

Mary DeDecker Botanical Grantee 2023 Progress Report
Kyle Rosenblad

Tasks for this project are grouped in five phases: leaf material collection, DNA extraction, DNA sequencing library preparation, DNA sequencing, bioinformatic analysis, and landscape genomic analysis. The study comprises 283 individual plants distributed across 49 montane meadows in the Sierra Nevada, San Bernardino, and southern Cascade mountain ranges.

Collection of leaf material is complete.

DNA extraction is approximately 80% complete. This phase of the project created a substantial bottleneck for approximately six months, but the problem is now resolved. During this bottleneck, I repeatedly tested numerous commonly used plant DNA extraction methods (e.g. variations on the classic CTAB method, Qiagen DNEasy) and found either unusably low DNA concentrations, as quantified by Qubit, or unusable DNA fragment size distributions after digestion with restriction enzymes (see library preparation below), as quantified by Bioanalyzer. After dozens of refinements, I have developed a highly modified version of the CTAB method that reliably produces DNA extractions with high concentrations, and which leads to suitable fragment size distributions for ddRADseq analysis in subsequent phases. My modified protocol requires an extended period of DNA precipitation in the freezer, which spreads the protocol out over several days. With the assistance of undergraduate researchers, I have now obtained suitable DNA extractions from 225 of the 283 plants in the study. This spring semester, the students I trained in the modified CTAB method are continuing to assist with the project, and I estimate that they will complete the extractions for the remaining 58 plants within the next month.

Using a subset of the DNA extractions that are already complete, I have begun piloting the DNA sequencing library preparation phase of the project. I have identified a pair of restriction enzymes (MspI and PstI) that produces a suitable distribution of DNA fragment sizes for downstream ddRADseq analysis, and I have begun training undergraduate research assistants in the techniques that will be applied following Peterson et al. (2014—*Diversity*). As DNA extractions are completed, the students and I will be aiming to complete library preparation and submit the libraries for sequencing by the end of March 2023. I have consulted with the staff at UC Berkeley's QB3 Genomics sequencing facility to arrange a plan for DNA sequencing to be implemented when library preparation is complete.

I have also begun preparing the bioinformatics pipeline that will be used to generate SNP (single nucleotide polymorphism) data from the DNA sequencing output. My study species, *Salix lemmonii*, is tetraploid and will thus require special bioinformatic tools beyond those typically used to call SNPs from short-read Illumina sequencing data. I have identified an R software package, "polyRAD", that includes the set of functionalities I will need. polyRAD allows users to specify multiple possible modes of chromosomal inheritance, from fully tetrasomic (likely resulting from a species' autopolyploid origin) to fully disomic (likely resulting from allotetraploid origin).

Similarly, I have begun preparing the analysis scripts that will be needed for the FOLDS analysis of climate change vulnerability among *Salix lemmonii* populations, as described in my proposal.

I intend to complete all analyses by the end of the spring 2023 semester, and to prepare the project's public-facing outputs this summer—i.e., a web application, a journal article, and—pending scheduling availability—a presentation to the Bristlecone CNPS Chapter.