Estimation of Geographic Origin from Dust Using Plant DNA Metabarcoding

Chelsea Lennartz
MIT Lincoln Laboratory
Joel Kurucar
MIT Lincoln Laboratory
Stephen Coppola
MIT Lincoln Laboratory
Janice Crager
MIT Lincoln Laboratory
Johanna Bobrow
MIT Lincoln Laboratory
Laura Bortolin
Mercy Bioanalytics
James Comolli (james.comolli@ll.mit.edu)
MIT Lincoln Laboratory

Research Article

Keywords: Metabarcoding, geographic attribution, geolocation, provenance, forensic, environmental DNA, eDNA, plant

Posted Date: May 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-523688/v1

Version of Record: A version of this preprint was published at Scientific Reports on August 10th, 2021.
See the published version at https://doi.org/10.1038/s41598-021-95702-3.

Estimation of Geographic Origin from Dust Using Plant DNA Metabarcoding

Chelsea Peragallo[1], Joel Kurucar[1], Stephen Coppola[1], Janice Crager[1], Johanna Bobrow[1], Laura Bortolin[2] and James Comolli[1*]

[1] MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02421

[2] Current address: Mercy Bioanalytics, 700 Main Street, Cambridge, MA 02139

* Corresponding author: james.comolli@ll.mit.edu; 781-981-1956

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.
This material is based upon work supported under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

ABSTRACT

Information obtained from the analysis of dust, particularly biological particles such as pollen, plant parts, and fungal spores, has great utility in forensic geolocation. As an alternative to manual microscopic analysis, we developed a pipeline that utilizes the environmental DNA (eDNA) from plants in dust samples to estimate previous sample location(s). The species of plant-derived eDNA within dust samples were identified using metabarcoding and their geographic distributions were then derived from occurrence records in the USGS Biodiversity in Service of Our Nation (BISON) database. The distributions for all plant species identified in a sample were used to generate a probabilistic estimate of the sample source. With settled dust collected at four U.S. sites over a 15-month period, we demonstrated positive regional geolocation (within 600 km^2 of the collection point) with 47.6% (20 of 42) of the samples analyzed. Attribution accuracy and resolution was dependent on the number of plant species identified in a dust sample, which was greatly affected by the season of collection. In dust samples that yielded a minimum of 20 identified plant species, positive regional attribution improved to 66.7% (16 of 24 samples). Using dust samples collected from 31 different U.S. sites, trace plant eDNA provided relevant regional attribution information on provenance in 32.2%. This demonstrated that analysis of plant eDNA in dust can provide an accurate estimate regional provenance within the U.S., and relevant forensic information, for a substantial fraction of samples analyzed.

KEYWORDS

Metabarcoding, geographic attribution, geolocation, provenance, forensic, environmental DNA, eDNA, plant

INTRODUCTION

The ability to match a person or object with a particular location (geographic attribution) can be a crucial part of a forensic investigation. Geographic attribution often relies on the analysis of dust because of its ubiquity, abundance, adherence to most surfaces, and wide variety of component particles that includes bacterial and fungal cells, spores, minerals, soil, plant and animal components, and products of combustion and other human activities (1). Each of these particle types can be indicative of the local exposure history of the sampled object or person, and their analysis and interpretation has been used for almost a century as a tool in forensic and investigative geographic attribution (2-4).

Biological components of dust and pollen, in particular, have proven to be a valuable investigative tool. Pollen is ubiquitous, stable, taxonomically diverse in structure, and able to discriminate geographically due to plant biogeography (5-7). Since its release varies seasonally, pollen can also provide information on the timing of dust accumulation (8). Historically, the characterization of pollen has relied heavily on analysis of the exine structure using microscopy (9). Proper pollen preparation for microscopy and analysis is time consuming, which limits sample testing capacity (10). In addition, accurate taxonomic identification and geographic inference from pollen images requires significant technical expertise and, even with that, is difficult for many pollen types. This has limited widespread application of the technique for forensic investigation as well as other applications (11). A number of approaches have been taken in an attempt to address this, including the Pollen Identification and Geolocation Technology (PIGLT) system. PIGLT is a standardized digital database of pollen images and software to augment taxonomic geographic distribution information to reduce the amount of expertise required (12, 13) and an automated system for pollen microscopic imaging and characterization (14).

DNA barcoding has been used for taxonomic identification for over two decades in environmental tracking, biodiversity studies, and product authentication (15-17), and has more recently been applied to pollen

classification (18). Barcoding targets a genomic DNA region that is common among taxa so it can be readily amplified, but whose sequence differs sufficiently to enable discrimination. Taxonomic identification is typically performed by matching a generated barcode sequence to a database of barcode sequences from identified taxa. Barcodes can be specific to a particular family or genus, but "universal" DNA barcodes have been implemented to help identify taxa in broad phylogenetic categories such as bacteria/archaea, fungi, and animals, including those from mixed samples in a process called metabarcoding. Compared to DNA metabarcoding of other taxa, plant DNA metabarcoding is challenged by the lack of truly universal barcode that can effectively discriminate most plant species (19-22). For this reason, several different plant barcodes are utilized, depending on the application. This includes the chloroplast gene encoding the large subunit of ribulose 1,5-bisphosphate carboxylase gene (*rbcL*), the maturase K gene (*matK*), or the intron region of the chloroplast tRNA gene (*trnL*), as well as the nuclear genome encoded internally transcribed spacer region (ITS2) and others (21).

Though initially used to characterize separated pollens (23), the combination of DNA metabarcoding with next generation sequencing has enabled characterization of pollen mixtures for allergen analysis in airborne samples (24, 25) characterization of honey (26-29), and has been further developed for forensic geolocation (18) and even pollen quantitation (30). Though not without limitations, DNA metabarcoding has the potential to rapidly advance many pollen application areas (31). In this study, we demonstrated that DNA metabarcoding of pollen and plant eDNA in dust samples can be used for rapid, high throughput estimation of sample origin. We intentionally focused on available datasets, National Center for Biotechnology Information (NCBI) Genbank for species identification of plant DNA barcodes and the Biodiversity in Service of Our Nation (BISON) data repository for estimation of the geographic species distribution, to demonstrate the utility of these resources. Our hypothesis was that plant remnants in dust were sufficient to provide valuable

information on sample provenance rapidly using DNA metabarcoding and automated analysis using currently-available taxonomic and species distribution datasets.

RESULTS

Characterization of Plant DNA from Dust

Plants were the focus because of the extensive, available barcode sequence and species distribution data. Though a greater diversity of bacterial and fungal DNA barcodes could be generated with most dust samples, the lack of matching sequence and species distribution data limited their utility using our methods. Our pipeline for estimating geographic origin (summarized in Figure 1) was tested using total eDNA isolated from settled environmental dust collected from four U.S. locations.

Using a relatively high stringency for OTUs (> 2 reads, present in 3 of the 3 sequencing replicates, relative abundance > 10^{-4}), there was an average of 42.2 ITS2 and 40.4 *rbcL*-3A OTUs per MA dust sample, with a large variance in the OTU number that depended on the date (time of year) of dust accumulation and collection (Figure 2). There was a significant difference ($p = 1.6\text{x}10^{-5}$) in the number of OTU found in each season. Over 200 total ITS2 and *rbcL*-3A OTUs present in triplicate samples were generated in mid to late spring (April and May 2016). The number of OTUs decreased in samples over the progression from summer to fall and reached a minimum in winter, when samples collected between November 2016 to February 2017 yielded 20 or fewer total OTUs, including two samples with zero OTU. Overall, 2 to 6 times fewer OTUs were recovered from winter samples. In 2017, the OTU number per sample started to increase with the onset of spring. As expected, there was significant seasonal variation in the presence of most individual OTUs, with the date of maximum abundance dependent on taxon. This is illustrated in Figure 3, which shows the abundance of detected OTUs with the highest read counts according to time of year. There appeared to be three main groups

of OTU: those most prevalent in early spring, those most prevalent in late spring/early summer, and those most prevalent in late summer.

Surprisingly, the length of time that slides were exposed to the environment had only a minor impact on the number of OTUs per sample, with no significant difference ($p = .51$) in the number of OTUs in dust collected after coincident 14-, 28-, or 56-day exposure (Figure 2). This indicated that substantial plant material deposited on a slide in 14 days or less. As expected from this data, the most prevalent OTUs at any one collection time were found in coincident 14, 28, and 56 days samples with few exceptions (Supplementary Figure 1), and differences between these samples were largely represented by lower abundance OTU.

The average number of plant OTUs recovered from dust samples from MA was 82.6 ± 73.5, while that from FL, NM, or SC averaged 133.2 ± 105.4, 135.5 ± 72.7, and 132.0 ± 138.1, respectively. Though MA dust samples yielded fewer OTU than those from other collection sites, this difference was not statistically significant ($p = .68$). The number of OTUs recovered in samples from NM and SC displayed similar seasonal variation to those from MA, with more recovered in early spring to early summer than in December - February (Supplementary Figure 2). However, OTUs in samples from FL showed a seasonal variation with a different pattern relative to the other sites, with the most OTU/sample recovered in December through February.

Estimation of Geographic Origin

The taxonomy of the plant OTUs derived from DNA extracted from dust were assigned by matching to plant barcode records in NCBI Genbank using a 100% identity threshold. Only 38.3 ± 19.2% of the total OTUs per sample were defined at the species level, the remainder were discarded. 43.6% of the ITS2 OTUs matched a single species and 69.2% matched 5 or fewer species, while 24.7% and 60.3% of the *rbcL*-3A OTUs matched one and less than five species, respectively (Supplementary Figure 3). This showed that ITS2 performed better that *rbcL*-3A in assigning single plant species using the defined parameters.

A majority of the point-to-grid maps, which showed the percentage of OTU from that sample that had at least two occurrence records within each 250 km^2 grid, indicated the region of sample origin within the U.S., meaning that grids with the highest percentage of plant OTUs from that sample co-located within the region of the collection site. Examples of point-to-grid maps generated from each collection site are shown in Figure 4. Gaussian mixture modeling (GMM) was applied to the point-to-grid map generated from 14-day samples from the four sites to enable quantitation of accuracy (TP) and resolution (AT5PE) (Figure 4B). TP indicated the percentage of grids that contain fewer OTU than the grid with the actual sample location and was used with a cutoff of 90%. At that threshold, the grid containing the truth point contained more mapped OTUs than 90% of all grids. AT5PE indicated the mean distance (in km) of the truth site from the five highest probability peaks as determined by our analysis. We utilized an AT5PE of 600 km as a threshold, and considered a positive for regional attribution to hit both thresholds, >90% TP and <600 km AT5PE.

When 14-day dust samples across all four sites were analyzed, the TP was greater than 90% in 20 of 42 (47.6%), the AT5PE was < 600 km in 24 of 42 (57.1%) and 20 of 42 (47.6%) were deemed to produce positive regional geolocation (Table 1).

The geolocation accuracy varied significantly by collection site, with MA (53.5%), NM (60.0%), and SC (50.0%) showing a similar percentage of positives, while FL samples yielded none. The number of mapped OTUs found in a dust sample appeared to significantly impact the TP and AT5PE (Figure 5). With both metrics, samples that yielded fewer than 20 mapped OTUs showed greatly reduced attribution accuracy (TP and AT5PE) compared to those with 20 or more OTUs (Table 1). 18 of 42 samples contained fewer than 20 OTUs, 13 of which were collected October – February. When considering only those samples with more than 20 OTUs, 16 of the 24 (66.7%) yielded positive attribution, with a similar variance by the site of sample collection. The duration of dust accumulation did not have a significant impact on attribution accuracy (Supplementary

Table 1) and there was a similar threshold of 20 OTUs with samples where dust was collected for more than 14 days.

Though dependent of the location of dust collection, these data demonstrated that plant DNA from dust combined with available species distribution information could accurately estimate the region of origin from a significant proportion of samples. With sufficient OTUs, two thirds of the dust samples yielded positive regional attribution. Samples that did not produce correct regional attribution generally had a dispersed OTU distribution, i.e. poor resolution, and did not define an incorrect possible region or origin. These would have been indicated by < 90% TP and 600 km or less AT5PE.

Dust Samples from Other U.S. Locations

To characterize the achievable geographic attribution accuracy and resolution with dust samples collected from a broader set of locations, metabarcoding was performed on 31 environmental dust samples from different U.S. locations that were collected as part of the Wild Life Our Homes (WLOH) citizen science project (32). When these samples were analyzed using our pipeline, 10 of the 31 dust samples (32%) generated minibarcodes that resulted in positive attribution (TP > 90%, AT5PE < 600 km) (Figure 6A). The number of OTUs appeared to have less of an impact on the attribution accuracy with this sample set, and the percentage of positive attribution improved only slightly when samples with more than 20 mapped OTU were considered. Mapping the 31 samples to assess the impact of the region of sample origin on the attribution accuracy and resolution showed that the highest attribution accuracy and precision were achieved with samples from Montana, Texas, and the Middle Atlantic Region. Samples from the west coast of the U.S. and Midwest produced reasonable accuracy (75% or higher truth percentage) but generally low resolution (Figure 6B). This suggested that OTUs derived from samples from these locations may be less informative, though many more samples would have to be processed to generate significance. It is worth noting that the WLOH sampling

method differed from that used for the louvered shelters in that dust accumulation in the WLOH samples was not standardized so that accumulation occurred over a variable duration longer than 14 days. In addition, exposure to environmental factors, which can affect DNA stability, was less controlled.

DISCUSSION

We have utilized currently available plant sequence and species distribution data to demonstrate a streamlined system for exploiting plant eDNA in dust for forensic attribution. Plant barcodes generated by standard metabarcoding methods were fed into a data processing pipeline that demonstrated trace plant DNA in dust can provide an accurate estimate of regional geographic attribution (within 600 km or less) from nearly half of samples collected.

One unique aspect of our pipeline was the use of publicly available biogeographic data from BISON for determination of the geographic distribution of the species found in the samples. Our objective was to determine if currently-available reference data like that in BISON could be applied to attribution determination to avoid the cost and time needed for sample collection and analysis to create a new reference dataset. The BISON dataset has over 400 million species observation records from across the U.S., and thus can provide widely applicable georeferenced information. While this study focused on attribution of samples gathered within the U.S. using BISON, the technique could be extended to other global regions with a high density of plant species occurrence information through the use of the Global Biodiversity Information Facility (GBIF) (33) or other datasets. For our pipeline, the total area covered by observation records of a plant species was used as an indicator of its geographic distribution, then individual species distributions were normalized and merged into OTU distributions, which were then overlaid using a geographic information system (GIS). This generated a map that indicated the geographic areas with the largest OTU distribution overlap to provide an estimate of the

sample origin. Using this attribution system, plant minibarcodes provided more provenance information than animal minibarcodes, which generated fewer OTU of less diversity, or than fungal or bacterial barcodes and generated numerous OTUs that were less able to match the NCBI Genbank reference database or have available species distribution information in BISON.

Fungal DNA analysis has been demonstrated to be able to delineate soil samples (34, 35) as well as be informative in unguided geographic attribution. Fungal ITS1 minibarcode OTUs from over 900 WLOH dust samples from different U.S. locations, a subset of which were used in this study, enabled the estimation of geographic provenance with a median prediction margin of 230 km (36). A similar approach was used to determine the worldwide country of origin from dust samples (37, 38). These approaches avoid many of the pitfalls of taxonomic identification and species distribution estimation but, unlike our approach, require generation of a new reference dataset. The analysis of bacterial barcode sequences has also been used for forensic attribution, particularly to link soil or other samples to a source location (39, 40). Their use for unguided geographic attribution is challenged by the tremendous local and worldwide diversity (41, 42) as well as a lack of reference data.

Using our attribution system, roughly 40% of the 80 dust samples collected at four different U.S. locations and 32% of the 31 dust samples collected from different U.S. locations provided accurate regional attribution estimates. This percentage increased if samples with less than a threshold of 20 detected OTUs were excluded. 18 of the 29 samples that had fewer than 20 OTUs were collected between the months of November and February, which showed that dust collected in the winter was less able to support accurate attribution. This confirmed our expectation that the amount of available airborne plant eDNA from plant-derived particles such as pollen, seeds, and spores released by plants is reduced in winter. Snow cover in some sites may also have prevented aerosol dispersion of ground particles. It should be noted that, for all samples, there was no evidence of erroneous attribution, where the site of origin was estimated to be in an incorrect regional location. In

samples that did not yield positive attribution, the estimated area of attribution was broad and undefined, meaning there were no samples with a low TP and AT5PE <600 km^2.

The data also indicated that 14 days was a sufficient time for dust collection to capture the most prevalent OTU, which implied that an object needed to reside at a location for only 14 days or less (depending on the season) to accumulate an identifiable signature. Further investigation into samples with outdoor environmental exposure shorter than 14 days is needed to determine the minimum amount of time required for attributable signature accumulation, but preliminary studies have indicated sufficient OTUs can be generated with dust that accumulates in as little as 3 days.

The geographic attribution accuracy and resolution was significantly impacted by the number of mapped OTUs (Figures 4 and 5). This is partly due to system design, where OTUs are mapped but the subset with narrow geographic distributions are most informative. The OTU number is mainly affected by the amount and variety of plant eDNA deposited on the slide surface, which is dependent on the local plant abundance and diversity, air flow in the shelter, the length of time of dust accumulation, the season of dust accumulation, and exposure to environmental factors such as light and precipitation that could impact DNA stability. Our method ensured collection of dust that was fresh and relatively protected from sunlight and precipitation. Attribution capability would likely be degraded with samples exposed to the environment, where the degradation rate of eDNA could vary by several orders of magnitude depending on the matrix and environmental conditions (43). Dust from more exposed environmental surfaces, such as the door tops in the WLOH samples or from exposed surfaces in the same locations at the louvered shelters, confirmed this. Our slide preparation and transport methods also ensured that only dust originating from the collection site accumulated on slides. Objects of forensic interest are more likely to have traveled to, or been used in, more than one location where dust can accumulate. Plant OTUs from more than one location could be indicative of multiple source locations, though the ability to discriminate more than one site would be complicated by the duration of residence in each

location, the exposure to environmental elements, season of exposure, the distance between locations, and the time since the object was in the previous location(s). The effects of these parameters on attribution accuracy need to be better characterized to determine the applicability of the current pipeline to dust from objects that have resided in more than one location.

The number of mapped OTUs was also significantly impacted by inefficiencies in matching to the NCBI Genbank and BISON databases. Recent estimates are that barcode sequences from 25 to 40% of the roughly 390,000 plant species in the world (44) are represented in NCBI Genbank, with the estimated coverage of the 51,000 U.S. plant species higher (45). However, these estimates include entries representing all plant barcodes, meaning that for any one barcode there are significant gaps in taxonomic coverage. In addition, the short sequence length of minibarcodes, necessary for compatibility with next generation sequencing, limited their ability to discriminate among plant species records in the NCBI Genbank database. One OTU matched a single species record 50 to 60% of the time, with the majority of the remainder matching 2 to 10 species. In fact, we used a similarity threshold of 100% because a lower threshold increased the number of OTUs assigned to the same species while not substantially increasing the number of new species identified. Recently developed sequence alignment algorithms that are alternatives to nBLAST may enable improved plant taxonomic assignment using minibarcodes (46), as may use of longer barcode sequences generated through the use of improved sequence chemistry, amplification and sequencing of long barcode amplicons using nanopore-type sequencing, chloroplast genome sequencing, or genome skimming (31, 47).

The ITS2 and *rbcL*-3A plant minibarcode primer pairs were selected for OTU yield and taxonomic identification after comparison to other primer sets targeting the chloroplast loci *trnL*, *rbcL*, or *matK* or the nuclear ITS region due to their representation in NCBI Genbank. These had been previously utilized in studies involving metabarcoding, taxonomic identification, or *in silico* studies (19, 48-52). Chloroplast barcodes typically amplify better due to multiple genome copies but have more limited discrimination of related species,

while nuclear-based genome barcodes often have more difficulty in amplification and recovery. Using our protocol, the primer sets for the ITS and *matK* minibarcode regions did not amplify well, while the *trnL* minibarcode regularly produced more reads and OTUs when compared directly to other primer sets but was less able to define OTUs to the species level. Two different ITS2 minibarcode primer sets, including the pair used in this study, and *rbcL* minibarcode primer pairs targeting both the 5' and 3' regions of the gene amplified most consistently and produced the most mapped OTU (data not shown). Inclusion of additional minibarcode primer sets would likely improve plant OTU detection at the cost of having multiple OTUs representing the same species.

The quality and availability of plant biogeographic reference data is perhaps the most important factor affecting attribution applicability, accuracy, and resolution. BISON provided a tremendous wealth of information for determining species distributions to enable a proof-of-concept demonstration, but limitations in taxonomic, geographic, and temporal coverage (53) impacted the achievable attribution resolution and accuracy of our pipeline. Incomplete taxonomic coverage affected the ability to fully characterize the biodiversity distributions of sample OTUs. This was exacerbated by the difficulty in harmonizing the different taxonomic nomenclature systems used by NCBI Genbank and BISON, which likely resulted in an inability to retrieve occurrence records for some species. Geographic coverage, or how well the actual distribution of a species is documented by the occurrence records, and uneven spatial distribution of occurrence records also likely impacted the accuracy of attribution predictions (54). The species occurrence records may also be temporally skewed if a species' range has significantly changed in response to environmental shifts. To mitigate these issues, occurrence records can be augmented with data from other available sources, or with species distribution modeling, to enhance both taxonomic and geographic coverage. Lastly, though BISON data is significantly curated, there are possible data quality issues due to incorrect taxonomic assignment or duplicate or erroneous

entries. The same is true of NCBI Genbank, which is known to have sequence and taxonomic errors. Curation of these reference datasets could provide a significant improvement in attribution accuracy.

This analysis demonstrated that plant eDNA in dust from a significant percentage of samples is capable of reproducibly defining the U.S. region of sample origin within a radius of 600 km or less. The capability to acquire a regional estimate of provenance in many trace samples rapidly, without specialized expertise, can have value in many types of forensic investigation. We believe that, by streamlining metabarcoding protocols (using multiplexing, for instance) and automating data analysis, attribution information could be gained from hundreds of dust samples in days.

METHODS

Dust Collection

Dust was collected on standard 72x25x1 mm glass microscope slides (e.g., VWR VistaVision 16004-422) that were cleaned with glass cleaner, twice rinsed with distilled water, and dried with compressed air. Nine slides were secured with magnets onto three platforms of a louvered shelter (SRS100LX radiation shield, Ambient Weather, Chandler, AZ) mounted on a tripod one meter off the ground and at least 10 meters from buildings or other structures that could impede airflow (Supplementary Figure 4). The louvered shelter enabled particle settling, passive collection, on the slides while protecting them from precipitation. Dust was collected on slides at four U.S. locations, Lexington, MA, Panama City, FL, Socorro, NM, and Edgewood, SC, between March 2016 to June 2017. At the FL, NM, and SC sites, dust was collected from slides left undisturbed for 14 days once per season, which resulted in 4 or 5 slide sets per site. At the MA location, slide sets were collected after 14, 28, or 56 days of concurrent environmental exposure resulting in 28 14-day, 14 28-day, and 7 56-day sets. After environmental exposure, slides were removed from the collection rig and stored at 4°C within 5 days

of collection. Settled particles, comprised of pollen, spores, plant fragments, and inorganic material, were gathered from two slides from each set with a single Puritan DNA-free PurFlock Ultra Tipped Applicator with Transport Tubes (Puritan #253306UTTFDNA) wetted with isopropanol. These were air dried and stored at 4°C until further processing.

Wild Life Our Homes Samples

In addition to passively collected samples, 31 environmental dust samples from different U.S. locations that were collected as part of the Wild Life Our Homes (WLOH) citizen science project (32) were analyzed. These were a portion of a larger set of paired indoor and outdoor dust samples collected by enrolled citizen scientists between March 2012 and May 2013. Dust was collected by swabbing the upper door trim on an interior door in the main living area of a resident's home or the upper door trim on the outside surface of an exterior door. The *trnL* plant minibarcode information from 459 of these samples was previously described (55).

Metabarcoding and Sequencing

DNA was extracted from swabs using the MoBio/Qiagen PowerSoil htp-96 Well Isolation Kit (catalog #12955-4) according to the manufacturer's recommended protocol. Plant-specific minibarcodes targeting the ITS2 and *rbcL* regions (Supplementary Table 2) were used for their compatibility with next-generation sequencing (NGS) technology and for improved amplification of eDNA. Two minibarcodes were utilized to improve the number of plant species detected, and the OTU obtained with each primer set were combined and analyzed as a group. Minibarcodes were amplified and sequenced by a commercial vendor (Jonah Ventures, LLC, Boulder, CO) using primers with appended 5' adapter sequences. Each 25 µl amplification reaction consisted of 12.5 µl GoTaq PCR mastermix (Promega cat # M5133), 0.5 µl each forward and reverse primer

(0.2 μM final concentration), 1 μl extracted eDNA, and 10.5 μl DNase/RNase-free water. DNA was PCR amplified by denaturation at 94 °C for 5 minutes, followed by 40 cycles of 30 sec at 94 °C, 30 sec at 56 °C, and 45 sec at 72 °C, and a final elongation at 72 °C for 10 min. PCR amplifications were confirmed using agarose gel electrophoresis. Amplicons were then purified using MoBio UltraClean-htp 96 well PCR Clean-Up Kit (catalog # 12596-4) and each combination of sample and primer pair was assigned its own sequencing barcode. Pools were sequenced in triplicate using an Illumina MiSeq using the Reagent Kit v2 300-cycle kit (catalog # MS-102-2002). Sequences were demultiplexed using Golay barcodes (56) via QIIME v1.9.1 (57) and merging of paired end reads and trimming were performed with USEARCH (58). CUTADAPT v1.8.1 was then used to identify and remove remaining primer and adapter regions (59). Sequences were quality trimmed to have a maximum expected number of errors per read of less than 0.5. General quality filtering and OTU construction was completed as per the UPARSE pipeline (60) with de novo clustering at 99% sequence similarity. These parameters help to ensure that individual reads are correctly mapped to their respective OTU. Merged reads from ITS2 and single reads from *rbcL*-3A were clustered into OTUs (99% similarity using a de novo method). OTUs that had fewer than 3 reads, those that were not present in 3 of the 3 sequencing replicates, or those that had a relative abundance less than 10^{-4} were culled. This eliminated most of the OTUs (60.9% to 92.4%) from each sample, particularly those of lower abundance. The blank (buffer only) and negative control (clean slide) samples yielded no OTUs post-stringency filtering. Statistical analysis on the number of OTUs was performed nu analysis of variance (ANOVA).

Taxonomic Assignment and Species Distribution Determination

Taxa were assigned using the Genbank nBLAST homology inquiry tool using a query threshold of 100%. 40-50% of OTUs matched more than one species using a 100% homology threshold and, in this case, all species were retained. However, a plant species was only represented once per sample, even if it matched more

than one OTU. Genbank taxonomic nomenclature was not completely consistent with that of BISON, so assigned taxa were edited by trimming to only their genus and species, i.e. removing subspecies and variety names, then processed using the R package taxize (61, 62) to better align species names to those present in BISON. As a final step, manual curation corrected misspellings and removed unassigned taxa (uncultured, environmental sample, etc.).

Occurrence data was retrieved from the BISON database by using the taxonomic serial number (TSN), since ~98% of entries in BISON had an associated TSN, then by genus and species. Record retrieval from the BISON database was initially performed using a custom R script that retrieved records from the BISON application program interface (API) using the rbison package (63). The records retrieved were modified to exclude groups based on certain data fields, for instance records that have been flagged for having apparently invalid or mismatched latitude/longitude coordinates, countries, or continents. Up to 10^5 species occurrence records were retrieved for a single query. Some species produced no recorded occurrences with a geographic reference due to the lack of complete taxonomic coverage of the BISON biogeography database and the inability to resolve all nomenclature inconsistencies.

Mapping

The ArcGIS 10 geographic information system (GIS) package (Environmental Systems Research Institute (ESRI, Redlands, CA) was used to estimate the geographic attribution achieved from the species distributions associated with each sample. The primary output was a point-to-grid U.S. map that, within each grid, displayed the percentage of OTU from a sample with 2 or more occurrence records in that grid. To do this, the total observation records for each species assigned to an OTU were merged to generate an OTU-based geographic distribution. This step improved the attribution accuracy of the pipeline compared to consideration of each species within a sample independently. OTU-based geographic distributions were converted to an

analytical layer that was intersected with a global 250-km grid map displaying the number of occurrences per grid. Species prevalence in each grid was normalized by positive (two or more occurrences) or negative (less than two occurrences) designation. The maps of the normalized OTU-based geographic distributions for every OTU in a sample were overlaid and, for each grid, the proportion of the total OTU was calculated. This method minimized the impact of OTUs with wide geographic distributions (with occurrence records in many grids) that were less informative for geographic attribution, and enabled detection of OTU with a more localized distribution, which were more informative for geographic attribution. These steps were merged into a custom Python script, that utilized the ArcGIS library Arcpy (2014, ESRI) to enable automated analysis of multiple data sets and high throughput mapping.

To better compare point-to-grid maps, geographic attribution metrics were generated from a Gaussian Mixture Model (GMM) fitted to the point-to-grid map (**Error! Reference source not found.**). The GMM was fitted with Scikit-learn using the variational inference extension to the expectation maximization (EM) algorithm with the Dirichlet process to determine the number of OTUs in the mixture (64). This incorporated the analytical advantages of having a probabilistic model for the PtG data while retaining the robustness to low quality data (normalization) provided by the PtG method. The primary metrics utilized truth data, in this case the actual collection location, to determine the accuracy and resolution of the geographic attribution estimated from the plant OTU. Accuracy, designated truth percentage (TP), indicated how closely an attribution map came to measuring the location of sample origin by measuring the likelihood percentile of the truth point in the data, i.e. the percentage of grids with less than the OTU count value in the grid containing the location of sample collection. A higher TP reflected better accuracy. Attribution resolution described the spatial precision of the data, with the primary metric calculated by determining the distance(s) between the truth point (location of the sample origin) and top 5 points (map grids) with highest likelihood/OTU number. This was referred to as the average top 5 peaks error (AT5PE), and was the standard indicator of attribution resolution.

REFERENCES

1. Stoney D, Bowen A, Stoney P. Inferential source attribution from dust: review and analysis. Forensic Sci Rev. 2013;25:107-42.

2. Locard E. The analysis of dust traces. Am J Police Sci. 1930;1:276.

3. Wickenheiser RA. Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. Journal of Forensic Science. 2002;47(3):442-50.

4. Adams-Groom B. Frequency and abundance of pollen taxa in crime case samples from the United Kingdom. Grana. 2015;54(2):146-55.

5. Bryant VM, Jones GD. Forensic palynology: Current status of a rarely used technique in the United States of America. Forensic Science International. 2006;163(3):183-97.

6. Laurence AR, Bryant VM. Forensic palynology and the search for geolocation: Factors for analysis and the Baby Doe case. Forensic science international. 2019;302:109903.

7. Mildenhall D, Wiltshire PE, Bryant VM. Forensic palynology: why do it and how it works. Elsevier; 2006.

8. Taylor B, Skene K. Forensic palynology: spatial and temporal considerations of spora deposition in forensic investigations. Australian Journal of Forensic Sciences. 2003;35(2):193-204.

9. Halbritter H, Ulrich S, Grímsson F, Weber M, Zetter R, Hesse M, et al. Methods in Palynology. Illustrated Pollen Terminology: Springer; 2018. p. 97-127.

10. Stillman E, Flenley JR. The needs and prospects for automation in palynology. Quaternary Science Reviews. 1996;15(1):1-5.

11. Walsh KA, Horrocks M. Palynology: its position in the field of forensic science. Journal of forensic sciences. 2008;53(5):1053-60.

12. Christou C, Jacyna G, Goodman F, Deanto D, Masters D, editors. Geolocation analysis using Maxent and plant sample data. 2015 IEEE International Symposium on Technologies for Homeland Security (HST); 2015: IEEE.

13. Goodman F, Doughty J, Gary C, Christou C, Hu B, Hultman E, et al., editors. PIGLT: a pollen identification and geolocation system for forensic applications. Technologies for Homeland Security (HST), 2015 IEEE International Symposium on; 2015: IEEE.

14. Lagerstrom R, Arzhaeva Y, Bischof L, Haberle S, Hopf F, Lovell D, editors. A comparison of classification algorithms within the classifynder pollen imaging system. AIP Conference Proceedings; 2013: American Institute of Physics.

15. Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. Systematic biology. 2005;54(5):852-9.

16. Kress WJ. Plant DNA barcodes: Applications today and in the future. Journal of Systematics and Evolution. 2017;55(4):291-307.

17. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. A plea for DNA taxonomy. Trends in Ecology & Evolution. 2003;18(2):70-4.

18. Bell KL, Burgess KS, Okamoto KC, Aranda R, Brosi BJ. Review and future prospects for DNA barcoding methods in forensic palynology. Forensic Science International: Genetics. 2016;21:110-6.

19. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS one. 2010;5(1):e8613.

20. Fahner NA, Shokralla S, Baird DJ, Hajibabaei M. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution, and annotation of four DNA markers. PloS one. 2016;11(6):e0157505.

21. Group CPW, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. Proceedings of the National Academy of Sciences. 2009;106(31):12794-7.

22. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research. 2006;35(3):e14-e.

23. Matsuki Y, Isagi Y, Suyama Y. The determination of multiple microsatellite genotypes and DNA sequences from a single pollen grain. Molecular Ecology Notes. 2007;7(2):194-8.

24. Kraaijeveld K, De Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, et al. Efficient and sensitive identification and quantification of airborne pollen using next‐generation DNA sequencing. Molecular ecology resources. 2015;15(1):8-16.

25. Müller-Germann I, Vogel B, Vogel H, Pauling A, Fröhlich-Nowoisky J, Pöschl U, et al. Quantitative DNA analyses for airborne birch pollen. PloS one. 2015;10(10).

26. Bell KL, Fowler J, Burgess KS, Dobbs EK, Gruenewald D, Lawley B, et al. Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. Applications in Plant Sciences. 2017;5(6):1600124.

27. Keller A, Danner N, Grimmer G, Ankenbrand M, von der, Von Der Ohe K, Von Der Ohe W, et al. Evaluating multiplexed next‐generation sequencing as a method in palynology for mixed pollen samples. Plant Biology. 2015;17(2):558-66.

28. Sickel W, Ankenbrand MJ, Grimmer G, Holzschuh A, Härtel S, Lanzen J, et al. Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. BMC ecology. 2015;15(1):20.

29. Prosser SW, Hebert PD. Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. Food Chemistry. 2017;214:183-91.

30. Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N. Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. Scientific Reports. 2020;10(1):1-9.

31. Bell KL, de Vere N, Keller A, Richardson RT, Gous A, Burgess KS, et al. Pollen DNA barcoding: current applications and future prospects. Genome. 2016;59(9):629-40.

32. Barberán A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, et al. Continental-scale distributions of dust-associated bacteria and fungi. Proceedings of the National Academy of Sciences. 2015;112(18):5756-61.

33. Lane MA, Edwards JL. The global biodiversity information facility (GBIF). Biodiversity Databases. 1: ROUTLEDGE in association with GSE Research; 2007. p. 1-4.

34. Giampaoli S, Berti A, Di Maggio R, Pilli E, Valentini A, Valeriani F, et al. The environmental biological signature: NGS profiling for forensic comparison of soils. Forensic science international. 2014;240:41-7.

35. Young J, Austin J, Weyrich L. Soil DNA metabarcoding and high-throughput sequencing as a forensic tool: considerations, potential limitations and recommendations. FEMS microbiology ecology. 2017;93(2).

36. Grantham NS, Reich BJ, Pacifici K, Laber EB, Menninger HL, Henley JB, et al. Fungi identify the geographic origin of dust samples. PLoS One. 2015;10(4):e0122605.

37. Allwood JS, Fierer N, Dunn RR, Breen M, Reich BJ, Laber EB, et al. Use of standardized bioinformatics for the analysis of fungal DNA signatures applied to sample provenance. Forensic Science International. 2020:110250.

38. Grantham NS, Reich BJ, Laber EB, Pacifici K, Dunn RR, Fierer N, et al. Global forensic geolocation with deep neural networks. arXiv preprint arXiv:190511765. 2019.

39. Damaso N, Mendel J, Mendoza M, von Wettberg EJ, Narasimhan G, Mills D. Bioinformatics approach to assess the biogeographical patterns of soil communities: the utility for soil provenance. Journal of forensic sciences. 2018;63(4):1033-42.

40. Lenehan CE, Tobe SS, Smith RJ, Popelka-Filcoff RS. Microbial composition analyses by 16S rRNA sequencing: A proof of concept approach to provenance determination of archaeological ochre. PloS one. 2017;12(10):e0185252.

41. Badgley AJ, Jesmok EM, Foran DR. Time Radically Alters Ex Situ Evidentiary Soil 16S Bacterial Profiles Produced Via Next‐Generation Sequencing. Journal of forensic sciences. 2018;63(5):1356-65.

42. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology. 2017;15(10):579.

43. Thomsen PF, Willerslev E. Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation. 2015;183:4-18.

44. Kew RBG. The state of the world's plants report–2016. Royal Botanic Gardens, Kew. 2016.

45. Ulloa CU, Acevedo-Rodríguez P, Beck S, Belgrano MJ, Bernal R, Berry PE, et al. An integrated assessment of the vascular plant species of the Americas. Science. 2017;358(6370):1614-7.

46. Li H, Bai H, Yu S, Han M, Ning K. Holmes-ITS2: Consolidated ITS2 resources and search engines for plant DNA-based marker analyses. bioRxiv. 2018:263541.

47. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. Molecular ecology. 2016;25(7):1423-8.

48. Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC genetics. 2014;15.

49. Dunning LT, Savolainen V. Broad-scale amplification of matK for DNA barcoding plants, a technical note. Botanical Journal of the Linnean Society. 2010;164(1):1-9.

50. Han J, Zhu Y, Chen X, Liao B, Yao H, Song J, et al. The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. BioMed research international. 2013;2013.

51. Little DP. A DNA mini‐barcode for land plants. Molecular ecology resources. 2014;14(3):437-46.

52. Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M. A universal DNA mini-barcode for biodiversity analysis. BMC genomics. 2008;9(1):214.

53. Meyer C, Weigelt P, Kreft H. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters. 2016;19(8):992-1006.

54. Meyer C, Kreft H, Guralnick R, Jetz W. Global priorities for an effective information basis of biodiversity distributions. Nature Communications. 2015;6:8221.

55. Craine JM, Barberán A, Lynch RC, Menninger HL, Dunn RR, Fierer N. Molecular analysis of environmental plant DNA in house dust across the United States. Aerobiologia. 2017;33(1):71-86.

56. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012;6(8):1621.

57. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods. 2010;7(5):335.

58. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.

59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):pp. 10-2.

60. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods. 2013;10(10):996-8.

61. Chamberlain S, Szocs E, Boettiger C, Ram K, Bartomeus I, Baumgartner J, et al. Taxize: taxonomic information from around the web. Version 0.7. 8. 2016.

62. Chamberlain SA, Szöcs E. taxize: taxonomic search and retrieval in R. F1000Research. 2013;2.

63. Chamberlain SA, Boettiger C. R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints; 2017. Report No.: 2167-9843.

64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12:2825-30.

DECLARATIONS

This material is based upon work supported under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

ACKNOWLEDGEMENTS

We would like to thank Joe Craine (Jonah Ventures), Gerald “Stinger” Guala (USGS), Healy Hamilton (NatureServe), and Jerry Sellers for helpful discussions. We’d also like to acknowledge Joshua Dettman and Natalie Damaso for thoughtful review of the manuscript and Noah Fierer for access to WLOH dust samples.

AUTHOR CONTRIBUTIONS

CP developed methods and software to process, obtain, and curate the metabarcode data and to interface with sequence and biogeographic databases She also analyzed the data and assisted significantly with data interpretation. JK developed the geographic attribution metrics and methods/software for their acquisition from the attribution data. SC developed methods and software to process data and generate point-to-grid maps in ArcGIS. JB was critical in sample collection and processing. JC designed, assembled, and tested the dust sampling stations, collected samples, and organized sample collection from various sites. LB developed the methodology to integrate species distribution data by OTU. JC planned the studies, assisted with sample collection, prepared samples, analyzed and interpreted the data, and drafted the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests

ETHICS DECLARATION

Not applicable

CONSENT TO PUBLISH

Not applicable

DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

FIGURE LEGENDS

Figure 1. Diagram summarizing the plant eDNA geographic attribution pipeline used in this study. Starting from settled dust, DNA is extracted then subjected to metabarcoding with ITS2 and *rbcL*-3A, sequencing, and data processing to obtain an estimate of the site of origin.

Figure 2. Number of total OTU per sample generated with ITS2 and *rbcL*-3A minibarcodes from dust samples collected in Lexington, MA. Samples were collected on the date indicated after 14 (blue), 28 (orange), or 56 (gray) days of environmental exposure prior to analysis.

Figure 3. Heat map of the 40 most abundant OTUs found in 14-day dust samples collected in Lexington, MA. The OTU number, barcode used, and assigned genus according to NCBI Genbank are listed, as is the relative abundance in the sample collected on the date indicated. The shade of color indicates the number of reads from a minimum of 10 (light green) to roughly 50,000 (dark green).

Figure 4. (A) Point-to-grid maps displaying overlaid distributions from OTU using ITS2 and *rbcL*-3A minibarcodes generated from 14-day dust samples collected from four locations on the date indicated. Grid color represents the percentage of total OTU in that sample that had a threshold number of point occurrence

records within that grid. The location of sample collection is indicated (pink star). (B) Maps showing the result of Gaussian modeling model fitting to the data generated in the point-to-grid map. The location of sample collection is indicated (pink star) as are the locations of the top five peaks used to calculate AT5PE.

Figure 5. Correlation of total mapped OTU in a sample to its attribution accuracy metrics. Dust samples collected after 14-day exposure in MA (gray), FL (blue), NM (red), or SC (orange) were subject to plant metabarcoding and the resulting OTUs were mapped using our attribution pipeline. (A) Relationship between mapped OTU and TP, with a 90% TP cutoff indicated. (B) Relationship between mapped OTU and AT5PE, with a 600 km cutoff indicated.

Figure 6. Attribution achieved with plant eDNA derived from 31 WLOH dust samples. (A) Correlation of TP to AT5PE in WLOH dust samples in samples with 20 or more (blue) or less than 20 (red) mapped OTUs. Cutoffs for 90% TP and 600 km AT5PE are shown. (B) Map of the site of sample collection of the 31 WLOH samples in addition to the attribution accuracy (TP) and resolution (AT5PE) determined from assessment of plant eDNA from dust samples. The color indicates a TP of greater (green) or less than (brown) 90%. The shape of the marker indicates AT5PE of < 100 km (star), 100 – 500 km (circle), 500 – 1000 km (square), or > 1000 km (diamond).

Table 1. Attribution metrics for 14-day dust samples.

	All samples				Samples with ≥ 20 OTUs			
	#	≥ 90%TP	< 600 km AT5PE	True positive	#	≥ 90% TP	< 600 km AT5PE	True positive
FL	5	0	2	0	2	0	0	0
MA	28	15	17	15	16	13	13	12
NM	5	3	3	3	4	2	2	2
SC	4	2	2	2	2	2	2	2
all sites	42	20	24	20	24	20	17	16

Total number of 14-day dust samples, or the subset of 14-day dust samples with ≥ 20 OTU, that yielded a ≥ 90% TP, less than 600 km AT5PE, or true positive geographic attribution from analysis of constituent plant eDNA from samples collected at the sites indicated. A true positive attribution was defined ≥ 90% TP with < 600 km AT5PE.

Figures

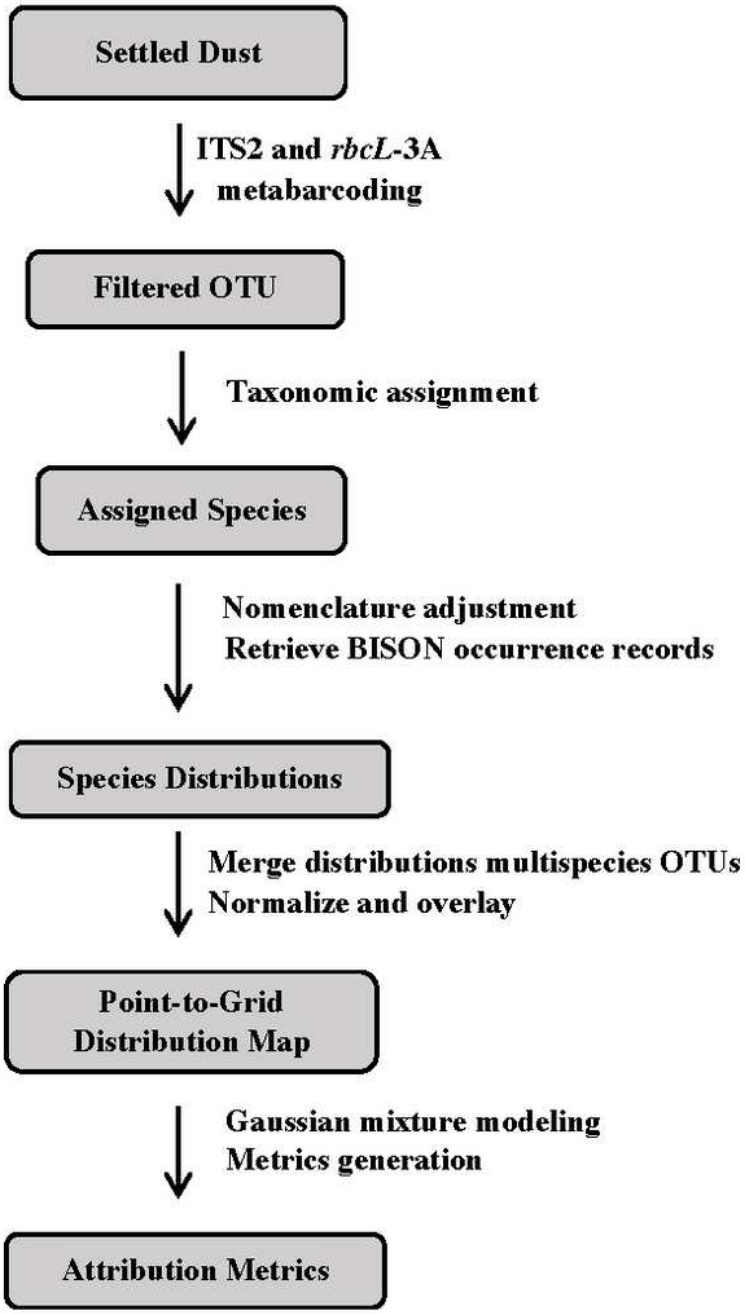

Figure 1

Diagram summarizing the plant eDNA geographic attribution pipeline used in this study. Starting from settled dust, DNA is extracted then subjected to metabarcoding with ITS2 and rbcL-3A, sequencing, and data processing to obtain an estimate of the site of origin.

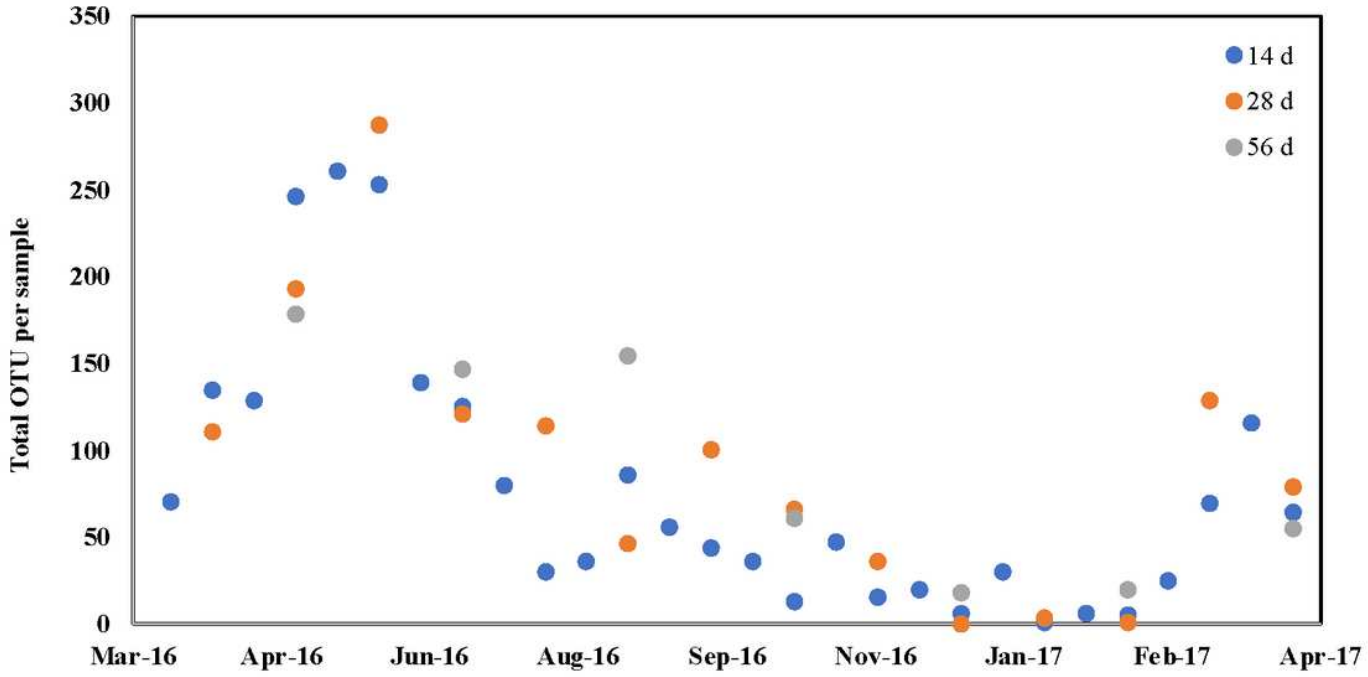

Figure 2

Number of total OTU per sample generated with ITS2 and rbcL-3A minibarcodes from dust samples collected in Lexington, MA. Samples were collected on the date indicated after 14 (blue), 28 (orange), or 56 (gray) days of environmental exposure prior to analysis.

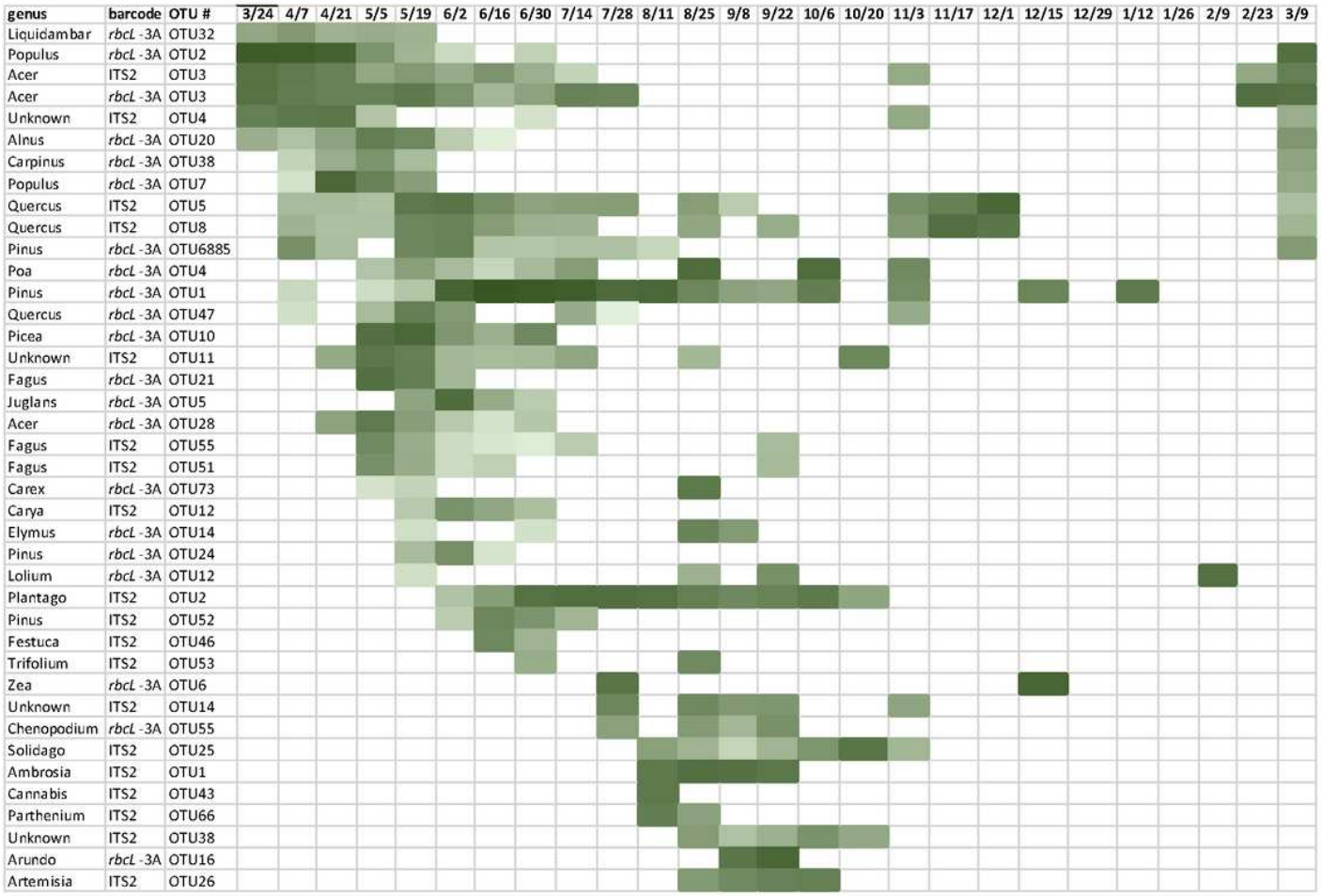

Figure 3

Heat map of the 40 most abundant OTUs found in 14-day dust samples collected in Lexington, MA. The OTU number, barcode used, and assigned genus according to NCBI Genbank are listed, as is the relative abundance in the sample collected on the date indicated. The shade of color indicates the number of reads from a minimum of 10 (light green) to roughly 50,000 (dark green).

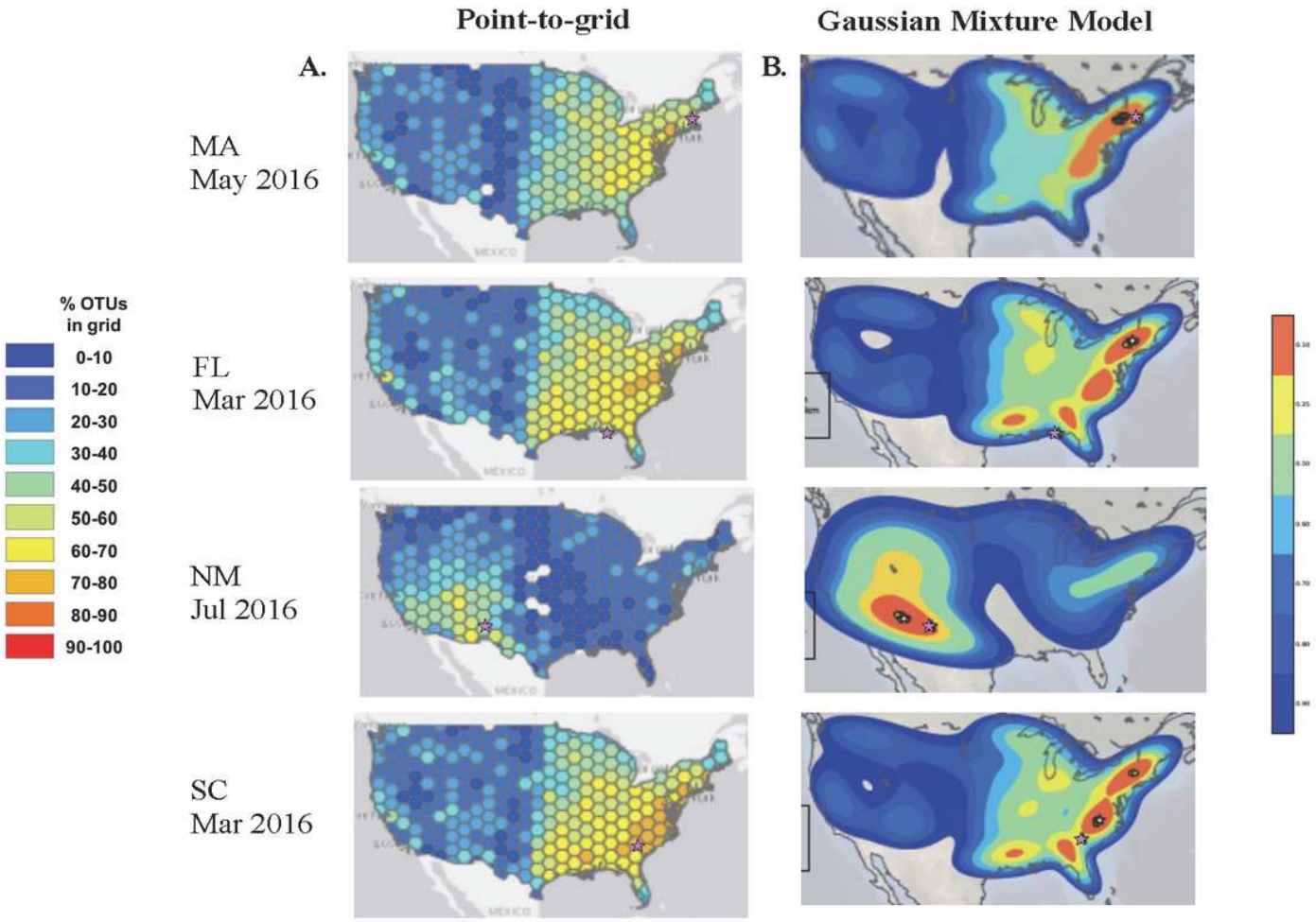

Figure 4

(A) Point-to-grid maps displaying overlaid distributions from OTU using ITS2 and rbcL-3A minibarcodes generated from 14-day dust samples collected from four locations on the date indicated. Grid color represents the percentage of total OTU in that sample that had a threshold number of point occurrence records within that grid. The location of sample collection is indicated (pink star). (B) Maps showing the result of Gaussian modeling model fitting to the data generated in the point-to-grid map. The location of sample collection is indicated (pink star) as are the locations of the top five peaks used to calculate AT5PE.

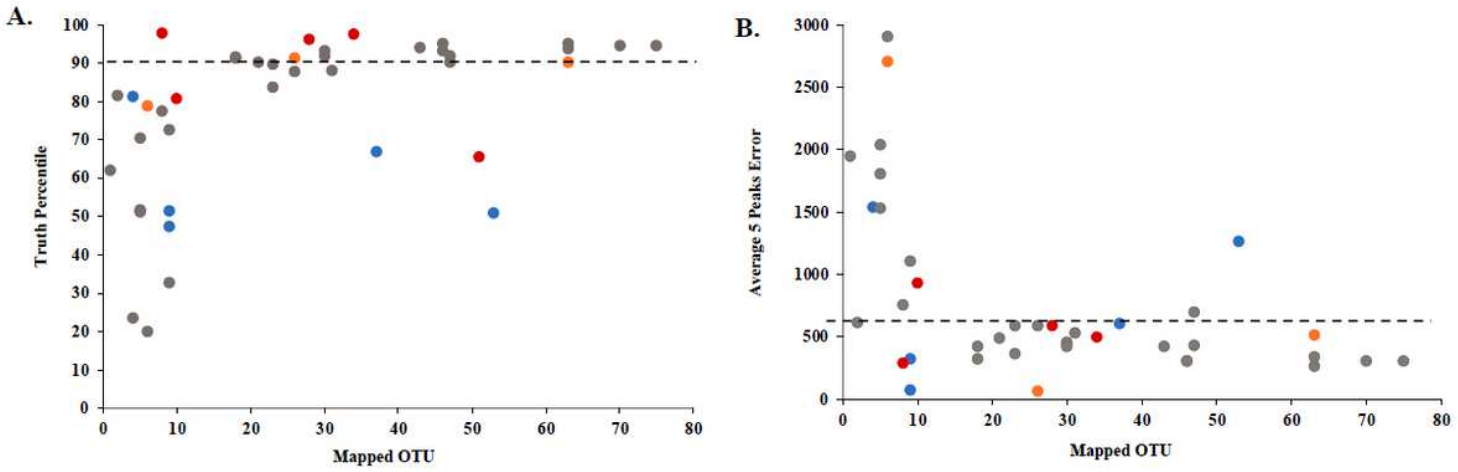

Figure 5

Correlation of total mapped OTU in a sample to its attribution accuracy metrics. Dust samples collected after 14-day exposure in MA (gray), FL (blue), NM (red), or SC (orange) were subject to plant metabarcoding and the resulting OTUs were mapped using our attribution pipeline. (A) Relationship between mapped OTU and TP, with a 90% TP cutoff indicated. (B) Relationship between mapped OTU and AT5PE, with a 600 km cutoff indicated.

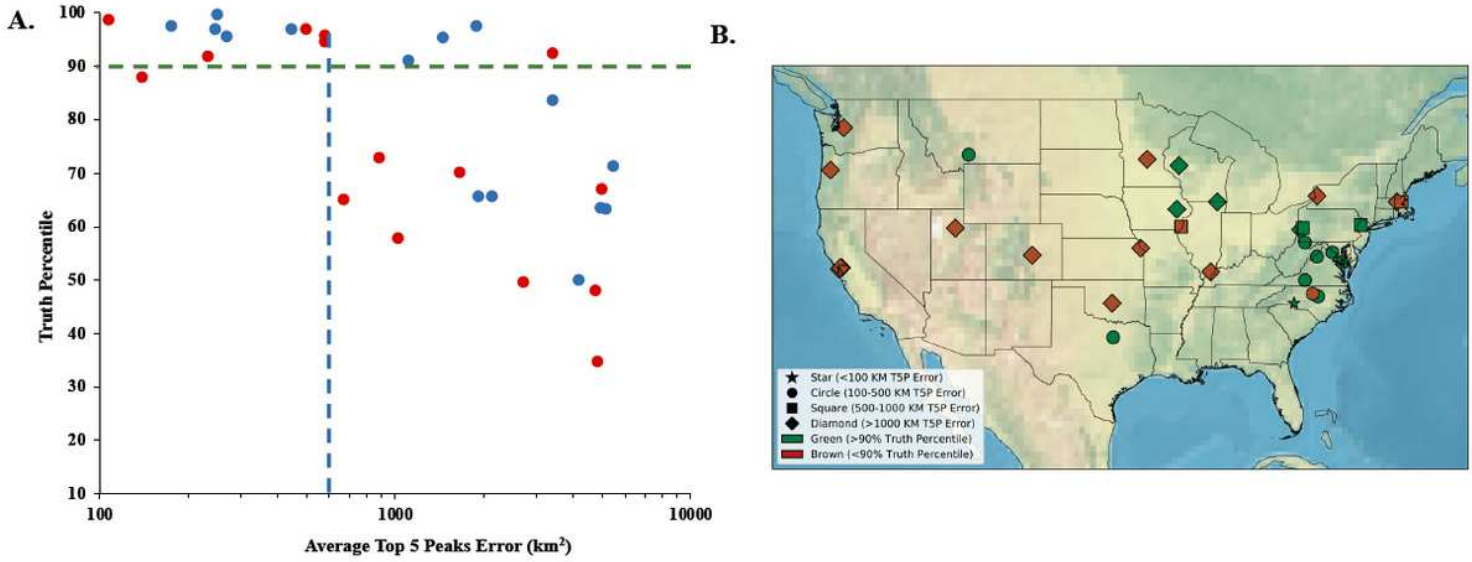

Figure 6

Attribution achieved with plant eDNA derived from 31 WLOH dust samples. (A) Correlation of TP to AT5PE in WLOH dust samples in samples with 20 or more (blue) or less than 20 (red) mapped OTUs. Cutoffs for 90% TP and 600 km AT5PE are shown. (B) Map of the site of sample collection of the 31 WLOH samples in addition to the attribution accuracy (TP) and resolution (AT5PE) determined from assessment of plant eDNA from dust samples. The color indicates a TP of greater (green) or less than

(brown) 90%. The shape of the marker indicates AT5PE of < 100 km (star), 100 – 500 km (circle), 500 – 1000 km (square), or > 1000 km (diamond).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Comollietalsupplementaryinfo13May21.pdf