A Phylogeny of the Ginseng Genus (*Panax* L., Araliaceae) based on Whole Plastome and 935 Targeted Nuclear DNA Sequences

Regina Fairbanks1, 2, Jun Wen2, Elizabeth Zimmer2

1. Department of Biology, University of Pennsylvania
2. Department of Botany, National Museum of Natural History

Introduction

*Panax* L. (*Araliaceae*), the ginseng genus, is one of the most medicinally important plants in China, where traditional medicine has utilized ginseng root for thousands of years. In addition to capturing interest in its medicinal applications, ginseng intrigues researchers as one of ~65 flowering plant genera with a classical eastern Asian and eastern North American disjunct geographic distribution5, 7, 12. It contains approximately 18 species, which are asymmetrically distributed with only 2 species native to North America5. Several molecular phylogenetic studies have attempted to resolve the evolutionary relationships within the genus5, 13, 14, 15, 16. However, relationships within the genus, and within the *P. bipinnatifidus* species complex especially, have been difficult to resolve. This project attempts to assess the utility of whole plastome and targeted nuclear DNA sequences produced with Next-Generation sequencing in improving phylogenetic resolution for this genus.

Methods

Molecular Experiments:

DNA was extracted from silica-dried leaves using the Qiagen DNeasy kit. DNA was sheared using a sonicator, then libraries were prepared with the NEBNext Ultra II kit. DNA was enriched for over 938 target nuclear genes before being sequenced on an Illumina HiSeq platform at NovoGene. The baits for the 938 nuclear genes were designed based on two genomes (*P. ginseng* and *P. notoginseng*) and two transcriptomes (*Hederis helic* and *Polyscias fruticosa*).

Sequence Assembly and Phylogenetic Analyses:

(using the HybPiper pipeline6 and performed on Hydra, the High Performance Computing Cluster at the Smithsonian, and Genomenius)

- **Raw reads** were processed with Trimomatic2 and evaluated with FastQC7.
- **HybPiper** python scripts8, BWA9, Spades10, and Exonerate11 were used to map, assemble, and retrieve the targeted nuclear sequences. 935 of the 938 targeted genes were recovered successfully. Chloroplast genomes were assembled with both reference-based and de novo approaches.
- MAFFT12 was used to align sequences.
- The 935 targeted nuclear genes were concatenated with AMAS12, and the best model for analysis was determined with PartitionFinder13.
- *MrBayes*14 and IQ-TREE15 were used to generate phylogenetic analysis.
- The 935 targeted nuclear genes and chloroplast sequences were analyzed as separate datasets.

Results & Conclusions

- The whole plastome and targeted nuclear gene phylogenies produced in this study are largely congruent with each other. Although the positions of *Polyscias australiana* and *Osmoxylon novoguineense* are not consistent between the phylogenies, this is likely due to the limited sampling. Furthermore, they are congruent with previous phylogenies of *Panax* based on plastid sequences9, 10 and ITS sequences5, 12.

- The sister relationship of *Aralia* and *Panax* shown by this study, as well as other molecular phylogenies, is supported by morphological data.

Analysis demonstrated *Panax siamensis*, a previously unsampled new species from northern Thailand, has many unique mutations in its chloroplast genome compared to other *Panax* species. The phylogenetic analyses suggest it forms a clade with *Panax vietnamensis*, another species from Southeast Asia. This hypothesis will be tested with a much larger dataset of 62 *Panax* samples which are currently being sequenced.

- *Panax notoginseng* is sister to the clade of the other four *Panax* species included in this study.

References & Acknowledgements

Acknowledgements

We thank Chun Su, Jing Liu, and Gabriel Johnson for laboratory assistance, and Zetong Nie and Bin-Bin Li for assistance in analysis. We also thank Gene Hunt, Liz Costrell, and Virginia Power, for their support of the NHRE program. Funding was provided through NSF REU site OCE-1560088 and Smithsonian Institution Barcoding Network (SiBNet).

References