147 mammal species (Parr et al., 2009), including many large carnivores: cheetah (*Acinonyx jubatus*), lions (*Panthera leo*), leopards (*Panthera pardus*), and African wild dogs (*Lycaon pictus*), which are threatened by persecution (Swanepoel et al., 2016), poaching (Everatt et al., 2019), and infection by multi-host pathogens such as *Mycobacterium bovis* (Caron et al., 2003; Higgitt et al., 2019). Poaching and infectious diseases also pose severe threats for large herbivores in the park, such as critically endangered black rhinoceros (*Diceros bicornis*) (Ferreira et al., 2018), hippopotamus (*Hippopotamus amphibius*), elephant (*Loxodonta africana*), and giraffe (Bengis and Erasmus, 1988). Conceived in the era of big-game hunting, the KNP and has a long history of water provisioning, including the construction of a series of artificial waterholes beginning in the 1930s (Redfern et al., 2005). These waterholes were intended to increase game numbers by stabilizing water availability year-round and are frequently visited by a diversity of birds and mammals. However, these artificial water sources have resulted in the expansion and overabundance of water-dependent species such as impala (*Aepyceros melampus*), elephants, and zebra (*Equus burchellii*), negatively impacting rarer drought-tolerant species (sable and roan antelopes), resulting in many being decommissioned since the 1990s (Smit et al., 2007).

With support from South African National Parks (SANParks), we conducted focused biodiversity surveys using camera trap and eDNA approaches to describe the mammal communities associated with waterholes in the KNP. Using camera trapping, we can observe the species that visit each waterhole, and quantify relative frequencies of visitation for species at each site. By pairing these data with COI sequences amplified from eDNA present in the waterholes, we can determine the sensitivity of eDNA to capture species visitation patterns inferred from the cameras, and identify the methodological, environmental and ecological factors that influence the efficacy of eDNA-based approaches. Our results allow us to make recommendations for the use of these approaches for documenting mammal diversity via waterholes. Here we assess the utility of eDNA matabarcoding in a semi-controlled environment, halfway between mock communities and surveys of natural water bodies. In addition to describing diversity as measured by species richness, we examine the ability of eDNA metabarcoding to document species interactions via waterhole visitation.

Materials & Methods

Study Site

We surveyed the mammalian communities associated with waterholes in the Kruger National Park (KNP), South Africa, a large protected savanna ecosystem and a global hotspot for large mammal diversity (Parr et al., 2009). Intensive sampling was conducted through June and July 2015, during the the middle of the dry season when natural sources of surface water are largely dry and watering hole use by medium and large vertebrates is highest. Notably, 2015 coincided with an exceptional drought in the KNP and a significant die-off of mega-herbivores (Malherbe et al., 2020). A subset of small concrete waterholes in the KNP were chosen for sampling. These waterholes when full hold roughly 2000 litres of water, are predominantly supplied with water from boreholes, and less commonly filled with river water piped long distances underground. Owing to their small sizes, a complete inventory of animal visitation may be achieved by setting just one or two cameras per site. In addition, the water in these small concrete troughs is likely to be well mixed and experience limited inflow from nearby surface waters. This reduces potential for input of DNA by animals other than those visiting the sites, and increases homogeneity of the water, reducing the number of samples needed for eDNA based surveys. The small, isolated, and replicated nature of the artificial waterholes coupled with available species inventories and reference sequences make this

system ideal for exploring the utility of eDNA metabarcoding.

Across the southern half of the park, six concrete bottom artificial waterholes were selected for sampling based on accessible distance to the laboratory to allow for sample processing within 12 hours of collection. Each of the waterholes is equipped with a ball-valve, which regulates water levels, re-filling the trough from nearby reservoirs when water levels drop. Other than intermittent clogging of the intake pipes, these troughs are stable water sources throughout the dry season. Sampled waterholes varied in shape, with some mimicking the contours of natural pans, making volume estimations difficult, but the overall design included longer and shorter axes, with comparable dimensions across waterholes. Four of these sites are filled with groundwater via boreholes, while two are fed by a pipeline that diverts water from the Olifants river, within the park.

Camera Trapping & Annotation

At each site, 2015 model Bushnell Aggressor Trophy Cam HD Low-Glow cameras were set up either in trees, on poles, or placed in metal cases and attached to the concrete ball-valve housing. When possible, one camera was used, but in some cases two cameras were needed to get complete coverage of a waterhole. Cameras were set to take a photo when triggered by motion, then wait one minute before being triggered again, so as to limit the number of photos taken of the same individual visitation, and increase battery life. In addition to motion triggering, cameras were set to take time lapse photos at five minute intervals, regardless of motion triggering. Individuals present in photographs were annotated to species level, and only individuals in contact with the waterhole, or directly adjacent to the waterhole were counted as using the waterhole (i.e. animals in the background, or passing through the field of view but without orientation towards the waterhole were excluded). For photos with more than 20 individuals per species, the number of individuals was counted three times and the mean recorded.

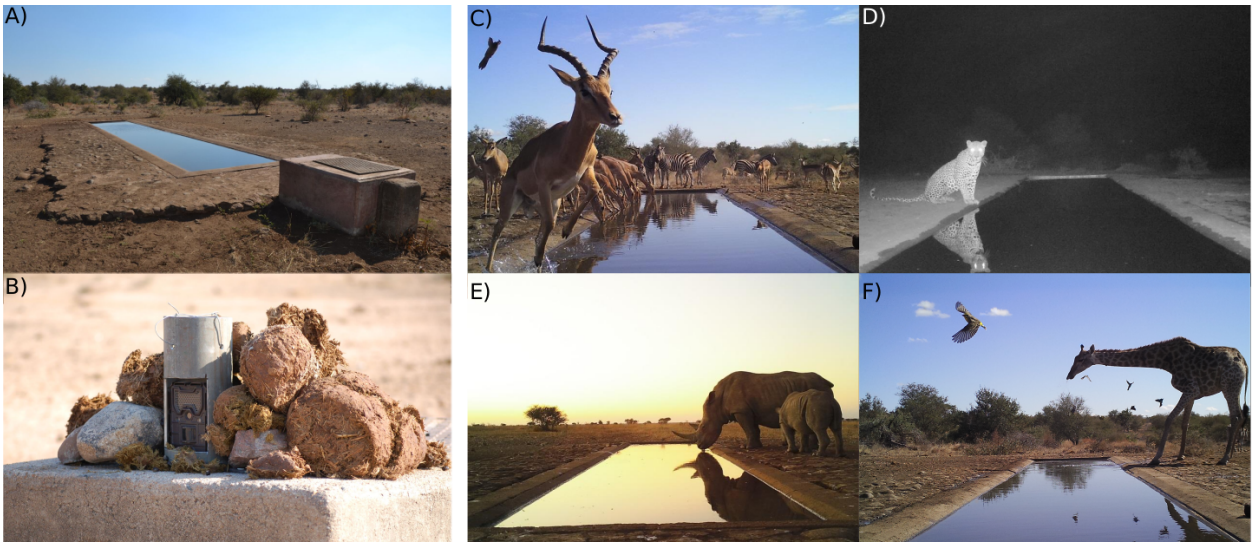

Figure 1: Example of a trough-shaped artificial waterhole (A) with associated camera trap housing attached to the valve housing (B), and examples of photographs (C-F).

Water Sampling

Following initial camera placement, we sampled each waterhole once per week for two weeks, resulting in a study design of two water samples and associated weeks worth of camera samples per site. Each sampling event consisted of gathering two replicate 1L water samples in autoclaved, UV sterilized glass jars from

opposite ends of the waterhole. These replicate samples were taken along the waterholes longest axis that maximized the distance and upwind-downwind gradient between them, if a strong wind was present. Water samples were placed on icepacks in a plastic cooler and kept between 4-8° Celsius until returning to the laboratory, where they were refrigerated until being filtered. Immediately after gathering water samples, water quality parameters were measured at each site using a YSI 650QS multi-parameter sonde. Temperature (C°), conductivity (mS/cm), dissolved oxygen (% saturation), and pH were recorded. Water quality was measured three times along the longest axis to account for variability in water quality within each waterhole.

Water Sample Processing

In the laboratory, water collection bottles were the washed with ELIMINase® (Decon Labs) and rinsed with deionized (DI) water to minimize sample cross-contamination. For each sample replicate, we sub-sampled 150 mL of water. We then filtered these sub-samples through gamma-irradiated 0.2 Supor® hydrophilic polyethersulfone membranes (Pall no. 66234). The filtration apparatus consisted of three 300mL Advantec polysulfone 47mm filter funnels fitted to a Pall vacuum manifold with vacuum pressure maintained by a Pall filtration vacuum/pressure pump (model no. 13158). Prior to filtration, all funnels components and tweezers used to manipulate the filters were sterilized by soaking with 10% bleach for 10 minutes, rinsing with DI water, washing with ELIMINase®, rinsing with DI water, and exposure to UV radiation for a minimum of 30 minutes. Gloves were worn at all times and changed between samples to minimize cross-sample contamination. All samples were processed within 12 hours of collection. Filters were stored in sterile 15mL Falcon tubes and placed in a freezer at -60°C.

DNA Extraction, Amplification, and Sequencing

DNA was isolated from filter papers using MO BIO PowerWater® DNA isolation kits. A cocktail of three primer sets targeting mammal and vertebrate species (BR5, mam_ckt, mam_ckt_230; see SM 1.2 for sequences) tagged with an Illumina adapter sequence were used to amplify variable regions of Cytochrome C Oxidase I (COI) through polymerase chain reaction (PCR). The PCRs used a standard mix of 17.8μL molecular grade water, 2.5μL 10× reaction buffer (200mM Tris HCl, 500mM KCl, pH 8.4), 1μL $MgCl_2$ (50mM), 0.5μL dNTP (10mM), 0.5μL forward primer (10mM), 0.5μL reverse primer (10mM), 0.2μL Platinum Taq DNA polymerase (Invitrogen), and 2μL DNA as template for a total volume of 25μL. PCRs underwent the following cycler conditions: initial 94°C for 5 minutes, then 30 cycles of 94°C for 40 seconds, 46°C for 1 minute, 72°C for 30 seconds, and a final temperature of 72°C for 2 minutes. Amplification success was confirmed through gel electrophoresis, using a 1.5% agarose gel. PCR products were purified using MinElute PCR purification kit (Qiagen), and quantified through flurometry using a Quant-iT PicoGreen dsDNA assay kit (Invitrogen). Samples were normalized, then multiplexed with the Nextera XT Index kit (96 indexes) (Illumina) and sequenced on an Illumina MiSeq flowcell using a V2 sequencing chemistry kit (2 x 250) making up approximately 1/8th of the run.

Sequence Processing

Across all samples, we generated a total of 2,709,713 Illumina reads. Sequences were separated by primer set and primer sequences removed using the trim.seqs function in mothur (Schloss et al., 2009). Reads were then processed in R (version 3.5.2) (R Development Core Team, 2008) using the package *dada2* version X (Callahan et al., 2016) following a modified version of the DADA2 Bioconductor workflow (Callahan et al., 2017a) and online tutorials and workflow for Big Data (benjjneb.github.io/dada2/tutorial.html). Separate sequence processing pipelines were developed for each primer set (BR5, mam_ckt, mam_ckt_230), although the overall process was similar for each region. All reads were filtered by quality, removing sequences with

maximum expected error (maxEE) greater than 4 for both forward and reverse reads, and reads with any base pair having Q of 6 or lower. Reads were truncated based on dropoffs in quality profiles. For BR5, reads were truncated to a length of 200 and 150 bp for forward and reverse reads respectively, 230 bp and 140 bp for mam_ckt, 220 bp and 140 bp for mam_ckt_230. Since the samples were sequenced across four different runs, subsequent steps of learning error rates, dereplication, denoising and Amplicon Sequence Variant (ASV) calling (Callahan et al., 2017b) using pooled samples, and merging of paired reads were performed separately for each run. For mam_ckt and mam_ckt_230 merging was done via concatenation. Tables of ASV sequences per sample within each run were then combined and chimera detection using all pooled samples was performed. In total 921,086 reads were retained, representing 2,986 ASVs.

Assembling the Kruger Vertebrate COI Reference Library

We assembled a COI KNP reference library by first compiling a complete species lists for mammals, birds, reptiles, amphibians, and fish maintained by the Kruger National Park Museum, returning a total of 854 named species. We then identified species synonyms via the *taxize* package in R (Scott Chamberlain and Eduard Szocs, 2013; Chamberlain et al., 2020) returning 1,283 Latin binomials. We downloaded representative COI sequences from GenBank via the R package *rentrez* using the following search string in the nuccore databate: "COX1[gene] OR cox1[gene] OR coxI[gene] OR CO1[gene] OR COI[gene] OR Cytochrome c oxidase subunit I[gene] OR cytochrome c oxidase subunit I[gene] OR cytochrome oxidase subunit 1[gene] OR Cytochrome oxidase subunit 1[gene]) AND 0:5000[Sequence Length] AND ∗SPECIES∗ [ORGN]", where ∗SPECIES∗ was replaced with the species Latin binomial. Since barcode sequences avalaible in the BOLD public database (Ratnasingham and Hebert, 2007) are not always found in GenBank (Porter and Hajibabaei, 2018a), COI sequences were also downloaded from BOLD via the R package *bold* (Chamberlain, 2017) by searching for each species and downloading available COI-5P sequences. Finally, to gather additional COI sequences for species which were not available via the search strategies described above, we excised COI sequences from whole annotated mitochondrial genomes available on GenBank using the *PrimerMiner* R package (Elbrecht and Leese, 2017) functions *Download_mito* and *Mito_GB2fasta*. To clean and format the reference sequences for use with *dada2*, we concatenated individual FASTAs per species into a single file, changing multi-line FASTAs to single-line, removing leading or trailing ambiguous bases (denoted by "N" or "-"), and excluding sequences that still included ambiguous base assignments.

We performed manual cleaning and curation of the concatenated FASTA sequence matrix, which involved sequence alignment via MUSCLE (Edgar, 2004) including an additional reference sequence used to identify the Folmer region (GBMA3852-12). After alignment, reference sequences were trimmed to the Folmer region in Aliview (Larsson, 2014), then the Folmer reference sequence and any duplicate sequences removed. Sequences were then realigned with MUSCLE and poorly aligned sequences removed. Alignment gaps in sequences were removed and all sequences filtered to include only those longer than 249bp. We then performed another iteration of removal of duplicated sequences, alignment, manual cleaning of poorly aligned sequences, and removal of alignment gaps. We used this output as our custom KNP vertebrate COI reference library for subsequent taxonomic annotation. We next generated a dada2-compatible taxonomy mapping file based on the KNP species inventories. Scripts (bash and R) to download sequences and format the resulting set of FASTAs are available in the Data & Code Supplement (to be archived via Zenodo upon acceptance), and at github.com/maxfarrell/eDNAcamtrap.

Taxonomy Assignment

Taxonomy assignment was conducted on the merged sequence table from across the three primer sets. Taxonomy assignment from Phylum to Species was performed using the RDP classifier (Wang et al., 2007), using the *dada2* assignTaxonomy function and our custom KNP Vertebrate Reference Library with mini-

mum bootstrap values of 50, 80, 95, and 98. To evaluate the sensitivity of taxonomic assignment to choice of reference library, we also assigned taxonomy using the Porter COI reference library (Porter and Hajibabaei, 2018b), and the MIDORI COI reference library (Machida et al., 2017), both with a minimum bootstrap value of 80%, see Wang et al. (2007). We provide scripts to format these databases for use with DADA2.

To explore how estimates of species richness varied with camera trapping effort, and how closely these are mirrored in eDNA detection, we counted the cumulative number of species observed across set time intervals: from three hours to 168 hours before sampling (Fig. SM1). For each time window, we then counted the number of species that were documented in both the camera traps and in the eDNA sample for that site. While our study design aimed to capture a minimum of one week of uninterrupted camera trapping data per sample, we were able to capture between two to three weeks of continuous data at some sites, allowing us to extend our species richness estimates beyond one week.

Modeling eDNA Detection

We used a hierarchical Bayesian regression with Bernoulli response to model species detections by eDNA metabarcoding across samples. We included as predictors the total visitation per species per sample measured as the numbers of individuals identified at the waterhole edge across all photos, the amount of time between sampling and the last instance each species visited a site, average female body mass in kilograms taken from the PanTHERIA mammal trait database (Jones et al., 2009), and sample level environmental predictors including temperature, conductivity, dissolved oxygen, and pH. In addition to these continuous predictors, we included hierarchical predictors for site and species effects, each with adaptively regularizing priors to allow for partial pooling across respective groups (see SM 1.3 for model formula).

We fit the model in Stan version 2.18.0 (Carpenter et al., 2017) via the R package *brms* version 2.7.0 (Bürkner, 2017). We assumed weakly informative priors assessed through prior predictive simulation. We ran the model across four chains, with 10,000 iterations per chain. For each chain, the first 5000 iterations were used as burn-in and discarded. The remaining iterations were thinned to retain every fourth iteration, resulting in a total of 5000 posterior draws. We diagnosed convergence by visual inspection of traceplots and observation of Rhat values equal to 1.0 for all estimated parameters, and assessed model fit using posterior predictive checks. To aid interpretation of the model we explored conditional effects plots for key predictors, and summarized posterior predictions from the model to examine the probability of eDNA detection across species.

Species Co-occurrence Patterns

To explore patterns in species co-occurrence at waterholes, we investigated the overlap in waterhole use per species pair, as determined by eDNA metabarcoding and camera trapping. For eDNA metabarcoding, we counted the number of samples in which each pair of species was detected. To parallel this with camera trap data, we counted the number of site-weeks in which pairs of species were detected with camera trapping, throughout the sampling period, and subset to 36 hours before water sampling. Our approach is not an attempt to quantify the physical co-occurrence of species pairs at an exact time and location, but rather identifies a temporal window in which pairs of species overlap in space.

Results

Kruger National Park Vertebrate COI Reference Library

Our curated reference library contained 4965 unique sequences for 391 species. This library covers 86 of the 147 mammals documented in the park, and although coverage is low for some small-bodied and diverse

orders (23/41 Chiroptera, 7/23 Rodentia, 1/9 Insectivora), the orders most likely to actively use the waterholes are well represented (23/26 Cetartiodactyla, 17/27 Carnivora, all 5 Primates, and all 3 Perissodactyla) (see Table SM1 for counts by mammalian order).

Taxonomic assignment

Total detected species richness from eDNA samples ranged between 13 and 30 species across reference libraries and RDP minimum bootstrap values (Fig. SM3, Table SM2). The previously compiled COI reference libraries for chordata (Porter–"TP") and metazoans (MIDORI–"MID") identified fewer species compared to our custom library (Fig. SM3, Table SM2). Taxonomic assignment using our curated library and a minimum bootstrap of 50% (KNP50) returned 30 species. These largely comprised mammals (19 species), but also included seven birds, two reptiles, one amphibian, and one fish species (Fig. 2, Table SM2). Increasing the minimum bootstrap value to 80% returned fewer species (14 mammals and two birds), with no species-level detections for reptiles, amphibians, or fish. The MIDORI and Porter reference libraries returned single fish and amphibian species which were different than those identified by the KNP50 approach (Table SM2) and not known to be present in the KNP (though for the amphibian, two congeneric species are documented in the park). It is possible that these sequences represent species related to those found in the KNP, while the actual KNP species are missing from the reference libraries, leading to high confidence being placed on the incorrect species. To quantify variation in taxonomic assignment across methods, we partitioned Sorensen's beta diversity across the species lists generated by each approach and find that the majority of differences across approaches is due to turnover rather than nestedness (Fig. SM2). This means that across the reference libraries, different species are being identified rather than each method identifying smaller subsets of the same set of species. For all subsequent analyses we used the KNP vertebrate reference library with an 80% minimum bootstrap (KNP80) as this provided the same species richness as higher bootstrap values using the same reference library (Fig. SM3), but only included mammals that were documented in the camera traps, thus minimizing false positives – errors of commission.

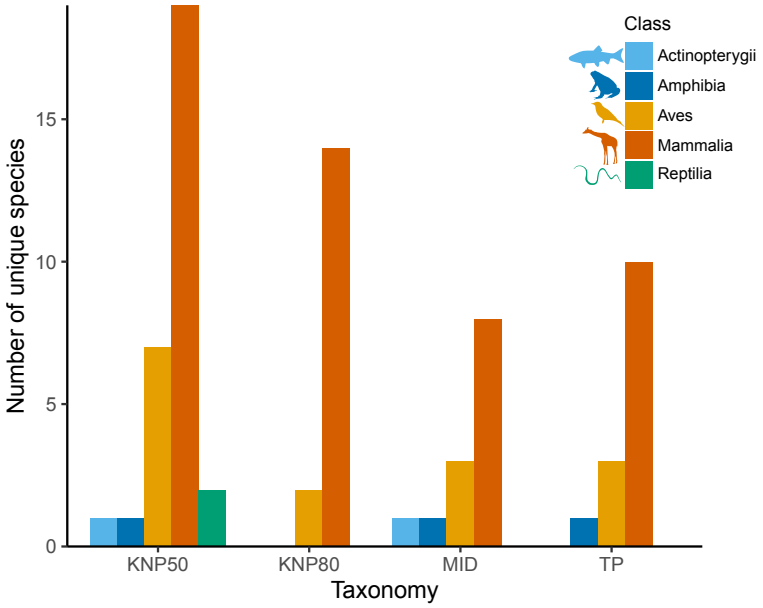

Figure 2: Number of species detected grouped by Class, according to the KNP reference library at the 50 (KNP50) and 80 % minimum bootstrap values, and the Porter Chordata reference library (TP) and the MIDORI reference library (MID) at 80% minimum bootstrap values.

Comparing eDNA and camera trap detections

Due to camera failures, at two sites we were only able to gather one complete week of camera trap data. However, for another site, we were able to obtain three uninterrupted weeks of camera trap data along with associated weekly water samples. In total, we were able to collect data from 11 weeks of uninterrupted camera trapping with associated water samples, resulting in 16,027 annotated photographs. In total, we identified 21 mammal species by camera traps, with 14 of these species detected by eDNA (Fig. 3). Two of the species documented in camera traps (*Mungos mungo* & *Paraxerus cepapi*) were not detected by eDNA because they did not have representative sequences in the KNP vertebrate reference library. Across all species which had reference sequences, 25% of species by sample combinations identified by camera trapping were recovered by eDNA. When repeating this calculation considering only species seen in the last 12 hours before sampling, 50% of species by sample combinations identified by camera trapping were detected by eDNA. When comparing short sampling times (3-12 hours before eDNA sampling), camera traps and eDNA return similar diversity estimates (Fig. SM1). However, eDNA-based richness plateaus after approximately 24 hours, indicating that the genetic signals of animals visiting waterholes longer than one day prior to water collection are unlikely to be detected via eDNA. To further explore temporal sensitivity in eDNA detection, we examined the relationship between the total visitation and time since last visit of a species across samples (Fig. 4). We find a negative relationship between total visitation and time since last visit, which indicates that species visiting waterholes often or in large numbers are more likely to visit relatively recently before a water sampling event, and are thus more likely to be detected by eDNA.

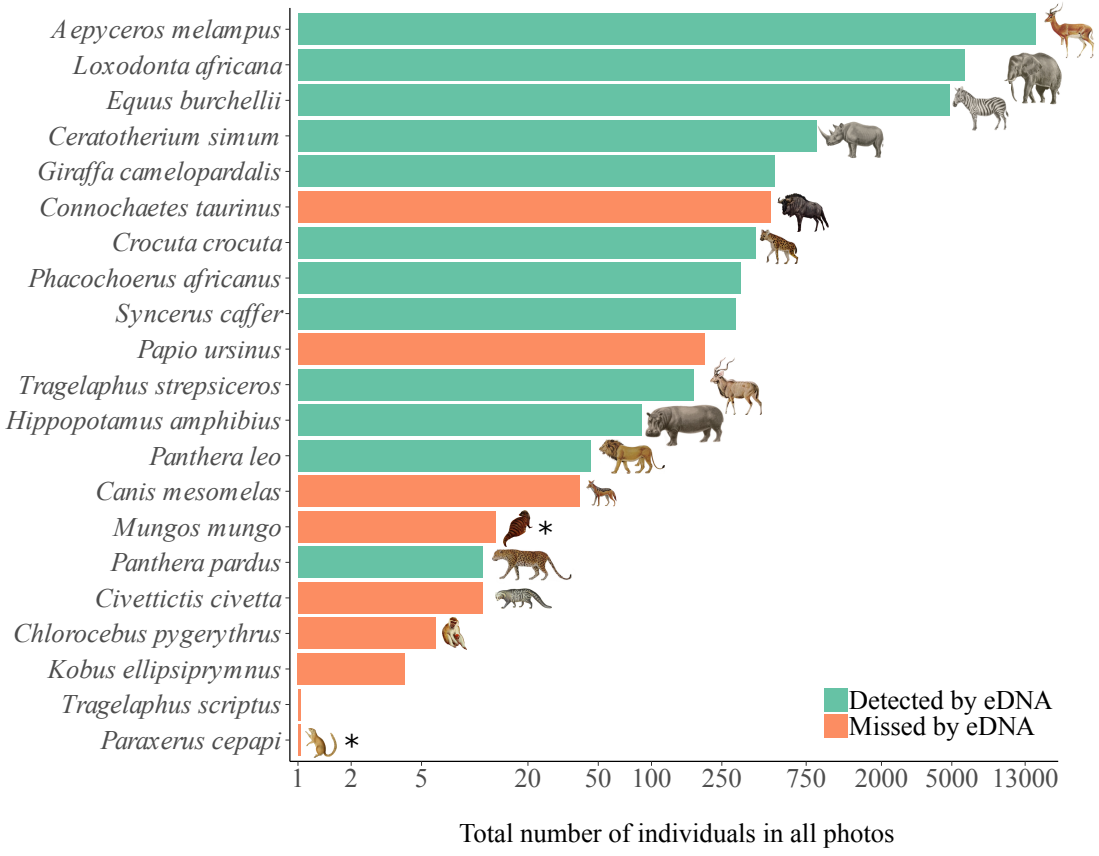

Figure 3: Number of individuals per species across all photographs. Coloring represents detection by eDNA across any of the samples. *species without representative sequences in the KNP vertebrate reference library.

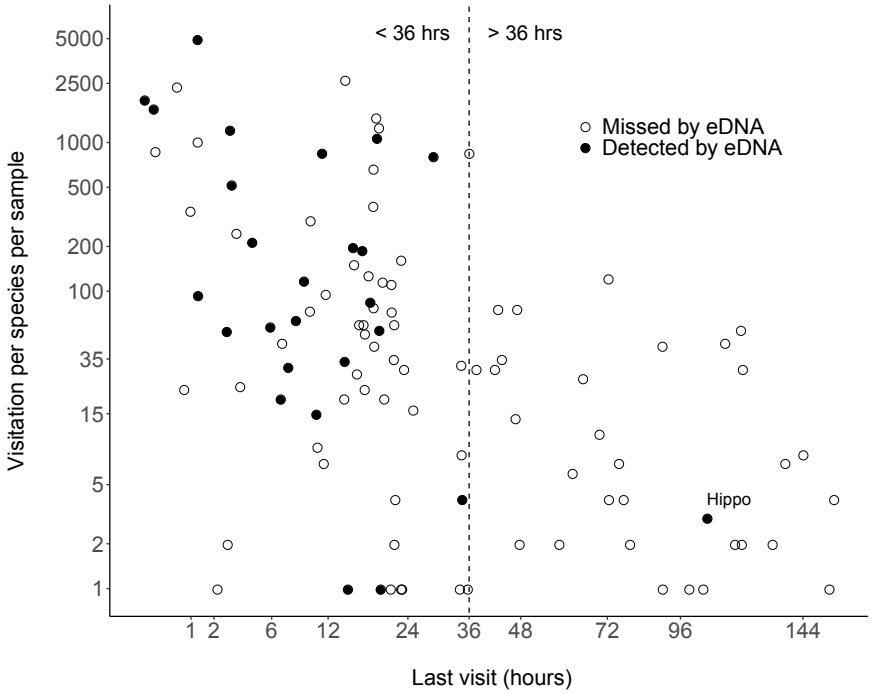

Figure 4: Raw data plot showing the relationship between the number of hours before sampling that a species last visited the site and the total count of visitations per sample over the previous week. Filled points indicate samples which have positive eDNA detection for the given species.

Variation in eDNA detection among species

We found large variation in the raw eDNA detection rates among species (comparing species identified by eDNA to those seen at each site in the previous week; Table SM4), with spotted hyena (*Crocuta crocuta*) having the highest eDNA detection rate (62% - five eDNA detections out of the eight samples in which it was seen by camera traps). Two species (baboons (*Papio ursinus*) and wildebeest (*Connochaetes taurinus*)) were never detected by eDNA despite visiting waterholes in relatively high numbers (Fig. 3). The clearest statistical predictor of detection probability was time since last visit (Fig. 5A-B, Table SM3), which indicates a rapid drop in detection probability within 36 hours. Body mass was also an important predictor of detection success. Smaller bodied species visit waterholes less often than larger mammals (Fig. SM4); however, body mass also had a relatively large positive effect in our eDNA detection model, with an 80% credible interval excluding zero (Fig. 5A, Table SM3), indicating that, after adjusting for total visitation and time since last visit, larger species may have behaviours or physiologies that make them shed eDNA at higher rates. Examining the raw data, we were never able to detect eDNA from species with an average mass less than 50 kg.

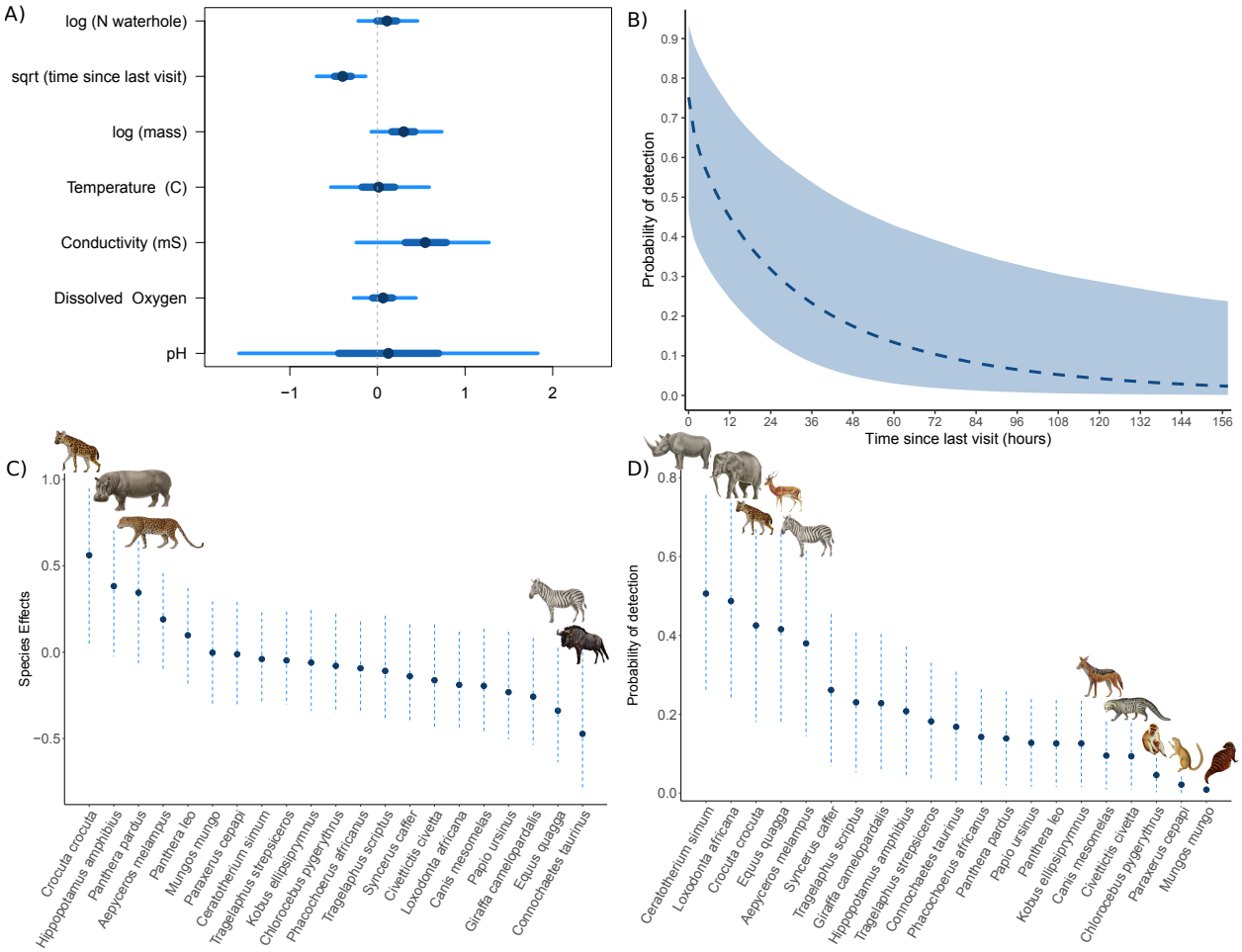

Figure 5: A) Estimated regression coefficients for continuous predictors from the eDNA detection model. Points represent mean estimates with thick and thin lines representing 50% and 95% credible intervals respectively. B) Conditional effect of time since a species last visited a site before sampling on the probability of eDNA detection. Conditional effects are estimated using the mean for continuous variables. The dashed line represents the mean effect, with the shaded area indicating the 95% credible interval. C) Estimated hierarchical effects for species, presented on the log-odds scale. These indicate the influence of species identity on eDNA detection after adjusting for other predictors in the model. Our model indicates that hyena and hippopotamus are detected more often than we would expect from their size and waterhole visitation patterns alone. Points indicate posterior means, with bars representing 50% credible intervals. Species are arranged by decreasing mean. D) Posterior predictions of species detection by eDNA metabarcoding – these predictions show that, on average, white rhino and elephant are the species most likely to be detected in any given water sample. Points indicate posterior means, with bars representing +/- the variance. Species are arranged by decreasing mean.

Species-level hierarchical effects (Fig. 5C) capture variation among species unexplained by the continuous predictors after adjusting for the other predictors in the model (time since last visit, body size, water quality parameters, etc.). These may help to identify additional unaccounted for variables. We also used posterior predictions from our Bayesian model to characterize the mammalian waterhole community as detected from eDNA (Fig. 5D). Unlike the species-level effects, these represent predicted species detection rates based on posterior predictions from the fit model. Here we show that white rhinos (*Ceratotherium*

simuim) and elephants (*Loxodonta africana*) have the highest mean probabilities of being detected, followed by hyena (*Crocuta crocuta*), zebra (*Equus quagga*), and impala (*Aepyceros melampus*). After these species there is a relative drop in the predicted probability of eDNA detection. Interestingly, wildebeest (*Connochaetes taurinus*) and baboons (*Papio anubis*), which were never detected by eDNA, have posterior probabilities of detection comparable to other species, indicating that given their waterhole ecology it is unusual that we did not find a positive eDNA detection in our study.

Waterhole Activity Patterns

Last, we examined species activity patterns at waterholes, and the ability of eDNA metabarcoding to capture species co-occurrences identified through camera trapping. Our camera trap data illustrates that species show strong signals of waterhole use across weeks (Fig. SM6), likely reflecting species specific differences in water needs. We can also detect variation in species co-occurrences in our eDNA profiles (Fig. SM7), though the lower overall species richness and detection rate of eDNA allows only for a coarser understanding of these patterns. We compared species co-occurrence matrices inferred via cameras and eDNA after subsetting camera trap data to photographs taken less than 36 hours before water sampling (removing bias introduced by the temporal decay in eDNA detection, shown above). When comparing only species detected by both methods, there is a correlation of 0.57 between the two co-occurrence matrices (Fig. 6A-B). It is likely that eDNA misses some co-occurrences detected with cameras, but rarely implies false co-occurrence, we therefore suggest that it may represent a conservative estimate of recent species overlaps across sites.

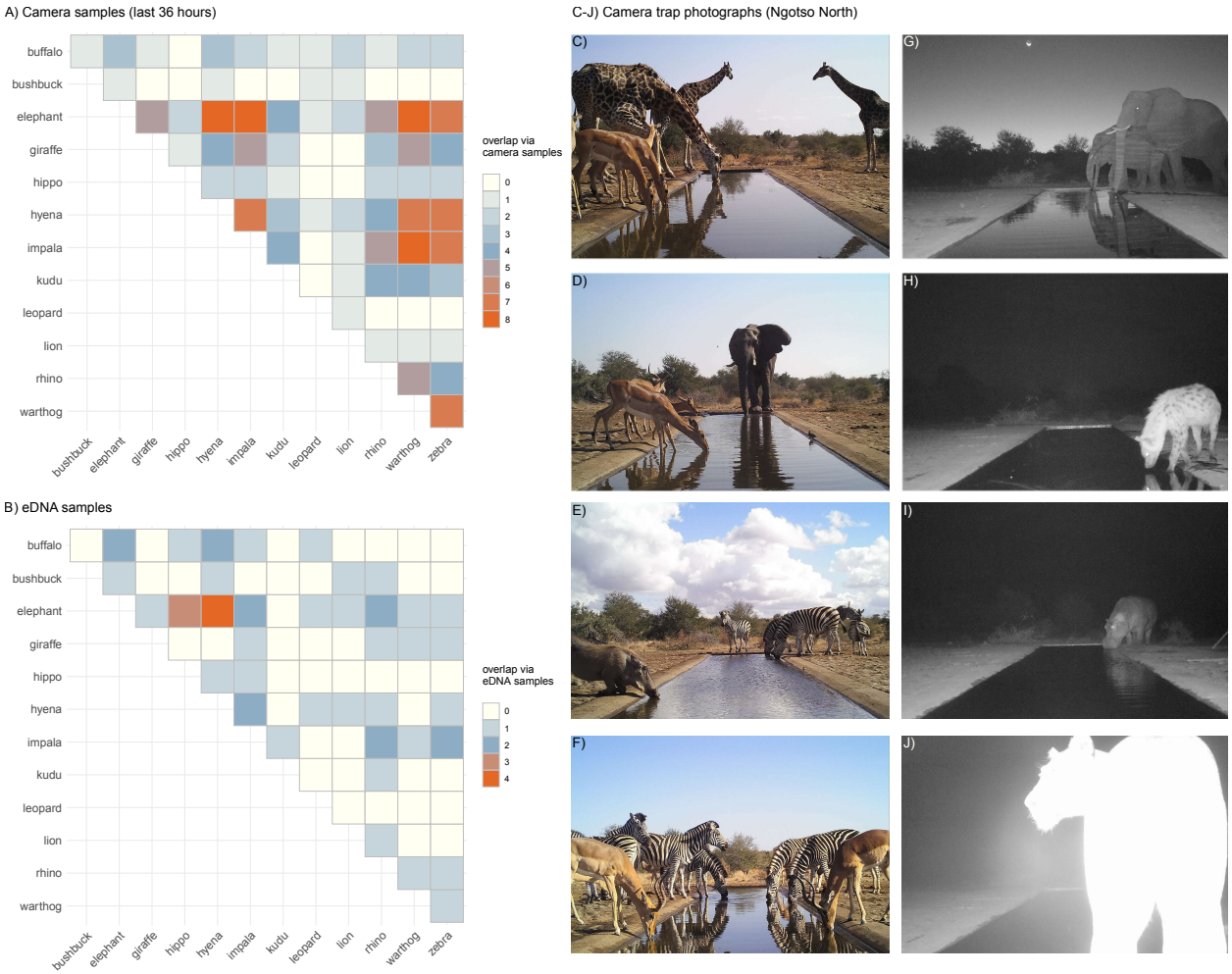

Figure 6: Comparison of overlap in waterhole usage documented by (A) camera traps (restricted to detections in the 36 hours prior to water sampling) and (B) by eDNA metabarcoding. After subsetting, the correlation between the two heatmaps is 0.57. Camera trap photographs from one site (Ngotso North) illustrate species co-occurrences (C-F) and single species visitations which are common during hours of darkness (G-J).

Discussion

Here we explored the efficacy of environmental DNA (eDNA) metabarcoding for documenting waterhole visitation of multiple mammals in a biodiverse savanna ecosystem, and contrast eDNA profiles to contemporaneous data collected from camera traps. The Kruger National Park (KNP) is considered one of the world's premier savanna reserves, and is home to 147 species of mammal, including stable populations of critically endangered species (Parr et al., 2009). We show that eDNA metabarcoding can successfully recover the signature of waterhole visitation for medium to large mammals, and accurately captures the water-dependent mammal community in this ecosystem. We also show that eDNA metabarcoding can provide a useful index of temporal overlap of species across space. However, the performance of eDNA metabarcoding is highly sensitive to the choice of reference library used for sequence annotation. In our study system, for which complete vertebrate species inventories are available, we obtain best results when using a site specific set of reference sequences as compared to publicly available reference libraries. By comparing eDNA detections to

species visitations documented by camera traps, we find that detection of eDNA from waterholes is temporally limited, with rates rapidly decaying since a species last visits a given site. We also illustrate sensitivity in detection rates across species, likely reflecting variability in the ability of molecular approaches to detect specific taxa, and differences in eDNA deposition due to physiological or behavioural traits influencing the ecology of waterhole use among species.

The ability to successfully detect species through eDNA metabarcoding is contingent on the availability of a library of representative sequences across species of interest. The relatively long history of conservation and biodiversity monitoring in the KNP allowed us to start with complete species lists for vertebrate groups in the park, and generate a curated set of reference sequences. We show substantially improved performance of eDNA detection across species when using our custom reference sequences as compared to existing publicly available reference libraries. Our results caution against using pre-configured reference libraries for eDNA-based biodiversity assessments, especially in areas for which putative species lists are available.

Even with our highly curated reference library, we still returned some "false" positives - instances when mammals were detected by eDNA but were not seen visiting the waterhole in the previous week; however, it is possible that some of these apparent false positives might have ecological explanations. For example, using the 80% minimum bootstrap we documented two "false" positives: buffalo (*Syncerus caffer*) in one sample, and bushbuck (*Tragelaphus scriptus*) in another. Both sites recording these positive detections were also visited by scavengers or predators (hyena, jackal, lion) within an hour of water collection, suggesting potential contamination due to indirect tissue deposition by predation or scavenging.

As the applications of eDNA metabarcoding expands, we suggest that future efforts focus on leveraging expertise of park scientists and existing natural history collections to further improve sequence reference libraries. Our species inventories for the park were provided by the collections manager for the KNP natural history museum, which includes a number of well preserved specimens across all taxonomic groups found in the park. As methods for the sequencing of DNA from preserved specimens continues to increase in sensitivity (Prosser et al., 2016; Düx et al., 2020), local collections present a valuable untapped resource for gathering expert-identified tissue samples of cryptic species currently missing from major sequence databases. For example, we identified two bird species: the critically endangered white-backed vulture (*Gyps africanus*) – which lives in high densities in our study area (Murn et al., 2013), and the helmeted guineafowl (*Numida meleagris*). Both species were detected by eDNA, and seen during sample collection and in the camera traps. There could be potential, therefore, to expand future surveys to include ground birds and threatened scavenging birds (Murn et al., 2013) via waterhole eDNA.

We found eDNA detection to be temporally limited in our study system, with highest detection probability for species visiting waterholes 24-36 hours prior to water sampling. We observed intense usage of these waterholes, and frequently waterholes were nearly emptied due to drawn down after visitation by large numbers of elephants and impala. It appears possible that eDNA deposited by visiting individuals is being actively removed through the act of drinking, rather than degrading due to unfavorable abiotic conditions. One consideration that appears to be largely overlooked by other eDNA studies of terrestrial mammals is both the time of day and frequency at which samples are collected. In our system, the time since a species' last visit is a stronger predictor of detection than the magnitude of waterhole use. Some species that infrequently visit waterholes, such as lion, are successfully detected when they visit within 24 hours prior to sampling, while other species that visit in higher numbers may be missed by eDNA simply because they did not visit just prior to sampling (see Fig. SM5 for illustrative time series). We only documented one instance of eDNA detection when a species last visited more than 36 hours prior to sampling. This single point represents a hippopotamus (*Hippopotamus amphibius*). While we did observe hippopotamus visiting artificial waterholes (Figs. 3, 6I), this particular instance of positive detection was at a site fed by a pipeline which brings diverted water from the Olifants river in which there are large groups of resident hippopotamus. It is therefore possible that this detection does not represent direct eDNA input from waterhole visitation.

In the Kruger, the provisioning of water via reservoirs and artificial waterholes is a key management

issue (Purdon and Aarde, 2017; Smit et al., 2007; Redfern et al., 2005) as it has promoted the increase of water-dependent species such as impala, zebra, and elephants, largely to the detriment of many drought-tolerant species such as roan (*Hippotragus equinus*) and sable antelopes (*Hippotragus niger*) which had suffered from competition for forage (Harrington et al., 1999). Through camera trapping we show that the KNP waterholes are dominated by these three water-dependent species, which we observed orders of magnitude higher visitation rates compared to other species using the waterholes. These same species are reflected as having high probability of detection with eDNA. While this extreme waterhole use make these species good targets for eDNA monitoring, the drinking ecology of several additional species, including rhinos and hyenas, also provide good targets for eDNA sampling.

Beyond time since last visit and mass as explanatory factors, our hierarchical model identified additional variation in eDNA detection among species. Hyena, hippopotamus, leopard, impala, and lion were estimated to have slightly elevated species effects, indicating that we have a higher probability of detecting these species after adjusting for other model predictors. Conversely, we find a lower probability of detecting particular species including wildebeest, zebra, giraffe, and baboons. The species level effects capture other qualitative observations, most obviously, our inability to detect baboons or wildebeest via eDNA, despite having reference sequences in our KNP library. These low detection rates may reflect molecular limitations, such as primer bias. They may also be the outcome of unmeasured ecological or behavioural traits, such as the potential for species to be "messy" or "clean" drinkers which influences the amount of saliva and epithelial cells that a species deposits. Because of such intrinsic species differences, it is likely that even well-designed sampling strategies will miss some species, and in systems where there is uncertainty about true species occurrences, absence of evidence should not be taken as evidence of absence. We should, however, note that camera trap surveys also suffer from similar biases, with variable detection across sites and species. For camera trap data, it is now common practice to use occupancy models, which explicitly include variability in detection across levels of organization (Burton et al., 2015), and parallel approaches show great promise for eDNA-based surveys (Schmelzle and Kinziger, 2016; Dorazio and Erickson, 2018; McClenaghan et al., 2020).

While the temporal limitation of eDNA may necessitate frequent sampling to capture the full species richness at a site, it offers a unique opportunity to study species associations. In addition to documenting mammal co-occurrences, paired eDNA metabarcoding of vertebrate and microbial sequences could be used to detect species associations across kingdoms. When comparing mammal species associations detected by both eDNA and camera trapping in the 36 hours before sampling, there is a correlation of 0.57 between the two co-occurrence matrices, indicating that eDNA may provide a crude proxy for species spatial overlap within time windows relevant for potential cross-species microbial or pathogen transmission via waterholes. Given the short half-life of eDNA detectability, we suggest that a large proportion of microbial taxa in these sites is likely to be animal-associated, partly explaining the high temporal and spatial turnover of bacterial communities in across these waterholes found previously (Farrell et al., 2019). These taxa may represent pathogens, or members of species oral, fecal, or skin microbiomes deposited through contact while drinking. The presence of artificial waterholes has dramatically altered mammal communities across the park, they have also likely infectious disease dynamics, particularly among water-dependent species, and for pathogens such as foot and mouth disease, bovine tuberculosis, and cryptosporiduim, which may be transmitted via shared water sources. In addition to pathogens, artificial waterholes may be sources for horizontal transfer of mutualist or commensal taxa, or antibiotic resistance plasmids carried from outside of the park (Mariano et al., 2009). Thus, while the short persistence time of eDNA in waterholes limits our ability to characterise waterhole usage for more than a couple of days before sampling, it offers an opportunity for identifying host-associated microbes, and the construction of association networks to study these dynamics through non-invasive sampling.

Camera traps offer unparalleled detail on species abundances, temporal visitation patterns, co-occurrences, and the potential for the identification of individuals (Burton et al., 2015). For example, only camera trap

data provide sufficient resolution to elucidate fine scale temporal interactions, such as direct or indirect competition for waterhole access. However, we found that cameras were prone to frequent technical failure resulting in cameras unexpectedly powering off, susceptible to tampering by animals (and people), and required substantial effort and time to position correctly. eDNA offers many unique advantages. Water sampling requires a very small amount of time in the field, meaning that large distances and multiple sites could be covered in a single day. Perhaps most importantly, eDNA-based approaches allow for the identification of cryptic species, and cross-kingdom biodiversity surveys. We suggest the two approaches are highly complementary, and that when used together provide rich data for describing and tracking trajectories of communities over time, helping guide applied conservation decisions in the KNP, which looks to embrace the heterogeneity and natural variation in its ecological systems (du Toit et al., 2003).

The temporal limits of eDNA for mammal detection in this system may mean that as an approach it is limited in describing total or long term diversity, though this may be overcome with increased sampling frequency or volumes, or extension of sampling to other waterhole types such as the large earthen dams also abundant in the KNP. Through camera trapping we observed that species tended to have variable patterns in terms of the times of day they visited waterholes, indicating that there may be optimal times of day to collect water samples to have the best chance of detecting a particular species. This consideration is not likely to be a factor for systems in which eDNA persists for prolonged periods of time, but may be critical for increasing detection rates for rare species in the KNP and similar systems. In the KNP, the temporal decay in the probability of species detection by eDNA metabarcoding means that eDNA approaches may be most useful for documenting patterns of waterhole use by species that are water-dependent, such as elephants, zebra, and impala, and those with consistent daily patterns of waterhole use, such as hyenas and rhinoceros. However, eDNA may miss species that visit waterholes less frequently and with relatively small body mass ($<$ 50kg) such as jackals, civets, and vervet monkeys.

Conclusion

Our study shows the utility of eDNA for biodiversity monitoring, and how it offers an added understanding of species ecology through pairing with camera trapping. As both of these biodiversity sampling approaches scale up and include increased automation, we will have further opportunity to understand the power and limitations of molecular approaches for documenting biodiversity patterns from local to regional scales. For the Kruger and other African savanna ecosystems, we hope our study lays the groundwork for future exciting work exploring cryptic biodiversity, the impacts of water provisioning on rare and threatened taxa, and the dynamic nature of species associations in these vital conservation areas.

Acknowledgements

This work would not have been possible without the logistical support of the African Centre for DNA Barcoding and Purvance Shikwambana, molecular work by Shannon Eagle, photographic identifications by Cody Danaher, and the incredible knowledge and protection provided by Kruger game guards Thomas, Velly, and Martin. We also thank Steven Kembel, his lab, Alexis Carteron, Joanne Littlefair, and Benjamin Callahan for constructive feedback on bioinformatic analyses. Special thanks are due to B.S. Bezeng for volunteering his friendship, dedication, and long days in the field. This work was performed with the support and permission of SANParks (project DAVIJ1256). MJF was supported by a Vanier NSERC CGS, the Vineberg Fellowship in Freshwater Biology/Limnology, the CIHR Systems Biology Training Program, and project funding support from the Quebec Centre for Biodiversity Science, the McGill Biology Department, and an NSERC Discovery Grant awarded to TJD.

Data & Code Accessibility

Scripts to generate the reference library, format public references for use with DADA2, perform all analyses, and generate the figures are available at github.com/maxfarrell/eDNAcamtrap (to be archived via zenodo upon acceptance of manuscript), along with data necessary to reproduce the analyses. Raw sequence data per sample are available via the Sequence Read Archive (SRA BioProject PRJNA490450 Study SRP161599).

Author Contributions

MJF and TJD designed the study with input from all co-authors. MJF gathered the data and performed water filtrations with support from DG and MvdB. MH designed the molecular workflow and performed DNA extractions, amplifications, and sequencing. MJF designed and performed all data analyses. MJF wrote the paper with input from TJD and DG.

Acknowledgements

This work would not have been possible without the logistical support of the African Centre for DNA Barcoding and Purvance Shikwambana, molecular work by Shannon Eagle, photographic identifications by Cody Danaher, species inventories from the Kruger National Park Museum staff, and the incredible knowledge and protection provided by Kruger game guards Thomas, Velly, and Martin. We also thank Steven Kembel, his lab, Alexis Carteron, Joanne Littlefair, and Benjamin Callahan for constructive feedback on bioinformatic analyses. Special thanks are due to B.S. Bezeng for volunteering his friendship, dedication, and long days in the field. This work was performed with the support and permission of SANParks (project DAVIJ1256). MJF was supported by a Vanier NSERC CGS, the Vineberg Fellowship in Freshwater Biology/Limnology, the CIHR Systems Biology Training Program, and project funding support from the Quebec Centre for Biodiversity Science, the McGill Biology Department, and an NSERC Discovery Grant awarded to TJD.

References

Amoroso, C. R., Kappeler, P. M., Fichtel, C., and Nunn, C. L. Temporal patterns of waterhole use as a predator avoidance strategy. *Journal of Mammalogy*, 101(2):574–581, 2020.

Bengis, R. and Erasmus, J. M. Wildlife diseases in South Africa : a review. *Revue scientifique et technique (International Office of Epizootics)*, 7(4):807–821, 1988.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., and de Bruyn, M. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in ecology & evolution*, 29(6):358–367, 2014.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., and Boutin, S. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, 52(3):675–685, 2015.

Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017.

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research*, 5(3):1492, 2017a.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583, 2016.

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, 11(12):2639–2643, 2017b.

Caron, A., Cross, P. C., and Toit, J. T. d. Ecological Implications of Bovine Tuberculosis in African Buffalo Herds. *Ecological Applications*, 13(5):1338–1345, 2003.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*, 76(1):1–32, 2017.

Chamberlain, S. bold: Interface to Bold Systems API. R package version 0.5.0., 2017.

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon, M., Li, G., and Grenié, M. *taxize: Taxonomic information from around the web*, 2020.

Day, K., Campbell, H., Fisher, A., Gibb, K., Hill, B., Rose, A., and Jarman, S. Development and validation of an environmental DNA test for the endangered Gouldian finch. *Endangered Species Research*, 40: 171–182, 2019.

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

Dorazio, R. M. and Erickson, R. A. ednaoccupancy: An r package for multiscale occupancy modelling of environmental dna data. *Molecular Ecology Resources*, 18(2):368–380, 2018.

du Toit, J., Rogers, K., and Biggs, H. *The Kruger Experience: Ecology and Management of Savanna Heterogeneity*. 2003.

Düx, A., Lequime, S., Patrono, L. V., Vrancken, B., Boral, S., Gogarten, J. F., Hilbig, A., Horst, D., Merkel, K., Prepoint, B., Santibanez, S., Schlotterbeck, J., Suchard, M. A., Ulrich, M., Widulin, N., Mankertz, A., Leendertz, F. H., Harper, K., Schnalke, T., Lemey, P., and Calvignac-Spencer, S. Measles virus and rinderpest virus divergence dated to the sixth century BCE. *Science*, 368(6497):1367–1370, 2020.

Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004.

Elbrecht, V. and Leese, F. Primerminer: an r package for development and in silico validation of dna metabarcoding primers. *Methods in Ecology and Evolution*, 8(5):622–626, 2017.

Everatt, K. T., Kokes, R., and Lopez Pereira, C. Evidence of a further emerging threat to lion conservation; targeted poaching for body parts. *Biodiversity and Conservation*, 28(14):4099–4114, 2019.

Farrell, M. J., Govender, D., Hajibabaei, M., van der Bank, M., and Davies, T. J. Bacterial diversity in the waterholes of the Kruger National Park: an eDNA metabarcoding approach. *Genome*, 62(3):229–242, 2019.

Ferreira, S. M., Bissett, C., Cowell, C. R., Gaylard, A., Greaver, C., Hayes, J., Hofmeyr, M., Moolman-van der Vyver, L., and Zimmermann, D. The status of rhinoceroses in South African National Parks. *Koedoe*, 59:1 – 11, 2017.

Ferreira, S. M., Greaver, C., Nhleko, Z., and Simms, C. Realization of poaching effects on rhinoceroses in Kruger National Park, South Africa. *African Journal of Wildlife Research*, 48(1), 2018.

Franz, M., Kramer-Schadt, S., Greenwood, A. D., and Courtiol, A. Sickness-induced lethargy can increase host contact rates and pathogen spread in water-limited landscapes. *Functional Ecology*, 32(9):2194–2204, 2018.

Furlan, E. M., Davis, J., and Duncan, R. P. Identifying error and accurately interpreting environmental DNA metabarcoding results: A case study to detect vertebrates at arid zone waterholes. *Molecular Ecology Resources*, 00(n/a):1–18, 2020.

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., and Calvignac-Spencer, S. Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. *Environmental DNA*, 2(1):63–76, 2020.

Harper, L. R., Lawson Handley, 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., and Hänfling, B. Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. *Biological Conservation*, 238:108225, 2019.

Harrington, R., Owen-Smith, N., Viljoen, P., Biggs, H., Mason, D., and Funston, P. Establishing the Causes of the Roan Antelope Decline in the Kruger National Park, South Africa. *Biological Conservation*, 90: 69–78, 1999.

Harrison, J. B., Sunday, J. M., and Rogers, S. M. Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, 286 (1915):20191409, 2019.

Higgitt, R. L., Schalkwyk, O. L. v., Klerk-Lorist, L.-M. d., Buss, P. E., Caldwell, P., Rossouw, L., Manamela, T., Hausler, G. A., Hewlett, J., Mitchell, E. P., Helden, P. D. v., Parsons, S. D. C., and Miller, M. A. Mycobacterium bovis Infection in African Wild Dogs, Kruger National Park, South Africa. *Emerging Infectious Diseases*, 25(7), 2019.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A., Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M., and Purvis, A. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90(9):2648–2648, 2009.

Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278, 2014.

Leempoel, K., Hebert, T., and Hadly, E. A. A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proceedings of the Royal Society B: Biological Sciences*, 287(1918): 20192353, 2020.

Machida, R. J., Leray, M., Ho, S.-L., and Knowlton, N. Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, 4(1):170027, 2017.

Malherbe, J., Smit, I. P., Wessels, K. J., and Beukes, P. J. Recent droughts in the Kruger National Park as reflected in the extreme climate index. *African Journal of Range & Forage Science*, 37(1):1–17, 2020.

Mariano, V., McCrindle, C. M. E., Cenci-Goga, B., and Picard, J. A. Case-Control Study To Determine whether River Water Can Spread Tetracycline Resistance to Unexposed Impala (Aepyceros melampus) in Kruger National Park (South Africa). *Applied and Environmental Microbiology*, 75(1):113–118, 2009.

McClenaghan, B., Compson, Z. G., and Hajibabaei, M. Validating metabarcoding-based biodiversity assessments with multi-species occupancy models: A case study using coastal marine edna. *PLoS One*, 15 (3):1–17, 2020.

Murn, C., Combrink, L., Ronaldson, G. S., Thompson, C., and Botha, A. Population estimates of three vulture species in Kruger National Park, South Africa. *Ostrich*, 84(1):1–9, 2013.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., and Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25):E5716–E5725, 2018.

Parr, C. L., Woinarski, J. C. Z., and Pienaar, D. J. Cornerstones of biodiversity conservation? Comparing the management effectiveness of Kruger and Kakadu National Parks, two key savanna reserves. *Biodiversity and Conservation*, 18(13):3643, 2009.

Porter, T. M. and Hajibabaei, M. Over 2.5 million COI sequences in GenBank and growing. *PLoS One*, 13 (9):1–16, 2018a.

Porter, T. M. and Hajibabaei, M. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, 27(2):313–338, 2018b.

Prosser, S. W. J., deWaard, J. R., Miller, S. E., and Hebert, P. D. N. DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources*, 16(2):487–497, 2016.

Purdon, A. and Aarde, R. J. v. Water provisioning in Kruger National Park alters elephant spatial utilisation patterns. *Journal of Arid Environments*, 141:45 – 51, 2017.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.

Ratnasingham, S. and Hebert, P. D. N. Bold: The Barcode of Life Data System. *Molecular ecology notes*, 7(3):355–364, 2007.

Redfern, J., Grant, C., Gaylard, A., and Getz, W. Surface water availability and the management of herbivore distributions in an African savanna ecosystem. *Journal of Arid Environments*, 63(2):406–424, 2005.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., and Gough, K. C. The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5):1450–1459, 2014.

Sales, N. G., Kaizer, M. d. C., Coscia, I., Perkins, J. C., Highlands, A., Boubli, J. P., Magnusson, W. E., Da Silva, M. N. F., Benvenuto, C., and Mcdevitt, A. D. 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, 2020.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., and Weber, C. F. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23): 7537–7541, 2009.

Schmelzle, M. C. and Kinziger, A. P. Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, 16(4):895–908, 2016.

Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Yu, D. W., Wilting, A., and Gilbert, M. T. P. 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, 2015.

Scott Chamberlain and Eduard Szocs. taxize - taxonomic search and retrieval in R. *F1000Research*, 2013.

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

Smit, I. P., Grant, C. C., and Devereux, B. J. Do artificial waterholes influence the way herbivores use the landscape? Herbivore distribution patterns around rivers and artificial surface water sources in a large African savanna park. *Biological Conservation*, 136(1):85–99, 2007.

Smit, I. P., Peel, M. J., Ferreira, S. M., Greaver, C., and Pienaar, D. J. Megaherbivore response to droughts under different management regimes: lessons from a large African savanna. *African Journal of Range & Forage Science*, 37(1):65–80, 2020.

Swanepoel, L. H., Balme, G. A., Williams, S. T., Power, R. J., Snyman, A., Gaigher, I. G., Senekal, C., Martins, Q., and Child, M. F. A conservation assessment of Panthera pardus. 2016.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., and Packer, C. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, 2(1):150026, 2015.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, 21(8):2045–50, 2012.

Turner, W. C., Kausrud, K. L., Beyer, W., Easterday, W. R., Barandongo, Z. R., Blaschke, E., Cloete, C. C., Lazak, J., Van Ert, M. N., Ganz, H. H., Turnbull, P. C. B., Stenseth, N. C., and Getz, W. M. Lethal exposure: An integrated approach to pathogen transmission via environmental reservoirs. *Scientific Reports*, 6(1):27311, 2016.

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., and Miya, M. Environmental DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*, 17(6): e63–e75, 2017.

Valeix, M., Fritz, H., Loveridge, A. J., Davidson, Z., Hunt, J. E., Murindagomo, F., and Macdonald, D. W. Does the Risk of Encountering Lions Influence African Herbivore Behaviour at Waterholes? *Behavioral Ecology and Sociobiology*, 63(10):1483–1494, 2009.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16): 5261–5267, 2007.

Environmental DNA for tracking waterhole visitation in savannah ecosystems
Supplementary Materials

Maxwell J. Farrell[1*], Danny Govender[2,3],
Mehrdad Hajibabaei[4], Michelle van der Bank[4,5], T. Jonathan Davies[5,6]

[1]Ecology & Evolutionary Biology, University of Toronto
[2]SANParks Scientific Services
[3]University of Pretoria
[4]University of Guelph
[5]African Centre for DNA Barcoding, University of Johannesburg, South Africa
[6]Botany, Forest & Conservation Sciences, University of British Columbia
*To whom correspondence should be addressed; E-mail: maxwell.farrell@utoronto.ca

November 3, 2020

1 Supplementary Methods

1.1 Assembling Kruger Vertebrate COI Reference Library

This is a textual outline for generating the Kruger Vertebrate COI Reference Library. For scripts to reproduce the automated sections of this workflow see the Data & Code supplement (github.com/maxfarrell/eDNAcamtrap, to be archived via Zenodo upon acceptance).

1. Mammal latin binomials and taxonomy from KNP inventories
2. Identify synonyms with R package *taxize*
3. Download COI sequences from Gen Bank (via R package *rentrez*)

   Search term in nuccore database:
   "COX1[gene] OR cox1[gene] OR coxI[gene] OR CO1[gene] OR COI[gene] OR Cytochrome c oxidase subunit I[gene] OR cytochrome c oxidase subunit I[gene] OR cytochrome oxidase subunit 1[gene] OR Cytochrome oxidase subunit 1[gene]) AND 0:5000[Sequence Length] AND ∗SPECIES∗ [ORGN]" where ∗SPECIES∗ was replaced with the species Latin binomial.
4. Download COI sequences from BOLD public database (via R package *bold*)

   Search for species name in and pull COI-5P sequences.
5. Excise COI sequences whole mitochondrial genomes using (via R package *PrimerMiner* functions Download_mito and Mito_GB2fasta)
6. Merging and cleaning reference library
   (a) Concatenate all individual fastas per species into one file
   (b) Change multi-line fasta sequences to single-line
   (c) Remove trailing ambiguous bases (N or -)
   (d) Remove leading ambiguous bases (N or -)
   (e) Remove sequences that have any ambiguous bases (N or -) and the preceding header line
   (f) Generate taxonomy mapping file
   (g) Prune FASTA sequences to only those in the taxonomy
   (h) Remove any duplicated sequence IDs in the fasta (they seem to come from GB)
   (i) Remove and lines with only ">"
   (j) Alignment with Muscle for visual check of sequences
   (k) Add reference to identify folmer region (GBMA3852-12)
   (l) Alignment with Muscle
   (m) Manual trimming of sequences to folmer region in AliView
   (n) Remove duplicated sequences
   (o) Alignment with Muscle
   (p) Manual cleaning of poorly aligned sequences in AliView
   (q) Alignment with Muscle
   (r) Manual cleaning of poorly aligned sequences in AliView

(s) Remove alignment gaps

(t) Filter sequences to those longer than 249 bp

(u) Alignment with Muscle

(v) Remove duplicated sequences

(w) Alignment with Muscle

(x) Manual cleaning of poorly aligned sequences in AliView

(y) Remove alignment gaps

(z) save as "final_Oct15_2019.fasta"

Order	Species in library	Species in park	Coverage
Carnivora	17	27	63 %
Cetartiodactyla	23	26	89 %
Chiroptera	23	41	56 %
Hydracoidea	1	2	50 %
Insectivora	1	9	11 %
Lagomorpha	1	3	33 %
Macroscelidea	2	3	66 %
Perissodactyla	3	3	100 %
Pholidota	1	1	100 %
Primates	5	5	100 %
Proboscoidea	1	1	100 %
Rodentia	7	23	30 %
Tubulidentata	1	1	100 %

Table SM1: Counts of mammal species per order included in the Kruger Vertebrate Reference Library, and present in the park.

1.2 Primers

Primer Set	Primer Name	Source	Sequence
mam ckt F + 230R	CO1 Minibar-MamF1	This paper	AAY CGM TGA YTR TTY TCI ACC AAY C
	CO1 Minibar-MamF2	This paper	AAY CGM TGA YTR TTY TCI ACC AAC C
	CO1 Minibar-MamF3	This paper	AAC CGA TGA CTA TTC TCT ACC AAT C
	230R	Gibson et al. 2014	CTTATRTTRTTTATICGIGGRAAIGC
mam ckt (F + R)	CO1 Minibar-MamF1	This paper	AAY CGM TGA YTR TTY TCI ACC AAY C
	CO1 Minibar-MamF2	This paper	AAY CGM TGA YTR TTY TCI ACC AAC C
	CO1 Minibar-MamF3	This paper	AAC CGA TGA CTA TTC TCT ACC AAT C
	CO1 Minibar-MamR1	This paper	AAR ATT ATY ACG AAR GCR TGR GCT G
	CO1 Minibar-MamR2	This paper	AAR ATT ATY ACG AAR GCR TGR GCR G
	CO1 Minibar-MamR3	This paper	AAR ATT ATY ACG AAI GCR TGI GCR G
BR5	B-F	Hajibabaei et al. (2012)	CCI GAY ATR GCI TTY CCI CG
	R5	Gibson et al. (2014)	GTR ATI GCI CCI GCI ARI ACI GG

1.3 Statistical Model for detection by eDNA

We used a hierarchical Bayesian regression with Bernoulli response to model species detections by eDNA metabarcoding across samples (species by sample detections denoted by i subscript). As predictors we included the total visitation per species per sample measured as the numbers of individuals identified at the waterhole edge across all photos ("N_waterhole"), the amount of time between sampling and the last instance each species visited a site ("last visit"), using the camera trap data and assuming no observation error. We additionally incuded species average female body mass in kilograms ("mass") taken from the PanTHERIA mammal trait database (Jones et al., 2009), and sample level environmental predictors including temperature, conductivity, dissolved oxygen, and pH. In addition to these continuous predictors, we included hierarchical predictors for site and species effects, each with adaptively regularizing priors to allow for partial pooling across respective groups. The statistical model follows:

$$\text{detection}_i \sim \text{Bernoulli}\,(p_i)$$

$$\begin{aligned}
\text{logit}(p_i) = \alpha &+ \beta_1\ \text{N_waterhole}_i \\
&+ \beta_2\ \text{last visit}_i \\
&+ \beta_3\ \text{mass}_i \\
&+ \beta_4\ \text{temperature}_i \\
&+ \beta_5\ \text{conductivity}_i \\
&+ \beta_6\ \text{dissolved oxygen}_i \\
&+ \beta_7\ \text{pH}_i \\
&+ \alpha_{\text{species}[i]} \\
&+ \alpha_{\text{site}[i]}
\end{aligned}$$

$$\begin{aligned}
\alpha_{\text{species}[i]} &\sim \mathcal{N}(0,\ \sigma_p^2) \\
\alpha_{\text{site}[i]} &\sim \mathcal{N}(0,\ \sigma_s^2)
\end{aligned}$$

$$\begin{aligned}
\alpha &\sim \mathcal{N}(0, 1) \\
\beta_{1-7} &\sim \mathcal{N}(0, 1.5) \\
\sigma_{p,s}^2 &\sim \mathcal{N}(0, 1)
\end{aligned}$$

Supplementary Results

Method	Class	Species
KNP50	Actinopterygii	*Micropanchax johnstoni*
KNP50	Amphibia	*Tomopterna cryptotis*
KNP50	Aves	*Anhinga rufa*
KNP50	Aves	*Butorides striatus*
KNP50	Aves	*Cinnyris mariquensis*
KNP50	Aves	*Gyps africanus*
KNP50	Aves	*Numida meleagris*
KNP50	Aves	*Sagittarius serpentarius*
KNP50	Aves	*Struthio camelus*
KNP50	Mammalia	*Aepyceros melampus*
KNP50	Mammalia	*Atilax paludinosus*
KNP50	Mammalia	*Ceratotherium simum*
KNP50	Mammalia	*Crocuta crocuta*
KNP50	Mammalia	*Elephantulus brachyrhynchus*
KNP50	Mammalia	*Equus quagga*
KNP50	Mammalia	*Gerbilliscus leucogaster*
KNP50	Mammalia	*Giraffa camelopardalis*
KNP50	Mammalia	*Hippopotamus amphibius*
KNP50	Mammalia	*Loxodonta africana*
KNP50	Mammalia	*Neamblysomus julianae*
KNP50	Mammalia	*Panthera leo*
KNP50	Mammalia	*Panthera pardus*
KNP50	Mammalia	*Phacochoerus africanus*
KNP50	Mammalia	*Rhinolophus hildebrandtii*
KNP50	Mammalia	*Sylvicapra grimmia*
KNP50	Mammalia	*Syncerus caffer*
KNP50	Mammalia	*Tragelaphus scriptus*
KNP50	Mammalia	*Tragelaphus strepsiceros*
KNP50	Reptilia	*Prosymna stuhlmannii*
KNP50	Reptilia	*Psammophis mossambicus*
KNP80	Aves	*Gyps africanus*
KNP80	Aves	*Numida meleagris*
KNP80	Mammalia	*Aepyceros melampus*
KNP80	Mammalia	*Ceratotherium simum*
KNP80	Mammalia	*Crocuta crocuta*
KNP80	Mammalia	*Equus quagga*
KNP80	Mammalia	*Giraffa camelopardalis*
KNP80	Mammalia	*Hippopotamus amphibius*
KNP80	Mammalia	*Loxodonta africana*
KNP80	Mammalia	*Panthera leo*
KNP80	Mammalia	*Panthera pardus*
KNP80	Mammalia	*Phacochoerus africanus*
KNP80	Mammalia	*Sylvicapra grimmia*
KNP80	Mammalia	*Syncerus caffer*
KNP80	Mammalia	*Tragelaphus scriptus*
KNP80	Mammalia	*Tragelaphus strepsiceros*
MID	Actinopterygii	*Zoarces gillii*
MID	Amphibia	*Xenopus muelleri*
MID	Aves	*Francolinus squamatus*
MID	Aves	*Gyps africanus*
MID	Aves	*Numida meleagris*
MID	Mammalia	*Ceratotherium simum*
MID	Mammalia	*Crocuta crocuta*
MID	Mammalia	*Equus burchellii*
MID	Mammalia	*Homo sapiens*
MID	Mammalia	*Hystrix cristata*
MID	Mammalia	*Loxodonta africana*
MID	Mammalia	*Panthera leo*
MID	Mammalia	*Panthera pardus*
TP	Amphibia	*Xenopus muelleri*
TP	Aves	*Francolinus squamatus*
TP	Aves	*Gyps africanus*
TP	Aves	*Numida meleagris*
TP	Mammalia	*Aepyceros melampus*
TP	Mammalia	*Equus asinus*
TP	Mammalia	*Giraffa camelopardalis*
TP	Mammalia	*Homo sapiens*
TP	Mammalia	*Hystrix cristata*
TP	Mammalia	*Phacochoerus africanus*
TP	Mammalia	*Sylvicapra grimmia*
TP	Mammalia	*Syncerus caffer*
TP	Mammalia	*Tragelaphus scriptus*
TP	Mammalia	*Tragelaphus strepsiceros*

Table SM2: Species identified per taxonomic assignment method)

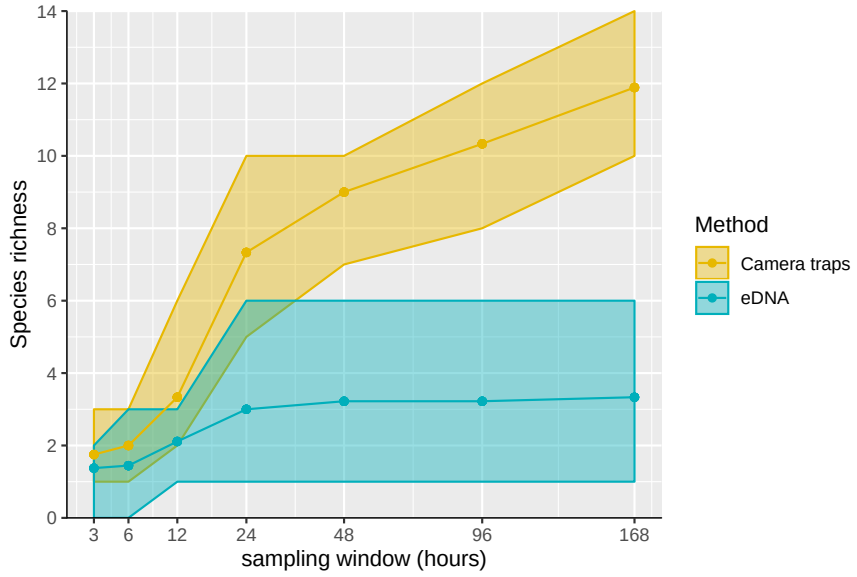

Figure SM1: Species richness estimates comparing camera traps to eDNA detection as a function of sampling window (time between sampling events). As the time between sampling events increases, camera traps detect increasing numbers of species, whereas the number of unique species detected by eDNA metabarcoding plateaus after roughly 24 hours, indicating that it may be most useful for detecting short-term visitation patterns.

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
log (N waterhole)	0.11	0.17	-0.22	0.10	0.46	22716.46	1.00
sqrt (last visit)	-0.40	0.14	-0.70	-0.39	-0.13	26697.80	1.00
log (mass)	0.30	0.21	-0.07	0.29	0.73	19399.64	1.00
Temperature	0.01	0.29	-0.53	0.01	0.59	23519.40	1.00
Conductivity	0.55	0.38	-0.24	0.55	1.27	17024.15	1.00
Dissolved Oxygen	0.06	0.18	-0.27	0.06	0.44	15372.11	1.00
pH	0.12	0.87	-1.58	0.13	1.83	22765.78	1.00
sigma species	0.61	0.41	0.03	0.55	1.54	8048.31	1.00
sigma site	0.54	0.45	0.02	0.43	1.67	8818.97	1.00

Table SM3: Estimated model parameters for eDNA detection model

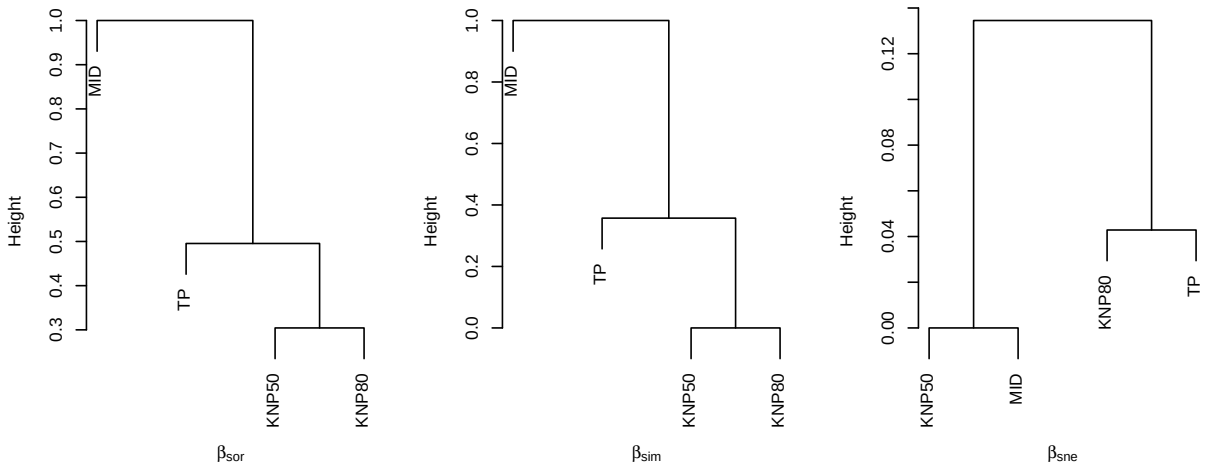

Figure SM2: Partitioning of Sorensen's beta diversity (β_{sor}) into turnover component (Simpson's diversity (β_{sim}), and nestedness (β_{sne}) components. Analyses conducted using the "betapart" R package (cite Baselga et al. 2012)

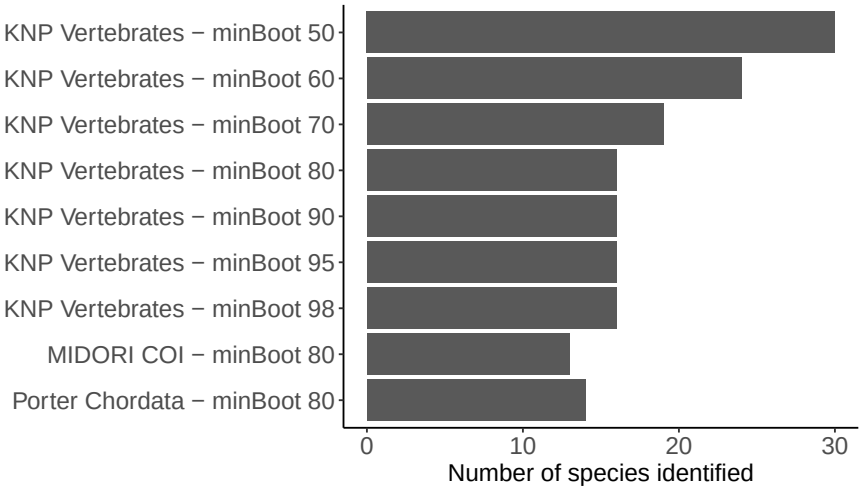

Figure SM3: Species richness across all samples by taxonomic reference library and minimum bootstrap (minBoot) value for the RDP classifier.

Species	eDNA detections	Camera detections	Rate
Crocuta crocuta	5	8	0.62
Ceratotherium simum	3	6	0.50
Panthera pardus	1	2	0.50
Aepyceros melampus	4	9	0.44
Loxodonta africana	4	9	0.44
Hippopotamus amphibius	2	5	0.40
Equus quagga	2	7	0.29
Syncerus caffer	1	5	0.20
Panthera leo	1	6	0.17
Tragelaphus strepsiceros	1	6	0.17
Giraffa camelopardalis	1	7	0.14
Phacochoerus africanus	1	9	0.11
Connochaetes taurinus	0	7	0.00
Canis mesomelas	0	5	0.00
Chlorocebus pygerythrus	0	4	0.00
Civettictis civetta	0	4	0.00
Papio ursinus	0	4	0.00
Kobus ellipsiprymnus	0	1	0.00
* *Mungos mungo*	0	1	0.00
* *Paraxerus cepapi*	0	1	0.00
Tragelaphus scriptus	0	1	0.00

Table SM4: Per species rates of eDNA (KNP Vertebrate Referecen Library with RDP classifier at 80% minimum bootstrap) and camera detections by sample. * denotes species which did not have representative sequences in the KNP vertebrate reference library. Across all species that had reference sequences, 25% of species by sample combinations identified by camera trapping were recovered by eDNA.

Species	eDNA detections	Camera detections	Rate
Hippopotamus amphibius	1	1	1.00
Syncerus caffer	1	1	1.00
Crocuta crocuta	4	6	0.67
Ceratotherium simum	3	5	0.60
Aepyceros melampus	2	4	0.50
Equus quagga	2	4	0.50
Loxodonta africana	2	4	0.50
Canis mesomelas	0	3	0.00
Civettictis civetta	0	1	0.00
Giraffa camelopardalis	0	1	0.00

Table SM5: Per species rates of eDNA and camera detections by sample, subset to camera detections within 12 hours prior to sampling. Across all species 50% of species by sample combinations identified by camera trapping were recovered by eDNA.

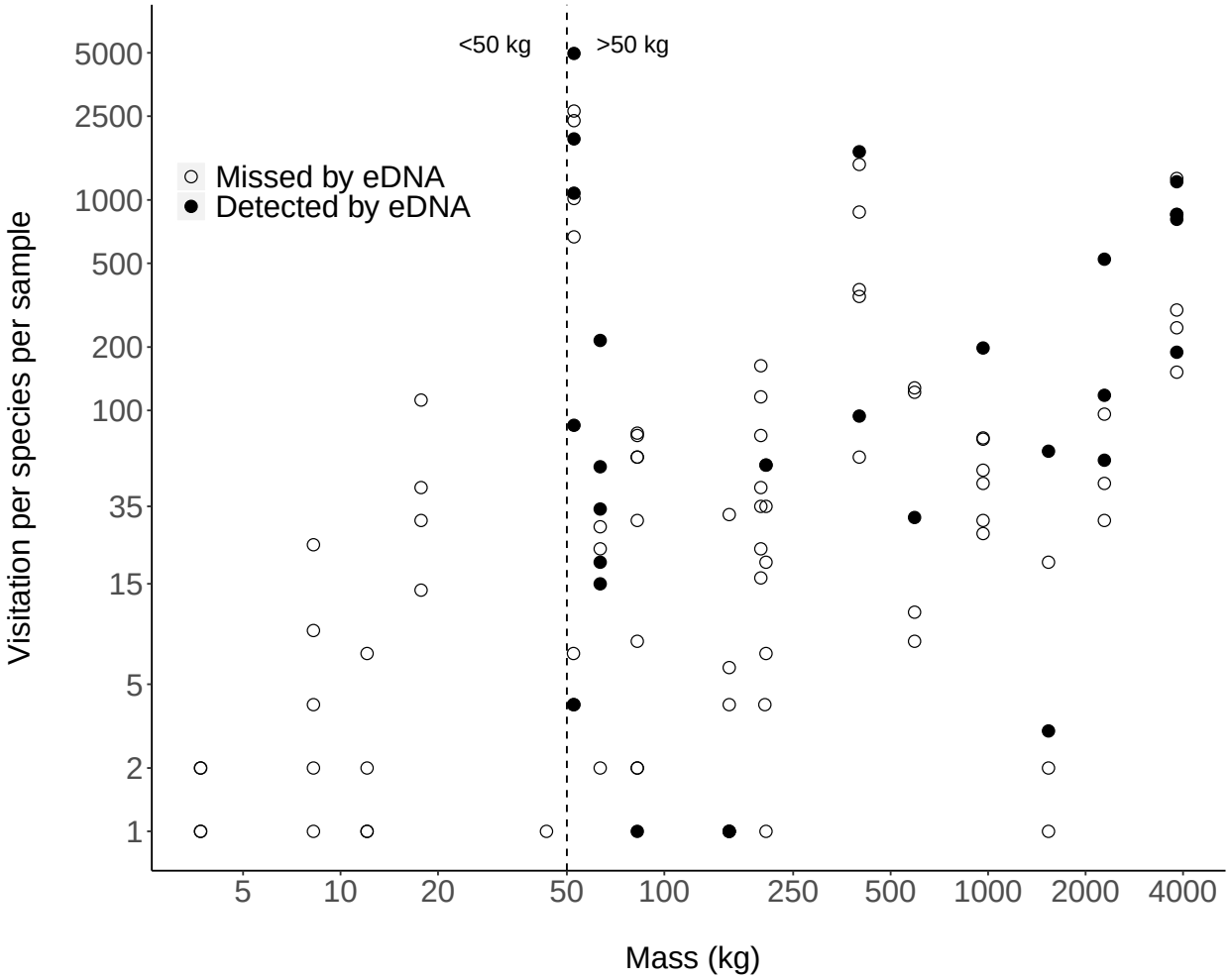

Figure SM4: Raw data plot showing the relationship between species mass (average adult female mass in kilograms) and the total number of animals per sample. Filled points indicate samples which have positive eDNA detection for the given species.

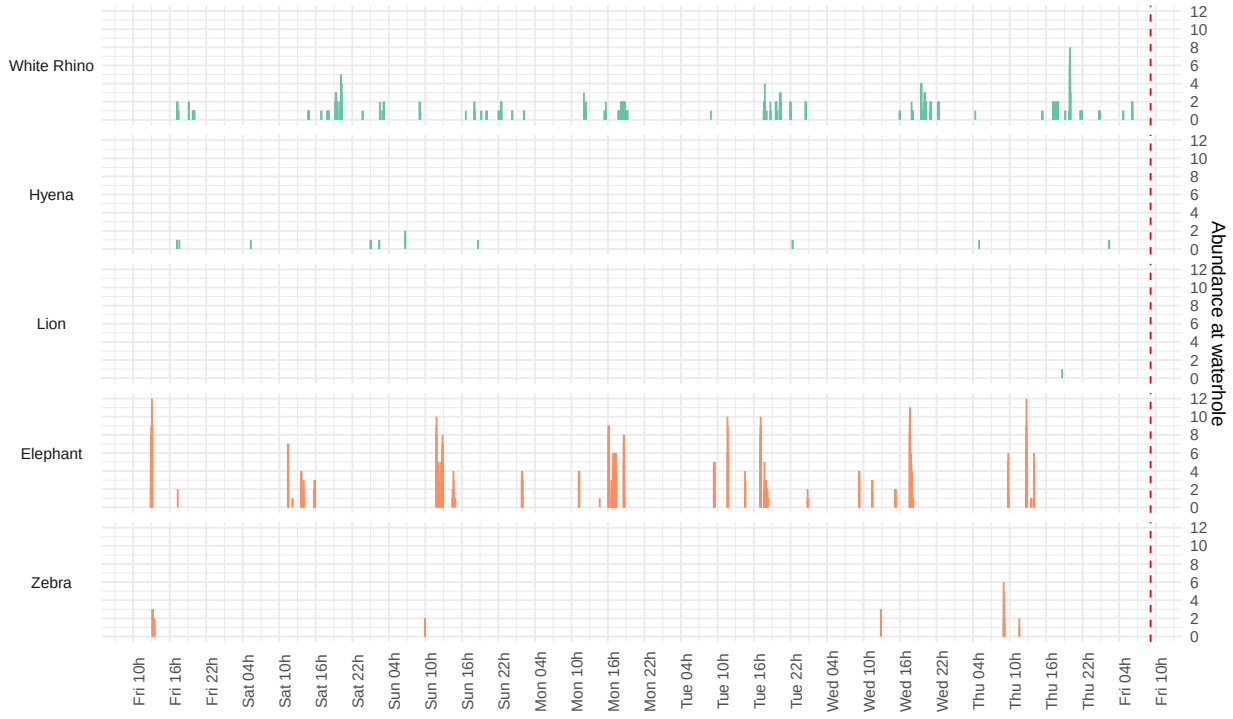

Figure SM5: Example of visitation timeseries for five species based a week of camera trapping at a single site. The red dashed line indicates the timing of water sampling. Orange indicates species that were not detect by eDNA, with turquiose representing species detected by eDNA, as in Fig. 3

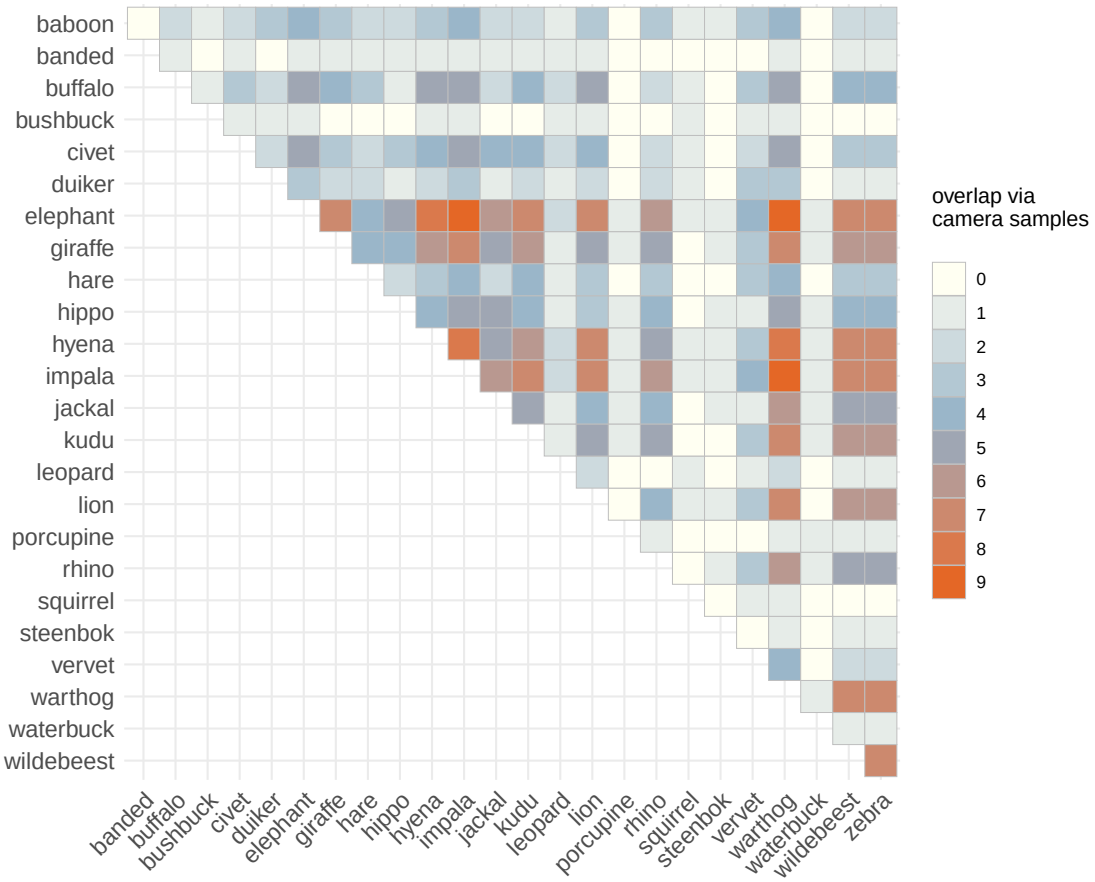

Figure SM6: Heatmap of camera-trap inferred pairwise overlap. Overlap among species pairs is determined by the number of weekly samples for which both species were documented by camera traps.

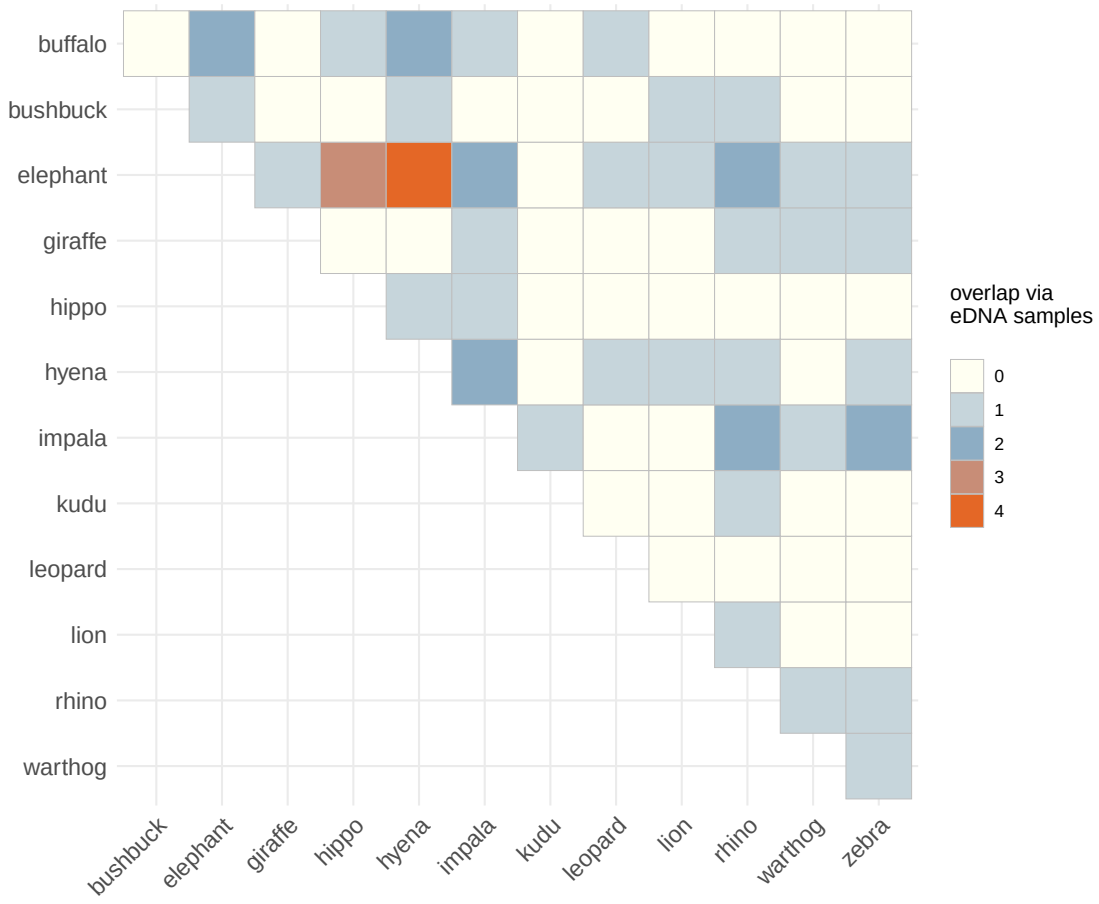

Figure SM7: Heatmap of eDNA inferred pairwise overlap among species present at waterholes. Overlaps were calculated using the number of samples in which both metabarcoding simultaneously detected both species.

References

Gibson, J., Shokralla, S., Porter, T. M., King, I., Konynenburg, S. v., Janzen, D. H., Hallwachs, W., and Hajibabaei, M. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, 111(22): 8007–8012, 2014.

Hajibabaei, M., Spall, J. L., Shokralla, S., and van Konynenburg, S. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12(1):28, 2012.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., Teacher, A., Bininda-Emonds, O. R. P., Gittleman, J. L., Mace, G. M., and Purvis, A. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90(9):2648–2648, 2009.