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

Regina Fairbanks[1, 2], Jun Wen[2], Elizabeth Zimmer[2]

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

Smithsonian Institution

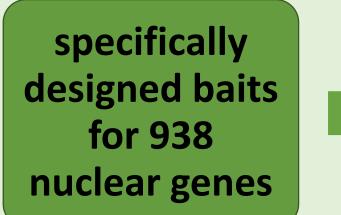NATIONAL MUSEUM *of* NATURAL HISTORY

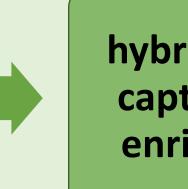NSF — REU Site, OCE-1560088

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 distribution[7,13]. It contains approximately 18 species, which are asymmetrically distributed with only 2 species native to North America[7]. Several molecular phylogenetic studies have attempted to resolve the evolutionary relationships within the genus[7,13,14,15]. 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 nuclear target 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 (*Hedera helix* and *Polyscias fruticosa*).

specifically designed baits for 938 nuclear genes → hybridization capture and enrichment → Illumina sequencing and data analysis

Figure 1: Target Enrichment Process

Sequence Assembly and Phylogenetic Analyses:

(using the HybPiper pipeline[4] and performed on Hydra, the High Performance Computing Cluster at the Smithsonian, and Geneious)

- Raw reads were processed with Trimmomatic[2] and evaluated with FastQC[1].
- HybPiper python scripts[4], BWA[8], Spades[10], and Exonerate[11] 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.
- MAFFT[5] was used to align sequences.
- The 935 targeted nuclear genes were concatenated with AMAS[3], and the best model for analysis was determined with PartitionFinder[6].
- MrBayes[12] and IQ-TREE[9] were used to generate phylogenetic trees.
- The 935 targeted nuclear genes and chloroplast sequences were analyzed as separate datasets.

Results

A. Nuclear genes — B. Plastome

- *Sinopanax formosanus*
- *Dendropanax arboreus*
- *Osmoxylon novoguineense*
- *Polyscias australiana*
- *Aralia parasitica*
- *Aralia finlaysoniana*
- *Panax wangianus*
- *Panax omeiensis*
- *Panax siamensis*
- *Panax vietnamensis*
- *Panax notoginseng*
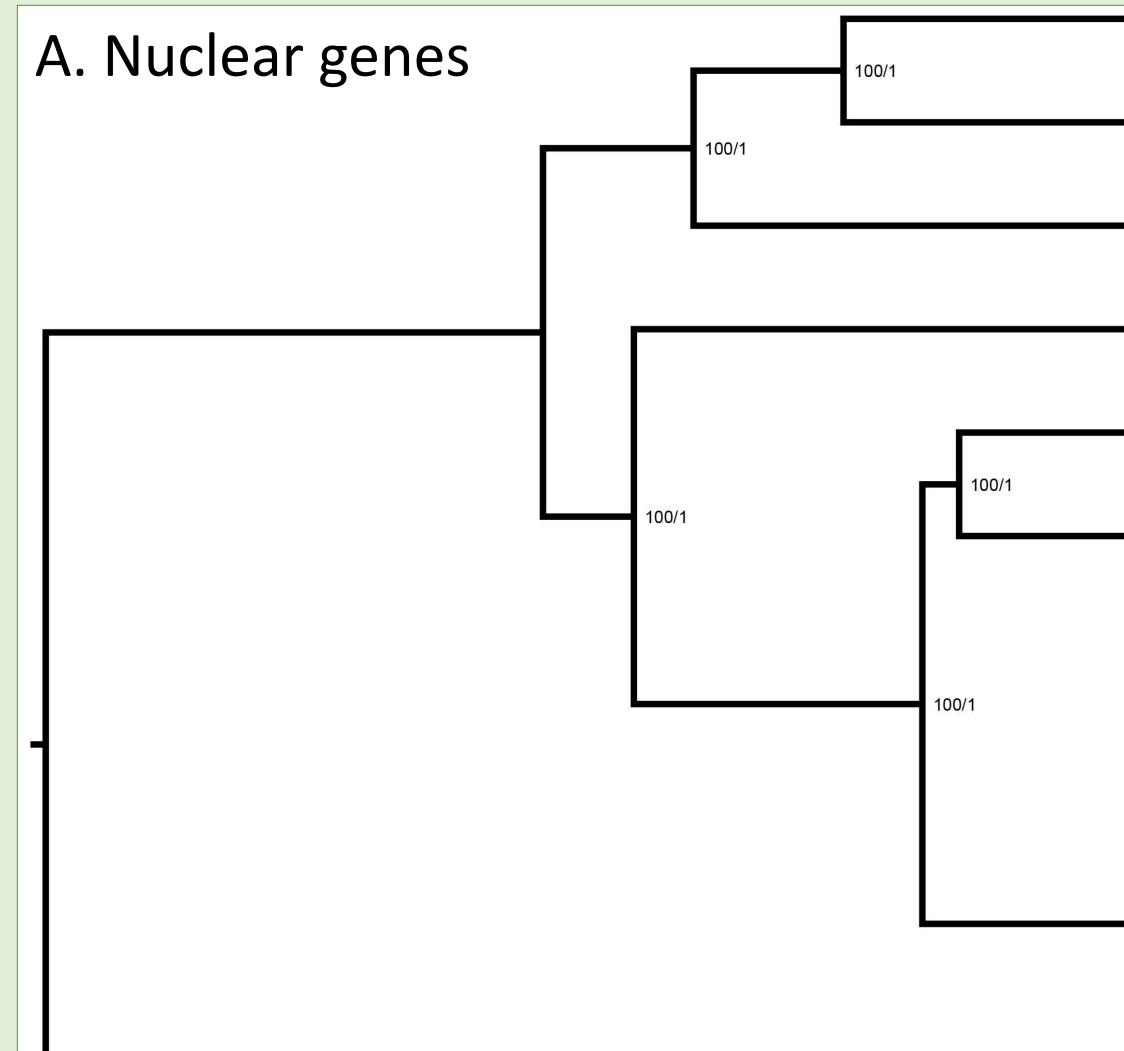- *Harmsiopanax ingens*

0.008   0.002

Figure 2: A. Phylogenetic tree using 935 targeted nuclear sequences. B. Phylogenetic tree based on whole plastome sequences. Trees were produced based on maximum likelihood (ML) using IQ-TREE and Bayesian inference (BI) using MrBayes. ML and BI trees were congruent with each other. At each node are the bootstrap values (left) and posterior probabilities (right).

Figure 3: Representatives of the ginseng genus:
(A) *P. wangianus*
(B) *P. bipinnatifidus*
(C) *P. pseudoginseng*
(D) *P. ginseng*
(E) *P. quinquefolius* root
(F) *P. notoginseng*
Photos courtesy of Jun Wen

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 chloroplast sequences[7] and ITS sequences[7,13].
- 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

References
1. Andrews, S. FASTQC. A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
2. Bolger, A. M., M. Lohse, and B. Usadel, (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.
3. Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660.
4. Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett, (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.*, 4(7):1600016.
5. Katoh, K. and D. M. Standley (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30:772-780.
6. Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.*, 34(3):772-773.
7. Lee, C. and J. Wen (2004) Phylogeny of Panax using chloroplast trnC-trnD intergenic region and the utility of trnC-trnD in interspecific studies of plants. *Mol. Phylogenet. Evol.*, 31:894-903.
8. Li, H. and R. Durbin (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60. [PMID: 19451168]
9. Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies.. *Mol. Biol. Evol.*, 32:268-274.
10. Nurk, S. et al. (2013) Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng M., Jiang R., Sun F., Zhang X. (eds) Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science, vol 7821. Springer, Berlin, Heidelberg
11. Slater, G. S., and E. Birney (2005) Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6:31.
12. Ronquist, F. and J. P. Huelsenbeck. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19:1572-1574.
13. Wen, J. and E. A. Zimmer (1996) Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): Inferences from ITS sequences of nuclear ribosomal DNA. *Mol. Phylogenet. Evol.*, 6: 167-177.
14. Zuo, Y.J., Z.J. Chen, K. Kondo, T. Funamoto, J. Wen, and S.L. Zhou (2011) DNA barcoding of *Panax* species. *Planta Med.*, 77:182-187.
15. Zuo, Y.J., J. Wen, J.S. Ma, and S.L. Zhou (2015) Evolutionary radiation of the *Panax* bipinnatifidus species complex (Araliaceae) in the Sino-Himalayan region of eastern Asia as inferred from AFLP analysis, *J. Syst. Evol.*, 53:210-220.