MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1966–1979 doi: 10.1111/j.1365-294X.2011.05261.x

Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity

KENNETH ANDERSEN,* KAREN LISE BIRD,† MORTEN RASMUSSEN,* JAMES HAILE,‡ HENRIK BREUNING-MADSEN,† KURT H. KJÆR,* LUDOVIC ORLANDO,* M. THOMAS P. GILBERT* and ESKE WILLERSLEV*

**Centre for GeoGenetics, University of Copenhagen, Øster Voldgade 5-7, DK-1350, Denmark, †Institute of Geography and Geology, University of Copenhagen, Øster Voldgade 10, DK-1350, Denmark, ‡Ancient DNA Research Laboratory, Murdoch University, South Street, WA 6150, Australia*

Abstract

DNA molecules originating from animals and plants can be retrieved directly from sediments and have been used for reconstructing both contemporary and past ecosystems. However, the extent to which such 'dirt' DNA reflects taxonomic richness and structural diversity remains contentious. Here, we couple second generation high-throughput sequencing with 16S mitochondrial DNA (mtDNA) meta-barcoding, to explore the accuracy and sensitivity of 'dirt' DNA as an indicator of vertebrate diversity, from soil sampled at safari parks, zoological gardens and farms with known species compositions. PCR amplification was successful in the full pH range of the investigated soils (6.2 ± 0.2 to 8.3 ± 0.2), but inhibition was detected in extracts from soil of high organic content. DNA movement (leaching) through strata was evident in some sporadic cases and is influenced by soil texture and structure. We find that DNA from the soil surface reflects overall taxonomic richness and relative biomass of individual species. However, one species that was recently introduced was not detected. Furthermore, animal behaviour was shown to influence DNA deposition rates. The approach potentially provides a quick methodological alternative to classical ecological surveys of biodiversity, and most reliable results are obtained with spatial sample replicates, while relative amounts of soil processed per site is of less importance.

Keywords: meta-barcoding, 'dirt' DNA, environmental samples, DNA deposition, DNA leaching

Received 5 January 2011; revision received 19 July 2011; accepted 27 July 2011

Introduction

DNA preserved in no more than a few grams of sediments has been shown to yield significant information about the biodiversity of past ecosystems (Willerslev *et al.* 2003), providing access to the composition of, for example, plant, mammal, bird, fungus, bacteria and insect communities, even in the absence of macrofossils (Willerslev *et al.* 2003, 2007; Lydolph *et al.* 2005; Thomsen *et al.* 2009). Although first studied intensively using DNA from permafrost representing ancient environments, the approach has also been applied successfully to sediments from non-frozen sites (Hofreiter *et al.* 2003; Willerslev *et al.* 2003; Haile *et al.* 2007; Hebsgaard *et al.* 2009)—notably much younger in age and not without possible problems related to DNA leaching through strata (Haile *et al.* 2007). Recently, the approach has been used for assessments of fungal biodiversity from surface soil (Geml *et al.* 2009) and to detect amphibian and fish species in freshwater environments (Ficetola *et al.* 2008; Jerde *et al.* 2011). While the exact sources of this ancient and modern 'dirt' DNA remain largely unknown, faeces, hair, skin flakes and urine from mammals are hypothesised as main contributors (Willerslev *et al.* 2003; Lydolph *et al.* 2005; Haile *et al.* 2007).

With the introduction of second generation sequencing platforms coupled to 5′-tagging (Binladen *et al.*

Correspondence: Eske Willerslev,
E-mail: ewillerslev@snm.ku.dk

2007) of generic primers, it is possible to sequence DNA amplicons at unprecedented depth, out-competing traditional cloning/sequencing approaches both with regards to data generation and thus analysis, with possible detection of rare sequences and quantification of relative sequence abundances (Binladen *et al.* 2007; Deagle *et al.* 2009, 2010; Pegard *et al.* 2009; Rasmussen *et al.* 2009; Soininen *et al.* 2009; Valentini *et al.* 2009, 2010; Bohmann *et al.* 2011; Kowalczyk *et al.* 2011). Such 'dirt' DNA meta-barcoding approaches (Valentini *et al.* 2008) have successfully been used to time the last occurrence dates of extinct species with woolly mammoth (*Mammuthus primigenius*) and horse (*Equus caballus*) surviving in mainland Alaska at least 3500 years longer than previously thought, thereby questioning common theories on causes of Ice Age mammalian extinctions (Haile *et al.* 2009).

Despite the power of the analysing 'dirt' DNA, a number of fundamental assumptions remain untested, including to what extent 'dirt' DNA qualitatively and quantitatively reflects plant and animal biodiversity. This is particular interesting for mammals that have conceivable differences in biomass, mobility and behaviour characteristics, all of which may introduce major sources of bias in the results. Similarly, while DNA leaching through strata has not been observed in frozen sediments (Willerslev *et al.* 2004, 2007; Hansen *et al.* 2006; Arnold *et al.* 2011), the extent to which leaching affects biodiversity estimates under non-frozen conditions remains unclear, with different studies showing contradictory findings (Haile *et al.* 2007; Gilbert *et al.* 2008a; Hebsgaard *et al.* 2009; Rasmussen *et al.* 2009).

Here, we investigate the potential for profiling the diversity of vertebrate communities using 'dirt' DNA meta-barcoding based on high-throughput sequencing of tagged PCR amplicons. Soil from safari parks, ostrich farms and zoological gardens were selected due to detailed records of the presence of exotic species, offering the unique opportunity to monitor both method accuracy and sensitivity. Additionally, we explored different sediment types for their susceptibility to DNA leaching, and we define general guidelines as to when leaching is likely to affect diversity estimates of 'dirt' DNA from ancient sediments.

Materials and methods

Sampling strategies and study sites

Six locations were specifically selected due to the presence of animals that are not part of the contemporary Danish fauna (Fig. 1); thus any DNA match to these species can only originate from the animals actually observable (or historically present) at the study sites (Table 1).

Two different sampling approaches were applied at these six sites (Fig. 1a). The first was designed in order to test if vertical movement of DNA molecules takes place through soil profiles (Site A–C). For this purpose, soil profiles were excavated at three sites (the elephant enclosure: Site A1; the tiger enclosure: Site B1–B2; and the ostrich enclosure: Site C1). The second sampling approach was designed to investigate the spatial distribution and deposition rates of dirt DNA from vertebrate species present at the study sites (D–F). For this purpose, a quadrat soil sampling approach was applied at three sites (the savanna enclosure: Site D and Site E; and the lion enclosure: Site F), as demonstrated in Fig. 1b–c. Soil was sampled from a regular dispersed

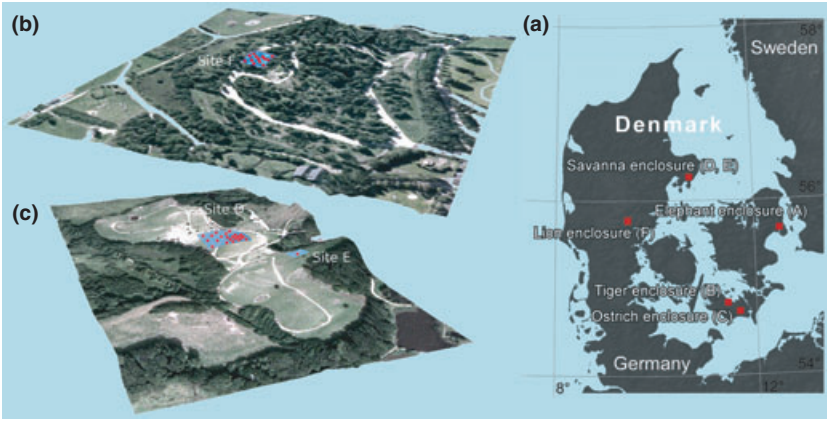

Fig. 1 (a) Location of the study sites A–F; elephant enclosure (Site A); tiger enclosure (Site B); ostrich enclosure (Site C), savanna enclosure (Site D–E), lion enclosure (Site F). (b) Position of quadrats (blue) and individual subsamples (red) of the 20 × 20 m quadrat at Site F sampled in the 1 ha lion enclosure, and (c) of the 20 × 20 m and 40 × 40 m quadrats at Site D, and the 10 × 10 m quadrat at Site E, sampled in the 5 ha savanna enclosure.

Table 1 Overview of sampling locations and procedures, the management history and introduction period of species present at the study sites

Code	Sampling procedures	mtDNA region	Species currently/historically present in the enclosures	Introduction period
Copenhagen Zoo, elephant enclosure				
A1	Sampling depth (cm): 10-30-50-70	16S (68–71 bp)	Asian elephant (*E. maximus*)†	1914–Sep. 2009
	Sample size (g): 6.5 g each	CR (165 bp)	African elephant (*L. africana*)†	1930–1970
Knuthenborg Park and Safari, tiger enclosure				
B1 & B2	Sampling depth (cm): 0-10-20-30-40	16S (68–71 bp)	Tiger (*P. tigris*)	1969–2009
	Sample size (g): 6.5 g each	D-loop (76 bp)		
		Cytb (124 bp)		
Døllefjelde ostrich farm				
C1	Sampling depth (cm): 0-2-4-6-8-60	D-loop (78 bp)	Ostrich (*S. camelus*)	1997–2009
	Sample size (g): 6.5 g each			
Ree Park, savanna enclosure (5 ha)				
D1	*Quadrats sampled at the surface:*	16S (68–71 bp)	Blackbuck (*A. cervicapra*)	2003–2009
	0.5 g aliquoted from 13 samples (13/400 m^2)	D-loop (78 bp)	Blesbuck (*D. pygargus*)	2003–2009
D2	1.3 g aliquoted from five samples (5/400 m^2)		Blue wildebeest (*C. taurinus*)	2003–2009
D3	6.5 g from five individual samples (5/400 m^2)		Common eland (*T. oryx*)	2003–2009
D4	6.5 g from 13 individual samples (13/400 m^2)		Giraffe (*G. camelopardalis*)	Jan.–Mar. 2009
D5	6.5 g from 13 individual samples (13/1600 m^2)		Ostrich (*S. camelus*)	2003–2009
	Quadrats sampled in 10 cm depth:		Watusi cow (*B. taurus taurus watusi*)	2003–2009
D6	0.5 g aliquoted from 13 samples (13/400 m^2)		Zebra (*E. burchellii*)	2003–2009
D7	1.3 g aliquoted from five samples (5/400 m^2)			
D8	6.5 g from five individual samples (5/400 m^2)			
Ree Park, savanna enclosure (5 ha)				
E1	*Quadrats sampled in 10 cm depth:*	16S (68–71 bp)	Blackbuck (*A. cervicapra*)	2003–Jan. 2009
	5 samples only used for soil chemical analysis		Blesbuck (*D. pygargus*)	2003–Jan. 2009
E2	0.5 g aliquoted from 13 samples (13/100 m^2)		Blue wildebeest (*C. taurinus*)	2003–Jan. 2009
	–		Common eland (*T. oryx*)	2003–Jan. 2009
	–		Ostrich (*S. camelus*)	2003–Jan. 2009
	–		Watusi cow (*B. taurus taurus watusi*)	2003–Jan. 2009
	–		Zebra (*E. burchellii*)	2003–Jan. 2009
	–		Bactrian camel (*C. bactrianus*)†	–2003
	–		Chital (*Cervus axis*)†	–2003
	–		Kulan (*Equus hemionus kulan*)†	–2003
	–		Onager (*Equus hemionus onager*)†	–2003
Givskud Zoo, lion enclosure (1 ha)				
F1	*Quadrat sampled at the surface*	16S (68–71 bp)	Lion (*P. leo*)	1969–2009
	0.5 g aliquoted from 13 samples (13/400 m^2)			
	Quadrat sampled in 10 cm depth			
F2	0.5 g aliquoted from 13 samples (13/400 m^2)			

†Species not present at the time of sampling.

point pattern over an area of 10 × 10 m (Site E), 20 × 20 m (Site D and Site F) or 40 × 40 m quadrats (Site D). From each quadrat plot, a number of 13 soil samples were collected, both at the soil surface and in 10 cm depth.

Soil samples for DNA analysis were collected in DNA-free 50 mL falcon tubes, and kept refrigerated for 1–2 days before long-term storage at –20 °C. When sampling in soil profiles, all samples collected for soil particle and chemical analyses were immediately adjacent to the soil samples collected for DNA analysis. From two quadrats sampled in the savanna enclosure (Site D–E), five subsurface samples (*ca.* 10 cm depth) were collected immediately adjacent to the soil samples collected for DNA analysis (Fig. 2).

'Dirt' DNA extraction and amplification

Soil DNA extraction and PCR setup was undertaken in laboratory facilities dedicated to ancient DNA (aDNA) studies, following established procedures (Willerslev & Cooper, 2005). DNA extraction was performed on 6.5 g of soil, using the PowerMax™ Soil DNA isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA). From the samples the wet-weight-soil (Table 1) was subjected to DNA extractions, following established protocols for

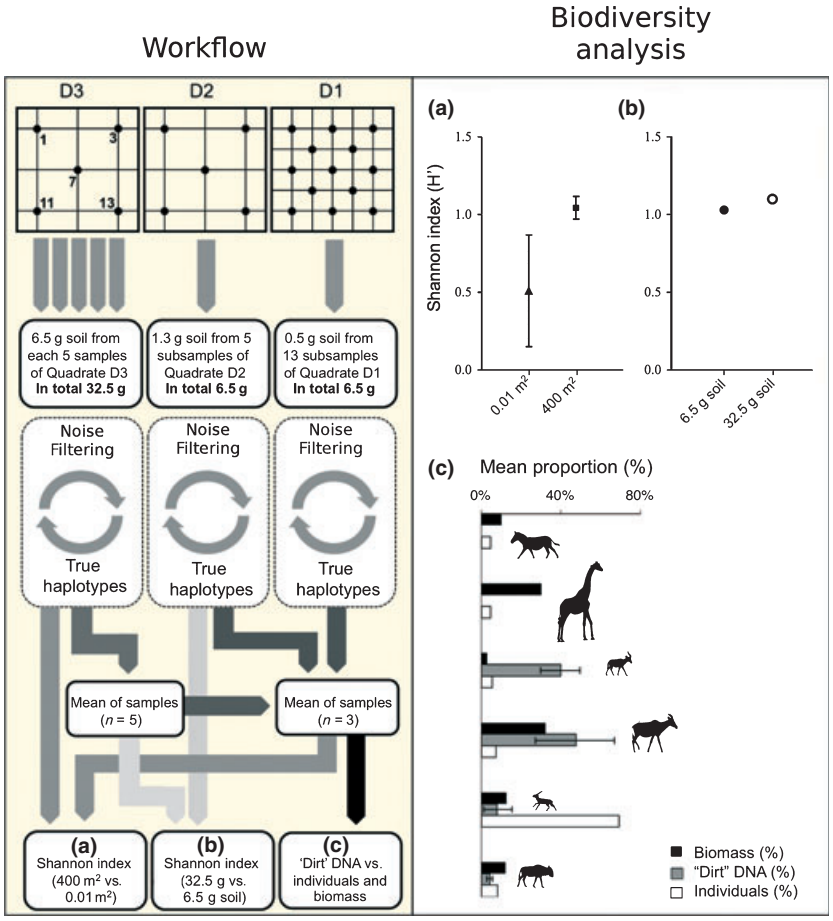

Fig. 2 *Workflow*: the position of samples within the 20 × 20 m quadrat (D1–D3); the number of subsamples; and amount of soil processed. Post-sequencing, filtering was performed to remove sequencing noise. The remaining true haplotypes formed the basis for calculating the proportional distribution of haplotypes within the samples. *Biodiversity analysis:* In (a) the mean Shannon–Weaver biodiversity index are shown (±1 SD) for five individually extracted samples (▲) and of the three sampling replicates (■) of the 400 m^2 area. (b) The Shannon–Weaver biodiversity index, based on DNA extraction from 1.3 g soil from five subsamples in total 6.5 g (●), and based on DNA extraction from 6.5 g soil from each subsample in total 32.5 g (○). In (c) the proportions of 'dirt' DNA from individual species is calculated as the mean (±2 SD) of the three replicates of the 400 m^2 quadrat plot; proportions of total biomass and number of individuals of the species present at Site D. From the top: zebra, giraffe, blesbuck, common eland, blackbuck and blue wildebeest. The position of the five samples collected in 10 cm depth for soil chemical analysis is similar to the five samples in quadrat D3.

sediment ancient DNA (sedaDNA) (Haile *et al.* 2009). One extraction control was included for every eight extractions.

Mammalian DNA was screened in the extracts using generic primers (Table 2), '16S A&M Fv2 69' and '16S A&M Rv2 short' (Rasmussen *et al.* 2009), to PCR amplify a short fragment of the gene encoding the 16S ribosomal RNA, whose length is species specific which ranges between 28 and 31 bp (68–71 bp including primers). These primers were made compatible for multiplexed parallel deep-sequencing by addition of an 8 bp specific nucleotide tag at the 5' end of the primer (Binladen *et al.* 2007). Furthermore, the primers were modified into Roche 'Fusion' primers that contain the FLX adaptors A and B. Human blocking probes were used (Rasmussen *et al.* 2009) to inhibit the PCR amplification of human DNA that might have been present in the extract. Each 25 μL PCR reaction contained 5 μL DNA extract, 1.0U Platinum® *Taq* High Fidelity DNA Polymerase (Invitrogen, Carlsbad, CA, USA), 1X High Fidelity PCR buffer, 2 mM Mg_2SO_4, 0.2 mM of each dNTP, 0.4 μM of each primer, 4 μM of each blocker '16S A&M Human block 1' and '16S A&M Human block 2' (Rasmussen *et al.* 2009) (Table 2) and 2 μg bovine serum albumin (BSA). From samples of the soil profiles, four positive PCR reactions were obtained from each DNA extract, and from the quadrat samples two positive PCR reactions were obtained from each DNA extract.

Table 2 Primers and blocking probes used (5′ fusion sequence and barcodes are not shown)

Primers	mtDNA	Temp. (°C)	Sequence 5′–3′	Length (bp)	References
16S A&M Fv2 69	16S	52	CCCCGAAACCAGACGAGCTA	68–71	Rasmussen *et al.* (2009)
16S A&M Rv2 short	16S	52	TCACTATTTTGCNACATAGA	68–71	Rasmussen *et al.* (2009)
16S A&M Human block 1	16S	–	GAGCTACCTAAGAACAGCTA	–	Rasmussen *et al.* (2009)
16S A&M Human block 2	16S	–	TTTGCTACATAGACGGGTGT	–	Rasmussen *et al.* (2009)
16S fv2 long	16S	55	ACCTTTTGCATAATGAATTAAC	180–184	This study
16S rv2 long	16S	55	GGACAACCAGCTATCACCA	180–184	This study
Panthera F	d-loop	58	TCCAATCCTTCAACTTTCTCA	76	This study
Panthera R5	d-loop	58	ACAGTTATGTGTGATCATGGGC	76	This study
Struds F5	d-loop	60	CGCTAGTTTCATGACCATTC	78	This study
Struds R5	d-loop	60	CCCTGACTTAGGAACCAGT	78	This study
CR13u	Cytb	60	AAATTCTCACCGGCCTCTTTCTA	124	Burger *et al.* (2004)
CR131	Cytb	60	TTGGCGTGTAGGTACCGGATAA	124	Burger *et al.* (2004)
Mammoth CRR1	CR	56	TGAGAAATCTCTAGTCATCATG	165	Haile *et al.* (2009)
Mammoth CRF1	CR	56	CATGCTTATAAGCAAGTACTGT	165	Haile *et al.* (2009)

One PCR blank was included every eight reactions. PCR conditions consisted of initial 4 min incubation at 94 °C, followed by 55 cycles of 30 s at 94 °C, 40 s at 52 °C and 40 s at 68 °C, followed by a final extension step for 10 min of 72 °C. In addition to the 16S generic primers, a number of specific primers were used on samples from the six sites (Table 1). PCR setup and thermal programs, applied for these primer sets were similar as described above, with annealing temperatures as indicated in Table 2.

DNA sequencing

The tagged PCR products were sequenced on a Roche FLX platform (Roche-454, Branford, CT, USA), in two different sequencing runs. The ten nucleotide-tags were used twice in the first setup (samples from Site D, E and F), but run in separate lanes on the PicoTiterPlate during sequencing. In the second setup, the 10 tags were used once (samples from Profiles A1 and B2).

To minimise the sequencing of non-specific PCR amplicons, PCR products were gel purified prior to sequencing. The PCR products derived from each extract were initially pooled, and mixed with loading dye in proportions 1:5, heated to 70 °C for 5 min, placed on ice until cool, then run on a 3% gel. During UV visualisation, the correct size of the amplified products was determined and a gel slice containing DNA fragments of the expected length was cut with a sterile scalpel. DNA was further extracted from gel slices using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA, USA), following the manufacturer's instructions. The DNA concentration of each sample was subsequently assayed using quantitative real-time PCR (qPCR) against a known molecular standard, following Meyer *et al.* (2008). Lastly all uniquely labelled PCR products were pooled at equimolar concentrations and sequenced.

Analysis of sequence diversity

The sequencing error rate was estimated following the guidelines in Huse *et al.* (2007), and was used to calculate expected error distributions, which approximates a binominal distribution (Supporting information). Our probabilistic model assumes a conservative 0.45% per base error rate for the GS-FLX sequencing system (Supporting information), and was used as a threshold to delimit boundaries of sequence clusters. We discarded sequences with lower than expected frequencies in the interval k mismatched ($k \leq 4$) to the locally most frequent haplotype within a sequence cluster. Starting with the globally most frequent haplotype, the true number of sequence clusters within a sample was estimated, by consecutively identifying the locally most frequent haplotype outside the expected error distribution (Supporting information). Similar approaches have recently been reported to remove noise generated during regular PCR, emulsion PCR and sequencing-by-synthesis, hereby reducing the number of haplotypes to a level actually present in constructed samples (Zagordi *et al.* 2010; Quince *et al.* 2011). The remaining authentic haplotypes were globally aligned against an in-house 16S mtDNA reference database, containing 780 mammal sequences downloaded from GenBank or generated for this purpose (Supporting information). Subsequently, the identified haplotypes were assigned to the taxonomic level of species, genus or family, depending on similarity to reference sequences with informative polymorphisms, which they fully match in length and sequence.

From samples of the soil profiles (A1 and B2) the proportions of elephant and tiger haplotypes within

samples were calculated using the contamination with cow (*Bos taurus*) as an internal standard (Supporting information), which was present at a consistent level in our results, and likely derives from the BSA (Taylor 1996) that was used in similar concentration in all PCRs.

From the quadrats sampled in the savanna enclosure (Site D–E) and the lion enclosure (Site F), the proportions within samples were calculated based on frequencies of the identified haplotypes. These calculations were performed after excluding sequences assigned to taxa commonly observed as contaminants in reagents and consumables (*Homo sapiens, B. taurus, Sus scrofa, Capra aegagrus, Ovis aries* and *Mus spp.*) (Leonard *et al.* 2007; Haile *et al.* 2009). Additionally, for the samples from the savanna enclosure (Site D), the mean of five individual subsamples, and the mean of three replicates of the 20 × 20 m quadrat were calculated (Fig. 2). The obtained qualitative and quantitative sequence data was compared to the known management history and population parameters of the animal stock. Furthermore, the diversity of haplotypes identified in the soil surface at the savanna enclosure (Site D) was calculated, using the Shannon–Weaver biodiversity index (H'), both for individual subsamples and for the three replicates of the 20 × 20 m quadrat plot.

Soil analysis

The soil samples were all dried at 70 °C overnight and sieved through a 2 mm sieve, discarding the fraction > 2 mm. For profiles A1, B1 and B2, the pH(H_2O), particle size distribution, soil exchangeable Mg^{2+}, Ca^{2+}, K^+ and Na^+ ions and soil organic matter (SOM) were determined. Similar analyses, except particle size distribution, were performed for five soil samples from the subsurface layer of both plots sampled in the savanna enclosure, Site D and E. From the particle size distribution the surface area per particle volume (As) was estimated, following Segal *et al.* (2009) (Supporting information).

Results

Short mtDNA fragments originating from mammals present at the study sites could be PCR amplified from DNA extracted from surface soil, and in a few cases in the subsurface layers, using generic mammalian primers and subsequent high-throughput sequencing of the amplicons. The initial filtering of sequences, within the expected error distribution of more abundant and highly similar sequences, was found to reduce the number of haplotypes to a level close to the species richness actually observable at the sites. This effect has also recently been reported, based on sequencing of constructed samples, using identical sequencing technology (Quince *et al.* 2011).

The processed data was used to investigate four key questions: (i) how does the taxonomic richness of DNA sequences relate to the species present at the study sites? (ii) how does the spatial variation of 'dirt' DNA diversity in the soil surface relate to the populations present at the site? (iii) does the DNA from the surface leach down into the soil? (iv) does soil chemistry, texture and structure affect the extent of DNA leaching and DNA recovery efficiency?

Detection and quantification of taxonomic richness from 'dirt' DNA

The identified haplotypes were in most cases assigned with an exact match to the reference sequence database. Exceptions include zebra (*Equus burchellii*), blue wildebeest (*Connochaetes taurinus*) and common eland (*Taurotragus oryx*). In these cases sequence identification was inferred by similarity to the nearest relative for which a reference sequence is available, namely horse (exact match), black wildebeest (*Connochaetes gnou*: exact match) and bongo (*Tragelaphus eurycerus*: one substitution difference). All *Bovidae* genera that were identified at Site D and E are well-defined *molecular operational taxonomic units* (MOTUs) in the sense that they are all unique and distinguishable from other genera represented in the reference database, with only one haplotype identified per genus (Valentini *et al.* 2008).

In samples from the elephant enclosure (A1) a total of seven distinct haplotypes were identified as authentic using the performed filtering. Of these haplotypes four matched exactly to published sequences of *Elephas maximus* (two haplotypes), *Loxodonta africana* and *M. primigenius* (one shared haplotype) and *Mammut americanum* (one unique haplotype). Furthermore, three unknown haplotypes with high similarity to extinct and extant diversity (2–3 substitutions) was detected. This excessive level of diversity is likely due to artificially generated diversity (Taylor 1996) or unreported mtDNA diversity in contemporary elephants (see below). The seven identified elephant haplotypes each have a characteristic error rate (mean: 0.28% per base) and the proportions of erroneous sequences showing k mismatches ($k \leq 4$) to these haplotypes approximate the expected error distribution (Supporting information). These sequences were collapsed using the Elephantidae family as a cut-off taxonomic level. By deep-sequencing of samples from the tiger enclosure (B2) we only detected one true tiger (*Panthera tigris*) haplotype that also has characteristic error rates (mean: 0.16% per base).

The proportions of 'dirt' DNA within samples are shown for the soil profiles (A–C) in Table 3, and for the

Table 3 Soil profiles from sites A–C: Soil parameters; soil surface area per particle volume (A_s), SOM, pH and exchangeable cations ($cmol_C(+)/kg$; composed of Mg^{2+}, Ca^{2+}, K^+ and Na^+); Proportions (%) of haplotypes, inferred from high-throughput sequencing of 'dirt' DNA; number of assigned reads for each sample and presence/absence data for ostrich[(1)] and tiger[(2)]. (+) detected, (–) not detected, * Not determined

Code	Sample	As (cm^2/cm^3)	SOM (mass%)	pH (H_2O) (±0.2)	$cmol_C$ (+)/kg	Ostrich D-loop[(1)]	Elephant 16S	Tiger 16S/ d-loop[(2)]	Cow 16S	Reads
A1	*Elephant enclosure*									
	10 cm	6260	4.2	8.3	44.1		84.0		16.0	12 573
	30 cm	7149	*	*	*		91.1	0.1	8.8	887
	50 cm	3643	2.3	8.2	38.3		30.0		70.0	1195
	70 cm	357	*	*	*		11.5		88.5	157
B1	*Tiger enclosure, below trail*									
	0 cm	13 467	*	*	*			(+)		
	10 cm	17 131	*	*	*			(–)		
	20 cm	18 073	*	*	*			(–)		
	30 cm	*	*	*	*			(–)		
	40 cm	*	*	*	*			(–)		
B2	*Tiger enclosure, below latrine*									
	0 cm	*	10.7	6.5	25.5			98.1	1.9	2513
	10 cm	15 742	8.3	6.4	31.1		0.1	82.0	17.9	6976
	20 cm	15 535	5.6	6.2	19.7			54.6	45.2	2042
	30 cm	19 475	4.2	6.0	15.8			0.0	100	129
	40 cm	19 574	1.2	7.4	15.8			0.0	100	554
C1	*Ostrich farm*									
	0–2 cm	*	*	*	*	(+)				
	2–4 cm	*	*	*	*	(–)				
	4–6 cm	*	*	*	*	(–)				
	6–8 cm	*	*	*	*	(–)				
	8–10 cm	*	*	*	*	(–)				
	10–12 cm	*	*	*	*	(–)				
	60 cm	*	*	*	*	(–)				

spatial samples (D–F) in Table 4. To overcome inhibition by soil compounds, two samples of surface soil from Site B were diluted 1:10 before PCR amplification. Frequency distributions of the sequence diversity present in these samples were corrected accordingly before the fraction of 'dirt' DNA could be estimated.

Spatial variation of 'dirt' DNA in the soil surface

We used a quadrat sampling approach to evaluate the spatial variation of 'dirt' DNA deposition in the soil surface and to investigate how much the results are dependent on the quantity of soil processed (Fig. 2). The Shannon-Weaver index (±1SD) was calculated, based on within sample haplotype proportions of the amplified 16S mammalian sequences, for five individual samples from the savanna enclosure (Site D) (value, 0.52 ± 0.35), and based on the three replicates of the 400 m^2 quadrat (value, 1.05 ± 0.07). We find that the haplotype richness increases when spatial replicates are available (Fig. 2a) and that the range in diversity estimates is lower (0.98–1.13) compared to the single sample extractions (range, 0.15–1.01). Furthermore, we find that DNA extractions from the same five samples performed on 32.5 and 6.5 g of soil, respectively, produce similar results (Fig. 2b). The average proportion of 'dirt' DNA from individual haplotypes, recovered by the three replicates of the 400 m^2 quadrat in the savanna enclosure (Site D), was contrasted to number of individuals and relative amount of biomass present at the site (Fig. 2c).

Furthermore, we explored the relationship between quadrat size and the likelihood of detecting a species, using quadrat sizes of 20 × 20 and 40 × 40 m. Positive amplification of a 78 bp d-loop mtDNA fragment was obtained, using ostrich (*Struthio camelus*) specific primers, from the soil surface in the savanna enclosure, at Site D (Table 4). However, ostrich was infrequently detected (Fig. S2b, Supporting information), as only 7/13 (54%) samples were PCR positive from the 1600 m^2 quadrat (D5), in contrast to 4/13 (31%) from the 400 m^2 quadrat (D4).

Detection of 'dirt' DNA below the soil surface

'Dirt' DNA, of 68–71 bp length, was detected below 10 cm depth at two sites. In the elephant enclosure

Table 4 Quadrat plots at Site D–F: Proportions (%) of haplotypes; presence/absence data for ostrich[(1)] (+) detected, (–) not detected; Shannon-Weaver biodiversity index (H') (±1SD); per species proportion of total individuals (%) and biomass (%) in the savanna enclosure, as the species holding at the beginning of 2009. Soil parameters: the mean of five samples (±2SD) from Site D and E (SOM, pH and exchangeable cations (cmolC(+)/kg; composed of Mg^{2+}, Ca^{2+}, K^{+} and Na^{+} ions). *Not determined

Code	Sample description	SOM (mass%)	pH(H_2O) (±0.2)	$cmol_C$ (+) /kg	Ostrich D-loop[(1)]	Lion 16S	Camel 16S	Giraffe 16S	Blackbuck 16S	Blesbuck 16S	Wildebeest 16S	Eland 16S	Horse/ Zebra 16S	H'	Reads
Savanna enclosure, surface soil															
D1	13 subsamples/400 m^2				(+)			0.0	5.5	34.6	3.9	55.7	0.4	1.00	1964
D2	5 subsamples/400 m^2				(+)			0.0	5.9	39.7	4.9	49.5	0.0	1.03	1370
D3	Mean proportion				2/5			0.0	12.5	45.6	4.5	37.4	0.0	1.13	7688
D4	Mean proportion				4/13			*	*	*	*	*	*		
D5	Mean proportion	*	*	*	7/13			*	*	*	*	*	*		
D1–D3	Mean proportion of area				(+)			0.0	8.0	40.0	4.4	47.5	0.1	1.07	11 022
Savanna enclosure, diversity of 'dirt' DNA															
D3	Mean diversity													0.52 (±0.35)	
D1–D3	Mean diversity of area													1.05 (±0.07)	
Savanna enclosure, Species holding															
-	Proportion of individuals							5	69	6	8	7	5	0.89	
-	Proportion of biomass							30	13	3	12	32	10	1.28	
Savanna enclosure, 10 cm depth															
D6	13 subsamples/400 m^2				(-)			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
D7	5 subsamples/400 m^2				(-)			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
D8	Mean proportion	3.2 (±2.5)	6.0 (±0.4)	6.6 (±4.5)	0/5			0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
Savanna enclosure, 10 cm depth															
E1	5 subsamples/100 m^2	1.2 (±0.4)	6.2 (±0.8)	2.9 (±2.5)			*		*	*	*	*	*		
E2	13 subsamples/100 m^2						35.1		23.9	0.0	21.5	19.5	0.0		615
Lion enclosure															
F1	soil surface					70.0		f					30.0		3002
F2	10 cm depth	*	*	*		0.0							100.0		1078

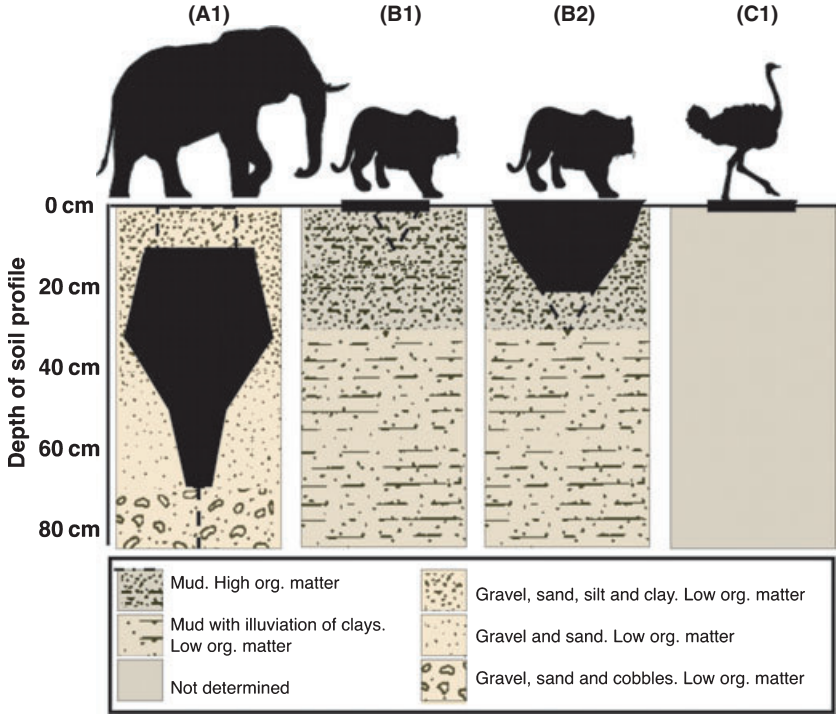

Fig. 3 'Dirt' DNA profiles obtained by high-throughput sequencing (A1 and B2), or using species-specific primers (B1 and C1). (A1) Elephant enclosure (10–70 cm). (B1) Tiger enclosure, below trail (0–40 cm). (B2) Tiger enclosure, below latrine (0–40 cm). (C1) Ostrich enclosure (0–60 cm).

(A1) elephant was detected in soil sampled from 10 to 70 cm depths (Fig. 3). However, Elephantidae specific primers amplifying a 165 bp mtDNA fragment did not yield PCR products from these samples. In the tiger enclosure DNA leaching was detected below the latrine (B2) down to 20 cm depth. DNA leaching was not detected below the trail (B1), although tiger DNA was detected in the surface soil. Furthermore, positive PCR amplification of a 76 bp mtDNA fragment, but not a 124 bp mtDNA fragment, was obtained from these samples using tiger-specific primers. At the ostrich farm (C1), positive amplification of a 78 bp mtDNA fragment was only obtained from the soil surface (< 2 cm).

Even though all but one of the species present in the savanna enclosure (Site D) were detected in the soil surface (D1–D3) none of this diversity was detected in 10 cm depth by any of the three sampling intensities (D6–D8). In another area of the savanna enclosure (Site E) several species that had been absent for 2 months (Table 1) were detected in 10 cm depth. This includes Camelidae, which had been absent from the area for the past 6 years. In the lion (*Panthera leo*) enclosure (Site F) both predator and prey (horse) were detected in the soil surface, but only prey in 10 cm depth.

Soil chemistry, texture and structure

A series of edaphic parameters (geometric surface area per particle volume, pH(H_2O), soil exchangeable Mg^{2+}, Ca^{2+}, K^+ and Na^+ ions and SOM) were determined and contrasted to DNA recovery efficiency in the soil profiles (Table 3) and quadrats (Table 4), where DNA was recovered below the surface.

All soil samples used in this study had values from pH 6.2 ± 0.2 to pH 8.3 ± 0.2, and positive PCR amplification was obtained across this pH range. Inhibition of PCR amplification was detected in soil from Profile B1 and B2, most probably as a result of soil compounds co-extracted with DNA, as these soil samples show maximal high SOM content. However, this inhibition could easily be overcome by diluting DNA extracts ten-fold, in DNA-free water. No definitive relationships could be drawn from the results obtained for soil cations and DNA recovery efficiency. In Fig. 4, the proportions of elephant sequences, within the four samples from Site A, are plotted against surface area per particle volume (As). These results reveal that larger proportions of 'dirt' DNA are obtained from samples with larger soil particle surface area. Furthermore, the change of soil structure, coinciding with the illuvial clay horizon in 30 cm depth below the tiger

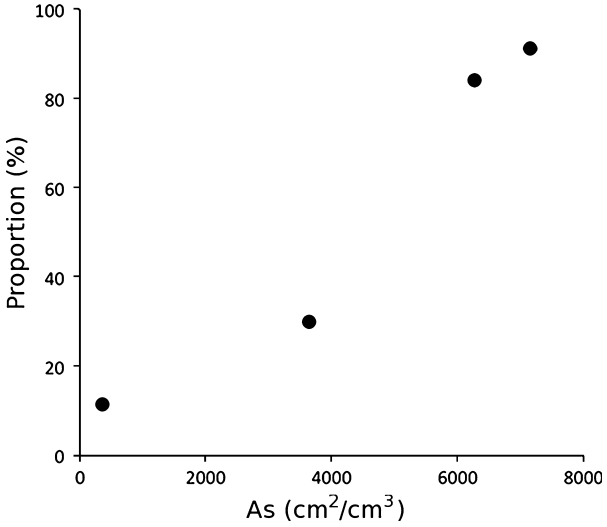

Fig. 4 Proportion (%) of elephant (Elephantidae) 'dirt' DNA recovered from four samples (A1, 10–70 cm depth) plotted against soil geometric surface area per particle volume (As).

latrine (B2), influence the extent of the 'dirt' DNA leaching, whereas leaching in the elephant enclosure (A1), a coarse textured soil, was detected down to 70 cm depth (Fig. 3).

Discussion

'Dirt' DNA as an indicator of taxonomic richness

The assignment of DNA sequences into taxonomic groups is particularly challenging in the case of 'dirt' DNA because of both the short sequence lengths produced, and accumulating errors arising during regular PCR, emulsion PCR and sequencing-by-synthesis (Huse *et al.* 2007; Quince *et al.* 2011). If these errors are not accounted for they lead to inflated estimates of taxonomic richness (Quince *et al.* 2011). Exactly how sensitive biodiversity estimates inferred from high-throughput sequencing of 'dirt' DNA is to such artefacts has not previously been addressed *in situ*, since in the absence of any recognizable remains in the soil no prior knowledge on diversity is available. We evaluated the accuracy of 'dirt' DNA meta-barcoding *sensu lato* (Valentini *et al*. 2008) as an indicator of taxonomic richness, using the unique framework offered by detailed records of animal holdings at safari parks, zoological gardens and farms.

By contrasting the observed and expected error distributions we identified authentic molecular diversity. Among the filtered sequences of the expected fragment length (28–31 bp) several taxonomic identifications showed a 100% match to the mammal reference sequences (data not shown). These discarded taxonomic identifications, are likely a result of artificial mutations. Importantly, the remaining haplotypes could all be related to species that have been/are living in the enclosures, through sequence similarity to orthologous sequences in the reference database. The sensitivity of the method was found to be high since only one genus (*Giraffa*), which was actually present at the study sites, was not detected. However, giraffe (*Giraffa camelopardalis*) had only been present for 2 months in the savanna enclosure (Site D). Interestingly, in the lion enclosure (Site F) both predator (*Panthera*) and prey (*Equus*) were detected.

In most cases one haplotype per species was obtained per site, with the notable exception of the elephant enclosure (A1) where seven haplotypes were identified with four having an exact match to previously reported diversity within both extinct and extant Elephantidae (Supporting information). DNA damage is known to lead to nucleotide misincorporation during PCR (Hansen *et al*. 2001). Consequently, relatively few damaged DNA molecules acting as starting templates during PCR could possibly explain the observed diversity of Elephantidae haplotypes; however, such a phenomenon cannot explain why a high proportion of SNPs occurs at the same nucleotide positions in sequences recovered from independent samples. Another possibility is that matches to DNA of extinct relatives simply reflect true, but previous unreported, diversity in the very short mtDNA fragment amplified. An alternate explanation, of laboratory-based contamination entering the extracts is unlikely, as mastodon has not been studied in the laboratory in which the analyses were performed. This excessive level of haplotype diversity did not preclude correct taxonomic identification as elephant was detected, in agreement with the Asian elephant that was present in the enclosure 2 months before sampling took place. However, for the reasons outlined above the detection of African elephant cannot be authenticated and again highlights an important issue regarding the use of 'dirt' DNA meta-barcoding as an indicator of taxonomic richness. Potential bias can be generated either by (i) low resolution of the barcoding marker leading to an underestimation of the taxonomic richness, or by (ii) intraspecific polymorphism of the barcoding marker leading to an overestimation of the taxonomic richness (Valentini *et al*. 2008).

'Dirt' DNA as an indicator of biodiversity

The composition of the 'dirt' DNA pool is determined by the rate of input and decay in the environment, which is influenced by several aspects including biotic, edaphic and climatic factors (Levy-Booth *et al*. 2007). In

order to evaluate whether biotic factors such as behaviour and physiology influence the deposition rate of 'dirt' DNA from mammals, we contrasted the quantitative 'dirt' DNA profiling with population parameters of the exotic animal stock in the savanna enclosure. Such differences are likely to influence biodiversity estimates inferred using the 'dirt' DNA meta-barcoding approach. Intriguingly, the biomass of the individual populations present on site appears as a better proxy for 'dirt' DNA deposition rates than the population census size (Fig. 2c). If the same situation applies to natural systems, the approach would show great potential as a quick methodological alternative to classical ecological surveys of biodiversity. We find that the spatially sampled data have a higher biodiversity, and a lower variation in biodiversity than that obtained from individual soil samples (Fig. 2a), indicating that the 'dirt' DNA is not uniformly dispersed in the sampled area. Furthermore, we find that the sensitivity of the method is not improved by increasing the volume of soil used for extraction (Fig. 2b).

The optimal quadrat size will likely be species-specific and influenced by social organisation, territorial behaviour and physiology of individual species. For one species (ostrich) presence/absence data was obtained from 13 individual samples from two quadrat sizes (20×20 and 40×40 m). While the expected species was detected in both cases (Table 4, D4–D5), our data highlight that several samples per quadrat are required in order to obtain meaningful estimates of biodiversity from terrestrial ecosystems. In the soil surface at the savanna enclosure (Fig. 2c), DNA sequences from blackbuck (*Antilope cervicapra*), blue wildebeest and common eland were found in proportions that approximate the relative amount of biomass present at the time of sampling. However, blesbuck (*Damaliscus pygargus*) that make up to 6% of total individuals and 3% of total biomass, were frequently identified in the soil surface (40%). This is most likely a result of a unique behaviour in habitat occupation: the sampled 400 m^2 area was positioned on a hill, which is used by the species for territorial marking. By contrast, while zebras had been present for 6 years, the genus was only detected in low amounts (0.4%), using the highest sampling intensity (D1). This could be related to a difference in physiology (zebra is the only non-ruminant species) or behaviour influencing the amount of time spent in the sampled area. Furthermore, while giraffe had been present for 2 months it was not detected.

Additional rare haplotypes will possibly be detected, and accurate estimation of haplotype frequencies will be obtained, by deep-sequencing multiple PCRs of the individual sample DNA extract. Nevertheless, the higher diversity (and lower variation) of the spatial sample replicates (Fig. 2a) indicate that extra spatial sampling, rather than a more detailed description of the individual sample DNA extracts, improves the estimates of structural diversity inferred from 'dirt' DNA profiling of vertebrate communities.

'Dirt' DNA cycling in the environment

The turnover rate of the 'dirt' DNA pool is likely to show climatic and seasonal variation, and even differences in DNA sources and soil microclimates may affect conditions for DNA preservation (Levy-Booth *et al.* 2007). Such differences can potentially introduce some bias in biodiversity estimates inferred using the 'dirt' DNA meta-barcoding approach. From mammals, likely important sources contributing to the 'dirt' DNA pool include faeces, urine, skin flakes and hair (Willerslev *et al.* 2003; Lydolph *et al.* 2005; Haile *et al.* 2007).

Different DNA sources initially provide different quality and quantity of DNA, and whether or not long-term 'dirt' DNA preservation is affected by the original DNA source remains an open question. In this study we only detected mtDNA fragments from vertebrates of 68–78 bp in length. The preservation of six-year-old Bactrian camel (*Camelus bactrianus*) DNA, at higher levels than DNA deriving from species present until 2 months before sampling (Table 4), is possibly related to differences in respective DNA preservation conditions. Interestingly, Bactrian camels produce large amounts of hair, which has been shown to provide extremely well preserved short fragment DNA available for analysis in aDNA studies (Gilbert *et al.* 2004, 2007, 2008b; Rasmussen *et al.* 2010).

DNA preservation may also be affected by microenvironmental factors in the soil matrix. *On the soil surface* only species presently situated in the enclosures were detected (excluding domesticated animals) most likely as a result of the time dependent process of DNA decay (Lindahl 1993), and the constant input of new 'dirt' DNA into the environment. *Below the soil surface* Camelidae were detected after absence for 6 years, as well as several species present until 2 months before sampling (Site E). Likewise, elephant was detected below the surface (A1) after absence for 2 months.

'Dirt' DNA interactions with the soil matrix

The increasing use of 'dirt' DNA as a paleoecological proxy is dependent on temporal association of the DNA molecules with undisturbed sediments, which is compromised by movement (leaching) of DNA between strata. DNA leaching has been shown to occur in unsaturated temperate sediments (Haile *et al.* 2007), but how biotic and edaphic factors influencing the extent of DNA leaching remains unexplored.

We find that DNA had leached below 10 cm depth at two sites: in the elephant enclosure (A1) and underneath the tiger latrine (B2), whereas leaching was neither detected underneath a trail along a permanent fence in the tiger enclosure (B1), nor at the ostrich farm (C1). At both these sites, where leaching was demonstrated, this could be due to behavioural peculiarities resulting in a large amount of urine being deposited in a small area. In agreement with studies conducted *in vitro* (Frostegård *et al*. 1999; Lloyd-Jones & Hunter 2001; Cai *et al*. 2006) we find that the quantity of recovered 'dirt' DNA is proportional to the estimated soil particle surface area of samples from the elephant enclosure (A1), a coarse textured soil (Fig. 4). In contrast, DNA leaching was less pronounced in the tiger enclosure below the latrine (B2). At this site, leaching did not extend beyond the illuvial clay horizon in the soil (Fig. 3), which represents an abrupt change in soil structure and suggests that the clay horizon has acted as a 'shield' precluding further DNA leaching. The extent of DNA leaching was independent of soil chemistry. Of note, all soil samples had pH values above six which results in an overall negative charge of DNA molecules (Allemand *et al*. 1997) allowing further adsorption to soil particles through cations (Nguyen & Elimelech 2007). Therefore, we anticipate that soil chemistry, especially in cases where pH values are inferior to six, could represent an important contributor to DNA leaching in other soil surveys.

Differences in DNA leaching for different organisms/sources have previously been reported (Haile *et al*. 2007), possibly as a result of physiological differences or density (Hebsgaard *et al*. 2009). In the lion enclosure, horse and lion were both detected in the soil surface in proportions 3–7, respectively (Table 4). However, horse but not lion was detected at a depth of 10 cm at this site. Studies of the DNA quality in fresh faecal samples, show that digested prey provides low DNA yields compared to that of the predator (Deagle *et al*. 2006). Interestingly, this relationship suggests that the leached horse 'dirt' DNA does not originate from faeces, as lion was not detected below the surface. Body fluids saturating the soil below carcasses could be the source of leaching and explain the observed discrepancy. This implies that 'dirt' DNA leaching in unsaturated soil is not only dependent on relative input in the soil surface, but is also influenced by the 'dirt' DNA source.

Conclusion

The results are consistent with the findings of previous studies reporting that short mtDNA fragments from local animals can be recovered *in situ* from soils in temperate climatic regions. Soil texture affected the amount of adsorbed mtDNA, which was found to be larger for soils with larger soil particle surface area. This relationship may be used to predict the suitability of sites with respect to the analysis of 'dirt' DNA. Leaching of DNA was detected in cases of extreme population densities and below latrines. The findings further indicate that differences in leaching potential for different DNA sources exist, and that the extent of the leaching is affected by soil structure.

The proportional distribution of haplotypes in the soil surface reflected overall above ground taxonomic richness, but was influenced by the biomass of populations on site rather than the population census size. The performance of the 'dirt' DNA meta-barcoding approach was related to the spatial scale, and extra spatial sampling rather than sample volume was found to improve biodiversity estimates. This, as well as behavioural peculiarities in relation to territorial marking, is of importance for further implementations of the method in natural settings.

Acknowledgments

The research was supported by the Danish National Research Foundation 'GeoGenetics' grant. The authors thank Mads Frost Bertelsen and Jørgen Lindhart (Copenhagen Zoo), Lisbeth Høgh (Knuthenborg Park & Safari), Richard Østerballe (Givskud Zoo), Jesper Stagegaard (Ree Park) and Knud Nybro Jensen (Døllefjelde Strudsefarm) for allowing access to their facilities. The authors also thank Hans Siegismund and Rasmus Heller, from the Department of Biology at Copenhagen University, for supplying samples from their tissue collection, and Kim Andersen for technical support.

References

Allemand JF, Bensimon D, Jullien L, Bensimon A, Croquette V (1997) pH-dependent specific binding and combing of DNA. *Biophysical Journal*, **73**, 2064–2070.

Arnold JL, Roberts RG, MacPhee RDE, *et al*. (2011) Paper II—Dirt, dates and DNA: OSL and radiocarbon chronologies of perennially-frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. *Boreas*, **3**, 417–445.

Binladen J, Gilbert MTP, Bollback JP, *et al*. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, e197.

Bohmann K, Monadhem A, Lehmkuhl Noer C, *et al*. (2011) Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS ONE*, **6**, e21441.

Burger J, Rosendahl W, Loreille O, *et al*. (2004) Molecular phylogeny of the extinct cave lion *Panthera leo* spelaea. *Molecular Phylogenetics and Evolution*, **30**, 841–849.

Cai P, Huang Q, Jiang D, Rong X, Liang W (2006) Microcalorimetric studies on the adsorption of DNA by soil

colloidal particles. *Colloids and Surfaces B: Biointerfaces*, **49**, 49–54.

Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. *Frontiers in Zoology*, **3**. doi:10.1186/1742-9994-3-11

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.

Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Frostegård Å, Courtois S, Ramisse V, *et al.* (1999) Quantification of bias related to the extraction of DNA directly from soil. *Applied and Environmental Microbiology*, **65**, 5409–5420.

Geml J, Laursen GA, Timling I, *et al.* (2009) Molecular phylogenetic biodiversity assessment of arctic and boreal ectomycorrhizal *Lactarius* Pers. (*Russulales; Basidiomycota*) in Alaska, based on soil and sporocarp DNA. *Molecular Ecology*, **18**, 2213–2227.

Gilbert MTP, Wilson AS, Bunce M, *et al.* (2004) Ancient mitochondrial DNA from hair. *Current Biology*, **14**, R463–R464.

Gilbert MTP, Tomsho LP, Rendulic S, *et al.* (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*, **317**, 1927–1939.

Gilbert MTP, Jenkins DL, Götherström A, *et al.* (2008a) DNA from pre-Clovis human coprolites in Oregon, North America. *Science*, **320**, 786–789.

Gilbert MTP, Kivisild T, Grønnow B, *et al.* (2008b) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science*, **320**, 1787–1789.

Haile J, Holdaway R, Oliver K, *et al.* (2007) Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor?. *Molecular Biology and Evolution*, **24**, 982–989.

Haile J, Froese DG, MacPhee RDE, *et al.* (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences, USA*, **106**, 22363–22368.

Hansen A, Willerslev E, Wiuff C, *et al.* (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution*, **18**, 262–265.

Hansen AJ, Mitchell DL, Wiuf C, *et al.* (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics*, **173**, 1175–1179.

Hebsgaard MB, Gilbert MTP, Arneborg J, *et al.* (2009) 'The farm beneath the sand'—an archaeological case study on ancient 'dirt' DNA. *Antiquity*, **83**, 430–444.

Hofreiter M, Mead J, Martin P, Poinar H (2003) Molecular caving. *Current Biology*, **13**, R693–R695.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.

Jerde C, Mahon AR, Chadderton WL, Lodgel DM (2011) 'Sight-unseen' detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.

Kowalczyk R, Taberlet P, Coissac E (2011) Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *Forest Ecology and Management*, **261**, 821–828.

Leonard JA, Shanks O, Hofreiter M, *et al.* (2007) Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science*, **34**, 1361–1366.

Levy-Booth DJ, Campbell RG, Gulden RH, *et al.* (2007) Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, **39**, 2977–2991.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.

Lloyd-Jones G, Hunter DWF (2001) Comparison of rapid DNA extraction methods applied to contrasting New Zealand soils. *Soil Biology & Biochemistry*, **33**, 2053–2059.

Lydolph MC, Jacobsen J, Arctander P, *et al.* (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology*, **71**, 1012–1017.

Meyer M, Briggs AW, Maricic T, *et al.* (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research*, **36**, e5.

Nguyen TH, Elimelech M (2007) Adsorption of plasmid DNA to a natural organic matter-coated silica surface: kinetics, conformation, and reversibility. *Langmuir*, **23**, 3273–3279.

Pegard A, Miquel C, Valentini A, *et al.* (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from faeces. *Journal of Agricultural and Food Chemistry*, **57**, 5700–5706.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, **12**, 38.

Rasmussen M, Cummings LS, Gilbert MTP, *et al.* (2009) Response to comment by Goldberg *et al.* on 'DNA from pre-clovis human coprolites in Oregon, North America'. *Science*, **320**, 786–789.

Rasmussen M, Li Y, Lindgreen S, *et al.* (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, **463**, 757–762.

Segal E, Shouse PJ, Bradford SA, Skaggs TH, Corwin DL (2009) Measuring particle size distribution using laser diffraction: implications for predicting soil hydraulic properties. *Soil Science*, **174**, 639–645.

Soininen EM, Valentini A, Coissac E, *et al.* (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology*, **6**, 16.

Taylor PG (1996) Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Molecular Biology and Evolution*, **13**, 283–285.

Thomsen PF, Elias S, Gilbert MTP, *et al.* (2009) Non-destructive sampling of ancient insect DNA. *PLoS ONE*, **4**, e5048.

Valentini A, Pompanon F, Taberlet P (2008) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110–117.

Valentini A, Miquel C, Nawaz MA, *et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources*, **9**, 51–60.

Valentini A, Miquel C, Taberlet P (2010) DNA barcoding for honey biodiversity. *Diversity*, **2**, 610–617.

Willerslev E, Hansen A, Binladen J, *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

Willerslev E, Hansen A, Poinar H (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution*, **19**, 141–147.

Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B*, **272**, 3–16.

Willerslev E, Cappellini E, Boomsma W, *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, **317**, 111–114.

Zagordi O, Geyrhofer L, Roth V, Beerenwinkel L (2010) Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *Journal of Computational Biology*, **17**, 417–428.

K.A. (Centre for GeoGenetics, Copenhagen University) is developing methods to facilitate the use of 'dirt' DNA preserved in sediments as a paleoecological proxy. J.H. (aDNA research laboratory, Murdoch University), M.R., L.O., M.T.P.G. and E.W. (Centre for GoeGenetics, Copenhagen University) are researching population genetics, paleoecology and evolution, using novel sources of ancient DNA. K.H.K. (GeoGenetics, Copenhagen University) researches glacial geology, stratigraphy and paleoclimate modelling. K.L.B. and H.B-M. (Geographical institute, Copenhagen University) researches soil chemistry and soil development.

Data accessibility

DNA sequences: NCBI SRA: SRX088549; SRX088551; SRX088552.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Position of samples in quadrate plots and soil profiles at the study locations.

Fig. S2 Sample numbers and position inside the 20×20 and 40×40 m quadrates of Site D, and position of samples where ostrich was identified using species-specific primers.

Table S1 Additional sequences of the 16S mammalian barcode region, obtained from tissue samples, or identified by deep sequencing of soil samples of Profile A1 and Site D.

Table S2 Nucleotide tags and numbers corresponding to sample names, and accession numbers for the three sequencing experiments uploaded to NCBI SRA.

Table S3 The distribution by number of errors, calculated for seven haplotypes identified as Elephantidae (11 598 reads), *Elephas* (8040 reads), *Loxodonta* (2897 reads) and *Panthera*. The per base error rate (p) were calculated from the frequency of mismatches ($k \leq 4$) including substitutions, insertions and deletions, and the proportion of errors distributed among these three groups.

Table S4 Frequency data obtained from high-throughput sequencing of samples from quadrate plots at Site D–F.

Table S5 Frequency data obtained from high-throughput sequencing of samples from Profiles A1 and B2.

Table S6 Measurements of cation concentrations [$cmol_c$ (+)/kg], total carbon (C %), SOM (mass %) and pH (H_2O) of soil samples from the study sites.

Table S7 Results for soil fraction analysis on soil samples from soil Profiles A1, B1 and B2.

Appendix S1 Materials and methods: Details on sampling locations and study sites, Specificity of the species-specific primers, Extension of reference sequence material, Haplotype identifications, Taxonomic identifications, Internal standard and soil analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.