Effect of Specimen Age on DNA Barcoding Success

Samiha Ahmed[1], Nana Opare-Addo[2] & Lydia Paradiso[3,4]
[1]Manhattan Center for Science and Mathematics, [2]Brooklyn Latin School, [3]New York Botanical Garden,[4]CUNY Graduate Center

Abstract

Plant DNA is well preserved in herbarium specimen samples, making them suitable for molecular and phylogenetic studies. However, ancient plant DNA poses difficulties due to its fragmented nature; this is a result of the degradation that DNA undergoes and the subsequent damage from chemical and physical factors. With this in mind, we investigated how the age of herbarium specimens would affect the success rate of DNA Barcoding, using specimens of different ages from 7 different species. We were able to extract detectable amounts of DNA from all 42 specimens. We then amplified two DNA barcoding regions (*rbcL* and *matK*), using the polymerase chain reaction (PCR), and sequenced the samples that were successful. We obtained at least one of the barcoding sequences for 22 specimens, including all of the specimens less than 20 years old, as well as a few of the specimens collected over 100 years ago. We observed that although older specimens were generally less successful than newer ones, other factors—in addition to age—may affect sequencing success.

Introduction

Historically, botanists have aimed to correctly identify known species and discover specimens yet to be found in nature. These practices provide both biologists and the general public the tools to analyze endangered species, invasive species, and most importantly biological diversity. For a long time, identification of organisms was solely based on morphological characteristics such as size, shape, color, and structure of plant parts. However, there are various aspects that make this method of classification unreliable, including ambiguous structure, convergent evolution, or poorly preserved specimens.

Fortunately, advancements in genetics—particularly DNA Barcoding—has made it easier to accurately identify specimens (Hebert et al. 2003). The genes considered the "official" DNA barcode regions for plants are two chloroplast genes, *rbcL* and *matK* (CBOL et al., 2009). However, we must also consider how age may hinder the correct identification of a specimen. Ancient plant DNA undergoes severe degradation, resulting in subsequent damage from chemical and physical factors (Alaeddini et al. 2010). Outside of DNA degradation, poor treatment and lengthy preservation of specimens in jars may also hinder PCR amplification (Knölke et al. 2004). Significant age has typically impacted barcoding success, with the greatest decline being most noticeable in samples over 100 years old (Korpelainen et al. 2019).

With this in mind, we extracted and barcoded DNA from herbarium specimens of various plant species in order to study how DNA extraction and sequencing success may differ based on age; before testing, we hypothesized that as the age of the specimen increased, the DNA barcoding success rate would decrease. Additionally, we examined other specific factors that correlated with successful sequencing (e.g. plant family, physical state of the specimen).

Methods

Step 1: Specimen Selection

We defined 3 age categories:

- Specimens collected before 1920
- Specimens collected between 1920 and 2000
- Specimens collected after 2000

For 7 species which occur in New York City, we selected two specimens each from the 3 age categories. Specimens were selected from the herbarium at the New York Botanical Garden.


Acer saccharum (Sugar Maple)


Apios americana (American Groundnut)


Baptisia tinctoria (Wild Indigo)


Juniperus virginiana (Eastern Red Cedar)

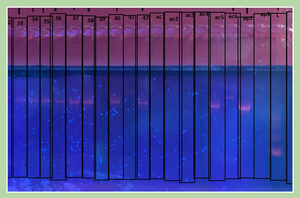

Osmunda regalis (Royal Fern)


Panicum virgatum (Switchgrass)


Pinus strobus (White Pine)

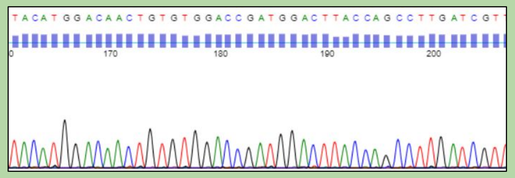
Step 2: Lab Work

- DNA extraction
- PCR to amplify *rbcL* and *matK* barcodes
- Confirm success with gel electrophoresis
- Sent out for sequencing

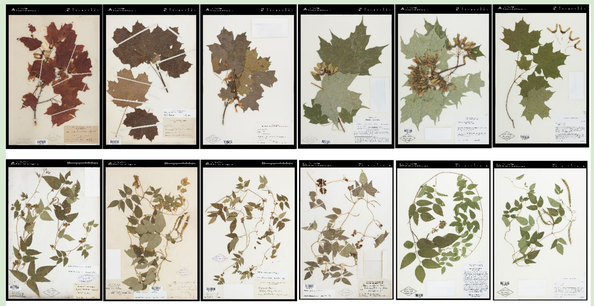


Step 3: Data Analysis

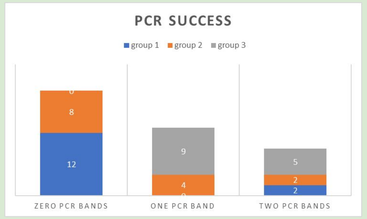
- Analyzed results using DNA Subway
- Compared success for PCR and sequencing based on age group of the specimens



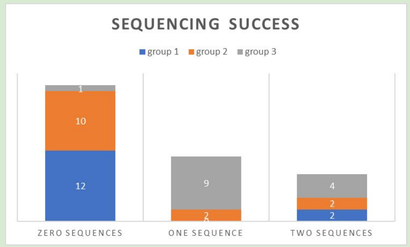
Results



DNA extraction was successful for all 42 specimens, with quantification of the DNA showing some amount of detectable DNA for all of the samples we extracted. The DNA concentration ranged from 1.2 ng/uL to 53.6 ng/uL.



PCR success: For the oldest age group, we saw PCR bands for both *rbcL* and *matK* for both specimens of *Apios americana*. For the middle age group, we saw PCR bands for specimens of three species, *Apios americana*, *Baptisia tinctoria*, and *Panicum virgatum*. For the newest age group, we saw PCR bands for every specimen: 4 for rbcL only and 5 for both genes. We sent a total of 27 PCR products for sequencing.



Sequencing success: For the oldest age group, we obtained high quality sequences for both specimens of *Apios americana* for both *rbcL* and *matK*. For the middle age group, we obtained sequences for one *Apios americana* (both genes), one *Baptisia tinctoria* (both genes), one *Baptisia tinctoria* (*rbcL* only), and one *Panicum virgatum* (*rbcL* only). The sequences received for both genes on the other *Apios americana*, as well as for *matK* for the other *Baptisia tinctoria*, were of too low quality. For the newest age group, we obtained *rbcL* sequences for all specimens except one *Acer saccharum* specimen. *matK* sequences were obtained for 4 samples: both *Apios americana* and both *Baptisia tinctoria*.

Four specimens which produced PCR bands (both *Juniperus communis* and both *Pinus strobus*) did not produce DNA sequences of high enough quality.

Discussion and Conclusions

Our hypothesis was that the age of the specimen would have an effect on PCR and sequencing success. We did see that older specimens had lower success rates in both PCR amplification and sequencing compared to more recent ones. However, for three species, we saw some success in both PCR and sequencing, *Apios americana*, *Baptisia tinctoria*, and *Panicum virgatum*.

It was interesting that, both American groundnut and wild indigo are from the same family, Fabaceae (the pea family). It is possible that in addition to the effect of age, there may be other important factors such as taxonomic placement that affect sequencing success. A potential explanation for the limited PCR and sequencing success observed for *matK* might be the use of the same primers for all sequences. Universal *matK* primers are known to underperform for certain groups, such as ferns, which were represented in our sample set. Given additional time, we could have further experimented with different primers or other PCR settings to enhance the success rate for *matK* amplification and sequencing.

It is an exciting result that we were able to obtain DNA sequences from two samples collected in 1874 and 1890! This demonstrates the viability of obtaining valuable genetic information from historical samples. Overall, our results do show a negative correlation between the age of the specimen and PCR and sequencing success. However, the success of a few species that are from an older age group allude to additional factors affecting DNA barcoding other than age. Future research can explore these factors, which may include preservation methods, taxonomic group, or molecular techniques used.

Works Cited

-Alaeddini R, et al. Forensic implications of genetic analyses from degraded DNA--a review. For. Sci Intl Genet. 2010;4:148–157..
-CBOL Plant Working Group 1, Hollingsworth, et al. PNAS, 106(31):12794-12797.
-Hebert PDN, et al. (2003) Biological identification through DNA barcodes. Proc Roy Soc B Biol Sci 270:313–321.
-Knölke, S., et al. (2004). A procedure for combined genitalia dissection and DNA extraction in Lepidoptera, Insect Systematics & Evolution, 35(4), 401-409.
-Korpelainen, H., et al. The effects of sample age and taxonomic origin on the success rate of DNA barcoding when using herbarium material. Plant Syst Evol 305, 319–324 (2019).

Acknowledgments

We want to thank the DNA Learning Center, in particular Dr. Alison Mayle and Arden Feil, for providing the materials we needed for this study. We also thank the New York Botanical Garden for use of the herbarium collections and the laboratory space. This work was carried out as part of Cold Spring Harbor's Urban Barcode Research Program (UBRP), which is funded by The Pinkerton Foundation and Science Sandbox: An Initiative of the Simons Foundation.