

Current Biology

Measuring biodiversity from DNA in the air

Highlights

- Environmental DNA can be collected from air samples collected in open environments
- Vertebrate eDNA carried in the air can be used to identify terrestrial animals
- Environmental DNA can be detected in air several hundred meters from the source
- Airborne environmental DNA can be detected from consumed prey following predation

Authors

Elizabeth L. Clare, Chloe K. Economou, Frances J. Bennett, ..., Benjamin McRobie, Rosie Drinkwater, Joanne E. Littlefair

Correspondence

eclare@yorku.ca

In brief

Environmental (e)DNA is shed by all species and can be collected to monitor biodiversity-revolutionizing research, particularly in aquatic ecosystems. Clare et al. collect vertebrate eDNA from the air to identify terrestrial diversity and predation. Airborne eDNA was recovered hundreds of meters from sources, indicating that populations may be monitored at a distance, significantly advancing terrestrial ecology.



Report

Measuring biodiversity from DNA in the air

Elizabeth L. Clare,^{1,2,4,6,*} Chloe K. Economou,¹ Frances J. Bennett,¹ Caitlin E. Dyer,¹ Katherine Adams,³ Benjamin McRobie,³ Rosie Drinkwater,¹ and Joanne E. Littlefair^{1,5}

¹School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

²Department of Biology, York University, Toronto, ON M3J 1P3, Canada

³Hamerton Zoo Park, Huntingdon PE28 5RE, UK

⁴Twitter: @Dr_bat_girl

⁵Twitter: @JELittlefair

⁶Lead contact

*Correspondence: eclare@yorku.ca

<https://doi.org/10.1016/j.cub.2021.11.064>

SUMMARY

The crisis of declining biodiversity¹ exceeds our current ability to monitor changes in ecosystems. Rapid terrestrial biomonitoring approaches are essential to quantify the causes and consequences of global change. Environmental DNA² has revolutionized aquatic ecology,³ permitting population monitoring⁴ and remote diversity assessments matching or outperforming conventional methods of community sampling.^{3–5} Despite this model, similar methods have not been widely adopted in terrestrial ecosystems. Here, we demonstrate that DNA from terrestrial animals can be filtered, amplified, and then sequenced from air samples collected in natural settings representing a powerful tool for terrestrial ecology. We collected air samples at a zoological park, where spatially confined non-native species allowed us to track DNA sources. We show that DNA can be collected from air and used to identify species and their ecological interactions. Air samples contained DNA from 25 species of mammals and birds, including 17 known terrestrial resident zoo species. We also identified food items from air sampled in enclosures and detected taxa native to the local area, including the Eurasian hedgehog, endangered in the United Kingdom. Our data demonstrate that airborne eDNA concentrates around recently inhabited areas but disperses away from sources, suggesting an ecology to airborne eDNA and the potential for sampling at a distance. Our findings demonstrate the profound potential of air as a source of DNA for global terrestrial biomonitoring.

RESULTS AND DISCUSSION

An environmental DNA (eDNA) assay for terrestrial life

DNA is shed from all organisms and deposited as eDNA. Coupled with DNA-based approaches for identifying species,^{6–8} eDNA has been used to analyze contemporary and past ecosystems for nearly 2 decades^{2,9,10} with a particular focus on aquatic systems. An explosion of interest in aquatic eDNA to assess populations and track invasive species has revolutionized aquatic science, management, and conservation⁴ with efficient non-invasive inventory assessment methods.^{3–5} Together with the growth of DNA reference databases, there is strong potential to transform our ability to monitor global ecosystems. However, while aquatic habitats are now commonly surveyed with eDNA and often carry signatures of nearby terrestrial biodiversity,^{11,12} a truly terrestrial targeted eDNA system has not been developed. On land, animal eDNA has been measured from where it settles in permafrost, blood, snow, soil, and honey,¹³ and recently by spraying foliage and collecting the runoff to gather settled surface eDNA¹⁴; but unlike the aquatic environment, no single approach has gained widespread use. Until now, collecting animal eDNA directly from the air, analogous to aquatic vertebrate eDNA sampling in water, has been mostly untested⁴ (see also Lynggaard et al.¹⁵). There has been a longer

history of sampling microbes¹⁶ and eDNA from plants^{17,18} and fungi¹⁹ in airborne dust, and recently from some chordates^{20,21} or animals under highly controlled laboratory conditions.²² Our goal was to collect eDNA from terrestrial vertebrates, laying the groundwork for global terrestrial biomonitoring of animals using air sampling. To test the hypothesis that vertebrate DNA is carried in the air, we collected 72 air samples from 20 locations within the grounds of Hamerton Zoological Park in Huntingdonshire, UK. Of these, 64 yielded DNA that was identified as belonging to non-human terrestrial vertebrates with multiple sources represented in most samples (Figures 1, S1, and S2; Tables 1 and 2). Low DNA extraction volumes and concentrations suggests a future role in pooling replicate samples, as is done in DNA biomonitoring using leeches.²³ This can increase positive hit rates, while reducing false negatives and sequencing costs for large-scale biosurveys. However, the high level of success detecting local species in this study establishes that sampling terrestrial vertebrate life from air can be conducted under natural conditions and provides extensive validation of this technique for global terrestrial biomonitoring.

Multiple DNA markers detect eDNA in the air

We targeted mitochondrial regions using 16S and cytochrome c oxidase I (COI) markers to recover airborne eDNA from terrestrial



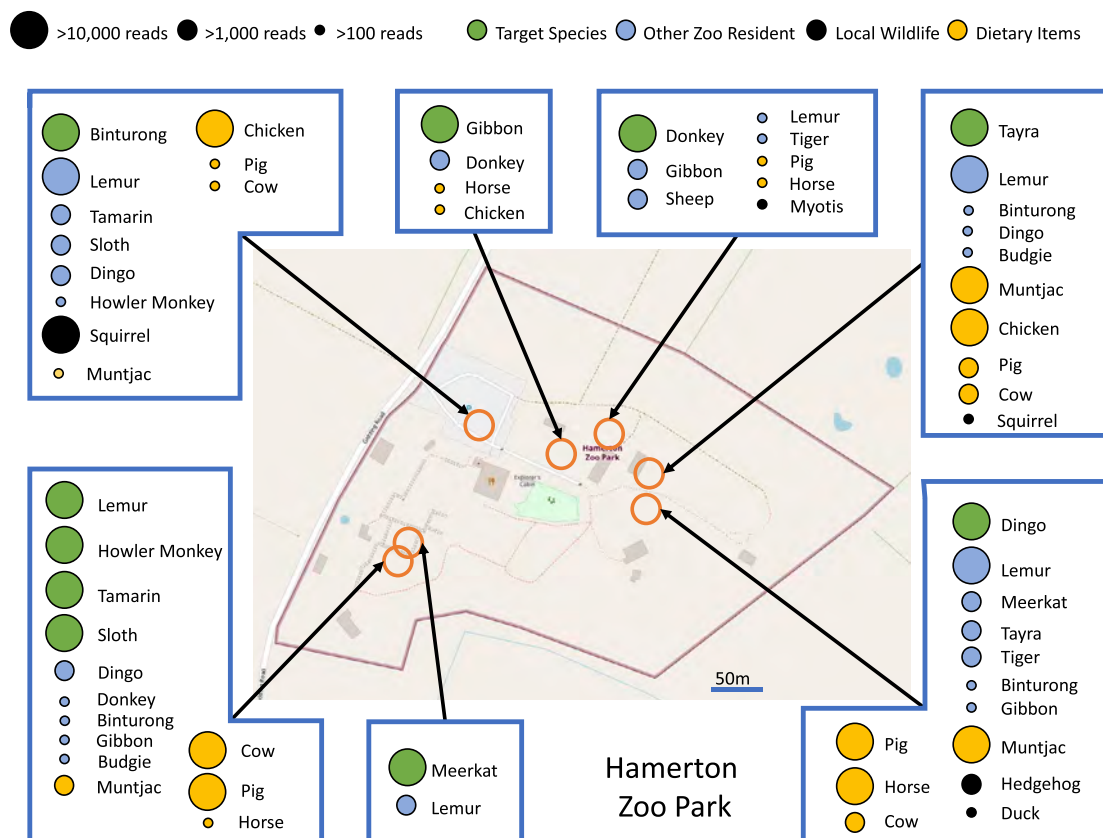


Figure 1. Species identified at 7 zoo locations using DNA collected from air sampling

Identifications are color-coded to indicate the taxonomic assignment of the DNA, and circles are scaled to represent approximate read abundance (low, medium, and high copy number). Orange rings indicate sampling location. Identifications with <100 copies were excluded from the figure. A 50 m distance scale is indicated by a blue line.

Full data with read counts for all locations are provided in [Tables 1 and 2](#), [Data S1](#), and [Figures S1 and S2](#).

mammals and compare this to the known fauna of the zoo park ([Figures 1, S1, and S2](#); [Data S1A–S1C](#)). We recovered 12,207,070 16S reads after the removal of adaptor sequences; these were used as input into the DADA2 bioinformatic pipeline. After length and quality filtering, paired-end merging, and removal of chimeras, 11,707,400 reads remained assigned to 335 amplicon sequence variants (ASVs). Taxonomic ID of the ASVs was assigned using BLAST and further refined with BASTA using a last common ancestor (LCA) algorithm and based on 97% sequence similarity to the 16S reference database ([Figure S2](#); see [STAR Methods](#) for details and parameters of *Eira barbara* identification). In three cases, identifications were further resolved to species based on the known local inventory ([STAR Methods](#)). These ASVs were resolved to 23 taxa (26.4% of the original reads), which included 11 targeted zoo mammals and 12 additional species from the zoo, local wildlife, and food items ([Figures 1, S1, and S2](#); [Data S1A and S1B](#)).

We recovered 6,167,294 COI reads from samples amplified by COI primers ([Data S1A](#)). These data were processed in the mBRAVE pipeline ([STAR Methods](#)). Filtered data included 1,061,857 reads that were compared to reference databases. From these, 361,889 reads (5.9% of the original reads) were assigned to non-human mBRAVE BINs²⁴ at >97% sequence similarity and resolved to species level based on sequence similarity

matches >99% in most cases ([Figure S2](#)), with the exception of *Canis*, in which species cannot be easily differentiated. We report these as dingo (*Canis lupus*), although it is also possible that domestic dog DNA is present on site. In addition to dingo, we recorded four other zoo mammals, one native bird, and four species that include likely dietary items and the wattled crane (*Grus carunculata*), which were present on site as zoo residents. We found a small but significant difference in community composition recovered from 16S, nested 16S, and nested COI amplicons ([STAR Methods](#)), likely driven by the greater richness detected by the 16S amplicon (PERMANOVA, $R^2 = 0.039$, $p = 0.0008$). Filter pore size had no effect on the recovered community composition (PERMANOVA, $R^2 = 0.013$, $p = 0.116$).

The movement of DNA across landscapes

Considerable research in aquatic systems has focused on the “ecology of eDNA,”²⁵ quantifying and understanding factors influencing eDNA detections beyond inventories alone. Zoo parks are ideal for the validation of airborne eDNA movement patterns because they contain captive colonies of non-native species whose identity and spatial location are known with certainty. We predicted that target species would be detected in confined spaces (i.e., inside enclosures that had solid walls limiting air movement) based on our laboratory proof of

Table 1. DNA-based identification of target zoo species at each location

Location	n	<i>Varecia variegata</i> (black and white lemur)	<i>Arctictis binturong</i> (Javan binturong)	<i>Canis</i> (dingo)	<i>Hylobates lar</i> (gibbon)	<i>Choloepus didactylus</i> (sloth)	<i>Alouatta caraya</i> (howler monkey)	<i>Eira barbara</i> (tayra)	<i>Saguinus oedipus</i> (cotton topped tamarin)	<i>Equus asinus</i> (donkey)	<i>Suricata suricata</i> (meerkat)	<i>Panthera tigris</i> (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	113,789	679	4,498	243	19,010	50,924	32	49,625	353	0	0
Tiger enclosure	12	5,702	337	4742	0	0	0	4,528	0	0	0	34
Dingo enclosure	12	23,528	580	284,203	1	0	0	680	0	521	3,025	1,319
Tayra enclosure	12	99,029		438	425	0	0	118,135	0	0	0	25
Meerkat colony	6	1,091	0	5	0	0	0	0	0	2	103,438	0
Sloth/ 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	242,780	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
Rubbish bins	6	8	0	0	0	0	0	0	0	0	0	0

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 0.45- μ m filters), except for the primate house, where 3 inside and 1 outside space were sampled (8 samples); the sloth/possum house, which only had an inside space sampled (two samples); and the meerkat colony, cat circle, bear enclosure, owl walkway, and rubbish bins, which were only sampled outside (2 samples). n values represent total number of pooled sequencing runs (samples $\times$ 3 PCR). See also [Data S1](#) and [Figures S1](#) and [S2](#).

Table 2. DNA based identification of non-target species at each location

Location	n	Non-target zoo animals					Food or farm				Native				Invasive or food
		<i>Grus carunculata</i> (wattled crane)	<i>Melopsittacus undulatus</i> (budgerigar)	<i>Ovis aries</i> (sheep)	<i>Capra guttata</i> (goat)	<i>Taeniopteryx</i> (zebra finch)	<i>Bos taurus</i> (cow)	<i>Equus caballus</i> (horse)	<i>Sus scrofa</i> (pig)	<i>Gallus gallus</i> (chicken)	<i>Erinaceus europaeus</i> (hedgehog)	Anatidae ^a (duck)	<i>Myotis</i> sp. (Bat)	<i>Sciurus</i> sp. (squirrel)	<i>Muntiacus reevesi</i> (muntjac deer)
Ring-tailed lemur enclosure	12 0	0	0	0	0	0	26	0	5,714	1	0	0	0	0	0
Binturong enclosure	12 0	0	0	0	0	0	487	20	322	27,471	39	0	0	108,799	588
Primate house	24 3	443	27	0	0	0	81,043	166	143,146	525	0	14	567	0	1,934
Tiger enclosure	12 0	0	0	0	0	0	38	240	394	342	77	0	0	0	3,120
Dingo enclosure	12 7	0	0	0	0	0	8,133	15,780	56,915	85	3,262	318	0	0	51,889
Tayra enclosure	12 1	442	0	0	0	0	1,864	0	1,252	87,466	0	0	0	718	28,579
Meerkat colony	6 165	0	0	0	0	0	0	0	0	7	0	0	0	0	0
Sloth/possum house	6 0	0	30	0	0	0	9,667	0	3,156	0	0	0	0	0	0
Lynx enclosure	12 245	0	0	0	90	15	0	12	27,690	0	0	0	0	0	0
Maned wolf enclosure	12 0	0	0	0	0	282	0	0	0	0	0	0	0	0	0
Cheetah enclosure	12 12,955	0	0	0	0	1,351	0	0	0	0	0	0	0	0	0
Gibbon enclosure	12 0	0	0	0	0	15	0	0	160	0	0	0	0	0	0
Camel enclosure	12 0	0	231	0	0	2,141	0	0	0	0	0	0	0	0	0
Wallaby enclosure	12 0	0	2,791	0	0	1211	0	0	0	0	0	0	0	0	0
Possum enclosure	12 0	0	14	0	63	3,551	0	0	0	0	0	0	0	0	0
Donkey enclosure	12 1	0	1,654	0	0	43	133	272	0	0	0	553	0	0	0
Cat circle	6 0	0	0	0	0	2,405	0	0	2	0	0	0	0	0	0
Bear enclosure	6 0	0	231	8	0	31,999	0	0	8	0	35	0	0	0	0
Owl walkway	6 0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rubbish bins	6 0	0	0	0	0	90	0	0	1	0	0	0	0	0	0

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 0.45- μ m filters) except for the primate house, where 3 inside and 1 outside space were sampled (8 samples); the sloth/possum house, which only had an inside space sampled (2 samples); and the meerkat colony, cat circle, bear enclosure, owl walkway, and rubbish bins, which were only sampled outside (2 samples). n values represent total number of pooled sequencing runs (samples $\times$ 3 PCRs). See also [Data S1](#) and [Figures S1](#) and [S2](#).

^aCould represent an identification to either zoo stock or wildfowl

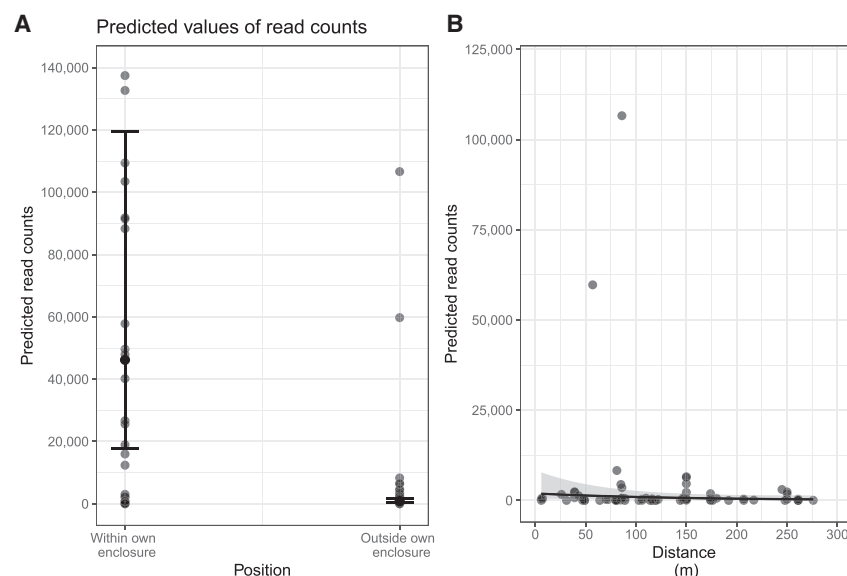


Figure 2. Read count variability with distance from known source

(A) Read counts significantly varied according to the sampling position relative to the animal's own enclosure. Read counts from samples within the animal's own enclosure were higher than from samples outside the animal's own enclosure (this also included the enclosures of other animals). Error bars represent 95% confidence intervals from model predictions.

(B) Read counts were not significantly affected by distance from the animal, once samples from the animal's own enclosure were excluded. Both plots show predicted read counts from negative binomial models.

DNA read counts are generally highest within the animal's own enclosure, whether open to air flow or fully enclosed, we picked up trace read counts in air samples hundreds of meters from the most likely source. For example, meerkat DNA from an outdoor colony was identified in air sampled at the dingo enclosure 245 m

concept,²² with cave and roost surveys as a natural field application. However, we also detected airborne eDNA in enclosures with open sides permitting air flow and in outside enclosures where the barrier was only a fence and thus not a limit on air movement. In fact, some "within own enclosure" read counts in Figure 2A are in open air animal housing without a physical barrier between the animal and local wind conditions. We also detected DNA away from target sources and across open areas subjected to air currents and local weather. In addition to the targeted taxa, we identified three species of mammals and three species of birds known to be housed at the zoo but where we did not have access to their enclosures (Table 2). These additional detections were frequently recovered at the closest sampling point to their actual residence. For example, the indoor exhibit housing budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taenopygia guttata*) was closed during the sampling period, but we detected their DNA in air samples collected at the adjacent primate house and possum enclosure. We found a significant effect of sampling location on recovered community composition, with samples taken at the same location often being more similar than samples taken at different locations (PERMANOVA, $R^2 = 0.475$, $p = 0.0002$). We also compared read counts with distance to the most likely source using two models (Figure 2; STAR Methods). There was a significant effect of the sample position relative to the animal's own enclosure on the read counts (Figure 2A, likelihood ratio statistic = 58.4, $df = 1$, $p < 0.001$), with read counts inside the animal's own enclosure being higher (model estimate: 46,174 reads, confidence limits: 17,621–120,998) than read counts outside the animal's own enclosure (model estimate: 842 reads, confidence limits: 386–1,833). When datapoints from within the animal's own enclosures were removed (i.e., zero distances), there was no relationship between read count numbers and distance from the enclosure (Figure 2B; likelihood ratio statistic = 2.51, $df = 1$, $p = 0.113$). Neither model was overdispersed.

Sampling took place in the presence of animals, but they were not permitted to interact with the sampling equipment. While

away and at the gibbon enclosure 122 m away. A similar observation was made at the Copenhagen Zoo where DNA was observed to concentrate around the source species but occasionally move the full length of the zoo grounds.¹⁵ While contamination between samples is theoretically possible, samples were collected and processed on different days, and we used multiple negative DNA extractions and PCR controls along with empty wells in the sequencing run. Negative well contamination following filtering was low (Data S1A–S1C). Some expected taxa based on sampling location produced read counts lower than these negative thresholds (e.g., tiger [*Panthera tigris*]); thus, we retain all of the data to indicate these very likely positives but treat low copy-number identifications with caution. Contamination in air sampling is likely to represent a significant issue going forward. Human DNA, in particular, will contaminate every lab, and specialized environmental DNA facilities may be required. Alternatively, bioinformatic removal of expected contamination (e.g., human, fish, and naked mole rat, in our case) can be conducted before using negative controls to filter the remaining samples. Positive controls indicated high PCR efficiency and very minimal evidence of contamination in negative extractions. PCR controls, or empty wells used as sequencing controls, following initial filtering suggest that detections represent real events. We used both standard and nested PCRs for the more successful 16S data to see whether some taxa would only be recovered using a more intensive nested approach. This effect was not realized with species such as tiger and sloth (*Saguinus oedipus*) appearing preferentially in the unnested PCR. While nesting can help recover very low-level trace material, it can also cause such species to be lost in mixes in which the dominant DNA signal takes over. Similarly, contamination is exponentially increased by this method (Data S1C). We suggest nesting only be used in the future as a supplement to standard PCR recovery.

Our ability to confirm exact sources and thus dispersal of eDNA highlight three key findings: (1) high dispersal of eDNA in air makes increased sampling key, (2) long-range dispersal at

hundreds of meters is possible, and (3) the lack of a relationship between read count and distance from target suggests that read counts should not be used to estimate the location of a DNA source. The next step will be to assess how the movement of air may affect detections.

Detection of predation and species of special ecological concern

More than one-third of the recovered sequences matched cow (*Bos taurus*), horse (*Equus caballus*), pig (*Sus scrofa*), or chicken (*Gallus gallus*). While we cannot preclude DNA drifting in from the surrounding countryside, it is likely that these represent food provided to the carnivores. Particularly high concentrations of chicken DNA were detected in the binturong (*Arctictis binturong*) and tayra (*E. barbara*) enclosures, while horse, cow, and pig were concentrated in samples from the dingo enclosure, correctly matched with dietary provisions by zoo staff (Table 2). Similarly at the Copenhagen Zoo, fish were detected in proximity to where they were provided as prey.¹⁵ We observed some unexpected concentrations of these DNA sources, perhaps reflecting the movement of people and materials throughout the zoo. For example, an unexpected concentration of pig and cow DNA inside the lemur enclosure could reflect the movement of people or equipment between animal houses.

While the aim of our study was an inventory of the zoo species, adjacent rural settings are a source for DNA from wildlife. We identified DNA associated with squirrels (*Sciurus*) and ducks (Anatidae) in several air samples (Table 2). Ducks are also kept as zoo stock, but we could not identify the genus or species with accuracy, so conservatively we classify this as wildfowl common to the area. We may have detected *Myotis* bats, although we also treat this with caution as many bat DNA samples are handled in the processing laboratory facility. Of special interest was the detection of the European hedgehog (*Erinaceus europaeus*) in three samples (Table 2). Hedgehogs are commonly observed on site by staff, although they are not as active in the winter; thus, their detection is particularly interesting. As of 2020, the hedgehog was listed as vulnerable to extinction in the United Kingdom (<https://www.mammal.org.uk/science-research/red-list/>), making it vital to develop additional methods to monitor and protect existing populations. UK species of special interest such as the great crested newt have been the model for the development of aquatic eDNA detection methods²⁶ and provide a framework for validating airborne eDNA for similar monitoring. One commonly cited application of eDNA approaches is the detection of invasive species. We detected muntjac deer (*Muntiacus reevesi*) in five samples. These muntjacs are native to China but became locally invasive after multiple releases in England in the 19th century.²⁷ They are now well established in eastern England, the location of the zoological park, and are frequently seen on site. They are also provided in food for several species; thus, the detection of muntjacs may reflect either food or wildlife.

Airborne eDNA for biodiversity monitoring

Rapid and accurate terrestrial biomonitoring techniques are essential to our attempts to quantify the causes and consequences of global change^{28,29} and to assist with focused, on-the-ground conservation efforts. Anthropogenic effects have

caused pervasive biodiversity declines across ecosystems,^{1,30,31} particularly from land-use change, habitat loss, and degradation,³² leading to the reorganization of global biodiversity patterns and processes.^{28,33} The inability to detect species and measure population dynamics rapidly and accurately is a fundamental challenge in quantifying our position relative to biodiversity and conservation targets.^{29,30} Detecting changes in diversity, abundance, and community composition as well as species range shifts are priorities highlighted by researchers, conservationists, and major international initiatives²⁹ such as the United Nations Sustainable Development Goals. New approaches that provide simple, large-scale, and automated monitoring techniques are an urgent requirement that are needed to address the often-intractable challenge of biodiversity monitoring.³³

Our study provides compelling evidence that air can be used as a source of DNA for biomonitoring. The detection of multiple taxa in air samples known to reside at the zoo without high false-positive detections strongly validates the local source of the DNA. The detection of species of conservation concern, as well as DNA from dietary items, possibly via the detection of aerosolized fecal material, is compelling and demonstrates the versatility of this genetic approach. Aquatic eDNA sampling was initially treated with caution due to the challenges of swift-moving currents and dilution in large bodies of water; however, both challenges have been largely overcome, and the subsequent rapid global uptake of aquatic eDNA as a biomonitoring tool highlights the versatility of eDNA sampling techniques. If airborne eDNA sampling can be successfully developed, it will have major implications for global terrestrial biomonitoring. The novel opportunities that eDNA approaches provide for tracking faunal composition, non-invasive monitoring of species of special ecological concern, and the detection of species invasion are extremely exciting and suggest that airborne eDNA could revolutionize the ways in which scientists study and monitor terrestrial biodiversity non-invasively on a global scale.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Air sampling
- METHOD DETAILS
 - Sample collection
 - DNA extraction
 - PCR amplification and sequencing
 - PCR visualization and sequencing
 - Bioinformatics methods for COI regions
 - Bioinformatics methods for 16S regions
 - Data filtering
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Comparison of primers and filter pore size
 - Analysis of distance from target enclosure

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2021.11.064>.

ACKNOWLEDGMENTS

This study was funded through an EPSRC impact accelerator account provided to Queen Mary University of London (grant no. EP/R511596/1). We appreciate the support of the Hamerton Zoo Park Staff and the Barts and the London Genome Centre for assistance during this project.

AUTHOR CONTRIBUTIONS

E.L.C. designed the experiment and performed field work; C.K.E. performed all of the laboratory work; F.J.B. performed the field work; C.E.D., R.D., and J.E.L. performed the analysis; and K.A. and B.M. facilitated the field work. All of the authors contributed to manuscript preparation.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 7, 2021

Revised: October 5, 2021

Accepted: November 26, 2021

Published: January 6, 2022

REFERENCES

- Diaz, S., Settele, J., Brondizio, E.S., Ngo, H.T., Agard, J., Armeth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., et al. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* 366, eaax3100.
- Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., and Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791–795.
- Bessey, C., Neil Jarman, S., Simpson, T., Miller, H., Stewart, T., Kenneth Keesing, J., and Berry, O. (2021). Passive eDNA collection enhances aquatic biodiversity analysis. *Commun. Biol.* 4, 236.
- Ruppert, K.M., Kline, R.J., and 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. *Glob. Ecol. Conserv.* 17, e00547.
- Mena, J.L., Yagui, H., Tejeda, V., Bonifaz, E., Bellemain, E., Valentini, A., Tobler, M.W., Sánchez-Vendizú, P., and Lyet, A. (2021). Environmental DNA metabarcoding as a useful tool for evaluating terrestrial mammal diversity in tropical forests. *Ecol. Appl.* 31, e02335.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., and Vogler, A.P. (2002). DNA points the way ahead in taxonomy. *Nature* 418, 479.
- Blaxter, M. (2003). Molecular systematics: counting angels with DNA. *Nature* 421, 122–123.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321.
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506, 47–51.
- Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M.B., Brand, T.B., Hofreiter, M., Bunce, M., Poinar, H.N., Dahl-Jensen, D., et al. (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317, 111–114.
- Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., et al. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol. Ecol. Resour.* 17, e63–e75.
- Sales, N.G., McKenzie, M.B., Drake, J., Harper, L.R., Browett, S.S., Coscia, I., Wangenstein, O.S., Baillie, C., Bryce, E., Dawson, D.A., et al. (2020). Fishing for mammals: landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *J. Appl. Ecol.* 57, 707–716.
- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., and de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29, 358–367.
- Valentin, R.E., Fonseca, D.M., Gable, S., Kyle, K.E., Hamilton, G.C., Nielsen, A.L., and Lockwood, J.L. (2020). Moving eDNA surveys onto land: strategies for active eDNA aggregation to detect invasive forest insects. *Mol. Ecol. Resour.* 20, 746–755.
- Lynggaard, C., Bertelsen, M.F., Jensen, C.V., Johnson, M.S., Frøslev, T.G., Olsen, M.T., et al. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. *Curr. Biol.* Published online January 6, 2022. <https://doi.org/10.1016/j.cub.2021.12.014>.
- Leppänen, H.K., Täubel, M., Jayaprakash, B., Vepsäläinen, A., Pasanen, P., and Hyvärinen, A. (2018). Quantitative assessment of microbes from samples of indoor air and dust. *J. Expo. Sci. Environ. Epidemiol.* 28, 231–241.
- Folloni, S., Kagkli, D.M., Rajcevic, B., Guimarães, N.C., Van Droogenbroeck, B., Valicente, F.H., Van den Eede, G., and Van den Bulcke, M. (2012). Detection of airborne genetically modified maize pollen by real-time PCR. *Mol. Ecol. Resour.* 12, 810–821.
- Johnson, M.D., Cox, R.D., and 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. *Environ. DNA* 1, 176–185.
- Banchi, E., Ametrano, C.G., Tordoni, E., Stanković, D., Ongaro, S., Tretiach, M., Pallavicini, A., and Muggia, L.; ARPA Working Group (2020). Environmental DNA assessment of airborne plant and fungal seasonal diversity. *Sci. Total Environ.* 738, 140249.
- Aalismail, N.A., Díaz-Rúa, R., Gerdali, N., Cusack, M., and Duarte, C.M. (2021). Diversity and sources of airborne eukaryotic communities (AEC) in the Global Dust Belt over the Red Sea. *Earth Syst. Environ.* 5, 459–471.
- Serrao, N.R., Weckworth, J.K., McKelvey, K.S., Dysthe, J.C., and Schwartz, M.K. (2021). Molecular genetic analysis of air, water, and soil to detect big brown bats in North America. *Biol. Conserv.* 261, 109252.
- Clare, E.L., Economou, C.K., Faulkes, C.G., Gilbert, J.D., Bennett, F., Drinkwater, R., and Littlefair, J.E. (2021). eDNAir: proof of concept that animal DNA can be collected from air sampling. *PeerJ* 9, e11030.
- Drinkwater, R., Schnell, I.B., Bohmann, K., Bernard, H., Veron, G., Clare, E., Gilbert, M.T.P., and Rossiter, S.J. (2019). Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Mol. Ecol. Resour.* 19, 105–117.
- Ratnasingham, S., and Hebert, P.D.N. (2013). A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE* 8, e66213.
- Barnes, M.A., and Turner, C.R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* 17, 1–17.
- Rees, H.C., Baker, C.A., Gardner, D.S., Maddison, B.C., and Gough, K.C. (2017). The detection of great crested newts year round via environmental DNA analysis. *BMC Res. Notes* 10, 327.
- Hemami, M.R., Watkinson, A.R., and 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. *For. Ecol. Manage.* 215, 224–238.
- Eriksson, B.K., and Hillebrand, H. (2019). Rapid reorganization of global biodiversity. *Science* 366, 308–309.
- Amano, T., Székely, T., Sandel, B., Nagy, S., Mundkur, T., Langendoen, T., Blanco, D., Soykan, C.U., and Sutherland, W.J. (2018). Successful

- conservation of global waterbird populations depends on effective governance. *Nature* 553, 199–202.
30. Johnson, C.N., Balmford, A., Brook, B.W., Buettel, J.C., Galetti, M., Guangchun, L., and Wilmschurst, J.M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science* 356, 270–275.
 31. Seibold, S., Gossner, M.M., Simons, N.K., Blüthgen, N., Müller, J., Ambarli, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J.C., et al. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature* 574, 671–674.
 32. Tilman, D., Clark, M., Williams, D.R., Kimmel, K., Polasky, S., and Packer, C. (2017). Future threats to biodiversity and pathways to their prevention. *Nature* 546, 73–81.
 33. Barlow, J., França, F., Gardner, T.A., Hicks, C.C., Lennox, G.D., Berenguer, E., Castello, L., Economo, E.P., Ferreira, J., Guénard, B., et al. (2018). The future of hyperdiverse tropical ecosystems. *Nature* 559, 517–526.
 34. Ivanova, N.V., Clare, E.L., and Borisenko, A.V. (2012). DNA barcoding in mammals. In *DNA Barcodes: Methods and Protocols* (Methods in Molecular Biology, W.J. Kress, and D.L. Erickson, eds. (Smithsonian National Museum of Natural History), pp. 153–182.
 35. Taylor, P.G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Mol. Biol. Evol.* 13, 283–285.
 36. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88.
 37. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.
 38. RStudio Team (2020). RStudio: integrated development for R. <http://www.rstudio.com/>.
 39. R Development Core Team (2021). R: a language and environment for statistical computing (R Foundation for Statistical Computing).
 40. Kahlke, T., and Ralph, P.J. (2019). BASTA – taxonomic classification of sequences and sequence bins using last common ancestor estimations. *Methods Ecol. Evol.* 10, 100–103.
 41. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R.R.B., Simpson, G.L., Solymos, M.P., et al. (2019). *vegan: community ecology package*. R package version 2.5-6. <https://cran.r-project.org/web/packages/vegan/index.html>.
 42. Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., and Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* 9, 378–400.
 43. Hartig, F. (2021). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.4. <https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html>.
 44. Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S.J., Rodriguez, M.J., Dorea, C.C., et al. (2017). Open the Sterivex™ casing: an easy and effective way to improve DNA extraction yields. *Limnol. Oceanogr. Methods* 15, 1015–1020.
 45. Hallam, J., Clare, E.L., Jones, J.I., and Day, J.J. (2021). Biodiversity assessment across a dynamic riverine system: a comparison of eDNA metabarcoding versus traditional fish surveying methods. *Environ. DNA* 3, 1247–1266.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
72 filtered air samples	Hamerton Zoological Park, Cambridgeshire, UK	N/A
Critical commercial assays		
DNeasy Blood and Tissue Kit	QIAGEN	Cat#69504
QIAshredder	QIAGEN	Cat#79656
QIAGEN Multiplex mix	QIAGEN	Cat#206143
Monarch DNA Gel Extraction Kit	NEB	Cat#T1020S
QuBit dsDNA HS Assay Kit	Invitrogen	Cat#Q32854
DNA D1000 TapeStation kit	Agilent	5067-5584
MiSeq v2 Chemistry 500 cycle	Illumina	MS-102-2003
Deposited data		
Raw amplicon sequence data for COI and 16S PCR's	This paper	NCBI SRA: PRNJA743788
Oligonucleotides		
COI primer -AquaF2: ATCACRACCATCAT YAAATRAARCC	34	Eurofins Genomics
COI primer - Vr1d: TAGACTTCTGGGTG GCCRAARAAYCA	34	Eurofins Genomics
Tagged COI primer - TSP1_AquaF2: TCTACACTCGTCGGCAGCGTCAGA TGTGTATAAGAGACAGATCACRAC CATCATYAAATRAARCC	34	Eurofins Genomics
Tagged COI primer - TSP2_ Vr1d: GTCTCGTGGGCTCGGAGATGTGTA TAAGAGACAGTAGACTTCTGGGTG GCCRAARAAYCA	34	Eurofins Genomics
16S primer - 16Smam_forward: CGGTTGGGGTGACCTCGGA	35	Eurofins Genomics
16S primer - 16Smam_reverse: GCTGTTATCCCTAGGGTAACT	35	Eurofins Genomics
Tagged 16S primer - TSP1_16Smam_forward: TCTACACTCGTCGGCAGCGTCAGATGTGTA TAAGAGACAGCGGTTGGGGTGACCTCGGA	35	Eurofins Genomics
Tagged 16S primer - TSP2_16Smam_reverse: GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGCTGTTATCCCTAGGGTAACT	35	Eurofins Genomics
Software and algorithms		
mBRAVE	N/A	http://www.mbrave.net/
AdapterRemoval Version2	36	https://adapterremoval.readthedocs.io/en/stable/
DADA2 Version 1.16	37	https://benjjneb.github.io/dada2/tutorial.html
R Studio Version 1.1.463	38	https://www.rstudio.com/
R Version 4.0.2	39	https://www.r-project.org/
BASTA Version 1.3.2.3	40	https://github.com/timkahlke/BASTA/wiki

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
vegan Version 2.5-6	41	https://cran.r-project.org/web/packages/vegan/index.html
glmmTMB Version 1.0.2.1	42	https://github.com/glmmTMB/glmmTMB
DHARMA Version 0.3.3.0.	43	http://florianhartig.github.io/DHARMA/
Other		
Sterivex-HV Pressure Filter Unit, 0.45 μ m pore size	Merck Millipore	SVHV010RS
Sterivex-HV Pressure Filter Unit, 0.22 μ m pore size	Merck Millipore	SVGP01050

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Elizabeth Clare (eclare@yorku.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw DNA amplicon data have been deposited at NCBI Sequence Read Archive and are publicly available as of the date of publications. The accession number is listed in the [Key resources table](#).

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Air sampling

This study used filtered air samples which were collected at Hamerton Zoo Park, a 25-acre conservation zoo in Huntingdonshire UK established in 1990 and containing approximately 100 species of animal, mostly mammals and birds of conservation concern and surrounded by a matrix of agricultural land in rural England. Most species live in enclosures which have free access to outside ranges allowing free air exchange. Air samples were taken from 20 locations each targeting a different area of the zoo. Samples were taken both inside and outside enclosures (where possible) using two filter types for a total of 72 samples. Inside enclosures were often open on one side permitting airflow. Some species were limited to either only outside or only inside spaces.

METHOD DETAILS

Sample collection

Air samples were collected using a peristaltic pump (Geotech) which moves air through a flexible tube. We inserted Sterivex-HV filters (Merck Millipore) with 0.22 μ m and 0.45 μ m filter sizes into the intake end of the tube so that air was drawn from the outside, through the filter and then into the hose. The filter was then placed inside or outside an animal enclosure so that air would be sampled continuously. We targeted 15 enclosures which contained zoo species represented in molecular reference collections. For each of these locations, we sampled air for 30 min at 300ml/min filter rate using each filter and we sampled inside an enclosure (e.g., the sleeping chamber) and outside in the open air enclosure (where species move about freely) within 5 m of the enclosure opening. We also sampled from general areas of the zoo including the Cat Circle, Owl walkway and near rubbish bins. In addition, we sampled at the Tasmanian Golden Possum Enclosure and Syrian Brown Bear Enclosure, but Golden Possums were not represented in reference databases and the bears were in their hibernation cycle and closed to close sampling (i.e., no indoor samples were taken). We treat these two areas as general areas for sampling. All filters were placed in sterile bags following sample collection and frozen until DNA extraction.

DNA extraction

DNA extraction and PCR were carried out within a biological safety cabinet under maximum flow. All extraction procedures followed^{22,44,45} as follows. All equipment was sterilized using UV, 10% bleach, 70% ethanol and ultrapure water between each sample. Following existing protocols,^{22,44,45} the filter was cracked open and the filter removed. DNA was extracted using a Blood and Tissue kit (QIAGEN UK) following manufacturer's protocol but with ATL buffer volumes increased to 450 µl to ensure the filters were submerged. We used 50 µl proteinase K, 500 µl buffer AL and 500 µl of 100% ethanol. We used multiple negative controls at the extraction, PCR and sequencing stages. Samples were lysed overnight using a platform shaking at 650 rpm at 56°C. The samples were then vortexed and transferred to fresh tubes for extraction. We used QIA shredder spin columns (QIAGEN UK) on the remaining filter paper and the flow-through was added to the rest of the sample at which point buffer AL was added. Extraction then followed manufacturer instructions but with centrifugation completed at 11,000 rpm for 3 min following AW2. DNA was eluted in 30 µl of elution buffer pre-heated to 70°C. Elution buffer was cycled through the column three times with 5 min incubation times in each cycle to increase DNA concentration.

PCR amplification and sequencing

Each DNA extract was subjected to three PCRs with cow DNA as a positive control as follows:

16S PCR

We amplified a small region of mammal mitochondria 16S gene using the mam1 and mam2 primers³⁵ modified with adaptors for the Illumina MiSeq sequencing platform. The PCR mix included 7.5 µl of QIAGEN multiplex mix, 1.5 µl ddH₂O, 5 µl of template DNA and 0.5 µl of each primer (10 µM stocks of each) and amplification used cycling conditions of 95°C for 15 min, 40 cycles of 94°C for 30 s, 55°C for 90 s, 72°C for 90 s, and a final 72°C for 10 min and a 10°C hold.

16S nested PCR

To increase amplification success for low yield samples we performed nested PCRs. For the nested 16S PCRs we first used non-tagged mam1 and mam2 primers. For these reactions, we used 3 µl of template DNA (adjusting the amount of water accordingly) and increased the annealing temperature to 59°C. We then used 1 µl of each PCR product from the first reaction as a template for a second PCR, again using the same 16S mam1 and mam2 Illumina MiSeq tagged primers. PCR conditions for the second PCR were as previously mentioned.

COI nested PCR

We amplified a small portion of the 5' end of the cytochrome oxidase gene using AquaF2 forward and VR1d reverse primers.³⁴ We employed a two-step nested PCR strategy. For the first stage PCR, the PCR mix we used comprised 7.5 µl QIAGEN multiplex mix, 3.5 µl ddH₂O, 3 µl of template DNA and 0.5 µl of each primer (10 µM stocks of each). For the majority of samples, we used 1 µl of PCR product from this first reaction as a template for a second PCR using AquaF2 and VR1d Illumina MiSeq tagged primers. For selected samples with significant non-target bands, we gel extracted the target band from the first PCR (Monarch DNA Gel Extraction Kit) and used 1 µl of this purified DNA in the second PCR. Reaction conditions for both first and second PCRs were as follows: 95°C for 15 min, followed by 40 cycles of 94°C for 30 s, 51°C for 90 s, and 72°C for 90 s and a final extension at 72°C for 10 min, then a hold at 10°C.

PCR visualization and sequencing

All products, including positive (cow DNA) and negative controls (no template) were visualized using a 1% agarose gel as an initial screening tool and then quantified using Qubit and TapeStation. Amplicons were sequenced on an Illumina MiSeq using unique 5' forward tags at the Barts and the London Genome Centre following standard protocol using bidirectional 250bp chemistry. The results were demultiplexed by tag for bioinformatics processing.

Bioinformatics methods for COI regions

COI read files were uploaded to the mBRAVE platform (<http://www.mbrave.net>). Paired end samples were assembled with a minimum overlap of 20bp and max substitution of 5bp. Samples were processed to maximize data retention for later steps with the following parameters, Trim Front = 38bp, Trim End = 26bp, Trim Length = 500bp, Min QV filter = 0, Min Length = 100bp, Max bases with low (< 20) QV = 75%, Max bases with ultra low QV (< 10) = 75%. ID threshold = 10%, Exclude from OTU at 10% MIN OTU size = 1 and OTU threshold = 2%.

The reads were compared to the "Hamerton Zoo 1" bespoke reference database consisting of 610 sequences representing 20 species known to reside at the zoo and targeted in our sampling. These data were taken from existing public data in the BOLD database. Sequences not identified by comparison to this bespoke reference collection were then screened in sequential order to system reference libraries:

- SYS-CRLCHORDATA (Chordata references) consisting of 40,565 species
- SYS-CRLAVES (Aves reference) consisting of 5832 species
- SYS-CRLBACTERIA (Bacteria reference) consisting of 2066 species
- SYS-CRLFUNGI (Fungi reference) consisting of 565 species
- SYS-CRLINSECTA (Insect reference) consisting of 217,994 species
- SYS-CRNONINSECTARTH (Non-Insect Arthropoda reference) consisting of 27,832 species

SYS-NONARTHINVERT (Non-Arthropoda Invertebrate reference) consisting of 34,927 species
SYS-CRLPROTISTA (Protista COI reference collection) consisting of 5250 species

Bioinformatics methods for 16S regions

We used AdapterRemoval V2³⁶ to first identify and then remove adaptor contamination, using the additional parameters—trimns and trimqualities, to remove Ns and runs of low quality bases. Read pairs were not collapsed at this step. We processed the remaining reads into amplicon sequence variants (ASVs) using the DADA2 pipeline³⁷ in R.^{38,39} We filtered the reads using DADA2 with the following parameters: truncate length after 100 bases in both directions (truncLen = c(100,100)), reads with any Ns were removed (maxN = 0), reads higher than expected error removed (maxEE = c(2,2)), truncate reads based on low quality scores (truncQ = 2) and discard phiX genes (rm.phix = TRUE). Each of the filtered read pairs were dereplicated, the amplicon error rate was estimated, and the core algorithm was used to calculate the true ASVs counts in the data. Finally, read pairs were merged, ASVs in each sample were counted and chimeric sequences were removed.

Final 16S ASVs were blasted against a local subset of the GenBank database (search term: “16S,” downloaded 23rd May 2021, 467,306 records), with > 97% identity and output hits limited to 15 sequences. We manually discarded hits with low query coverage (< 90%). We then applied BASTA (a last common ancestor algorithm) to the resulting hits, configured to return a majority taxonomy from 90% of the hits.⁴⁰ Because the tyra (*Eira barbara*) was not represented within the 16S reference data we reran this comparison allowing 96% matches to the nearest ancestor in the reference data, the wolverine, *Gulo gulo* (which is not present in the zoo) and assigned ASVs to *Eira barbara* if there was a 96% match to *Gulo gulo*. ASVs receiving higher level taxonomic assignments and were resolved as follows. ASVs designated as Artiodactyla were resolved to Reeve’s muntjac (*Muntiacus reevesi*) as the other similar match to a reference was bay duiker (*Cephalophus dorsali*) and is not present on site. Similarly, ASVs designated as Cervidae were a perfect match to muntjac and a lower match to pampas deer (*Ozotoceros bezoarticus*) which was not present on site. We retain muntjac for these as well. ASVs identified as Herpestidae were perfect or highly similar (> 99%) matches to meercat (*Suicata suiricata*, which was on site) and lower matches (97%) to other species not present, thus we designate these as meercat (*S. suricata*). An ASV identified as tamarin (Genus: *Saguinus*), was resolved to cotton-topped tamarin (*Saguinus oedipus*) based on matches > 99% to that species which was present in the zoo while other potential matches were < 98%.

Data filtering

For both COI and 16S data we excluded the mole-rats, *Heterocephalus glaber* or *Fukomys damerensis* and several fish identifications as expected contamination from the previous experiment using the same equipment²² and as a general presence in our laboratory air due to long term genomics projects. These data represented 2.2% of the assigned reads and while found in very low levels in many PCRs, primarily impacted two samples which otherwise produced almost no data. We similarly excluded all human sequences which are expected as a general contamination in all samples and controls. We then examined negative well contamination and recorded identifications in negative samples and the number of reads. We differentiated identifications which would remain if largest negative well ID number was used as a filter and treat each of the three amplifications separately (e.g., a negative well with 500 reads assigned as a contamination would cause us to flag any ID with 500 or fewer reads assigned, we treat this maximum read count filter separately for COI nested, 16S and 16S nested PCRs, [Data S1A–S1C](#)).

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison of primers and filter pore size

We investigated the impact of amplicon choice (16S, nested 16S, and nested COI) and filter pore size (0.22 μm or 0.45 μm) on the community composition recovered by metabarcoding. We first filtered out read counts which were indicated as possible contamination from their presence in the negative controls while retaining known true positives. We removed ASVs and samples where the total read count across the entire study summed to 0 as a result of this first step. We created a jaccard dissimilarity matrix with the R package “vegan”⁴¹ and visually analyzed the community composition based on this matrix with non-metric multidimensional scaling (NMDS),⁴¹ using 300 random starts, 4 dimensions, and setting the “weakties” argument to false to deal with tight clustering (high similarity) of points. We colored points using the ordiplot function in “vegan,” based on amplicon marker, filter pore size, and sample location (where the air was filtered from). We then performed PERMANOVA to statistically analyze the effects of amplicon marker, pore size, and sample location on the recovered community composition, using the adonis function from the “vegan” package, with 5000 permutations.

Analysis of distance from target enclosure

We then analyzed the effects of the sampling position relative to the animal’s own enclosure (i.e., inside ($n = 27$) or outside ($n = 78$) the animal’s own enclosure) on read counts. Read counts from all three PCR procedures were pooled and mean read counts/location for each identified zoo species were calculated. We also examined the relationship between read counts and distance from the animal’s enclosure in a second model. The distances between the sampling points to the originating enclosures were calculated as a straight line to the nearest meter using google maps satellite view (i.e., the distance between a sampling point which detected tiger DNA and the tiger enclosure). Distance varied from 0 – 276 m, but we excluded zero distance datapoints (i.e., datapoints from inside the

animal's own enclosure, $n = 27$), as this effect had already been examined by the first model. In both cases we used negative binomial mixed effects models using the glmmTMB package⁴² in R version 4.0.2,³⁹ with species and filter ID as random effects (filter ID was necessary as we treated read counts from different species from the same filter as different data points). The first model (measuring the effect of sample position inside or outside the animal's own enclosure) also included a zero inflation term. We checked for over-dispersion and patterns in the model residuals using the DHARMa package.⁴³ In both models, we tested the significance of the “sample position” and “distance” terms in explaining the read counts by calculating the likelihood ratio test using the “drop1” function with a chi-square distribution.

Current Biology, Volume 32

Supplemental Information

Measuring biodiversity from DNA in the air

Elizabeth L. Clare, Chloe K. Economou, Frances J. Bennett, Caitlin E. Dyer, Katherine Adams, Benjamin McRobie, Rosie Drinkwater, and Joanne E. Littlefair

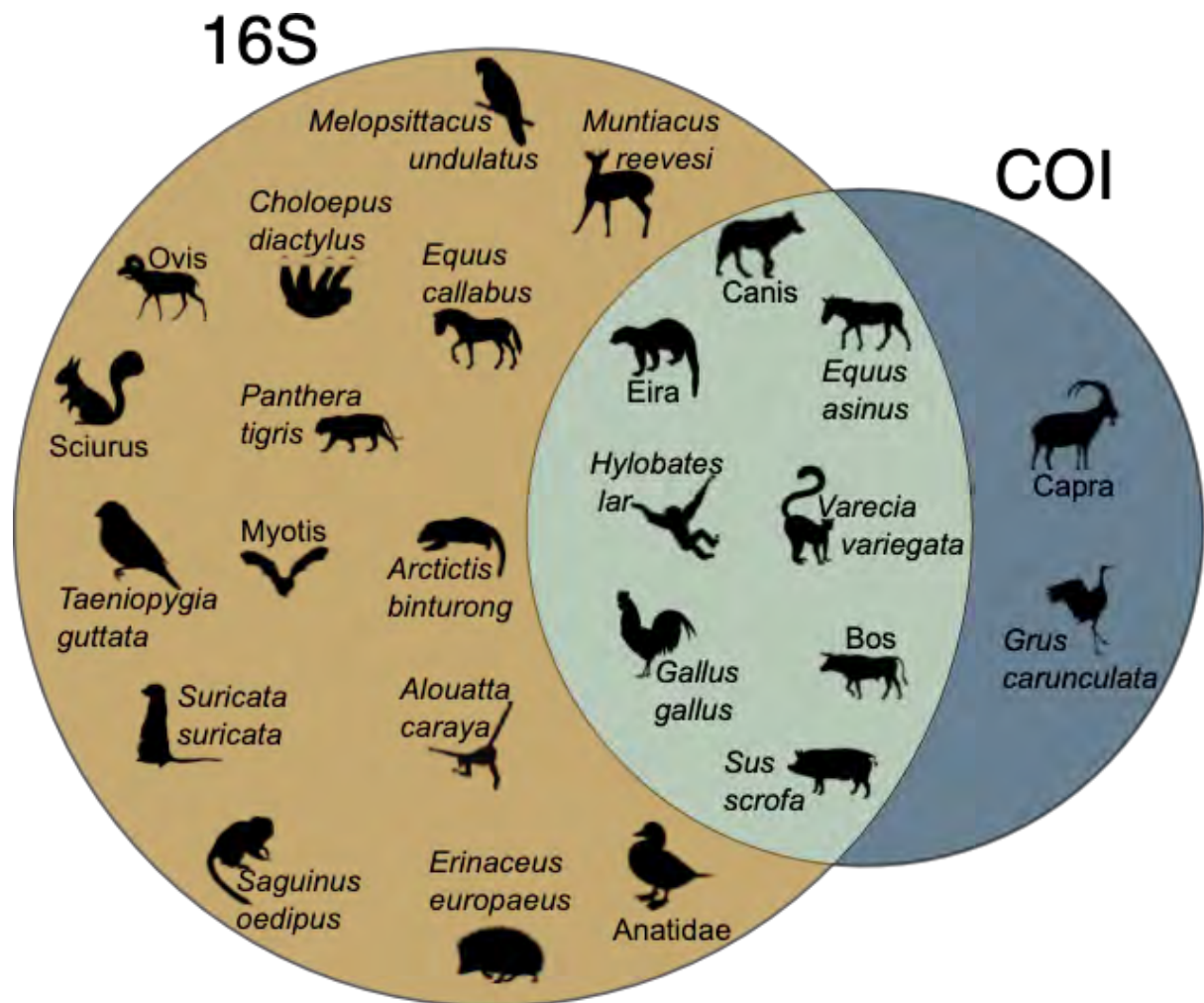


Figure S1. Venn diagram showing the overlap in detections between the two markers. Related to Figure 1, Table 1 and Table 2.

Taxa are identified by either the species or genus name assigned in the analyses with a pictorial representation. Taxa only in the yellow half of the circle indicate those detected using 16S primers (n = 15) while in the grey half are the taxa only detected with the COI primer set (n = 2). In the middle, is the overlap of taxa detected using both primers (n = 8). 16S data includes reads generated by both regular and nested PCR.

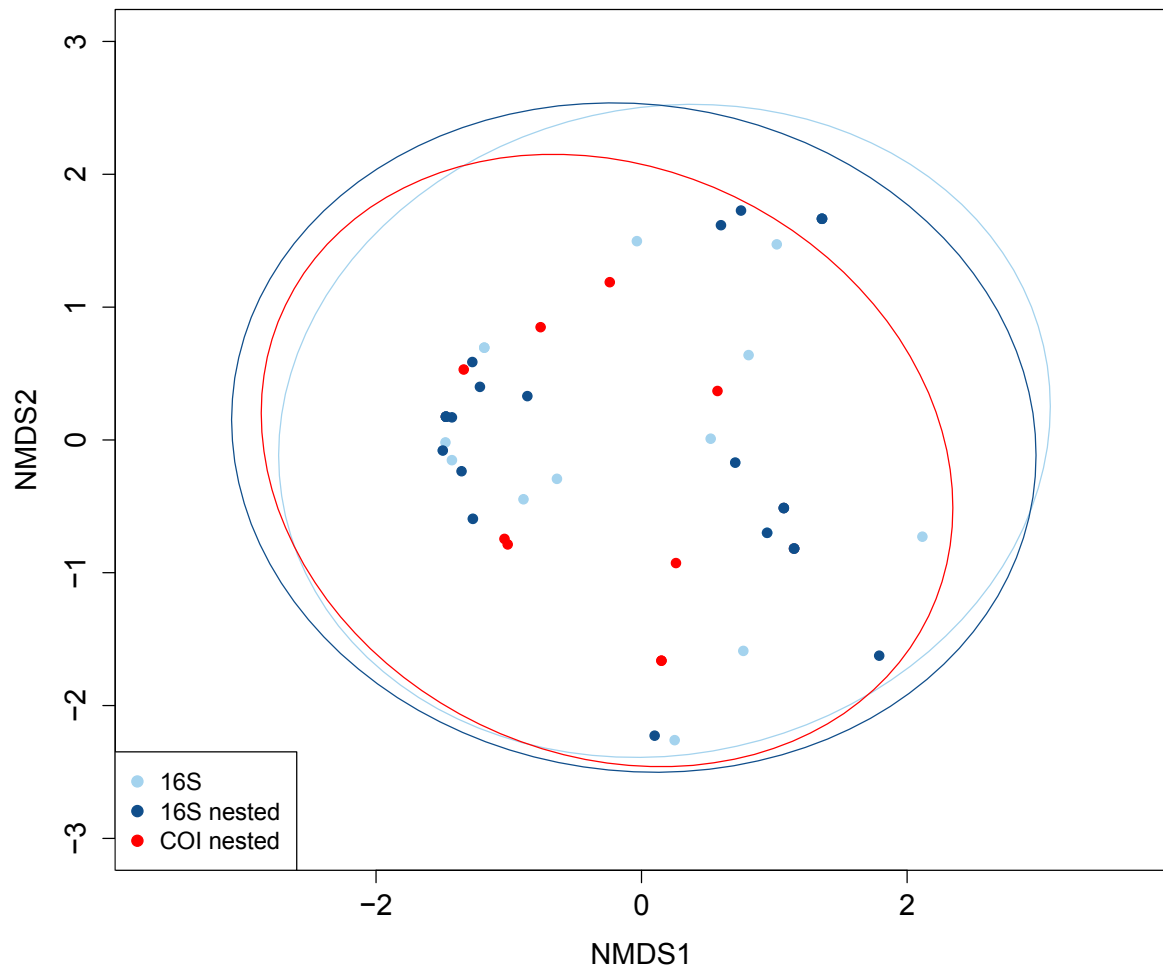


Figure S2. Non-metric multidimensional scaling (NMDS) ordination plot showing the overlap in community composition between the two markers, Related to Figure 1, Table 1 and Table 2.

Red dots indicate COI reads, dark blue dots indicate 16s nested PCR reads while light blue dots indicate regular 16s reads. Ellipses are 95% confidence intervals.