Open access • Posted Content • DOI:10.1101/2021.07.15.452392

Measuring biodiversity from DNA in the air — Source link

Elizabeth L. Clare, Elizabeth L. Clare, Chloe K. Economou, Frances J. Bennett ...+5 more authors

Institutions: Queen Mary University of London, York University

Published on: 15 Jul 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Environmental DNA, Biodiversity, Population, Aquatic biomonitoring and Endangered species

Related papers:

- Global terrestrial distribution of penguins (Spheniscidae) and their conservation by protected areas
- Biodiversity in Mediterranean-climate streams of California
- Community structure and diversity of tropical forest mammals: data from a global camera trap network.
- Estimating biodiversity changes in the Camargue wetlands: An expert knowledge approach.
- Synoptic tinkering: integrating strategies for large-scale conservation

Share this paper:

View more about this paper here: https://typeset.io/papers/measuring-biodiversity-from-dna-in-the-air-5b7a93vdmi

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1
2 ***Measuring biodiversity from DNA in the air***
3
4 Elizabeth L. Clare*,[1,2] Chloe K. Economou[1], Frances J. Bennett[1], Caitlin E. Dyer[1], Katherine
5 Adams[3], Benjamin McRobie[3], Rosie Drinkwater[1], Joanne E. Littlefair[1]
6
7 1. School of Biological and Chemical Sciences, Queen Mary University of London, London. UK,
8 E14NS
9 2. Department of Biology, York University, Toronto, Ontario, Canada, M3J1P3
10 3. Hamerton Zoo Park, Huntingdon, UK, PE28 5RE
11
12 **Corresponding author:** Elizabeth L Clare
13 Department of Biology, York University, Toronto, Ontario, Canada, M3J1P3
14 519-503-3898
15 **Email:** eclare@yorku.ca
16
17 Deposited as a preprint at bioRxiv
18
19 **Author Contributions:** ELC designed the experiment and performed field work, CKE performed
20 all laboratory work, FJB performed field work, CED, RD and JEL performed analysis, KA and BM
21 facilitated field work. All authors contributed to manuscript preparation.
22
23 **Competing Interest Statement:** Authors declare no competing interests.
24
25 **Classification:** Biological Sciences: Ecology
26
27 **Keywords:** airDNA, environmental DNA, eDNA, biosurvey
28
29 **This PDF file includes:**
30 Main Text
31 Figures 1 to 2
32 Tables 1 to 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Impacts of the biodiversity crisis far exceed our ability to monitor changes in terrestrial ecosystems. Environmental DNA has revolutionized aquatic biomonitoring, permitting remote population and diversity assessments. Here we demonstrate that DNA from terrestrial animals can now be collected from the air under natural conditions, a ground-breaking advance for terrestrial biomonitoring. Using air samples from a zoological park, where species are spatially confined and unique compared to native fauna, we show that DNA in air can be used to identify the captive species and their potential interactions with local taxa. Air samples contained DNA from 25 species of mammal and bird including 17 known (and distinct) terrestrial zoo species. We also identified food items from air sampled in enclosures and detected four taxa native to the local area, including the Eurasian hedgehog, endangered in the UK, and the muntjac deer, a locally established invasive species. Our data provide evidence that airDNA is concentrated around recently inhabited areas (e.g., indoor enclosures) but that there is dispersal away from the source suggesting an ecology to airDNA movement which highlights the potential for airDNA sampling at distance. Our data clearly demonstrate the profound potential of air as a source of DNA for global terrestrial biomonitoring and ecological analysis.

Significance Statement: The global decline in biodiversity requires rapid non-invasive biomonitoring tools applicable at a global scale. In this study we collect environmental DNA from mammals and birds from air samples collected in a natural setting. Using only air, we identified 25 species of mammal and bird known to be in the area. Our dataset detected species at risk of local extinction and several confirmed predator-prey interactions. This approach will revolutionize terrestrial biodiversity surveys.

Main Text

Introduction

Anthropogenic impacts have caused pervasive biodiversity declines across ecosystems (Díaz et al., 2019; C. N. Johnson et al., 2017; Seibold et al., 2019), particularly from land-use change, habitat loss and degradation (Tilman et al., 2017) leading to the reorganization of global biodiversity patterns and processes (Barlow et al., 2018; Eriksson & Hillebrand, 2019). Rapid and accurate biomonitoring techniques are essential to our attempts to quantify the causes and consequences of global environmental change (Amano et al., 2018; Eriksson & Hillebrand, 2019) and to assist with focused, on-the-ground conservation efforts. Our inability to detect species and measure population dynamics rapidly and accurately is often cited as a fundamental challenge in quantifying our position relative to biodiversity and conservation targets (Amano et al., 2018; C. N. Johnson et al., 2017). Indeed, detecting changes in diversity, abundance, and community composition as well as species range shifts are priorities highlighted by researchers, conservationists, and major international initiatives (Amano et al., 2018). New approaches that provide simpler, large-scale, and automated monitoring techniques are an urgent requirement, needed to address the often-intractable challenge of biodiversity monitoring (Barlow et al., 2018). Decades of development in molecular diagnostics have resulted in established DNA-based approaches for determining species (Blaxter, 2003; Hebert, Cywinska, Ball, & DeWaard, 2003; Tautz et al., 2002) and detecting ecological interactions (Pompanon et al., 2012). The development of DNA reference databases, which permit rapid species identification from unknown environmental samples of (even fragmentary) genetic material, has the potential to transform our ability to monitor global ecosystems. Yet biodiversity monitoring often still relies on capture of live specimens, which is both rate limiting and invasive (Singer, Fahner, Barnes, McCarthy, & Hajibabaei, 2019).

It is well documented that DNA is shed from all organisms and deposited as environmental (e)DNA. This material has been used to analyze contemporary and past ecosystems for nearly two decades (Willerslev et al., 2007, 2014, 2003). An explosion of interest in using aquatic eDNA to assay populations and track invasive species has revolutionized aquatic science, management, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

87 conservation (Ruppert, Kline, & Rahman, 2019). As the field matures, considerable research effort
88 now focuses on the "ecology of eDNA" (Barnes & Turner, 2016) – quantifying and understanding
89 factors influencing eDNA detections beyond inventories alone. Comparative studies have shown
90 that metabarcoding of aquatic eDNA matches or even outperforms conventional methods of
91 community sampling (Bessey et al., 2021; Mena et al., 2021; Ruppert et al., 2019). Additionally, an
92 indication of terrestrial biodiversity can also be obtained from eDNA analysis of water and
93 sediments sampled from aquatic systems (Sales et al., 2020; Ushio et al., 2017), though detections
94 may be biased towards semi-aquatic species. A comparison of tropical mammal detection methods
95 (Mena et al., 2021) including eDNA from lentic and lotic systems, live-trapping, pitfall traps, camera
96 traps, and mistnets found integrated methods provide best estimates of community composition.
97 Although aquatic eDNA alone recovered much of the diversity of mammals (Mena et al., 2021) this
98 would be limited when aquatic systems are not in the vicinity.
99
100 A truly terrestrial targeted eDNA system has not yet been developed. On land, eDNA has been
101 measured in permafrost, blood, snow, soil, and honey (Bohmann et al., 2014) and recently by
102 spraying foliage and collecting the runoff to gather eDNA from the surfaces (Valentin et al., 2020).
103 Collecting eDNA from the air, analogous to aquatic eDNA sampling, has remained mostly
104 theoretical (Ruppert et al., 2019) with a few demonstrations recovering DNA from plants (Folloni et
105 al., 2012; M. D. Johnson, Cox, & Barnes, 2019) or fungi (Banchi et al., 2020), although these were
106 based primarily on the analysis of collected dust. We recently demonstrated (Clare et al., 2021)
107 that animal DNA can be extracted directly from air under highly controlled laboratory conditions.
108 The potential for sampling life from air samples could revolutionize terrestrial biodiversity
109 assessments, but to date it has not been tested in the wild. The challenge with validating airDNA
110 methods is establishing an experimental design that permits spatial scales for detection without
111 confounding DNA sources. Zoological parks are ideal for this because they contain captive colonies
112 of mostly non-native species whose identity and spatial location are known with certainty. Indeed,
113 metabarcoding of soils from safari parks, zoological gardens and farms has been used to test the
114 efficacy of these approaches and results have reflected the overall taxonomic richness of terrestrial
115 vertebrates present (Anderson-Carpenter et al., 2011).
116
117 Our objective is to use air samples from a zoological park in Huntingdonshire UK to identify zoo
118 species and native wildlife in the first practical application of airDNA sampling under natural
119 conditions. This approach will greatly extend the validation of this technique for global terrestrial
120 biomonitoring and establishes the potential uses of airDNA in ecological systems.
121

122 Materials and Methods

123

124 Methods

125 *Sample Collection*

126 This study was conducted at Hamerton Zoo Park, a 25-acre conservation zoo in Huntingdonshire
127 UK established in 1990 and containing approximately 100 species of animal, mostly mammals and
128 birds of conservation concern. It is surrounded by a matrix of agricultural land in rural England.
129 Most species live in enclosures which have free access to outside ranges allowing free air
130 exchange. Air samples were collected using a peristaltic pump (Geotech) and Sterivex-HV filtered
131 (Merck Millipore) with 0.22 µm and 0.45 µm filter sizes. We targeted 15 enclosures which contained
132 zoo species represented in molecular reference collections. For each of these locations, we
133 sampled air for 30 min at 300ml/min filter rate using each filter and we sampled inside an enclosure
134 (e.g. the sleeping chamber) and outside in the open air enclosure (where species move about
135 freely) within 5m of the enclosure opening. We also sampled from general areas of the zoo including
136 the Cat Circle, Owl walkway and near rubbish bins. In addition, we sampled at the Tasmanian
137 Golden Possum Enclosure and Syrian Brown Bear Enclosure, but Golden Possums were not
138 represented in reference databases and the bears were in their hibernation cycle and closed to
139 close sampling (i.e. no indoor samples were taken). We treat these two areas as general areas for

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

140 sampling. All filters were placed in sterile bags following sample collection and frozen for DNA
141 extraction.
142
143 ***DNA extraction***
144 DNA extraction and PCR were carried out within a biological safety cabinet under maximum flow.
145 All extraction procedures followed (Clare et al., 2021). In general, all equipment was sterilized using
146 UV, 10% bleach, 70% ethanol and ultrapure water between each sample. Following existing
147 protocols (Cruaud et al., 2017), the filter was cracked open and the filter removed. DNA was
148 extracted using a Blood and Tissue kit (Qiagen UK) following manufacturer's protocol but with ATL
149 buffer volumes increased to 450µl to ensure the filters were submerged. We used 50µl proteinase
150 K, 500µl buffer AL and 500µl of 100% ethanol. We used multiple negative controls at the extraction,
151 PCR and sequencing stages. Samples were lysed overnight using a platform shaking at 650rpm at
152 56°C. The samples were then vortexed and transferred to fresh tubes for extraction. We used QIA
153 shredder spin columns (Qiagen UK) on the remaining filter paper and the flow-through was added
154 to the rest of the sample at which point buffer AL was added. Extraction then followed manufacturer
155 instructions but with centrifugation completed at 11,000 rpm for 3 min following AW2. DNA was
156 eluted in 30 µl of elution buffer pre heated to 70 °C. Elution buffer was cycled through the column
157 three times with 5 min incubation times in each cycle to increase DNA concentration.
158
159 ***PCR amplification and sequencing***
160
161 Each DNA extract was subjected to three PCRs as follows:
162
163 *16S PCR* **-** We amplified a small region of 16S mammal mitochondria using the mam1 and mam2
164 primers (Taylor, 1996) modified with adaptors for the Illumina MiSeq sequencing platform. The PCR
165 mix included 7.5µl of Qiagen multiplex mix, 1.5µl ddH$_2$O, 5µl of template DNA and 0.5µl of each
166 primer (10µM stocks of each) and amplification used cycling conditions of 95°C for 15min, 40 cycles
167 of 94°C for 30s, 55°C for 90s, 72°C for 90s, and a final 72°C for 10min and a 10°C hold.
168
169 *16S nested PCR* **-** To increase amplification success for low yield sample we performed nested
170 PCRs. For the nested 16S PCRs we first used non-tagged mam1 and mam2 primers. For these
171 reactions, we used 3µl of template DNA (adjusting the amount of water accordingly) and increased
172 the annealing temperature to 59°C. We then used 1µl of each PCR product from the first reaction
173 as a template for a second PCR, again using the same 16S mam1 and mam2 Illumina MiSeq
174 tagged primers. PCR conditions for the second PCR were as previously mentioned.
175
176 *COI nested PCR* **-** We amplified a small portion of the 5' end of the cytochrome oxidase gene using
177 AquaF2 forward and VR1d reverse primers (Ivanova, Clare, & Borisenko, 2012). We employed a
178 two-step nested PCR strategy. For the first stage PCR, the PCR mix we used comprised 7.5µl
179 Qiagen multiplex mix, 3.5µl ddH2O, 3µl of template DNA and 0.5 µl of each primer (10 µM stocks
180 of each). For the majority of samples, we used 1µl of PCR product from this first reaction as a
181 template for a second PCR using AquaF2 and VR1d Illumina MiSeq tagged primers. For selected
182 samples with significant non-target bands, we gel extracted the target band from the first PCR
183 (Monarch DNA Gel Extraction Kit) and used 1 µl of this purified DNA in the second PCR. Reaction
184 conditions for both first and second PCRs were as follows: 95°C for 15min, followed by 40 cycles
185 of 94°C for 30s, 51°C for 90s, and 72°C for 90s and a final extension at 72°C for 10min, then a hold
186 at 10°C.
187
188 ***PCR visualization and sequencing***
189 All products, including positive (cow DNA) and negative controls were visualized using a 1%
190 agarose gel as an initial screening tool and then quantified using Qubit and Tapestation. Amplicons
191 were sequenced on an Illumina MiSeq using unique 5' forward tags at the Barts and the London
192 Genome Centre following standard protocol using bidirectional 250bp chemistry. The results were
193 demultiplexed by tag for bioinformatics processing.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

194
195 ***Bioinformatics methods for COI regions***
196 COI read files were uploaded to the mBRAVE platform (http://www.mbrave.net). Paired end
197 samples were assembled with a minimum overlap of 20bp and max substitution of 5bp. Samples
198 were processed to maximize data retention for later steps with the following parameters, Trim
199 Front=38bp, Trim End=26bp, Trim Length=500bp, Min QV filter=0, Min Length=100bp, Max bases
200 with low (<20) QV=75%, Max bases with ultra low QV (<10)=75%. ID threshold=10%, Exclude from
201 OTU at 10% MIN OTU size=1 and OTU threshold=2%.
202
203 The reads were compared to the "Hamerton Zoo 1" bespoke reference database consisting of 610
204 sequences representing 20 species known to reside at the zoo and targeted in our sampling. These
205 data were taken from existing public data in the BOLD database. Sequences not identified by
206 comparison to this bespoke reference collection were then screened in sequential order to system
207 reference libraries:
208 SYS-CRLCHORDATA (Chordata references) consisting of 40,565 species
209 SYS-CRLAVES (Aves reference) consisting of 5832 species
210 SYS-CRLBACTERIA (Bacteria reference) consisting of 2066 species
211 SYS-CRLFUNGI (Fungi reference) consisting of 565 species
212 SYS-CRLINSECTA (Insect reference) consisting of 217,994 species
213 SYS-CRNONINSECTARTH (Non-Insect Arthropoda reference) consisting of 27,832 species
214 SYS-NONARTHINVERT (Non-Arthropoda Invertebrate reference) consisting of 34,927 species
215 SYS-CRLPROTISTA (Protista COI reference collection) consisting of 5250 species
216
217 ***Bioinformatics methods for 16S regions***
218 We used AdapterRemoval V2 (Schubert, Lindgreen, & Orlando, 2016) to first identify and then
219 remove adapter contamination, using the additional parameters --trimns and --trimqualities, to
220 remove Ns and runs of low quality bases. Read pairs were not collapsed at this step. We processed
221 the remaining reads into amplicon sequence variants (ASVs) using the DADA2 pipeline (Callahan
222 et al., 2016) in R ("R Development Core Team: R: A language and environment for statistical
223 computing," 2021; RStudio Team, 2020). We filtered the reads using DADA2 with the following
224 parameters: truncate length after 100 bases in both directions (truncLen=c(100,100)), reads with
225 any Ns were removed (maxN=0), reads higher than expected error removed (maxEE=c(2,2)),
226 truncate reads based on low quality scores (truncQ=2) and discard phiX genes (rm.phix=TRUE).
227 Each of the filtered read pairs were dereplicated, the amplicon error rate was estimated, and the
228 core algorithm was used to calculate the true ASVs counts in the data. Finally, read pairs were
229 merged, ASVs in each sample were counted and chimeric sequences were removed.
230
231 Final 16S ASVs were blasted against a local subset of the GenBank database (search term: "16S",
232 downloaded 23rd May 2021, 467,306 records), with >97% identity and output hits limited to 15
233 sequences. We manually discarded hits with low query coverage (<90%). We then applied BASTA
234 (a last common ancestor algorithm) to the resulting hits, configured to return a majority taxonomy
235 from 90% of the hits (Kahlke & Ralph, 2019). Because the Tyra (*Eira barbara*) was not represented
236 within the 16S reference data we reran this comparison allowing 96% matches to the nearest
237 ancestor in the reference data *Gulo gulo* (not present in the zoo) and assigned ASVs to *Eira barbara*
238 if there was a 96% match to *Gulo gulo*.
239
240 ***Data filtering***
241 For both COI and 16S data we excluded *Heterocephalus glaber* or *Fukomys damerensis*
242 identifications as expected contamination from the previous experiment using the same equipment
243 (Clare et al., 2021) and we excluded all human sequences which are expected as a general
244 contamination in all samples and controls. We then examined negative well contamination and
245 recorded identifications in negative samples and the number of reads. We differentiated
246 identifications which would remain if largest negative well ID number was used as a filter and treat
247 each of the three amplifications separately (e.g. a negative well with a 500 reads assigned as a

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

248 contamination would cause us to flag any ID with 500 or fewer reads assigned, we treat this
249 maximum read count filter separately for COI, 16S and 16S nested PCRs, Supplemental
250 Information).
251

252 ***Statistical analysis***

253 Read counts from all three PCR procedures were pooled (Supplemental Information 2) and mean
254 read counts/location for each identified zoo species were calculated. We first examined the effect
255 of the sampling position relative to the animal's own enclosure (i.e. inside (n=23) or outside (n=79)
256 the animal's own enclosure) on read counts. In a second model, we examined the relationship
257 between read counts and distance from the animal's enclosure. The distances between the
258 sampling points to the originating enclosures were calculated as a straight line to the nearest meter
259 using google maps satellite view (i.e. the distance between a sampling point which detected tiger
260 DNA and the tiger enclosure). Distance varied from 0 – 276 metres, but we excluded zero distance
261 datapoints (i.e. datapoints from inside the animal's own enclosure, n = 23), as this effect had already
262 been examined by the first model. In both cases we used zero-inflated negative binomial mixed
263 effects models using the glmmTMB package (Brooks et al., 2017) in R version 4.0.2 ("R
264 Development Core Team: R: A language and environment for statistical computing," 2021), with
265 species and filter ID as random effects (filter ID was necessary as we treated read counts from
266 different species from the same filter as different data points). We checked for overdispersion and
267 patterns in the model residuals using the DHARMa package (Hartig, 2021). In both models, we
268 tested the significance of the "sample position" and "distance" terms in explaining the read counts
269 by calculating the likelihood ratio test using the "drop1" function with a chi-squared distribution.
270

271 **Results**

272

273 ***Sample collection***

274 We collected 72 air samples from 20 locations around Hamerton Zoo Park. Of these 64 yielded
275 DNA which was identified as belonging to non-human terrestrial vertebrates with multiple sources
276 represented in most samples (Figure 1). All data produced is available on the NCBI short read
277 archive BioProject ID:PRJNA743788.
278

279 ***16S Data***

280 We recovered 12,207,070 reads after the removal of adapter sequences; these were used as input
281 into the DADA2 bioinformatic pipeline. After length and quality filtering, paired-end merging and
282 removing chimeras, 11,707,400 reads remained assigned to 335 amplicon sequence variants
283 (ASVs). Taxonomic ID of the ASVs was assigned using BLAST and further refined with BASTA
284 using a last common ancestor (LCA) algorithm and based on 97% sequence similarity to the 16S
285 reference database (see methods for *Eira barbara* identification parameters).
286

287 Several ASVs received higher level taxonomic assignments and were resolved as follows. ASVs
288 designated as Artiodacyla were resolved to *Muntiacus reevesi* as the other similar match to a
289 reference was *Cephalophus dorsali* (bay duiker) and is not possible on site. Similarly, ASVs
290 designated a Cervidae were a perfect match to muntjac and a lower match to *Ozotoceros*
291 *bezoarticus* which was not possible on site. We retain muntjac for these as well. ASVs identified as
292 Herpestidae were perfect or highly similar (>99%) matches to *Suircata suiricata* (which was on site)
293 and lower matches (97%) to other species not present thus we designate these as *S. suricata*. An
294 ASV identified as *Saguinus* was resolved to *Saguinus oedipus* based on matches >99% to that
295 species which was present in the zoo while other potential matches were <98%.
296

297 ***COI Data***

298 We recovered 6,167,294 reads from samples amplified by COI primers. These data were
299 processed in the mBRAVE pipeline. Filtered data included 1,061,857 reads that were compared to
300 reference databases. From these 361,889 reads were assigned to a non-human mBRAVE BINs
301 (Ratnasingham & Hebert, 2013) at >97% sequence similarity and resolved to species level based

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 on sequence similarity matches >99% in most cases with the exception of *Canis* where species
303 cannot be easily differentiated. We report these as dingo in Table 1, though it is also possible that
304 domestic dog DNA is present on site.
305
306 ***Negative controls for sequence filtering***
307 We used multiple negative controls at DNA extraction, PCR and as empty wells in the sequencing
308 run. Negative well contamination following filtering was very low. However, there was contamination
309 of black and white lemur (1,373 reads) in a negative sample from the 16S PCR negative, donkey
310 (16,836 reads) in a negative of the nested 16S PCR and chicken (513 reads) in a negative of the
311 COI nested PCR. Therefore, in the Supplemental Information for detections, we highlight any read
312 count larger than these to indicate higher support for the taxonomic assignment. Some expected
313 taxa based on sampling location produced read counts lower than these negative thresholds (e.g.
314 *Panthera tigris*) thus we retain all data in Tables 1,2&3 and Supplemental Information to indicate
315 these very likely positives but treat low copy number identifications with caution. All positive control
316 data (cow) was recovered indicating high PCR efficiency and there was very minimal evidence of
317 cow in negative extraction, PCR controls or empty wells used as sequencing controls suggesting
318 that detections in samples represent real dietary detections.
319
320 ***Statistical analysis***
321 We compared read counts with distance to most likely source using two models (Figure 2). There
322 was a significant effect of the sample position relative to the animal's own enclosure on the read
323 counts (Figure 2A, likelihood ratio statistic = 64.1, df = 1, $p < 0.001$), with read counts inside the
324 animal's own enclosure being higher (model estimate: 54,899 reads, confidence limits: 19225-
325 156766) than read counts outside the animal's own enclosure (model estimate: 689 reads,
326 confidence limits: 301-1578). When datapoints from within the animal's own enclosures were
327 removed (i.e., zero distances), there was no relationship between read count numbers and distance
328 from the enclosure (Figure 2B, likelihood ratio statistic = 3.01, df = 1, $p = 0.0828$). Neither model
329 was overdispersed.
330
331 **Discussion**
332
333 Our objective was to collect DNA from air samples and use these to assay for local biodiversity
334 under natural conditions. Using air samples collected at Hamerton Zoo Park we successfully
335 recovered nearly 2.7 million non-human vertebrate DNA sequences. While our laboratory proof of
336 concept (Clare et al., 2021) allowed us to predict that target species could be detected in confined
337 spaces (i.e., inside a sleeping enclosure), detecting airDNA outside enclosures, away from a
338 source, diluted by the air volume in open areas and subjected to wind and local weather
339 represented a far greater challenge. The success of our study shows the potential of conducting
340 biodiversity surveys in the wild, using airDNA, representing an exciting new frontier in biodiversity
341 monitoring.
342
343 In addition to the taxa we targeted as part of the known zoo stock, we identified three species of
344 mammal and three species of bird known to be housed at the zoo but in enclosures that we did not
345 have access to (Table 2, Table 3). These additional detections were frequently recovered at the
346 closest sampling point to their actual residence. For example, the indoor exhibit housing budgies,
347 *Melopsitacus undulatus*, and zebra finches, *Taenlopygla guttata*, was closed during the sampling
348 period but we detected their DNA in air samples collected at the adjacent primate house and
349 possum enclosure. While DNA read counts were generally highest within the enclosure where they
350 are expected, we picked up trace read counts in air samples taken more than 250 m from the most
351 likely source (Figure 2). While contamination between samples is theoretically possible, samples
352 were collected and processed on different days and high read counts were retained even after
353 stringent filtering by sequence quality and negative controls. For example, meerkat DNA from an
354 outdoor colony was identified in air sampled at the dingo enclosure 245 m away and at the gibbon

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

355 enclosure 122 m away. Copy number of recovered sequences was not related to distance from
356 source when enclosures were excluded (Figure 2).
357
358 More than a third of the recovered sequences matched cow, horse, pig or chicken. While we cannot
359 preclude DNA drifting in from the surrounding countryside, it is likely these represent food provided
360 to the carnivores. Particularly high concentrations of chicken DNA were detected in the binturong
361 and tayra enclosures while horse, cow and pig were concentrated in samples from the dingo
362 enclosure, so correctly associated with dietary preference (Table 2). Detecting species interactions
363 has been a special focus of environmental DNA approaches (Drinkwater et al., 2019; Pompanon
364 et al., 2012) but this is the first time species interactions have been detected from air. We also
365 observed some unexpected concentrations of these DNA sources perhaps reflecting the movement
366 of people and materials throughout the zoo. For example, an unexpected concentration of pig and
367 cow DNA inside the lemur enclosure could reflect the movement of people between animal houses.
368
369 While the primary aim of our study was an inventory of the zoo species, adjacent rural settings are
370 a source for DNA from wildlife *in naturalibus*. We identified DNA associated with squirrels and ducks
371 in several air samples. Several ducks are kept as zoo stock, but we could not identify the genus or
372 species with accuracy, so we classify this as wildfowl but with caution. We may have detected
373 *Myotis* bats, though we also treat this with caution as many bat DNA samples are handled in the
374 processing laboratory facility (Table 2).
375
376 Of special interest was the detection of the European hedgehog in three samples. Hedgehogs are
377 commonly observed on site by staff, though they are not as active in the winter thus their detection
378 is particularly interesting. As of 2020, the hedgehog was listed as vulnerable to extinction in the UK
379 (https://www.mammal.org.uk/science-research/red-list/), making it vital to develop additional
380 methods to monitor and protect existing populations. UK species of special interest such as the
381 great crested newt have been the model for the development of aquatic eDNA detection methods
382 (Rees, Baker, Gardner, Maddison, & Gough, 2017) and provide a framework for validating airDNA
383 for similar monitoring. Another commonly cited application of eDNA approaches is the detection of
384 invasive species. We detected muntjac deer, *Muntiacus reevesi*, in five samples. These muntjacs
385 are native to China but became locally invasive after multiple releases in England in the 19th
386 century (Hemami, Watkinson, & Dolman, 2005). They are now well established in the east of
387 England, the location of the zoological park. They are also provided in food for several species on
388 site thus the detection of *M. reevesi* may reflect either food or wildlife (Table 3).
389
390 Our study provides compelling evidence that air can be used as a source of DNA for biomonitoring.
391 The detection of multiple taxa in air samples known to reside at the zoo without high false positive
392 detections strongly validates the local source of the DNA. The detection of species of conservation
393 concern and invasive species, as well as DNA from dietary items, likely via the detection of
394 aerosolized fecal material, is compelling and demonstrates the versatility of this genetic approach.
395 High negative rates and low DNA extraction volumes and concentrations suggests a future role in
396 pooling replicate samples, as is done in DNA biomonitoring using leeches (Schnell et al., 2018).
397 This can increase positive hit rates while reducing sequencing costs. The rapid global uptake of
398 aquatic eDNA as a biomonitoring tool highlights the need for new sampling techniques. If airDNA
399 sampling is successfully developed it will have major implications for global terrestrial
400 biomonitoring. The novel opportunities this method provides for tracking faunal composition, non-
401 invasive monitoring of species of special ecological concern, and the detection of species invasion
402 are extremely exciting, and suggest that airDNA could revolutionize the ways in which scientists
403 study and monitor terrestrial biodiversity and could be implemented non-invasively at a global scale.
404
405 **Acknowledgments**
406 This study was funded through an EPSRC impact accelerator account provided by Queen Mary
407 University of London (Grant: EP/R511596/1). We appreciate the support of the Hamerton Zoo Park
408 Staff and the Barts and the London Genome Centre for assistance during this project.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

409

410 **References**

411 Amano, T., Székely, T., Sandel, B., Nagy, S., Mundkur, T., Langendoen, T., … Sutherland, W. J.
412 (2018). Successful conservation of global waterbird populations depends on effective
413 governance. *Nature*, *553*(7687), 199–202. doi:10.1038/nature25139
414 Anderson-Carpenter, L. L., McLachlan, J. S., Jackson, S. T., Kuch, M., Lumibao, C. Y., & Poinar,
415 H. N. (2011). Ancient DNA from lake sediments: bridging the gap between paleoecology
416 and genetics. *BMC Evolutionary Biology*, *11*, 30. doi:10.1186/1471-2148-11-30
417 Banchi, E., Ametrano, C. G., Tordoni, E., Stanković, D., Ongaro, S., Tretiach, M., … Lazzarin, S.
418 (2020). Environmental DNA assessment of airborne plant and fungal seasonal diversity.
419 *Science of The Total Environment*, *738*, 140249.
420 doi:https://doi.org/10.1016/j.scitotenv.2020.140249
421 Barlow, J., França, F., Gardner, T. A., Hicks, C. C., Lennox, G. D., Berenguer, E., … Graham, N.
422 A. J. (2018). The future of hyperdiverse tropical ecosystems. *Nature*, *559*(7715), 517–526.
423 doi:10.1038/s41586-018-0301-1
424 Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for
425 conservation genetics. *Conservation Genetics*, *17*, 1–17. doi:10.1007/s10592-015-0775-4
426 Bessey, C., Neil Jarman, S., Simpson, T., Miller, H., Stewart, T., Kenneth Keesing, J., & Berry, O.
427 (2021). Passive eDNA collection enhances aquatic biodiversity analysis. *Communications*
428 *Biology*, *4*(1), 236. doi:10.1038/s42003-021-01760-8
429 Blaxter, M. (2003). Molecular systematics: Counting angels with DNA. *Nature*, *421*(6919), 122–
430 124.
431 Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., … de Bruyn, M.
432 (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in*
433 *Ecology & Evolution*, *29*(6), 358–367. doi:10.1016/j.tree.2014.04.003
434 Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., …
435 Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-
436 inflated generalized linear mixed modeling. *R Journal*, *9*(2), 378–400. doi:10.32614/rj-2017-
437 066
438 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P.
439 (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature*
440 *Methods*, *13*(7), 581–583. doi:10.1038/nmeth.3869
441 Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F., Drinkwater, R., &
442 Littlefair, J. E. (2021). eDNAir: proof of concept that animal DNA can be collected from air
443 sampling. *PeerJ*, *9*, e11030. doi:https://doi.org/10.7717/peerj.11030
444 Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S. J., Rodriguez, M. J., Dorea, C. C., &
445 Culley, A. I. (2017). Open the SterivexTM casing: An easy and effective way to improve
446 DNA extraction yields. *Limnology and Oceanography: Methods*, *15*(12), 1015–1020.
447 doi:https://doi.org/10.1002/lom3.10221
448 Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., … Zayas, C. N. (2019).
449 Pervasive human-driven decline of life on Earth points to the need for transformative
450 change. *Science*, *366*(6471), eaax3100. doi:10.1126/science.aax3100
451 Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., … Rossiter, S. J.
452 (2019). Using metabarcoding to compare the suitability of two blood-feeding leech species
453 for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources*, *19*(1).
454 doi:10.1111/1755-0998.12943
455 Eriksson, B. K., & Hillebrand, H. (2019). Rapid reorganization of global biodiversity. *Science*,
456 *366*(6463), 308–309. doi:10.1126/science.aaz4520
457 Folloni, S., Kagkle, D.-M., Rajcevic, B., Guimaraes, N., Van Droogenbroeck, B., Valicente, F. H.,
458 … Van den Bulcke, M. (2012). Detection of airborne genetically modified maize pollen by
459 real-time PCR. *Molecular Ecology Resources*, *12*(5), 810–821.
460 doi:https://doi.org/10.1111/j.1755-0998.2012.03168.x
461 Hartig, F. (2021). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)
462 Regression Models. Retrieved from http://florianhartig.github.io/DHARMa/

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. doi:10.1098/rspb.2002.2218

Hemami, M.-R., Watkinson, A. R., & Dolman, P. M. (2005). Population densities and habitat associations of introduced muntjac Muntiacus reevesi and native roe deer Capreolus capreolus in a lowland pine forest. *Forest Ecology and Management*, *215*(1), 224–238. doi:https://doi.org/10.1016/j.foreco.2005.05.013

Ivanova, N. V., Clare, E. L., & Borisenko, A. V. (2012). *DNA barcoding in mammals*. *Methods in Molecular Biology* (Vol. 858). doi:10.1007/978-1-61779-591-6_8

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science*, *356*(6335), 270–275. doi:10.1126/science.aam9317

Johnson, M. D., Cox, R. D., & Barnes, M. A. (2019). Analyzing airborne environmental DNA: A comparison of extraction methods, primer type, and trap type on the ability to detect airborne eDNA from terrestrial plant communities. *Environmental DNA*, *1*(2), 176–185. doi:https://doi.org/10.1002/edn3.19

Kahlke, T., & Ralph, P. J. (2019). BASTA – Taxonomic classification of sequences and sequence bins using last common ancestor estimations. *Methods in Ecology and Evolution*, *10*(1), 100–103. doi:10.1111/2041-210X.13095

Mena, J. L., Yagui, H., Tejeda, V., Bonifaz, E., Bellemain, E., Valentini, A., … Lyet, A. (2021). Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecological Applications*, e02335. doi:https://doi.org/10.1002/eap.2335

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, *21*(8), 1931–1950. doi:10.1111/j.1365-294X.2011.05403.x

R Development Core Team: R: A language and environment for statistical computing. (2021).

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: the barcode index number (BIN) system. *PloS One*, *8*(7), e66213. doi:10.1371/journal.pone.0066213

Rees, H. C., Baker, C. A., Gardner, D. S., Maddison, B. C., & Gough, K. C. (2017). The detection of great crested newts year round via environmental DNA analysis. *BMC Research Notes*, *10*(1), 327. doi:10.1186/s13104-017-2657-y

RStudio Team. (2020). RStudio: Integrated Development for R. Boston< MA: RSTUDIO.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. doi:https://doi.org/10.1016/j.gecco.2019.e00547

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., … McDevitt, A. D. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*, *57*(4), 707–716. doi:https://doi.org/10.1111/1365-2664.13592

Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C., Sinding, M.-H. S., … Gilbert, M. T. P. (2018). Debugging diversity – a pan-continental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool. *Molecular Ecology Resources*, *18*(6), 1282–1298. doi:https://doi.org/10.1111/1755-0998.12912

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*(1), 88. doi:10.1186/s13104-016-1900-2

Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., … Weisser, W. W. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature*, *574*(7780), 671–674. doi:10.1038/s41586-019-1684-3

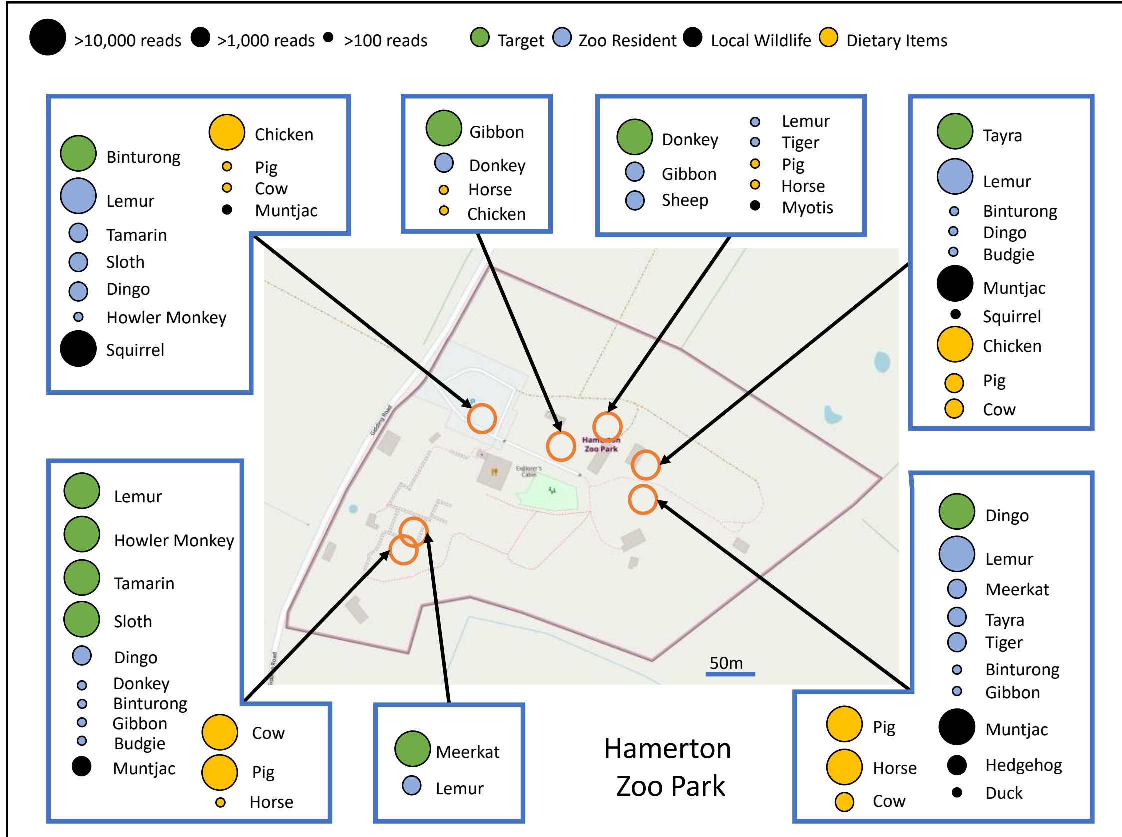
Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., & Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

517 study of eDNA metabarcoding seawater. *Scientific Reports*, *9*(1), 5991. doi:10.1038/s41598-
518 019-42455-9
519 Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., Vogler, A. P., & Anonymous. (2002). DNA
520 points the way ahead in taxonomy. *Nature*, *418*(6897), 479.
521 Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna.
522 *Molecular Biology and Evolution*, *13*(1), 283–285.
523 doi:10.1093/oxfordjournals.molbev.a025566
524 Tilman, D., Clark, M., Williams, D. R., Kimmel, K., Polasky, S., & Packer, C. (2017). Future
525 threats to biodiversity and pathways to their prevention. *Nature*, *546*(7656), 73–81.
526 doi:10.1038/nature22900
527 Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., … Miya, M. (2017).
528 Environmental DNA enables detection of terrestrial mammals from forest pond water.
529 *Molecular Ecology Resources*, *17*(6), e63–e75. doi:10.1111/1755-0998.12690
530 Valentin, R. E., Fonseca, D. M., Gable, S., Kyle, K. E., Hamilton, G. C., Nielsen, A. L., &
531 Lockwood, J. L. (2020). Moving eDNA surveys onto land: Strategies for active eDNA
532 aggregation to detect invasive forest insects. *Molecular Ecology Resources*, *20*(3), 746–
533 755. doi:https://doi.org/10.1111/1755-0998.13151
534 Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M. B., Brand, T. B., …
535 Collins, M. J. (2007). Ancient Biomolecules from Deep Ice Cores Reveal a Forested
536 Southern Greenland. *Science*, *317*(5834), 111–114. doi:10.1126/science.1141758
537 Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., … Taberlet, P.
538 (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486),
539 47–51. doi:10.1038/nature12921
540 Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., … Cooper,
541 A. (2003). Diverse Plant and Animal Genetic Records from Holocene and Pleistocene
542 Sediments. *Science*, *300*(5620), 791–795. doi:10.1126/science.1084114
543

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

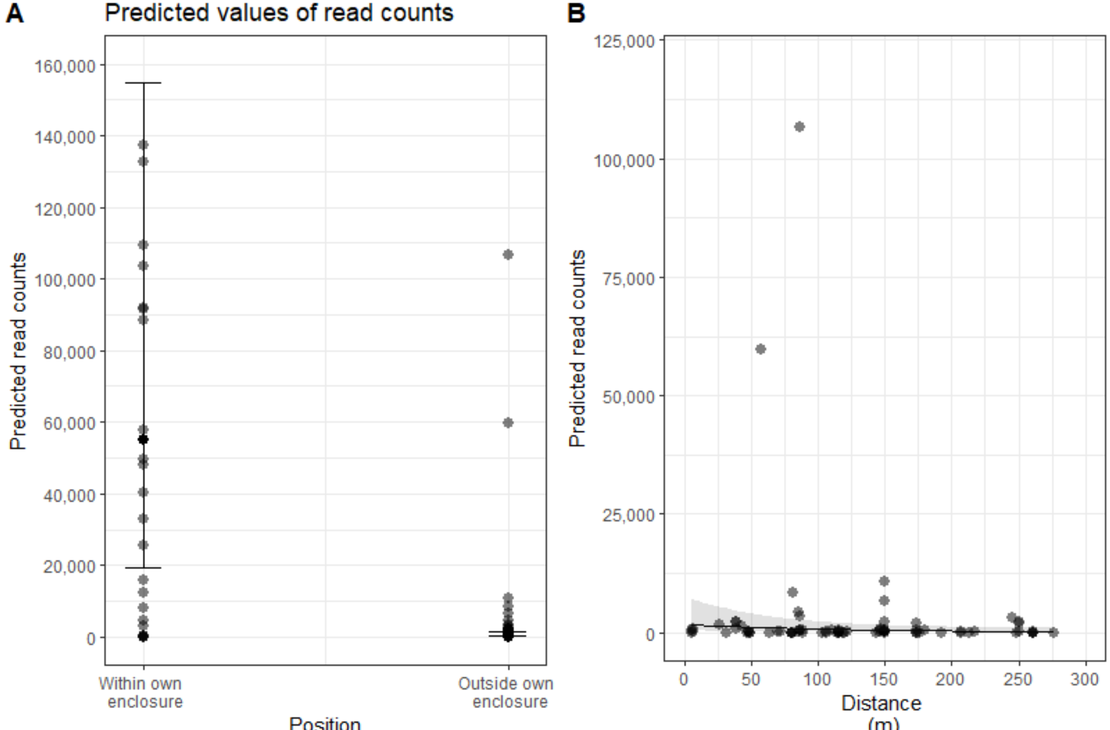
544 **Figures and Tables**
545



546
547 **Figure 1: Species identified at seven zoo locations using only DNA collected from air**
548 **sampling.** Identifications are colour coded to indicate the origin of the DNA and circles are scaled
549 to represent approximate read abundance (low, medium and high copy number). Orange rings
550 indicate sampling location. Identifications with <100 copies were excluded from the figure. Full data
551 with read counts for all locations are provided in Extended Table 1, 2 and 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

552



553

554 **Figure 2: Read count variability with distance from known source.** A) Read counts significantly
555 varied according to the sampling position relative to the animal's own enclosure. Read counts from
556 samples within the animal's own enclosure were higher than from samples outside the animal's
557 own enclosure (this also included the enclosures of other animals). B) Read counts were not
558 significantly affected by distance from the animal, once samples from the animal's own enclosure
559 were excluded. Both plots show predicted read counts from zero-inflated negative binomial models.

Table 1: DNA based identification of target zoo species at each location. Cell values represent total read counts from pooled COI, 16s and nested 16s amplifications. Each location was sampled 4 times (inside and outside using 0.25 and 0.45µm filters) with the exception of the Primate House where three inside and one outside space were sampled (eight samples), the Sloth and Possum House which only had an inside space (two samples) and the Meerkat Colony, Cat Circle, Bear Enclosure, Owl Walkway and Bins which were only outside (two samples). N-values represent total number of pooled sequencing runs (samples x 3 PCRs).

Location	*N*	*Varecia variegate* (Black and White Lemur)	*Arctictis binturong* (Javan Binturong)	*Canis* (Dingo)	*Hylobates lar* (Gibbon)	*Choloepus didactylus* (Sloth)	*Alouatta caraya* (Howler Monkey)	*Eira barbara* (Tayra)	*Saguinus oedipus* (Cotton topped Tamarin)	*Equus asinus* (Donkey)	*Suricata suricatta* (Meerkat)	*Panthera tigris* (Tiger)
Ring-Tailed Lemur Enclosure	12	55,183	279	0	0	0	11,064	2	72	554	0	0
Binturong Enclosure	12	100,174	183,201	2,612	0	2,595	533	0	6,557	1	0	0
Primate House	24	892,893	403	8,322	243	19,010	45,136	20	49,627	353	0	0
Tiger Enclosure	12	29,163	337	147,394	1	0	0	4,528	0	521	0	34
Dingo Enclosure	12	99,050	580	137,858	424	0	0	680	0	0	3,025	1,319
Tayra Enclosure	12	230,386	276	309	1	0	0	118,135	0	0	0	25
Meerkat Colony	6	1,091	0	5	0	0	0	0	0	2	103,438	0
Sloths and Possum House	6	8	0	0	25	0	0	0	0	0	0	0
Lynx Enclosure	12	85	0	5,442	133	0	0	0	0	10	0	220
Maned Wolf Enclosure	12	208	0	59,739	1,186	0	0	0	0	514	0	0
Cheetah Enclosure	12	77	0	0	370	0	0	0	0	6	0	0
Gibbon Enclosure	12	28	0	0	113,889	0	0	0	0	8,309	261	8
Camel Enclosure	12	1,009	0	3	0	0	0	0	0	1	0	0
Wallaby Enclosure	12	0	0	87	0	0	0	0	0	1	0	0
Possum Enclosure	12	2,860	0	22	0	0	0	0	0	0	1,669	0
Donkey Enclosure	12	354	0	0	319	0	0	0	0	242,392	0	618
Cat Circle	6	429	0	341	0	0	0	0	0	0	0	0
Bear Enclosure	6	12	0	0	8	0	0	0	0	0	0	0
Owl Walkway	6	10	0	0	0	0	0	0	0	0	0	0
Bins	6	8	0	0	0	0	0	0	0	0	0	0

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

574 **Table 2: DNA based identification of non-target zoo species at each location.** Cell values
575 represent total read counts from COI, 16s and nested 16s amplifications. Each location was
576 sampled 4 times (inside and outside using 0.25 and 0.45µm filters) with the exception of the Primate
577 House where three inside and one outside space were sampled (eight samples), the Sloth and
578 Possum House which only had an inside space (two samples) and the Meerkat Colony, Cat Circle,
579 Bear Enclosure, Owl Walkway and Bins which were only outside (two samples). N-values represent
580 total number of pooled sequencing runs (samples x 3 PCRs).

		Non-Target Zoo Animals				
Location	*N*	*Grus carunculate* (Wattled Crane)	*Melopsittacus undulatus* (budgerigar)	*Ovis aries* (Sheep)	*Capra* Goat)	*Taenlopygla guttata* (Zebra Finch)
Ring-Tailed Lemur Enclosure	12	0	0	0	0	0
Binturong Enclosure	12	0	0	0	0	0
Primate House	24	3	443	27	0	0
Tiger Enclosure	12	0	0	0	0	0
Dingo Enclosure	12	7	0	0	0	0
Tayra Enclosure	12	1	442	0	0	0
Meerkat Colony	6	165	0	0	0	0
Sloths / Possum House	6	0	0	30	0	0
Lynx Enclosure	12	245	0	0	0	90
Maned Wolf Enclosure	12	0	0	0	0	0
Cheetah Enclosure	12	12,955	0	0	0	0
Gibbon Enclosure	12	0	0	0	0	0
Camel Enclosure	12	0	0	231	0	0
Wallaby Enclosure	12	0	0	2,791	0	0
Possum Enclosure	12	0	0	14	0	63
Donkey Enclosure	12	1	0	1,654	0	0
Cat Circle	6	0	0	0	0	0
Bear Enclosure	6	0	0	231	8	0
Owl Walkway	6	0	0	0	0	0
Bins	6	0	0	0	0	0

581

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.15.452392; this version posted July 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: DNA based identification of non-zoo species at each location. Cell values represent total read counts from COI, 16s and nested 16s amplifications. Each location was sampled 4 times (inside and outside using 0.25 and 0.45µm filters) with the exception of the Primate House where three inside and one outside space were sampled (eight samples), the Sloth and Possum House which only had an inside space (two samples) and the Meerkat Colony, Cat Circle, Bear Enclosure, Owl Walkway and Bins which were only outside (two samples). N-values represent total number of pooled sequencing runs (samples x 3 PCRs).

		Food or Farm				Native				Invasive Or Food
Location	*N*	*Bos taurus* (Cow)	*Equus caballus* (Horse)	*Sus scrofa* (Pig)	*Gallus gallus* (Chicken)	*Erinaceus europaeus* (Hedgehog)	*Anatidae** (Duck)	*Myotis sp.* (Bat)	*Sciurus sp.* (Squirrel)	*Muntiacus reevesi* (Muntjac Deer)
Ring-Tailed Lemur Enclosure	12	26	0	5,714	1	0	0	0	0	0
Binturong Enclosure	12	487	20	322	27,426	39	0	0	108,799	588
Primate House	24	81,030	166	171,441	75	0	14	0	0	1,946
Tiger Enclosure	12	38	240	394	342	77	0	0	0	3,120
Dingo Enclosure	12	8,133	15,780	56,915	85	3,262	318	0	0	51,889
Tayra Enclosure	12	1,864	0	1,252	87,012	0	0	0	718	28,579
Meerkat Colony	6	0	0	0	7	0	0	0	0	0
Sloths / Possum House	6	9,667	0	3,156	0	0	0	0	0	0
Lynx Enclosure	12	15	0	12	27,690	0	0	0	0	0
Maned Wolf Enclosure	12	282	0	0	0	0	0	0	0	0
Cheetah Enclosure	12	1,351	0	0	0	0	0	0	0	0
Gibbon Enclosure	12	15	0	0	160	0	0	0	0	0
Camel Enclosure	12	2,141	0	0	0	0	0	0	0	0
Wallaby Enclosure	12	623	0	0	0	0	0	0	0	0
Possum Enclosure	12	3,551	0	0	0	0	0	0	0	0
Donkey Enclosure	12	43	133	272	0	0	0	553	0	0
Cat Circle	6	2,405	0	0	2	0	0	0	0	0
Bear Enclosure	6	31,905	0	0	8	0	35	0	0	0
Owl Walkway	6	0	0	0	0	0	0	0	0	0
Bins	6	90	0	0	1	0	0	0	0	0

* could represent an identification to either zoo stock or wildfowl

Supplementary Information is available for this paper

Supplementary File 1 contains a detailed table associated with read counts for each PCR reaction.