

Environmental DNA as a management tool for tracking artificial waterhole use in savanna ecosystems

Maxwell J. Farrell^{1*}, Danny Govender^{2,3},
Mehrdad Hajibabaei⁴, Michelle van der Bank^{4,5}, T. Jonathan Davies^{5,6}

¹Ecology & Evolutionary Biology, University of Toronto

²SANParks Scientific Services

³University of Pretoria

⁴Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph

⁵African Centre for DNA Barcoding, University of Johannesburg, South Africa

⁶Botany, Forest & Conservation Sciences, University of British Columbia

*To whom correspondence should be addressed; E-mail: maxwell.farrell@utoronto.ca

Keywords: eDNA metabarcoding, biodiversity survey, camera trap, megafauna, water management

1 Abstract

2 Game parks are the last preserve of many large mammals, and in savanna ecosystems, management of sur-
3 face waters poses a conservation challenge. In arid and semi-arid regions, water can be a scarce resource
4 during dry seasons and drought. Artificial waterholes are common in parks and reserves across Africa, but
5 can alter mammal community composition by favoring drought intolerant species, with consequences for
6 disease dynamics, and population viability of drought-tolerant species. Analysis of waterborne environ-
7 mental DNA (eDNA) is increasingly used to inform conservation of rare and invasive species, and conduct
8 large-scale biodiversity assessments. To explore the reliability of eDNA as an indicator of mammal wa-
9 terhole use in savannas, we compare eDNA metabarcoding and camera traps for documenting artificial
10 waterhole use in the Kruger National Park, South Africa, a global hotspot for mammal diversity. We show
11 that eDNA metabarcoding can recover the majority of mammal species detected by camera traps, including
12 a number of endangered species, but DNA signatures of mammal visitation are temporally limited, with best
13 performance when tracking water-dependent large bodied mammals visiting within two days of sampling.
14 Our results highlight limitation of eDNA based monitoring in these systems, including the lack of long-term
15 eDNA persistence in small and highly utilized waterholes, and variability in detection rates among species.
16 However, we demonstrate that eDNA-based approaches can be used to track mammals of conservation con-
17 cern, and reflect patterns of recent waterhole use and co-occurrence across water-dependent species, both of
18 which are crucial for making evidence-based decisions regarding water management and provisioning.

19 Introduction

20 Across Sub-Saharan Africa, artificial watering holes have been introduced in conservations areas and private
21 reserves (Weir, 1971; Hitchcock, 1996; Berry et al., 2001; Egeru et al., 2015; Chamaillé-Jammes et al.,
22 2016)). These waterholes were intended to increase game numbers by stabilizing water availability year-
23 round and are frequently visited by a diversity of birds and mammals. However, shifting availability of
24 water resources can modify predator-prey interactions (Valeix et al., 2009; Amoroso et al., 2020), influence
25 cross-species pathogen transmission (Turner et al., 2016; Franz et al., 2018), and drive large-scale shifts
26 in community composition (Redfern et al., 2005; Smit et al., 2007). The retention of these artificial sites
27 remains contentious as while they provide opportunities for game viewing that boost important tourism
28 revenue, this can conflict with conservation goals to promote biodiversity and vital ecosystem processes
29 through management strategies that maintain natural spatial and temporal variability in water availability
30 (Smit et al., 2020, 2007).

31 The Kruger National Park (KNP) is among the oldest game reserves in southern Africa, and a key con-
32 servation site for a number of threatened species such as white rhinoceros (*Ceratotherium simum*) (Ferreira
33 et al., 2017), a species of global conservation concern. Poaching and infectious diseases pose severe threats
34 for many species in the park including critically endangered black rhinoceros (*Diceros bicornis*) (Ferreira
35 et al., 2018), hippopotamus (*Hippopotamus amphibius*), elephant (*Loxodonta africana*), and giraffe (*Gi-*
36 *raffa camelopardalis*) (Bengis and Erasmus, 1988). The KNP began the introduction of artificial waterholes
37 in the 1930's, intending to increase game numbers by stabilizing water availability year-round (Redfern
38 et al., 2005). However, these artificial water sources have resulted in the expansion and overabundance of
39 drought-intolerant species such as impala (*Aepyceros melampus*), elephants, and zebra (*Equus burchellii*).
40 Conceived in the era of big-game hunting, the resulting community imbalance led to increased competition
41 for food, negatively impacting rarer drought-tolerant species, such as the sable (*Hippotragus niger*) and roan
42 antelopes (*Hippotragus equinus*), which have suffered marked declines in the park (Smit et al., 2007). Con-
43 sequently, many artificial waterholes were decommissioned, reducing numbers from over 300 at the peak in
44 the early 1990s (Smit et al., 2007) to roughly 160 in 2015, but their current usage by wildlife and impact on
45 wildlife community structure has not been systematically documented. Successful management of surface
46 waters can be strengthened by applying new tools to conduct biodiversity surveys at waterholes. Here we
47 explore the utility and limitations of two emerging approaches: sequencing of environmental DNA (eDNA),
48 and camera trapping.

49 Sequencing of eDNA from waterholes and ponds has been used successfully for surveying terrestrial
50 mammals in sub-Saharan Africa (Seeber et al., 2019) and Japan (Ushio et al., 2017), multiple vertebrates
51 in Australian arid zones (Furlan et al., 2020), and both semi-aquatic and terrestrial mammals in the United
52 Kingdom (Harper et al., 2019; Sales et al., 2020b). These studies indicate broad utility, but show that efficacy
53 varies across environmental conditions and species due to differences in abundance, body size, trophic
54 position, drinking behaviour, and tolerance to drought (Ushio et al., 2017; Seeber et al., 2019; Harrison
55 et al., 2019; Harper et al., 2019; Sales et al., 2020a,b; Leempoel et al., 2020; Jo et al., 2020; Lyet et al.,
56 2021). Therefore, despite the promise of eDNA, there remain questions about and how eDNA biomonitoring
57 compares to other survey methods including camera trapping (Harper et al., 2019; Wearn and Glover-Kapfer,
58 2019; Sales et al., 2020a,b; Leempoel et al., 2020; Lyet et al., 2021).

59 In contrast to eDNA approaches, camera traps have a longer history of use and have gained increased
60 popularity due to improved picture quality, decreased costs of digital cameras, and progress in citizen sci-
61 ence and machine learning for species identification (Norouzzadeh et al., 2018). Camera trap surveys are
62 now also being conducted at larger scales, with savanna ecosystems being the focus of one of the most
63 intensive co-ordinated global camera trap monitoring efforts (Swanson et al., 2015). Camera traps can pro-
64 vide a range of important data relevant for waterhole analyses including abundances, activity patterns, and
65 identification of individuals (Burton et al., 2015). However, camera traps are also subject to numerous lo-

66 gistic challenges including technical failure, theft and animal interference, inconsistent performance across
67 habitats and camera models, and informatic challenges of data storage, curation, and image classification
68 (Glover-Kapfer et al., 2019; Wearn and Glover-Kapfer, 2019). Thus, both camera trapping and eDNA-based
69 approaches have strengths and weaknesses (Harper et al., 2019; Leempoel et al., 2020; Sales et al., 2020a).

70 Here we evaluate the efficacy of eDNA metabarcoding in documenting waterhole visitation and tracking
71 endangered mammals in the KNP. With support from South African National Parks, we pair camera trap and
72 eDNA approaches to describe the mammal communities associated with artificial waterholes. We contrast
73 camera trap data on species visitation with Cytochrome C Oxidase I (COI) sequences amplified from eDNA
74 present in waterholes, and identify factors influencing the efficacy of eDNA-based approaches. Our study
75 not only explores the utility of eDNA metabarcoding for biodiversity monitoring, but also for quantifying
76 patterns of species co-occurrence, and tracking of endangered species in savanna ecosystems. This work
77 will help inform conservation planning and the management of artificial waterholes and surface water avail-
78 ability in arid and semi-arid regions by evaluating complimentary survey methods, and their utility for future
79 long-term monitoring programs. We suggest eDNA and camera trapping together are likely to become fun-
80 damental tools in biodiversity surveys as they can be conducted at greater spatial scales, broader taxonomic
81 scales, and finer temporal scales than previous approaches to document the use of surface waters in these
82 ecosystems.

83 Materials & Methods

84 Study Site

85 Sampling was conducted in the KNP through June and July 2015, when natural sources of surface water
86 are largely dry and waterhole use is highest. Notably, 2015 coincided with an exceptional drought in the
87 KNP and a significant die-off of mega-herbivores (Malherbe et al., 2020). Across the southern half of the
88 park, six concrete bottom artificial waterholes were selected based on distance to the laboratory that allowed
89 for sample processing within 12 hours of collection. These small waterholes hold approximately 2,000L
90 liters of water. Each is equipped with a ball-valve that regulates water levels (Fig. 1). Waterholes varied in
91 shape, with some mimicking the contours of natural pans. Four of the sites were filled with groundwater via
92 boreholes, while two were fed by a pipeline that diverts water from the nearby Olifants river.

93 Camera Trapping & Annotation

94 At each site, 2015 model Bushnell Aggressor Trophy Cam HD Low-Glow cameras were set up either in
95 trees, on poles, or placed in metal cases and attached to the concrete ball-valve housing. Cameras were set
96 to take time lapse photos at five-minute intervals, and also when triggered by motion with a minimum one
97 minute interval between photographs. Individuals present in photographs were annotated to species level,
98 and individuals in contact with the waterhole, or directly adjacent to the waterhole were counted as using the
99 waterhole (animals in the background, or passing through the field of view but without orientation towards
100 the waterhole were excluded).

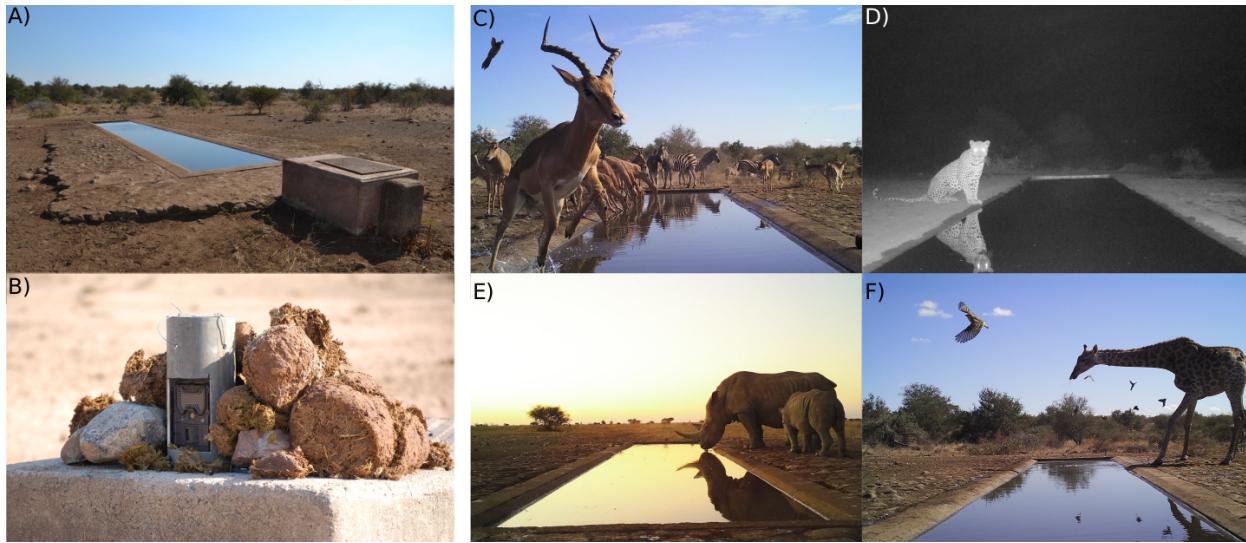


Figure 1: Example of a trough-shaped artificial waterhole (A) with associated camera trap housing attached to the valve housing (B), and examples of photographs (C-F).

101 Water Sampling

102 We took water samples after one week and two weeks following camera placement, providing two water
103 samples and two weeks of camera samples per site. At each sampling, we collected two replicate 1L water
104 samples in autoclaved, UV sterilized glass jars from opposite ends of the waterhole. Water samples were
105 placed on icepacks and kept between 4-8° Celsius until until filtering. Because eDNA degradation can be
106 influenced by physicochemical water properties (Jo et al., 2020; Curtis et al., 2021), we also collected data
107 on water temperature (C°), conductivity (mS/cm), dissolved oxygen (% saturation), and pH, using a YSI
108 650QS multi-parameter sonde.

109 Collection jars were washed with ELIMINase® (Decon Labs) and rinsed with deionized (DI) water to
110 minimize sample cross-contamination. Because of water turbidity at some sites, filtering the entire collected
111 volume was not feasible. Therefore, from each initial 1L sample, we sub-sampled 150mL for all down-
112 stream analyses and filtered this through gamma-irradiated 0.2 µm Supor® hydrophilic polyethersulfone
113 membranes (Pall no. 66234) using 300-mL Advantec polysulfone 47-mm filter funnels and a Pall manifold
114 with vacuum pressure maintained by a Pall filtration pump (model 13158). Prior to filtration, all funnel
115 components and tweezers were sterilized by soaking with 10% bleach for 10 minutes, rinsing with DI wa-
116 ter, washing with ELIMINase®, rinsing with DI water, and exposure to UV radiation for a minimum of
117 30 minutes. To identify potential contamination, negative controls were generated twice during sampling
118 by filtering 1L of deionized water used in the laboratory and subjected to extraction, amplification, and se-
119 quencing as for all other samples. Filters were stored in sterile 15mL Falcon tubes and placed in a freezer at
120 -60°C. All samples were filtered within 12 hours of collection.

121 DNA Extraction, Amplification, and Sequencing

122 DNA was isolated from filter papers using MO BIO PowerWater® DNA isolation kits. Three cocktails of
123 primers targeting mammal and vertebrate species (Supplementary Materials (SM) 1.2) tagged with Illumina
124 adapter sequences were used to amplify variable regions of the COI barcode gene – currently the gene with
125 the largest taxonomic coverage across animals (Porter and Hajibabaei, 2018a) – through polymerase chain
126 reaction (PCR). The mam ckt F + R and mam ckt F + 230R primer sets included three new minibarcode

127 primers based on methods described in Meusnier et al. (2008) and modified to better target mammalian
128 species. Primer3 software was used to ensure proper binding properties of the novel primers. The third
129 primer set (BR5) combined forward and reverse primers developed by Hajibabaei et al. (2012) and Gibson
130 et al. (2014).

131 Two-step PCRs were performed, with Illumina-tagged primers being used in the second step. For each
132 primer set/cocktail, individual primers were pooled (outlined in SM 1.2), and a separate PCR was con-
133 ducted. A mask was worn during all PCRs, and one reaction was done for each primer set/cocktail (i.e. no
134 replicate PCRs). The primer sets were pooled prior to sequencing. PCRs used a standard mix of 17.8 μ L
135 molecular grade water, 2.5 μ L 10x reaction buffer (200mM Tris HCl, 500mM KCl, pH 8.4), 1 μ L MgCl₂
136 (50mM), 0.5 μ L dNTP (10mM), 0.5 μ L forward primer (10mM), 0.5 μ L reverse primer (10mM), 0.2 μ L Plat-
137 inum Taq DNA polymerase (Invitrogen), and 2 μ L DNA as template for a total volume of 25 μ L. PCRs cycler
138 conditions varied by primer set (see SM 1.2 for details).

139 Amplification success was confirmed through gel electrophoresis, using a 1.5% agarose gel. PCR prod-
140 ucts were purified using MinElute PCR purification kit (Qiagen), and quantified through fluorometry using
141 a Quant-iT PicoGreen dsDNA assay kit (Invitrogen). Samples were normalized, then multiplexed with the
142 Nextera XT Index kit (96 indexes) (Illumina) and sequenced on an Illumina MiSeq flowcell using a V2 se-
143 quencing chemistry kit (2 x 250). In addition the two negative controls of deionized water described above,
144 one negative extraction control, and one negative PCR control were included. The first two negative controls
145 were sequenced, but the extraction and PCR controls were clean and thus sequencing was not necessary.

146 Bioinformatic Analyses

147 Across all samples, we generated a total of 2,709,713 Illumina reads. Sequences were separated by primer
148 set and primer sequences removed using the trim.seqs function in mothur (Schloss et al., 2009). Reads were
149 then processed in R (version 3.5.2) (R Development Core Team, 2008) using the package *dada2* version
150 X (Callahan et al., 2016) following the DADA2 Bioconductor workflow (Callahan et al., 2017b) and the
151 workflow for Big Data (benjneb.github.io/dada2/tutorial.html). Separate pipelines were developed for each
152 primer set. All reads were filtered by quality, removing sequences with maximum expected error greater
153 than 4 for both forward and reverse reads, and reads with any base pair having Q of 6 or lower. Reads
154 were truncated based on dropoffs in quality profiles. For primer set BR5, reads were truncated to a length
155 of 200 and 150 bp for forward and reverse reads respectively, 230 bp and 140 bp for the mam ckt F + R
156 primer set, 220 bp and 140 bp for the mam ckt F + 230R primer set. As samples were sequenced across four
157 different runs, learning error rates, dereplication, denoising and Amplicon Sequence Variant (ASV) calling
158 (Callahan et al., 2017a) using pooled samples, and merging of paired reads were performed separately for
159 each run. For mam ckt F + R and mam ckt F + 230R merging was done via concatenation. Tables of ASV
160 sequences per sample within each run were then combined and chimera detection using all pooled samples
161 was performed. In total 921,086 reads were retained, representing 2,986 ASVs.

162 Taxonomic assignment from Phylum to Species was performed using the RDP classifier (Wang et al.,
163 2007) via the *dada2* assignTaxonomy function and a custom KNP Vertebrate COI Reference Library (see
164 SM 1.1 for details) with minimum bootstrap values of 50, 80, 95, and 98. To evaluate the sensitivity of
165 taxonomic assignment to choice of reference library, we also assigned taxonomy using the Porter COI
166 reference library (Porter and Hajibabaei, 2018b), and the MIDORI COI reference library (Machida et al.,
167 2017), both with a minimum bootstrap value of 80%, see Wang et al. (2007).

168 Modeling eDNA Detection

169 We used a hierarchical Bayesian regression with Bernoulli response to model species detections by eDNA
170 metabarcoding across samples. As predictors, we calculated total visitation per species per sample (the

171 number of individuals identified at the waterhole) and the time between sampling and the last instance each
172 species visited a site based on the camera trap photographs. In addition, we included average female body
173 mass as reported in Jones et al. (2009), and environmental predictors including temperature, conductivity,
174 dissolved oxygen, and pH. We also included hierarchical predictors for site and species effects, each with
175 adaptively regularizing priors to allow for partial pooling across groups (see SM 1.3 for model formula).

176 We fit the model in Stan 2.18.0 (Carpenter et al., 2017) via the R package *brms* 2.7.0 (Bürkner, 2017).
177 We assumed weakly informative priors assessed through prior predictive simulation. The model was run
178 across four chains, with 10,000 iterations per chain and the first 5,000 iterations discarded as burn-in and
179 remaining iterations thinned to retain every fourth iteration. Convergence was checked by visual inspection
180 of traceplots and Rhat equal to 1.0 for all estimated parameters. Model fit was assessed using posterior
181 predictive checks.

182 Species Co-occurrence Patterns

183 To explore patterns in species co-occurrence, we investigated overlap in waterhole use as determined by
184 eDNA metabarcoding and camera trapping. For eDNA, we counted the number of samples in which each
185 species pair was detected. To match this with camera trap data, we counted the number of site-weeks in
186 which pairs of species were detected with camera trapping, throughout the sampling period, and subset to
187 36 hours before water sampling (see Results). Our approach is not an attempt to quantify the physical co-
188 occurrence of species pairs at an exact time and location, but rather identifies a temporal window in which
189 pairs of species overlap in space.

190 Results

191 Taxonomic assignment

192 Total detected species richness from eDNA ranged between 13 and 30 species across reference libraries and
193 RDP minimum bootstrap values (Fig. SM3, Table SM2). The previously compiled COI reference libraries
194 for chordata (Porter–“TP”) and metazoans (MIDORI–“MID”) identified fewer species compared to our
195 custom library (Fig. SM3, Table SM2). Taxonomic assignment using our curated library and a minimum
196 bootstrap of 50% (KNP50) returned 30 species. These largely comprised mammals (19 species), but also
197 included seven birds, two reptiles, one amphibian, and one fish species (Fig. 2, Table SM2). Increasing
198 the minimum bootstrap value to 80% returned fewer species (14 mammals and two birds), with no species-
199 level detections for reptiles, amphibians, or fish. For all subsequent analyses we used the KNP vertebrate
200 reference library with an 80% minimum bootstrap (KNP80) as this provided the same species richness as
201 higher bootstrap values using the same reference library (Fig. SM3), but only included mammals that were
202 documented in the camera traps, thus minimizing false positives – errors of commission. Using the KNP80
203 approach, negative controls identified a total 54 ASVs (5,066 reads total) across both controls, eight of
204 which were present in both negative controls. However, 52 ASVs could only be assigned to Phylum level,
205 while the remaining two could only be assigned to Class. Since no ASVs found in the negative controls
206 could be assigned to species level, we concluded no significant cross-contamination, and thus there was no
207 need to use the negative controls to filter ASVs identified in the rest of our samples (as all of our subsequent
208 analyses involve analysis of sequences identified to the species level).

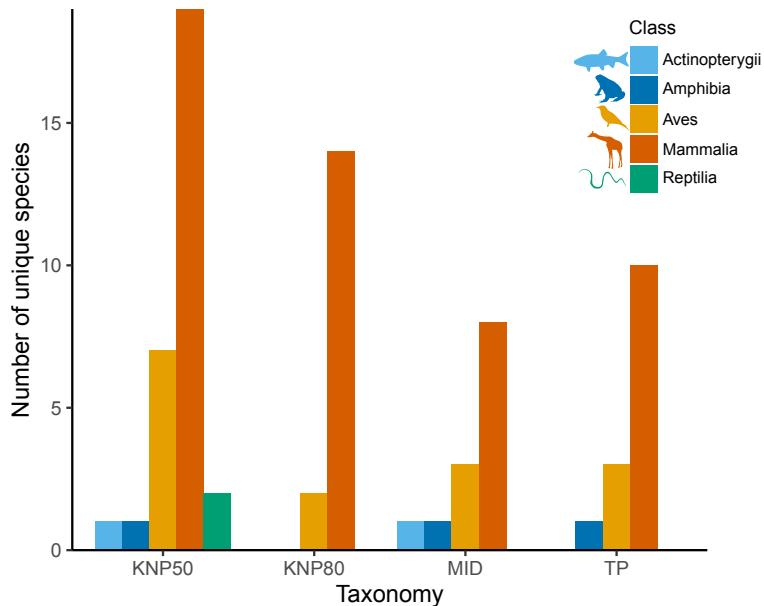


Figure 2: Number of species detected grouped by Class, according to the KNP reference library at the 50 (KNP50) and 80 % minimum bootstrap values, and the Porter Chordata reference library (TP) and the MIDORI reference library (MID) at 80% minimum bootstrap values. The MIDORI and Porter reference libraries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

209 Comparing eDNA and camera trap detections

210 We collected 11 weeks of uninterrupted camera trapping with associated water samples, resulting in 16,027
211 annotated photographs. We identified 21 mammal species by camera trap, 14 of which were detected by
212 eDNA (Fig. 3). Two species documented by camera traps (*Mungos mungo* and *Paraxerus cepapi*) were
213 not detected by eDNA because they did not have representative sequences in the KNP vertebrate reference
214 library. Across all species with reference sequences, 25% of species by sample combinations identified by
215 camera were recovered by eDNA. When restricting to only species seen in the last 12 hours before sampling,
216 50% of species by sample combinations identified by camera were detected by eDNA. When comparing
217 shorter camera trap sampling times (3-12 hours before eDNA sampling), cameras and eDNA return similar
218 diversity estimates (Fig. SM1). While we observed a negative relationship between total visitation and time
219 since last visit of a species, species last visiting a site >36 hours before sampling were rarely detected,
220 irrespective of overall visitation frequency (Fig. 4).

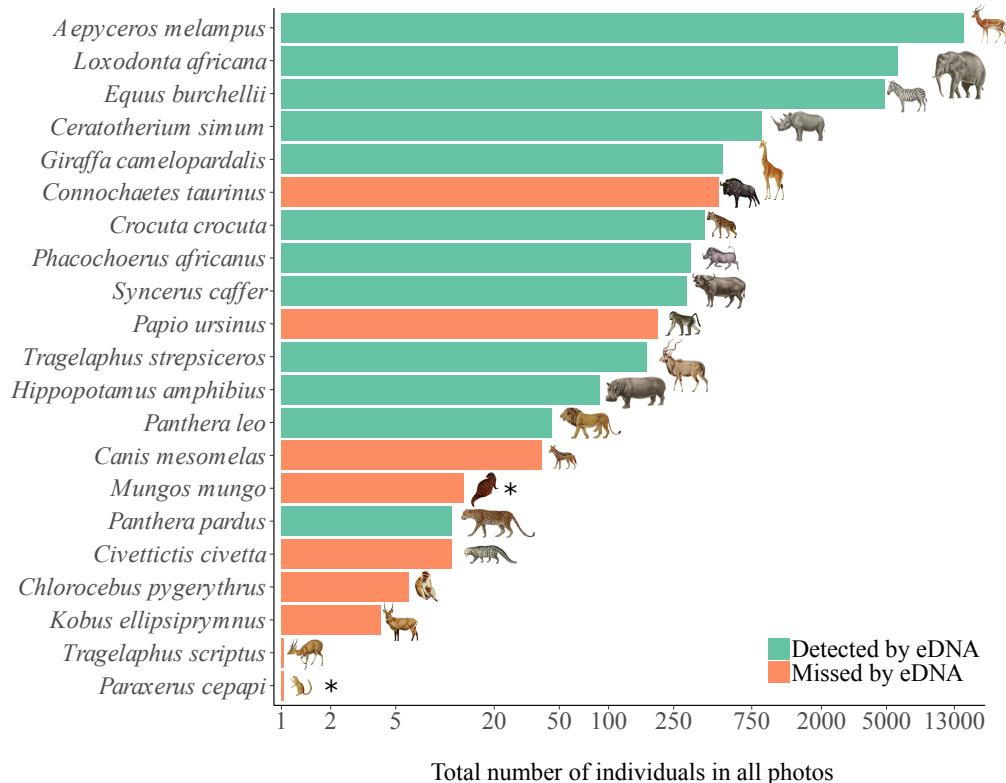


Figure 3: Number of individuals per species across all photographs. Coloring represents detection by eDNA across any of the samples. *species without representative sequences in the KNP vertebrate reference library.

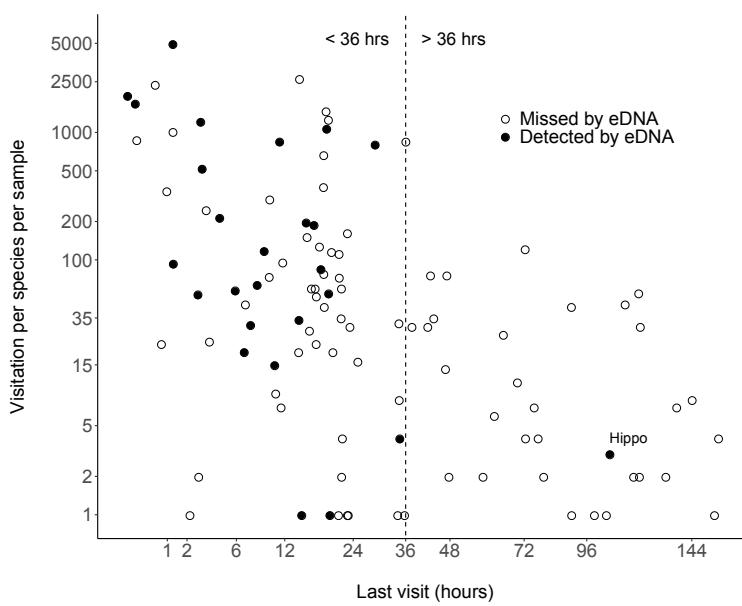


Figure 4: Raw data plot showing the relationship between the number of hours before sampling that a species last visited the site and the total count of visitations per sample over the previous week. Filled points indicate samples which have positive eDNA detection for the given species.

221 Variation in eDNA detection among species

222 We found large variation in the raw eDNA detection rates among species (Table SM4). Two species (baboons
223 (*Papio ursinus*) and wildebeest (*Connochaetes taurinus*)) were never detected by eDNA despite visiting
224 waterholes in relatively high numbers (Fig. 3). The clearest species-level predictor of detection probability
225 was time since last visit (Fig. 5A-B, Table SM3), and there was a rapid drop in detection probability after
226 36 hours (Fig. 4). Body mass was also an important predictor of detection success. While smaller bodied
227 species visit waterholes less often than larger mammals (Fig. SM4), body mass also had a relatively large
228 positive effect in our eDNA detection model adjusting for total visitation and time since last visit, (Fig. 5A,
229 Table SM3), indicating that larger species may have behaviours or physiologies that make them shed eDNA
230 at higher rates. We were never able to detect eDNA from species with an average mass less than 50 kg.

231 We used posterior predictions from our model to characterize the mammalian waterhole community
232 as detected by eDNA (Fig. 5D). Unlike the species-level effects, these represent species detection rates
233 based on predictions from the fit model. We find that white rhinos (*Ceratotherium simuim*) and elephants
234 (*Loxodonta africana*) have the highest mean probabilities of being detected, followed by hyena (*Crocuta*
235 *crocuta*), zebra (*Equus quagga*), and impala (*Aepyceros melampus*). Wildebeest (*Connochaetes taurinus*)
236 and baboons (*Papio anubis*), which were never detected by eDNA, have posterior probabilities of detection
237 comparable to other species, indicating that given their waterhole ecology it is unexpected that we did not
238 find a positive eDNA detection in our study. While baboons fell below the approximate size threshold (50kg)
239 for detection in this system, we were surprised not to detect wildebeest as they are large bodied and appear to
240 have visitation patterns similar to other large antelopes that we detected via eDNA. Primer bias is one of the
241 largest sources of taxonomic bias in eDNA metabarcoding studies (van der Loos and Nijland, 2021; Peixoto
242 et al., 2021). We conducted *post hoc* BLAST matches between our primers and archived COI sequences for
243 wildebeest, which suggested strong alignment, but it is possible other molecular biases, such as sub-optimal
244 annealing temperatures for this species (van der Loos and Nijland, 2021), limited our ability to amplify these
245 sequences.

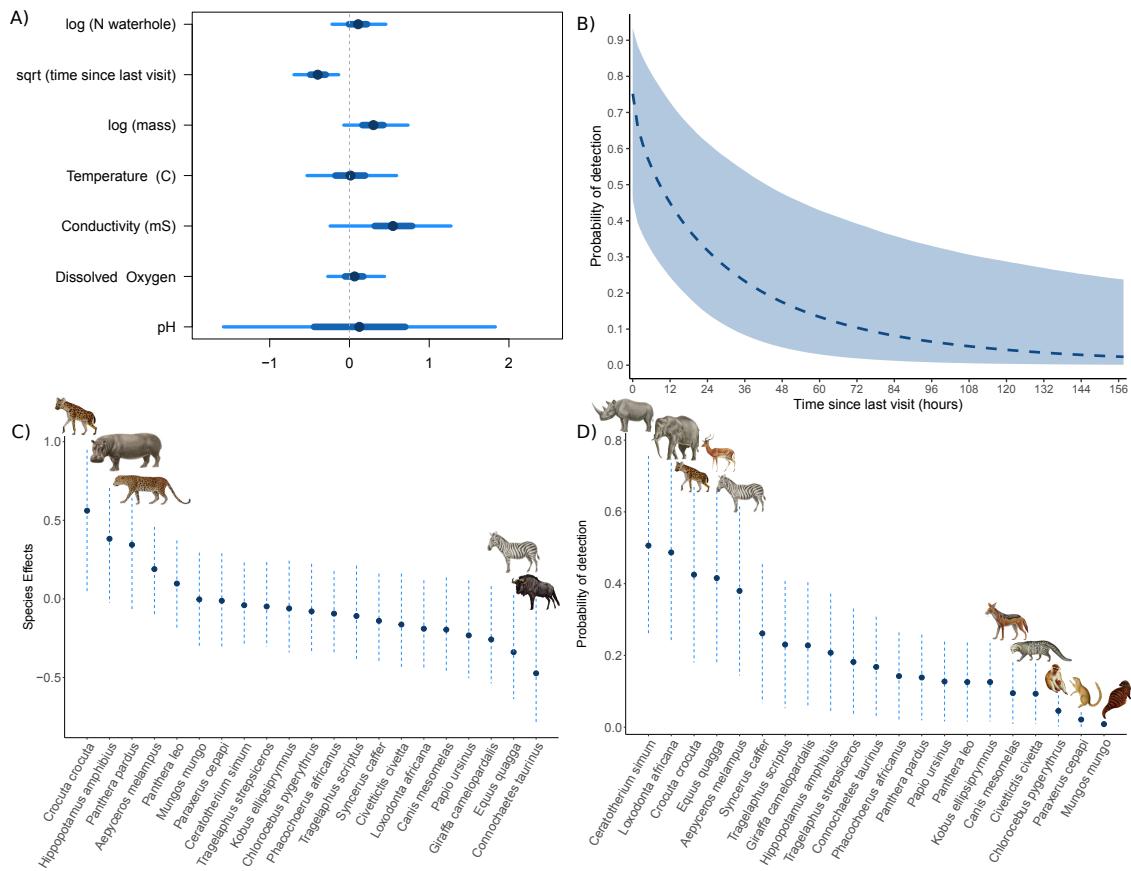


Figure 5: A) Estimated regression coefficients for continuous predictors from the eDNA detection model. Points represent mean estimates; thick and thin lines represent 50% and 95% credible intervals respectively. B) Conditional effect of time since a species last visit on the probability of eDNA detection. 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 that hyena and hippopotamus are detected more often than we would expect from their size and waterhole visitation patterns alone. 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. For panels C and D, points indicate posterior means, with bars representing +/- the variance and species are arranged by decreasing mean. Animal images are used to help illustrate top and bottom ranked species in these panels.

246 Waterhole Activity Patterns

247 Camera traps showed that species have large differences in the timing of waterhole use across weeks (Fig. SM6), likely reflecting species specific differences in water needs. We also detect this variation in species 248 co-occurrences via our eDNA profiles (Fig. SM7), though the lower species richness and detection rates of 249 eDNA only allow for a coarser understanding of these patterns. Comparing species co-occurrences inferred 250 via eDNA and camera trap data subset to photographs taken less than 36 hours before water sampling 251 (removing bias introduced by the temporal decay in eDNA detection), and for only species detected by both 252 methods, there is a correlation of 0.57 ($p < 0.001$) between the two co-occurrence matrices (Fig. 6A-B). 253

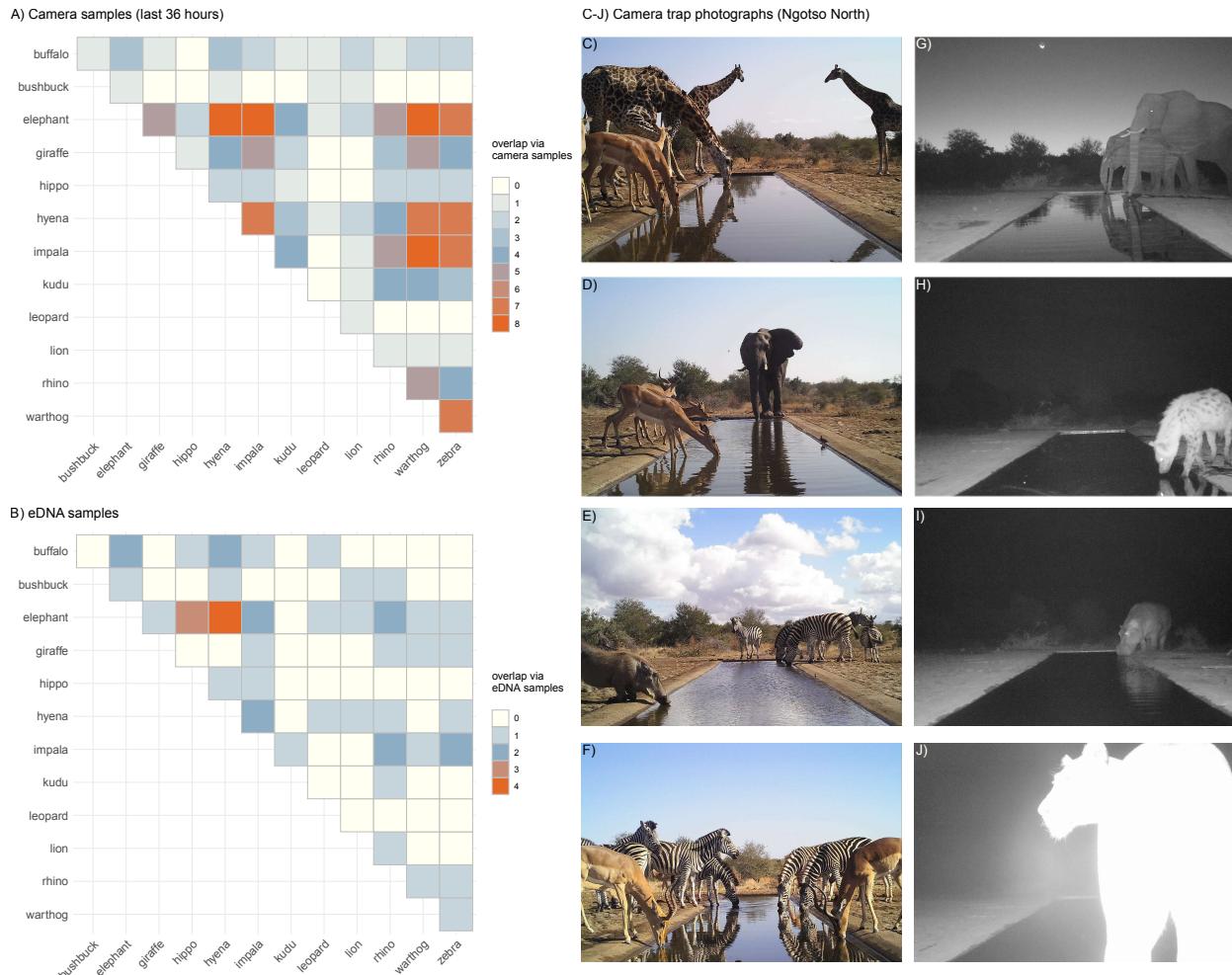


Figure 6: Comparison of overlap in waterhole usage documented by (A) camera traps (restricted to detections in the 36 hours prior to water sampling) and (B) by eDNA metabarcoding. After subsetting, the correlation between the two heatmaps is 0.57 ($p < 0.001$). Camera trap photographs from one site (Ngotso North) illustrate species co-occurrences (C-F) and single species visitations which are common during hours of darkness (G-J).

254 Discussion

255 In the KNP, provisioning of water via reservoirs and artificial waterholes are a key management issue (Smit
 256 et al., 2007; Redfern et al., 2005). Artificial waterholes have promoted water-dependent species such as
 257 impala, zebra, and elephants, largely to the detriment of many drought-tolerant species which have suffered
 258 from competition for forage (Harrington et al., 1999). Our camera trapping confirms that waterholes in the
 259 KNP are dominated by these three water-dependent species, and demonstrate that eDNA metabarcoding
 260 can recover the signature of waterhole visitation for medium to large mammals, accurately capturing the
 261 water-dependent mammal community. Our results reinforce the growing consensus that artificial waterholes
 262 contribute to the taxonomic dissonance of wildlife communities in arid lands, and we highlight the efficacy
 263 of waterhole eDNA for wildlife monitoring. While frequent waterhole use make elephants, impala, and
 264 zebra good targets for eDNA detection, the drinking ecology of several other species, including rhinos and
 265 hyenas, also make them promising targets for eDNA-based monitoring. However, we find that eDNA per-

266 performance also varies across species, is sensitive to the choice of genomic reference library and bioinformatic
267 pipeline, and that detection rates can rapidly decay over time.

268 Compared to eDNA metabarcoding, camera traps offer unparalleled detail on species abundances, tem-
269 poral visitation patterns, co-occurrences, and the potential for the identification of individuals (Burton et al.,
270 2015). However, we show that eDNA metabarcoding can also provide a useful index of spatio-temporal
271 overlap of species, which may be used to inform management of waterborne diseases in artificial surface
272 waters. In addition, eDNA approaches offer many unique advantages. Water sampling requires a very small
273 amount of time in the field, meaning that multiple distant sites can be covered in a single day. Importantly,
274 eDNA-based approaches also allow for the identification of cryptic species, such as pathogenic microbes
275 (see Farrell et al. (2019)). We suggest the two approaches are thus highly complementary, and when used
276 together can track trajectories of communities over time, providing rich data to help guide applied conser-
277 vation management decisions in the KNP and other savanna ecosystems across sub-Saharan Africa.

278 The long history of conservation and biodiversity monitoring in the KNP allowed us to start with com-
279 plete species lists for vertebrate groups in the park, and generate a curated set of reference sequences. We
280 show improved performance of eDNA detection across species when using our custom reference sequences
281 as compared to existing reference libraries. Our results caution against using pre-configured reference li-
282 braries for eDNA-based biodiversity assessments, especially in areas where comprehensive species lists
283 may be obtained. We suggest that future eDNA metabarcoding efforts focus on leveraging expertise of park
284 scientists and existing natural history collections to improve sequence reference libraries. As methods for
285 sequencing DNA from preserved specimens increase in sensitivity (Prosser et al., 2016), local collections
286 present an increasingly valuable DNA source for expert-identified samples.

287 We found eDNA detection to be temporally limited in our system, with highest detection probability for
288 species visiting waterholes within 24-36 hours of sampling. For example, some species that infrequently
289 visit waterholes, such as lion, are successfully detected when they visit within 24 hours prior to sampling,
290 while other species that visit in higher numbers may be missed by eDNA simply because they did not
291 visit just prior to sampling (see Fig. SM5 for illustrative time series). We observed intense usage of these
292 waterholes, with near complete emptying in some cases, and it appears that previously deposited eDNA is
293 being actively removed through subsequent drinking, rather than molecular degradation. Thus increasing
294 both the frequency of sampling and the volumes filtered would be needed to increase the rate of species
295 detections with eDNA sequencing. In addition, it may be possible to target particular species by sampling
296 at times of day that coincide with daily activity patterns.

297 Beyond time since last visit, we identified additional species variation in eDNA detection. Hyena, hip-
298 popotamus, leopard, impala, and lion were estimated to have slightly elevated species effects, indicating that
299 we have a higher probability of detection. Conversely, we find a lower probability of detecting wildebeest,
300 zebra, giraffe, and baboons. The species level effects capture other qualitative observations, most obviously,
301 our inability to detect wildebeest, and species with relatively small body mass (< 50kg), such as baboons,
302 jackals, civets, and vervet monkeys, despite having representative sequences in our reference library. Low
303 detection rates may reflect molecular limitations, such as primer bias, environmental factors such as water
304 conductivity (Collins et al., 2018), and/or unmeasured ecological or behavioural traits, such as the poten-
305 tial for species to be “messy” or “clean” drinkers which influences rates of DNA shedding. Because of
306 such intrinsic site and species differences, it is likely that even well-designed sampling strategies will miss
307 some species, and in systems where there is uncertainty about true species occurrences, absence of evidence
308 should not be taken as evidence of absence.

309 We attribute the temporal decay of eDNA in our study in large part to the overall visitation rates and rapid
310 removal and turnover of water through drinking. In the Kruger, artificial waterholes are a key management
311 issue as they have led to an overabundance of drought-intolerant species to the detriment of many drought-
312 tolerant (Smit et al., 2007). Through camera trapping we show that the artificial waterholes in the KNP are
313 dominated by three drought-intolerant species (impala, zebra, and elephants), which had orders of magnitude

314 higher visitation rates compared to other species observed at the waterholes. While this extreme skew
315 in waterhole use is likely reducing the ability for eDNA to capture longer term dynamics of waterhole
316 visitation, the drinking ecology of other species, such as rhinos and hyenas, make them good targets for
317 eDNA sampling. Even with the imperfect detection of eDNA metabarcoding, it provides an effective method
318 for sampling a large number of sites, requiring less time in the field, and with less potential for mechanical
319 failure or tampering, as was common with our camera traps.

320 While high temporal turnover of eDNA in these sites may necessitate increased sampling frequency to
321 capture total species richness at a site, it offers a unique opportunity to study cryptic species associations.
322 We find a significant, if noisy, correlation when comparing detections by both eDNA and camera trapping
323 across the full time window, which increases when restricting to animals seen within 36 hours before sam-
324 pling, indicating that eDNA may provide a crude proxy for species spatial overlap within time windows
325 relevant for potential cross-species microbial or pathogen transmission via waterholes (see (Farrell et al.,
326 2019)). We suggest that a large proportion of microbial taxa in these waterholes may be animal-associated,
327 representing pathogens, or part of the oral, fecal, or skin microbiomes of visiting host species. In addition
328 to pathogens, artificial waterholes may be sources for horizontal transfer of mutualist or commensal taxa, or
329 antibiotic resistance plasmids carried from outside of the park (Mariano et al., 2009). Thus, while fine-scale
330 visitation patterns are better assessed by camera traps and the short persistence time of eDNA in waterholes
331 limits our ability to characterize waterhole usage beyond a few days of activity, eDNA sequencing offers
332 an opportunity to use non-invasive sampling to identify host-associated microbes, and examine how water
333 provisioning may be impacting pathogen and microbiome dynamics.

334 Our study illustrates the utility of eDNA as a tool for biodiversity monitoring, and how it offers an
335 added understanding of species ecology through pairing with camera trapping. As methods improve with
336 respect to choice of marker gene, primers, reference libraries, and sequencing depth we expect increased
337 detection ability (Ushio et al., 2017; Lyet et al., 2021), but similar biases to those highlighted here will
338 likely persist. Nonetheless, we have shown that easily-collected water samples can provide reliable data
339 for tracking species of conservation concern in the KNP, such as rhinoceros Ferreira et al. (2015), and the
340 critically endangered white-backed vulture (Murn et al., 2013). For the Kruger and other African savanna
341 ecosystems, we hope our study lays the groundwork for future exciting work exploring cryptic biodiversity,
342 the impacts of water provisioning on rare and threatened taxa, and the dynamic nature of species associations
343 in these vital conservation areas.

344 **Data & Code Accessibility**

345 Scripts to generate the reference library, format public references for use with DADA2, perform all analy-
346 ses, and generate the figures are available at github.com/maxfarrell/eDNAcamtrap (archived at 10.5281/zen-
347 odo.6977480), along with data necessary to reproduce the analyses. Raw sequence data per sample are
348 available via the Sequence Read Archive (SRA BioProject PRJNA490450).

349 **Author Contributions**

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

354 **Acknowledgements**

355 This work would not have been possible without logistical support from the African Centre for DNA Bar-
356 coding and Purvance Shikwambana, molecular work by Shannon Eagle, photographic identifications by
357 Cody Danaher, species inventories from the Kruger National Park Museum staff, and the knowledge and
358 protection provided by Kruger game guards Thomas, Velly, and Martin. We also thank Steven Kembel, his
359 lab, Alexis Carteron, Joanne Littlefair, and Benjamin Callahan for constructive feedback on bioinformatic
360 analyses. Special thanks to B.S. Bezeng for his friendship, dedication, and long days in the field. This work
361 was conducted with support and permission from SANParks (project DAVIJ1256). MJF was supported by
362 a Vanier NSERC CGS, the Vineberg Fellowship, the CIHR SBTP, and project funding from the Quebec
363 Centre for Biodiversity Science, McGill Biology, and an NSERC Discovery Grant awarded to TJD.

364 References

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

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

369 Berry, H., Fox, V., and Berry, P. (2001). Synchrony of drinking in Double-banded Sandgrouse, *Pterocles*
370 *bicinctus*, at Etosha National Park, Namibia. *Ostrich*, 72(1-2):109–113.

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

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

375

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

378 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016).
379 DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–
380 583.

381 Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. (2017b). Bioconduc-
382 tor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research*,
383 5(3):1492.

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

387 Chamaillé-Jammes, S., Charbonnel, A., Dray, S., Madzikanda, H., and Fritz, H. (2016). Spatial Distribution
388 of a Large Herbivore Community at Waterholes: An Assessment of Its Stability over Years in Hwange
389 National Park, Zimbabwe. *PLOS ONE*, 11(4):e0153639.

390 Collins, R. A., Wangensteen, O. S., O’Gorman, E. J., Mariani, S., Sims, D. W., and Genner, M. J. (2018).
391 Persistence of environmental DNA in marine systems. *Communications Biology*, 1(1):1–11.

392 Curtis, A. N., Larson, E. R., and Davis, M. A. (2021). Field storage of water samples affects measured
393 environmental DNA concentration and detection. *Limnology*, 22(1):1–4.

394 Egeru, A., Wasonga, O., MacOpiyo, L., Mburu, J., Tabuti, J. R. S., and Majaliwa, M. G. J. (2015). Piospheric
395 influence on forage species composition and abundance in semi-arid Karamoja sub-region, Uganda. *Pas-
396 toralism*, 5(1):12.

397 Farrell, M. J., Govender, D., Hajibabaei, M., van der Bank, M., and Davies, T. J. (2019). Bacterial diversity
398 in the waterholes of the Kruger National Park: an eDNA metabarcoding approach. *Genome*, 62(3):229–
399 242.

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

403 Ferreira, S. M., Greaver, C., Knight, G. A., Knight, M. H., Smit, I. P. J., and Pienaar, D. (2015). Disruption
404 of rhino demography by poachers may lead to population declines in kruger national park, south africa.
405 *PLoS One*, 10(6):1–18.

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

408 Franz, M., Kramer-Schadt, S., Greenwood, A. D., and Courtiol, A. (2018). Sickness-induced lethargy
409 can increase host contact rates and pathogen spread in water-limited landscapes. *Functional Ecology*,
410 32(9):2194–2204.

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

414 Gibson, J., Shokralla, S., Porter, T. M., King, I., Konynenburg, S. v., Janzen, D. H., Hallwachs, W., and
415 Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample
416 of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*,
417 111(22):8007–8012.

418 Glover-Kapfer, P., Soto-Navarro, C. A., and Wearn, O. R. (2019). Camera-trapping version 3.0: current
419 constraints and future priorities for development. *Remote Sensing in Ecology and Conservation*, 5(3):209–
420 223.

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

424 Harper, L. R., Lawson Handley, L., Carpenter, A. I., Ghazali, M., Di Muri, C., Macgregor, C. J., Logan,
425 T. W., Law, A., Breithaupt, T., Read, D. S., McDevitt, A. D., and Häfling, B. (2019). Environmental
426 DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority
427 mammals. *Biological Conservation*, 238:108225.

428 Harrington, R., Owen-Smith, N., Viljoen, P., Biggs, H., Mason, D., and Funston, P. (1999). Establishing the
429 Causes of the Roan Antelope Decline in the Kruger National Park, South Africa. *Biological Conservation*,
430 90:69–78.

431 Harrison, J. B., Sunday, J. M., and Rogers, S. M. (2019). Predicting the fate of eDNA in the environment
432 and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*,
433 286(1915):20191409.

434 Hitchcock, D. (1996). Wildlife observed in Kutse Game Reserve, Botswana, at pans with either artificial or
435 natural water sources. *African Journal of Ecology*, 34(1):70–74.

436 Jo, T., Arimoto, M., Murakami, H., Masuda, R., and Minamoto, T. (2020). Estimating shedding and decay
437 rates of environmental nuclear DNA with relation to water temperature and biomass. *Environmental DNA*,
438 2(2):140–151.

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

444 Leempoel, K., Hebert, T., and Hadly, E. A. (2020). A comparison of eDNA to camera trapping for as-
445 sessment of terrestrial mammal diversity. *Proceedings of the Royal Society B: Biological Sciences*,
446 287(1918):20192353.

447 Lyet, A., Pellissier, L., Valentini, A., Dejean, T., Hehmeyer, A., and Naidoo, R. (2021). eDNA sampled from
448 stream networks correlates with camera trap detection rates of terrestrial mammals. *Scientific Reports*,
449 11(1):11362.

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

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

454 Mariano, V., McCrindle, C. M. E., Cenci-Goga, B., and Picard, J. A. (2009). Case-Control Study To
455 Determine whether River Water Can Spread Tetracycline Resistance to Unexposed Impala (*Aepyceros*
456 *melampus*) in Kruger National Park (South Africa). *Applied and Environmental Microbiology*, 75(1):113–
457 118.

458 Meusnier, I., Singer, G. A., Landry, J.-F., Hickey, D. A., Hebert, P. D., and Hajibabaei, M. (2008). A
459 universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, 9(1):214.

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

462 Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., and Clune, J.
463 (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with
464 deep learning. *Proceedings of the National Academy of Sciences*, 115(25):E5716–E5725.

465 Peixoto, S., Chaves, C., Velo-Antón, G., Beja, P., and Egeter, B. (2021). Species detection from aquatic
466 eDNA: Assessing the importance of capture methods. *Environmental DNA*, 3(2):435–448.

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

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

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

473 R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Founda-
474 tion for Statistical Computing, Vienna, Austria.

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

477 Sales, N. G., Kaizer, M. d. C., Coscia, I., Perkins, J. C., Highlands, A., Boubli, J. P., Magnusson, W. E.,
478 Da Silva, M. N. F., Benvenuto, C., and McDevitt, A. D. (2020a). Assessing the potential of environmental
479 DNA metabarcoding for monitoring Neotropical mammals: a case study in the Amazon and Atlantic
480 Forest, Brazil. *Mammal Review*, 50(3):221–225.

481 Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., Wangensteen, O. S.,
482 Baillie, C., Bryce, E., Dawson, D. A., Ochu, E., Häfling, B., Lawson Handley, L., Mariani, S., Lambin,
483 X., Sutherland, C., and McDevitt, A. D. (2020b). Fishing for mammals: Landscape-level monitoring of
484 terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*,
485 57(4):707–716.

486 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A.,
487 Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J.,
488 and Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported
489 software for describing and comparing microbial communities. *Applied and Environmental Microbiology*,
490 75(23):7537–7541.

491 Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., and Greenwood, A. D.
492 (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African
493 waterholes. *Molecular Ecology Resources*, 19(6):1486–1496.

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

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

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

503 Turner, W. C., Kausrud, K. L., Beyer, W., Easterday, W. R., Barandongo, Z. R., Blaschke, E., Cloete,
504 C. C., Lazak, J., Van Ert, M. N., Ganz, H. H., Turnbull, P. C. B., Stenseth, N. C., and Getz, W. M.
505 (2016). Lethal exposure: An integrated approach to pathogen transmission via environmental reservoirs.
506 *Scientific Reports*, 6(1):27311.

507 Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T.,
508 Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., and Miya, M. (2017). Environmental
509 DNA enables detection of terrestrial mammals from forest pond water. *Molecular Ecology Resources*,
510 17(6):e63–e75.

511 Valeix, M., Fritz, H., Loveridge, A. J., Davidson, Z., Hunt, J. E., Murindagomo, F., and Macdonald, D. W.
512 (2009). Does the Risk of Encountering Lions Influence African Herbivore Behaviour at Waterholes?
513 *Behavioral Ecology and Sociobiology*, 63(10):1483–1494.

514 van der Loos, L. M. and Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities
515 and the methodology involved. *Molecular Ecology*, 30(13):3270–3288.

516 Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid as-
517 signment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*,
518 73(16):5261–5267.

519 Wearn, O. R. and Glover-Kapfer, P. (2019). Snap happy: camera traps are an effective sampling tool when
520 compared with alternative methods. *Royal Society Open Science*, 6(3):181748.

521 Weir, J. (1971). The effect of creating additional water supplies in a Central African National Park. *The*
522 *Scientific Management of Animal and Plant Communities for Conservation*. Blackwell Scientific Publica-
523 *tions, Oxford*, pages 367–385.

Environmental DNA as a management tool for tracking artificial waterhole use in savanna ecosystems

Supplementary Materials

Maxwell J. Farrell^{1*}, Danny Govender^{2,3},
Mehrdad Hajibabaei⁴, Michelle van der Bank^{4,5}, T. Jonathan Davies^{5,6}

¹Ecology & Evolutionary Biology, University of Toronto

²SANParks Scientific Services

³University of Pretoria

⁴Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph

⁵African Centre for DNA Barcoding, University of Johannesburg, South Africa

⁶Botany, Forest & Conservation Sciences, University of British Columbia

*To whom correspondence should be addressed; E-mail: maxwell.farrell@utoronto.ca

1 Supplementary Methods

1.1 Assembling the Kruger Vertebrate COI Reference Library

We assembled a COI KNP reference library by first compiling 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 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. Since barcode sequences available in the BOLD public database (Ratnasingham and Hebert, 2007) are not always found in GenBank (Porter and Hajibabaei, 2018), COI-5P sequences were also downloaded from BOLD via the R package *bold* (Chamberlain, 2017). 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 sequence removed. Sequences were then realigned with MUSCLE and poorly aligned sequences removed. Alignment gaps were removed and sequences were filtered to include only those longer than 249 bp. 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. 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.

1.1.1 Outline of the KNP Reference Library Assembly

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"

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).

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

Table SM1: Counts of mammal species per order included in the Kruger Vertebrate Reference Library, present in the park according to KNP Library species lists, and the percentage of species in the park which are represented in the reference library.

1.2 Primers

Primer Set	Primer Name	Source	Sequence
mam ckt F + 230R (230 bp)	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)	CTTATRTTTRTTATICGIGGRAAIGC
mam ckt (F + R) (130 bp)	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 (310 bp)	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

The PCR cycler conditions for each of the cocktails are as follows:

- mam ckt F + 230R: 94°C for 1m, five cycles of (94°C for 30s, 48°C for 40s, 72°C for 1m) and 30 cycles of (94°C for 30s, 50°C for 40s, 72°C for 1m) 72° for 5m, hold at 4°C.
- mam ckt F + R: 94°C for 1m, five cycles of (94°C for 30s, 50°C for 40s, 72°C for 1m) and 35 cycles of (94°C for 30s, 55°C for 40s, 72°C for 1m) 72° for 5m, hold at 4°C
- BR5: 95°C for 5min, 35 cycles of 94°C for 40s, 46°C for 1min, and 72°C for 30s, and a final extension at 72°C for 5min

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

Table SM2: Table of species identified by eDNA metabarcoding, according to taxonomic assignment method (i.e. according to the KNP reference library at both 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.)

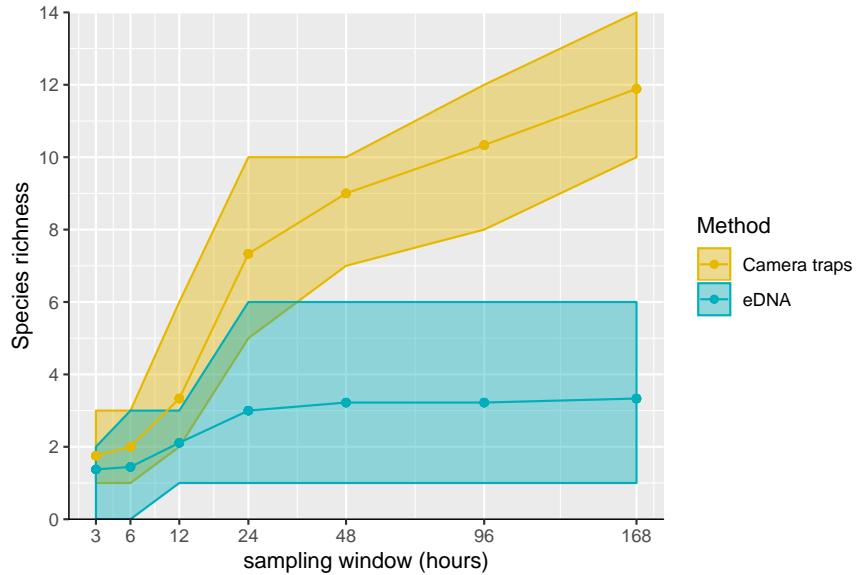


Figure SM1: Species richness by camera traps (yellow) and compared with the number of species also detected by eDNA metabarcoding (blue). Richness is calculated per site and across increasing sampling windows starting from the time of first sampling (e.g. 12 hours would represent all animals seen by camera traps within the first 12 hours after sampling a given site, and the number of these same species also detected in the corresponding eDNA sample). As the time between sampling events increases, camera traps detect increasing numbers of species, whereas the number of unique species also 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%	effective sample size	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

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.

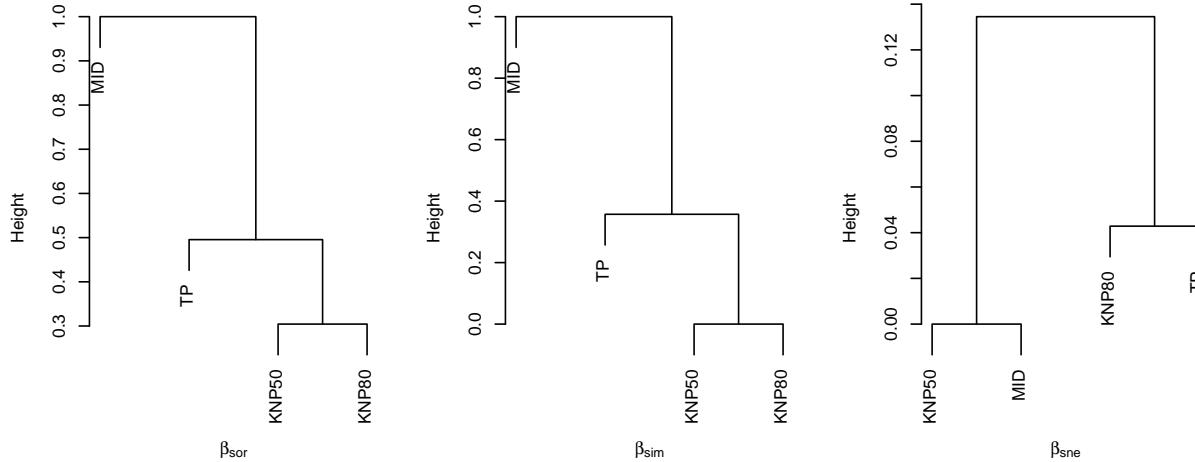


Figure SM2: Similarity in species identified via eDNA metabarcoding as a function of reference library approach. Differences between sets of species identified with each method are analyzed by hierarchical clustering using Sorensen's beta diversity (β_{sor}), which breaks down into a turnover component (Simpson's diversity (β_{sim}), and a nestedness (β_{sne}) component. This approach shows that the two KNP reference library approaches (KNP50 and KNP80) identify similar species, which more closely resembles the results of the Porter Chordata reference library (TP) compared to the MIDORI reference library (MID). The other two metrics indicate that this similarity is largely attributed turnover rather than nestedness, indicating that the KNP reference library is identifying different species, rather than a subset of species identified by the other approaches. Analyses conducted using the “betapart” R package (Baselga and Orme, 2012).

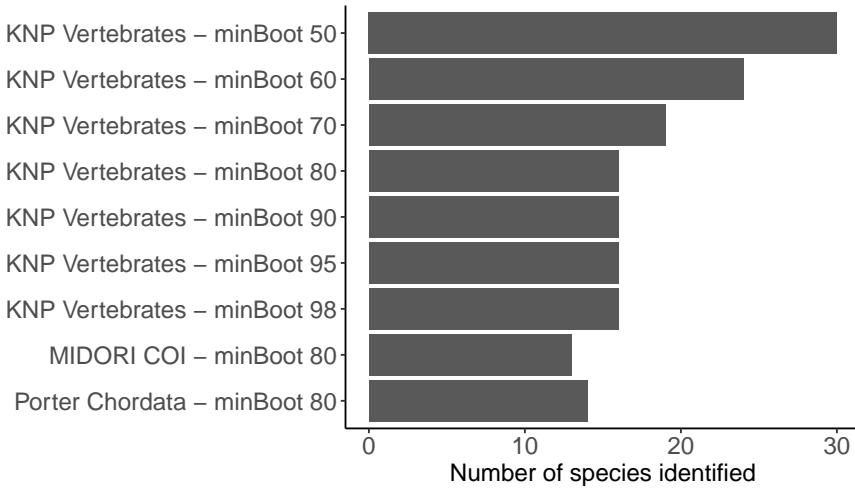


Figure SM3: Total vertebrate species richness identified by eDNA metabarcoding across all samples according to taxonomic reference library (Kruger Reference Library (KNP), MIDORI COI (MID), and Porter Chordata (TP), and the minimum bootstrap (minBoot) value used when applying the RDP classifier via dada2.

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

Table SM4: Per species rates of eDNA (KNP Vertebrate Reference 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
<i>Hippopotamus amphibius</i>	1	1	1.00
<i>Syncerus caffer</i>	1	1	1.00
<i>Crocuta crocuta</i>	4	6	0.67
<i>Ceratotherium simum</i>	3	5	0.60
<i>Aepyceros melampus</i>	2	4	0.50
<i>Equus quagga</i>	2	4	0.50
<i>Loxodonta africana</i>	2	4	0.50
<i>Canis mesomelas</i>	0	3	0.00
<i>Civettictis civetta</i>	0	1	0.00
<i>Giraffa camelopardalis</i>	0	1	0.00

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

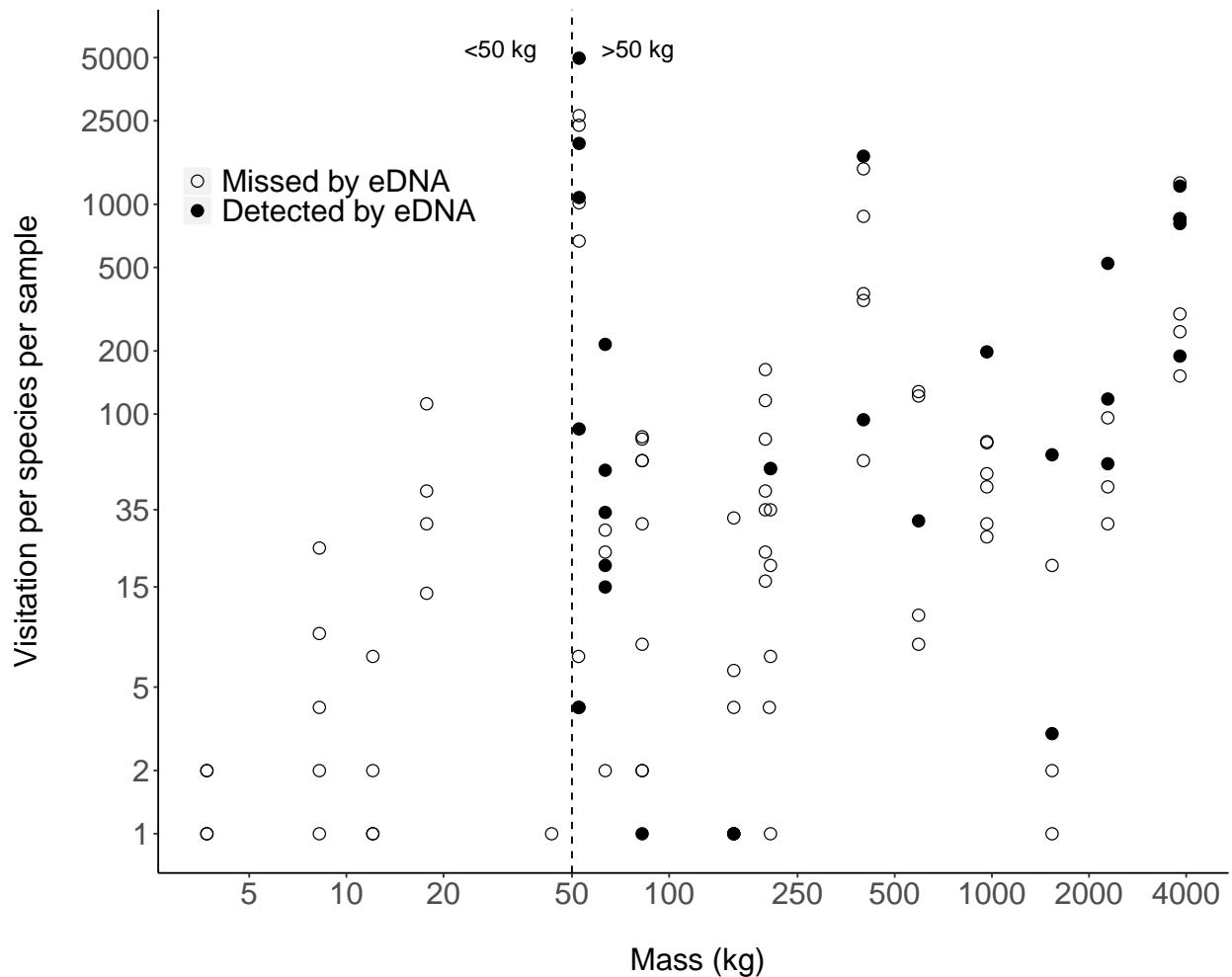


Figure SM4: Raw data plot showing the relationship between species mass (average adult female mass in kilograms) and the total number of animals of that species observed via camera trapping. Filled points indicate samples which have positive eDNA detection for the given species.



Figure SM5: Example of visitation timeseries for five species based one week of camera trapping. Data from a single site is used for demonstration. The red dashed line indicates the time water was sampled. Orange lines indicate abundances of species that were seen in the camera traps but not detected by eDNA, while turquoise lines represent species detected by both camera trapping and eDNA, as in Fig. 3

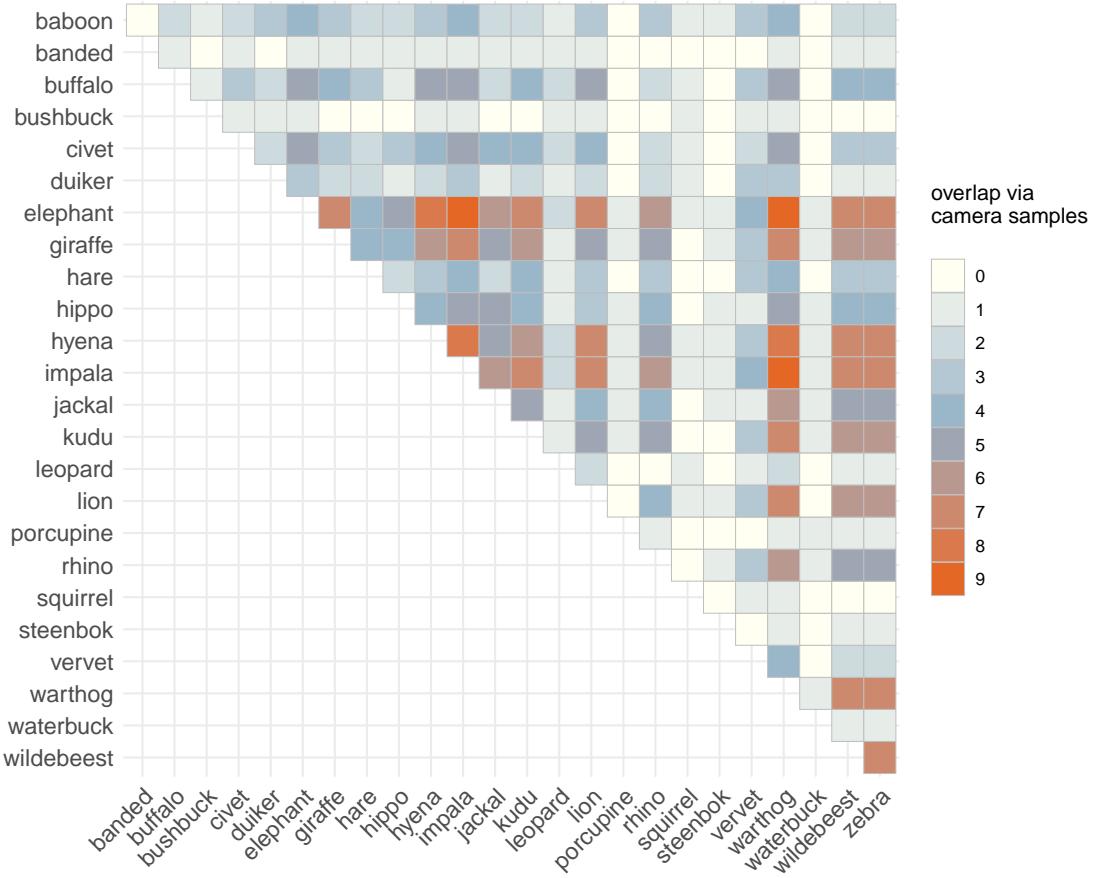


Figure SM6: Heatmap of camera-trap observed overlap among pairs of mammals. Overlap among species pairs is quantified by the number of weekly samples for which both species were documented by camera traps. For example, impalas and elephants were seen at the same sites in 9 weekly samples whereas baboons and banded mongooses were never seen at any of the same sites.

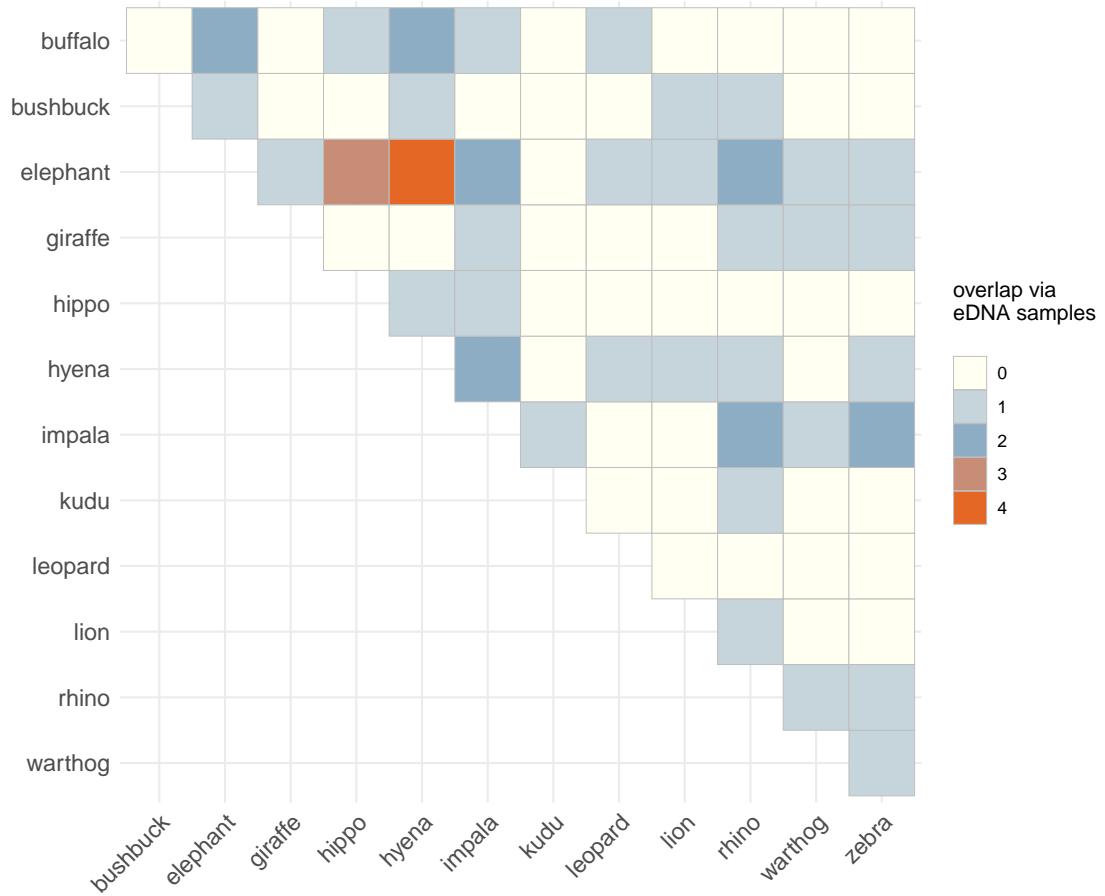


Figure SM7: Heatmap of eDNA inferred pairwise overlap among species present at waterholes. Overlap among species pairs is quantified by the number of eDNA samples in which metabarcoding simultaneously detected both species. For example, eDNA from elephants and impala were detected in 4 weekly samples, whereas eDNA from baboons and banded mongooses were never detected in any of the same samples.

References

Baselga, A. and Orme, C. D. L. betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, 3(5):808–812, 2012.

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

Chamberlain, S., Szoechs, 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.

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.

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.

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

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

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

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