Forensic Science International 270 (2017) 153–158

Contents lists available at ScienceDirect

Forensic Science International

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

Microbial soil community analyses for forensic science: Application to a blind test

Sandrine Demanèche[a,*], Leif Schauser[b], Lorna Dawson[c], Laure Franqueville[a], Pascal Simonet[a]

[a] Université de Lyon, École Centrale de Lyon, Laboratoire Ampère (CNRS UMR5005), Environmental Microbial Genomics, Ecully, France
[b] QIAGEN Aarhus A/S, Silkeborgvej 2, DK-8000 Aarhus, Denmark
[c] The James Hutton Institute, Craigiebuckler, Aberdeen, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 21 October 2016
Accepted 3 December 2016
Available online 9 December 2016

Keywords:
Soil DNA
Bacteria
Forensic geoscience
High throughput sequencing
Ribosomal Intergenic Spacer Analysis

ABSTRACT

Soil complexity, heterogeneity and transferability make it valuable in forensic investigations to help obtain clues as to the origin of an unknown sample, or to compare samples from a suspect or object with samples collected at a crime scene. In a few countries, soil analysis is used in matters from site verification to estimates of time after death. However, up to date the application or use of soil information in criminal investigations has been limited. In particular, comparing bacterial communities in soil samples could be a useful tool for forensic science. To evaluate the relevance of this approach, a blind test was performed to determine the origin of two questioned samples (one from the mock crime scene and the other from a 50:50 mixture of the crime scene and the alibi site) compared to three control samples (soil samples from the crime scene, from a context site 25 m away from the crime scene and from the alibi site which was the suspect's home). Two biological methods were used, Ribosomal Intergenic Spacer Analysis (RISA), and 16S rRNA gene sequencing with Illumina Miseq, to evaluate the discriminating power of soil bacterial communities. Both techniques discriminated well between soils from a single source, but a combination of both techniques was necessary to show that the origin was a mixture of soils. This study illustrates the potential of applying microbial ecology methodologies in soil as an evaluative forensic tool.

1. Introduction

Soil adheres to and can persist on footwear, tires, tools, clothing *etc*. which makes it a particularly valuable trace material in forensic investigations to yield clues as to the origin of an unknown sample, or to compare samples from a suspect with those from a crime scene [1–3]. Apart from palynology, the organic (plant and animal derived) components found in soil have been given much less attention than the comparison of soil inorganic (*i.e.* mineral and elemental values) components to differentiate between samples [4–6]. Although these inorganic methods have already been applied successfully in some forensic investigations, they generally suffer from a lack of sensitivity with an insufficient discriminatory potential: geology and mineralogy are relatively homogeneous and very similar [7].

However, at a much more detailed scale (below the cm level), soil is a complex matrix where heterogeneity favors its colonization by a huge diversity of microorganisms, allowing differentiation of soil samples from very close locations by the specificity of their microbiota [8–10]. The strong discriminating power of microbiomes has already been exploited in other circumstances, to relate cell phones to their owners for instance [11]. In soil, the huge bacterial diversity level combined to the "transferable" properties of soil particles offers particularly promising opportunities to develop new approaches for forensic purposes.

The soil microbiota is now investigated using the metagenomics approach based on the direct extraction and exploitation of microbial DNA from soil samples [12,13]. However, there is no generally accepted DNA extraction method among soil microbiologists, several protocols must be tested for each soil type [14–16]. But there is agreement that metagenomics suffers from numerous biases [17]. These limitations justify further validation experimentally of the methods used in terms of sensitivity, specificity and reproducibility for forensic use.

For that purpose, a collaborative inter-laboratory blind test was performed in the context of a European project named MiSAFE (http://forensicmisafe.wix.com/misafe). The efficiency of several

* Corresponding author.
E-mail address: sandrine.demaneche@ec-lyon.fr (S. Demanèche).

http://dx.doi.org/10.1016/j.forsciint.2016.12.004
0379-0738/

methods and genes for forensic applications was previously tested [18] and the conclusions were used to carry out a blind test. A mock crime scene was designed, based on a real case example, with a scenario that a local farmer discovered an area of disturbed soil on his land, and subsequently a packet of drugs was found buried underground. Three locations were chosen to use in the collaborative laboratory assessment exercise. These included locations with both similar and different characteristics to the crime scene (site 1). One site was chosen very close (site 2, 25-m away) to where the original mock crime scene experiment was set up, and another was an alibi site (site 3, 0.8 km away from the mock crime scene) close to the farm (the suspect's home). In addition, two unknown soil samples were taken on two spades found at the suspect home, one of these samples being subsequently (and covertly to extend the blind test context) mixed with other soil samples in order to evaluate the efficiency of each method in the case of complex samples.

We present here the part of the results of this large multidisciplinary study, (involving 12 participants and a broad range of techniques) that aimed to compare and determine the efficiency of two DNA-based technologies applied to soil extracted DNA to discriminate between soil samples: a global quick and cheap profile-based approach with Ribosomal Intergenic Spacer Analysis (RISA) and a detailed but more expensive sequence-based approach with 16S rRNA gene sequencing (Illumina Miseq) were used.

2. Material and methods

2.1. Soil samples

Three control, Site_1_11 (site 1, the crime scene), Site 2_19 (site 2, the context site) and Site_3_05 (site 3, the alibi site), and two unknown, GT A_14 (GTA) and GT B_03 (GTB) refrigerated soil samples were sent in a fresh condition to our lab in October 2014 and stored for two days at 4 °C before DNA extraction with the Nucleospin Soil kit (Macherey-Nagel, France). Three extractions were performed on 350 mg of unknown soil samples and two extractions were performed on 350 mg of the three control soil samples according to the manufacturer's instructions. Final DNA elution was performed with 50 μL of elution buffer (Tris 10 mM). Concentrations (Table 1) were determined with a fluorometric method (Qubit, Invitrogen).

Table 1
DNA concentrations of soil samples.

Sample name	Pure DNA (ng μL^{-1})	Diluted DNA (ng μL^{-1})
GT A_14		
1A	156	13.8
1B	152	13.8
1C	175	14.8
GT B_03		
2A	164	13.2
2B	140	12.4
2C	133	12.8
Site 1_11		
3A	173	13.2
3B	168	15.1
Site 2_19		
4A	168	14.2
4B	137	12.5
Site 3_05		
5A	187	15
5B	169	14.8

2.2. Ribosomal Intergenic Spacer Analysis (RISA)

Ribosomal intergenic sequences were amplified with primers S-D-Bact-1522-b-S-20 (5′-TGCGGCTGGATCCCCTCCTT-3′) and L-D-Bact-132-a-A-18 (5′-CCGGGTTTCCCCATTCGG-3′) [19]. Amplifications were performed using the Hotstart mix RTG kit (GE Healthcare, France). Freeze-dried beads (with Taq polymerase, dNTPs and buffer) were rehydrated with 23 μL of the primer solution (0.8 μM final) and 2 μL of diluted DNA (Table 1) or water for the negative sample were added. The PCR program included an initial denaturation at 94 °C for 5 min, then 30 cycles with denaturation at 94 °C for 45 s, annealing for 30 s at 55 °C and extension at 72 °C for 1 min, and a final extension at 72 °C for 2 min. Amplification profiles were checked on 1% agarose gel and analyzed with a DNA 1000 chip on a Agilent 2100 Bioanalyzer (Agilent, USA). The RisaAligner software was used to transform fluorescence data so that the lowest negative value becomes 0 (as negative values have no biological meaning) and to normalize data [20]. Correlation between variables (Pearson correlation) was verified (XLSTAT v2016-03-30882, Addinsoft) and data were then analyzed by Principal Component Analysis (PCA) followed by Between Group Analysis (BGA) computed using Ade4TkGUI package of R software (version 3.2.2) [21,22]. Ascendant Hierarchical Clustering (AHC) was calculated with XLSTAT and bootstrap probability values were obtained from pvclust package (version 2.0-0) of R software [23].

2.3. High throughput sequencing (HTS)

DNA samples were subjected to HTS by Illumina MiSeq technology according to Illumina recommendations (16S Metagenomic Sequencing Library Preparation, Part #15044223_B) with the V3 chemistry (2 × 300 bp). Pooled libraries at 6 pM were spiked with 5% PhiX DNA (Illumina, UK) for sequencing in our lab. Data were analyzed using the Microbial Genomics Module of the CLC Genomics Workbench software (Qiagen, Denmark) following the dedicated tutorial "OTU Clustering and Analysis of Microbial Communities" available on their website (http://www.clcbio.com/files/tutorials/OTU_Clustering_Microbial_Analysis.pdf). During the trimming process, sequences were filtered on length to remove reads with less than 200 nt and more than 550 nt. Other steps were performed as described in the tutorial. After OTU clustering, a filtering step was inserted to suppress OTUs which combined abundance was below 12 reads (corresponding to the number of samples). Filtered OTU table was exported in text format. As data did not follow a normal distribution, it was log-transformed (after having added one to all values). Ascendant Hierarchical Clustering (AHC) was computed with XLSTAT. ANOVA was performed with PAST3 software [24]. PCA was computed with STAMP software [25].

3. Results

3.1. Blind test results according to RISA

The RISA data set was subjected to Between Group Analysis (BGA) using Principal Component Analysis (PCA) with R software. The first axis separated sites 2 and 3 from site 1, GTA and GTB samples (Fig. 1A), 36% of the total variance was explained by the first axis on the PCA (data not shown). Site 1 and sample GTA were separated from sample GTB by the second axis (22% on the PCA), as well as from sites 2 and 3. The cluster analysis (Fig. 1B) confirmed that site 2 and site 3 samples were different from other samples (bootstrap values = 100%), that site 1 was not significantly different from soil sample GTA and that soil sample GTB could be related to site 1, but not affiliated to (bootstrap value = 100%).