Capturing airborne eDNA in an old growth forest fragment in New York City

Authors: Kailani Gaynor[1], Alexa Leiva[1], Stephanie Marino[1], and Zachary Vetsch[1]
Mentors: Allison Granberry[1] and Damon P. Little[2]
[1]Hostos Lincoln Academy; [2]New York Botanical Garden

Abstract

Environmental DNA, or eDNA, is DNA that is shed in air, water, or soil by organisms. Sampling eDNA from a forest is useful because eDNA potentially samples an entire population of plants without directly collecting from each individual. The forest at the New York Botanical Garden is home to 261 extant native species and 229 non–native species. We tested the viability of sequencing plant eDNA from air samples in the forest. Air samples were collected in winter. 4 of the 8 air filters produced sequenceable *rbcL* D mini–barcode amplicons. 32 sequence types were detected from these 4 sites. Three sequence types accounted for more than 17% of the sequencing reads. Microscopic examination of an air filter revealed cuticle, epicuticular wax, and trichomes, but no pollen or spores. This protocol could be used to sample vegetation when plants are physically inaccessible or when plants are sterile/dormant.

Introduction

Environmental DNA, or eDNA, is DNA that is shed in air, water, or soil by organisms. It can be used to detect the presence of a species even if the individual that shed the eDNA is not directly observed. When shed, eDNA can be encapsulated inside a cell, an organelle, or it can be free. Sampling eDNA from a forest is useful because eDNA potentially samples an entire population without directly collecting DNA from each individual or damaging the forest.

Airborne terrestrial plant eDNA has been collected using various types of dust traps[1,2]. In these studies, two genera were assayed using targeted quantitative Polymerase Chain Reaction (qPCR). Clare et al.[3] demonstrated that air, sampled from a confined space using a peristaltic pump and a Sterivex-HV filter, is a viable source of eDNA for the identification of animals.

The 20 hectare primary forest at the New York Botanical Garden (NYBG) is home to a diverse flora[4]. The forest has 261 extant native species and 229 non–native species.

We: (1) determined the viability of a plant airborne eDNA assay and **(2)** compared the number of species found via eDNA to the list of known species in the NYBG forest.

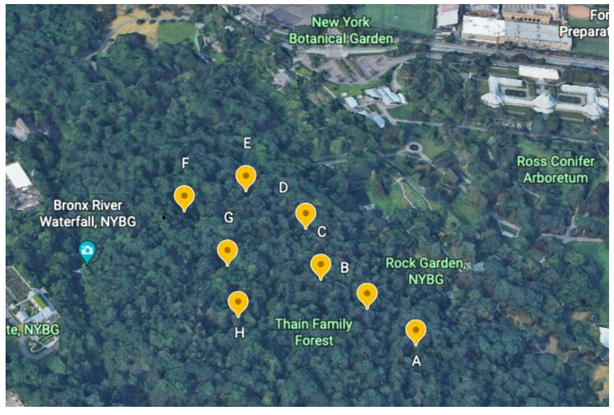

Fig. 1: Sampling sites within the NYBG forest.

Material & Methods

Sampling—Eight sites along a trail that bisects the NYBG forest were selected (Fig. 1). At each site air was sampled for 15 minutes at 229 mL/s using a sterile Sterivex 0.22 µm filter (SVGV010RS) and a peristaltic pump (Solinst 410).

eDNA isolation—DNA was extracted from filter membranes by proteinase K digestion and silica mini spin column (EconoSpin) isolation and purification.

eDNA sequencing—Mini–barcode *rbcL* D was amplified[5]. PCR products were combined by site, purified with the Qiagen PCR purification kit, quantified (Agilent DNA 1000 kit), and sequenced using Amplicon EZ (150–paired end).

Reference database—A *rbcL* barcode reference database for vascular plants reported from the NYBG forest[4] was constructed from GenBank and BOLD: stop–codon free *rbcL* sequences were downloaded and aligned using MAFFT[6].

Sequence analysis—Low–quality reads were removed and paired reads merged using fastp[7], primer sequences were removed with pTrimmer[8], low–quality assemblies were filtered with fastp[7], and BRONX[9] was used to compare assembled sequences to the reference database.

Scanning Electron Microscopy (SEM)—A filter membrane was cut into 1 cm^2 fragments, mounted on aluminum stubs, and sputter coated with gold/palladium (DeskV HP). Samples were examined with a Hitachi SU3500 in high–vacuum mode at 5–10 kV using the secondary electron detector.

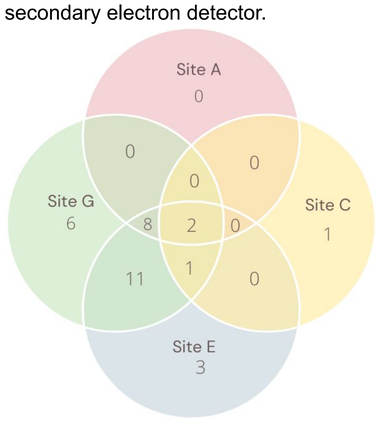

Fig. 2: Venn diagram of eDNA sequence types by sampling site.

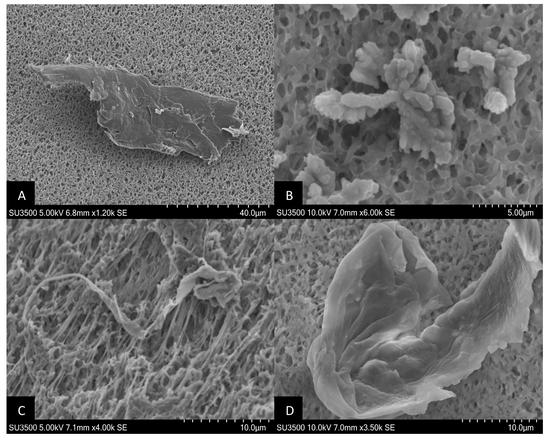

Fig. 3. Scanning Electron Microscope (SEM) observations of plant fragments on filter membrane from sampling site H (Fig. 3). **A:** Plant epidermis with wax; **B:** Wax crystals; **C:** Plant trichome; **D:** Plant cuticle with wax.

Results

Table 1. eDNA sequences detected in the NYBG forest.

Sequence type	Site(s)	Reads (% of total)	Reference identification(s)
1	A, C, E, G	12.1536%	*Datura stramonium*
2	A, E, G	2.2422%	*Arisaema triphyllum, Hypochaeris radicata*
3	A, C, E, G	0.5527%	*Arctium minus*
4	A, E, G	0.6234%	*Dichanthelium clandestinum, Dichanthelium depauperatum, Digitaria sanguinalis, Echinochloa crus, Echinochloa walteri, Panicum dichotomiflorum, Setaria faberi, Setaria pumila, Setaria viridis*
5	A, E, G	0.0150%	*Arisaema triphyllum*
6	A, E, G	0.0951%	*Digitaria sanguinalis*
7	A, E, G	3.1351%	*Carya glabra, Fraxinus americana, Monotropa uniflora, Oxalis stricta, Quercus alba, Quercus bicolor, Quercus macrocarpa, Quercus palustris, Quercus rubra, Quercus velutina*
8	A, E, G	0.0738%	*Eleusine indica, Oxalis corniculata*
9	A, E, G	0.0013%	*Ginkgo biloba*
10	A, E, G	0.0096%	*Juniperus virginiana*
11	C, E, G	0.8740%	*Platanus occidentalis*
12	C	0.0042%	*Dennstaedtia punctilobula*
13	E, G	0.9059%	*Carya alba, Carya cordiformis, Carya glabra, Carya ovata, Oxalis stricta, Plantago major*
14	E, G	0.1897%	*Liquidambar styraciflua*
15	E	0.0195%	*Carya cordiformis*
16	E, G	0.0046%	*Euonymus alatus*
17	E, G	0.0057%	*Quercus rubra*
18	E	0.0008%	*Populus deltoides*
19	E, G	0.0079%	*Salix discolor*
20	E, G	0.1167%	*Allium vineale*
21	E, G	0.0209%	*Cleome houtteana*
22	E, G	0.0415%	*Peltandra virginica*
23	E, G	0.2076%	*Aesculus sylvatica, Alliaria petiolata, Amelanchier arborea, Barbarea vulgaris, Capsella bursa, Cardamine concatenata, Cardamine diphylla, Cardamine flexuosa, Cardamine hirsuta, Diplotaxis tenuifolia, Draba verna, Ipomoea purpurea, Lepidium didymum, Lepidium virginicum, Solanum carolinense*
24	E, G	0.0069%	*Potamogeton pusillus*
25	E	0.0003%	*Carya alba, Carya cordiformis, Carya glabra, Carya ovata, Decodon verticillatus, Oxalis stricta, Plantago major*
26	E, G	0.0016%	*Arisaema triphyllum, Hypochaeris radicata, Peltandra virginica*
27	G	0.0097%	*Arabidopsis thaliana*
28	G	0.0010%	*Aesculus sylvatica, Alliaria petiolata, Amelanchier arborea, Barbarea vulgaris, Brassica rapa, Capsella bursa, Cardamine concatenata, Cardamine diphylla, Cardamine flexuosa, Cardamine hirsuta, Diplotaxis tenuifolia, Draba verna, Ipomoea purpurea, Lepidium didymum, Lepidium virginicum, Solanum carolinense*
29	G	0.0009%	*Cardamine flexuosa, Cardamine pratensis*
30	G	0.0006%	*Cleome houtteana, Sisymbrium officinale*
31	G	0.0036%	*Allium vineale, Galanthus nivalis*
32	G	0.0004%	*Aesculus sylvatica, Alliaria petiolata, Amelanchier arborea, Barbarea vulgaris, Capsella bursa, Cardamine concatenata, Cardamine diphylla, Cardamine flexuosa, Cardamine hirsuta, Diplotaxis tenuifolia, Draba verna, Ipomoea purpurea, Lepidium didymum, Lepidium virginicum, Rorippa indica, Solanum carolinense*

Reference database—A database of 8,333 *rbcL* D sequences from 484 species was constructed. Sequences of six species reported from the NYBG forest[4] were unavailable.

PCR and eDNA sequencing—Amplicons of *rbcL* D were detected for 4 of 7 sampling sites. 4,045,172 paired–end reads were produced; 1,227,139 reads passed quality filtering. 32 distinct sequence types with 10 or more high–quality reads were detected (Table 1; Fig. 2).

SEM—Plant fragments observed on filter membrane H included cuticle, epicuticular wax, and trichomes (Fig. 3). No plant or fungal pollen/spores were observed.

Discussion

The short (140 bp) and highly–conserved nature of *rbcL* D makes species–level identification within the NYBG forest difficult: 13 of the 32 sequence types were assigned to more than one species (Table 1). Future studies may consider using a more variable, universal marker.

The three most frequent sequence types (Table 1) are weedy herbaceous species (1, 2) and dominant tree species (mostly, 7). Together these represent more than 17% of the high–quality reads. The remaining 29 sequence types were detected at very low levels and are primarily composed of common weedy species. Sequence type 24 (*Potamogeton pusillus*) is an aquatic plant found in the Bronx River. Its DNA was detected at two sites (interior to the forest and near the river)—perhaps in the form of dispersed pollen transported via water and wind.

Two sequence types (1, 3) were common to all four sites (Fig. 2). Sites E and G share the most sequence types (11); they also have the greatest total number of sequence types—perhaps due to the constant wind from the Bronx River.

No pollen or spores were found in the SEM examination of the filter from site H (Fig. 3). Previous plant eDNA studies[1,2] presumed, but did not conclusively demonstrate, that pollen was the major source of eDNA—that assumption does not hold for these samples.

This protocol, with additional sampling sites, could be used to quickly produce vegetation inventories during seasons (e.g. winter) that traditional protocols cannot because plants are sterile or dormant; or in cases when collecting from plants is physically challenging.

References

1. Johnson, M. D., Cox, R. D. & Barnes, M. A. Analyzing airborne environmental DNA: A comparison of extraction methods, primer type, and trap type on the ability to detect airborne eDNA from terrestrial plant communities. *Environ. DNA* **1**, 176–185 (2019).
2. Johnson, M. D., Cox, R. D. & Barnes, M. A. The detection of a non-anemophilous plant species using airborne eDNA. *PLOS ONE* **14**, e0225262 (2019).
3. Clare, E. L. *et al.* eDNAir: proof of concept that animal DNA can be collected from air sampling. *PeerJ* **9**, e11030 (2021).
4. Atha, D. E. *et al.* The historic and extant spontaneous vascular flora of The New York Botanical Garden. *Brittonia* **68**, 245–277 (2016).
5. Little, D. P. A DNA mini-barcode for land plants. *Mol. Ecol. Resour.* **14**, 437–446 (2014).
6. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
7. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
8. Zhang, X. *et al.* pTrimmer: An efficient tool to trim primers of multiplex deep sequencing data. *BMC Bioinformatics* **20**, 236 (2019).
9. Little, D. P. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS ONE* **6**, e20552 (2011).

Acknowledgements

We thank Eliot Nagele for allowing us to collect samples in the NYBG forest; Tynisha Smalls and Adelin Figueroa for expediting our sequencing order; and Barbara Ambrose for assistance with NYBG bureaucracy.