Nanopore Sequencing as a Supplementary Tool for DNA Barcode Identification of Plant Samples with Failed Sanger Sequencing: A Case Study of Wild Plants from Bailu Garden, Dushu Lake

Authors: Sean Yang (杨树)[1]
Mentor: David A. Micklos[1], Xinyue Wang[2]

[1] Cold Spring Harbor Asia DNA Learning Center, Suzhou

Abstract

DNA barcoding enables reliable plant species identification via standardized genetic markers. Field plant samples frequently fail PCR and Sanger sequencing after rapid chromatography paper-based DNA extraction, caused by insufficient DNA yield and incomplete tissue lysis. This study tested whether silica-based re-extraction combined with Nanopore sequencing could rescue failed samples. Eight wild plant specimens were collected from Bailu Garden, Dushu Lake. Four samples obtained valid Sanger sequences with ≥98% BLAST similarity. The other four failed samples underwent silica re-extraction and Nanopore sequencing; only one achieved reliable identification, representing a 25% recovery rate. Silica purification can rescue a subset of difficult samples, yet efficiency remains limited. Optimizing extraction and tissue lysis is vital to boost DNA barcoding success.

Keywords: DNA barcoding; plant identification; rapid DNA extraction; silica-based re-extraction; Nanopore sequencing; Sanger sequencing

Introduction

DNA barcoding employs standardized DNA markers for species identification. Sanger sequencing yields accurate barcode data, yet success relies heavily on DNA template quality. The chromatography paper-based rapid extraction method offers convenient DNA isolation, though many field plant samples still fail PCR amplification. Silica-based purification may enhance DNA quality and rescue unsuccessful samples. Nanopore sequencing supports amplicon analysis but suffers from higher raw error rates, limiting its performance on short plant barcodes. This study explores whether silica re-extraction can recover barcode data from rapid-extraction failures. We hypothesize this method improves template quality and retrieves usable sequences for part of failed samples.

Materials & Methods

1. Eight fresh wild plant leaf samples were collected from Bailu Garden, Dushu Lake (Figure 1) and stored at −20 °C to prevent DNA degradation. Post-extraction PCR screening separated samples into Group 1 (4 samples with visible target bands) and Group 2 (4 repeatedly PCR-negative samples).
2. Two DNA extraction methods were used. All samples initially adopted Cold Spring Harbor's rapid filter-paper protocol. Group 2 samples underwent silica-column re-extraction with liquid nitrogen grinding, thermal lysis, repeated washing and heated elution. DNA concentration and purity were measured by spectrophotometry.
3. Universal plant barcode primers were used in 25 μL PCR reactions. Thermal cycling included 94 °C initial denaturation, 35 amplification cycles and a final extension. Amplicons were detected using 2% agarose gel electrophoresis (Figure 2).
4. Purified PCR products of Group 1 received bidirectional Sanger sequencing. Raw chromatograms were checked manually to assemble consensus barcode sequences.
5. Re-extracted DNA of Group 2 was amplified, built into libraries and sequenced on a portable Nanopore device. Raw reads were quality-filtered, and target-length sequences were reserved (Figure 3). Analyses were completed on DNA Subway.
6. Qualified sequences were annotated via BLAST (Altschul et al. 1990). Top-hit similarity ≥98% represented successful species identification for both sequencing methods. MUSCLE alignment was performed on all five successfully recovered barcode sequences.

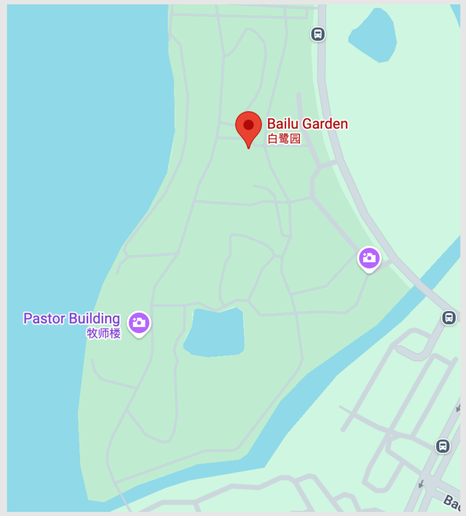

Figure 1 Google Map of Dushu Lake Bailue Garden

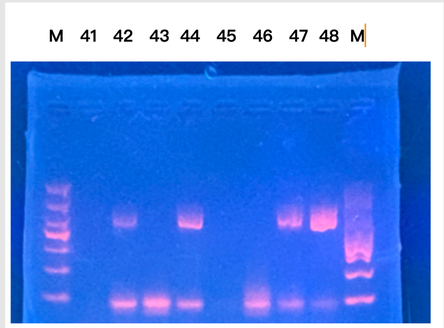

Figure 2 RBCL PCR amplification result

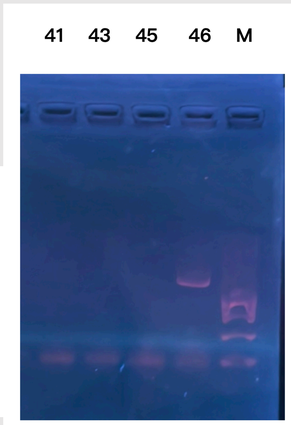

Figure 3 Repeated dna isolation with silica method

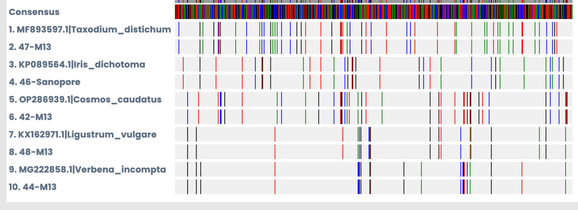

Figure 4. MUSCLE multiple sequence alignment of the five successfully identified plant DNA barcode sequences.

Results

1. Eight wild plant samples from Bailu Garden were processed with rapid DNA extraction and PCR amplification. Gel electrophoresis results (Figure 2) divided samples into two groups. Half of the samples yielded clear target bands, while the other half showed no visible amplicons after repeated PCR trials, indicating extraction and amplification failure.
Sanger Sequencing Performance

2. All four PCR-positive samples obtained high-quality consensus sequences via bidirectional Sanger sequencing. BLAST analysis on DNA Subway confirmed 100% identification success, with every sequence reaching ≥98% similarity and achieving accurate species assignment (Altschul et al. 1990).
Nanopore Sequencing Recovery Outcomes

3. The four PCR-failed samples were re-extracted using a silica-based method and analyzed by Nanopore sequencing (Figure 3). Only one sample met the 98% similarity threshold for valid identification. The other three contained severe base errors and indels, producing unreliable, biologically inconsistent BLAST results.

4. Overall Identification Success Rate
Sanger sequencing group: 4/4 samples successfully identified
Silica-recovered Nanopore group: 1/4 samples successfully identified
Total successful barcode identification rate across all samples: 5/8 (62.5%)

Discussion

This study hypothesized that silica-based re-extraction combined with Nanopore sequencing could rescue barcode identification from samples failing rapid DNA extraction. The results partially supported this hypothesis, with only one out of four reprocessed samples achieving valid species identification, yielding a 25% recovery rate.
The limited recovery was mainly caused by poor initial template quality. Several failed samples, including Metasequoia and Hydrilla verticillata, possess tough and complex tissues. Inadequate grinding and incomplete cell lysis during rapid extraction reduced DNA yield and purity, causing initial PCR failure. Although silica-column purification improves DNA purification efficiency, it cannot fully compensate for severely low template quantity.
Furthermore, the intrinsically high raw error rate of Nanopore sequencing impairs identification accuracy on short plant barcode fragments (Loit et al. 2019). Most re-extracted samples produced severe base substitutions and indels, resulting in unreliable BLAST matches. Consistent with previous studies, simply replacing sequencing platforms cannot resolve fundamental defects in early DNA extraction (Bronzato Badial et al. 2018).
This study indicates that optimizing tissue grinding and lysis is more critical than switching sequencing methods for difficult aquatic and woody plant samples. This research has obvious limitations due to its small sample size. Future studies will expand sample diversity, optimize silica extraction protocols, and further define the practical boundary of Nanopore sequencing in low-quality plant barcode identification.

References

1.Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3):403–410.
2.Bronzato Badial A, Sherman V, Stone A, Rivera Y, Munoz-Aguirre P, Melgar J, De Wolf E. 2018. Nanopore sequencing as a surveillance tool for plant pathogens. Plant Dis 102(8):1648–1652.
3.CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc Natl Acad Sci U S A 106(31):12794–12797.
4.Loit K, Bahram M, Polme S, Tedersoo L. 2019. High raw error rate limits nanopore species identification for short plant barcode fragments. Phytopathology 109(11):1892–1901.
5.Loman NJ, Watson M. 2015. Successful test of the nanopore MinION sequencer for amplicon sequencing. Nat Methods 12(10):961–965.

Acknowledgements

I sincerely thank Dr. David A. Micklos from Cold Spring Harbor Laboratory DNA Learning Center for continuous guidance on this project. Special gratitude goes to Dr. Uwe Hilgert and Dr. Linnea Fletcher for offering valuable research ideas and academic support. I also appreciate Mrs. Xinyue Wang and Miss Nancy for preparing laboratory reagents and providing essential experimental assistance. Additionally, I thank Cold Spring Harbor Laboratory DNA Learning Center for supplying research resources and access to the DNA Subway platform for sequence analysis.