Using Deep Genome Skimming to resolve phylogenetic relationships in *Fuchsia* and *Circaea*

Lucia Anderson[a,b], Warren Wagner[a], Jun Wen[a], Gabriel Johnson[a]

[a]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

[b]The Pennsylvania State University, University Park, PA, USA

Introduction

Fuchsia is the third largest genus in family Onagraceae, and the only genus with berries and biporate pollen. Species of *Fuchsia* range from Central and South America, with some species native to Hispaniola and the Pacific Islands (New Zealand and Tahiti)[1].

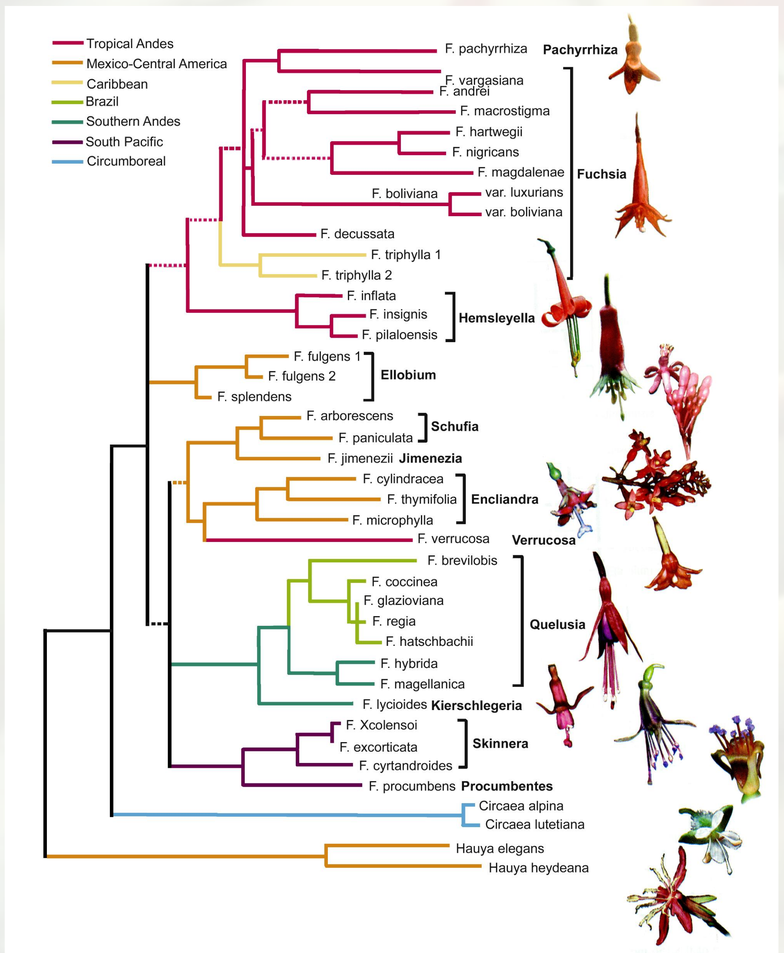


Figure 1: *Fuchsia* molecular phylogeny from Berry et al. (2004) with range distribution, sectional divisions, and flower morphology. Dotted lines represent unresolved phylogenetic relationships[1].

Molecular analyses have shown that *Fuchsia* and the widespread circumboreal genus, ***Circaea***, are sister genera[2] and have confirmed the character-based sectional groups within *Fuchsia*. These analyses, however, have lacked the resolution and sampling depth to determine early evolutionary history, biogeographical diversification, and the inter-sectional relationships of the genus[1], leaving many unanswered questions regarding *Fuchsia's* origins.

Research Goals

- **Resolve the early diversification within *Fuchsia***
- **Evaluate the success of Deep Genome Skimming within Onagraceae systematics as an alternative to Target Enrichment**
- **Further explore the evolutionary relationships between *Circaea* and *Fuchsia***

Herbarium Sampling

Sampling Aims:

Represent every section/clade	Obtain highest quality leaf material
Captures diversity of genera	Expands on previous studies

Image courtesy of Department of Botany Collections Database, Smithsonian Institution

Figure 3: Herbarium sheet of *Fuchsia boliviana* (Wood 5205 US).

Figure 2: Researchers discussing herbarium sheets for sampling.

Photo by James Di Loreto, Smithsonian Institution

Sampling Breakdown:

- Species Sampled: 44
  - 33 *Fuchsia*, 9 *Circaea,* 2 *Huaya*, 1 *Megacorax* (outgroups)
- Herbarium specimen from US National Herbarium

Molecular Methods

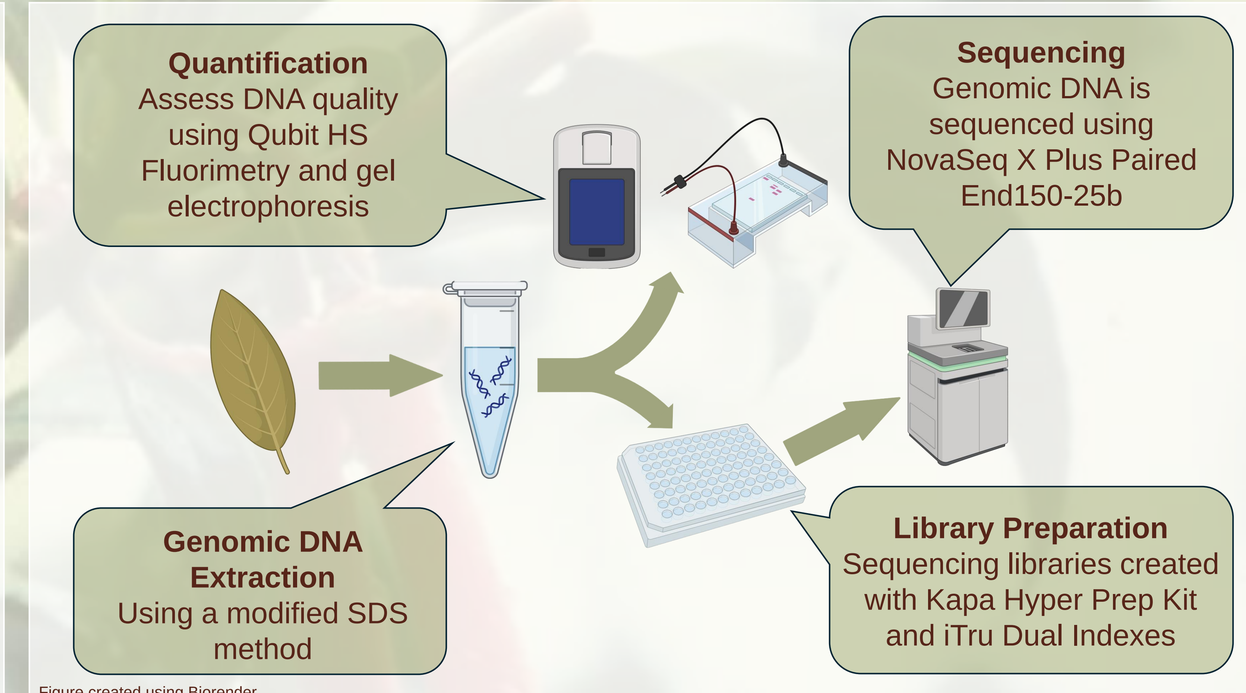


Figure created using Biorender.

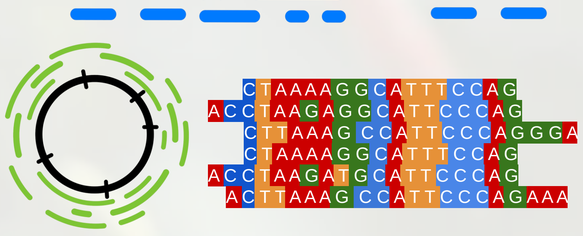
Bioinformatic Analysis

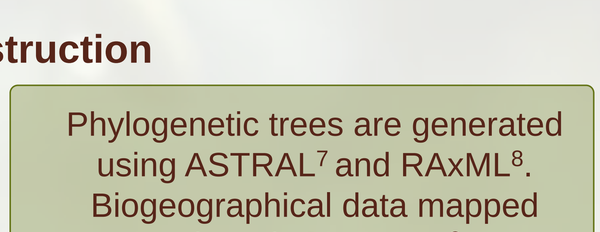
Genome Trimming and Quality Control

FastP[3] is used to removed adapters and low-quality reads from the whole genome reads.

Genome Assembly and Alignment

The plastome is assembled (*de novo* and reference mapped) using GetOrganelle4. Low copy nuclear genes are assembled via target reference mapping to ~300 loci using HybPiper[6]. Genomes are aligned using and MAFFT[5].





Tree Construction

Phylogenetic trees are generated using ASTRAL[7] and RAxML[8]. Biogeographical data mapped using BioGeoBears[9].

Next Steps

- **Finish collecting molecular data.**
  - 28 samples extracted and sequenced with 16 remaining.
- **Conduct bioinformatic analysis and create phylogenetic trees.**
- **Map morphological data, biogeographic data, and fossil record data to phylogenetic results.**
- **Expand sampling effort of Circaea to build upon previous analyses.**

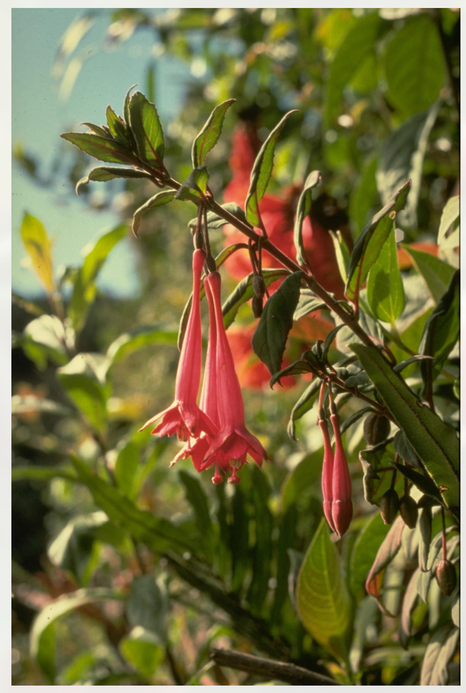


Figure 4: *Fuchsia caucana* in Nario, Colombia (Berry 3252 MO). Photo by Paul E. Berry, University of Michigan.

References and Acknowledgements

[1]Berry, P. E., W. J. Hahn, K. J. Sytsma, J. C. Hall, and A. Mast. 2004. Phylogenetic relationships and biogeography of Fuchsia (Onagraceae) based on noncoding nuclear and chloroplast DNA data. Amer. J. Bot. 91: 601–614.
[2]Bult, C. J., and E. A. Zimmer. 1993. Nuclear ribosomal RNA sequences for inferring tribal relationships within Onagraceae. Systematic Botany 18: 48–63.
[3]FASTP: Shifu Chen. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2: e107.
[4]GetOrganelle: Jian-Jun Jin*, Wen-Bin Yu*, Jun-Bo Yang, Yu Song, Claude W. dePamphilis, Ting-Shuang Yi, De-Zhu Li. **GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes.** *Genome Biology* **21**, 241 (2020).
[5]Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).
[6]Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C, and Wickett, N. J. (2016). HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. Applications in Plant Sciences, 4(7), 1600016.
[7]Zhang, Chao, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. 2018. "ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees." BMC Bioinformatics 19 (S6): 153.
[8]Alexey M. Kozlov, Diego Darriba, Tomáš Flouri, Benoit Morel, and Alexandros Stamatakis (2019) **RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference.** *Bioinformatics, 35 (21), 4453-4455*
[9]Matzke, Nicholas J. (2018). BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts. version 1.1.1, published on GitHub on November 6, 2018.

We would like to thank the NHRE program advisors, Jessica Johnson and Virginia Power, and directors, Vanessa González and Ioan Lascu for their support. We thank Alice Tangerini and Michelle-Marie Nelson for their help with figure design. We thank Jun Wen and Warren Wagner for the funding for this project, and NSF for the funding of this REU program. Background image courtesy of Paul E. Berry, University of Michigan.

NATIONAL MUSEUM of NATURAL HISTORY
Smithsonian

NSF
REU Site, EAR-2244445