Australian Journal of Forensic Sciences

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tajf20

Forensic microbiology and geographical location: a systematic review

Bruna Moitas, Inês Morais Caldas & Benedita Sampaio-Maia

To cite this article: Bruna Moitas, Inês Morais Caldas & Benedita Sampaio-Maia (30 Mar 2023): Forensic microbiology and geographical location: a systematic review, Australian Journal of Forensic Sciences, DOI: 10.1080/00450618.2023.2191993

To link to this article: https://doi.org/10.1080/00450618.2023.2191993

AUSTRALIAN JOURNAL OF FORENSIC SCIENCES
https://doi.org/10.1080/00450618.2023.2191993

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Forensic microbiology and geographical location: a systematic review

Bruna Moitas[a], Inês Morais Caldas[b,c,d] and Benedita Sampaio-Maia[b,e,f]

[a]Departamento de Ciências da Saúde Pública e Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto, Porto, Portugal; [b]Faculdade de Medicina Dentária da Universidade do Porto, Porto, Portugal; [c]CFE - Centre of Functional Ecology, University of Coimbra, Coimbra, Portugal; [d]TOXRUN – Toxicology Research Unit, University Institute of Health Sciences, CESPU, CRL, Gandra, Portugal; [e]i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; [f]INEB - Instituto Nacional de Engenharia Biomédica, Universidade do Porto, Porto, Portugal

ABSTRACT
Establishing geographical location is a crucial aspect of forensic sciences, distinguishing between primary and secondary crime scenes, linking an individual to a crime scene, or detecting sources of disease. Microorganisms can be used as geolocation indicators since microbial communities vary according to climatic factors (e.g. temperature, humidity, soil properties, altitude). Therefore, this systematic review aimed to investigate whether the human or environmental microbiomes help to determine a crime's geolocation. Articles were searched in PubMed,Scopus and Web of Science using keywords and data fields. The final selection included seven (of 172) manuscripts. The results showed that the microbial profile of either human or environmental samples have the potential to link a cadaver or a crime scene to a given location, highlighting microbes' usefulness in obtaining information from geographical locations (e.g. soil samples from a suspect's shoe matched to a source). However, research is required before applying this forensic strategy to real scenarios. For instance, optimizing and standardizing the microbiome analysis methods and determining several factors that may influence the results.

ARTICLE HISTORY
Received 23 July 2022
Accepted 13 March 2023

KEYWORDS
Bacterial DNA; forensic microbiology; forensic sciences; geographical location; human microbiome; soil microbiome

Introduction

Forensic microbiology applies microbiological methods to criminal and medicolegal investigation[1–5]. This forensic science is responsible for analyzing and interpreting microbial evidence, mainly in biocrime, bioterrorism, human identification, estimation of postmortem interval, and geolocation[1–8]. Microorganisms are a phylogenetically diversified group, subdivided into bacteria, archaea, fungi, microalgae, protozoa, metazoans, and viruses[9,10]. These microorganisms form a complex and ubiquitous community, establishing commensal, symbiotic, or pathogenic relationships with man. Therefore, microorganisms contribute to the balance between health and

CONTACT Inês Morais Caldas ✉ icaldas@fmd.up.pt
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00450618.2023.2191993

disease[11,12]. For geographical location determination, the human microbiome analysis presents advantages over other methods. Microbes are present in all seasons and habitats, including the most extreme ones[3,5,6,9]. Most human associated microorganisms are resistant to degradation due to the presence of the cell wall and the ability to form biofilms, namely bacteria and fungi[13]. The diversity of the microbial communities is as diverse as a fingerprint, distinguishing even monozygotic twins[15].

For geolocation, microbes can be used to distinguish between primary and secondary crime scenes, locate clandestine graves, and identify suspects. These associations are possible because microbial communities differ in composition and function depending on geographical locations[14–16], climate (precipitation rates, altitude, temperature, and soil properties) and host properties or energy sources available in the environment[17]. The association between the victim's or suspect 's microbiome can establish links between them and the crime scene in cases of human trafficking or source of disease[2–4,18].

This systematic review aims to assess how microbiology can be used as a forensic tool to establish a geographical location, linking the victims or suspects to a location, linking the victims or suspects to a location using the human microbiome.

Materials and methods

This review followed the Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) tool and guide[19]. The review has also been submitted for registration on Prospero, and has the following registration ID CRD42022376240.

The scientific articles chosen for this review were selected from PubMed, Scopus and Web of Science database between September and November 2022, with the query (forensic or forensics) AND ('microbiology*' or 'microbiom*' or 'microorganism*' or 'microbe*' or 'microbiota' or 'microflora' or 'microbia*' or 'bacteria' or 'fungi' or 'yeast' or '*Streptococcus*' or '*Candida*') AND ('geographic* location' or 'geolocation*')This query intended to respond to the following PICO question: 'In human samples, how can microbiology assist forensic science practice through the analysis of the transmission of microorganisms between the subject and a place to establish the geolocation of a crime scene?'

First, articles corresponding to reviews, systematic reviews, and metaanalyses were excluded. Afterwards, articles were selected progressively, starting by reading the title, then the abstract and, finally, by reading the full article (Supplementary Figure S1). The eligibility assessment for each article was made independently by the three authors. Disagreements were resolved by consensus, excluding all those who did not meet the established inclusion criteria.

Data were extracted from each primary study and organized into a table, including title, authors, year of publication, population (number and type of participants), the type of study, the main objective, the intervention (microbial group assessed and method of analysis), and the outcome (significant findings and quantitative results). For risk of bias analysis in the individual studies, the Joanna Briggs InstituteFaculty of Health and Medical Sciences at the University of Adelaide protocol was followed[20]. In all included articles, the analysis was separately conducted by the authors. For each question in the protocol, articles were classified for the risk of bias as 'no', 'yes' or 'unclear'. For each yes, a point was given, and articles scoring six or more were selected for this review (Supplementary table S1).

Results and discussion

The selected studies were published in English and intended to evaluate the potential of microbial samples as tools for forensic investigation[16,17,21–24] (Table 1). From a total of 172, the final selection included seven manuscripts. In these studies used two different strategies to link the microbial profile to the geolocation. Three papers used the analysis of the microbiome of human samples: paper, saliva[23], stool and saliva[21], and the last up to 54 different body sites[16]. The other four papers used the microbial communities of environmental samples: one did not specify[17], one used dust at regional, national and global level[25], and the other two studies used soil samples[22,24]. Regarding microbial assessment methodology, high throughput 16S rRNA gene sequencing was performed in all studies but Habtom et al.[22] also used terminal restriction fragment length polymorphism (TRFLP). In the following two sections the studies using either human or environmental samples are described.

Human samples

The work of Clarke and colleagues[21] wants to understand how microbiome samples (oral and stool) from four geographically divergent populations could be used to distinguish between them, even for populations living in the same city. Microbiota profiling was performed targeting the V4 region of the 16S rRNA gene from the 206 samples, and the female participants were from one of four geographic regions: Barbados, Santiago (Chile), Pretoria (South Africa), and Bangkok (Thailand). By analysing the stool microbiome, the top five dominant taxa were *Bacteroides, Prevotella_9, Faecalibacterium, Alistipes*, and unclassified *Eubacterium*. Differences were found in geographical divergent populations, with, for example, higher abundance of *Faecalibacterium* in South African individuals and lower abundance in the individuals from Thailand. The data suggested that in populations with similar diets, the most geographically distinct taxa was in lower abundance in the stool (10.4% of the total gut microbiome), so they analysed the influence of the lifestyle on stool microbiota. Smoking was correlated with never smoking, living with a smoker or being an exsmoker. Also, BMI categories and diet (corn/corneal) had little influence, and only on stool microbiota. The oral microbiome was analysed, and the phyla's Bacteroidetes and Proteobacteria demonstrated significant differential abundance between countries. The most dominant genera among oral microbiota were *Prevotellaceae, Pasteurellaceae_unclassifed, Haemophilus, Streptococcus, Gemelia, Veillonella*, and *Neisseria*. Furthermore, oral microbiota could differentiate geographic locations with 16% variation between countries, where Chilean communities were the most geographically distinct, so that oral microbial community composition may vary according to the lifestyle of populations, as well as stool microbiota. Differences in oral microbiome composition between different geographical regions were also found in the study by Liang and colleagues[23], who analysed 70 saliva samples by 16S rRNA gene V3V4 sequencing. This authors also compared samples of other fluids and tissue: vaginal secretions, semen and skin; and the results showed that the dominant genus in each body region is different, being for vaginal secretions the genus *Lactobacillus* (69.02%), *Corynebacterium* (16.38%) for semen and *Cutibacterium* (70.13%) for skin. It has also been shown that body fluids can be clearly distinguished, as 49 of the 50 samples analysed

Table 1. Summarization of the information obtained from the articles under analysis.

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
21	206 stool and oral samples from four globally diverse populations: Barbados (n = 32); Santiago, Chile (n = 69); Pretoria, South Africa (n = 37); and Bangkok, Thailand (n = 68)	To understand how the oral and stool microbiome can be used to distinguish between populations in different countries and the influence of diet and lifestyles.	16S rRNA gene V4 sequencing.	Regarding stool microbiome, the top five most dominant taxa identified were *Bacteroides, Prevotella_9, Faecalibacterium, Alistipes*, and unclassified *Eubacterium. Faecalibacterium* was observed at significantly higher abundance in South African individuals and lower abundance in the Tai individuals. Chilean stool microbiota correlates with having never smoked; Pretorians correlates with BMI categories and eating corn/cornmeal; Tai correlated with living with a smoker and stool microbiota of Barbadian population is not significantly correlated with any. In the oral microbiota, the dominant taxa are *Prevotellaceae, Pasteurellaceae_unclassifed, Haemophilus, Streptococcus, Gemelia, Veillonella*, and *Neisseria*. The oral microbiota is also diet-related and therefore contains on average more geo-specific bacteria (16%), compared to the stool microbiota. Chilean oral microbial communities were 17% taxa of variation, 9% to Pretorian and 4% for Barbadian for geographical distinction. To evaluate the intra-region variation, Chilean and Barbadian were subdividing in two regions: The Chilean neighbourhoods do not have a significant difference between the oral or faecal microbiomes, only *Family XI Gemella* demonstrated differences in abundance; No taxa in the Barbados samples were identified as significantly abundant, with the exception of one faecal taxon (*Prevotellaceae Prevotella v9*).	YES

(*Continued*)

Table 1. (Continued).

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
[25]	1816 dust samples (continental USA n = 1301; North Carolina's tricounty n = 116 and global data n = 399)	Test a new algorithm that estimates the intensity surface of a spacial point pattern and determine the scales at which geolocation is feasible, using microbioma data	DeepSpace (geolocation algorithm using deep neural network classifiers)	Compared the models: Spatial NN; Spatial RF; Spatial Net; DeepSpace; BDA; Area DNN and partitioning schemes: coarse; fine; mixed and none. Regarding, National data, North Caroline contributes the most samples despite being the ninth most populous state. Reduction in prediction error as the complexity of spacial classifier increases (Spatial NN to Spatial RF to Spatial Net to DeepSpace). DeepSpace outperforms Spatial NN, Spatial RF and Spatial Net in predicting the origin of a sample.Soatial RF and DeepSpace are biased towards populated urban area: Spatial RF does not detect quite as many regional patterns as does Deep Space (identifies more areas with low prediction error throughout the Northeast, Midwest and Western USA. At regional level, BDA and county Area DNN achived the lowest prediction errors, however Area DNN has the highest accuracy compared with the spatial models. In global scale, among the three portioning schemes, spatial models perform best with the miced partitions which achieved country classification rates of 62,7% for Spatial NN, 74,9% for Spatial RF, 84,2% for Spatial Net and 89,5% for DeepSpace. The models had difficulties in distinguishing samples from the bordering countries of Uruguay and Argentina (DeepSpace misclassifies 3 samples and Area DNN misclassifies 10 samples). Samples for Croatia were often misclassified with countries nearby, by DeepSpace, only Macedonia (close to Croatia) was always correctly classified but Area DNN misclassified samples from both. DeepSpace often predicted samples from Omar to be from Qatar and Area DNN did not. This suggests that are regional patterns within and between countries.	YES

(*Continued*)

Table 1. (Continued).

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
[22]	Soil samples from five research sites in Israel around a 260 km line from the desert to Mediterranean climatic region: 2–4 different soil types were sampled per site. Six soil types were address: loess, desert Skeletal soil, sand, Mediterranean mountain soil, rendzina, and terra rossa.	Evaluate bacterial distribution and assess their resolution at local to regional scales, between and within different soil types.	Terminal restriction fragment length polymorphism (TRFLP) and 16S rRNA gene V4 sequencing.	TRFLP analysis detected 447 terminal restriction fragments, 29 were found in all sites and soil types and 55 were specific to a single soil type at specific site. At local scale (25–1000 m), differences between bacterial communities were primarily driven by soil type. At regional-scale distances (1–260 km), the bacterial communities from different soil types were still significantly different and the physico-chemical environment was the dominant factor determining the community profile, namely the annual precipitation, and the soil levels of sodium and ammonium. Distance-decay relationship: The greater the distance between two geographic locations, the greater the difference between the bacterial communities present. This was observed from distances to ten metres to 10 km (2 m apart: not significant differences). Although less important than geographical distance, soil type also relevant for microbial community. Main phylum in terra rossa, rendzina and sand are Proteobacteria and Actinobacteria. At genus level, the most abundant genera were *Rubrobacter, Microvirga* and unidentified Acidobacteria in rendzima and terra rossa; in the sand was *Microvirga, Arthrobacter* and *Bacillus*.	Yes

(*Continued*)

Table 1. (Continued).

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
[17]	305 environmental samples from 16 cities* for the training dataset (from the Metagenomics and Metadesign of Subways and Urban Biomes – MetaSUB). Plus 61 mystery samples from unsampled cities. *Auckland, Berlin, Bogota, Hamilton, Hong Kong, Ilorin, London, Marseille, New York, Offa, Porto, Sacramento, Sao Paulo, Sofia, Stockholm, Tokyo	Machine learning framework to determine the geolocations from metagenomics profiling of microbial samples.	Metagenomics DNA Illumina sequencing	Number of samples per city varies between 10 and 26. Taxa at species level play a significant role in the selected features (41 out 50 are at species level). 38% are human pathogens and bacterial phytopathogens enriched at species (*Salmonella enterica subsp. Enterica serovar Choleraesuis Staphyloccus capitis subsp. Capitis, Fusobacterium hwasookii, Bacillus flexus, Streptococcus salivarius)*. Other prevalent bacterial clades are involved on carbon and nitrogen of metabolism (*Lactobacillus delbrueckii subsp. Bulgaricus, Rubrobacter xylanophilus DSM 9941, Cellulomonas flavigena, Comamonas aquatica)*. This geolocation prediction framework successfully assigned a certain number of samples to their cities of origin with high probabilities. Samples achieving high probabilities on their ground truth cities, the resulting distribution is significantly lower than those obtained from real training data. Interpolation results from biological coordinates show much higher confidence compared to those from geographic coordinates.	Yes

(*Continued*)

Table 1. (Continued).

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
23	70 saliva samples from Guangdong, Qinghai, Henan, Zhejiang, and Jilin (14 from each location)	Understand how regional location influences saliva microbial profile	16S rRNA gene V3-V4 sequencing	OTU abundance index of bacteria in Henan samples was the highest. Firmicutes represented the most abundant bacteria in all samples, followed by Proteobacteria, Actinobacteria, and Bacteroidota. Regarding genus, *Streptococcus* was the dominant bacterium in saliva, followed by *Neisseria, Rothia, Porphyromonas*, and *Granulicatella*. Relative abundance of *Rothia* in Jilin and Qinghai was higher than in other regions. Significant differences in microbial community abundance at the genus level were found between the different regions, including *Streptococcus, Neisseria, Rothia, Granulicatella, Prevotella, Gemella, Haemophilus, Actinomyces, Veillonella, TM7x, Leptotrichia, Salmonella, Acinetobacter, and Capnocytophaga*. The dominant genus of each body fluid was different: the dominant bacteria in the vaginal secretions were *Lactobacillus, Corynebacterium* for semen, and *Cutibacterium* for skin. The four body fluids and tissues could be clearly distinguished. Saliva bacteria show potential in geographical inference.	Yes

(*Continued*)

Table 1. (Continued).

Reference	Samples	Main goal	Method of microbial assessment	Major findings	Geolocation determination
[24]	90 soil samples from ten sites within the Greater Wellington region, each site with three patches (each patch, three samples).	Determine whether the characterization of soil bacterial and fungal community structures can be used to discriminate soils	16S rRNA sequence for bacteria and ITS for fungal	Bacterial community was significantly influenced by site: only three of 45 site comparisons could not be discriminated and one site discriminated at an elevated level from all other sites. Clear differences in bacterial community structure between some sites while other sites shared a more similar community structure. Twenty-seven sites could be discriminated at the highest significance level ($p \leq 0.1\%$), nine at $p \leq 1\%$, and six at $p \leq 5\%$. The fungal community was significantly influenced by site. Forty-four of the 45 comparisons could be discriminated; Taita and Otari were the only two parks that could not be discriminated between. Thirty-three of 45 discriminated from each other at the highest level of significance ($p \leq 0.1\%$) with nine discriminating at $p \leq 1\%$ and two at $p \leq 5\%$. Combining bacterial and fungal profiles did little to discriminate further between patches than was provided by fungi alone, showing 44 of the 45 site comparisons some significant level of discrimination, with Taita and Otari being the only two sites that could not be discriminated between.	Yes
[16]	16S rRNA sequences from 20,820 human samples obtained from 54 different body sites of individuals from 138 cities of 35 countries.	Forensic microbiome database (FMD) provide the scientific and non-scientific communities with data and tools to explore the possibilities of microbiomes to answer forensic questions and serve as model for any future databases.	16S rRNA sequencing	The majority of 16S rRNA data was obtained from stool samples, followed by saliva and other oral locations. Accuracy is 80,5% for cities, 81,5% for state/region, and 92,1% for countries. The accuracy ranges from 61% for retro auricular crease to 93% for saliva samples. Stool samples reached a 78% prediction accuracy. Incorrectly predicted vagina samples which constitute 13% of all vagina samples. The database and the website will facilitate exploration of the taxonomic underpinnings of geolocation signals, both through dynamics explorations of the taxonomic distributions of microbiomes from different geographic locations through comparison of the data samples in combination with user supplied metadata.	Yes

were correctly associated with the fluid/tissue of origin, only one semen sample was mistaken for vaginal. Additionally, the random forest model was used to predict the value of microbial salivary markers to distinguish regionally, using 21 saliva samples from the five regions and the five main eigenvalues are *Oribacterium, Peptostreptococcus, Haemophilus, Veillonella* and Saccharimonadaceae. The results showed that 16 of the 21 test samples were correctly classified, demonstrating the potential of the model to recognize saliva samples from the regions under study.

In conclusion, the analysis of oral and stool microbiome can provide important information regarding the geographically location and the influence of populations' diets, behaviours, and lifestyles. Furthermore, the dominant genus in each body region is different. However, it is necessary to collect comprehensive microbial data from different geographical locations, distinct soil types, from local to continental levels, with larger sample sizes, and from different body parts, but also to understand how external factors (e.g. diet, environment) may influence the structure of each microbiota community. After this, the development of a reference database for subsequent comparison with samples of unknown origin, is necessary for the use of the microbiome as a tool to make inferences about geographical location. Additionally, for machine learning algorithms, the sample size of the study by Liang and colleagues[23] is considered reduced. At a sample level, the criteria established for the selection of participants can condition the results and the influence of other factors that affect the oral microbiota. More studies with a larger sample are needed to overcome these limitations and exhaustively record individual oral characteristics, including diet and other habits that may influence salivary microbial structure. Despite this, this study allowed us to conclude that the analysis of the microbial community in saliva can give us important information about body fluid traceability and geographic inference.

Singh and colleagues[16] introduce the forensic microbiome database (FMD). This dataset provides data and tools to explore the possibilities of microbiomes to answer forensic questions, serving as a model for other databases. This work uses 20,820 samples collected from different body sites from people from diverse geographic location (35 different countries corresponding to 138 cities). Most samples (approximately 50%) were obtained from stool samples, followed by saliva and other oral locations. All samples were subjected to 16S sequencing. The results showed that, when body sites with more than 150 samples in the database were considered (96% of the data), the accuracy was 80.5% for cities, 81.5% for state/region, and 92.1% for countries. Also, the prediction accuracy ranged from 61% for retroauricular crease to 93% for saliva samples. It was observed that similar body sites are cross predicted. This crossprediction is negligible between the oral cavity, skin, vagina, and stool, demonstrating the unique microbiome composition of different body sites. At last, the incorrect samples (20.5%) were analysed to understand the impact of distance on incorrect predictions. In the case of incorrectly predicted vagina samples (13% of all vagina samples), the average distance is 7000 km. The vagina samples predicted as stool samples were dominated by the same genus, suggesting either cross-contamination or biological/technical contamination, which explains the considerable variation in incorrect samples' distance.

The significant limitations to the use of the microbiome as forensic evidence are associated with the lack of standardized collection and storage protocols, the influence of external factors that may induce sample degradation or contamination, the sensitivity and species discrimination by sequencing techniques, the privacy of subjects' genetic data, and the accuracy and robustness of statistical data[2,26].

Environmental samples

Grantham and colleagues[25] used 1816 dust samples collected at three levels (global (30 countries on 6 continents), national (continental USA), and regional) to propose a new forensic geolocation algorithm for estimating the strength of a spatial point pattern using deep neural network classifiers trained on random Voronoi partitions of the spatial domain. This algorithm can approximate the conditional continuous distribution of a sample's provenience given its microbial composition by adapting a deep neural network classifier to each partition and calculating the average over the partitions. Also, it was compared with other geolocation algorithms (Spatial NN; Spatial RF; Spatial Net; BDA and Area DNN) and partitioning schemes (coarse; fine; mixed and none).

At the regional scale, spatial models perform similarly across partitioning schemes, with BDA and Area DNN models achieving lowest prediction errors and Area DNN having the highest accuracy rate (53.4%). By focusing on a small geographic area, we can insulate the capacity of the models to predict the source of a sample when biogeographic differences are kept roughly constant and as demonstrated in the countrylevel analysis, DeepSpace is able to learn regional patterns in the data, which could improve its classification accuracy.

Starting at the national level, the results indicate that spatial models, DeepSpace outperforms Spatial NN, Spatial RF and Spatial Net in predicting the geographic origin of samples (state, county, and city). Regarding the partitioning schemes, fine has high prediction errors but slightly lower coverage probabilities, coarse has the lowest prediction error but coverage probabilities are overestimated, and mixed partitioning is the balance between the above. There is a decrease in prediction error as the complexity of the spatial classifier increases (Spatial NN to Spatial RF to Spatial Net to DeepSpace). In the coarse partition, Spatial RF and DeepSpace are biased towards populated urban areas since more data are available than for the surrounding rural areas, and DeepSpace detects more regional patterns than Spatial RF. On a national scale, differences in fungal occupancy are likely to reflect both biogeographic differences in terms of which taxa occur were and differences in local habitats.

At the global level, the models were compared for their capacity to capture the country of provenance of the sample. Area DNN achieves a high classification rate of 84.7%. Among the three partitioning schemes, the spatial models perform best with the mixed partitions achieving country classification rates of 89.5% (DeepSpace), 84.2% (Spatial Net), 74.9% (Spatial RF) and 62.7% (Spatial NN). The models have a difficulty in distinguishing among samples from the countries bordering Uruguay and Argentina, but DeepSpace misclassifies only three samples in comparison to 10 samples misclassified by Area DNN. Samples for Croatia were often misclassified as being from neighbouring countries by DeepSpace. Interestingly, Macedonia in contrast, which is very close to Croatia, was always properly classified by DeepSpace and misclassified by Area DNN. In other regions, the performance of the models was inverted: DeepSpace often predicted Oman samples to be from Qatar, while Area DNN did not. Overall, DeepSpace achieves noticeably fewer errors than the Area DNN when classifying country. This suggests that there are regional patterns within and between countries that a pointlevel model can exploit for more accurate forensic geolocations.

In conclusion, DeepSpace is a geolocation algorithm that combines random spatial partitions with deep learning classification and has been applied to three spatial scales: regional, national, and global. At regional level, the results obtained were not satisfactory since none of the methods reaches accuracy percentages higher than 53% but, on the contrary, at national and global level the results were very good, presenting error rates lower than 100 km at continental USA level and 90% correct classification in 28 countries at global level. Despite these results, limitations were identified at the sampling level:[1] the dust samples used were neither randomly nor systematically collected at any spatial level and[2] samples were collected from rarely disturbed surfaces (exterior doors and windows). Future studies could combine the analysis of fungi and bacteria, including other characteristics (besides spatial coordinates) such as soil type or seasonality improving the results obtained.

Habtom and colleagues[22] collected samples from five different sites across the rainfall gradient of Israel, to assess the bacterial community present in the different soil types and to understand the differences at local (metres) and regional (kilometres) level. The analysis of the bacterial community by TRFLP detected 447 TRFs, of which 6.5% were found in all soil types, 12.3% were specific to one site and soil type, and not found in other samples. Also, this analysis showed that site location is more important than soil type in determining the microbial community structure, as geographic locations formed clusters without regard to soil types. In contrast, even soil types from different locations did not form clusters. Despite this, bacterial communities differed significantly depending on soil type, bearing in mind that precipitation is highly correlated with soil community composition at all sites. Physico-chemical analyses were performed demonstrating that microbial structure is correlated with sodium and ammonium levels in the soil. The soil parameters with the most significant influence on community structure differed by region; the following factors were important: water saturation, levels of sodium, potassium, phosphorus and organic matter.

This work also analysed the differences between bacterial communities depending on the distances between them and determined that there is a significant correlation in the sense that the further away two communities are, the more distinct they will be, and this relationship is not significant at 2 metres of distance. However, it is observed from ten metres to ten kilometres, in all types of soil. In addition, the bacterial composition of three adjacent soils (rendzina, terra rossa and sand) was analysed, collecting five samples of each type, in one region. Starting at the phylum level, Proteobacteria and Actinobacteria are the main ones with significantly higher Bacilli levels in the sand and significantly lower planctomycetacia and verrucomicrobia, not being able to distinguish between rendzina and terra rossa, and sand was not well differentiated from them. At the level of taxonomic genus, the communities of the three soils are significantly different from each other, with sand showing less diversity compared to rendzina and terra rossa (more similar to each other) but with more significant variability between communities of the same type of soil (sand). The most abundant genera in rendzina and terra rossa soils were Rubrobacter, Microvirga and unidentified Acidobacteria while in sand they were Microvirga, Arthrobacter and Bacillus. Lastly, this work also addressed the forensic evaluation of microbial evidence by likelihood ratio (LR): six types of soil (loess, desert skeletal soil, sand, Mediterranean mountain soil, rendzina, and terra rossa) were used as a test set for evaluation of soil microbial evidence through modelling and validation of the LR, getting the accuracy of Cllr = 0.57 (measured as the cost of log-likelihood ratio).

Also, it was possible to statistically and repeatedly distinguish between different geographic locations in the same soil type and different soil types in the same geographic location, thus demonstrating the potential use of the soil microbiome as a tool in forensic sciences. However, it is still necessary to overcome temporal limitations (collecting samples at different points), sample storage and lack of standardized protocols to analyse soil DNA and the impossibility of analysing samples that have soil mixtures.

Huang and colleagues[17] presented a machine-learning framework to determine the geolocations from metagenomics profiling microbial samples. This work included 305 environmental samples from 16 cities in the training set, plus 61 mystery samples from other cities. The data set originated from the multi-source microbiome data from MetaSUB International Consortium used in the CAMDA 2019 Metagenomic Forensics challenge. Results showed that samples from one city generally cluster in a distinct group. Therefore, different cities corresponded to separated groups based on specific species profiles. As the multiclass classifier only presents probabilities of sampled cities, these cases present a limitation to the study. To overcome this limitation, the authors resorted to Kriging Interpolation (originated by geostatistics to estimate the probabilities of 'filling' spatial locations between sampled cities) to produce the optimal linear unbiased prediction of intermediate values under the assumption of broad sense stationary of covariance on the map. Using Kriging's interpolation, the authors assumed that 'everything is related to everything else, but near things are more related than distant things' (Tobler's First Law of Geography). This assumption is only sometimes valid since, in some cases, geographically further away cities may have more similar abundance profiles than closer cities. This represents another limitation overcome by setting a biological coordinate system on which the distance between cities better reflects their similarity in terms of biological differences. Based on the biological coordinates of all sampled cities, the biological coordinates of unsampled cities can be derived by applying affine transformation between the biological and geographical coordinate systems, using the coordinates of sampled cities as anchor points. Another limitation was the poorly performance of algorithm on some testing samples. The low performance was possibly due to the limited number of cities from the training set compared to the size of the geographic coverage of those cities, as well as the small number of available training samples for each city, which may not cover all potential microbiome from the city.

In conclusion, this geolocation prediction framework successfully assigned samples to their cities of origin with high probabilities. Also, the interpolation results from biological coordinates show much higher confidence than those from geographic coordinates. This implies that biological coordinates better reflect the deviation of the abundance profiles of metagenomic samples of different locations for deducting the geolocation of unsampled sites. The proposed method provides accurate predictions of the geolocation of microbial samples using selected abundance profiles as features.

Macdonald and colleagues[24] tested the ability of terminal restriction fragment length polymorphism (T-RFLP) to distinguish between soils, to determine if profiling of bacterial and fungal community structures could be used to discriminate between soils collected from distinct locations in Greater Wellington region of New Zealand.

Variations in the microbial community between ten sites were assessed. Within each site, three distinct patches ranging from five to sixty metres apart were identified, with a total of 90 samples. Bacterial communities were analysed by 16S rRNA sequencing and fungi by ITS region, between 18S and 23S regions. This worked showed that the biogeographic location influenced the structure of bacterial and fungal communities. For bacteria, 27 sites could be discriminated at the highest significance level ($p \leq 0.1\%$), nine at $p \leq 1\%$, and six at $p \leq 5\%$. For fungal community, between 17 and 31 TRFs were obtained which showed that some sites share a more similar fungal community while others show obvious differences. When combining the analysis of the communities 44 of 45 site comparison showed level of discrimination.

No discrimination between vegetation classes was identified for bacteria or fungi. Within each site, samples were collected from three different patches (A, B, and C) which were also subjected to comparison. For most regions, there is a significant effect of patch on microbial community structure within the site, both bacteria and fungi. Although the degree of discrimination varied between sites and it was not possible to discriminate between all soils, results demonstrated that samples obtained from different soils within the Greater Wellington region could be discriminated to different degrees based on microbial profiles, even though they have underlying geological similarities, with fungal community structure generally demonstrating the high discriminatory potential between sites than bacterial. The combination of bacterial and fungal analysis only added less information than the fungal community structure provided alone, as variability within each patch is more significant in the bacterial community than in fungi.

In sum, this study has shown differences in the structure of soil bacterial and fungal communities and that the structure varies along environmental gradients, even at local and regional scales. Therefore, soil microbial community analysis can be helpful when it is necessary to associate a soil trace found in a forensic context with a geographical location. However, much research is needed to overcome several limitations or questions that remain, including further studies to understand how specific taxonomic groups can provide better discriminatory information, how vegetation structurally influences the soil microbial community, and exhaustive sampling in different habitats and soils, to understand variation at different scales better.

Conclusion

Although only seven geolocation studies were included, we can infer that it is possible to obtain information about a location through either human or environmental microbiome analysis. However, further studies are needed to overcome some limitations for the microbiome to be used in forensic analysis for geographic location. These studies should allow a better knowledge of sampling and analytical limitations, elucidating the influence of external factors (e.g. climate and diet) on the microbiome, and obtaining more data to relate the microbiome from a cadaver to a given location. Although further studies are required, this work demonstrates that the analysis of microorganisms within and surrounding the cadaveric island can reveal relevant information for the establishment or estimation of the geographical location.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

1. Garcia MG, Perez-Carceles MD, Osuna E, Legaz I. Impact of the human microbiome in forensic sciences: a systematic review. Appl Environ Microbiol. 2020;86(22):22. doi:10.1128/AEM.01451-20.
2. Clarke TH, Gomez A, Singh H, Nelson KE, Brinkac LM. Integrating the microbiome as a resource in the forensics toolkit. Forensic Sci Int Genet. 2017;30:141–147.
3. Oliveira M, Amorim A. Microbial forensics: new breakthroughs and future prospects. Appl Microbiol Biotechnol. 2018;102(24):10377–10391. doi:10.1007/s00253-018-9414-6.
4. Schmedes SE, Sajantila A, Budowle B, Kraft CS. Expansion of microbial forensics. J Clin Microbiol. 2016;54(8):1964–1974. doi:10.1128/JCM.00046-16.
5. Metcalf JL, Xu ZZ, Bouslimani A, Dorrestein P, Carter DO, Knight R. Microbiome tools for forensic science. Trends Biotechnol. 2017;35(9):814–823. doi:10.1016/j.tibtech.2017.03.006.
6. Finley SJ, Benbow ME, Javan GT. Microbial communities associated with human decomposition and their potential use as postmortem clocks. Int J Legal Med. 2015;129(3):623–632. doi:10.1007/s00414-014-1059-0.
7. Metcalf JL. 2019. Estimating the postmortem interval using microbes: knowledge gaps and a path to technology adoption. Forensic Sci Int Genet. 38:211–218. doi:10.1016/j.fsigen.2018.11.004.
8. Singh B, Minick KJ, Strickland MS, Wickings KG, Crippen TL, Tarone AM, Benbow ME, Sufrin N, Tomberlin JK, Pechal JL, et al. Temporal and spatial impact of human cadaver decomposition on soil bacterial and arthropod community structure and function. Front Microbiol. 2017;8:2616. doi:10.3389/fmicb.2017.02616.
9. Alan G, Sarah JP. Microbes as forensic indicators. Trop Biomed. 2012;29(3):311–330.
10. Massey SE, Cano RJ, Toranzos GA. Comparative microbial genomics and forensics. Microbiol Spectr. 2016;4(4):4. doi:10.1128/microbiolspec.EMF-0001-2013.
11. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804–810. doi:10.1038/nature06244.
12. Lederberg J, McCray AT. ‘Ome sweet ‘omics - a genealogical treasury of words. Scientist. 2001;15(7):8.
13. Leake SL, Pagni M, Falquet L, Taroni F, Greub G. The salivary microbiome for differentiating individuals: proof of principle. Microbes Infect. 2016;18(6):399–405. doi:10.1016/j.micinf.2016.03.011.
14. Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, Gilbert JA. Forensic analysis of the microbiome of phones and shoes. Microbiome. 2015;3(1):21. doi:10.1186/s40168-015-0082-9.
15. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M, Wang L, Wang G, Wang F, Xu J, Cao H, et al. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. ISME J. 2015;9(9):1979–1990. doi:10.1038/ismej.2015.11.
16. Singh H, Clarke T, Brinkac L, Greco C, Nelson KE. 2021. Forensic microbiome database: a tool for forensic geolocation meta-analysis using publicly available 16S rRNA microbiome sequencing. Front Microbiol. 12:644861. doi:10.3389/fmicb.2021.644861.
17. Huang L, Xu C, Yang W, Yu R. A machine learning framework to determine geolocations from metagenomic profiling. Biol Direct. 2020;15(1):27. doi:10.1186/s13062-020-00278-z.
18. Santiago-Rodriguez TM, Cano RJ, Cano RJ, Toranzos GA. Soil microbial forensics. Microbiol Spectr. 2016;4(4). doi:10.1128/microbiolspec.EMF-0007-2015.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

20. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc. 2015;13(3):132–140. doi:10.1097/XEB.0000000000000055.
21. Clarke T, Brinkac L, Greco C, Alleyne AT, Carrasco P, Inotroza C, Tau T, Wisitrasameewong W, Torralba MG, Nelson K, et al. Sampling from four geographically divergent young female populations demonstrates forensic geolocation potential in microbiomes. Sci Rep. 2022;12 (1):18547. doi:10.1038/s41598-022-21779-z.
22. Habtom H, Pasternak Z, Matan O, Azulay C, Gafny R, Jurkevitch E. 2019. Applying microbial biogeography in soil forensics. Forensic Sci Int Genet. 38:195–203. doi:10.1016/j.fsigen.2018.11.010.
23. Liang X, Han X, Liu C, Du W, Zhong P, Huang L, Huang M, Fu L, Liu C, Chen L, et al. Integrating the salivary microbiome in the forensic toolkit by 16S rRNA gene: potential application in body fluid identification and biogeographic inference. Int J Legal Med. 2022;136(4):975–985. doi:10.1007/s00414-022-02831-z.
24. Macdonald CA, Ang R, Cordiner SJ, Horswell J. Discrimination of soils at regional and local levels using bacterial and fungal T-RFLP profiling. J Forensic Sci. 2011;56(1):61–69. doi:10.1111/j.1556-4029.2010.01542.x.
25. Grantham NS, Reich BJ, Laber EB, Pacifici K, Dunn RR, Fierer N, Gebert M, Allwood JS, Faith SA. Global forensic geolocation with deep neural networks. J R Stat Soc C, R Stat Soc. 2020;69 (4):909–929. doi:10.1111/rssc.12427.
26. Hanssen EN, Avershina E, Rudi K, Gill P, Snipen L. 2017. Body fluid prediction from microbial patterns for forensic application. Forensic Sci Int Genet. 30:10–17. doi:10.1016/j.fsigen.2017.05.009.