Combining methods for non-invasive fecal DNA enables whole genome and metagenomic analyses in wildlife biology

Alida de Flamingh[1]*, Yasuko Ishida[2], Patrícia Pečnerová[3], Sahara Vilchis[4], Hans R. Siegismund[3], Rudi J. van Aarde[5], Ripan S. Malhi[1,4]* and Alfred L. Roca[1,2]*

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [2]Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [3]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark, [4]Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [5]Department of Zoology and Entomology, Conservation Ecology Research Unit, University of Pretoria, Pretoria, South Africa

Non-invasive biological samples benefit studies that investigate rare, elusive, endangered, or dangerous species. Integrating genomic techniques that use non-invasive biological sampling with advances in computational approaches can benefit and inform wildlife conservation and management. Here, we used non-invasive fecal DNA samples to generate low- to medium-coverage genomes (e.g., >90% of the complete nuclear genome at six X-fold coverage) and metagenomic sequences, combining widely available and accessible DNA collection cards with commonly used DNA extraction and library building approaches. DNA preservation cards are easy to transport and can be stored non-refrigerated, avoiding cumbersome or costly sample methods. The genomic library construction and shotgun sequencing approach did not require enrichment or targeted DNA amplification. The utility and potential of the data generated was demonstrated through genome scale and metagenomic analyses of zoo and free-ranging African savanna elephants (*Loxodonta africana*). Fecal samples collected from free-ranging individuals contained an average of 12.41% (5.54−21.65%) endogenous elephant DNA. Clustering of these elephants with others from the same geographic region was demonstrated by a principal component analysis of genetic variation using nuclear genome-wide SNPs. Metagenomic analyses identified taxa that included *Loxodonta*, green plants, fungi, arthropods, bacteria, viruses and archaea, showcasing the utility of this approach for addressing complementary questions based on host-associated DNA, e.g., pathogen and parasite identification. The molecular and bioinformatic analyses presented here contributes towards the expansion and application of genomic techniques to conservation science and practice.

Introduction

Non-invasive biological samples benefit studies that investigate rare and elusive (Ferreira et al., 2018; Franklin et al., 2019), endangered (Ang et al., 2020) and/or dangerous species (Bellemain and Taberlet, 2004; Mondol et al., 2009). Non-invasively collected samples can be obtained without directly interacting with animals, so that collection of samples do not impact the wellbeing of the animal from which it is collected (Taberlet et al., 1999; Lefort et al., 2019; Sappington, 2019). This approach allows researchers to increase sample representation, by increasing geographic distribution and/or taxonomic diversity, or by substituting for physical sample collection procedures that may be arduous, expensive, or potentially harmful to the animal (e.g., chemical immobilization). The use of non-invasive samples has been complemented by advances in molecular techniques (Andrews et al., 2021) that progressively allow for smaller quantities of sample (and associated DNA template) to be sufficient for complex molecular analyses (Glocke and Meyer, 2017; Rohland et al., 2018; Xavier et al., 2021). Wildlife conservation genomics, the application of genomic techniques to inform conservation and management of species (Allendorf et al., 2010; Supple and Shapiro, 2018; Hohenlohe et al., 2021), can benefit from these technological advances, especially when used in synergy with non-invasive biological samples (Andrews et al., 2021; Hohenlohe et al., 2021). For example, combining non-invasive sampling with molecular analyses can benefit research and management of endangered species that may be difficult to sample due to low abundance (Palomares et al., 2002; Baker et al., 2013; Carroll et al., 2018), or where invasive sampling can impact behavior and/or sociality (DelGiudice et al., 2001; Pelletier et al., 2004; Arnemo et al., 2006; Becciolini et al., 2019).

We here combine established DNA collection, extraction and sequencing protocols, allowing for the accessible use of non-invasive fecal DNA samples to generate low- to medium-coverage whole genome and metagenomic sequences. In this study we aimed to combine the following: 1) a simplified and accessible DNA collection and extraction protocol that uses a standard DNA preservation card, which does not require refrigeration and is easy to transport and store; 2) a protocol for genomic library construction and shotgun sequencing that does not require enrichment or targeted DNA amplification; and 3) evaluate the utility and potential of the data generated through the application of genome-scale analysis and metagenomics of zoo and free-ranging elephants in their native environments. The molecular and bioinformatic combination of methods used here precludes additional efforts associated with host-specific DNA enrichment and allows for the simultaneous analysis of host-associated DNA (e.g., DNA that is not from the host genome, but is extracted as part of the fecal sample, for example, fecal microbiome data) concomitant with host-specific DNA.
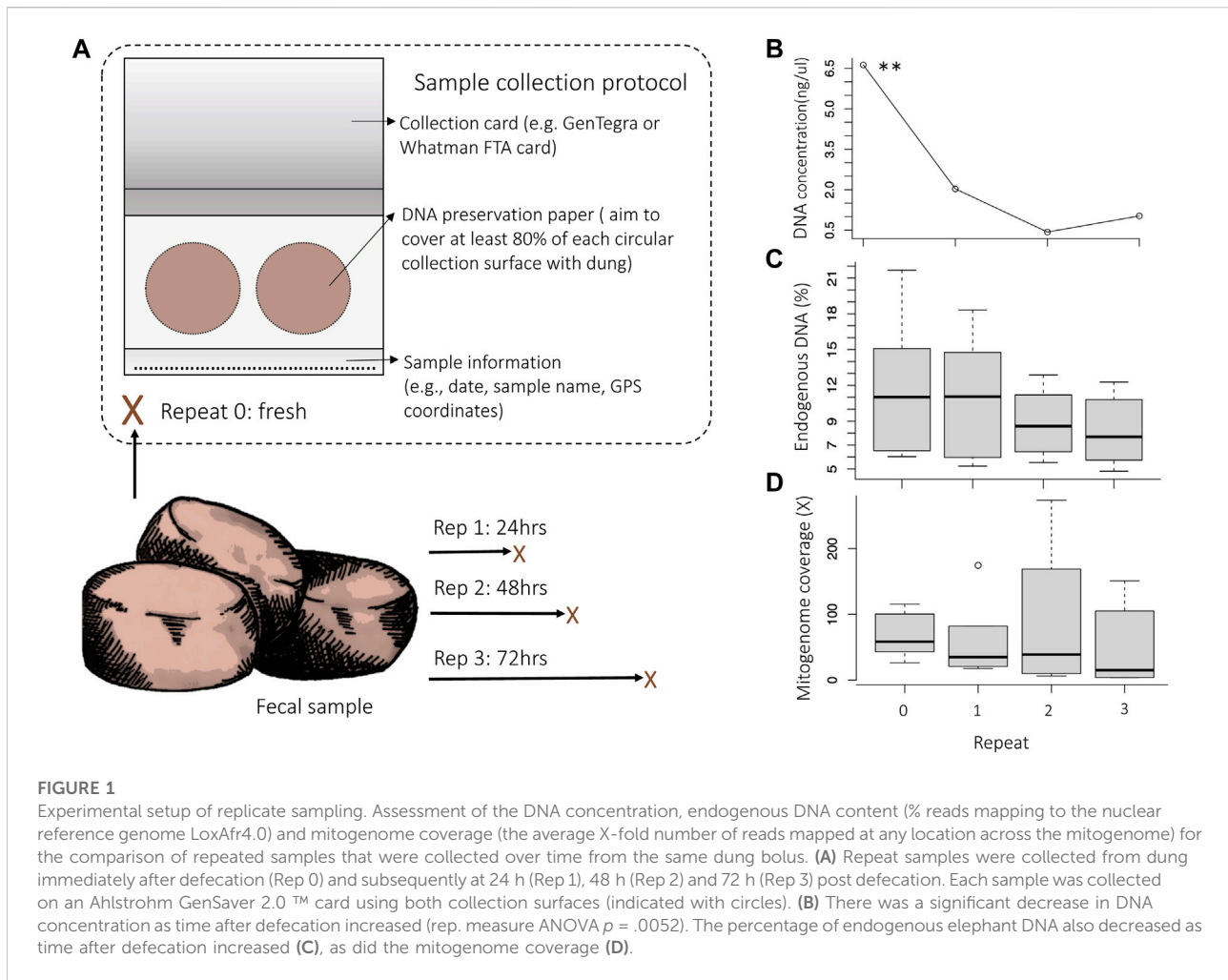
Elephants (*L. africana*) are ideal candidates for testing a non-invasive approach because invasive sampling can be unsafe and expensive (Jacobson et al., 1988; Kock et al., 1993), and they occur across very large and sometimes difficult to access geographic areas (Gray et al., 2014). Non-invasive samples have been used extensively to study elephants, e.g., to establish relatedness and demography (Munshi-South, 2011), investigate hybridization between forest and savanna elephants (Bonnald et al., 2021), study population structure and gene flow (de Flamingh et al., 2018), and estimate population size (Gray et al., 2014). We used samples from zoo individuals to investigate DNA content and preservation relative to dung freshness. We quantified DNA preservation over time using repeated sampling of the same fecal bolus across a 3-day (72 h, Figure 1) time-series in zoo individuals. We then verified the practicality and effectiveness of the combination of methods in free-ranging elephants from five localities in South Africa.

Methods and materials

Study design

We tested the use of DNA preservation cards (paper treated with chemicals that prevent microbial growth) in zoo elephants using a time-series of repeated samples (Figure 1). Three independent sets of samples were used (Supplementary Table S1); the first set included 24 fecal samples from six elephants under human care from two zoos in the United States of America (Dataset 1), the second set included 13 fecal DNA samples from 13 free-ranging elephants in five geographic localities in South Africa (Dataset 2), and the third set included reference samples from known geographic locations and for which genomic data had been previously generated from high quality tissue and blood samples (Dataset 3).

For Dataset 1, samples were collected using the GeneSaver2™ Card (see Supplementary File 2–DNA collection protocol) for three elephants housed at Jacksonville Zoo and Gardens, Florida, and three elephants from Dallas Zoological Gardens, Texas. For each elephant, four repeated samples were collected over time from the same dung bolus to quantify DNA degradation and to estimate how dung freshness may impact target-species DNA preservation. Repeated samples were collected from the same dung bolus (Figure 1) over 3 days: immediately after defecation (Rep 0), and then at 24 h (Rep 1), 48 h (Rep 2), 72 h after defecation (Rep 3). These time-points were selected based on

FIGURE 1
Experimental setup of replicate sampling. Assessment of the DNA concentration, endogenous DNA content (% reads mapping to the nuclear reference genome LoxAfr4.0) and mitogenome coverage (the average X-fold number of reads mapped at any location across the mitogenome) for the comparison of repeated samples that were collected over time from the same dung bolus. **(A)** Repeat samples were collected from dung immediately after defecation (Rep 0) and subsequently at 24 h (Rep 1), 48 h (Rep 2) and 72 h (Rep 3) post defecation. Each sample was collected on an Ahlstrohm GenSaver 2.0 ™ card using both collection surfaces (indicated with circles). **(B)** There was a significant decrease in DNA concentration as time after defecation increased (rep. measure ANOVA p = .0052). The percentage of endogenous elephant DNA also decreased as time after defecation increased **(C)**, as did the mitogenome coverage **(D)**.

previous studies that investigated elephant fecal DNA degradation across similar timeframes (Vidya and Sukumar, 2005; Bourgeois et al., 2019). Elephant dung was exposed to conditions that mimic natural environments between sampling events, e.g., samples from Jacksonville Zoo and Gardens were exposed to sun for half the day and on one sampling day there was light rain (staggered collection meant that rain was on Rep 3 and Rep 2 for samples; Supplementary Figure S1).

For Dataset 2, samples from 13 free-ranging elephants were collected from five geographic areas within South Africa (Supplementary Table S1; Supplementary Figure S2). Collectors targeted fresh dung, or dung that still had wet mucus (see Supplementary File 2–DNA collection protocol).

Dataset 3 included 24 high-quality DNA samples that were geographically provenanced. These samples were used as references to evaluate the data that we generated in this study. This geo-referenced dataset included genomic data from four tissue samples from Kruger National Park, South Africa (Pečnerová et al., *in prep*), one blood sample from Kenya and one blood sample from Kruger National Park (Palkopoulou et al., 2018), and 18 blood samples from Gorongosa National Park, Mozambique (Campbell-Staton et al., 2021; Supplementary Table S1; Supplementary Figure S2).

Fecal DNA collection, extraction, and genomic library construction

Fecal sample collection used DNA preservation cards, which have also been used previously for conservation genomic (Régnier et al., 2011; Ward et al., 2019; Morrison et al., 2020) and microbiome research (Song et al., 2016; Yarlagadda et al., 2022). We used Ahlstrom GenSaver 2.00™ cards, which are paper cards that have been treated with a chemical to minimize environmentally induced DNA degradation and the growth of microorganisms, allowing for collection, transport, and long-term preservation of DNA from biological samples at ambient temperature (Ahlstrom Corporation, Helsinki, Finland). Fecal

DNA was collected (Supplementary Figure S3) using a standardized collection protocol (Supplementary File 2–DNA collection protocol). For free-ranging elephants, samples represented individuals from different herds, geographic locations, and different times of the day.

DNA was extracted using the QIAamp® PowerFecal® Pro DNA Kit (Qiagen, Ann Arbor, United States of America). Modifications to the standard DNA extraction protocol included the use of approximately one quarter of the collection card (0.5 cm × 1.0 cm) as the starting material (Supplementary Figure S3; card was further cut into small pieces to maximize cell lysis), after which sample fecal matter and 800 μl CD1 solution were vortexed for 10 minutes. A maximum of four samples were processed per round of extraction to avoid cross-sample contamination. The final DNA elution step was repeated twice, each time using 50 μl of Elution Buffer, resulting in a final DNA elute volume of 100 μl.

Extracted DNA concentration was quantified using the Qubit Broad-Range dsDNA platform (ThermoFisher Scientific, Waltham, United States of America) and DNA concentration was compared across repeated samples (Dataset 1) using a repeated measures ANOVA in R (Core Team R, 2013) and a Bonferroni correction (source code available from https://github.com/adeflamingh/de_Flamingh_et_al_2022_FiG). DNA concentration and endogenous content (see details below; Supplementary Table 2) were compared between fresh (Rep 0) samples from zoo individuals (Dataset 1) and samples from free-ranging individuals (Dataset 2) using a Welch Two Sample *t*-test. In addition, each repeat (Rep 1, 2, and 3) of Dataset 1 was also compared with Dataset 2 and a Benjamini-Hochberg correction was applied to account for multiple comparisons (Benjamini and Hochberg, 1995) in the program R (https://github.com/adeflamingh/de_Flamingh_et_al_2022_FiG).

Genomic libraries were constructed using a tagmentation library building approach as part of the Illumina DNA Prep Kit (ThermoFisher Scientific, Waltham, United States of America) and using the protocol described by Yarlagadda et al. (2022). Reps 2 and 3 of one of the zoo individuals did not show evidence of successful DNA extraction based on Qubit quantitation and amplification of the shorter 500 bp mitochondrial sequence. To account for possible contamination, we also included negative control libraries with all rounds of sample processing. These libraries were constructed using blank GenSaver 2.00™ cards and the same reagents that were used for the fecal DNA libraries. We used IDT for Illumina DNA/RNA Unique Dual Indexes (Catalog number 20027213, Illumina, San Diego, Unites States), and shotgun-sequenced the pooled samples at the Roy. J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign (UIUC), United States of America. Genomic library fragment size distribution was assessed by means of AATI (Advanced Analytical Technologies, Inc. Fragment Analyzer). Samples were sequenced in three independent runs (Supplementary Table S3, Supplementary Table S4) using the Illumina NovaSeq 6,000 sequencing platform; the first two rounds of sequencing included the zoo, free-ranging and control libraries, and the third round included repeated sequencing (re-sequencing) of libraries of three individuals to increase depth and breadth of genome coverage (see below, Extended genomic analysis and prospective applications for details).

Bioinformatic analyses

Quality control and genome alignment

Bioinformatic analysis used the Biocluster2 high-performance computing system at the Carl R. Woese Institute for Genomic Biology, UIUC, United States of America. Samples were de-multiplexed and the reads trimmed to a minimum length of 25 bp using the program FastP v.0.19.6 (Chen et al., 2018). Reads were aligned to a reference African savanna elephant mitochondrial genome (mitogenome; GenBank accession number JN673264, Brandt et al., 2012) and nuclear genome (LoxAfr 4.0; Meyer et al., 2017) using bowtie2 (Langmead and Salzberg, 2012) and BWA (Li and Durbin, 2010). Both alignment programs produced a similar number of reads mapping to mitogenome and nuclear genome reference sequences, and we used alignment files that were generated using the BWA-mem algorithm and default parameters in BWA for subsequent analysis. Alignments were transformed into BAM format using SAMtools v. 1.1 (Li et al., 2009), and filtered to remove unmapped reads and reads with a minimum alignment quality score less than 30. Filtered BAM files were then sorted and indexed, with PCR duplicates marked and removed with the Picard Toolkit v. 2.10.1 (Picard Toolkit, 2019, Broad Institute). Genome alignment statistics were calculated for each individual sample; we calculated the breadth (%) of each complete genome covered by reads; and the depth of coverage (X-fold) as the average number of reads mapping to each position on the complete genome using SAMtools (source code available at github.com/adeflamingh). The genomic percentage of endogenous elephant DNA was calculated as the fraction of reads mapping to the elephant nuclear genome (LoxAfr 4.0) to the total number of reads per sample.

Metagenomic analysis and taxonomic classification

Unprocessed fastQ files were quality filtered and trimmed using Trimmomatic (Bolger et al., 2014): we removed adapters, leading and tailing low quality bases, and "N" bases with quality below three; we scanned the reads with a 4 bp sliding window, cutting reads where the average quality per bp dropped below 15; to increase specificity for our metagenomic classification, we

removed reads that were less than 50 bp long after trimming and only reads that were properly paired were retained for subsequent metagenomic analysis.

We used Kraken2 (Wood et al., 2019) to classify reads into taxonomic units. Kraken2 is a metagenomic sequence classifier that uses an ultrafast k-mer based approach to classify and assign taxonomic labels to short DNA reads (Wood et al., 2019). We compiled a custom search database to which reads were compared using the "kraken2- build" command. In addition to the African savanna elephant genome (LoxAfr 4.0, Meyer et al., 2017), our custom database included the RefSeq libraries archaea, bacteria, plasmid, viral, human, fungi, plant, and protozoa, the NCBI non-redundant nucleotide database, and UniVec_Core (NCBI database of vector, adapter, linker, and primer sequences). We concatenated pairs of reads together using the "--paired" function in Kraken2, parallelized the classification to run using 24 computing threads, and used a confidence scoring threshold of 0.05 to minimize erroneous classification and false positive rates (see Supplementary Figure S4 for details on confidence scoring threshold value). We also included the option for generating a summary report "—report" which was used estimate proportional read composition for comparison across repeated samples. To account for possible collector bias (e.g., different quantities of sample collection; Supplementary Figure S3), the comparative visualizations of compositional taxonomic groups were plotted separately for Dallas Zoological Gardens, Jacksonville Zoo and Gardens, and for samples collected from free-ranging elephants.

Extended genomic analysis and prospective applications

We assessed whether the data generated by the fecal DNA card collection and library methods used in this study may allow for genome-level analysis. We evaluated the quality of our data by means of a phylogeographic analysis approach that was developed for low-coverage shotgun sequencing data (Yao et al., 2020). This approach uses genotype likelihoods rather than called genotypes to estimate single nucleotide polymorphisms (SNPs) in ANGSD (Korneliussen et al., 2014). Using a principal component analysis (PCA) in PCAngsd (Meisner & Albrechtsen, 2018), we determined whether the genomic data obtained from fecal cards cluster with high-quality genomes originating from the same geographic location. To generate higher coverage genomic data for fecal samples, we re-sequenced genomic libraries for three fecal card samples: SAA10 from Addo Elephant National Park, SAT04 from Tembe Elephant Park, and WNP01 from the Knysna forest. These samples represented a range of DNA quality and quantity (both in DNA concentration and percentage endogenous DNA; Supplementary Table S3 and Supplementary Table S5). Re-sequenced libraries were

independently aligned to the reference genome, we then merged and deduplicated DNA alignments for each individual using SAMtools and compared these three samples to 24 geographically referenced samples (Dataset 3), resulting in a total of 27 samples. We were primarily interested in demonstrating the utility of the approach by assessing clustering patterns in the PCA, and thus did not focus on inter- and intra-population differences in genetic differentiation and composition. Following Li et al. (2020), our phylogeographic approach relied on ANGSD to filter and compile a SNP dataset; we only retained SNPs that were present in at least half (14) of the 27 individuals and which had a p-value of 0.01 or less (indicative of the statistical likelihood of the position being a variable site). A total of 3,112,723,441 sites were analyzed, 29,116,747 sites were retained after ANGSD filtering, and 12,715,208 sites were used for subsequent PCA analysis. We used PCAngsd to perform a PCA analysis, and R to visualize the clustering patterns (source code available at github.com/adeflamingh). To investigate whether small population size and inbreeding might explain the PCA clustering patterns, specifically the clustering patterns observed for the elephants from Gorongosa (see Discussion), we estimated genome-wide heterozygosity (GWH) in ANGSD as the proportion of heterozygous genotypes (analogous to theta-based estimates). Estimates of GWH have been used as a proxy for inbreeding, where inbreeding increases the homozygosity across the genome of an individual (Hansson and Westerberg, 2002; Balloux et al., 2004). WNP01 was excluded from the GWH analysis to remove low genome coverage-associated bias, and we included five-hundred million sites ("nSites = 500,000,000") in the GWH calculation for all other individuals. In addition to the phylogeographic analysis, we also used the Rx method developed by de Flamingh et al. (2020) to estimate the biological sex of the re-sequenced elephant with the lowest coverage (WNP01). We chose to estimate the sex of the elephant with the lowest coverage to set a lower-bound for sequencing coverage needed to successfully estimate sex, and because the sex of this individual is known based on photographic and observational evidence (Moolman et al., 2019a) while the sex of the other re-sequenced free-ranging individuals is unknown. We did not estimate the sex of other individuals as we were only interested in demonstrating the applicability of our method, rather than investigating sex-associated genomic patterns across the datasets.

Results

Collection card comparison

Preliminary testing compared two types of DNA preservation cards, the Whatman Flinders Technology Associates (FTA) card (Whatman plc, Maidstone,

United Kingdom) and the Ahlstrom GenSaver 2.00™ card (GenTegra, Pleasanton, United States of America). Both types of cards resulted in similar extracted DNA quantity and quality as quantified using Qubit DNA quantitation and based on targeted PCR amplification of a 500 bp elephant mitochondrial DNA fragment (Nyakaana and Arctander, 1999). We used only Ahlstrom GenSaver 2.00™ cards for subsequent DNA collection.

Method validation (dataset 1)

To determine and quantify the effectiveness of the card collection and library building protocols, we assessed DNA concentration and endogenous elephant DNA content using repeated samples that were collected across a time interval (Dataset 1; Figure 1). The concentration of the extracted DNA was examined using broad-range (BR) Qubit quantitation and ranged from undetectable to 16 ng/μL (Supplementary Table S2). There was a significant decrease in DNA concentration as time after defecation increased (rep. measure ANOVA $p = .0052$). Consistent with the decrease in DNA concentration, the percentage endogenous DNA (Figure 1C) and the mitogenome coverage (Figure 1D) also decreased as time after defecation increased. For some samples, the DNA concentration was below the level detectable by BR Qubit quantification. However, mitochondrial genomes could be reconstructed for all sequenced samples, even for those that had DNA concentrations that were too low to be quantified with BR Qubit quantitation (<1 ng/μL for our samples), suggesting that DNA was still preserved in the collection card even when DNA quantitation failed. This was supported by DNA fragment analysis, which showed DNA fragments of variable length were present in the genomic libraries (see Supplementary Figure S5 for an example of a DNA fragment distribution curve for one of the samples that had unquantifiable DNA). We recommend that samples be evaluated using DNA fragment analysis prior to sequencing to verify that DNA is present in the genomic library, or alternatively, using reagents with higher sensitivity (e.g., Qubit dsDNA high-sensitivity quantitation) or using quantitative PCR may allow for quantitation and quality assessment of genomic libraries that have very low DNA concentrations (<1 ng/μL). More importantly, even samples collected 72 h after defecation (the longest collection time examined) contained enough endogenous DNA to allow for high-coverage complete mitogenome reconstruction (Supplementary Table S2). For example, the complete mitogenome for the Rep 3 sample from "Jenny" could be reconstructed to 151 X-fold coverage.

Metagenomic classification of the reads in Dataset 1 into taxonomic units revealed that the composition of the samples changed as time after defecation increased for elephants from the Dallas Zoological Gardens (Supplementary Figure S6) and from

the Jacksonville Zoo and Gardens (Supplementary Figure S7); in general, the proportional contribution of bacterial DNA increased, and other classified taxonomic groups (except fungi) decreased as the dung sample aged. There was also an increase in number of reads that could not be classified as time after defecation increased (Supplementary Figure S6 and Supplementary Figure S7; dark grey). The proportional contribution of endogenous elephant DNA was consistently higher in the samples that were collected from fresh dung (Rep 0) and decreased with sample age. Taxonomic classification showed that most samples contained reads that originated from bacteria, endogenous elephant DNA, arthropods, other eukaryotes, and viruses.

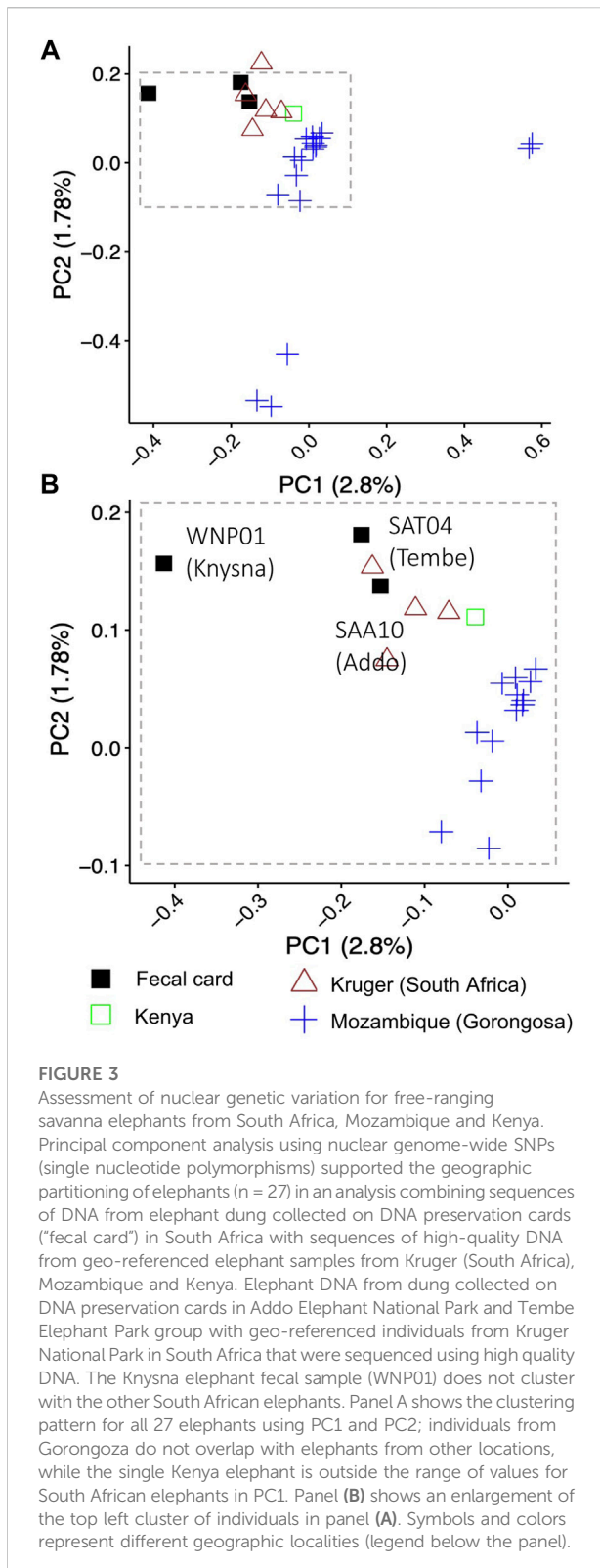## Reproducibility in the field (dataset 2)

From free-ranging elephants, collectors targeted fresh dung, and placed on the cards a larger volume of fecal matter than was collected for zoo elephants (Supplementary Figure S3). The DNA concentration and percentage endogenous elephant DNA did not differ significantly between fresh zoo samples (Rep 0) and samples from free-ranging elephants (Supplementary Table S5; Welch t-test for concentration, $p = 0.34$; Welch t-test for endogenous content, $p = 0.87$). DNA concentrations in fecal samples from zoo elephants that were collected 24 h, 48 h and 72 h after defecation (Rep 1–3) were significantly lower (Welch t-test; Rep 1: $p = 0.032$; Rep 2: $p = 0.01$; Rep 3: $p = 0.014$) than the DNA concentrations for samples from free-ranging individuals.

For Dataset 2 which included all of the free-ranging individuals, we were able to reconstruct complete mitochondrial genomes (≥98% breadth of coverage), with a depth of coverage ranging from 6.9 to 156.6 X-fold (average depth = 80 X-fold; Supplementary Table S3). We were also able to generate low-coverage genomic data for partial nuclear genomes, ranging from 3.97 to 35.98% of the breadth of the nuclear genome at a coverage depth of 0.01–0.60 X-fold (Supplementary Table S3). Based on the proportion of reads aligning to the African elephant reference genome, samples in Dataset 2 contained an average of 12.41% (5.54–21.65%) endogenous (elephant) DNA. Metagenomic analysis found that on average 21.42% (0.02–34.46%) of classified reads were assigned to the genus *Loxodonta* (Supplementary Table S6).

Genome alignment statistics for re-sequenced libraries

Three genomic libraries (SAA10, SAT04 and WNP01) were submitted for an additional round of sequencing, and alignments of the independent libraries were merged for each sample (Supplementary Table S3). For SAA10, nuclear genome coverage increased from 0.36 X-fold representing 17.85% of

FIGURE 2
Metagenomic classification of fecal DNA samples from 13 free-ranging elephants from South Africa. Compositional taxonomic contributions of classified reads are summarized as a pie chart for each individual (unclassified reads are not shown). Taxonomic groups have been color coded and corresponding colors are indicated in the legend in the top row of the figure. Taxonomic classification showed that most samples contained reads that originated from bacteria, endogenous elephant DNA, green plants, arthropods, and other eukaryotes.

the breadth of the genome, to 5.8 X-fold representing 93.1% of the genome. For SAT04, nuclear genome coverage increased from 0.45 X-fold representing 30.56% of the genome, to 4.2 X-fold representing 90% of the genome. For WNP01, nuclear genome coverage increased from 0.2 X-fold representing 4.01% of the genome to 2.4 X-fold representing 14.96% of the genome. There was also a corresponding increase in coverage observed for the mitogenome, with SAA10, SAT04 and WNP01 respectively having mitogenome coverages of 1248 X-fold, 668 X-fold and 193 X-fold reads.

Extended genomic analysis and prospective applications (Dataset 1, 2 and 3)

Metagenomics

Metagenomic analysis revealed that fecal samples from free-ranging elephants contained reads that originated from the same main classification groups (*Loxodonta*, green plants, fungi, arthropods, bacteria, viruses, archaea) that were detected in zoo individuals. Zoo and free-ranging elephant samples did not contain reads of human origin except for a single zoo sample (Thandi, Rep 2) and a single free-ranging elephant sample (SAA01), both of which had ≤0.01% of total reads

assigned to *Homo sapiens*. The average number of classified reads assigned to *Loxodonta* was higher in free-ranging (21.42%) than zoo (7.9%) individuals, and the relative contribution of bacteria was lower in free-ranging individuals (Figure 2, brown portion) while other taxonomic groups such as arthropods (Figure 2, black portion) were more abundant in free-ranging than zoo samples.

Most samples (zoo and free-ranging) contained reads that originated from arthropods. For example, samples of fresh dung (Rep 0) from "Jenny" and "Mlilo" from Dallas Zoological Gardens contained reads originating from Arthropoda (Supplementary Figure S6) that respectively contributed 0.68% and 0.89% of all reads (including unclassified reads), or 3.2% and 4.1% of classified reads. Of these classified reads, a substantial proportion (45% and 42.7% of reads classified as Arthropoda) was identified as belonging to the lepidopteran clade Ditrysia. Similar patterns and taxon identifications were observed for samples collected from free-ranging individuals (e.g., SAA03, SAA06 and SAT02 contained a substantial proportion of reads originating from Ditrysia).

In some individuals a proportion of the reads originated from Archaea. For example, "Thandi" from the Jacksonville Zoo and Gardens (Supplementary Figure S7) had a substantial portion of reads assigned to the genus *Methanobrevibacter*, a dominant gut-associated archaeon (Hansen et al., 2011). There were relatively

FIGURE 3
Assessment of nuclear genetic variation for free-ranging savanna elephants from South Africa, Mozambique and Kenya. Principal component analysis using nuclear genome-wide SNPs (single nucleotide polymorphisms) supported the geographic partitioning of elephants (n = 27) in an analysis combining sequences of DNA from elephant dung collected on DNA preservation cards ("fecal card") in South Africa with sequences of high-quality DNA from geo-referenced elephant samples from Kruger (South Africa), Mozambique and Kenya. Elephant DNA from dung collected on DNA preservation cards in Addo Elephant National Park and Tembe Elephant Park group with geo-referenced individuals from Kruger National Park in South Africa that were sequenced using high quality DNA. The Knysna elephant fecal card sample (WNP01) does not cluster with the other South African elephants. Panel A shows the clustering pattern for all 27 elephants using PC1 and PC2; individuals from Gorongoza do not overlap with elephants from other locations, while the single Kenya elephant is outside the range of values for South African elephants in PC1. Panel **(B)** shows an enlargement of the top left cluster of individuals in panel **(A)**. Symbols and colors represent different geographic localities (legend below the panel).

few reads classified as originating from Fungi across all samples (zoo and free-ranging), with the highest proportion of fungal reads present in the samples for which the most time had passed

since deposition (e.g., Rep 3 samples for Ali and Thandi; Supplementary Figure S7). Metagenomic analysis also identified reads originating from viruses, e.g., multiple samples (SAA05, SAT04; Figure 2) contained contributions from Caudovirales.

PCA and genetic sex

A PCA using nuclear genome-wide SNPs revealed geographical partitioning of the elephants (Supplementary Figure S2). Elephants for which fecal card samples were collected from Addo Elephant National Park and Tembe Elephant Park in South Africa grouped most closely with geo-referenced elephants (Dataset 3) that were also from South Africa (e.g., Kruger) (Figure 3B; cluster at top left of panel). The Knysna elephant from South Africa (fecal card sample WNP01) did not cluster with any of the geo-referenced elephants when considering PC1 (Figure 3B, Supplementary Figure S8A). However, all three elephants from South Africa from which DNA was collected using fecal cards did cluster together when PC2 and PC3 were considered (Supplementary Figure S8B). Each of the principal components contributed very little to the overall genomic variation observed (PC1: 2.8%, PC2: 1.78%, PC3: 1.67%). Elephants from Gorongosa National Park, Mozambique were separated by PC1 and PC2 into three clusters (Figure 3A, blue symbols). The separate PCA positioning of elephants from Gorongosa National Park did not correspond to genome-wide heterozygosity (Supplementary Figure S9), and the overall clustering pattern for all elephants was not concomitant with genome sequence coverage (Supplementary Figure S10; all clusters included a range of coverages). Finally, we examined whether the biological sex of the elephants could be determined using DNA sequencing reads from fecal cards. We examined the elephant with the lowest genome coverage of the three South African elephants, the Knysna elephant (WNP01), and were able to identify this elephant as a female (Rx: 0.8274169, $p < .05$) using the approach of de Flamingh et al. (2020).

Discussion

We present a combination of methods that includes a simplified and accessible DNA sample collection and extraction protocol using a DNA preservation card that does not require refrigeration after sample collection, and is easy to transport and store in the field and in the laboratory. Other fecal DNA sample collection methods rely on refrigeration and/or chemical preservation (Murphy et al., 2002; Nsubuga et al., 2004). Access to refrigeration is often not feasible due to mode of transport (e.g., itinerant and/or trekked field expeditions) or because of limited access to electricity in undeveloped/rural areas. Previous means of chemical preservation of elephant fecal samples for DNA analyses have involved chemicals that

may be cumbersome to handle, may be dangerous, and may require refrigeration, e.g., ethanol or preservation buffers (Wasser et al., 2004; Gobush et al., 2008; Gray et al., 2011). In addition to these risks, the international transport or shipment of chemicals (e.g., ethanol) is often regulated and sometimes prohibited. Such methods of DNA sample collection may therefore hinder research projects. The use of fecal DNA preservation cards avoids these issues and provides an easy and accessible way for in-field researchers and conservation practitioners to collect and transport many samples at once.

We tested and allowed for DNA analysis of fecal samples that were up to 72 h (3 days) old, meaning that samples could be opportunistically collected in the field even days after animals have left an area (Figure 1). We were able to generate complete mitogenomes for all shotgun-sequenced zoo individual samples collected 72 h after defecation (>90% of the mitogenome with an average coverage of 55.7 X-fold for the samples). However, samples that were collected shortly after defecation more consistently resulted in high DNA concentration, high mitogenome coverage, and high endogenous target-species DNA content. To maximize the possibility of successful DNA collection, sample collectors should ensure that samples are completely dry before long-term storage, as DNA is prone to degradation if samples are not completely desiccated (Murphy et al., 2002; Nsubuga et al., 2004). We recommend storing DNA preservation cards with a desiccant (Supplementary File 2–Step 5). The preservation card collection protocol follows five easy steps, and no specialized training is required, therefore, samples may also be collected by persons who do not have formal scientific research training. In addition, the DNA preservation cards used are relatively inexpensive (at the time of this study the cost was less than US$ 5.00 per card with two collection surfaces for each bolus of dung). Although the percentage of endogenous DNA did not differ significantly between samples with different amounts of fecal matter (Supplementary Figure S3), we recommend that at least 80% of each collection surface should be covered by fecal matter to maximize the possibility of obtaining and preserving elephant DNA (see Supplementary File 2: Collection protocol).

Our study did not require host targeted DNA enrichment (Chiou and Bergey, 2015) for the sequencing of complete mitochondrial genomes and low-coverage nuclear genomes. On average our method resulted in 10–20% endogenous host-specific DNA (based on bioinformatic and metagenomic statistics; Supplementary Table S3, Supplementary File 4). We show that library re-sequencing can produce datasets that represent >90% of the complete nuclear genome of the target species, with coverages up to six X-fold and potentially higher. However, the resulting coverage of re-sequenced libraries is dependent on the diversity of available DNA template in the original library (Daley and Smith, 2014), as limited DNA template diversity will result in replicate sequencing of the same DNA molecule, resulting in PCR duplicate reads rather than unique reads that contribute to the breadth (%) and depth (X-fold) of genome coverage. For example, initial sequencing of WNP01 resulted in a comparatively smaller percentage of nuclear genome coverage (4% compared to ~17% in the other two re-sequenced individuals), and the re-sequencing of this library was less productive (~15% of the nuclear genome was recovered, compared to >90% for the other two re-sequenced libraries), likely due to limited DNA template molecule diversity. We therefore encourage researchers to conduct an initial sequencing screen to calculate genome coverage statistics that can be used to inform the choice of libraries to re-sequence. Initial screening of sequencing data can be conducted using software that predicts library complexity and prospective genome coverage, e.g., the program "preseq" (Daley et al., 2014).

Sequencing efforts would need to be increased to obtain target-species genome coverages similar to those of other methods that use bait or other enrichment protocols. However, the time, effort, and reagent cost associated with complex DNA- and RNA-bait design and laboratory enrichment may offset the potential cost of increased sequencing, especially as the costs of sequencing may continue to decrease. The molecular and bioinformatic combination of methods reported in this study has the potential to be used for additional analyses not pursued here, for example, for quantifying gene flow and relatedness using genotype likelihood scores; see Korneliussen and Moltke (2015) for a software tool for estimating pairwise relatedness from genotype likelihoods. However, researchers may seek to address specific research aims that do not require a holistic approach such as the one presented here. For example, the use of microsatellite analysis may be preferred to whole genome sequencing for establishing parentage (Thitaram et al., 2008) or pedigree (Okello et al., 2005) or geographic profiling using an existing database (Wasser et al., 2015). However, the approach used here allows for host-associated DNA to be investigated concomitant with host-specific DNA.

The ability to generate nuclear genome-wide data from non-invasively collected fecal samples can open up new possibilities for genomic analysis. We demonstrated the utility and potential application of the combination of methods used by generating and analyzing data for African savanna elephants. This may also be directly applicable to other taxa. We showed that the nuclear SNP dataset generated for non-invasive fecal DNA from free-ranging elephants is sufficient for complex molecular and bioinformatic analyses. For example, our analysis indicated that the fecal card samples grouped with other individuals from South Africa (e.g., individuals from Kruger), supporting the expectation of clustering of individuals that are from the same geographic region, as reported by earlier studies that used nuclear DNA microsatellites (Ishida et al., 2011; Wasser et al., 2015). However, there are unresolved patterns (e.g., the positioning of the Mozambique elephants, Figure 3) that would need to be further explored through a comparison of Dataset 3 to continent-

wide genomic data, although beyond the scope of the current study with its primary focus on fecal DNA sample collection and analysis. The generation of genome-wide data from fecal samples may enable other broad-scale genomic questions to be addressed, for example, the evolutionary history of elephants and other species, and estimation of the age of inter-species hybridization zones (Tonzo et al., 2020; Bonnald et al., 2021).

In the PCA, the geo-referenced elephants from Gorongosa National Park in Mozambique were separated into three clusters when considering PC1 and PC2 (Figure 3). These elephants were part of a study that investigated tusklessness as an X chromosome–linked dominant, male-lethal trait (Campbell-Staton et al., 2021). The elephant population in Gorongosa has fluctuated drastically during the last 3 decades, with a >95% decrease in population size as a consequence of civil war (1977–1992), and limited recovery of the population since the end of the civil war (Stalmans et al., 2019; Campbell-Staton et al., 2021). In addition, the population was augmented through the translocation of six elephant bulls from Kruger National Park in 2008 (Stalmans et al., 2019), which may have affected clustering in the PCA. We investigated the possibility that the PCA clustering pattern of Gorongosa elephant genetic variation may be a consequence of small population size and associated inbreeding (Allendorf et al., 2013) by estimating genome-wide heterozygosity (GWH) for Gorongosa and other elephants in this study (except WPN01 that was excluded due to low genome coverage). The GWH was not associated with the clustering patterns observed for Gorongosa elephants (Supplementary Figure S7), nor by the tusk status of these elephants (Campbell-Staton et al., 2021). Future research that compares the nuclear genomes of these elephants to geo-referenced nuclear genomes from elephants across different regions in Africa on a continental scale, including higher representation of regions surrounding Mozambique and more individuals from Kruger National Park, might shed light on the observed pattern of genetic variation.

In addition to broad-scale conservation genomic questions (e.g., phylogeography and genetic structuring), the approaches used in this study may enable the study of rare or elusive species, or of species or populations where low animal density or small population size make traditional methods of sample collection (e.g., trapping/immobilization) difficult or impossible. For example, using the approach developed by de Flamingh et al. (2020), we estimated the genomic sex of the only remaining Knysna elephant in Wilderness National Park (WNP; Moolman et al., 2019a). This elephant which lives in the Afromontane Knysna Forest of South Africa is the sole survivor of a population that once comprised thousands of elephants, and there is ongoing discussion on management actions, e.g., reintroduction of other elephants (Patterson, 2012; Moolman et al., 2019a; Moolman et al., 2019b). Consistent with photographic evidence (Moolman et al., 2019a), we confirmed genetically that this individual is a

female. Stakeholders should take this into consideration when developing management plans. This individual does not group with other georeferenced elephant samples. This may represent a unique relic genomic signature, but this clustering pattern could also be driven by low genome coverage rather than genetic variation.

Unlike previous studies that sought to generate nuclear genome data from non-invasive samples (Perry et al., 2010; Chiou and Bergey, 2015; Snyder-Mackler et al., 2016; Taylor et al., 2021), our DNA extraction and genomic library construction approaches did not target host-specific DNA. Non-target enriched shotgun sequencing data have been used to reconstruct mitochondrial genomes (Bon et al., 2012; Srivathsan et al., 2019; Ang et al., 2020), but low endogenous/ target DNA content has so far precluded the generation of nuclear genome-wide datasets without targeting host-specific DNA (Perry et al., 2010; Chiou and Bergey, 2015; Snyder-Mackler et al., 2016; Taylor et al., 2021). To overcome low endogenous DNA in non-invasive samples, researchers have used DNA and RNA baits and methylation patterns to enrich genomic libraries for host species DNA prior to sequencing (Perry et al., 2010; Chiou and Bergey, 2015; Snyder-Mackler et al., 2016). Bait-based and other enrichment protocols are expensive, labor-intensive and time consuming, and may result in capture-biases where portions of the genome may be overrepresented in the sequencing pool after enrichment (George et al., 2011; Chiou and Bergey, 2015). The analyses used in this study avoided such obstacles by shotgun-sequencing total genomic DNA using standard DNA extraction and library construction protocols, and allowed us to also capitalize on the availability of concomitant sources of DNA that form part of the collected fecal sample. Sequencing host-associated DNA decreases the relative contribution of endogenous host-specific DNA when samples are not enriched, but conversely this allows for the study of complementary aspects that can inform the conservation and management of species. Because parasites and pathogens may be identified in fecal samples (Srivathsan et al., 2019), our combination of methods has the potential for monitoring individual and population health, which may be especially beneficial for rare, elusive, and endangered species whose health cannot be assessed using traditional approaches (e.g., blood sample screening). The ability to assess and monitor microbiome communities, and therefore individual and population health, may inform species conservation and management (Kophamel et al., 2022). For example, our metagenomic analysis of host-associated DNA showed that multiple samples contained contributions from Caudovirales, an order of bacteriophages that has been associated with gut health (Lepage et al., 2008; Minot et al., 2011; Norman et al., 2015). Fecal samples can also be used to examine intestinal parasite presence and abundance (e.g., Srivathsan et al., 2019), and fecal DNA analysis is an effective and reliable method for

studying parasite infections (da Silva et al., 1999; Blessmann et al., 2003; Yoshikawa et al., 2011). In our free-ranging elephants (Supplementary File 4) reads were identified as originating from the phylum *Platyhelminthes*, which includes flatworms that are predominantly parasitic (Park et al., 2007).

The combination of methods presented here may expand the application of genomic techniques to conservation science. However, researchers should consider the taxon and study system of interest when developing research projects as taxon-specific factors might impact study outcomes. For example, Zoelzer et al. (2021) show that fecal microbiome diversity is significantly higher in herbivores than in carnivores, whereas intraspecific microbiome variation is high in carnivores but not in herbivores. Considering taxon-specific factors (e.g., diet and ecological context of the environment in which organisms occur) can therefore help ensure the success of research initiatives. The metagenomic analysis that we applied in this study showcases the potential utility of the molecular and bioinformatic approaches used in this study for addressing complementary questions based on host-associated DNA that can also inform species conservation.

Data availability statement

The genomic data for the zoo and free-ranging elephants in the study are available on the NCBI Short-Read Archive (SRA) under bioproject number PRJNA910109, metagenomic result files for replicate sampling of zoo elephant dung are available on DRYAD (https://doi.org/10.5061/dryad.80gb5mktv). Bioinformatic code is available from https://github.com/adeflamingh/de_Flamingh_et_al_2022_FiG.

Ethics statement

Ethical review and approval was not required for the animal study because only non-invasive fecal DNA samples were used. Collection and export/import of elephant dung was sanctioned by appropriate authorities prior to collection, including South African National Parks, the Department of Agriculture, Forestry and Fisheries in South Africa and the United States Department of Agriculture in the United States of America.

Author contributions

AdF, RJvA, RSM and ALR conceptualized and designed this study; RJvA provided free-ranging elephant samples; HRS, PP and YI provided access to high quality geographically referenced data; AdF and SV carried out the molecular analysis of fecal samples; AdF and YI carried out bioinformatic and statistical analyses; AdF, RJvA, RSM and ALR interpreted the data and results; all authors contributed to and approved the final version of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1021004/full#supplementary-material.

References

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat. Rev. Genet.* 11, 697–709. doi:10.1038/nrg2844

Allendorf, F. W., Luikart, G., and Aitken, S. N. (2013). *Conservation and the genetics of populations. 2nd editio.* West Sussex, UK: Blackwell Publishing, Wiley and Sons.

Andrews, K. R., de Barba, M., Russello, M. A., and Waits, L. P. (2021). "Advances in using non-invasive, archival, and environmental samples for population genomic studies," in *Population genomics: Wildlife.* Editors P. A. Hohenlohe and O. P. Rajora (Cham: Springer International Publishing), 63–99. doi:10.1007/13836_2018_45

Ang, A., Roesma, D. I., Nijman, V., Meier, R., and Srivathsan, A. (2020). Faecal DNA to the rescue: Shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be Critically Endangered species. *Sci. Rep.* 10, 1–16. doi:10.1038/s41598-020-66007-8

Arnemo, J. M., Ahlqvist, P., Andersen, R., Berntsen, F., Ericsson, G., Odden, J., et al. (2006). Risk of capture-related mortality in large free-ranging mammals: Experiences from scandinavia. *Wildl. Biol.* 12, 109–113. doi:10.2981/0909-6396(2006)12[109:rocmil]2.0.co;2

Baker, C. S., Hamner, R. M., Cooke, J., Heimeier, D., Vant, M., Steel, D., et al. (2013). Low abundance and probable decline of the critically endangered Maui's dolphin estimated by genotype capture-recapture. *Anim. Conserv.* 16, 224–233. doi:10.1111/J.1469-1795.2012.00590.X

Balloux, F., Amos, W., and Coulson, T. (2004). Does heterozygosity estimate inbreeding in real populations? *Mol. Ecol.* 13, 3021–3031. doi:10.1111/j.1365-294X.2004.02318.x

Becciolini, V., Lanini, F., and Ponzetta, M. P. (2019). Impact of capture and chemical immobilization on the spatial behaviour of red deer *Cervus elaphus* hinds. *Wildl. Biol.* 2019, 1–8. doi:10.1111/2981/wlb.00499

Bellemain, E., and Taberlet, P. (2004). Improved noninvasive genotyping method: Application to Brown bear (*Ursus arctos*) faeces. *Mol. Ecol. Notes* 4, 519–522. doi:10.1111/j.1471-8286.2004.00711.x

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Blessmann, J., Ali, I. K. M., Ton Nu, P. A., Dinh, B. T., Ngo Viet, T. Q., le Van, A., et al. (2003). Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. *J. Clin. Microbiol.* 41, 4745–4750. doi:10.1128/jcm.41.10.4745-4750.2003

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bon, C., Berthonaud, V., Maksud, F., Labadie, K., Poulain, J., Artiguenave, F., et al. (2012). Coprolites as a source of information on the genome and diet of the cave hyena. *Proc. Biol. Sci.* 279, 2825–2830. doi:10.1098/rspb.2012.0358

Bonnald, J., Utge, J., Kuhner, M. K., Wasser, S. K., Asalu, E., Okimat, J. P., et al. (2021). Who are the elephants living in the hybridization zone? How genetics may guide conservation to better protect endangered elephants. *Glob. Ecol. Conservation* 32, e01917. doi:10.1016/j.gecco.2021.e01917

Bourgeois, S., Kaden, J., Senn, H., Bunnefeld, N., Jeffery, K. J., Akomo-Okoue, E. F., et al. (2019). Improving cost-efficiency of faecal genotyping: New tools for elephant species. *PLoS One* 14, e0210811. doi:10.1371/journal.pone.0210811

Brandt, A. L., Ishida, Y., Georgiadis, N. J., and Roca, A. L. (2012). Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. *Mol. Ecol.* 21, 1175–1189. doi:10.1111/J.1365-294X.2012.05461.X

Campbell-Staton, S. C., Arnold, B. J., Gonçalves, D., Granli, P., Poole, J., Long, R. A., et al. (2021). Ivory poaching and the rapid evolution of tusklessness in African elephants. *Sci. (1979)* 374, 483–487. doi:10.1126/science.abe7389

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., et al. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evol. Appl.* 11, 1094–1119. doi:10.1111/EVA.12600

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Chiou, K., and Bergey, C. (2015). *FecalSeq: Methylation-based enrichment for noninvasive population genomics from feces.* bioRxiv, 032870. doi:10.1101/032870

Core Team, R. (2013). *R: A language and environment for statistical computing.* Vienna: R Foundation for statistical computing.

da Silva, A. J., Bornay-Llinares, F. J., Moura, I. N. S., Slemenda, S. B., Tuttle, J. L., and Pieniazek, N. J. (1999). Fast and reliable extraction of protozoan parasite DNA from fecal specimens. *Mol. Diagn.* 4, 57–64. doi:10.1016/s1084-8592(99)80050-2

Daley, T., Helus, V., and Smith, A. (2014). *The preseq Manual.*

Daley, T., and Smith, A. D. (2014). Modeling genome coverage in single-cell sequencing. *Bioinformatics* 30, 3159–3165. doi:10.1093/bioinformatics/btu540

de Flamingh, A., Roca, A. L., and van Aarde, R. J. (2018). Origin and phylogeography of African savannah elephants (*Loxodonta africana*) in Kruger and nearby parks in southern Africa. *Conserv. Genet.* 19, 155–167. doi:10.1007/s10592-017-1005-z

de Flamingh, A., Coutu, A., Roca, A. L., and Malhi, R. S. (2020). Accurate sex identification of ancient elephant and other animal remains using low-coverage DNA shotgun sequencing data. *G3* 10, 1427–1432. LP – 1432. doi:10.1534/g3.119.400833

DelGiudice, G. D., Mangipane, B. A., Sampson, B. A., and Kochanny, C. O. (2001). Chemical immobilization, body temperature, and post-release mortality of white-tailed deer captured by Clover trap and net-gun. *Wildl. Soc. Bull.* 29, 1147–1157. doi:10.2307/3784138

Ferreira, C. M., Sabino-Marques, H., Barbosa, S., Costa, P., Encarnação, C., Alpizar-Jara, R., et al. (2018). Genetic non-invasive sampling (gNIS) as a cost-effective tool for monitoring elusive small mammals. *Eur. J. Wildl. Res.* 64, 46–15. doi:10.1007/s10344-018-1188-8

Franklin, T. W., McKelvey, K. S., Golding, J. D., Mason, D. H., Dysthe, J. C., Pilgrim, K. L., et al. (2019). Using environmental DNA methods to improve winter surveys for rare carnivores: DNA from snow and improved noninvasive techniques. *Biol. Conserv.* 229, 50–58. doi:10.1016/j.biocon.2018.11.006

George, R. D., McVicker, G., Diederich, R., Ng, S. B., MacKenzie, A. P., Swanson, W. J., et al. (2011). Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. *Genome Res.* 21, 1686–1694. doi:10.1101/gr.121327.111

Glocke, I., and Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* 27, 1230–1237. doi:10.1101/GR.219675.116

Gobush, K. S., Mutayoba, B. M., and Wasser, S. K. (2008). Long-term impacts of poaching on relatedness, stress physiology, and reproductive output of adult female African elephants. *Conserv. Biol.* 22, 1590–1599. doi:10.1111/j.1523-1739.2008.01035.x

Gray, T. N. E., Vidya, T. N. C., Maxwell, A. L., Bharti, D. K., Potdar, S., Channa, P., et al. (2011). *Using fecal-DNA and capture-mark-recapture to establish a baseline Asian elephant population for the Eastern Plains Landscape, Cambodia.* India: WWF Greater Mekong Programme-Cambodia Programme, Phnom Penh, Cambodia and Jawaharlal Nehru Center for Advanced Scientific Research.

Gray, T. N. E., Vidya, T. N. C., Potdar, S., Bharti, D. K., and Sovanna, P. (2014). Population size estimation of an Asian elephant population in eastern Cambodia through non-invasive mark-recapture sampling. *Conserv. Genet.* 15, 803–810. doi:10.1007/s10592-014-0579-y

Hansen, E. E., Lozupone, C. A., Rey, F. E., Wu, M., Guruge, J. L., Narra, A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4599–4606. doi:10.1073/pnas.1000071108

Hansson, B., and Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* 11, 2467–2474. doi:10.1046/j.1365-294x.2002.01644.x

Hohenlohe, P. A., Funk, W. C., and Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Mol. Ecol.* 30, 62–82. doi:10.1111/MEC.15720

Ishida, Y., Oleksyk, T. K., Georgiadis, N. J., David, V. A., Zhao, K., Stephens, R. M., et al. (2011). Reconciling apparent conflicts between mitochondrial and nuclear phylogenies in African elephants. *PLoS One* 6, e20642. doi:10.1371/journal.pone.0020642

Jacobson, E. R., Kollias, G. v., Heard, D. J., and Caligiuri, R. (1988). Immobilization of African elephants with carfentanil and antagonism with nalmefene and diprenorphine. *J. Zoo Animal Med.* 19, 1–7. doi:10.2307/20094842

Kock, M. D., Martin, R. B., and Kock, N. (1993). Chemical immobilization of free-ranging African elephants (*Loxodonta africana*) in Zimbabwe, using etorphine (M99) mixed with hyaluronidase, and evaluation of biological data collected soon after immobilization. *J. Zoo Wildl. Med.* 24, 1–10.

Kophamel, S., Illing, B., Ariel, E., Difalco, M., Skerratt, L. F., Hamann, M., et al. (2022). Importance of health assessments for conservation in noncaptive wildlife. *Conserv. Biol.* 36, e13724. doi:10.1111/COBI.13724

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). Angsd: Analysis of next generation sequencing data. *BMC Bioinforma.* 15, 1–13. doi:10.1186/s12859-014-0356-4

Korneliussen, T. S., and Moltke, I. (2015). NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* 31, 4009–4011. doi:10.1093/bioinformatics/btv509

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

Lefort, M.-C., Cruickshank, R. H., Descovich, K., Adams, N. J., Barun, A., Emami-Khoyi, A., et al. (2019). *Blood, sweat and tears: A review of non-invasive DNA sampling.* bioRxiv, 385120.

Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., and Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: A role for bacteriophages? *Gut* 57, 424–425. doi:10.1136/gut.2007.134668

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210, 719–731. doi:10.1534/genetics.118.301336

Meyer, M., Palkopoulou, E., Baleka, S., Stiller, M., Penkman, K. E. H., Alt, K. W., et al. (2017). Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. *Elife* 6, e25413. doi:10.7554/eLife.25413

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., et al. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625. doi:10.1101/gr.122705.111

Mondol, S., Karanth, K. U., Kumar, N. S., Gopalaswamy, A. M., Andheria, A., and Ramakrishnan, U. (2009). Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biol. Conserv.* 142, 2350–2360. doi:10.1016/j.biocon.2009.05.014

Moolman, L., de Morney, M. A., Ferreira, S. M., Ganswindt, A., Poole, J. H., and Kerley, G. I. H. (2019a). And then there was one: A camera trap survey of the declining population of african elephants in Knysna, South Africa. *Afr. J. Wildl. Res.* 49, 16–26. doi:10.3957/056.049.0016

Moolman, L., Ferreira, S. M., Gaylard, A., Zimmerman, D., and Kerley, G. I. H. (2019b). The decline of the Knysna elephants: Pattern and hypotheses. *S Afr. J. Sci.* 115, 1–7. doi:10.17159/sajs.2019/4805

Morrison, C. E., Johnson, R. N., Grueber, C. E., and Hogg, C. J. (2020). Genetic impacts of conservation management actions in a critically endangered parrot species. *Conserv. Genet.* 21, 869–877. doi:10.1007/s10592-020-01292-4

Munshi-South, J. (2011). Relatedness and demography of African forest elephants: Inferences from noninvasive fecal DNA analyses. *J. Hered.* 102, 391–398. doi:10.1093/jhered/esr030

Murphy, M. A., Waits, L. P., Kendall, K. C., Wasser, S. K., Higbee, J. A., and Bogden, R. (2002). An evaluation of long-term preservation methods for Brown bear (*Ursus arctos*) faecal DNA samples. *Conserv. Genet.* 3, 435–440. doi:10.1023/a:1020503330767

Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., et al. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell.* 160, 447–460. doi:10.1016/j.cell.2015.01.002

Nsubuga, A. M., Robbins, M. M., Roeder, A. D., Morin, P. A., Boesch, C., and Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Mol. Ecol.* 13, 2089–2094. doi:10.1111/j.1365-294X.2004.02207.x

Nyakaana, S., and Arctander, P. (1999). Population genetic structure of the african elephant in Uganda based on variation at mitochondrial and nuclear loci: Evidence for male-biased gene flow. *Mol. Ecol.* 8, 1105–1115. doi:10.1046/j.1365-294x.1999.00661.x

Okello, J. B. A., Wittemyer, G., Rasmussen, H. B., Douglas-Hamilton, I., Nyakaana, S., Arctander, P., et al. (2005). Noninvasive genotyping and Mendelian analysis of microsatellites in African savannah elephants. *J. Hered.* 96, 679–687. doi:10.1093/jhered/esi117

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., et al. (2018). A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2566–E2574. doi:10.1073/pnas.1720554115

Palomares, F., Godoy, J. A., Piriz, A., O'Brien, S. J., and Johnson, W. E. (2002). Faecal genetic analysis to determine the presence and distribution of elusive carnivores: Design and feasibility for the iberian lynx. *Mol. Ecol.* 11, 2171–2182. doi:10.1046/J.1365-294X.2002.01608.X

Park, J.-K., Kim, K.-H., Kang, S., Kim, W., Eom, K. S., and Littlewood, D. T. J. (2007). A common origin of complex life cycles in parasitic flatworms: Evidence from the complete mitochondrial genome of *Microcotyle sebastis* (monogenea: Platyhelminthes). *BMC Evol. Biol.* 7, 11–13. doi:10.1186/1471-2148-7-11

Patterson, G. (2012). *The secret elephants*. Penguin Random House South Africa.

Pelletier, F., Hogg, J. T., and Festa-Bianchet, M. (2004). Effect of chemical immobilization on social status of bighorn rams. *Anim. Behav.* 67, 1163–1165. doi:10.1016/j.anbehav.2003.07.009

Perry, G. H., Marioni, J. C., Melsted, P., and Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Mol. Ecol.* 19, 5332–5344. doi:10.1111/j.1365-294X.2010.04888.x

Picard Toolkit (2019). in *Broad Institute, G. Repository*. Available at: http://broadinstitute.github.io/picard/ (B. I. No Title.*Picard Toolkit*

Régnier, C., Gargominy, O., Falkner, G., and Puillandre, N. (2011). Foot mucus stored on FTA® cards is a reliable and non-invasive source of DNA for genetics studies in molluscs. *Conserv. Genet. Resour.* 3, 377–382. doi:10.1007/s12686-010-9345-8

Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* 13, 2447–2461. doi:10.1038/s41596-018-0050-5

Sappington, T. W. (2019). Words matter: Extensive misapplication of" non-invasive" in describing DNA sampling methods, and proposed clarifying terms. *Peer Community Ecol.* 2, 1–19. doi:10.24072/pci.ecology.100029

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., et al. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics* 203, 699–714. doi:10.1534/genetics.116.187492

Song, S. J., Amir, A., Metcalf, J. L., Amato, K. R., Xu, Z. Z., Humphrey, G., et al. (2016). Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems* 1, 1–12. doi:10.1128/mSystems.00021-16

Srivathsan, A., Nagarajan, N., and Meier, R. (2019). Boosting natural history research via metagenomic clean-up of crowdsourced feces. *PLoS Biol.* 17, e3000517. doi:10.1371/journal.pbio.3000517

Stalmans, M. E., Massad, T. J., Peel, M. J. S., Tarnita, C. E., and Pringle, R. M. (2019). War-induced collapse and asymmetric recovery of large-mammal populations in Gorongosa National Park, Mozambique. *PLoS One* 14, e0212864. doi:10.1371/journal.pone.0212864

Supple, M. A., and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biol.* 19, 131. doi:10.1186/S13059-018-1520-3

Taberlet, P., Waits, L. P., and Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends Ecol. Evol.* 14, 323–327. doi:10.1016/s0169-5347(99)01637-7

Taylor, R. S., Manseau, M., Redquest, B., Keobouasone, S., Gagné, P., Martineau, C., et al. (2021). Whole genome sequences from non-invasively collected caribou faecal samples. *Conserv. Genet. Resour.* 14, 1–16. doi:10.1007/S12686-021-01235-2/FIGURES/6

Thitaram, C., Thongtip, N., Somgird, C., Colenbrander, B., van Boxtel, D. C. J., van Steenbeek, F., et al. (2008). Evaluation and selection of microsatellite markers for an identification and parentage test of Asian elephants (*Elephas maximus*). *Conserv. Genet.* 9, 921–925. doi:10.1007/s10592-007-9406-z

Tonzo, V., Papadopoulou, A., and Ortego, J. (2020). Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Mol. Ecol.* 29, 2254–2268. doi:10.1111/MEC.15475

Vidya, T. N. C., and Sukumar, R. (2005). Amplification success and feasibility of using microsatellite loci amplified from dung to population genetic studies of the Asian elephant (*Elephas maximus*). *Curr. Sci.* 88, 489–492.

Ward, A., Hide, G., and Jehle, R. (2019). Skin swabs with FTA® cards as a dry storage source for amphibian DNA. *Conserv. Genet. Resour.* 11, 309–311. doi:10.1007/s12686-018-1018-z

Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B., and Stephens, M. (2004). Assigning african elephant DNA to geographic region of origin: Applications to the ivory trade. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14847–14852. doi:10.1073/pnas.0403170101

Wasser, S. K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C., et al. (2015). CONSERVATION. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. *Sci. (1979)* 349, 84–87. doi:10.1126/science.aaa2457

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 1–13. doi:10.1186/s13059-019-1891-0

Xavier, C., Eduardoff, M., Bertoglio, B., Amory, C., Berger, C., Casas-Vargas, A., et al. (2021). Evaluation of DNA extraction methods developed for forensic and ancient DNA applications using bone samples of different age. *Genes.* 12, 146. doi:10.3390/GENES12020146

Yao, L., Witt, K., Li, H., Rice, J., Salinas, N. R., Martin, R. D., et al. (2020). Population genetics of wild *Macaca fascicularis* with low-coverage shotgun sequencing of museum specimens. *Am. J. Phys. Anthropol.* 173, 21–33. doi:10.1002/ajpa.24099

Yarlagadda, K., Zachwieja, A. J., de Flamingh, A., Phungviwatnikul, T., Rivera-Colón, A. G., Roseman, C., et al. (2022). *Geographically diverse canid sampling provides novel insights into pre-industrial microbiomes.* Biology: Proceedings of the Royal Society.

Yoshikawa, H., Dogruman-Ai, F., Turk, S., Kustimur, S., Balaban, N., Sultan, N., et al. (2011). Evaluation of DNA extraction kits for molecular diagnosis of human Blastocystis subtypes from fecal samples. *Parasitol. Res.* 109, 1045–1050. doi:10.1007/s00436-011-2342-3

Zoelzer, F., Burger, A. L., and Dierkes, P. W. (2021). Unraveling differences in fecal microbiota stability in mammals: From high variable carnivores and consistently stable herbivores. *Anim. Microbiome* 3, 77–16. doi:10.1186/s42523-021-00141-0