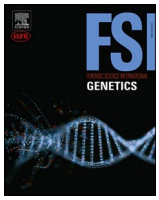
Forensic Science International: Genetics 67 (2023) 102931

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



The secret hidden in dust: Assessing the potential to use biological and chemical properties of the airborne fraction of soil for provenance assignment and forensic casework

Nicole R. Foster [a,*], Duncan Taylor [a,b], Jurian Hoogewerff [c], Michael G. Aberle [c], Patrice de Caritat [c,d], Paul Roffey [c,e], Robert Edwards [a], Arif Malik [f], Michelle Waycott [f], Jennifer M. Young [a]

[a] College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia
[b] Forensic Science SA, GPO Box 2790, Adelaide, SA 5001, Australia
[c] National Centre for Forensic Studies, University of Canberra, Bruce Australian Capital Territory 2617, Australia
[d] Geoscience Australia, GPO Box 378, Canberra Australian Capital Territory 2601, Australia
[e] Australian Federal Police, GPO Box 401, Canberra Australian Capital Territory 2601, Australia
[f] School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia

ARTICLE INFO

Keywords:
Environmental DNA
Trace DNA
Forensic biology
Dust
Metabarcoding

ABSTRACT

The airborne fraction of soil (dust) is both ubiquitous in nature and contains localised biological and chemical signatures, making it a potential medium for forensic intelligence. Metabarcoding of dust can yield biological communities unique to the site of interest, similarly, geochemical analyses can uncover elements and minerals within dust that can be matched to a geographic location. Combining these analyses presents multiple lines of evidence as to the origin of dust collected from items of interest. In this work, we investigated whether bacterial and fungal communities in dust change through time and whether they are comparable to soil samples of the same site. We integrated dust metabarcoding into a framework amenable to forensic casework, (i.e., using calibrated log-likelihood ratios) to predict the origin of dust samples using models constructed from both dust samples and soil samples from the same site. Furthermore, we tested whether both metabarcoding and geochemical/mineralogical analyses could be conducted on a single swabbed sample, for situations where sampling is limited. We found both analyses could generate results from a single swabbed sample and found biological and chemical signatures unique to sites. However, we did find significant variation within sites, where this did not always correlate with time but was a random effect of sampling. This variation within sites was not greater than between sites and so did not influence site discrimination. When modelling bacterial and fungal diversity using calibrated log-likelihood ratios, we found samples were correctly predicted using dust 67% and 56% of the time and using soil 56% and 22% of the time for bacteria and fungi communities respectively. Incorrect predictions were related to within site variability, highlighting limitations to assigning dust provenance using metabarcoding of soil.

1. Introduction

Dust was first proposed as a forensic evidence tool back in 1929 by Edmond Locard in his paper "The analysis of Dust Traces" published in the American Journal of Police Sciences. This analysis involved examining dust particles by hand to find key components that could be related to a crime scene [1]. Fortunately, environmental DNA (eDNA) has opened the door for faster, easier, and more sensitive methods to draw associations between trace material and crime scenes. eDNA encompasses all DNA found in environmental material such as water, soil and scats, and advancements in Massive Parallel Sequencing (MPS) technologies has seen this field proliferate [2]. The process of identifying organisms within environmental samples using eDNA is termed metabarcoding and this involves the amplification of specific universal regions of DNA from environmental samples using Polymerase Chain Reaction (PCR) and sequencing this DNA on MPS platforms [3]. This

* Corresponding author.
E-mail address: nicolefoster007@gmail.com (N.R. Foster).

https://doi.org/10.1016/j.fsigen.2023.102931
Received 30 September 2022; Received in revised form 19 August 2023; Accepted 21 August 2023
Available online 22 August 2023

tool is already being applied to forensic science across a range of applications including sampling DNA from soils or trace materials and applying this to determine provenance [4]. Metabarcoding has successfully recovered bacterial and fungal communities from trace materials such as dust [5,6], which, due to its ubiquitous nature, could be a key medium for intelligence purposes [4]. Furthermore, geochemical and mineralogical analyses may be able to provide complementary evidence to metabarcoding given these analyses have been undertaken on dust [7] and have yielded provenance potential in soil samples [8,9]. The combination of metabarcoding and biogeochemistry of dust has not yet been undertaken, including from dust swabs. Given the small sample size of dust collected from an item or location/surface of interest, it is unclear whether both analyses could be carried out on the same swabbed sample.

Dust collected on items of clothing or materials, could enable unique insights into where a person of interest has travelled. The bacterial and fungal signatures of dust have been shown to be localised to the source location and are geographically distinct [7,10]. Additionally, this has been demonstrated for soil samples [11], whereby community profiles can be different even at small geographic scales [12]. Metabarcoding of soil samples is used far more broadly than dust metabarcoding for a range of ecological applications [13]. This means metabarcoding data pertaining to soil samples from wide geographic origins are available in public repositories e.g., National Centre for Biotechnology Information (NCBI), European Nucleotide Archive (ENA) and these can be compiled to create reference databases e.g. Australian microbiome: https://www.australianmicrobiome.com/; international project Earth microbiome: http://www.earthmicrobiome.org/. The ability to match dust samples collected from items of interest to soil samples present in existing metabarcoding databases, would be useful to integrate dust metabarcoding as a forensic intelligence tool. This exact relationship has not yet been quantified, although Badgley, Jesmok and Foran [14] did find that bacterial community profiles of soil samples changed with the length of time that an item was separated from the sampling location. However, this was not to the extent that these items did not cluster with their source location in ordinations. Therefore, more research needs to be done to assess whether soil metabarcoding can be applied to dust sample provenance.

Alongside using soil samples to determine dust sample origin, provenance estimates of dust need to be integrated into forensic casework. Relationships between samples using metabarcoding data are often analysed using distance-based algorithms and plotted using clustering or ordination such that relationships can be visualised [4]. While this is useful, the European Network of Forensic Science Institutes (ENFSI) guidelines recommend a Bayesian framework for these analyses [15]. The most agreed upon way to present such results is in a likelihood ratio (LR) framework. This is shown by the equation below:

$$LR = \frac{p(E|Hp)}{p(E|Hd)}$$

Hp represents the proposition of the prosecution and Hd represents the proposition of the defence. Depending on the circumstances of the case, these propositions will vary to best address the question of interest. In broad terms, the propositions could be set to answer questions of common source or specific source. For common source propositions, the questions will be around whether two samples have a common (but unknown) origin (Hp), or whether they have a different origin (Hd). For specific source, the questions are whether a questioned sample comes from a nominated population or site (Hp) or from a population other than the one that has been nominated (Hd). This latter type of question is what we explore in our work, and leads to a LR which, if described in words, can be presented as:

$$LR = \frac{Probability\ of\ obtaining\ the\ result\ if\ the\ dust\ sample\ came\ from\ a\ site\ of\ interest}{Probability\ of\ obtaining\ the\ result\ if\ the\ dust\ sample\ came\ from\ a\ different\ site}$$

A LR greater than one supports the prosecution proposition that the sample came from the site of interest, while a value less than one supports the defence proposition that the sample came from a different site. Our ability to present results in this way, rather than ordinations, is important if dust metabarcoding is to become a tool in criminal investigations [4]. For complex trace measurements, where the dependencies of measurements are not fully modelled or understood, it is common to develop a LR based on 'scores' (called a score-based LR) [16]. These scores can be assigned by an algorithm designed to identify similarities and need not have a real-world meaning. The distribution of scores for same-site comparisons and different-site comparisons can then be used as population data for evaluating the significance of a single score in a case [17]. A LR can also be assigned based on the results of logistic regression that assigns a probability to samples coming from the same site.

In this study, we examined whether both biological and chemical properties of dust could be ascertained from a single swabbed sample and whether the results would prove to be discriminatory between sites. Further, we undertook an in-situ experiment to collect dust at different case study locations to determine whether the length of time an item spent at a location influenced the biological signal, and ability to discriminate between sites using metabarcoding. We also developed a framework to assess community composition of dust samples in the context of forensic casework by employing score-based calibrated log-likelihood ratios. By implementing this framework, we could then predict dust sample origin and examine whether the relationship between soil and dust samples is sufficiently correlated to utilise existing soil metabarcoding reference databases for future provenance estimates of dust. Thus, three main questions were addressed in this study; 1) Can biological and chemical properties be obtained from the same dust swab? 2) Does the ability to determine dust sample origin using metabarcoding change with the length of time spent at a location? And 3) Can we integrate dust metabarcoding into forensic casework and use soil samples as predictors for dust sample origin?

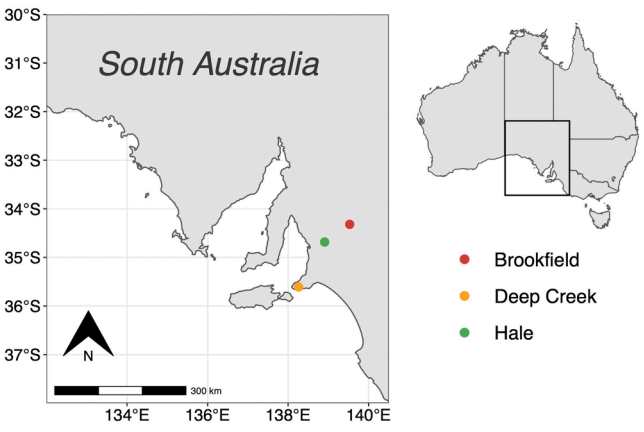


Fig. 1. Map of the three study sites where soil was collected for DNA extraction and experiments were set up for dust collection.

2. Methods

2.1. Study site

Three sites with differing soil properties were chosen across South Australia to conduct experimental dust collection, the choice of which was guided by samples collected by the Terrestrial Ecological Research Network (TERN) https://www.tern.org.au/ given soil properties had been established for these sites. The sites were Hale Conservation Park (34.68 S, 138.91E - a clayey-sand soil with a pH of 5.41), Deep Creek Conservation Park (35.61 S, 138.26E- a loam soil with a pH of 5.87) and Brookfield Conservation Park (34.31 S, 139.53E- a sandy loam soil with a pH of 7.1). See Fig. 1 for site map. Experiments were set up at each location and on the same day (week = 0), soil samples (~ 500 g) were collected from a 1 m^2 area in the centre of the experimental set up (at each location) and only the top 0.5 cm of soil was sampled using a sterilised trowel (cleaned with bleach, water and ethanol prior to use). Three replicates of 250 mg of homogenised soil samples from each site were used for the DNA extraction and soil model development. Additionally, rainfall for the region during the sampling period was recorded because, despite a tarpaulin being used for preservation of deposited dust particles, excessive rainfall can impact dust accumulation (Fig. S1).

2.2. Experimental design

Within each site, three independent dust collection experiments were conducted. These consisted of 12, 5 cm × 5 cm tiles placed on three circular boards (N = 36) at a height of 0.5 cm off the ground – being the optimal height for dust collection [18]. Each board supporting 12 tiles each, had an outer metal cage to prevent animal intervention and a tarpaulin cover overlaid on top to protect the sample from rain (although, we acknowledge in real world scenarios rain and other weather factors will impact dust accumulation i.e. UV degradation of DNA). The three dust collection boards at each site were spaced ~5 m apart to obtain spatial homogeneity at each site. The tiles were cleaned with bleach, water and ethanol before being placed at each site to collect dust, week = 0. After 1, 4, 8 and 12 weeks, three tiles from each of the three boards were selected randomly to account for edge effects and swabbed for DNA (3 tiles per board, 9 tiles per site, per week sampled, N = 108). Swabbing consisted of a cotton or nylon swab [19] with 60 μL of Triton -X buffer pipetted onto the top of the swab, prior to rubbing along the entire surface of the tile. The tile was disposed of after swabbing to prevent resampling the same tile. Swabs were retained in sterilized tubes and kept at room temperature until extraction [19]. An example of dust accumulation on the tiles can be found in Fig. S2.

2.3. Environmental DNA sampling, extraction, and sequencing

All dust samples (N = 108), soil samples (N = 9) and the addition of blank tubes as controls (N = 12) underwent DNA extraction using the DNeasy PowerSoil Pro Kit (QIAGEN, Hilden, Germany), adhering to the manufacturer's instructions aside from replacing the PowerBead Pro tubes with 2 mL Microcentrifuge tubes and zirconium beads (mixture of 1.4 mm and 2.8 mm). For the dust samples, an additional step was conducted prior to DNA extraction using this kit. Firstly, 600 μL CD1 solution was added to the sample tube containing the swab and this was incubated on a ThermoMixer C (Eppendorf, Germany) at 65ºC for 10 mins at 650 rpm. The swab was then removed using sterilised tweezers and placed in a Casework Spin Basket (Promega Corporation, Sydney, Australia) that was then placed back into the tube and centrifuged for 15,000 x g for 2 mins. The spin basket containing the swab was then disposed of and the solution was transferred into a 2 mL microcentrifuge tube with zirconium beads and proceeded through the same extraction process as the soil samples, adhering to the DNeasy PowerSoil Pro Kit instructions. After the addition of the initial lysis buffer step of this extraction protocol, followed by mechanical lysis and centrifugation, the remaining supernatant proceeded through the extraction protocol, but the pellet or remnants of the extraction for a subset of samples collected in week 12 (3 from each site, N = 9), were retained for further biogeochemical analyses. Following DNA extraction, a random subset of samples (N = 14) at different times (including soil samples) were quantified using a Quantus™ Fluorometer and QuantiFluor dsDNA System (Promega) to roughly assess DNA recovery. From the DNA concentration values, all soil samples underwent a 1:10 dilution with molecular grade water to ensure the following PCRs were not hindered by the high concentrations of DNA that were measured.

DNA extracts were then amplified via PCR for fungi using the forward primer segment (ITS1F) 5'- TTGGTCATTTAGAGGAAGTAA – 3' and the reverse primer segment (ITS2) 5'- CGTTCTTCATCGATGC-3' [20]. The bacterial 16 S rRNA v4 region was amplified using forward primer 5'-GTGCCAGCMGCCGCGGTAA-3' and reverse 5'- GGACTACHVGGGTWTCTAAT – 3' [20]. All PCRs were done in triplicate to minimise PCR bias. All primers were modified to include both unique dual barcodes for each extract so that samples could be pooled [21] and Illumina sequencing adapters. PCR amplification was performed in triplicate 12.5 μL comprised of; 2 mM $MgSO_4$, 0.6 mM dNTPs, 0.4 μM of each primer, 0.3 μL Platinum™ Taq DNA Polymerase High Fidelity in 10 x reaction buffer (Invitrogen™, Carlsbad, CA, USA), and 1.25 μL DNA extract. The PCR amplification protocol for 16 S was 3 mins at 94ºC, followed by 35 cycles of 94ºC for 45 s, 50ºC for 60 s and 68ºC for 90 s, and a final extension of 68ºC for 10 mins. The PCR amplification protocol for ITS was 1 min at 94ºC followed by 35 cycles of 94ºC for 30 s, 52ºC for 30 s and 68ºC for 30 s, and a final extension of 68ºC for 7 mins. All triplicate PCRs were combined and then visualised using gel electrophoresis (1 × TE buffer, 1.5% agarose gel for 25 min at 80 V). All samples were then pooled according to concentration estimates (determined via visual inspection of gel) into batches of 24 samples and then purified using AMPure XP beads (at 0.7 x volume concentration) to remove remaining primers and other impurities. Sample batches were then quantified on an Agilent High Sensitivity D1000 ScreenTape (Agilent) at 250–600 bp. All sample batches were then pooled to a final concentration of 2 nM and were sent for Illumina sequencing at the South Australian Genomics Centre with 2 × 150 bp paired end sequencing using Illumina MiSeq v2 chemistry.

2.4. Qualitative Elemental and Mineralogical Composition

A subset of dust samples (3 from each site, N = 9) that underwent DNA extraction after 12 weeks in the field were set aside after the initial step of DNA extraction (addition of lysis buffer, centrifuge and supernatant removed). The remaining sample is termed the 'pellet' of the DNA extract, and this was sent for analysis to determine elemental and mineralogical composition. In addition, nylon swabs (which were used to collect dust on the same subset of samples) that had not undergone DNA extraction, were also sent for analysis (to quantify any signals from particles of nylon which may have made it into the pellet). Zirconium beads were removed from the pellet of the extraction prior to analysis but these were also quantified separately in case of residue contamination. Sample extraction and preliminary preparation was carried out at the National Centre for Forensic Studies, University of Canberra. Approximately 1 mL of ultrapure Milli-Q water was added to the microcentrifuge tube containing the sample, and vortexed briefly until particles were sufficiently suspended. The solution was passed through a 500 μm nylon screen mesh, fitted in a glass funnel, into a 2 mL microcentrifuge tube. The mesh and funnel were washed with ultrapure Milli-Q water. The 2 mL microcentrifuge tube containing the wash solution (and any extracted particles) was centrifuged at 4000 rpm for 5 mins. The supernatant was carefully decanted, and the above procedure repeated to maximise recovery. Once recovery was considered adequate by visual inspection, the supernatant (ultrapure Milli-Q water) was removed and replaced with 100 – 200 μL of ethanol. By vortexing momentarily, and pipetting up-and-down several times, the pellet was

suspended in the liquid and transferred to an agate mortar. Particles were gently ground using an agate pestle by hand to produce a grain size more amenable to X-ray diffraction (XRD). The agate pestle was washed with ethanol, to recover any adhered particles. The ethanol solution in the mortar was left to passively evaporate briefly (< 2 – 3 mins), and by pipetting up-and-down several times, the pellet was suspended in the liquid and deposited onto a low background holder, being careful to minimise potential grain size separation. The sample was allowed to air dry in a fume cupboard, prior to being mounted and presented to the XRD instrument.

Powder XRD analysis was carried using a Malvern PANalytical Empyrean Series 3 X-ray diffractometer equipped with Bragg-Brentano HD divergent beam optic and a PIXcel3D detector. The diffractometer was fitted with a Co X-ray tube and operated at an applied voltage and current of 40 kV and 40 mA respectively. The Bragg-Brentano HD fixed optics were configured to allow maximum irradiation coverage of the circular sample holders, while avoiding beam spill-over. To improve particle statistics, a spinning stage was used to rotate the sample during measurement. Diffraction patterns were recorded by continuous scanning from 4 to 85 degrees 2-theta (°2θ), at a step size of 0.0131 °2θ and counting for 97 s/step, with the sample spun at 2 rotations per second. Under these conditions, one sample took 43 min to run. Qualitative mineral identification was carried out using Bruker's DiffracPlus EVA® version 12 software (2005; Karlsruhe, Germany) and the International Centre for Diffraction Data's Powder Diffraction File-2 database (2004; Newtown Square, PA, USA). Upon completion of XRD measurements, the dry sample particles from the low background holder were scraped into a 50 mL centrifuge tube. Any adhered particles were recovered by washing the low background holder with ultrapure Milli-Q water into the 50 mL tube. These were subsequently centrifuged at 2000 rpm for 10 min, the supernatant was then removed, and dried in a laboratory oven at 40°C with loosely placed lids.

To each 50 mL tube, 3 mL of 1:1 Aqua Regia (65% HNO_3, 35% HCl, both SupraPure Merck) was added, and heated on a heat block at 90°C for digestion over a one-hour period. The digest was diluted to 50 mL with Milli-Q water. From here, 0.2 mL was further diluted to 10 mL with 1% HNO_3 (SupraPure Merck) and 9 ng/L Internal Standard for ICPMS analysis. Three lab QC samples and three method blanks were added to the batch. Calibration standards were prepared consisting of a biforked dilution series of the Merck VI multi element standard, also diluted in 1% HNO_3 (SupraPure Merck) with 9 ng/L Internal Standard. Samples were analysed on an Agilent 7900 ICPMS under normal operating conditions.

3. Data analysis

3.1. Environmental DNA

Sequence reads were analysed using the DADA2 [22] pipeline separately for bacteria and fungi with default settings in R version 1.4.1 [23], all subsequent analyses were also conducted in R. This pipeline conducts sequence trimming, filtering, merging of paired reads and the construction of an amplicon sequence variant (ASV) table. A decision was made not to rarefy sequence reads given the evidence in McMurdie and Holmes [24]. Sequence variants were therefore assigned taxonomy using the SILVA database version 138.1 [25] and the UNITE ITS database general FASTA release [26] for 16S or ITS data, respectively. The parameter *minBoot* was set to 80 for the *assignTaxonomy* function in DADA2. Taxonomic assignment and sample data were then converted into a 'phyloseq' object [27] and an Operational Taxonomic Unit (OTU) table was constructed, *decontam* [28] was then used to filter taxa from samples that were detected in the control samples. The full code from raw sequences to a 'phyloseq' object is available in the Supplementary Material.

Further implementing the use of 'phyloseq', Bray-Curtis dissimilarities were calculated between sites and sampling times to determine beta diversity estimates using *vegdist* from the 'vegan' package [29]. Homogeneity of group (site) dispersions (variances) was calculated using the *betadisper* function from vegan. Significant results for the homogeneity of variance test meant we then used the modification to PERMANOVA pseudo F-test statistic developed by [30] to test significant differences between sites and sampling times for data with non-homogenic variance. We used the functions *f_permanova* and *f_permanovaPW* developed by [31] within the fathom toolbox in 'matlab' [32] to conduct these analyses with p-values adjusted using the Bonferroni method. Bray-Curtis dissimilarities were ordinated using non-metric Multi-Dimensional Scaling (nMDS) to visualise differences in diversity between sites and sampling time and this was done using 'phyloseq' [27]. To check how well the ordination fit the data, we plotted ordination distance against observed dissimilarity using the *stressplot* function in 'vegan' [29] and these are shown in the Supplementary Fig. S5. Species richness (Shannon diversity index) was also assessed within sites at the different sampling times and between the different boards set up to collect dust using 'phyloseq' [27]. An analysis of variance (ANOVA) using a linear model fit with sampling boards nested within time was applied to analyse species richness data.

3.2. Elemental and mineralogical analyses

It has already been established that chemical analyses can be conducted on filtered dust samples [7], therefore, we sought to assess whether geochemistry and metabarcoding could be conducted on a swabbed dust sample and yield results that are different between sites. Due to the small sample size of the material, absolute quantifications of elemental composition could not be achieved and so the decision was made to normalise elemental composition for each sample by summing all the measured elements and determining the percentage of each, N = 9 samples. One way t-tests were conducted in Microsoft Excel between each of the different sites (Brookfield, Hale and Deep Creek) for each measured element. Mineralogical data was simply qualified as either detected, not detected or maybe detected – the latter refers to a possible peak in the diffractogram, but this was not clear enough to the analyst to confidently assign presence. The relationship between elemental and mineralogical composition at the three different sites tested (Brookfield, Hale and Deep Creek) was explored using a redundancy analyses. This was conducted in R version 1.4.1 [23] using the 'vegan' package [29] with elements and minerals as explanatory variables and sites as response variables.

For these subset of dust samples that underwent elemental and mineralogical analyses (collected from week 12, 3 samples per site; N = 9), they also underwent metabarcoding and so Bray-Curtis dissimilarities were calculated to compare site clustering. These were visualised using Multi-Dimensional Scaling (MDS) separately for bacterial and fungal profiles. Additionally, we compared the recovered elemental and mineralogical compositions of the DNA extraction pellet to reference samples from the same site [33]. These reference samples were from the dust fraction of soil (< 75 μm) from these same sites and were analysed the same way – they were also normalised for each sample. For each site, one way t-tests were conducted to compare relative elemental composition between the pellet sample and the reference material.

3.3. Likelihood ratio calculations

To integrate dust metabarcoding data into forensic casework and assess whether soil samples can be used as predictors for dust provenance, we modelled both soil samples and dust samples collected in week 12 of the study (a random 3 from each site, N = 9 to match the number of soil samples collected at week = 0) separately for bacterial and fungal communities. Soil samples here are used as a proxy for soil metabarcoding reference databases given these are the same types of samples that would be deposited in databases. Choosing dust samples from week 12 was done purposefully as this was the maximum time

between collecting the soil samples which is more alike to a real-world scenario e.g., soil samples from reference sequence databases will be collected on different time scales to dust samples. The total number of soil and dust samples modelled was nine each; three sites and three samples per site. Bray-Curtis dissimilarities (BC) were calculated between soil or dust samples of the same site and then again between samples of different sites, and these values were used to calculate score-based LRs. Inflated beta distributions were fit to the data using 'gamlss' [34] in 'R' version 1.4.1 [23] to generate same site and different site distributions. The probability density function (PDF) was then evaluated for each sample's BC scores for the same site model and the different site model using *dBEINF* in the 'gamlss' package [34] and the log ratio of these values was taken to determine a log-LR for each sample. These log-LRs were then calibrated using guidelines in [17] to determine scores which is defined as "log-likelihood-ratio like in that it indicates the degree of similarity of a pair of samples while taking into consideration their typicality with respect to a model of the relevant population". These scores were determined as the average log-LR across the different sites. For example, each site has three samples (for soil and dust) and so for a given sample there are two BC scores calculated for same site distributions and six BC scores for different site distributions. Thus, one sample has two PDF values and subsequent log-LRs for the same site and six PDF values and log-LRs for the different sites. The log-LRs were then averaged over sites giving one score for the same site distribution and two scores for the different site distribution (one for each of the other two sites), totalling 3 scores for each sample. Logistic regression was then used to calibrate these scores and convert them to a likelihood ratio that can be interpreted in forensic casework. This was carried out in R version 1.4.1 [23] using package 'arm' [35], specifying same-origin as 1 and different-origin as 0, and generating probability values for each score. Probabilities were then converted to log-LRs as in Morrison [17].

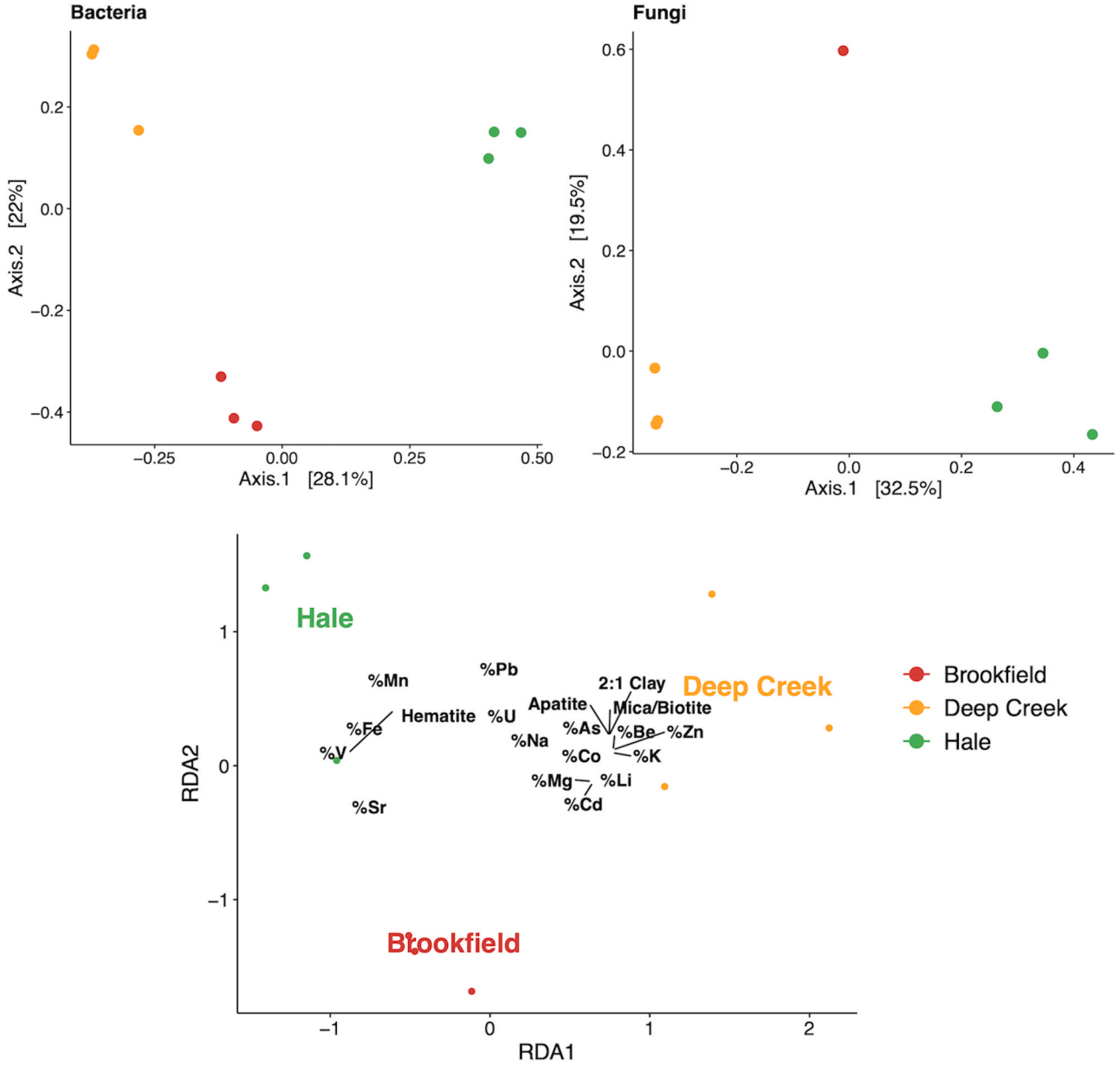


Fig. 2. Statistical biplots for dust samples collected in week 12 with both eDNA (bacterial and fungal communities), elemental and mineralogical analyses on the same samples. Axis 1 and axis 2 of the MDS plot for Bray-Curtis dissimilarities are shown for bacterial communities (top left) and fungal communities (top right). The first two dimensions (RDA1 and RDA2) of the ordination space for the redundancy analysis are shown in the bottom graph. Explanatory variables (elemental and mineralogical composition) are written in black and response variables (sites; Brookfield, Hale and Deep Creek) are coloured.

A random sample from each site, for sampling weeks 1, 4 and 8; N = 9 of the remaining dust samples were used to test the prediction capacity of the soil and dust models generated above. Selected dust samples were individually compared to each sample used to generate the soil and dust models; BC scores were generated for each comparison and PDF values were determined for each sample comparison using the inflated beta models of same-site and different-site distributions. Log-LRs were then calculated for each sample comparisons, and scores were generated by averaging log-LRs over samples for the same site and different site distributions. Logistic regressions were then constructed using these score values and probability values were calculated. These score-based probabilities were then converted into log-LRs and plotted using the known origin of the tested dust samples to ascertain which model generated the correct log-LR interpretation.

4. Results

The results are presented in three subsections, according to the specific question we tried to address.

4.1. Can biological and chemical properties be obtained from the same dust swab?

Both elemental and mineralogical analyses could be undertaken on the same swabbed dust sample that also underwent DNA extraction, and the results had discriminatory capabilities between sites (Fig. 2). Unfortunately, two of the three fungal profiles for the dust samples from Brookfield did not yield adequate sequencing reads and so failed to pass filtering and trimming, so there is inconsistency in the comparisons. Nevertheless, all nine samples worked for bacteria eDNA profiles and all elemental and mineralogical analyses. We found the nylon from the swab produced two broad "humps" in the diffractogram, however, we concluded this was non-issue given the qualitative nature of results but could become more of an issue if results were to be quantified. We also found a distinguishable pattern in the zirconium beads diffractogram which was not identified in sample diffractograms, highlighting residual contamination of the sample did not occur. We were still able to generate elemental signals for all samples despite undergoing the first step of the DNA extraction process. Of the elements measured, we only report on those that were significantly different between sites, and these are shown in Supplementary Table S1. Briefly, Deep Creek had a significantly greater percentage of elements Li, Be, Na, K, Zn, As and Pb than Brookfield which had greater percentages of Sr. Deep Creek also showed greater percentages of Li, Be, Na, K, Co, Zn, As and Cd than Hale which had a greater percentage of V, Fe and Sr. Finally, Hale had a higher percentage of elements V, Mn, Fe, Sr, Pb, U than Brookfield, whereas Brookfield had a higher percentage of Mg and Cd. In the mineralogical results, 2:1 clay, apatite and mica/biotite were only detected at Deep Creek and Hematite was only detected at Hale. Talc was also only detected at Hale and Deep Creek not Brookfield (Supplementary Fig. S3 (top)).

Comparing these results to reference dust material from the same site, Supplementary Table S2, we observed significant differences in elemental composition between the pellet and the reference samples. While we would expect some differences given different sample volume and sample type (dust versus pellet), these differences should not be significant if the elemental results are to be used in conjunction with reference data. There were also some differences in detected minerals between the two methods across all sites (Supplementary Fig. S3 (bottom)). This highlights that while we can extract elemental and mineralogical information from the same sample that has undergone eDNA analysis (i.e., the pellet of the DNA extraction), the results may not be comparable to reference material from the same sites and so may not be useful to estimate provenance.

4.2. Does the ability to determine dust sample origin using metabarcoding change with the length of time spent at a location?

Dust samples collected from different sites and analysed for bacterial and fungal communities showed significant differences in Bray-Curtis dissimilarity measures between sites, Fig. 3 (P = 0.003; Supplementary Table S3 & S4). Pairwise comparisons between sampling weeks within each site, showed significant community differences but these were not standardised for particular weeks, in other words, the weeks that were significantly different to one another were random across sites (Supplementary Table S5 & S6). For example, bacterial communities at the Deep Creek site were significantly different between all sampling times whereas Brookfield did not show significant differences for weeks 1 vs 4 nor weeks 8 vs 12 and Hale showed non-significant differences only for weeks 4 vs 8. Similarly for fungi profiles, we saw significant differences between all sampling times at the site Hale except weeks 4 vs 8, whereas Brookfield showed no differences across any sampling period and Deep Creek only showed significant differences for the week 1 comparisons. Looking at the number of reads and subsequent OTU's assigned to sequences, did not explain the random differences observed between weeks (Supplementary Fig. S4). Similarly, soil samples were significantly different to dust samples for some of the sampling weeks but this was random across sites and communities (bacteria or fungi) and was not explained by read depth. Notably, Deep Creek soil and dust samples were significantly different across all weeks for fungi communities and for Brookfield there were no significant differences across any weeks for bacteria communities. Stress plots (Supplementary Fig. S5)

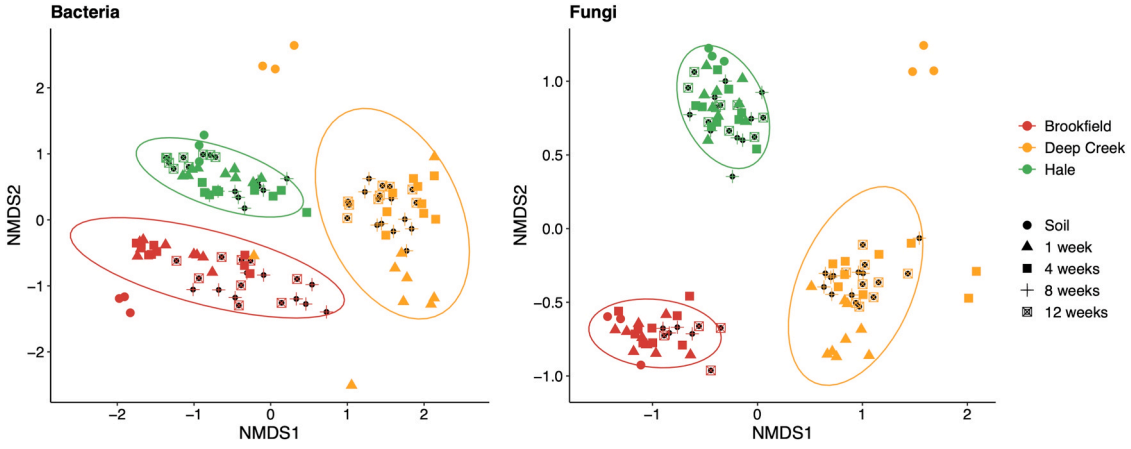


Fig. 3. 16 S (left) and ITS (right) Bray-Curtis dissimilarity measures with 95% confidence ellipses for soil and dust sampled over a three-month period using swabs (n = 36 at each site, N = 108). Shapes indicate either soil samples or dust sampling duration; colours indicate the site.

show that the ordinations using Bray-Curtis dissimilarities summarise the data well given R^2 values are > 0.9 and stress values are < 0.2 for both communities [36].

Species richness was found to be variable across time within the different sites measured. In the bacteria profiles, Brookfield samples in week 8 and 12 showed reduced species richness compared to weeks 1 and 4, which supports the community differences observed above, and the site Hale had significantly higher species richness in week 12 compared to the other weeks ($P < 0.05$). Species richness within fungal communities remained consistent through time at all sites (Supplementary Fig. S6). We also noted differences in species richness between soil and dust samples across both bacteria and fungi profiles for all sites except Hale. Further, significant differences ($P < 0.05$) in species richness were detected between the different boards set up to collect dust within the different time periods, for all sites and communities, except for bacteria communities from the site Hale, demonstrating spatial heterogeneity of biological communities within sites.

4.3. Can we integrate dust metabarcoding into forensic casework and use soil samples as predictors for dust sample origin?

Figs. 4 and 5 show the integration of Bray-Curtis dissimilarities to log-likelihood ratios so that they could be interpreted in a forensic context. The top graphs show inflated beta distributions between Bray-Curtis dissimilarities for samples of the same and different sites. The middle graphs show logistic regressions fitted to scores and the bottom graphs show the linear relationship of log-LR values calculated for scores. Dust sample comparisons (Fig. 5 (right)) never show complete dissimilarity in fungal composition i.e. BC $\neq 1$. Thus, probability calculations using logistic regression show at least a ~12% chance that a dust sample from a different site is predicted to be from the same site, as opposed to a 0% chance calculated for bacterial dust sample comparisons.

Using these models to predict sample origin for the subset of dust samples collected in weeks 1, 4 and 8 for both bacterial and fungal community profiles, we found several instances where samples were predicted to be from different sites (LR < 1) than the site of interest (origin) (LR > 1), i.e., the provenance prediction was incorrect. This occurred more for the soil model than the dust model; out of the nine possible correct predictions per test we saw four incorrect predictions for the soil model using bacteria profiles and seven for fungi. Using the dust model, we saw three incorrect predictions using bacteria and four for fungi (Supplementary Table S6). However, for all these incorrect predictions, this was a matter of not generating a prediction to one of the three sites at all, rather than an incorrect prediction to the wrong site (Fig. 6). Further, when samples were tested against non-origin sites all tests returned correct predictions that the sample was from a different site to the non-origin sites i.e., across both models tested and biological communities no sample was detected to be from a site different to the site of origin. In total, for bacteria communities, the dust model predicted 67% of samples correctly compared to 56% for the soil model. For fungal communities, the dust model predicted 56% of samples correctly compared to 22% for the soil model. Overall, bacteria appear to be a better predictor of dust sample origin than fungi, similarly, the dust model is a more effective predictor of provenance than the soil model. Even so, for bacterial communities across the three sites tested, the soil

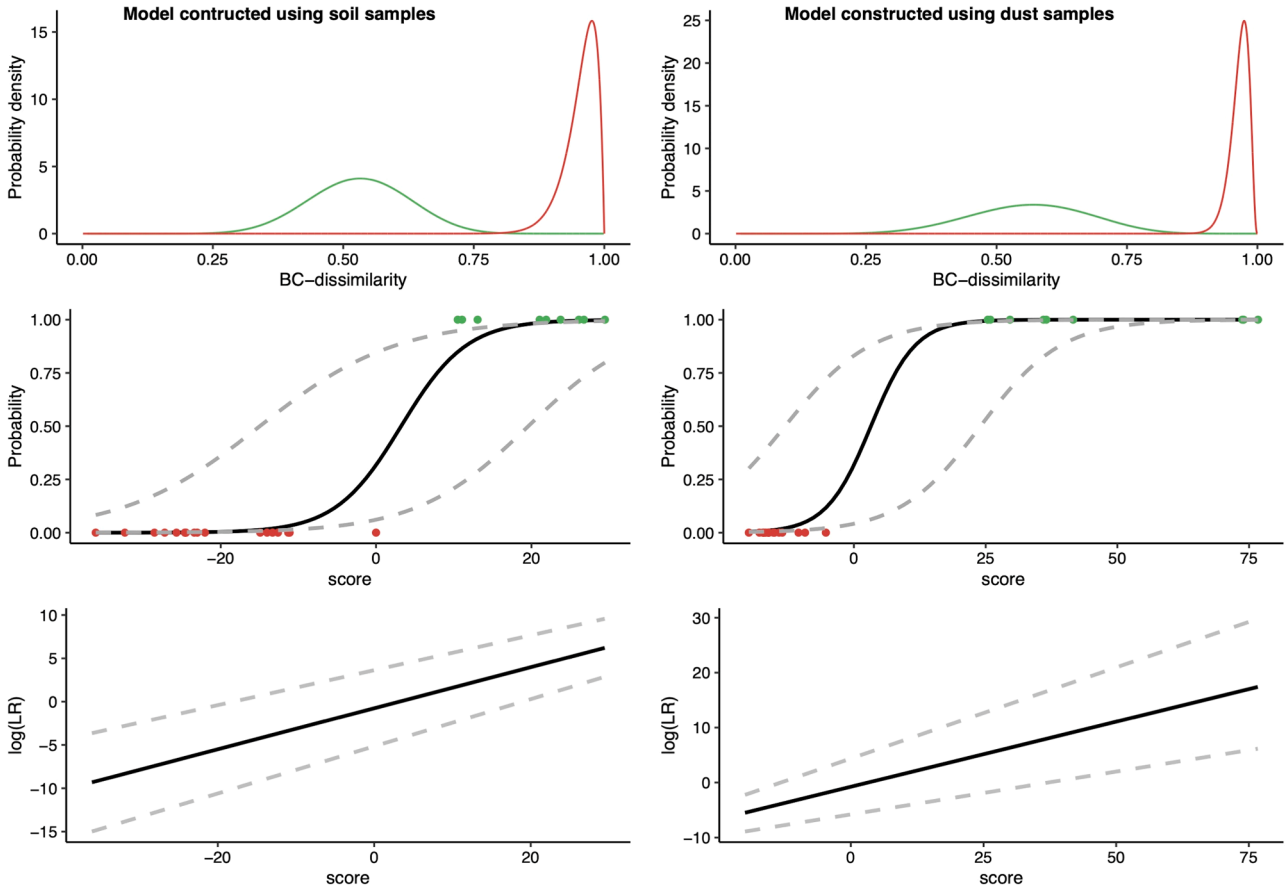


Fig. 4. Conversion of Bray-Curtis dissimilarity values to log-likelihood ratios for bacterial communities. Beta distributions are shown (top) between same site and different site distributions. Scores are plotted using Logistic Regressions (middle) and converted to log-likelihood ratios (bottom). Two models were constructed using either soil samples (left) or dust samples from week 12 of the study (right). Colors represent samples from the same site (green) and different site (red) and the dashed lines are the 95% confidence intervals.

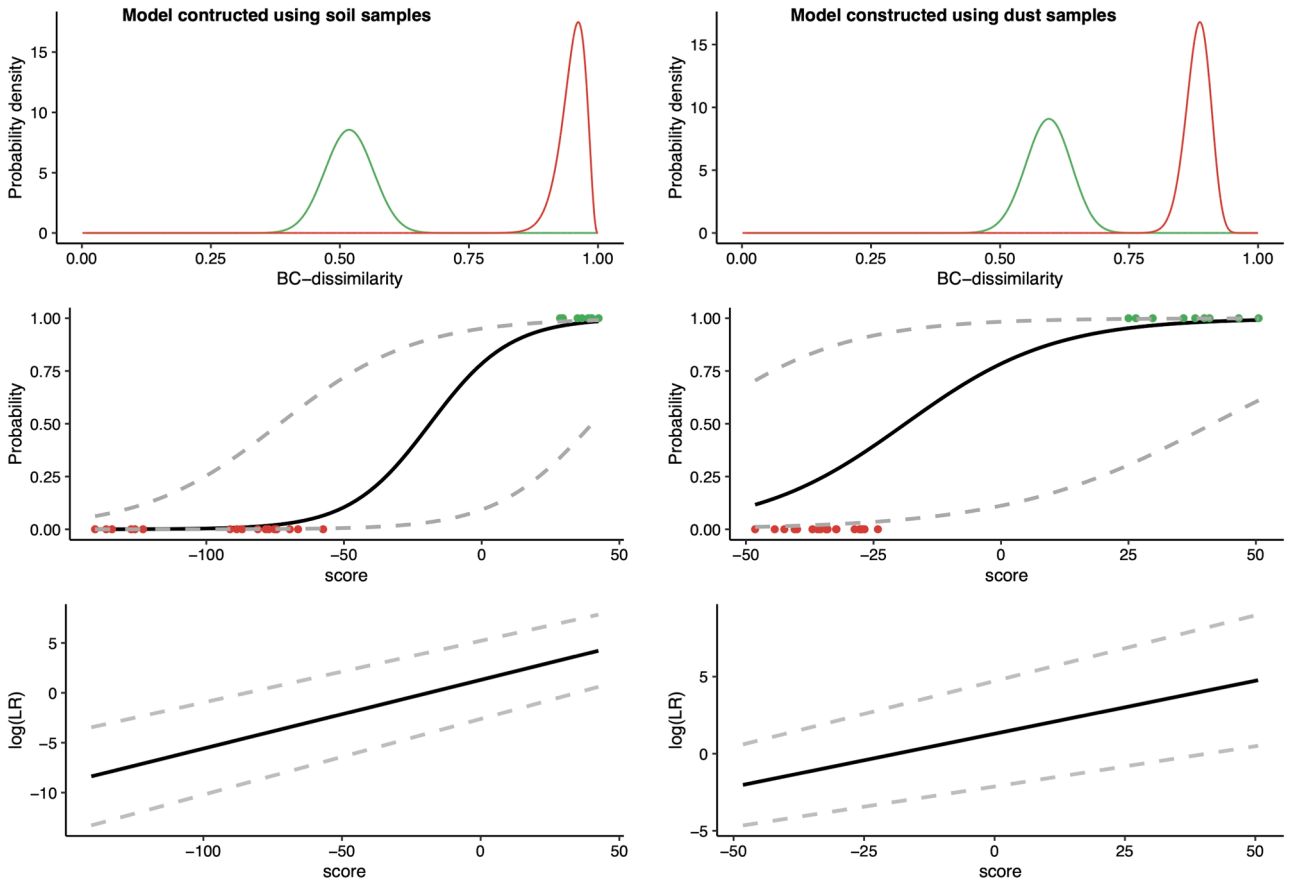


Fig. 5. Conversion of Bray-Curtis dissimilarity values to log-likelihood ratios for fungal communities. Beta distributions are shown (top) between same site and different site distributions. Scores are plotted using Logistic Regressions (middle) and converted to log-likelihood ratios (bottom). Two models were constructed using either soil samples (left) or dust samples from week 12 of the study (right). Colors represent samples from the same site (green) and different site (red) and the dashed lines are the 95% confidence intervals.

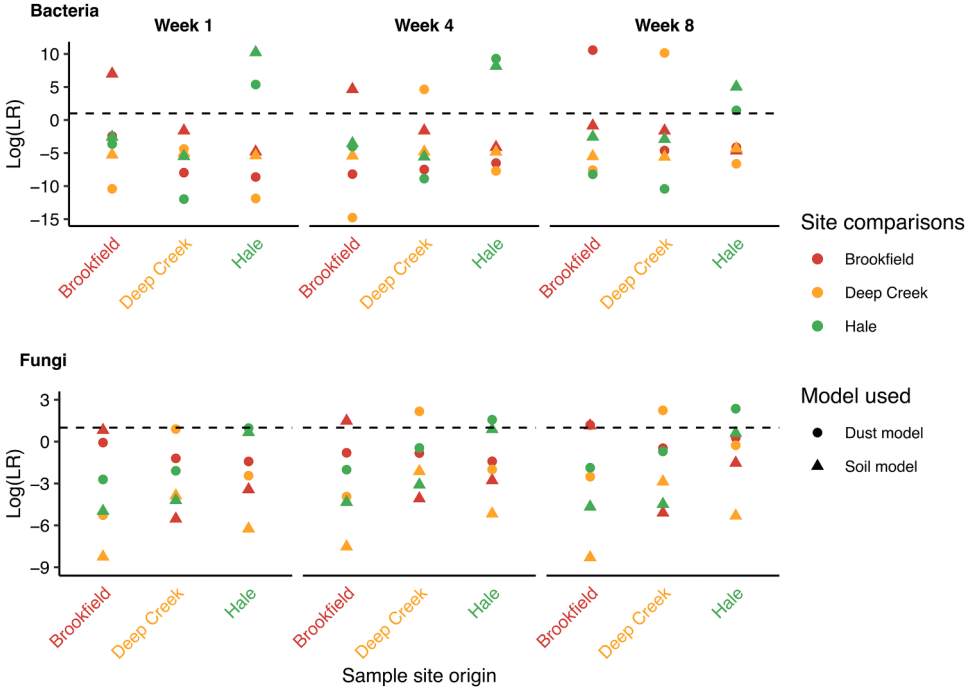


Fig. 6. Provenance testing for dust samples collected in week 1,4 and 8 (N = 9) of this study using soil and dust models for both bacteria (top) and fungi (bottom). Colors indicate sites and shapes indicate the type of model used. The dashed line marks a log-likelihood ratio of 1 where a value > 1 supports that the sample came from the site of interest (true origin) and < 1 indicates the sample came from a different site.

model predicted samples were from the site of interest (LR > 1) better in week 1 (67% correct predictions) but the dust model did a better job in week 4 and 8 (67%, and 100% correct predictions respectively), similarly, fungal communities were better predicted by the dust model in weeks 4 and 8 (67% and 100% correct predictions respectively) and neither model was a good predictor of origin in week 1 (zero samples were predicted correctly) (Fig. 6).

5. Discussion

This proof-of-concept study has furthered our understanding of whether dust can be employed as a forensic intelligence tool. We found dust samples from different sites contained significantly different communities of bacteria and fungi. However, we documented within site variability which could not be attributed to the length of time the items spent at a location accumulating dust – as differences in community composition did not vary successively with time but were random. This variation could be explained by the differences in species richness that occurred spatially with sampling, but due to the small sample size of this study, more testing is required to understand the influence of time and spatial heterogeneity on biological community composition of dust, as well as other potential contributing factors. This within site variation, however, was not greater than between sites such that sites still clustered separately in ordinations and could be discriminated. We did find that this variation influenced the ability to predict dust provenance using a forensic framework and employing the use of a calibrated log-LRs. We found provenance predictions did not always accurately predict that a questioned dust sample was from the site of origin leading to false negative predictions (e.g., no site of origin predicted), but we did not document any false positives (e.g., an incorrect site of origin predicted). We also demonstrated that both biological and chemical properties of dust can be generated from a single swabbed sample and yield discriminatory results between sites, however, these may not be comparable to reference material from the same site, given the differences in sample material (dust to DNA extraction pellet).

5.1. Can chemical and biological properties be obtained from the same dust swab?

In addition to biological properties of dust, elemental/mineralogical properties can be used to build additional evidence for dust sample provenance. Undertaking these two analyses on a single swabbed sample is important to assess whether this is possible in situations when sample material is limited and if so, whether the results are useful. From a swabbed dust sample that had undergone the first step of the DNA extraction process, we were able to document the presence and absence of multiple minerals and the percentage composition of multiple elements. These results proved congruent to the eDNA analysis in that sites could be discriminated based on ordinations. We were able to overcome signal interference from both the swab and the initial DNA buffer to generate elemental and mineralogical information. However, due to the small sample size (being the leftover pellet of a DNA extraction), we could not quantify absolute elemental composition. In addition, when we compared these results to reference samples from the same site, which had also been standardised by elemental composition, we found significant differences between elements recovered from the DNA extraction pellet and those recovered directly from dust material (soil sample sieved to < 75 μm). We also noted different minerals were observed, which overall makes predicting dust sample origin difficult using existing geochemical reference databases or comparing to the suspected site of origin. Despite these limitations, the information obtained from the pellet of the DNA extraction can still provide insight into the environmental conditions of the site where the dust was collected [37]. This can then be applied alongside metabarcoding data to support sample origin predictions using already established predictive frameworks [38,39]. Further, there may be some dependence between element and mineral composition of soils and detected communities of bacteria and fungi, given the associations between microbiome and soil properties [40], which could be applied to verify metabarcoding data if these relationships are quantified. The successful analysis of a dust sample for element and mineral composition, that has undergone DNA extraction, could be important in situations when sample quantity is limited. However, more studies need to be conducted to fully ascertain the uncertainties between the pellet from the DNA extraction and reference material from the same site.

5.2. Does the ability to determine dust sample origin using metabarcoding change with the length of time spent at a location?

Biological profiles of dust samples were shown to be significantly different between sites for both bacteria and fungi, highlighting the site discrimination capabilities of metabarcoding dust samples. The impact of sampling time on community composition was less clear. We observed differences in community composition between different time points sampled within the same sites, but this was not consistent between sites and occurred randomly. It is possible changes in bacteria and fungi are driven by factors other than time spent in a location. We compared rainfall patterns over sampling times (Supplementary Fig. S1) but found that this is not a likely explanation for the differences observed, as high rainfall events occurred prior to sampling in weeks 4 and 12 but not 8. We also investigated airstreams using HYSPLIT trajectories [41], noting airstream changes in weeks 8 and 12 compared to 1 and 4, which may only explain the differences observed for the bacteria profiles at the Brookfield site where we observed reduced species richness for weeks 8 and 12 compared to 1 and 4. Overall, we expected to see a closer relationship between dust and soil metabarcoding data at week 1 and more discrepancy as time went on due to DNA degradation, weather and other environmental factors that influence bacteria and fungi communities, as observed in [14], instead we saw that this was highly variable.

Other factors are likely driving changes in composition of bacteria and fungi communities, such as spatial heterogeneity in our sampling. Species richness was shown to be significantly different between sampling boards within samples of the same site and time period. This points to there being small scale diversity differences within the same site [42] but could have also been an artifact of board placement, where some boards may have been better positioned to collect more dust leading to higher species richness. The observed differences in species richness between time periods was only observed in bacteria, showing that these communities may be more variable over time than fungi. We also observed differences in species richness between soil and dust samples which is likely due to sampling material type (soil vs. dust). Overall, both species richness and community diversity across our different sites show no clear pattern that time has influenced community composition but rather, within site variability may play a bigger role. We did not, however, investigate seasonal influences on these biological communities and testing these may prove to influence community composition of bacteria and fungi when left for longer time periods.

5.3. Can we integrate dust metabarcoding into forensic casework and use soil samples as predictors for dust sample origin?

For both bacteria and fungi communities, we found provenance estimates were more accurate for dust models and less accurate for soil models, where this did vary slightly between sites. This finding aligns with the community composition differences observed between soil and dust above. However, model development could have also influenced this result whereby the soil model was based on samples collected in week = 0 and the dust model based on samples collected in week = 12. This then could bias results towards the soil model performing better in week 1 and the dust model better in week 8, but this is a trend we see only for bacteria not fungi – fungi were better predicted using the dust model as the soil model did not predict sample origin correctly in week 1

(Fig. 6). Bacteria communities may be more influenced by the temporal differences between the sample in question and the reference material used to construct the model, and fungi communities may be more influenced by whether the dust or soil fraction is used to predict origin. This means dust provenance may not be predicted well using existing soil metabarcoding databases for fungi communities, similarly, this may be also the case for bacteria communities given differences in sampling time between the dust from a crime scene and soil data deposited in reference databases. While this is unfortunate given the plethora of samples available in soil metabarcoding data repositories, dust may still be useful in forensic intelligence. If the variation between soil and dust is due to temporal differences, then soil samples could still be collected from crime scenes and compared to dust from items of interest where this may work better for bacteria than fungi. The closer in time samples are collected to the committed crime may improve the ability to assign correct sample origins. If the variation is due to differences in material type, then dust samples could be collated as references instead of soil samples to improve consistency between dust and soil sample diversity. Fungal community diversity measures from dust could also be improved by increasing the sequencing depth as we implemented 2×150 bp sequence chemistry in this study which may not have captured the variable length of ITS regions for all fungi. Improving the length of sequencing would increase the paired-end read overlap leading to more confidence in read alignment and possibly improving sequence assignment and model development.

While we were able to develop a predictive framework to ascertain dust sample provenance, we acknowledge there are limitations in our model design which could have influenced the predictive capacity of our models and conclusions above. We used nine samples to construct both models (bacteria and fungi) and the same site models (Figs. 4 and 5, top graph), were less reliable given the within site variation in species richness. This is supported by the fact our log-LR values, while > 1, do not provide strong support for the same origin hypotheses as many log-LR values are < 10 [15], especially within the fungal communities. Ultimately, within site variation in bacteria and fungi communities likely influenced whether dust samples were correctly predicted to be from their site of interest (origin) (Hp) or a different site (Hd), but did not influence whether the sites were statistically different based on BC dissimilarities. This highlights that distance-based analyses alone do not necessarily reflect the same results for sample provenance that a log-LR framework does. This is an important finding if provenance estimates based on metabarcoding are to advance from simple ordinations to casework [4]. Despite the limitations outlined above, our framework has shown that applying soil or dust samples to determine dust provenance will not overstate the strength of the evidence. If the community profiles between soil and dust samples have diverged, either due to time or stochasticity in the collection or analysis, this will not point to an incorrect site (false positive), but rather will not confidently assign provenance to any site (false negative). This study is a pilot experiment, based on three sites and a small sampling size and as such, we propose further testing prior to ruling out applying dust for use in forensic intelligence. We have highlighted how dust metabarcoding data can be integrated into forensic casework using our developed model, where sites of interest (crime scene) and alternative sites (alibi locations) can be compared to samples of unknown origin. This provides a more intuitive interpretation of the data than ordinations and can be applied in a more extensive study.

6. Conclusions

This proof-of-concept study found that metabarcoding of dust samples can generate bacteria and fungi community profiles that are unique to sites. We found that these profiles did not vary consistently with sampling time as we expected, but instead varied randomly which could be due to the within site variability rather than time spent at a location. This within site variability, while not greater than between sites, is likely to have contributed to incorrect provenance assignments across tested samples using a calibrated log-likelihood framework. We did find elemental and mineralogical analyses can be conducted on the same swabbed sample as eDNA analysis, meaning these could be used to assist dust provenance estimates. However, the differences in elemental and mineral compositions between samples (extraction pellet) and reference material (dust fraction of soil), may prevent the use of this approach in provenance assignment and could be better applied to generate information around environmental conditions of a site. Beyond conducting a more extensive study with more sites and replicates, a further avenue of work in this field is to investigate samples containing dust from multiple locations. Real world samples will often contain dust that is a mixture of multiple locations and disentangling this to predict sample origin is important. Authors in [43] were able to show the possibility of this using bacterial profiles of soil samples and [6] looked at detecting sample origin from artificial mixtures of dust, but this is yet to be tested in a real-world scenario. While there is still a long way to go in this field, both metabarcoding and biogeochemical analyses of dust samples show potential for applications in forensic science.

Declaration of Competing Interest

None declared.

Acknowledgements

The authors would like to thank the funding and support provided by the Defence Innovation Partnership through a collaborative research fund awarded to Dr. Jennifer Young [InFoDust project]. We would like to thank Denny Murphy and Belinda Martin for their assistance in the field and the Terrestrial Ecosystem Research Network (TERN) for providing site data. Patrice de Caritat publishes with the permission of the Chief Executive Officer, Geoscience Australia. The authors acknowledge the South Australian Genomics Centre which provided sequencing services. The SAGC is supported by the National Collaborative Research Infrastructure Strategy (NCRIS) via BioPlatforms Australia and by the SAGC partner institutes. The authors would also like to thank the reviewers who contributed a significant amount of time and provided helpful comments to improve this work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2023.102931.

References

[1] J.S. Allwood, N. Fierer, R.R. Dunn, The future of environmental DNA in forensic science, Appl. Environ. Microbiol. 86 (2) (2020) e01504–19.
[2] D. Garlapati, B. Charankumar, K. Ramu, P. Madeswaran, R. Murthy, A review on the applications and recent advances in environmental DNA (eDNA) metagenomics, Rev. Environ. Sci. Biotechnol. 18 (3) (2019) 389–411.
[3] K. Deiner, H.M. Bik, E. Mächler, M. Seymour, A. Lacoursière-Roussel, F. Altermatt, S. Creer, I. Bista, D.M. Lodge, N. de Vere, Environmental DNA metabarcoding: transforming how we survey animal and plant communities, Mol. Ecol. 26 (21) (2017) 5872–5895.
[4] J.M. Young, A. Linacre, Massively parallel sequencing is unlocking the potential of environmental trace evidence, Forensic Sci. Int Genet 50 (2021), 102393.
[5] A. Barberan, J. Ladau, J.W. Leff, K.S. Pollard, H.L. Menninger, R.R. Dunn, N. Fierer, Continental-scale distributions of dust-associated bacteria and fungi, Proc. Natl. Acad. Sci. USA 112 (18) (2015) 5756–5761.
[6] J.S. Allwood, N. Fierer, R.R. Dunn, M. Breen, B.J. Reich, E.B. Laber, J. Clifton, N. S. Grantham, S.A. Faith, Use of standardized bioinformatics for the analysis of fungal DNA signatures applied to sample provenance, Forensic Sci. Int 310 (2020), 110250.
[7] P. De Deckker, C.I. Munday, J. Brocks, T. O'Loingsigh, G.E. Allison, J. Hope, M. Norman, J.-B.W. Stuut, N.J. Tapper, S. van der Kaars, Characterisation of the major dust storm that traversed over eastern Australia in September 2009; a multidisciplinary approach, Aeolian Res 15 (2014) 133–149.
[8] D. Pirrie, L. Dawson, G. Graham, Predictive geolocation: forensic soil analysis for provenance determination, Episodes, Int. J. Geosci. 40 (2) (2017) 141–147.

[9] P. Sangwan, T. Nain, K. Singal, N. Hooda, N. Sharma, Soil as a tool of revelation in forensic science: a review, Anal. Methods 12 (43) (2020) 5150–5159.
[10] N.S. Grantham, B.J. Reich, K. Pacifici, E.B. Laber, H.L. Menninger, J.B. Henley, A. Barberan, J.W. Leff, N. Fierer, R.R. Dunn, Fungi identify the geographic origin of dust samples, PLoS One 10 (4) (2015), e0122605.
[11] E.M. Jesmok, J.M. Hopkins, D.R. Foran, Next-generation sequencing of the bacterial 16S rRNA gene for forensic soil comparison: a feasibility study, J. Forensic Sci. 61 (3) (2016) 607–617.
[12] N. Damaso, J. Mendel, M. Mendoza, E.J. von Wettberg, G. Narasimhan, D. Mills, Bioinformatics approach to assess the biogeographical patterns of soil communities: the utility for soil provenance, J. Forensic Sci. 63 (4) (2018) 1033–1042.
[13] K.M. Ruppert, R.J. Kline, M.S. Rahman, Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA, Glob. Ecol. Conserv 17 (2019), e00547.
[14] A.J. Badgley, E.M. Jesmok, D.R. Foran, Time radically alters Ex Situ evidentiary soil 16S bacterial profiles produced via next-generation sequencing, J. Forensic Sci. 63 (5) (2018) 1356–1365.
[15] I. ENoFS, ENFSI guideline for evaluative reporting in forensic science. Strengthening the Evaluation of Forensic Results Across Europe (STEOFRAE), European Network of Forensic Sciences Institutes, The Hague, Netherlands, 2015.
[16] S. Uitdehaag, F. Quaak, I. Kuiper, Soil comparisons using small soil traces, a case report. Soil in Criminal and Environmental forensics, Springer, 2016, pp. 61–69.
[17] G.S. Morrison, Tutorial on logistic-regression calibration and fusion:converting a score to a likelihood ratio, Aust. J. Forensic Sci. 45 (2) (2013) 173–197.
[18] J.M. Robinson, C. Cando-Dumancela, C. Liddicoat, P. Weinstein, R. Cameron, M. F. Breed, Vertical stratification in urban green space aerobiomes, Environ. Health Perspect. 128 (11) (2020), 117008.
[19] P. Kanokwongnuwut, P. Kirkbride, A. Linacre, Visualising latent DNA on swabs, Forensic Sci. Inter 291 (2018) 115–123.
[20] W. Walters, E.R. Hyde, D. Berg-Lyons, G. Ackermann, G. Humphrey, A. Parada, J. A. Gilbert, J.K. Jansson, J.G. Caporaso, J.A. Fuhrman, A. Apprill, R. Knight, Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys, mSystems 1 (1) (2016).
[21] J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, P.D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform, Appl. Environ. Microbiol. 79 (17) (2013) 5112–5120.
[22] B.J. Callahan, P.J. McMurdie, S.P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis, ISME J. 11 (12) (2017) 2639–2643.
[23] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021.
[24] P.J. McMurdie, S. Holmes, Waste not, want not: why rarefying microbiome data is inadmissible, PLoS Comput. Biol. 10 (4) (2014), e1003531.
[25] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Res. 41 (D1) (2012) D590–D596.
[26] K. Abarenkov, A. Zirk, T. Piirmann, R. Pöhönen, F. Ivanov, R. Nilsson, U. Kõljalg, UNITE QIIME release for Fungi 2. Version 10.05. 2021, UNITE Community 7 (2021) 1264763.
[27] P.J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, PloS One 8 (4) (2013), e61217.
[28] N.M. Davis, D.M. Proctor, S.P. Holmes, D.A. Relman, B.J. Callahan, Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data, Microbiome 6 (1) (2018) 226.
[29] J. Oksanen, R. Kindt, P. Legendre, B. O'Hara, M.H.H. Stevens, M.J. Oksanen, M. Suggests, The vegan package, Community Ecol. Package 10 (631–637) (2007) 719.
[30] M.J. Anderson, D.C. Walsh, K. Robert Clarke, R.N. Gorley, E. Guerra-Castro, Some solutions to the multivariate Behrens–Fisher problem for dissimilarity-based analyses, Aust. N. Z. J. Stat. 59 (1) (2017) 57–79.
[31] D. Jones, The Fathom Toolbox for Matlab: multivariate ecological and oceanographic data analysis. College of Marine Science, University of South Florida, St. Petersburg, FL, 2012.
[32] S.M. Toolbox, Matlab, Mathworks Inc, 1993.
[33] N.R. Foster, B. Martin, J. Hoogewerff, M.G. Aberle, P. de Caritat, P. Roffey, R. Edwards, A. Malik, P. Thwaites, M. Waycott, J. Young, The utility of dust for forensic intelligence: Exploring collection methods and detection limits for environmental DNA, elemental and mineralogical analyses of dust samples, Forensic Sci. Int 344 (2023), 111599.
[34] R.A. Rigby, D.M. Stasinopoulos, Generalized additive models for location, scale and shape, J. R. Stat. Soc. Ser. C. Appl. Stat. 54 (3) (2005) 507–554.
[35] A. Gelman, J. Hill, Data Analysis Using Regression and Multilevel/hierarchical Models, Cambridge University Press, 2006.
[36] K.R. Clarke, Non-parametric multivariate analyses of changes in community structure, Australian, J. Ecol. 18 (1) (1993) 117–143.
[37] X. Bi, S. Liang, X. Li, A novel in situ method for sampling urban soil dust: particle size distribution, trace metal concentrations, and stable lead isotopes, Environ. Pollut. 177 (2013) 48–57.
[38] C. Flojgaard, T.G. Froslev, A.K. Brunbjerg, H.H. Bruun, J. Moeslund, A.J. Hansen, R. Ejrnaes, Predicting provenance of forensic soil samples: linking soil to ecological habitats by metabarcoding and supervised classification, PLoS One 14 (7) (2019), e0202844.
[39] P. de Caritat, B. Woods, T. Simpson, C. Nichols, L. Hoogenboom, A. Ilheo, M. G. Aberle, J. Hoogewerff, Forensic soil provenancing in an urban/suburban setting: a simultaneous multivariate approach, J. Forensic Sci. 67 (3) (2022) 927–935.
[40] C.L. Lauber, M.S. Strickland, M.A. Bradford, N. Fierer, The influence of soil properties on the structure of bacterial and fungal communities across land-use types, Soil Biol. Biochem 40 (9) (2008) 2407–2415.
[41] A. Stein, R.R. Draxler, G.D. Rolph, B.J. Stunder, M. Cohen, F. Ngan, NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull Amer Meteor, Society 96 (12) (2015) 2059–2077.
[42] C.A. Macdonald, R. Ang, S.J. Cordiner, J. Horswell, Discrimination of soils at regional and local levels using bacterial and fungal T-RFLP profiling, J. Forensic Sci. 56 (1) (2011) 61–69.
[43] S. Demaneche, L. Schauser, L. Dawson, L. Franqueville, P. Simonet, Microbial soil community analyses for forensic science: application to a blind test, Forensic Sci. Int 270 (2017) 153–158.