Determination of a criminal suspect using environmental plant DNA metabarcoding technology

Yanlei Liu[1], Chao Xu[2], Wenpan Dong[3], Xueying Yang[4], and Shi-Liang Zhou[2]

[1]Institute of Botany Chinese Academy of Sciences
[2]Institute of Botany, Chinese Academy of Sciences
[3] Institute of Botany, Chinese Academy of Sciences
[4]Institute of Forensic Science, Ministry of Public Security

March 30, 2022

Abstract

There are criminal cases that no frequently used evidence, for example, DNAs from the criminal, is available. Such cases usually are unresolvable. With the advent of DNA metabarcoding, evidences are mined from environmental DNA and such cases become resolvable. This study reports how a criminal suspect was determined by environmental plant DNA metabarcoding technology. A girl was killed in a rural wet area in China without a witness or video record. Pants with dried mud was found from one of her boyfriend's house. The mud was removed from the pants and 11 more mud or soil samples surrounding murder scene were collected. DNA was extracted from the soil. Chloroplast rbcL gene fragments were amplified and sequenced on a next generation sequencing platform. Of the 2980 ZOTUs in total from the 12 samples, 1495 ZOTUs were identified to species, genera or families based on the existing public database. The feast analysis based on either taxa or taxa plus abundance data demonstrated that the mud on the suspect's pants was from the criminal scene. The suspect finally made a clean breast of his crime. This case implies that plant DNA in the environment soil is a new source of evidence in determination of suspects using DNA metabarcoding technology and has high potentials of extensive applications in criminal cases.

Determination of a criminal suspect using environmental plant DNA metabarcoding technology

Yanlei Liu[1,2], Chao Xu[1], Wenpan Dong[1], Xueying Yang[3*] and Shiliang Zhou[1*]

1. State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. National Engineering Laboratory for Forensic Science, Key Laboratory of Forensic Genetics, Institute of Forensic Science, Ministry of Public Security, Beijing 100038, China

*Correspondence: Shiliang Zhou, Fax: 86-10-62590843, E-mail: slzhou@ibcas.ac.cn;

Xueying Yang, Fax: 86-10-66269514, E-mail: yxystyhhp@163.com

Abstract

There are criminal cases that no frequently used evidence, for example, DNAs from the criminal, is available. Such cases usually are unresolvable. With the advent of DNA metabarcoding, evidences are mined from environmental DNA and such cases become resolvable. This study reports how a criminal suspect was determined by environmental plant DNA metabarcoding technology.

A girl was killed in a rural wet area in China without a witness or video record. Pants with dried mud was found from one of her boyfriend's house. The mud was removed from the pants and 11 more mud or soil samples surrounding murder scene were collected. DNA was extracted from the soil. Chloroplast *rbcL* gene fragments were amplified and sequenced on a next generation sequencing platform. Of the 2980 ZOTUs in total from the 12 samples, 1495 ZOTUs were identified to species, genera or families based on the existing public database. The feast analysis based on either taxa or taxa plus abundance data demonstrated that the mud on the suspect's pants was from the criminal scene. The suspect finally made a clean breast of his crime. This case implies that plant DNA in the environment soil is a new source of evidence in determination of suspects using DNA metabarcoding technology and has high potentials of extensive applications in criminal cases.

Key words : DNA barcoding; forensics; DNA metabarcoding; environmental plant DNA

Introduction

Human deoxyribonucleic acid (DNA) has been widely used in human individual identification (Ambers et al., 2018; Lygo et al., 1994; Meng et al., 2019), paternity identification (Bertoglio et al., 2020; Habibi et al., 2019) and other applications in forensics. However, human DNA is not always available. Under this situation, we have to resort to environmental DNA in the crime scene to narrow the search scope for criminal suspects and find out the truth.

Environmental materials such as soil, dust, water, etc., are very likely to be taken away unintentionally by suspects on his or her skin, shoes, clothes, hair or even in the nail seams. Among them, soil, usually contaminated by plant fragments or pollen grains, is the material the police can get in most criminal cases. Plant DNA is quite suitable for the forensic source tracking because of its ubiquity, stability and proper variability.

Plant DNA has a high potential providing definitive evidence during criminal investigations. With the advent of DNA metabarcoding, it has recently been used to find out body dumping site (Yang et al., 2015), residence of unknown human body (Liu et al., 2019), drowning site (Fang et al., 2019), and confirmation of suspected drowning (Kakizaki et al., 2018). Unfortunately, such applications are still very rare due to three main challenges. The first one is the difficulties in species identification of plant DNA in the environmental materials. Past projects (e.g., BARCODE 500K (https://ibol.org), BIOSCAN (Hobern & Hebert, 2019), ISHAM-ITS (Irinyi et al., 2016)) have enriched the pool of DNA barcodes, though the reference library for DNA barcoding is rather not comprehensive. Only less than 5.0% species of flowering plants have their *matK* or *rbcL* sequences deposited in GenBank (Liu et al. 2021).

The second challenge is that the Sanger sequencing method is not applicable to environmental DNA because the amplicons are a mixture of many species. Fortunately, next generation sequencing (NGS) platforms meet the requirement of environmental DNA metabarcoding and a very easy data processing method is now available (https://github.com/YanleiLiu1989/Cotu-master).

The last challenge is lack of an "ideal" DNA barcode for DNA metabarcoding (Ferri et al., 2015). DNA barcode is a short DNA sequence for species recognition and discrimination. DNA barcoding is a commonly used biotechnology in biology, environmental science, forensics, etc (Ferri et al., 2015; Hebert et al., 2003). It is a powerful molecular diagnostic method for specimen identification. Finding the best DNA barcodes (Dong et al., 2014; Dong et al., 2015; Kress & Erickson, 2007; Li et al., 2011) or developing new technical improvements (Yu et al., 2011; Xu et al., 2015) was one of the main themes for plant DNA barcoding during the past decade. Unfortunately, there is not a single ideal DNA barcode suitable for all plant species identification, and plant group-specific DNA barcodes seem more realistic. For example,*rbcL* is much less variable than *ycf1* in flowering plants, but acceptable as a DNA barcode for lower plants (Dong et al., 2015; Liu et al., 2020a).

The lower plants (algae) instead of higher plants (mosses, ferns and seed plants) play a very important role in investigation of wet environment-related criminal cases and *rbcL* has been proposed as a DNA barcode

of diatoms (Liu et al., 2020a). The variability of*rbcL* is much higher in lower plants than in higher plants and*rbcL* is one of the few choices of DNA barcodes for lower plants for its relatively higher species coverage of existing sequences and universal PCR primers (Ferri et al., 2015).

In this paper, we demonstrate how to use mud collected from a criminal suspect's pants to determine the real criminal in a murder case happened in China based on DNA metabarcoding of diatom using chloroplast*rbcL* gene fragments. The diatom communities in the mud provided solid evidence of the suspect's appearance in the murder scene.

Materials and methods

The whole study procedure includes five parts: soil DNA acquisition, amplicon preparation, amplicon sequencing, NGS data processing and suspect's mud tracking (Fig. S1).

Soil DNA acquisition

A total of 12 soil samples were collected, including one mud soil sample collected from the criminal suspect's pants, three soil samples from the center of crime scene and eight soil samples surrounding crime scene (Table 1, Fig. 1).

Approximately 25 mg of fully mixed soil of each sample was ground into powder in a grinder mill (MM400, Retsch GmbH, Germany) equipped with a zirconium magnetic bead at 29 Hz for two minutes at 30-second intervals to minimize DNA damage. Total soil DNA was extracted using the mCTAB method (Li et al., 2013). DNA was resuspended in 100 μL of TE buffer, visually checked on 1.5% agarose gels, and quantified on a Nanodrop2000c sectrophotometer (Thermo Fisher Scientific Inc., USA).

Amplicon preparation

Since this murder case happened in a small canal, we chose *rbcL*gene of diatom as our DNA barcode. The primer pair BacirbcL2f and BacirbcL2r (Liu et al., 2020a) was used for this case. DNA fragments from the same sample were labeled with a unique DNA oligo by PCR (Table 2, Fig. 2). A unique eight-nucleotide oligo for each sample was attached to the 5' end of both forward and reverse primers. PCR with a 10 μL mixture was conducted on Eppendorf instrument Mastercycler proS following Dong et al. (2015). The PCR products were checked by electrophoresis with a 1.5% agarose gel containing ethidium bromide under ultraviolet transilluminator.

The DNA-labelled PCR products were mixed, purified using a purification kit (Aidlab Biotechnologies Co., Ltd, China) on a 2% agarose gel, and quantified on a Nanodrop2000c spectrophotometer.

Amplicon sequencing

A sequencing library of the final PCR mixture was constructed for Ion Torrent platform using NEBNext® Fast DNA Library Prep Set for Ion Torrent (New England BioLabs, USA) and the library was sequenced at Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences on Ion Torrent S5xl Chip400.

NGS data processing

Quality control and demultiplexing. NGS data quality control was carried out using the NGS QC toolkit with the default parameters (Ravi et al., 2012). After quality control, the NGS data from Ion torrent S5xl were demultiplexed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) according to the sample labels and primers (Table 2).

Label and primer sequence removal. Low quality sequences and sequences shorter than 200 bp were discarded using NGS QC toolkit. Artificially added sequences, such as DNA labels and primers were trimmed off using Cutadapt software (https://cutadapt.readthedocs.io/en/stable/).

ZOTU generation and annotation. We followed the Unoise3 protocol (*http://www.drive5.com/usearch*) (Edgar, 2016) to generate ZOTUs of all 12 samples. All unique ZOTU sequences were identified using BLAST (Altschul, 2012) against NCBI database to assign scientific names to ZOTU sequences if possible.

Organism abundance. Each of all ZOTU sequences was used as a reference and the reads from each soil sample were mapped to the reference under a similarity of 0.97 using Usearch (Edgar, 2013). The number of reads matching each ZOTU was recorded as abundance of the ZOTU (organism).

Suspect's mud tracking

The potential origins of diatoms found in the mud from the suspect's pants were tracked to the 11 candidate soil samples by fast expectation-maximization for microbial source tracking (FEAST, Shenhav et al., 2019). FEAST is a software developed for deducing the potential origin(s) of a microorganism community. FEAST estimates the fraction of organisms from the potential source as well as the other sources as unknown source, which helps to verify the true or false source of microorganism community in the mud from the suspect's pants. FEAST is currently implemented in R and easy to run following the instructions online (*https://github.com/cozygene/FEAST*).

Results

Data from the Ion Torrent S5xl platform

After sequencing on the Ion Torrent S5xl platform, a total of 2,917,507 raw reads was obtained. After quality control and read length selection, 2,754,982 clean data (94.43%) were retained. The mean sequencing depth of soil samples is 229,581 reads.

Total and annotated number of ZOTUs

A total of 2,980 ZOTU sequences were created using Usearch. The mean ZOTU sequences per sample were 961. Among 12 samples Site 5-1 was the most abundant sample with 1,176 ZOTU sequences, while Site 3-1 was the least abundant sample with 725 ZOTU sequences (Table 1).

Among all ZOTU sequences, 1,495 ZOTU sequences (50.17%) could be annotated to genera or lower taxonomy level. The mean number of identified taxa per sample is 542. Among the 12 samples Site 2-1 was the most abundant sample with 762 known taxa, while Site 4-1 was the least abundant sample with 274 known taxa (Table 1). Of the 1,495 identifiable ZOTU sequences, 1243 ZOTU sequences were identified to 134 species of diatoms and 221 ZOTU sequences were identified to 77 genera of diatoms (Fig S2). *Amphora* , *Aulacoseira* , *Diadesmis* ,*Gomphonema* , *Lemnicola* , *Melosira* , *Placoneis* ,*Planothidium* , *Sellaphora* , and *Stauroneis* are the dominant genera in the 12 samples, and *Diadesmis confervacea* ,*Gomphonema parvulum* , *Lemnicola hungarica* , *Melosira varians* , *Placoneis elginensis* , *Sellaphora pupula* , and*Stauroneis kriegeri* are the dominant diatom species.

The mud on the suspect's pants was from the criminal scene

All results came to the conclusion that the mud on the suspect's pants was from the criminal scene (Fig. 3). When all ZOTUs were used and singletons were considered, about 56.19% of the diatom ZOTUs in the mud on the suspect's pants were trackable to sample Site 1-1, 7.44% to Site 1-2, 23.90% to Site 1-3, and only 12.47% to other sources (Fig. 3A). If singletons were not considered, the percentages were slightly higher, 65.75% to Site 1-1, 1.34% to Site 1-2, 25.18% to Site 1-3, and only 7.73% to other sources (Fig. 3B).

When only annotated ZOTUs were used and singletons were considered, the percentages were 65.99% to Site 1-1, 2.31% to Site 1-2, 25.91% to Site 1-3, and only 7.73% to other sources (Fig. 3C). When singletons were not considered, the figures were 67.08% to Site 1-1, 0.81% to Site 1-2, 24.42% to Site 1-3, and only 7.69% to other sources (Fig. 3D).

Discussion

Environmental plant DNA, a new source of evidence for difficult crime cases

Before the emergence of DNA metabarcoding technology, it is unrealistic to extract evidences from environmental DNA for forensic purpose because of high DNA cloning and sequencing costs. With the DNA metabarcoding technology, cold cases without witness, video record or human DNA become resolvable now.

DNA metabarcoding powered by next generation sequencing (NGS) technology has now become a powerful approach in forensic evidence collection from environmental samples (Young et al., 2017). This study provides one more example demonstrating successful tracking of the source of mud on the suspect's pants via diatom community. Suspect-related environmental samples including diatoms or pollen are new sources of material evidences. Special attentions should be paid to some technical aspects such as contaminations, DNA barcodes, and data processing methods when using DNA metabarcoding data as evidence.

Organism contamination and false positive

Plant particles such as leaf fragments, pollen grains and spores can be carried away by wind and contaminations to experimental samples are very likely to arise while collecting and processing samples. Although higher plants cannot move, spores and pollen grains can move for a long distance (Rousseau et al., 2006). Such particles accumulate on surfaces on experimental benches, apparatuses and even clothes (Thomsen & Willerslev, 2015). Despite the amount of contaminants are very very small, they are detectable when amplified by PCR and sequenced on NGS platforms. A biological contaminant free laboratory is ideal for this purpose. To lower the risk of contaminants, only organisms specific to unique environments can be considered, for example, diatoms in wet environment, forage herbs in prairies, crops in crop fields. Diatoms are of high species diversity and ubiquitous in wet environments. In this study, we barcoded diatoms in the soil because the accident happened in a canal and diatoms are very good indicators.

Suitable DNA barcode for DNA metabarcoding

Whether a soil sample could be traced back to its original source localities depends on the suitable DNA barcode to be used. Although there is incompatibility between the universality of primers and the resolution of barcodes for higher plants, the universality of primers is more important for DNA metabarcoding of lower plants in soil samples because almost all DNA barcodes are variable enough to resolve most known taxonomic units (Liu et al., 2020b). Lower plants have shorter lifetime, evolve much more quickly and accumulate more genetic variations in their genomes than higher plants. However, lower plants are the least known creatures to taxonomists and quite large of them can only be identified to genus or even family levels. For example,*rbcL* is one of the least variable gene of seed plants, but its variability in lower plants is much higher (Dong et al., 2014) and can serve as a DNA barcode for lower plants such as diatoms (Liu et al., 2020a).

Another advantage of using *rbcL* as a DNA barcode for lower plants is that this gene locates in plastid genome, implying that contaminations of plastid genome free organisms bring no trouble to the data analyses. There is usually a vast range of microorganisms in soil samples, for example, insects, fungi, bacteria, etc.. When using DNA barcodes from nuclear genome such as 18S, amplification of these organisms is usually inevitable, which needs quite large experimental and analysis resources.

A sequence reference library is not a prerequisite, but it is something better than nothing

Soil sample source tracking using DNA metabarcoding (or any other methods) is based on a set of data (here considered a local library) and operation taxonomic units (OTUs) instead of species names are used. This means that a universal reference sequence library is not necessary for forensics. As exemplified in this study, results based on the total OTUs came to the same conclusion as that based on the annotated OTUs. However, if the OTUs were annotated to species, genera or families, extra information such as morphological characters could be used and the evidences would be more solid.

To annotate the OTUs, a well-curated sequence reference library is indispensable. The reference library helps to exclude data of experimental artifacts (such as chimeras) and non-target species. Although some efforts have been made, the DNA barcode reference library is still far from being satisfactory due to low species coverage (Lou et al., 2010; Ratnasingham & Hebert, 2007; Tnah et al., 2019), especially for lower plants. For example, there are about 8397 known species worldwide and only 889 species have their *rbcL* sequences deposited in GenBank (4116 accessions. accessed on Jan. 9, 2021,*rbcL* in plastid genomes were not considered.).

Data analysis methods

DNA metabarcoding is NGS platform-based. Correct extraction of sequences is crucial for successful source tracking. There are several NGS data process pipelines (such as OTU, DADA2, COTU and etc.) and each of them has its own advantages and disadvantages. OTU pipeline, the earliest one, groups reads at a certain similarity (usually 0.97) and creates OTUs using very short computing time. DADA2 and Unoise3 do not adopt a subjective similarity value. COTU method, a recently proposed strategy, updates the OTU method by elongating the consensus sequence to be created (Liu et al. 2021) at the cost of computing time. Although there are some comparative studies on the pipelines (Prodan et al., 2020; Xiong & Zhan, 2018), it is still too early to say which one is the best.

The other important issue concerning data analysis is how soil samples can be reliably tracked back to the original place based on OTUs and their abundances. SourceTracker (Knights et al., 2011) and FEAST (Shenhav et al., 2019) are two most popular software packages for allocating components in a microorganism community to potential sources and the latter was claimed to be quicker and more accurate.

In this study, we tested source tracking accuracies of four kinds of data sets (four combinations between inclusion/exclusion of singletons and total/annotated OTUs) using FEAST. The results are nearly the same, indicating the high power of FEAST and reliability of the conclusion.

Conclusion

By using these results, the police successfully unmask the lie that the suspect has never been to the crime scene. Using this evidence as a breakthrough, the suspect finally made a clean breast of his crime. This case implies that plant DNA in the environment soil is a new source of evidence in determination of suspects using DNA metabarcoding technology and has high potentials of extensive applications in criminal cases.

Acknowledgement

This study was partly supported by the funds from the National Key R&D Program of China (2017YFC0803803), the open project of Institute of Forensic Science, Ministry of Public Security (2018FGKFKT04), Strategic Priority Research Program of the Chinese Academy of Sciences (XDA23080204, XDA19050303), National Natural Science Foundation of China (NSFC31872679) and the fundamental research fund of the Central Public Service Research Institute (2018JB001).

Authors' contributions

Shiliang Zhou and Xueying Yang designed this study; Yanlei Liu and Chao Xu designed and completed the experiments; Yanlei Liu analyzed the data; Yanlei Liu, Wenpan Dong and Chao Xu drafted the manuscript, Shiliang Zhou revised the manuscript; all authors have read and approved the final manuscript.

Data Accessibility

Twelve demultiplexed NGS original data from Ion Torrent S5 have been submitted to NCBI and the accession number is SRR13203136.

Tables

Table 1. Soil samples collected from suspect's pants, crime scene and nearby areas.

Table 2. Primers for amplifying diatom *rbcL* fragments together with oligoes attached to the 5' ends of the primers for labeling 12 soil samples.

Figures

Fig. 1. Localities of soil samples collected as references for tracking the possible source of the mud on the suspect's pants and pictures of crime scene, mud from the canal, and suspect's pants. A: Google map of five sampling sites; B: close-up of Site 1 to 3; C: the crime scene, D: soil from the canal; and E: suspect's pants.

Fig. 2. Diagram showing how *rbcL* fragments were labeled by PCR, pooled together and sequenced on Ion Torrent S5xl platform.

Fig. 3. Percentages of the ZOTUs in the mud on suspect's pants tracked to the 11 candidate source sites by FEAST. A: total ZOTUs with singletons; B: total ZOTUs without singletons; C: annotated ZOTUs with singletons; and D: annotated ZOTUs without singletons.

Fig. S1. Experiment design. Five parts, soil DNA acquisition, amplicon preparation, NGS data sequencing, NGS data processing, and suspect's mud tracking.

Fig. S2. The ZOTU-rich diatom genera (A) and species (B) found in 12 soil samples.

Refenence

Altschul, S. F. (2012). Basic local alignment search tool (BLAST).*Journal of Molecular Biology, 215* (3), 403-410.

Ambers, A., Votrubova, J., Vanek, D., Sajantila, A., & Budowle, B. (2018). Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains.*International Journal of Legal Medicine, 132* (6), 1545-1553.

Bertoglio, B., Grignani, P., Di Simone, P., Polizzi, N., De Angelis, D., Cattaneo, C., . . . Previdere, C. (2020). Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck.*Forensic Science International-Genetics, 44* .

Dong, W. P., Cheng, T., Li, C. H., Xu, C., Long, P., Chen, C. M., & Zhou, S. L. (2014). Discriminating plants using the DNA barcode*rbcL* b: an appraisal based on a large data set. *Molecular Ecology Resources, 14* (2), 336-343.

Dong, W. P., Xu, C., Li, C. H., Sun, J. H., Zuo, Y. J., Shi, S., . . . Zhou, S. L. (2015). *ycf* 1, the most promising plastid DNA barcode of land plants. *Scientific Reports, 5* (1), 8348.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods, 10* (10), 996.

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* , 081257.

Fang, T., Liao, S. P., Chen, X. G., Zhao, Y. C., Zhu, Q., Cao, Y. Y., . . . Zhang, J. (2019). Forensic drowning site inference employing mixed pyrosequencing profile of DNA barcode gene (*rbcL*).*International Journal of Legal Medicine, 133* (5), 1351-1360.

Ferri, G., Alu, M., Corradini, B., & Beduschi, G. (2009). Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. *International Journal of Legal Medicine, 123* (5), 395-401.

Habibi, S., Ahmadi, A., Behmanesh, M., Miri, A., & Tavallaie, M. (2019). Evaluation of ten SNP Markers for Human Identification and Paternity Analysis in Persian Population. *Iranian Journal of Biotechnology, 17* (3), 68-71.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences, 270* (1512), 313-321.

Hobern, D., & Hebert, P. D. N. (2019). BIOSCAN - Revealing Eukaryote Diversity, Dynamics, and Interactions. *Biodiversity Information Science and Standards, 3* .

Irinyi, L., Lackner, M., de Hoog, G. S., & Meyer, W. (2016). DNA barcoding of fungi causing infections in humans and animals.*Fungal Biology, 120* (2), 125-136.

Kakizaki, E., Sonoda, A., Sakai, M., & Yukawa, N. (2018). Simple detection of bacterioplankton using a loop-mediated isothermal amplification (LAMP) assay: First practical approach to 72 cases of suspected drowning. *Forensic Science International, 289* , 289-303.

Kress, W. J., & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding*trnH-psbA* Spacer Region. *Plos One, 2* (6).

Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., Liu, J. Q., . . . Grp, C. P. B. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America, 108* (49), 19641-19646.

Li, J. L., Wang, S. S., Yu, J., Wang, L., & Zhou, S. L. (2013). A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany* , 48(1), 72-78.

Liu, M. Y., Liu, Y. L., Wu, P., Chen, Q., & Zhou, S. L. (2019). Determination of Place of Residence Using the Gene Information of Plants Carried by the Human Body. *Journal of Forensic Medicine, 35* (6), 710-715.

Liu, M. Y., Zhao, Y., Sun, Y. Z., Li, Y. N., Wu, P., Zhou, S. L., & Ren, L. (2020). Comparative study on diatom morphology and molecular identification in drowning cases. *Forensic Science International, 317* , 110552.

Liu, M. Y., Zhao, Y., Sun, Y. Z., Wu, P., Zhou, S. L., & Ren, L. (2020). Diatom DNA barcodes for forensic discrimination of drowning incidents. *Fems Microbiology Letters, 367* (17).

Liu, Y. L., Xu, C., Sun, Y. Z., Wu, P., Dong, W. P., Yang, X. Y., & Zhou, S. L. (2021). Method for quick DNA barcode reference library construction. *Journal of Systematics and Evolution* , underview.

Lou, S. K., Wong, K. L., Li, M., But, P. P. H., Tsui, S. K. W., & Shaw, P. C. (2010). An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). *Bmc Genomics, 11* .

Lygo, J. E., Johnson, P. E., Holdaway, D. J., Woodroffe, S., Whitaker, J. P., Clayton, T. M., . . . Gill, P. (1994). The Validation of Short Tandem Repeat (Str) Loci for Use in Forensic Casework.*International Journal of Legal Medicine, 107* (2), 77-89.

Meng, H. T., Guo, Y. X., Jin, X. Y., Chen, C., Cui, W., Shi, J. F., . . . Zhu, B. F. (2019). Internal validation study of a newly developed 24-plex Y-STRs genotyping system for forensic application.*International Journal of Legal Medicine, 133* (3), 733-743.

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., & Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *Plos One, 15* (1).

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (*www.barcodinglife.org*).*Molecular Ecology Notes, 7* (3), 355-364.

Ravi, K., Patel, Mukesh, & Jain. (2012). NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *Plos One* .

Rousseau, D. D., Schevin, P., Duzer, D., Cambon, G., Ferrier, J., Jolly, D., & Poulsen, U. (2006). New evidence of long distance pollen transport to southern Greenland in late spring. *Review of Palaeobotany and Palynology, 141* (3-4), 277-286.

Shenhav, L., Thompson, M., Joseph, T. A., Briscoe, L., Furman, O., Bogumil, D., . . . Halperin, E. (2019). FEAST: fast expectation-maximization for microbial source tracking. *Nature Methods, 16* (7), 627.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation, 183* , 4-18.

Thorne, R. F. (2002). How many species of seed plants are there?*Taxon, 51* (3), 511-512.

Tnah, L. H., Lee, S. L., Tan, A. L., Lee, C. T., Ng, K. K. S., Ng, C. H., & Nurul Farhanah, Z. (2019). DNA barcode database of common herbal plants in the tropics: a resource for herbal product authentication.*Food Control, 95* , 318-326.

Xiong, W., & Zhan, A. B. (2018). Testing clustering strategies for metabarcoding-based investigation of community–environment interactions. *Molecular Ecology Resources, 18* (6), 1326-1338.

Xu, C., Dong, W. P., Shi, S., Cheng, T., Li, C., Liu, Y. L., . . . Zhou, S. L. (2015). Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. *Molecular Ecology Resources, 15* (6), 1366-1374.

Yang, X. Y., Song, B. K., Pei, L., & Song, J. Y. (2015). DNA barcoding analysis of plant evidence. *Chinese Journal of Forensic Medicine, 30* (2), 189-190.

Young, J. M., Austin, J. J., & Weyrich, L. S. (2017). Soil DNA metabarcoding and high-throughput sequencing as a forensic tool: considerations, potential limitations and recommendations. *Fems Microbiology Ecology, 93* (2).

Yu, J., Xue, J. H., & Zhou, S. L. (2011). New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution, 49* (3), 176-181.

[dataset] Liu, Y. L; 2020; Determination of a criminal suspect using environmental plant DNA metabarcoding technology; NCBI; SRR13203136.

Hosted file

`Table 1.docx` available at https://authorea.com/users/308400/articles/562785-determination-of-a-criminal-suspect-using-environmental-plant-dna-metabarcoding-technology

Hosted file

`Table 2.docx` available at https://authorea.com/users/308400/articles/562785-determination-of-a-criminal-suspect-using-environmental-plant-dna-metabarcoding-technology

Hosted file

`Fig 1 Localities of soil samples.pdf` available at https://authorea.com/users/308400/articles/562785-determination-of-a-criminal-suspect-using-environmental-plant-dna-metabarcoding-technology

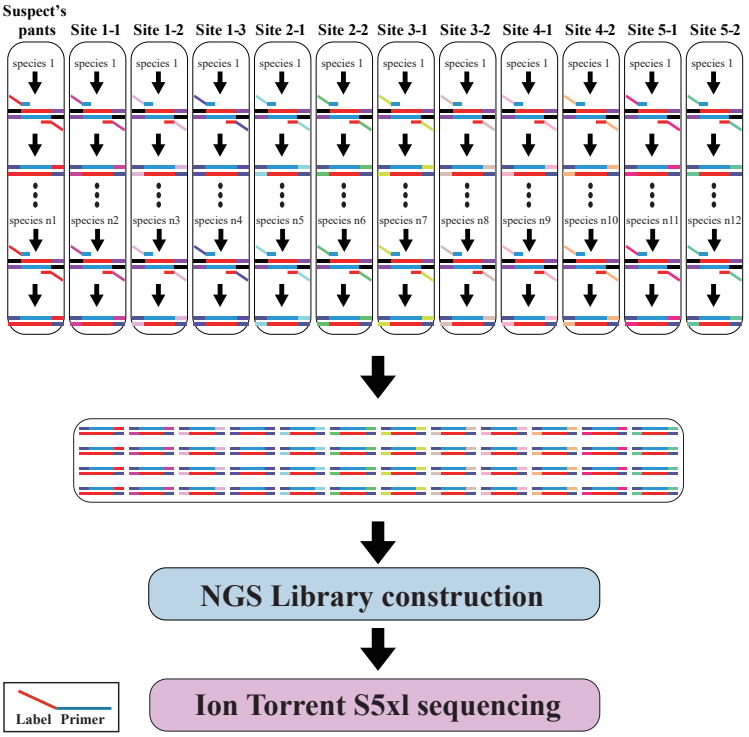

Fig. 2. Diagram showing how rbcL fragments were labeled by PCR, pooled together and sequenced on Ion Torrent S5xl platform.

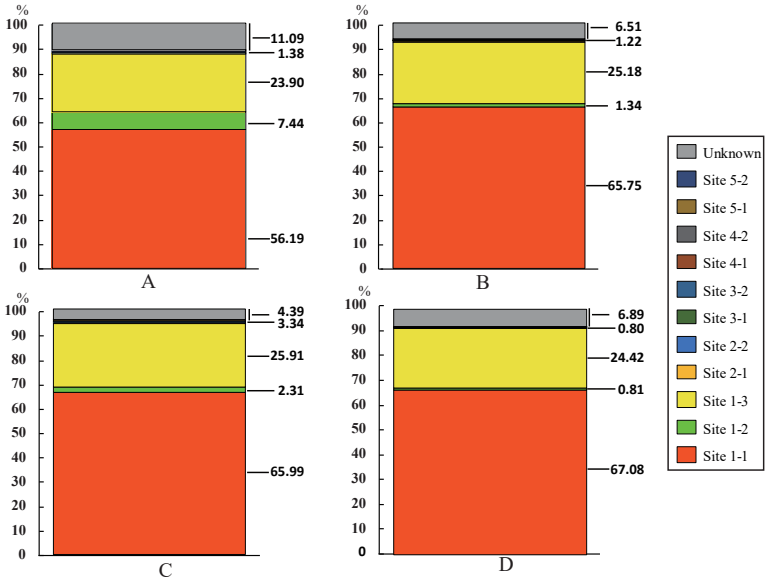

Fig. 3. Percentages of the ZOTUs in the mud on suspect's pants tracked to the 11 candidate source sites by FEAST. A: total ZOTUs with singletons; B: total ZOTUs without singletons; C: annotated ZOTUs with singletons; and D: annotated ZOTUs without single-

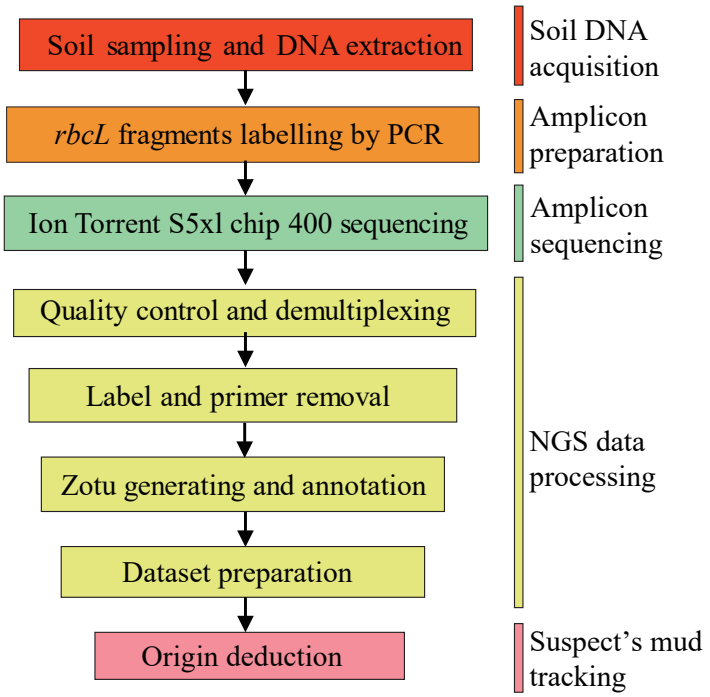

Fig. S1. Experiment design. Five parts, soil DNA acquisition, amplicon preparation, NGS data sequencing, NGS data processing, and suspect's mud tracking.

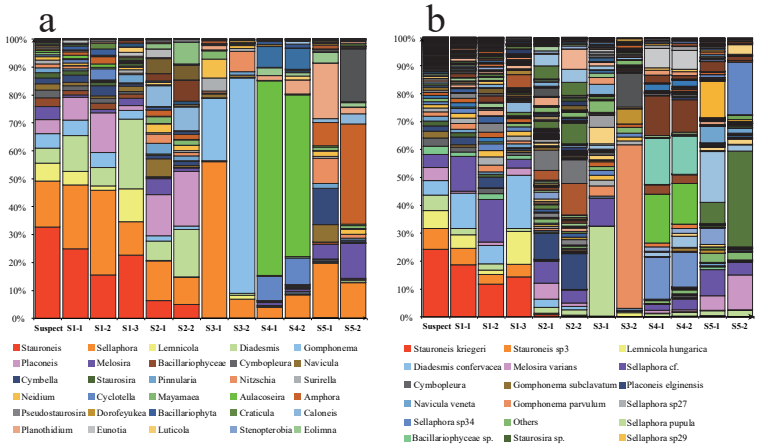

Fig. S2. The ZOTU-rich diatom genera (A) and species (B) found in 12 soil samples.