

Title: History of the Arctic Flora in North America

Authors: Eric G. DeChaine, Charles C. Davis

Western Washington University
516 High St.
Biology, Rm. 237
Bellingham, WA 98225

p. 360-650-6575
f. 360-650-3148
eric.dechaine@wwu.edu

Funding: NSF, OPP, Arctic Natural Sciences Program Award # 0714232

Overview: We developed a genetic dataset in an attempt to increase our understanding of how climate cycles of the Quaternary (the last 2.6 million years) impacted divergence in Arctic and alpine tundra flora. Specimens were collected in Beringia and the mountains of western North America, focusing on populations existing within (western and eastern Beringia, and the Pacific Northwest and Rocky Mountains, respectively), as well as sites between those refugia. Briefly, we sequenced DNA from chloroplast and nuclear loci for multiple species of tundra plants that inhabited the entire study region. Target species (*Campanula* species, *Kobresia myosuroides*, *Lloydia serotina*, *Pinguicula vulgaris*, *Poa alpina*, *Rhodiola* species, and *Saxifraga* species) were phylogenetically diverse and displayed a range of dispersal abilities, making them excellent candidates for evaluating general history of the tundra. Those genetic data make up this Arctic Flora dataset. From the genetic sequence data, gene trees are being reconstructed and used to estimate the patterns and timing of divergence, test which geologic and climatic events most impacted divergence, and infer how the process of genetic divergence differed between Arctic and alpine tundra refugia.

Time Period: Specimens were collected from 2007 through 2012

Location: The arctic and alpine tundras of North America

Data Collection: We generated a multi-locus DNA dataset of anonymous non-coding for estimating the evolutionary history of arctic-alpine plants in western North America. The use of multiple nuclear loci in studies of plant systematics improves estimates of species trees and the demographic history of constituent taxa. Non-coding markers were chosen as the basis for our phylogenetic and population genetic analyses because they are i) less likely to be under selection than coding genes, which would confound coalescent-based parameter estimation and ii) more likely to harbor informative nucleotide variation on the shallow timescales relevant to infraspecific studies. Further, our method of identifying loci allowed us to rapidly and randomly describe more non-coding loci and primers than could be obtained from the literature. In brief, we 1) extracted > 6 micrograms of genomic DNA from leaf tissue of a single representative specimen of each taxon using

Qiagen's (Valencia, CA, USA) DNeasy® Plant Extraction kit, 2) sheared DNA using a nebulizer, 3) blunt-end repaired the fragments, 4) electrophoretically separated DNA fragments on a 0.8% TAE gel, 5) extracted fragments ranging from 1 to 1.4 kilobase (kb) in length using Qiagen's QIAquick® gel extraction kit, 6) added an adenine overhang for ligation into Invitrogen's pCR®2.1-TOPO® TA vector and insertion into OneShot® TOP10 Competent cells, 7) grew cells on ampicillin-treated plates, and 8) screened colonies using standard blue-white screening. From the plates, we picked and lysed 32 colonies and used the lysate as template in PCR and sequencing reactions with M13 primers. Using Sequencher® v4.8 (Gene Codes Corp, Ann Arbor, MI, USA), vector sequence was removed from the raw sequences. To select the appropriate loci for use as ALPS, we ran blastn (nucleotide-nucleotide BLAST) and blastx (translated nucleotide-protein BLAST) searches against the nucleotide databases of the National Center for Biotechnology Information. We discarded DNA fragments matching organelle or protein coding sequences. Furthermore, we removed sequences that had long open reading frames. Fifteen sequences were selected for designing Anonymous Locus Primer Sets (ALPS). These primer sets were tested in temperature gradient PCR reactions to optimize annealing temperatures and to screen for non-target amplification. Ultimately, 5-10 ALPS reliably produced clean PCR products and were chosen for sequencing. In addition, we included the nuclear ITS region and the trnLF chloroplast region for some taxa.

Nuclear loci were amplified in 20 µl volumes, with final reagent concentrations of 1.5 mM MgCl₂, 0.5 µM primers, 0.4 mM dNTPs, 0.5 U Taq polymerase, and approximately 5 ng/µl template DNA. All reactions included a 3 min initial denaturation (94°C), 30 cycles of 30 sec denaturation (94°C), 30 sec annealing (55°C), and 1 min extension (72°C), and a final 10 min extension (72°C). All PCR products were cloned using TOPO® TA cloning kits and sequenced in both directions by the University of Washington's High-Throughput Genomics Service Unit. In preparation for downstream phylogenetic analyses, all genetic sequence datasets were edited on Sequencher (GeneCodes Corp.), aligned using ClustalX 2.0 (Larkin et al. 2007), and alignments were manually edited in MacClade 4.08a (Maddison and Maddison 2002).

Data Format: The data is archived in nexus files - the standard format for aligned sequence data ready to be used in a phylogenetic analysis. The files are separated by taxon and genetic locus within a given taxon. The nexus format provides details on the data type (DNA), number of samples, and length of sequence. Within a file, all the sequences are aligned but not interleaved, such that the 'sample name' is followed by the entire sequence for that sample.

References:

- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- Maddison, D. R. and W. P. Maddison, 2005. MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a. <http://macclade.org>.