



Soil microbial geolocation in forensic casework: a mini-review of methods, evidentiary strength, and admissibility

Abdulla Albastaki^{a,b,*} , Judith Smith^b

^a General Department of Forensic Science and Criminology, Dubai Police HQ, Dubai, United Arab Emirates

^b School of Law and Policing, University of Lancashire, Preston, United Kingdom

ARTICLE INFO

Keywords:

Forensic microbiology
Soil forensics
Microbial biogeography
16S rRNA
Geolocation
Source attribution
Daubert standard
Trace evidence

ABSTRACT

Soil is among the most frequently transferred trace materials in violent crime, and its examination draws on an established suite of physical, chemical and biological methods whose resolution at fine spatial scales is nonetheless limited. Over the past decade, advances in high-throughput sequencing of the bacterial 16S rRNA gene, paired with supervised machine-learning classifiers, have been proposed as a means of discriminating soils that these methods cannot readily separate. This mini-review synthesises current evidence for the use of soil microbial biogeography in forensic geolocation, with specific emphasis on three questions of direct relevance to casework: (i) what spatial resolution is achievable with current sequencing and classification pipelines; (ii) how stable are location-discriminating taxa under the storage, transport, and environmental conditions typical of forensic exhibits; and (iii) whether the resulting evidence, as currently presented, can satisfy the gatekeeping standards articulated in *Daubert v. Merrell Dow Pharmaceuticals* and comparable frameworks outside the United States. We distinguish geographic discrimination, closed-set classification and source attribution, and show that reported accuracies in excess of 90 % derive from closed-set classification against small reference panels rather than from validated source attribution. We outline a minimum dataset that forensic microbial geolocation reports should contain before being offered in court. We conclude that the field has demonstrated promising discriminatory capability under defined conditions, but that forensic source attribution remains insufficiently validated, and that microbial profiling is best used alongside established soil examination rather than in place of it. We set out the infrastructural and empirical priorities that would close the gap between laboratory demonstration and routine evidentiary use.

1. Introduction

Soil is recovered from footwear, tyres, vehicle wheel wells, clothing, tools, and human remains in a substantial proportion of outdoor violent crimes, and its evidential value has been recognised since the earliest systematic writings on forensic science [1,2]. Forensic soil examination is an established multi-proxy discipline. Current practice draws on colour recorded under standardised conditions, particle size distribution, mineralogy by X-ray diffraction and scanning electron microscopy, multi-element geochemistry, infrared spectroscopy and palynology, applied in combination and interpreted together [1–3]. These methods are supported by published protocols and, in several jurisdictions, by accredited casework practice. Their principal limitation is one of resolution rather than reliability. Two soils drawn from sites several kilometres apart may be physicochemically indistinguishable, and the

probative weight of a “match” has therefore often been constrained to exclusionary statements rather than source-level attributions [4].

The emergence of next-generation sequencing (NGS) and, in particular, the routine amplification and sequencing of the bacterial 16S ribosomal RNA (rRNA) gene has opened a new analytical axis. Soil bacterial communities are shaped by an interaction of climate, edaphic properties, vegetation, and land use, and this produces taxonomic signatures that vary at spatial scales relevant to criminal investigation from metres to continents [5,6]. When these signatures are combined with supervised classifiers such as random forests, support vector machines, or deep neural networks, published studies have reported correct assignment of soil samples to their district or city of origin at rates exceeding 90 % [7,8]. These figures are obtained under closed-set conditions, in which the true source is known in advance to lie within a small panel of candidate sites, and they are not equivalent to error

* Corresponding author. General Department of Forensic Science and Criminology, Dubai Police HQ, Dubai, United Arab Emirates.

E-mail addresses: aalbastaki1@lancashire.ac.uk (A. Albastaki), JASmith@lancashire.ac.uk (J. Smith).

<https://doi.org/10.1016/j.fsisyn.2026.100746>

Received 18 August 2026; Received in revised form 5 October 2026; Accepted 5 October 2026

Available online 7 October 2026

2589-871X/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

rates for source attribution in casework.

These laboratory results have outpaced the forensic community's capacity to evaluate them in evidentiary terms. Review articles to date have focused largely on analytical methodology or on the broader field of forensic microbiomics, in which soil geolocation is one application among many [9,10]. The specific question of how a soil microbial geolocation report would, or should, fare under the gatekeeping standards applied to expert scientific testimony has received comparatively little attention. This mini-review addresses that gap. We restrict our scope to bacterial community profiling of soil for the purpose of geographic source attribution, excluding post-mortem interval (PMI) estimation and human skin- or gut-microbiome individualisation, both of which raise distinct evidentiary issues and have been reviewed elsewhere [10,11]. We do not propose microbial profiling as a replacement for established soil examination, and Section 2.4 sets out the case for using the two together. Our aim is to provide practicing forensic scientists who will ultimately weigh this evidence, with a concise synthesis of what the method can currently do, what it cannot, and what remains to be validated before it should be offered in court.

2. The scientific basis for microbial geolocation of soil

Four distinct inferential tasks are often described interchangeably in this literature, and the distinction between them governs the weight that can properly be claimed. Geographic discrimination asks whether two samples can be told apart, and makes no claim about location. Closed-set classification assigns a sample to one of “k” candidate sites, all represented in the reference data and one of which is assumed to be the source; its chance baseline is $1/k$. Open-set provenance prediction estimates an origin for a sample whose true source may not be represented in the reference data at all. Source attribution is the forensic proposition that a questioned sample came from a specified location rather than from some other location within a relevant population of alternatives, and it requires an estimate of how often indistinguishable profiles arise among those alternatives. Demonstrated performance in one task does not transfer to another, and most published figures in this field, including those above 90 %, describe closed-set classification. We use these terms in this sense throughout.

2.1. Soil microbial biogeography

The premise of soil microbial geolocation is that bacterial and archaeal communities vary non-randomly across space. Large-scale surveys, most notably those associated with the Earth Microbiome Project, have demonstrated that soil pH is the single strongest predictor of bacterial community composition at global and continental scales, with secondary contributions from mean annual temperature, precipitation, and vegetation type [5,12]. Within a single climatic region, community composition continues to vary at scales of metres to kilometres, driven largely by local edaphic heterogeneity and land use. Habtom and colleagues, using terminal restriction fragment length polymorphism (TRFLP) profiles, demonstrated that soil bacterial profiles could distinguish locations separated by as little as 25 m, establishing a lower bound on achievable spatial resolution with community-level markers [13].

The forensic implication is that a given soil sample carries, in its microbial community composition, information about where on the landscape it most probably originated. Unlike mineralogical parameters, which reflect the underlying parent material and are therefore broadly stable across large geological units, microbial signatures respond to the full suite of local biotic and abiotic factors, providing a discriminating signal even between nearby sites that share geological substrate. These relationships describe average behaviour across particular sampling designs rather than the behaviour of any given landscape. Where land use, vegetation or disturbance history changes abruptly, communities can differ over metres; across homogeneous terrain they may be near-

indistinguishable over kilometres [14]. Achievable spatial resolution is therefore a property of the landscape under examination as much as of the method.

2.2. Sequencing and bioinformatic pipelines

Contemporary soil microbial geolocation studies rely almost exclusively on high-throughput amplicon sequencing of hypervariable regions of the 16S rRNA gene, most commonly the V3–V4 or V4 regions, on Illumina MiSeq or NovaSeq platforms [15]. DNA is typically extracted using bead-beating-based commercial kits (for example, the Qiagen DNeasy PowerSoil series), which have become the de facto standard and offer reasonable reproducibility across laboratories. Reads are denoised into amplicon sequence variants (ASVs) using DADA2 or a comparable pipeline, taxonomy is assigned against curated databases such as SILVA, Greengenes2, or RDP, and the resulting feature table forms the input to downstream classification [16]. Extraction chemistry, primer choice, denoising parameters and reference database version each influence the resulting feature table, so profiles produced under different pipelines are not directly comparable.

Shotgun metagenomic sequencing offers greater taxonomic and functional resolution and avoids the primer-bias inherent to amplicon approaches, but at substantially higher cost and with more demanding bioinformatic requirements. For routine casework, where cost per sample and laboratory throughput are real constraints, 16S amplicon sequencing remains the pragmatic choice; shotgun approaches are likely to be reserved for research validation and for cases where amplicon data prove equivocal [9].

2.3. Classification and source attribution

The statistical problem posed by soil microbial geolocation is a high-dimensional supervised classification: the input is a sparse feature vector of thousands of ASVs, and the output is a categorical or continuous geographic label. Random forest classifiers have been the workhorse of the field owing to their robustness to high dimensionality, tolerance of sparse data, and capacity to provide feature-importance rankings that are useful when articulating the basis of a conclusion to a court [7,17]. Deep neural network approaches have more recently been applied, with some studies reporting modest improvements in accuracy over random forests but at the cost of interpretability [18].

A critical limitation is that classification accuracy is highly dependent on the strength of the reference database and the geographic distance between candidate classes. Performance figures of 90 % or greater have generally been reported where the classifier is asked to choose among a small number of well-separated sites, each represented by multiple reference samples. Performance degrades, sometimes sharply, as candidate locations become more numerous, more proximate, or more environmentally similar [7]. This has direct implications for casework: the relevant question is rarely “which of these five sites did the sample come from?” but rather “could this sample have come from the suspect's garden, or from any of the thousands of plausible alternative sites in the region?” Reference databases capable of supporting the latter question are not yet widely available. Reported accuracy is therefore a property of the study design as much as of the method, being conditional on the number of candidate classes and the corresponding chance baseline, their separation, the number of reference samples per class, and the use of internal cross-validation rather than blind external testing. Such a figure should not be offered to a court as the error rate of the technique.

2.4. Relationship to established soil examination

Microbial profiling is best considered alongside established soil examination rather than as a replacement for it. The proxies carry different information: mineralogy and geochemistry reflect parent material and

long-term pedogenesis, palynology reflects vegetation and season, and microbial composition reflects contemporaneous local conditions. They also fail under different conditions, in that microbial communities change once soil is removed from its setting whereas elemental composition does not, while elemental composition may be uninformative where candidate sites share parent material. Agreement between independent proxies constrains the range of possible sources more effectively than any one of them alone, and concordance between independent lines of evidence is a recognised basis for strengthening an evaluative opinion.

Published casework reflects this. Concheri and colleagues combined multi-element analysis with bacterial community fingerprinting to link soil from a suspect vehicle to a crime scene, with neither analysis carrying the conclusion alone [19]; Young and colleagues combined high-throughput eukaryote sequencing with mid-infrared spectroscopy in a simulated crime scene scenario [20]. Whether microbial profiling offers resolution superior to established methods has not been established in direct comparison, and we do not claim it here. Its likely contribution is to cases in which established methods leave genuine ambiguity, for example between sites that share geology and land cover.

3. Representative studies and demonstrations of casework utility

A small but growing body of empirical work has begun to translate these principles into casework-relevant scenarios.

Foran and colleagues at Michigan State University, supported by the United States National Institute of Justice, have conducted one of the most sustained programmes of applied research in this area. Using 16S amplicon sequencing of soil samples collected from shoes, shovels, and clothing under simulated casework conditions, they demonstrated that supervised classifiers could associate recovered soils with their source site at statistically meaningful levels, and argued that the resulting objective error rates were sufficient to meet the Daubert criteria for admissibility [17,21]. These were closed-set comparisons among a panel of habitats sampled by the investigators, and the classification target was habitat of origin rather than a specific location.

Studies conducted in China on provincial and city-scale reference sets have reported classification accuracies of approximately 66.7 % at the district level and 90 % at the city level in blind tests [7]. These were closed-set assignments to panels of candidate districts and cities, and the fall in accuracy at the finer granularity is the general pattern in this literature. Comparable results have been obtained in European studies using metabarcoding approaches to link soil recovered from suspect items to candidate crime scenes. Giampaoli and colleagues, in particular, demonstrated that DNA metabarcoding of soil samples retained discriminating signal even after storage conditions approximating realistic exhibit handling [22]. Genuinely blind, externally administered evaluations remain rare; the blind test reported by Demanèche and colleagues is one of the few published examples of that design [23].

Of particular relevance to the evidentiary question, several studies have examined directly how well discriminating signatures persist under the conditions to which exhibits are routinely subjected. Time lapses of several weeks, exposure to fluctuating temperature, rainfall, and even desiccation of soil in a closed vehicle boot have not entirely destroyed the geolocation signal, although all of these conditions degrade accuracy to some extent [9,19]. Badgley and colleagues found that soils ageing on evidential items undergo consistent taxonomic change, drifting in multivariate space but not toward a different habitat, with 95 % of 364 evidentiary profiles still classified to the correct habitat [24]. These findings suggest that the method retains useful signal under realistic conditions, over the range of habitats, intervals and storage conditions so far tested, while also underscoring the need for case-specific validation and clearly stated confidence intervals.

It should be noted, however, that to the authors' knowledge no appellate court has yet ruled directly on whether soil microbial

geolocation evidence is admissible under Daubert or a similar standard. The scientific literature increasingly shows what the technique can do in casework, but it has not yet been tested in a contested trial.

4. Evidentiary strength and admissibility

4.1. The Daubert framework

In the United States, the admissibility of expert scientific testimony is governed by the framework articulated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), which assigns to the trial judge the role of gatekeeper charged with assessing whether proffered testimony is both reliable and relevant [25]. The Daubert Court identified several non-exclusive factors for evaluating reliability: whether the method has been tested, whether it has been subjected to peer review and publication, its known or potential error rate, the existence and maintenance of standards controlling its operation, and the degree of its acceptance within the relevant scientific community. Comparable, though not identical, frameworks apply in other common-law jurisdictions, including the criteria articulated in *R v. Bonython* (1984) in Australia and the approach taken by the courts of England and Wales following *R v. Dallagher* [2003] [26].

A soil microbial geolocation method that relies on published, peer-reviewed sequencing protocols and on widely-used machine-learning classifiers can satisfy several Daubert factors without major difficulty. The techniques have been tested, published, and (at the level of the component steps) widely accepted within microbial ecology. The points at which the framework becomes more demanding are those of standardisation and error rate.

4.2. Error rate and interpretation

Daubert requires that a method's known or potential rate of error be articulated. In microbial geolocation, the error rate is not a single number but a function of the reference database's granularity and geographic coverage, the classifier used, the environmental conditions to which the questioned sample has been subjected, and the verbal form of the conclusion offered. A report stating that a questioned soil sample is more consistent with one of five candidate sites than with the others is making a quite different claim, with a quite different error profile, than a report stating that the sample originated from a specific garden in a specific town. At present, most published studies report cross-validation accuracies on closed-set classification problems; these figures are informative but are not directly convertible into likelihood ratios of the kind that would properly support a source-level conclusion in court [10, 27]. Cross-validation accuracy measures how well a classifier reproduces labels within the dataset on which it was developed; a forensic error rate would describe how often the method, as applied in casework, reaches an incorrect conclusion on the proposition actually offered in evidence. No validated error rate of the second kind presently exists for soil microbial source attribution, and published classification accuracies should not be offered in its place.

The broader forensic science community has, over the past two decades, moved towards a likelihood-ratio framework for the evaluation of scientific evidence, in which the expert expresses the probability of the observed evidence under competing propositions (typically the prosecution's and the defence's). Applying this framework to microbial geolocation will require, at minimum, population-representative reference databases and agreed methods for estimating the denominator of the likelihood ratio, that is, the probability of observing the recovered microbial profile if the questioned soil did not come from the proposed source location. This is an area of active methodological development but is not yet, in this author's assessment, mature enough to support routine casework reporting at the source level.

4.3. Standardisation, validation, and accreditation

Even where the underlying science is sound, the admissibility of expert evidence depends on the integrity of the procedures by which the conclusion was produced. Forensic DNA typing achieved widespread admissibility not only because the underlying biology was robust but because a comprehensive infrastructure of validated kits, accredited laboratories, proficiency testing, and standardised reporting was developed in parallel [28]. Soil microbial geolocation currently lacks most of this infrastructure. There is no accredited reference database of comparable authority to the population genetic databases that underpin DNA typing; extraction and sequencing protocols, while converging, are not formally standardised across laboratories; and inter-laboratory proficiency schemes specific to forensic microbial geolocation are, to our knowledge, not yet in place.

These deficiencies do not render the evidence inadmissible in principle, but they do expose any expert report offered today to legitimate challenge on precisely the procedural grounds that Daubert and its analogues were designed to test. Addressing them will require sustained organisational investment. It would be wrong, however, to describe the remaining obstacles as purely infrastructural, since accreditation of a procedure does not validate an inference that the underlying data cannot support. Several of the open questions set out in Section 5 are scientific in character.

5. Limitations and open questions

Several limitations warrant explicit acknowledgement. First, the temporal stability of soil microbial communities imposes practical constraints. While studies have shown that discriminating signatures persist over weeks and under adverse storage conditions, seasonal turnover of community membership can, over longer periods, erode the match between a questioned sample and historical reference data [9]. Reference databases may therefore need to be updated at intervals, and questioned samples may need to be sequenced promptly after recovery. The interval between the offence, the recovery of the questioned sample and the collection of the known sample is accordingly an analytical variable rather than a procedural detail, and it should be reported as such.

Second, sampling design and environmental heterogeneity bear directly on what can be claimed. The number of samples taken per site, their spacing and the depth sampled all affect both the apparent within-site variance and the apparent separation between sites, and increasing within-site replication will improve reported accuracy without any change in the underlying biology. A crime scene is rarely a homogeneous plot: soil deposited on an item may sample a few square centimetres of a surface that varies at that scale, or may represent a mixture accumulated along a route. The comparison that matters in casework is amid within-site variance at the location in issue and between-site variance across the population of alternative sources, and this has rarely been reported explicitly [14].

Third, pinpointing a source at fine spatial scales is only possible where enough reference samples have been collected. If a candidate site is not represented in the reference set, that is not the same as having ruled it out, and a report must say which of the two it means: that the site was excluded on the basis of reference data, or that no reference data for it existed [27].

The size of the reference set also shapes reported performance in ways that are easy to miss. Two effects matter. First, the number of candidate sites sets the accuracy you would reach by guessing alone, so figures from studies using different numbers of sites cannot be compared unless that baseline is given. Second, taking only a few samples per site makes the sites look more distinct than they really are, because the overlap between them is never fully observed.

Fourth, the interpretability of machine-learning outputs, particularly those of deep neural networks, poses a real difficulty in the courtroom

setting. An expert asked to explain the basis of a conclusion to a jury must be able to articulate, in comprehensible terms, why the model reached the conclusion it did. Random forests and similar tree-based methods offer feature-importance outputs that can be used to identify the taxa driving a classification, and this transparency is a genuine advantage in evidentiary terms [17].

Fifth, contamination and chain-of-custody concerns, well-understood in forensic DNA typing, take on particular salience in microbial work because environmental microorganisms are ubiquitous and can be introduced at any stage of sample handling. Negative controls at collection, extraction, and sequencing stages are essential but are not yet universally reported in the published literature [10]. Reagent and consumable contamination is well documented in low-biomass work, and the contaminant profile differs between extraction kit lots and between laboratories [29]; soil is generally high-biomass, but trace recoveries from evidential items may not be. Batch effects arising from extraction date, plate position, library preparation and sequencing run deserve equal attention. Because a known sample is ordinarily collected after the questioned sample, batch and source are readily confounded, and a classifier may then discriminate processing history rather than provenance. Processing questioned and known samples within the same batch wherever the case permits is a straightforward mitigation that is not yet standard practice.

Finally, privacy and Fourth Amendment considerations, while peripheral to geolocation of soil from inanimate objects, become relevant when soil microbial evidence is combined with samples taken from a person's body or possessions, and these issues are beginning to attract scholarly attention [30].

6. Recommendations for forensic casework reporting

On the basis of the above, we propose a minimum dataset that any soil microbial geolocation report offered in support of expert testimony should contain. The report should specify: (i) the sampling and preservation history of the questioned and reference samples, with full chain-of-custody documentation; (ii) the DNA extraction kit and protocol, including version, and the results of extraction and no-template controls; (iii) the amplification primers, the sequencing platform, and read counts before and after quality filtering, together with whether the questioned and known samples were extracted and sequenced in the same batch; (iv) the denoising and taxonomic classification pipeline, with software versions and reference database versions; (v) the classifier used, including its training data, the validation scheme applied and whether it was internal cross-validation or blind external testing, and, where appropriate, feature-importance outputs; (vi) the set of candidate sites considered and the rationale for their selection, together with the density and geographic extent of the reference resource relied upon; (vii) an explicit statement of which of the four inferential tasks described in Section 2 the conclusion addresses; and (viii) a clearly-worded statement of the conclusion, with explicit acknowledgement of the method's limitations and of candidate locations for which no reference data were available.

Reports that fall short of this minimum dataset should, in our view, not be offered in support of source-level conclusions in court. Exclusionary or investigative-lead applications may be supportable on weaker foundations, but the evidentiary weight claimed must be proportionate to the methodological rigour demonstrated.

7. Conclusion

Soil microbial geolocation has progressed, in less than two decades, from a laboratory curiosity to a method demonstrably capable of discriminating soils that established physical and chemical methods cannot readily separate, and of assigning samples to sites within a closed reference panel at high rates. The underlying science is sound, the analytical pipelines are increasingly standardised, and the supervised-

learning classifiers employed are well-understood in other application domains. Substantial infrastructural work remains, in that accredited reference databases, validated and standardised pipelines, proficiency-tested laboratories and agreed frameworks for likelihood-ratio reporting are all absent. These are not, however, the only obstacles. Questions bearing directly on evidential weight remain open, among them the relationship between within-site and between-site variation at the scale of an actual scene, the confounding of source with processing batch, and the absence of error rates estimated under blind conditions.

The Daubert framework and its analogues are well-suited to evaluating this kind of evidence, and, far from being obstacles, they provide a constructive checklist against which the forensic microbial geolocation community can measure its readiness for courtroom deployment. On the present evidence, the field has demonstrated promising discriminatory capability under defined conditions, and microbial profiling is most defensible when used alongside established soil examination rather than in place of it; forensic source attribution on microbial evidence remains insufficiently validated to support source-level conclusions in court. Targeted investment in the missing infrastructure, combined with the empirical work identified in Section 5 and transparent reporting, offers a realistic route to changing that.

CRedit authorship contribution statement

Abdulla Albastaki: Writing – original draft, Investigation, Conceptualization. **Judith Smith:** Writing – review & editing, Supervision.

Declaration of generative AI use

During the preparation of this work, the authors used a large language model (Claude, Anthropic) to help identify relevant literature, to check grammar and language, and to flag unclear sentences for the authors to rewrite. All cited sources were independently located, read and verified by the authors, who reviewed all text and takes full responsibility for the content of the publication.

Funding

This work received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] K. Pye, Geological and Soil Evidence, CRC Press, 2007, <https://doi.org/10.1201/9781420004755>.
- [2] R.C. Murray, Evidence from the Earth : Forensic Geology and Criminal Investigation, Mountain Press Pub. Co., 2011, p. 200.
- [3] K. Pye, S.J. Blott, Development of a searchable major and trace element database for use in forensic soil comparisons, *Sci. Justice* 49 (3) (2009 Sep) 170–181, <https://doi.org/10.1016/j.scijus.2009.02.007>.
- [4] R.W. Fitzpatrick, M.D. Raven, How pedology and mineralogy helped solve a double murder case: using forensics to inspire future generations of soil scientists, *Soil Horiz.* 53 (5) (2012 Sep 18) 14–29, <https://doi.org/10.2136/sh12-05-0016>.
- [5] N. Fierer, R.B. Jackson, The diversity and biogeography of soil bacterial communities, *Proc. Natl. Acad. Sci.* 103 (3) (2006) 626–631, <https://doi.org/10.1073/pnas.0507535103>.
- [6] M. Delgado-Baquerizo, A.M. Oliverio, T.E. Brewer, A. Benavent-Gonzalez, D. J. Eldridge, R.D. Bardgett, et al., A global atlas of the dominant bacteria found in soil, *Science* (1979) 359 (6373) (2018) 320–325, <https://doi.org/10.1126/science.aap9516>.
- [7] Z. Wang, F. Zhang, L. Wang, H. Yuan, D. Guan, R. Zhao, Application of microbiome in forensics, *Genom. Proteom. Bioinform.* 21 (1) (2023) 97–107, <https://doi.org/10.1016/j.gpb.2022.07.007>.
- [8] N.S. Grantham, B.J. Reich, E.B. Laber, K. Pacifici, R.R. Dunn, N. Fierer, et al., Global forensic geolocation with deep neural networks, *J. R. Stat. Soc. Ser. C Appl. Stat.* 69 (4) (2020 Aug 1) 909–929, <https://doi.org/10.1111/rssc.12427>.
- [9] T.H. Clarke, A. Gomez, H. Singh, K.E. Nelson, L.M. Brinkac, Integrating the microbiome as a resource in the forensics toolkit, *Forensic Sci. Int. Genet.* 30 (2017) 141–147, <https://doi.org/10.1016/j.fsigen.2017.06.008>.
- [10] J.M. Robinson, Z. Pasternak, C.E. Mason, E. Elhaik, Forensic applications of microbiomics: a review, *Front. Microbiol.* (2021 Jan 13) 11, <https://doi.org/10.3389/fmicb.2020.608101>.
- [11] B. Moitas, I.M. Caldas, B. Sampaio-Maia, Forensic microbiology and geographical location: a systematic review, *Aust. J. Forensic Sci.* 56 (4) (2024 Jul 3) 416–431, <https://doi.org/10.1080/00450618.2023.2191993>.
- [12] L.R. Thompson, J.G. Sanders, D. McDonald, A. Amir, J. Ladau, K.J. Locey, et al., A communal catalogue reveals Earth's multiscale microbial diversity, *Nature* 551 (7681) (2017) 457–463, <https://doi.org/10.1038/nature24621>.
- [13] H. Habtom, S. Demaneche, L. Dawson, C. Azulay, O. Matan, P. Robe, et al., Soil characterisation by bacterial community analysis for forensic applications: a quantitative comparison of environmental technologies, *Forensic Sci. Int. Genet.* 26 (2017 Jan) 21–29, <https://doi.org/10.1016/j.fsigen.2016.10.005>.
- [14] E. Jurkevitch, Z. Pasternak, A walk on the dirt: soil microbial forensics from ecological theory to the crime lab, *FEMS Microbiol. Rev.* 45 (2) (2021 Mar 16), <https://doi.org/10.1093/femsre/fuaa053>.
- [15] J.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, et al., Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, *ISME J.* 6 (8) (2012) 1621–1624, <https://doi.org/10.1038/ismej.2012.8>.
- [16] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, S.P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data, *Nat. Methods* 13 (7) (2016 Jul 23) 581–583, <https://doi.org/10.1038/nmeth.3869>.
- [17] E.M. Jesmok, J.M. Hopkins, D.R. Foran, Next-Generation sequencing of the bacterial 16S *scp>rRNA</scp>* gene for forensic soil comparison: a feasibility study, *J. Forensic Sci.* 61 (3) (2016 May 18) 607–617, <https://doi.org/10.1111/1556-4029.13049>.
- [18] G. Xu, X. Teng, X.H. Gao, L. Zhang, H. Yan, R.Q. Qi, Advances in machine learning-based bacteria analysis for forensic identification: identity, ethnicity, and site of occurrence, *Front. Microbiol.* 21 (2023 Dec) 14, <https://doi.org/10.3389/fmicb.2023.1332857>.
- [19] G. Concheri, D. Bertoldi, E. Polone, S. Otto, R. Larcher, A. Squartini, Chemical elemental distribution and soil DNA fingerprints provide the critical evidence in murder case investigation, *PLoS One* 6 (6) (2011 Jun 3) e20222, <https://doi.org/10.1371/journal.pone.0020222>.
- [20] J.M. Young, L.S. Weyrich, J. Breen, L.M. Macdonald, A. Cooper, Predicting the origin of soil evidence: high throughput eukaryote sequencing and MIR spectroscopy applied to a crime scene scenario, *Forensic Sci. Int.* 251 (2015 Jun) 22–31, <https://doi.org/10.1016/j.forsciint.2015.03.008>.
- [21] J. Dawson, Solving Crimes with Soil Bacteria, National Institute of Justice, Office of Justice Programs [Internet], 2017 [cited 2026 Oct 5], <https://nij.ojp.gov/topics/articles/solving-crimes-soil-bacteria>.
- [22] S. Giampaoli, A. Berti, R.M. Di Maggio, E. Pilli, A. Valentini, F. Valeriani, et al., The environmental biological signature: NGS profiling for forensic comparison of soils, *Forensic Sci. Int.* 240 (2014 Jul) 41–47, <https://doi.org/10.1016/j.forsciint.2014.02.028>.
- [23] S. Demaneche, L. Schausser, L. Dawson, L. Franqueville, P. Simonet, Microbial soil community analyses for forensic science: application to a blind test, *Forensic Sci. Int.* 270 (2017) 153–158, <https://doi.org/10.1016/j.forsciint.2016.12.004>.
- [24] A.J. Badgley, E.M. Jesmok, D.R. Foran, Time radically alters *Ex Situ* evidentiary soil 16S bacterial profiles produced via next-generation sequencing, *J. Forensic Sci.* 63 (5) (2018 Sep 21) 1356–1365, <https://doi.org/10.1111/1556-4029.13753>.
- [25] v Daubert, Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, Supreme Court of the United States, 1993, p. 579.
- [26] G. Edmond, A. Roberts, Procedural fairness, the criminal trial and forensic science and medicine, *Sydney Law Review* [Internet] 33 (3) (2011) 359–394 [cited 2026 Oct 5], <https://classic.austlii.edu.au/au/journals/SydLawRw/2011/17.html>.
- [27] S.E. Schmedes, A. Sajantila, B. Budowle, Expansion of microbial forensics, *J. Clin. Microbiol.* 54 (8) (2016 Aug) 1964–1974, <https://doi.org/10.1128/JCM.00046-16>.
- [28] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, first ed., Academic Press (Elsevier), San Diego, CA, 2011 <https://doi.org/10.1016/C2011-0-04189-3>.
- [29] S.J. Salter, M.J. Cox, E.M. Turek, S.T. Calus, W.O. Cookson, M.F. Moffatt, et al., Reagent and laboratory contamination can critically impact sequence-based microbiome analyses, *BMC Biol.* 12 (1) (2014 Dec 12) 87, <https://doi.org/10.1186/s12915-014-0087-z>.
- [30] E. Murphy, Forensic DNA typing, *Annu Rev Criminol* 1 (1) (2018 Jan 13) 497–515, <https://doi.org/10.1146/annurev-criminol-032317-092127>.