Forensic Science International: Genetics 13 (2014) 176–184

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic soil DNA analysis using high-throughput sequencing: A comparison of four molecular markers

Jennifer M. Young *, Laura S. Weyrich, Alan Cooper

Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, 5005, Australia

ARTICLE INFO

Article history:
Received 8 May 2014
Received in revised form 25 July 2014
Accepted 28 July 2014

Keywords:
Soil
DNA
Forensic
Metagenomics
High-throughput sequencing (HTS)
DNA metabarcoding

ABSTRACT

Soil analysis, such as mineralogy, geophysics, texture and colour, are commonly used in forensic casework to link a suspect to a crime scene. However, DNA analysis can also be applied to characterise the vast diversity of organisms present in soils. DNA metabarcoding and high-throughput sequencing (HTS) now offer a means to improve discrimination between forensic soil samples by identifying individual taxa and exploring non-culturable microbial species. Here, we compare the small-scale reproducibility and resolution of four molecular markers targeting different taxa (bacterial 16S rRNA, eukaryotic18S rRNA, plant *trn*L intron and fungal internal transcribed spacer I (ITS1) rDNA) to distinguish two sample sites. We also assess the background DNA level associated with each marker and examine the effects of filtering Operational Taxonomic Units (OTUs) detected in extraction blank controls. From this study, we show that non-bacterial taxa in soil, particularly fungi, can provide the greatest resolution between the sites, whereas plant markers may be problematic for forensic discrimination. ITS and 18S markers exhibit reliable amplification, and both show high discriminatory power with low background DNA levels. The 16S rRNA marker showed comparable discriminatory power post filtering; however, presented the highest level of background DNA. The discriminatory power of all markers was increased by applying OTU filtering steps, with the greatest improvement observed by the removal of any sequences detected in extraction blanks. This study demonstrates the potential use of multiple DNA markers for forensic soil analysis using HTS, and identifies some of the standardisation and evaluation steps necessary before this technique can be applied in casework.

1. Introduction

In forensic science, soil can serve as powerful contact trace evidence because it is highly individualistic, has a high transfer and retention probability, and can be characterised using a number of techniques [1–4]. Standard analyses examine intrinsic properties of soil including mineralogy, geophysics, texture and colour [1–3,4]. However, soils can support a vast amount of organisms, which can be examined using DNA analysis. Soil DNA fingerprinting methods (reviewed in [5]) have been used in court to link a suspect to a crime scene [6]. However, many previous genetic approaches have relied on patterns of fragment length variation produced by amplification of unidentified microbial taxa in the soil extract and, these provide little resolution of spatially and temporally variable microbial communities [7,8]. In contrast, recent developments in massively-parallel high throughput DNA sequencing (HTS) technologies [9] now provide the ability to rapidly generate a detailed picture of soil microbial communities. Techniques such as DNA metabarcoding, where conserved genetic loci are used to simultaneously amplify and characterise a broad diversity of taxa, can drastically increase the power and resolution to distinguish soil samples, quickly identifying both microbial and non-microbial taxa from thousands of sequences produced in a single analysis [9–12].

For forensic soil discrimination, an ideal marker should target sequences endemic to the site of interest and thus isolate the most discriminative sequences. DNA metabarcoding [13,14] can be used to explore and isolate soil communities by targeting specific taxonomic groups, with the most commonly utilised markers in environmental samples being 16S ribosomal RNA (rRNA) gene region (bacteria), internal transcribed spacer I (ITS1) (fungi), 18S rRNA gene region (eukaryotes), and chloroplast *trn*L intron (plants). Previous comparison of bacterial and fungal DNA fingerprints demonstrated that fungi markers were more robust for soil forensic discrimination [15], and more resistant to biological, chemical and mechanical degradation than bacteria [16]. However, the

* Corresponding author. Tel.: +61 8 8313 8245.
E-mail address: jennifer.young@adelaide.edu.au (J.M. Young).

http://dx.doi.org/10.1016/j.fsigen.2014.07.014
1872-4973/

Table 1
Details of the four molecular markers used in the comparison.

Marker	Primer	Primer sequence (5′–3′)	Length (bp)	Target community
*trn*L	*trn*Lc	CGAAATCGGTAGACGCTACG		
	*trn*Lh	CCATTGAGTCTCTGCACCTATC	150	Plant
16S	341F	CCTACGGGAGGCAGCAG		
	518R	ATTACCGCGGCTGCTGG	150	Bacteria
ITS	ITS5	GGAAGTAAAAGTCGTAACAAGG		
	5.8S_fungi	CAAGAGATCCGTTGTTGAAAGTT	250	Fungi
18S	1391F Euk	GTACACACCGCCCGTC		
	EukBr	TGATCCTTCTGCAGGTTCACCTAC	200	Eukaryotes

discriminatory power of the common DNA metabarcodes using HTS has not yet been examined from a forensic casework perspective.

Reproducibility is a key area of concern when utilising HTS in forensic casework. Although DNA metabarcoding and HTS show potential for forensic science [17], strict validation standards must first be met [18], which includes demonstrating reproducibility, robustness, and reliability. Many steps of this approach can introduce bias [19,20]; both within a sample (DNA extraction or PCR amplification bias) and between samples collected at a specific site (sampling bias) because taxonomic groups can vary spatially even within a single site. The inexpensive methods and high sequence coverage offered by HTS approaches now provide the opportunity to test these biases, as multiple samples can be rapidly and efficiently processed in parallel [13,21,22]. Porter et al. recently examined the small-scale reproducibility of two soil core samples using HTS sequencing of mitochondrial 12S rDNA (vertebrates), 16S rDNA (Actinobacteria), mitochondrial COX1 (insects, amphibians) and chloroplast *trn*L intron (plants), and conclude that the *trn*L intron is the most reproducible, and COX1 the least reproducible, marker [23]. However, this study only used two samples from each core that were located 10 cm apart, so it is not clear how spatial variation or extraction and PCR bias were assessed. The reproducibility of markers and examination of spatial variation within different target taxa is crucial to ensure adequate sampling and representative diversity has been obtained.

In this study, we examined the potential of DNA metabarcoding and HTS for forensic soil analysis utilising four well-established metabarcoding markers: chloroplast *trn*L intron, ITS1, 16S rRNA and 18S rRNA (Table 1). We compared the discriminatory power of each marker and demonstrated the effects of removing the OTUs detected in the extraction blank controls from all DNA extracts. We also assessed each marker based on small-scale reproducibility and background DNA level. Based on these results, we have provided recommendations on the most promising markers and sampling methods for effective and reproducible forensic soil DNA analysis.

2. Materials and methods

2.1. Sample sites and DNA extraction

Two sites near Adelaide, South Australia were selected on the basis of contrasting soil properties (Fig. 1: site 1, Barker Inlet Wetland; site 2, Tennyson Dunes; 14 km distance between sites). Barker Inlet Wetland was characterised by dark, organic rich, saline soil and mangroves, whereas Tennyson Dunes was a coastal site characterised by sandy soil with low water, nutrient and organic content. At each site, four soil samples were collected 1 m apart along a 3 m transect using sterile 15 ml CELLSTAR® tubes. All samples were collected from the topsoil layer (0–2 cm) and were stored at 4 °C prior to DNA extraction within 24 h of collection. Prior to DNA extraction, visible plant roots were removed from the samples using sterile tweezers. From each sample, DNA was extracted from two 250 mg subsamples of starting material (Fig. S1) using the PowerSoil DNA Isolation kit (MOBIO, Carlsbad, CA, USA). All wetland samples were extracted separately from the coastal samples and an extraction blank control (EBC) was processed alongside the samples for each site. The DNA yield of each extract was quantified using the Nanodrop2000 Spectrometer (Thermo Scientific) by taking an average of two measurements.

2.2. PCR amplification and library preparation

DNA extracts and extraction blank controls were amplified using four metabarcoding markers: *trn*L intron, internal transcribed spacer (ITS1), 18S rRNA and 16S rRNA (Table 1). All primers were modified to include Ion Torrent sequencing adapters and a unique multiplex identifier (MID) tag [24] for each extract. PCR amplifications were performed in a 25 μl reaction mix containing 2.5 mM $MgCl_2$, 0.24 mM dNTPs, 0.24 μm of each primer, 0.4 mg/ml bovine serum albumin (0.2 M Betaine was used for *trn*L), 0.5 U Amplitaq Gold DNA polymerase in 10× reaction buffer (Applied Biosystems, Melbourne Australia), and 1 μl DNA extract. PCR amplification protocol for *trn*L and ITS was 9 mins at 94 °C, followed by 35 cycles of 94 °C for 30 s, 54 °C for 30 s, and 72 °C for 45 s, and a final extension at 72 °C for 7 min. For 18S and 16S, the same protocol was used, although the annealing steps were modified to 62 °C for 20s and 60 °C for 20s for 18S and 16S, respectively. PCR amplifications were performed in triplicate and pooled to minimise PCR bias, and a no-template control (PCR negative) was included for each MID tag. 2 μl of PCR product were subjected to agarose gel electrophoresis. PCR products were purified using Agencourt AMPure XP PCR Purification kit (Beckman Coulter Genomics, Lane Cove, NSW, Australia) and quantified using the HS dsDNA Qubit Assay on a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Purified PCR products were pooled to equimolar concentration and the library was diluted to 11 pM. Emulsion PCR and Ion Sphere Particle enrichment were performed on the Ion One Touch™ system (Life Technologies) using the Ion OneTouch™ 200 Template Kit v2 DL, before sequencing on the Ion Torrent Personal Genome Machine™ (Life Technologies) using the Ion PGM™ 200 Sequencing Kit and an Ion 316™ semiconductor chip (Life Technologies).

2.3. Data analysis

Base calling was performed using Torrent Suite v3.4.2 (Life Technologies) and the resulting reads were de-multiplexed using the sample MID tags, before the MID sequences were removed with the fastx_barcode_splitter tool (FASTX-toolkit v0.0.12; http://hannonlab.cshl.edu/fastx_toolkit). Cutadapt v1.1 [25] was subsequently used to trim primers from the raw reads using a strict zero mismatch threshold. Additionally, fastx_clipper tool was used to remove reads <100 bp (*parameters; –Q33 –l 100*), and reads with a Phred score less than 20 for 90% of the sequence were also removed using the fastq_quality_filter tool. The number of reads retained following each data filtering step has been summarised in Table S1. Filtered sequences were imported into QIIME (v.1.5.0.)[26], where

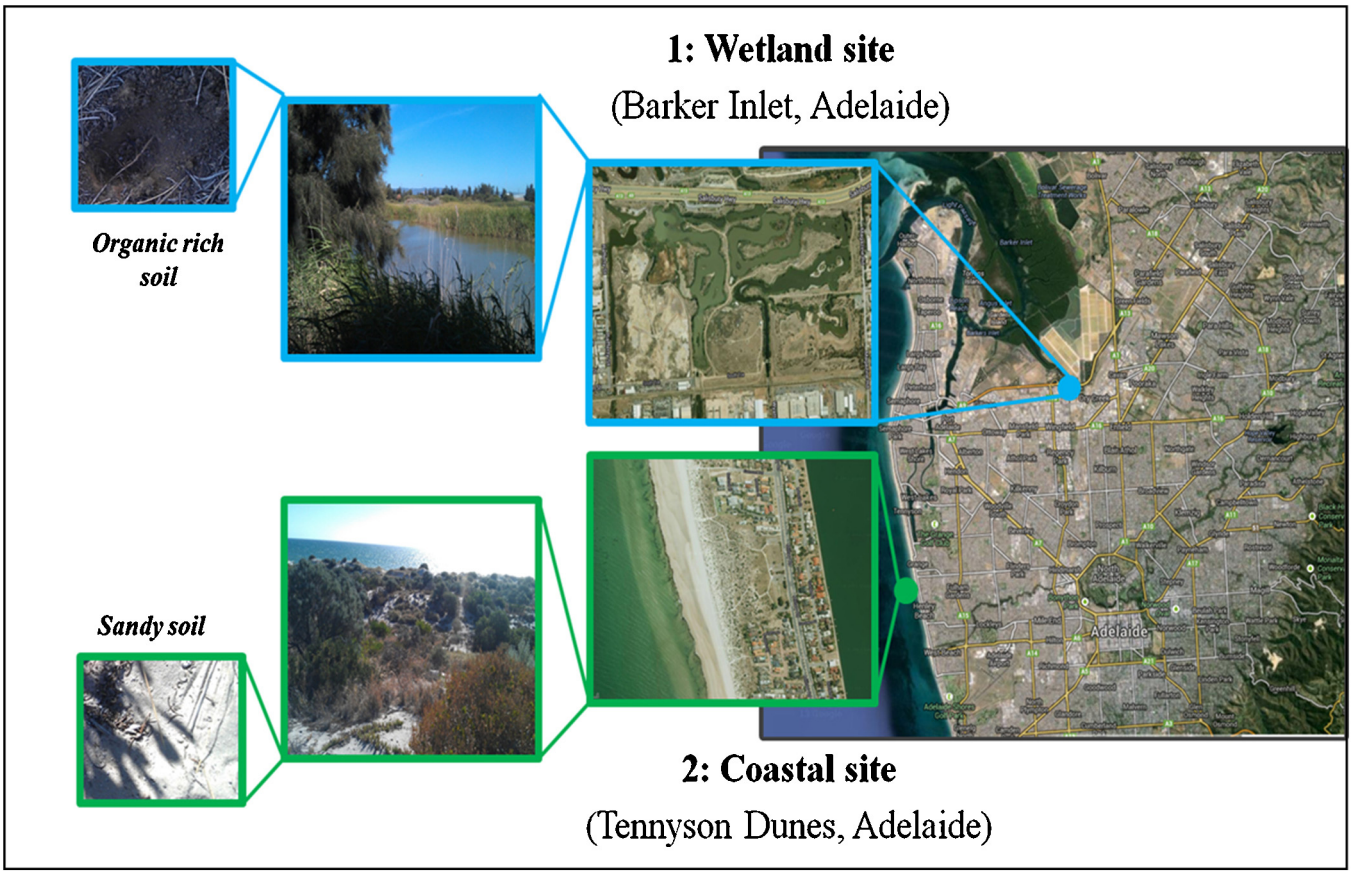

Fig. 1. Location of the two sites used in this study. At each site (1: wetland, 2: coastal), four soil samples were collected (1 m apart), and two DNA extracts were obtained from each sample.

similar sequences were binned into Operational Taxonomic Units (OTUs) using *de novo* UCLUST [27] at 97% similarity cut-off. The most abundant sequence was selected as the representative sequence for each OTU. For each marker, the data was rarefied to ensure an even read coverage for each sample and minimise skewed diversity estimates; *trn*L: 1097 sequences, ITS: 2980 sequences, 18S: 3000 sequences, 16S: 3000 sequences.

The rarefied OTU tables were used to examine the diversity within each sample (alpha diversity) and between different samples (beta diversity). The number of OTUs detected in each extract was calculated to compare the diversity detected by each marker, and examine habitat effects on the diversity of each target taxa. Beta diversity was visualised using Bray–Curtis (BC) distance cluster dendrograms and MDS plots, generated in PRIMER6 (*PRIMER-E Ltd.*, Luton, UK) with default parameters. For each marker, ANOSIM (*Analysis of Similarity*) was used to test for statistical differences in diversity between the samples from the two sites. The power of each marker to discriminate between the two sites was compared using Bray–Curtis dissimilarity (%) based on OTU abundance; a high dissimilarity value shows pronounced variation in OTU composition between the sites. We tested the effect of two additional filtering steps on resolution between sites: [1] removal of any OTUs detected in extraction blank controls to exclude any biological signal from laboratory reagents or contaminants and [2] removal of any OTUs common to all DNA extracts to enhance differences between samples. The Bray–Curtis dissimilarity value (%) was then calculated for each marker following each filtering step, and a second set of cluster dendrograms generated from the resulting OTU tables. To determine which marker provides the most reproducible OTU composition, the mean percent similarity (100-BC) was calculated from pair-wise comparisons of (a) duplicate extracts to assess within sample variation and (b) replicate samples (1 m apart) to assess within site variation. A reproducible OTU composition is indicated by a high percent similarity, and statistical differences in percent similarity and dissimilarity were performed in SPSS (SPSS Inc. Released 2009. PASW Statistics, Version 18.0. Chicago: SPSS Inc.).

3. Results

3.1. DNA Yield

Four samples from two sites were extracted in duplicate and the DNA yield of each extract was measured. There was no significant difference (*t*-test, $p = 0.371$) in DNA yield between extracts from the Wetland site (7.9 ± 1.8 ng/mg soil) and those from the Coastal site (4.5 ± 2.7 ng/mg soil). There was also no significant difference between samples within the Wetland site; however, sample C4 in the coastal site showed higher DNA yield than samples C1, C2 and C3 (Fig. S2). This was likely attributable to the presence of macroscopic plant material in C4 (Fig. S2). When this sample was excluded from analysis, the average DNA yield for the coastal site was 3.2 ± 0.5 ng/mg of soil, and a statistically significant difference was observed between the two sites (*t*-test, $p = 0.036$). The latter implies that DNA recovery from sandy soils may be more problematic, or that these soils contain less genetic material than organic rich soils. Despite this observation, C4 was retained for analysis to reflect realistic within site variation.

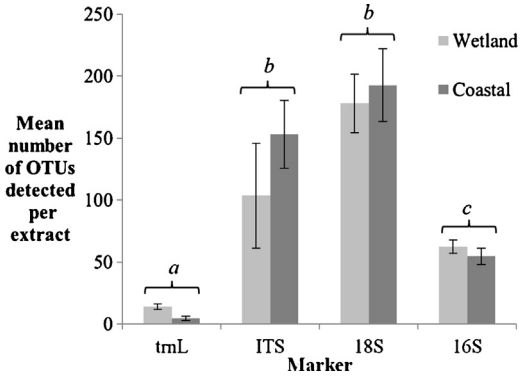

Fig. 2. Number of OTUs count detected by each marker. The mean number of OTUs detected per DNA extract from two sampling sites is displayed. Error bars represent standard deviation and letters indicate significant differences between markers: *trn*L detected significantly less OTUs than all other markers and both ITS and 18S detected significantly higher OTU count than *trn*L and 16S.

3.2. PCR amplification

PCR amplification of ITS, 16S and 18S was successful in all PCR reactions from all DNA extracts. However, amplification of *trn*L was inconsistent across triplicate reactions and duplicate DNA extracts, (Table S2). For the wetland site, amplification of *trn*L failed in three of eight DNA extracts (W2B, W3A and W4A) preventing further analyses of these samples for this marker. For ITS and *trn*L, no PCR product was visible in the extraction blank controls (EBCs) or PCR negatives after gel electrophoresis. The average concentration of non-purified PCR products from EBCs was 0.62 ± 0.06 ng/μL (*trn*L) and 0.92 ± 0.13 ng/μL (ITS). For 16S and 18S, EBC PCR products were detected; however, the corresponding no template control showed no PCR product. This suggests that background DNA was likely introduced during the DNA extraction process, like from reagents or labware [29,30]. Consequently, EBCs for 18S and 16S were sequenced in parallel with the samples.

3.3. Number of OTUs

Each marker detected a different number of OTUs (Fig. 2); the highest was detected using ITS and 18S (*one-way ANOVA: ITS*, $p = 0.017$, and *18S*, $p = 0.041$), whereas *trn*L detected the lowest (*one-way ANOVA*, $p < 0.01$, Fig. 2). The number of OTUs detected by the *trn*L marker was also dependant on site; the Wetland site was significantly higher than that in the coastal samples (*t*-test, $p = 0.001$). In contrast, no significant difference was observed between the two sites using ITS, 18S and 16S (*t*-test: *ITS*, $p = 0.105$; *18S*, $p = 0.472$; *16S*, $p = 0.118$). OTU count was most variable between duplicate extracts using ITS, 18S and 16S (range of SD; 2.83–20.51, 0.71–16.97, and 2.12–13.44, respectively), whereas the number of OTUs detected with *trn*L was more consistent (range of SD; 0.71–3.54), although the latter is likely related to the low number of OTUs detectable with this marker.

3.4. Discriminatory power

Using the Bray–Curtis distance, all extracts clustered according to site using ITS, 18S and 16S markers (Fig. 3 and Fig. S4). Dissimilarity values between the sites ranged from 29.9% (16S) to 85.6% (ITS), and the diversity was statistically different between the wetland and coastal samples (*one-way ANOSIM* ($p = 0.1\%$)*: ITS*, $R = 0.942$; *18S*, $R = 0.913$; *16S*, $R = 0.966$, Fig. S4). Discriminatory power was not directly related to the number of OTUs detected since ITS and 18S had comparable OTU counts, yet 18S had a lower dissimilarity than ITS (61.3% compared to 85.6%). The *trn*L marker failed to consistently differentiate between the sites since a single Wetland extract (W2) clustered more closely with the Coastal samples (Fig. 3 and Fig. S5). Nevertheless, the plant diversity detected in these two sites was statistically different despite sample W2 as an outlier (*one-way ANOSIM*, $p = 0.1\%$, *trn*L, $R = 0.933$).

3.5. Effects of OTU filtering on discriminatory power

To determine if the discriminatory power between the two sites could be improved by removing DNA sequences identified within extraction blank controls (EBC), any OTUs identified within the EBC for each marker were removed from the corresponding samples. In the ITS and *trn*L EBCs, no PCR product was visible following gel electrophoresis, while products were detected in those for 16S and 18S. Removal of the OTUs detected in the 16S and 18S EBCs led to an increase in discriminatory power between the sites in both cases (Fig. 4 and Table 2). A higher percentage of OTUs were excluded from the 16S data at this filtering step (17.7%) compared to the 18S dataset (6.6%), and resulted in the discriminatory power of the 16S marker increasing to a level (63.2%) comparable with 18S (63.6%). In addition, filtering of background OTUs improved the resolution between samples within a particular site. For example, prior to filtering duplicate wetland extracts from 16S analyses did not cluster by sample; however, following filtering, duplicate extracts were more similar to each other than any other sample from that site (Fig. 3). Grouping of duplicate extracts within a site indicates a heterogeneous distribution of taxa between the sampling points (1 m apart) within each site, and offers the potential to discriminate between samples at a much finer scale.

We also tested the effect of removing OTUs that were common to all extracts, and found this enhanced the discriminatory power between sites for all markers (Fig. 4 and Table 2). 16S displayed the largest increase in discriminatory power (11.1%) after the removal of shared OTUs, and indeed the discriminatory power of 16S (74.3%) exceeded that observed with 18S (70.5%). Nevertheless, ITS remained the most discriminative marker regardless of the filtering method applied (85.6% dissimilarity), although 18S and 16S appeared to be equally effective post-filtering.

3.6. Reproducibility and consistency

The OTU composition of each DNA extract was compared to determine which marker provides the most reproducible profile from (a) a single sample and (b) different samples within a single site. The *trn*L marker produced the most reproducible DNA profile both within a site and within a single sample (*one-way ANOVA*, $p < 0.01$) (Fig. 5). However, the values for *trn*L included only the 5/8 samples for which PCR amplification was successful in both extracts and the low diversity detected with this marker improved the chances of appearing more reproducible. Of the remaining three markers, the reproducibility of OTU composition within a site was comparable (48 ± 14% similarity, mean ± SD). However, the reproducibility of duplicate extracts from a single sample varied depending on the marker used. Duplicate extracts were more reproducible using ITS and 18S, compared to 16S (*one-way ANOVA*, $p < 0.01$) (Fig. 5A). Furthermore, the reproducibility of DNA profiles from duplicate extracts was not effected by habitat for ITS, 18S and 16S (*t*-test; $p = 0.225$, $p = 0.820$ and $p = 0.247$, respectively).

4. Discussion

We have demonstrated that DNA metabarcoding coupled with HTS has considerable potential to discriminate between forensic soil samples from different sites. By analysing four different DNA markers, we determined that non-bacterial ITS and 18S markers

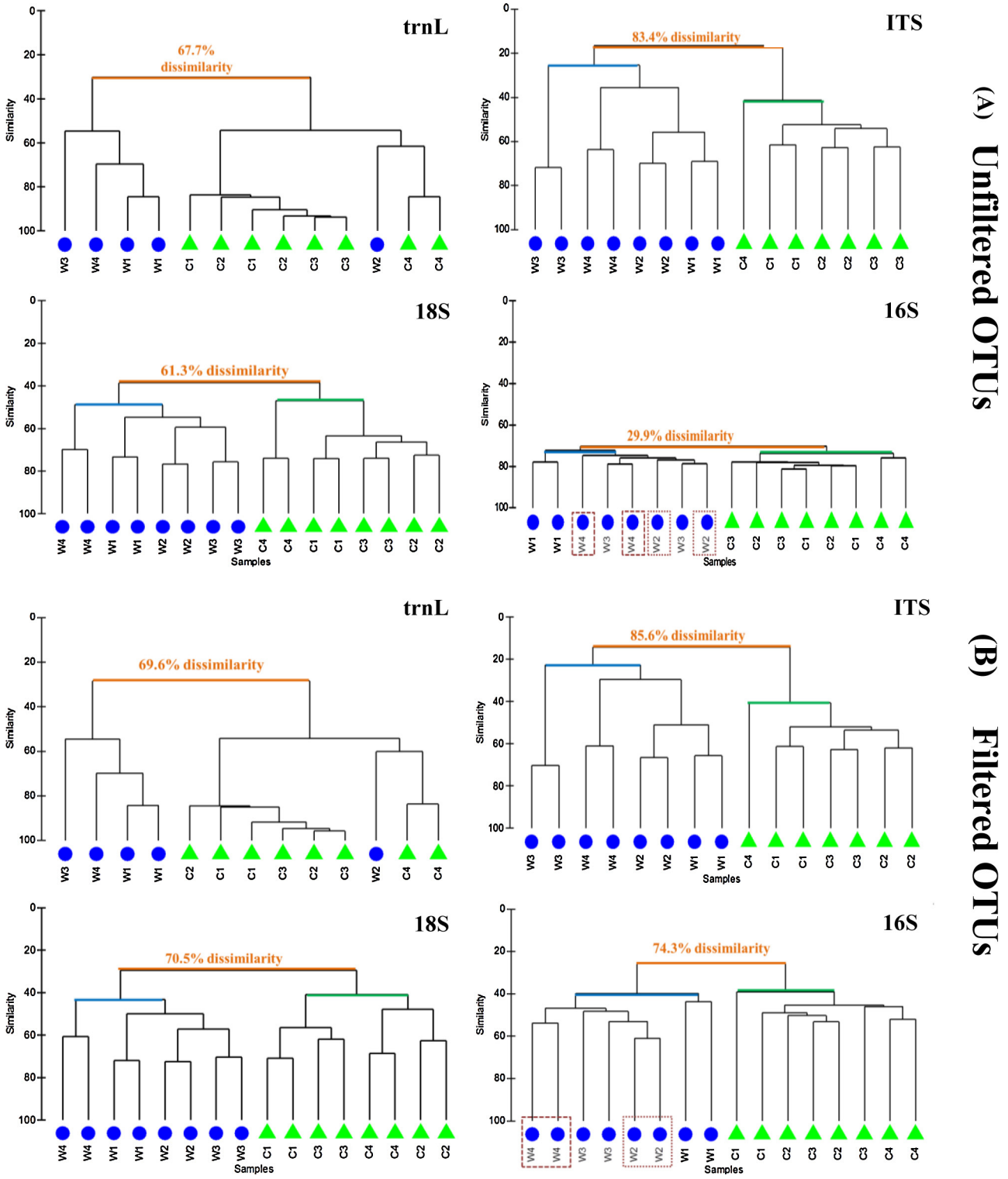

Fig. 3. The effect of OTU filtering on the discriminatory power of each marker. (A) Cluster dendrograms based on the Bray–Curtis distances were generated using the unfiltered OTU tables for each marker. OTU tables were then filtered in two consecutive steps to exclude (1) contaminant OTUs detected in the extraction blank controls and (2) any OTUs common to all DNA extracts for each marker. (B) A second set of cluster dendrograms was generated using the resulting filtered OTU tables. Boxes around 16S samples illustrate the improvement in sample resolution upon filtering of EBC.

appear most promising for forensic soil analysis, because each can amplify target DNA reliably and provide high discriminatory power between sites (Table 3). ITS achieved a higher discriminatory power (83.4% prior to OTU filtering and 85.6% post filtering) compared to 18S, (61.3% prior to filtering and 70.5% post filtering) and 16S (29.9% prior to OTU filtering and 74.3% post filtering). These results support previous T-RFLP studies that suggest fungi is the most robust target biomarker for soil discrimination [15].

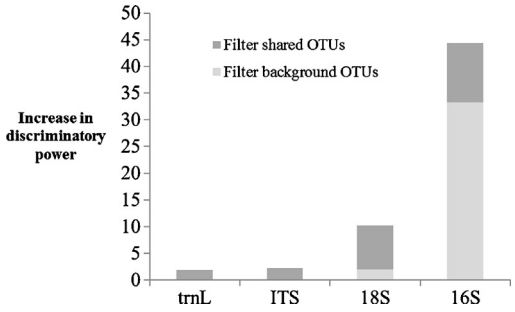

Fig. 4. Improvement in discriminatory power of each marker post filtering. The increase in Bray–Curtis percent dissimilarity values is plotted following two consecutive OTU filtering steps: (1) removal of contaminant OTUs detected in the extraction blank controls and (2) removal of OTUs shared between all DNA extracts. *PCR product was not visible in the agarose gel, so EBC contaminant filtering was not applied.

Although the most reproducible OTU profiles were generated from *trn*L libraries, likely due to the low number of OTUs detected by this marker, difficulties during the initial PCR amplification step indicate that this marker may not work well for forensic soil analysis. Inconsistent PCR amplification, particularly within the wetland site, suggested heterogeneous distribution of plant DNA across this site likely prevented detection in some samples. Nevertheless, current *trn*L analysis may work well in a specific case where detection of certain plant species could potentially link a suspect to a site [30]. A shorter 50 bp DNA fragment of the *trn*L intron [31] could be trialled as an alternative plant marker, although shorter read lengths may not distinguish between similar sequences.

Universal primers are designed to target a wide range of organisms; therefore, low levels of background DNA from laboratory reagents and labware may be detected by HTS when primers with a broad specificity are applied [28,29,30–34]. The primers utilised in this study target bacteria and fungi so, we would expect to detect background laboratory DNA in EBC and no-template PCR controls. In this study, background DNA was detected in the EBCs of 16S and 18S rRNA amplifications, suggesting the introduction of DNA occurred during the DNA extraction process. The EBCs were then sequenced to ensure background levels of DNA were analysed and excluded during the bioinformatic processing. This proved to be a vital control, as removal of reads observed in EBCs increased the resolution between sites as well as individual samples. Our EBC filtering process removed 17.4% from 16S and 6.6% from 18S OTUs, and increased the discriminatory power between the two sites by 29% and 2.3% for 16S and 18S, respectively. Furthermore, our study shows that different sample points within each site could be distinguished post-filtering, enabling soil discrimination at a finer scale. Porter et al. previously excluded 0.5% of 16S bacterial OTUs as contaminants from soil cores and found no contaminant reads using *trn*L [23]. This demonstrates the variation of background level DNA present in different laboratories and highlights the need to perform and sequence EBCs in parallel to experimental or forensic samples. In contrast to the EBC samples, amplification of PCR negatives did not result in visible amplicons, suggesting that this background DNA was not introduced from sample-to-sample contamination, and is from laboratory reagents. However, this does not eliminate the possibility that any sample-to-sample contamination did not occur during the extraction. This study processed samples in two separate batches to avoid sample-to-sample contamination between the two locations; however, this may not be practical for an analysis involving multiple locations. Therefore we would recommend an EBC between samples from each location so that any DNA transferred between locations can be detected and excluded. In addition, evidential samples should always be processed prior to, and independently of, reference samples. This also demonstrates the need to include no-template PCR negative controls in combination with EBCs in forensic studies. Currently, EBCs and no-template PCR controls are not routinely sequenced in metagenomic studies, which is a major concern for validation in forensic science. Despite efforts to minimise background contamination by carrying out experiments in dedicated PCR hoods, contributions of background DNA are likely inevitable, so proper control samples should be a key feature in metagenomic studies.

The removal of common OTUs present in all DNA extracts further improved the discriminatory power between the sites, offering a useful tool for improving resolution between samples expected to have similar soil profiles. A recent study demonstrated that soils analysed from similar land-use types shared comparable fungal compositions, despite being separated by up to 215 km [35]. Our results suggest that bacterial OTUs detected by 16S may be even more widespread (5.3% OTUs common to all extracts) than soil eukaryotes (0.5% ITS OTUs and 1.1% 18S OTUs common to all extracts). The unique biological signal from individual locations may be masked by OTUs found in many locations. Therefore, this filtering approach may be particularly advantageous for emphasising differences between samples from similar location types, and increase the discriminatory power.

It is possible that bacterial DNA may be more discriminative on a small scale to pinpoint specific areas within a site, whereas non-bacterial DNA may be preferable for identifying sites/localities at a broader scale. The reproducibility of 16S bacterial profiles from a single sample was lower than non-bacterial profiles generated using ITS (fungi) and 18S (eukaryotes). This is in agreement with Macdonald et al. who suggest higher local scale T-RFLP profile variation of bacteria compared to fungi [16]. Bacteria are thought to form micro-spatial niches within micro-aggregates of soil particles [36], therefore the distribution of bacterial taxa within a sample is more heterogeneous than non-bacterial taxa, and as a result, the latter are not as susceptible to potential biases with sampling and DNA extraction [37,38]. In this way, the variation in

Table 2
The number of OTUs removed during each filtering step and the resulting dissimilarity value (%).

		*trn*L	ITS	18S	16S
Unfiltered OTUs	# OTUs	19	654	699	299
	Dissimilarity	67.7%	83.4%	61.3%	29.9%
Filter background OTUs	# OTUs removed	0[a]	0[a]	46	52
	# OTUs remaining	19	654	653	247
	Dissimilarity	67.7%	83.4%	63.6%	63.2%
Filter shared OTUs	# OTUs removed	1	3	7	13
	# OTUs remaining	18	651	646	234
	Dissimilarity	69.6%	85.6%	70.5%	74.3%

[a] PCR product was not visible in the agarose gel for EBC so no contaminant filtering was applied.

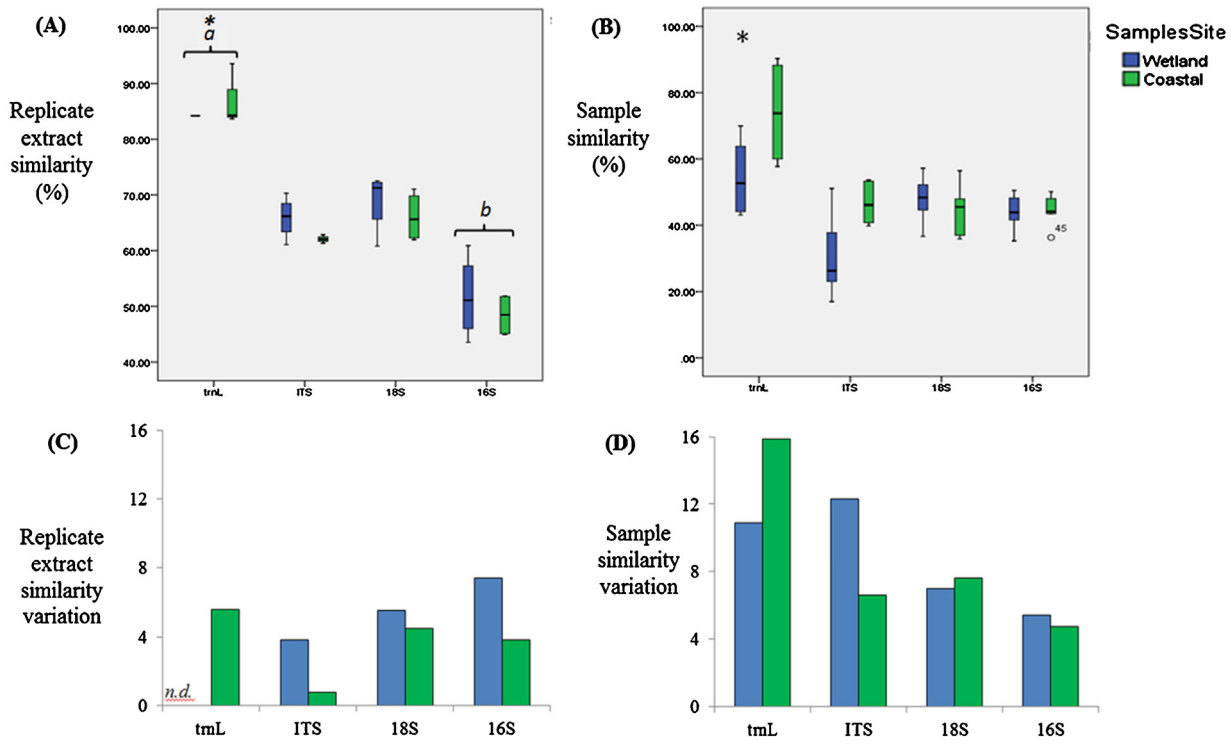

Fig. 5. Comparison of OTU composition reproducibility using each marker. Boxplots show the mean percent similarity (100-BC) from pair-wise comparisons of (A) replicate DNA extracts and (B) samples within each site. A reproducible OTU composition is represented by a high percent similarity. Figures (C) and (D) show the variation (standard deviation) of percent similarity across the pair-wise comparisons from (A) and (B) respectively. Letters indicate significant differences between markers. Both replicate extract similarity (%) and sample similarity (%) were significantly higher using *trn*L. For the other three markers, replicate extract similarity was significantly lower using 16S; however, sample similarity (%) was comparable between ITS, 18S and 16S. **trn*L data only includes the samples from which PCR amplifications were successful from both extracts (5/8 soil samples). n.d. represents no data available for analysis.

spatial distribution between bacteria and eukaryotes could enable a forensic scientist to select the most appropriate marker for a particular scenario. For example, to establish a link between soil from a shoe and a particular footprint, 16S may provide better results than ITS and 18S. Conversely, ITS and 18S may be more appropriate for linking a forensic sample to a site when only the general locality is known. Further, site-specific characteristics are likely to be valuable for selecting the most appropriate sampling approach to accurately capture a representative diversity at reference sites [39,40].

Comparative reproducibility of DNA profiles across a site suggested that two samples from the same location shared roughly 50% of the diversity detected. Superficially, this values seems low, yet it is consistent with a previous study which reported an average pairwise similarity of 0.518–0.773 from 16S bacterial T-RFLP analysis of soil samples collected 10 feet apart [41]. Similarly, a recent study using HTS reported 65% and 60% similarity between duplicate samples analysed using the short P6 loop of *trn*L intron (plant) and the 16S rDNA (bacteria), respectively [23]. Nevertheless, within-sample variation was significantly lower than between-sample/site variation. This indicates that DNA metabarcoding and HTS does hold promise for forensic application; however, because the methodology is in its infancy, further research is required to optimise the analytical methods to improve similarity estimates from replicate samples. The spatial distribution of taxa within soils means that sampling bias may contribute to reduced similarity values between duplicate extracts. Within each DNA extract, PCR bias and differences in copy number of DNA templates can skew estimates of taxonomic abundance [42,43]; however, PCR amplifications were carried out in triplicate in this study to minimise this effect. The use of quantitative PCR could improve estimates of taxonomic abundance within samples by

Table 3
Summary of the findings for each marker. Number of OTUs detected (Low: <20, Moderate: 50–70, High: >100). % OTUs common to all extracts (low: <5, moderate: 5–10, high: >10). Discriminatory power (low: <70%, moderate: 70–80% following filtering, High: >80% before filtering). Diversity within a sample (low: <60%, moderate: 60–80%, high: >80% similarity).

	*trn*L	ITS	18 S	16 S
PCR Amplification Success	Inconsistent	100%	100%	100%
Number of OTUs detected	Low	High	High	Moderate
Contaminant filtering	No[b]	No[b]	Yes	Yes
% OTUs common to all extracts	Low	Low	Moderate	High
Discriminatory power	Low	High	Moderate	Moderate
Replicate extract similarity	High[a]	Moderate	Moderate	Low
Sample similarity	High	Moderate	Moderate	Moderate

[a] Includes the samples from which PCR amplification was successful from both extracts (5/8 soil samples).
[b] A PCR product was not visible for the EBC control in the agarose gel, so contaminant filtering was not applied.

assessing DNA quality, DNA quantity and PCR inhibition during amplification [44]. Furthermore, despite efforts to exclude poor quality sequences using stringent filtering thresholds, sequencing errors associated with the Ion Torrent could also contribute somewhat to the low similarity values observed [34,45]. Currently, no algorithm is available to de-noise Ion Torrent data. In this regard, sequencing PCR products on Illumina platforms and filtering data using de-noising [46,47] and chimaera detection algorithms [48,49] should markedly improve similarity estimates. Furthermore, sequencing on the Illumina platform would also increase sequencing depth [11], allowing saturation of the diversity within each sample and thus minimise bias associated with limited coverage of the OTUs present in replicate samples [43].

This study demonstrates the potential use of HTS for forensic soil DNA analysis and highlights important practical and analytical steps required to obtain a robust DNA profile, given the wide spatial variability of taxa and sensitivity of HTS to contaminant DNA. However, before these novel forensic methods can be applied, strict validation standards must be met to define the robustness, confidence levels and potential weight of such evidence [50,51]. Future experimental trials analysing multiple sites are required to demonstrate broad-scale reliability and reproducibility, as well as error rates associated with this method [52–55]. In particular, potential limitations likely to be encountered during forensic casework must also be tested, since many factors could influence soil DNA profiles [56]. Temporal effects, such as rainfall and temperature, could impact the DNA profile obtained since a time lapse often exists between the criminal activity and the recovery of reference samples during investigations. Similarly, storage conditions of evidence samples could adversely affect the soil community profile. For example, if a soil is removed from the environment by transfer to an object, such as a shoe sole, conditions will be altered and in many cases such soil will dry-out. The impact of both soil moisture and sample drying will be critical in establishing the robustness of this method in practice, and further determining the most reliable target taxa. Transfer of soil to objects also introduces further limitations; (1) soil particles differ in persistence, with the larger aggregates being lost primarily, and (2) soil from the site of interest can be mixed with other soils, for example, a car tyre may contain multiple layers of soils from different locations. Each of these will alter the soil composition and presumably the biota detected within samples, therefore careful consideration is required prior to analysis and interpretation of data from such samples. Furthermore, sample quantity is often limited in casework, therefore the minimum sample size required to produce a representative diversity needs to be assessed to determine the potential value of trace samples. Research examining the ability to discriminate soil samples given each of these limitations, and determination of the most robust marker given certain circumstances, is required so that forensic scientists can make informed decisions on the suitability of this technique for particular cases.

Acknowledgements

This work was supported by the Australian Research Council, and the Australian Federal Police (ARC Linkage Grant LP34107393). We thank members of the Australian Centre for Ancient DNA for helpful comments on the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.07.014.

References

[1] R.M. Morgan, P.A. Bull, The philosophy, nature and practice of forensic sediment analysis, Progr. Phys. Geogr. 31 (2007) 43–58.
[2] K. Pye, Geological and Soil Evidence, Forensic Applications, CRC Press, Boca Raton, FL, 2007.
[3] A. Ruffell, Forensic pedology, forensic geology, forensic geoscience, geoforensics and soil forensics, Forensic Sci. Int. 202 (2010) 9–12.
[4] K.D. Lorna Ritz, David Miller, Criminal and Environmental Soil Forensics Criminal and Environmental Soil Forensics, Springer, 2009;
K. Ritz, L. Dawson, D. Miller (Eds.), Criminal and environmental soil forensics, Springer, Dordrecht, 2009.
[5] R.J. Cano, Molecular Microbial Forensics, Royal Soc Chemistry, Cambridge, 2010.
[6] G. Concheri, Chemical elemental distribution and soil DNA fingerprints provide the critical evidence in murder case investigation, PLoS ONE 6 (2011) e20222.
[7] J.B.H. Martiny, B.J.M. Bohannan, J.H. Brown, R.K. Colwell, J.A. Fuhrman, J.L. Green, M.C. Horner-Devine, M. Kane, J.A. Krumins, C.R. Kuske, P.J. Morin, S. Naeem, L. Ovreas, A.L. Reysenbach, V.H. Smith, J.T. Staley, Microbial biogeography: putting microorganisms on the map, Nat. Rev. Microbiol. 4 (2006) 102–112.
[8] S. Dequiedt, N.P.A. Saby, M. Lelievre, C. Jolivet, J. Thioulouse, B. Toutain, D. Arrouays, A. Bispo, P. Lemanceau, L. Ranjard, Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management, Global Ecol. Biogeogr. 20 (2011) 641–652.
[9] M.L. Metzker, Applications of Next generation sequencing – the next generation, Nat. Rev. Genet. 11 (2010) 31–46.
[10] J.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J.A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, ISME J. 6 (2012) 1621–1624.
[11] M.L. Metzker, Next generation technologies: basics and applications, Environ. Mol. Mutagen. 51 (2010), 691AQ.
[12] A. Magi, M. Benelli, A. Gozzini, F. Girolami, F. Torricelli, M.L. Brandi, Bioinformatics for next generation sequencing data, Genes 1 (2010) 294–307.
[13] P. Taberlet, E. Coissac, M. Hajibabaei, L.H. Rieseberg, Environmental DNA, Mol. Ecol. 21 (2012) 1789–1793.
[14] N.G. Yoccoz, The future of environmental DNA in ecology, Mol. Ecol. 21 (2012) 2031–2038.
[15] C.A. Macdonald, Discrimination of soils at regional and local levels using bacterial and fungal t-RFLP profiling, J. Forensic Sci. 56 (2011) 61–69.
[16] L.M. Macdonald, B.K. Singh, N. Thomas, M.J. Brewer, C.D. Campbell, L.A. Dawson, Microbial DNA profiling by multiplex terminal restriction fragment length polymorphism for forensic comparison of soil and the influence of sample condition, J. Appl. Microbiol. 105 (2008) 813–821.
[17] N. Fierer, C.L. Lauber, N. Zhou, D. McDonald, E.K. Costello, R. Knight, Forensic identification using skin bacterial communities, Proc. Natl. Acad. Sci. 107 (2010) 6477–6481.
[18] B. Budowle, P.C. Williamson (Eds.), Microbial Forensics, Academic Press, New York, 2010.
[19] N. Lombard, E. Prestat, J.D. van Elsas, P. Simonet, Soil-specific limitations for access and analysis of soil microbial communities by metagenomics, FEMS Microbiol. Ecol. 78 (2011) 31–49.
[20] P.-A. Schmidt, M. Bálint, B. Greshake, C. Bandow, J. Römbke, I. Schmitt, Illumina metabarcoding of a soil fungal community, Soil Biol. Biochem. 65 (2013) 128–132.
[21] A. Valentini, F. Pompanon, P. Taberlet, DNA barcoding for ecologists, Trends Ecol. Evol. 24 (2008) 110–117.
[22] P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann, E. Willerslev, Towards next-generation biodiversity assessment using DNA metabarcoding, Mol. Ecol. 21 (2012) 2045–2050.
[23] T.M. Porter, G.B. Golding, C. King, D. Froese, G. Zazula, H.N. Poinar, Amplicon pyrosequencing late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale reproducibility, Mol. Ecol. Resour. 13 (2013) 798–810.
[24] M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing, Cold Spring Harb. Protoc (2010), pdb.prot5448.
[25] M. Martin, R. Sven, From cutadapt to sequencetools (sqt): a versatile toolset for sequencing projects, EMBnet. journal 17.B (2012) 35.
[26] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335–336.
[27] R.C. Edgar, Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26 (2010) 2460–2461.
[28] C.E. Corless, M. Guiver, R. Borrow, V. Edwards-Jones, E.B. Kaczmarski, A.J. Fox, Contamination and sensitivity issues with a real-time universal 16S rRNA PCR, J. Clin. Microbiol. 38 (2000) 1747–1752.
[29] M.A. Tanner, B.M. Goebel, M.A. Dojka, N.R. Pace, Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants, Appl. Environ. Microbiol. 64 (1998) 3110–3113.
[30] G. Ferri, Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach, Int. J. Legal Med. 123 (2009) 395–401.
[31] P. Taberlet, E. Coissac, F. Pompanon, L. Gielly, C. Miquel, A. Valentini, T. Vermat, G. Corthier, C. Brochmann, E. Willerslev, Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding, Nucleic Acids Res. 35 (2007) e14.

[32] R. Schmieder, R. Edwards, Fast identification and removal of sequence contamination from genomic and metagenomic datasets, PLoS ONE 6 (2011) e17288.
[33] S. Champlot, C. Berthelot, M. Pruvost, E.A. Bennett, T. Grange, E.-M. Geigl, An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications, PLoS ONE 5 (2010) e13042.
[34] V. Kunin, A. Engelbrektson, H. Ochman, P. Hugenholtz, Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates, Environ. Microbiol. 12 (2010) 118–123.
[35] S. Kasel, L.T. Bennett, J. Tibbits, Land use influences soil fungal community composition across central Victoria, south-eastern Australia, Soil Biol. Biochem. 40 (2008) 1724–1732.
[36] L.G. Grundmann, F. Gourbiere, A micro-sampling approach to improve the inventory of bacterial diversity in soil, Appl. Soil Ecol. 13 (1999) 123–126.
[37] B.D. Lindahl, R.H. Nilsson, L. Tedersoo, K. Abarenkov, T. Carlsen, R. Kjøller, U. Kõljalg, T. Pennanen, S. Rosendahl, J. Stenlid, H. Kauserud, Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide, New Phytol. 199 (2013) 288–299.
[38] R. Blaalid, T.O.R. Carlsen, S. Kumar, R. Halvorsen, K.I. Ugland, G. Fontana, H. Kauserud, Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing, Mol. Ecol. 21 (2012) 1897–1908.
[39] K. Andersen, K.L. Bird, M. Rasmussen, J. Haile, H. Breuning-Madsen, K.H. KjÆR, L. Orlando, M.T.P. Gilbert, E. Willerslev, Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity, Mol. Ecol. 21.8 (2011) 1966–1979. , http://dx.doi.org/10.1111/j.1365-294X.2011.05261.x.
[40] K.L. Baker, S. Langenheder, G.W. Nicol, D. Ricketts, K. Killham, C.D. Campbell, J.I. Prosser, Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies, Soil Biol. Biochem. 41 (2009) 2292–2298.
[41] M.S. Meyers, D.R. Foran, Spatial and temporal influences on bacterial profiling of forensic soil samples, J. Forensic Sci. 53 (2008) 652–660.
[42] A.S. Amend, K.A. Seifert, T.D. Bruns, Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol. Ecol. 19 (2010) 5555–5565.
[43] A.J. Pinto, L. Raskin, PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets, PLoS ONE 7 (2012) e43093.
[44] D.C. Murray, M. Bunce, B.L. Cannell, R. Oliver, J. Houston, N.E. White, R.A. Barrero, M.I. Bellgard, J. Haile, DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches, PLoS ONE 6 (2011) e25776, http://dx.doi.org/10.1371/journal.pone.0025776.
[45] L.M. Bragg, G. Stone, M.K. Butler, P. Hugenholtz, G.W. Tyson, Shining a light on dark sequencing: characterising errors in ion torrent PGM data, PLoS Comput. Biol. 9 (2013) e1003031.
[46] B. Lee, J. Park, S. Yoon, Presented at the ICDM, 2012.
[47] C. Quince, A. Lanzen, R.J. Davenport, P.J. Turnbaugh, Removing noise from pyrosequenced amplicons, BMC Bioinform. 12 (2011) 38.
[48] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 27 (2011) 2194–2200.
[49] P.D. Schloss, D. Gevers, S.L. Westcott, Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies, PLoS ONE 6 (2011) e27310.
[50] S.E.S. Bruce Budowle, G. Breeze Roger, S. Keim Paul, A. Morse Stephen, Validation of microbial forensics in scientific, legal, and policy context, in: B. Budowle (Ed.), Microbial Forensics, second ed., Elsevier, 2010, pp. 649–663.
[51] S.E.S. Bruce Budowle, G. Breeze Roger, S. Keim Paul, A. Morse Stephen, Inferential validation and evidence interpretation, in: Microbial Forensics, second ed., Elsevier, 2010, pp. 527–544.
[52] D. Knights, E.K. Costello, R. Knight, Supervised classification of human microbiota, FEMS Microbiol. Rev. 35 (2011) 343–359.
[53] D. Knights, J. Kuczynski, E.S. Charlson, J. Zaneveld, M.C. Mozer, R.G. Collman, F.D. Bushman, R. Knight, S.T. Kelley, Bayesian community-wide culture-independent microbial source tracking, Nat. Methods 8 (2011) 761–763.
[54] K.M. Hewitt, M.F. Gonzalez, A. Chase, J.H.J.G. Caporaso, et al., Bacterial diversity in two neonatal intensive care units (NICUs), PLoS ONE 8 (1) (2013) e54703.
[55] K. Andersen, K.L. Bird, M. Rasmussen, J. Haile, H. Breuning-Madsen, K.H. Kjaer, L. Orlando, M.T.P. Gilbert, E. Willerslev, Meta-barcoding of 'dirt'DNA from soil reflects vertebrate biodiversity, Mol. Ecol. 21 (2012) 1966–1979.
[56] J.M. Young, N.J. Rawlence, L.S. Weyrich, A. Cooper, Limitations and recommendations for successful DNA extraction from forensic soil samples: a review, Sci. Justice (2014).