6-8-2021

Next Generation Sequencing Identifies Population Structure and Signatures of Local Adaptation in Red Alder (*Alnus rubra* Bong.)

Jacob Brent Loveless
Portland State University

Next Generation Sequencing Identifies Population Structure and Signatures of Local Adaptation in Red Alder (*Alnus rubra* Bong.)

by

Jacob Brent Loveless

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Biology

Dissertation Committee:
Daniel J. Ballhorn, Chair
Mitchel B. Cruzan
Jason Podrabsky
Rahul Raghavan
Catherine de Rivera

Portland State University
2021

Abstract

Red alder (*Alnus rubra* Bong.) is the dominant hardwood tree species in the Pacific Northwest. Red alders are only found in western North America, generally within 200 km of the coast and below 1000 m in elevation–although there are several disjunct populations in the northern Rocky Mountains in Idaho. Commonly a riparian species, the shade intolerant red alder generates most of the leaf litter in the streams and rivers it occupies which has been shown to greatly influence the decomposer communities. This, in turn, has a cascading effect throughout the entire riparian ecosystem greatly influencing multiple levels of the food chain. Red alders also aid in stream cooling by shading areas with its canopy which has been shown to reduce diel temperature swings. This reduction improves survivability in ecologically and economically important salmonid fish species (e.g. Coho Salmon, Rainbow Trout) in which the fry require cool freshwater streams and rivers for juvenile growth and development. When not occupying riparian habitat, red alder is a pioneer species that is well suited to occupy recently disturbed areas (e.g. timber harvesting, fire damage), or areas with poor soil conditions where essential nutrients may be limited. The success of red alder as an early successional species is facilitated by the symbiotic relationship it forms with the N_2-fixing actinorhizal bacteria *Frankia alni*. In this mutualism, *Frankia* bacteria fix nitrogen for red alder and in return they receive photosynthetic products produced by their symbiotic host. The process of bacterial nitrogen-fixation can be very expensive for the host plant since as much as 40% of all photoassimilates produced are funneled to the *Frankia* symbionts for this process. Red alders usually associate with specific strains of

Frankia–which exhibit low diversity over large special scales–with the choice being controlled primarily by the host. Resultantly, alders are able to improve soil conditions which has been shown to improve growth and success of other valuable tree species in the Pacific Northwest (e.g. Douglas fir, big leaf maple). Initially considered a nuisance weed or a pest species by local logging operations due to the enumerable red alder seedlings that occupy freshly logged areas, red alder has become one of the most commercially valuable species in the Pacific Northwest. The quick growth, short lifespan (< 100 years), and beautiful red tinted wood–paired with red alder's ability to fix nitrogen thereby eliminating the need for chemical fertilizers–has made red alder one of the most commercially valuable species in the Pacific Northwest with plantations in several states, and up into Canada as well.

Despite the invaluable ecosystem services and the high commercial value of red alder, very little is known about their genetics. In particular, studies utilizing next generation sequencing (NGS) technology are lacking. To help narrow the knowledge gap in red alder genetics I used a high-density single nucleotide polymorphism profile generated by genotype by sequencing to help fill in the gaps of knowledge in phylogeography and local adaptation to environmental conditions, and selection of specific *Frankia* strains as well. I use a suite of relatively recently developed bioinformatics programs to look for population structure in red alder, and then utilize these data to calculate dispersal dates into the disjunct populations in Idaho. I was able to identify four sub-populations of red alder (2 coastal, 2 inland), and five separate dates at which I estimate red alder was able to disperse into the inland range. This helped clear up

ambiguities in the timing and number of times that red alder was able to disperse into the inland disjunct populations.

A combination of microscopy and statistical analysis was then used to look for polyploidy in red alder, and determine if polyploidy is common in the species. Diploid and tetraploid cytotypes were identified, and the distribution throughout the range was relatively even. Finally, the disjunct populations encounter different abiotic conditions than those on the coast. This, as well as the lack of regular gene flow into the disjunct inland populations provide a good system for studying local adaptation via association studies in red alder. I also utilized association studies to look for markers involved in *Frankia* selection as well. I identified five marker SNPs as candidate loci for local adaptation to environmental variables. I was also able to identify two marker SNPs as candidate loci for selection of specific *Frankia* genotypes by the novel approach of using the *Frankia* genotype as a red alder phenotype in a genome wide association study. This finding has significant potential for future crop improvement approaches.

Future studies should look at the entire distribution of red alder including Canada, California, and the rest of the disjunct populations in Idaho. It is likely that more sub-populations and many more dispersal events inland would be identified. Particularly, the *Frankia* association study could be ran on thousands of SNPs in hundreds of individuals covering the whole range of red alder's distribution. If more detail could be elucidated on the genetics involved in this selection process it may be possible to engineer common crop species with the ability to form symbiotic associations to fix nitrogen for said crop

species. This would eliminate the need for the addition of chemical fertilizers which have been shown to have a devastating effect on freshwater ecosystems.

Dedication

For Tracy

Acknowledgements

So many people have played integral roles in my journey throughout the last ten years that it would be impossible for me to list them all. For that reason, I will keep my acknowledgements brief and to the point. First I would like to thank my family for their support. My wife Tracy in particular was essential. I would also like to thank Odin, Jupiter, and Brent Loveless. I would like to thank my dissertation committee for the flexibility needed to complete this in the age of Covid. I thank my advisor Daniel Ballhorn for his guidance. I would also like to thank the rest of my research lab mates in the Ballhorn lab. Special thanks to Emily Wolfe for assistance in bioinformatics. Thank you all for making this possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Species frequently have broad geographic distributions over large distances with members of the same species often occurring across major abiotic and geographic barriers (e.g. rivers, mountain ranges). Explaining current distributions of taxa requires knowledge of the biogeographical history of the taxa being investigated. For example, when a past barrier was formed (e.g. uplift of the Cascade mountain range). It is necessary to understand ancestral lineages on both sides of the barrier, and how dispersal affected these lineages (see Rosen 1978, Avise *et al*. 1987). The ebb and flow of ice sheets and glaciers during the Quaternary/Pleistocene was a significant factor in current plant distributions in the PNW (Soltis *et al*. 1997; Soltis *et al*. 2006; Jaramillo-Correa *et al*. 2009). The advance and retreat of these ice sheets lead to extirpations, reduced population sizes of species, and forced species to survive in restricted glacial refugia followed by mainly northward expansion as the glaciers retreated. This, in conjunction with the uplift of the Cascades, is responsible for the disjunct nature currently seen in the PNW (Jackson *et al*. 2000; Brunsfeld 2001). A disjunct distribution is when one or more groups of a species are separated geographically and no longer experience gene flow. Disjunct distributions can be multicausal and can result from both expansion and contraction of the host range (Tallis 1991). Two main hypotheses, the ancient vicariance (AV), and recent dispersal (RD) have been suggested to explain the current patterns observed in this disjunction (Brunsfeld *et al*. 2001). First, AV suggests that inland and coastal populations were separated ~2.5 mya when the Cascade mountain range uplifted. The resulting rain shadow eliminated suitable habitat allowing for separated populations

to evolve independently form each other. In this scenario, the populations on either side of the barrier should have a high degree of genetic differentiation due to loss of gene flow. Second, RD suggests that inland populations were extirpated by the uplift of the Cascades, and current distributions can be explained by recolonization of as glaciers began to retreat ~20 kya. Elucidation of the evolutionary and historical processes that explain current distributions has been the key goal of phylogeography since the late 1990's (Soltis *et al*. 1997; Mitton *et al.* 2000; Brunsfeld *et al*. 2001; Eckert *et al*. 2010; Breen *et al*. 2012). The advent of next generation sequencing technologies (NGS) has revolutionized how phylogeographic studies are conducted by allowing for the analysis of multiple loci at a reduced cost (McCormack *et al*. 2013).

The identification of molecular markers has become both fast and cost-effective with the advancement of NGS. Single nucleotide polymorphisms (SNPs) have become the marker of choice due to their biallelic nature, heritability, and ubiquity (Verma *et al*. 2015). Specifically, genotype by sequencing (GBS; Elshire *et al*. 2011) has proven to be an especially efficient strategy in SNP discovery in plants (Bird *et al*. 2017; Chen *et al*. 2017; Ruffley *et al*. 2018; Bagley *et al*. 2020). SNPs identified by GBS allow for the efficient analysis of potentially thousands on SNP's making them useful in population genetic studies looking at population structure (Siadjeu *et al*. 2018; Wang *et al*. 2018; Pereira-Dias *et al*. 2019), trait association (Sonah *et al* 2015; Huggins *et al*. 2019) and phylogeography (Ruffley *et al*. 2018; Bagley *et al*. 2020). The wide range of uses for SNP's identified by GBS would not be possible if not for the development of adequate bioinformatic tools for any given analysis. Advances in bioinformatics have produced

programs that use SNPs to assess population structure (STRUCTURE; Pritchard *et al*. 2000; ADMIXTURE; Alexander & Lange 2000), make common population genetics calculations (Genalex; Peakall & Smouse 2006, Adgenet; Jombart 2008), perform whole genome association studies and linkage mapping (TASSEL; Bradbury *et al*. 2007), associate SNPs with environmental variables (Bayescenv; Villemereuil & Gaggiotti 2015), and asses variation in ploidy levels (gbs2ploidy; Gompert & Mock 2017, Ploidyngs; dos Santos *et al*. 2016).

Red alder is the most common hardwood tree species in the Pacific Northwest. It is a monecious dicot that acts as a pioneer species characterized by shade intolerance, rapid juvenile growth, tolerance of wet conditions and riparian systems, and the ability to form a symbiotic relationship with the N_2-fixing Actinorhizal *Frankia* bacteria. Red alders grow rapidly eventually reaching heights of up to 40 m. They have been shown to grow more than a meter in the first year and can reach 3 meters a in 2-5-year-old plants (Harrington & Curtis 1986). Growth slows after the juvenile stage and alders mature at 60-70 years of age with the maximum age usually being about 100yrs (Worthington *et al*. 1962). Red alders are adept at growing in recently disturbed areas and often regenerate after burning. For decades red alder was considered a pest species due to low economic value, however the red alder has become the major hardwood species in the PNW and now its value is above that of even the Douglas fir. Red alder is commonly a lowland species with a range from southeastern Alaska to southern California. It is generally found below 750m (best below 350m) in elevation and within 200 km of the Pacific coast–although there are the disjunct populations in northern Idaho (Figure 1.1). Red

alder grows in wet and humid environments with annual precipitation ranges from 400-5600mm with most of the precipitation occurring in the winter months. Red alder can also tolerate a large range of temperatures from -30°C in Alaska and Idaho, to 46°C in parts of southern California. The ability to fix atmospheric nitrogen allows red alder to grow in poor soil conditions and soil moisture during the growing season appears to be one of the main conditions that influence where alder grows (Harrington 1994). Red alder can tolerate poor drainage and mild flooding making it a common species in swamps, marshes, or riparian systems. Generally, it does not grow in areas

where drought is common and in Idaho and southern California it seldom grows on south facing slopes and is usually confined to riparian systems or the borders of lakes. Pure alder stands are usually confined to riparian systems or disturbed areas with poor soils. Mixed stands are common with the alder frequently associated with Douglas fir, grand fir, western red cedar, and big leaf maple to name a few.

Red alders can reach sexual maturity in as few as 3 years (Stettler 1978). They are monecious in which both male and female catkins form on the previous years' twigs (Hitchcock *et al*. 1964). Pollen is small and produced in abundance with flowering usually occurring in early spring with peak pollen shedding usually preceding peak receptivity by a few days, however synchrony of shedding and receptivity has been found in some trees (Bormann 1985). Red alder is generally an outcrossing species but some self-pollination occurs (Stettler 1978**)**. Seeds are winged nutlets, small, and are formed in pairs on the bracts of woody strobili that is small and cone-like (Schopmeyer 1974).

Seeds lack endosperm and contain two small cotyledons (Brown 1986). Red alders are

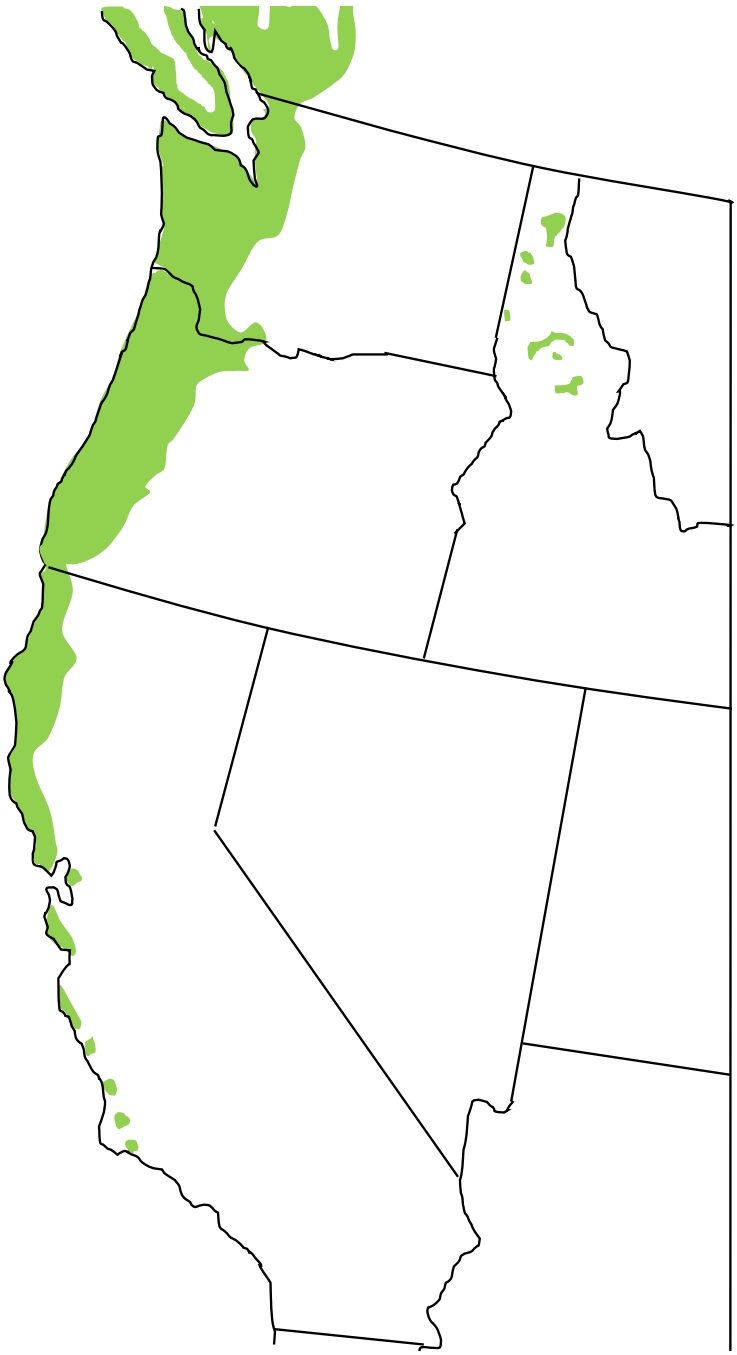


Figure 1.1. Map of Red Alder Distribution in the Pacific Northwest. Green shading indicated red alder distributions throughout the Pacific Northwest, southern Canada, and northern California.

prolific seed producers but seed production varies substantially from year to year and even tree to tree. For example, McGee (1988) found that whole stand seed production can increase as much as seven times from a bad year to a good year, and Brown (1985) reported a an in-stand variation of 0 seeds produced for some trees to 5.4 million seeds produced for others. Patterns of seed dispersal are not well documented however what little information that does exist suggests a similar pattern of dispersal to other members of Betulaceae (Zasada *et al*. 1991). Once catkins are mature, dispersal appears to be determined by dry weather conditions. Arid conditions dry the catkins which opens the scales and allows seeds to be disseminated by wind. The small, light nature (800-3000/g) of red alder seeds allows for efficient wind dispersal over long distances. Wind dispersal is the primary method for seed dissemination in red alder, however water dispersal has been shown (Brown 1986). Red alder seeds have been shown to be an important food source for some birds (White & West 1977). Birds can be passive or active seed dispersers by either disrupting catkins within the trees or by direct ingestion.

Red alder seeds germinate with little to no dormancy beginning in late February and is completed by mid-April in freshly disturbed areas (Haeussler & Tappeiner 1993). Despite the red alder being a prolific seed producer, there is no guarantee seeds will germinate or that seedlings will survive to the end of the growing season. For example, Harrington (1994) mentions an experiment where 1000-1500 seeds/m^2 were sowed in consecutive years and failed to produce a single seedling that survived to the end of the growing season. Seedling emergence is dependent on the type of site in which the seeds fall. For example, seedling emergence was twice as likely to occur in freshly disturbed

sites when compared to forested sites (Haeussler 1987). Seedling mortality was more prominent in forests where they were subject to damage by fungi and other pathogens that are not present in recently disturbed areas–seed mortality was also more pronounced in forested areas. As much as 60% of seeds deposited in forests are destroyed by soil organisms (Haeussler 1987).

Despite being an important pioneer species with high commercial value, very little is known about the genetics of red alder. Multiple studies have generated phenotypic data (Harrington 2006; Ballhorn 2017), however corresponding genotypic information is lacking (Harrington 2006). No varieties have been identified, however three possible subpopulations have been suggested based on morphological observations of herbarium specimens (Furlow 1979). At least four studies have looked at the history and distribution of red alder using genetic data (Strenge 1994; Soltis *et al.* 1997; Brumble 2008; Ruffley *et al*. 2017;). Three of the studies (Strenge 1994; Soltis *et al.* 1997; Brumble 2008;), used simple chloroplast markers and none of them were able to identify structure in genomic DNA despite evidence for distinct northern and southern genotypes being found. A fourth study (Ruffley *et al*. 2018) attempted to find genomic markers expand on the previous cpDNA findings using next generation sequencing (NGS) on herbarium specimens. Specifically, double digest restriction-site associated DNA sequencing (ddRAD) for SNP discovery to try and improve the quality and quantity of genetic data. The SNP data supported population structuring between the coastal and inland sites, unfortunately, a small sample size (49), as well as the decision to use plants from an herbarium collection greatly reduced the quality of genetic data. For example, only 648 usable SNPs were

identified, which is only a fraction of the nearly 6000 that they were expecting. Even though the number of SNP's identified was an order of magnitude less than expected, the authors were still able to obtain enough SNPs to test various scenarios pertaining to the current red alder distributions. Their results did not find the north/south sub-populations that were identified through cpDNA in the other studies, but they did find structuring between the inland and coastal alders. Interestingly, their results did not back either the AV or RD hypotheses either but rather a combination of AV with multiple migration events into the NRM.

The genus *Alnus* in general exhibits high variation in ploidy (Heuvel, 2011). *Alnus* chromosome number counts have yielded 2n=14, 28, 42, 56, 70, 84, and 112 (Furlow 1979; Oginuma *et al*. 2000; Heuvel 2011) with most species in the genus being tetraploid (Furlow 1979; Oginuma *et al*. 2000). *Alnus glutinosa*, *A. incana*, and *A. viridis* have been shown to have a base chromosome number of x=14 (Kovanda 1984; Furlow 1990) – which is typically considered the base chromosome number in *Alnus* (Furlow 1990). However, *A. pendula, A. senlatoides, A. japonica,* and *A. sieboidinna* were found to have a base chromosome count of x=7 (Oginuma *et al*. 2000). Thus, members of the genus with 28 chromosomes are frequently considered either diploid or tetraploid depending on the source (Furlow 1990). Red alder has a diploid chromosome count of 2n=14 (Wetzel 1929) placing it in the x=7 category. There is debate on the importance and role of polyploidy in speciation processes with many considering it an evolutionary dead end **(**see Soltis *et al*. 2014 for a review) – however polyploidization is common in plants and it is generally believed to have occurred at least one time in nearly all

angiosperm species (Jiao *et al*. 2011) suggesting polyploidy is likely a major driver in the evolution of plants (Adams & Wendel 2005; Chen 2007; Soltis *et al*. 2010). For example, it is estimated that 15% of all angiosperm speciation events have corresponded to an increase in ploidy, however benefits derived from the increase in ploidy are difficult to determine (Soltis *et al*. 2014). Polyploidy has also been an integral aspect in the domestication and proliferation of important crop species (Dubcovsky & Dvorak 2007). Despite the interest of polyploidy in both natural and commercial settings, many questions remain about how variation in ploidy affects the ecology and evolution of terrestrial plants (Soltis *et al*. 2010; Soltis *et al*. 2014)

The disjunct nature of plant distributions in many PNW species prevents regular gene flow into the NRM. Red alders in the NRM encounter different climatic conditions (e.g. temperature, temperature variation, precipitation) than those with coastal distributions. For example, temperature variation in coastal climates is generally between 10-15 ^{o}C where inland climates can vary as much as 40° C. Coastal climates also tend to have wet winters and dry summers while inland climates tend to have humid summers and dryer winters. The lack of gene flow and the difference in abiotic conditions between the coast and the NRM makes the red alder ideal for studying local adaptation. Understanding how a species responds to environmental changes is becoming more important as we face an unprecedented rate of the increase in global temperature. This is particularly important in plants because of their sessile nature (Alberto *et al*. 2013). Positive selection of traits beneficial for surviving in changing conditions is the primary

adaptive force driving plant evolution (Bose & Bartholomew 2013) and identification of these alleles is key to understanding adaptive evolution in plants.

The primary focus of this dissertation is using NGS to help elucidate the genetics involved in phylogeography, population genetics, polyploidy, and local adaptation of red alder, however it is impossible to talk about alders without also considering its N-fixing symbiotic partner *Frankia*. Red alders are typically colonized by specific *Frankia* strains (Lipus & Kennedy 2011), and the manner and specificity of this association is determined by the host plant (Pawlowsky & Demchenko 2012). In the PNW, alder-associated *Frankia* have been shown to exhibit low diversity over large spatial scales with as few as two genotypes constituting 96% of all identified samples (Kennedy *et al*. 2010a; Kennedy *et al*. 2010b). The low diversity of *Frankia* is not limited to the Pacific Northwest. For example, very little variation between *Frankia* populations was found on three mountain peaks in Arizona despite large spatial separation (Welsh *et al*. 2009). Since the discovery that root nodules provide an avenue for nitrogen fixation in legumes in the late 19th century, researchers have questioned why some plants can form these associations while others cannot, and whether or not this could be utilized in other plant systems (see Huisman & Geurts 2020 for a review). As NGS technologies have become more common we have increased our knowledge into the cellular and genetic processes involved in root nodule symbioses identifying several genes involved in nodule formation and the infection process. Despite many recent advances, the ability to engineer nitrogen-fixing nodulation ability in crop plants has yet to be achieved (Huisman & Geurts 2020).

For this dissertation my focus on *Frankia* is limited to their diversity, and whether or not markers in red alder can be identified that associate with their selection.

In this manuscript I use GBS to identify SNPs in 96 red alders sampled in 10 sites throughout the PNW (Figure 1.2). I use the bioinfomatic pipeline GBS-SNP-CROP (Melo *et al*. 2016) to identify SNPs from my sequencing results and create a high-density SNP profile for analysis. I then use a series of bioinformatics tools and statistical models to explore phylogeography, population genetics, polyploidy, and local adaptation in red alder.

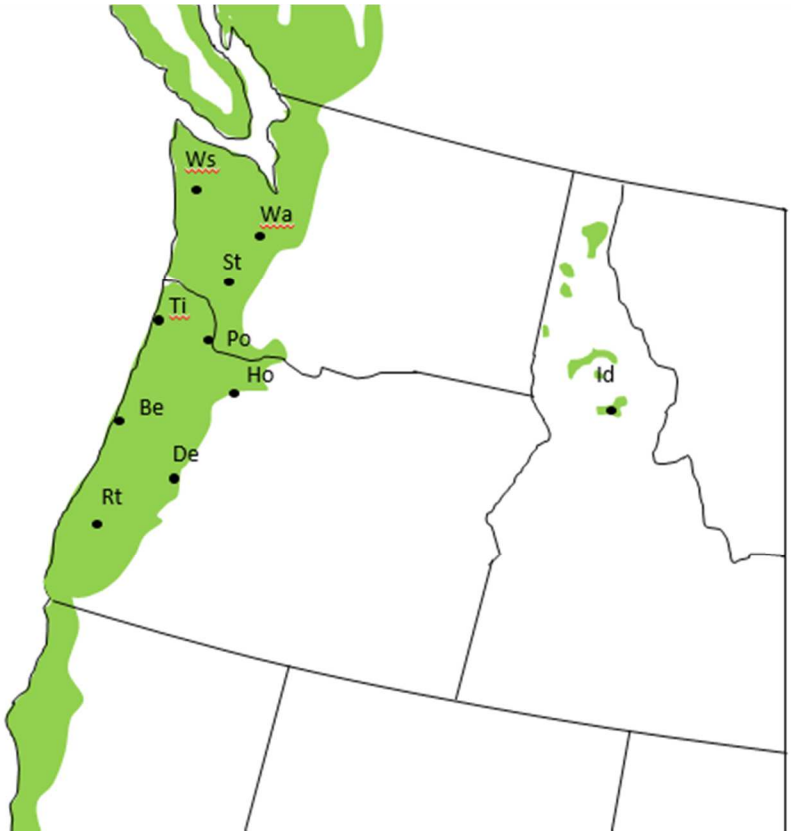


Figure 1.2. Map of Sampling Locations in the PNW. Map showing 10 red alder sampling sites in the Pacific Northwest.

In chapter 2 I use the high-density SNP profile called from GBS-SNP-CROP to look at population structure and test whether the alders' current distribution in the NRM is from AV or RD. I also use florescent microscopy on a subset of the 96 trees to identify polyploidy for the first time in red alder. I use DAPC and the program STRUCTURE to look for any population structuring in red alder of the PNW. Population structure refers to the patterns in genetic variation that result from the past or present departure from panmixia of a population. Understanding population structure can reveal when migration regimes changed in natural populations and identify areas of genetic diversity and characterizing the structure of current populations is key to conservation genetics. Both DAPC and STRUCTURE were chosen because they have been shown to be robust in handling mixed ploidy data. Population genetics calculations are also calculated in the program Genalex. The Bayesian program BEAST was used to make a timed phylogeny of red alder to help determine if there is evidence for AV or RD. Next I use SNP frequencies in the program gbs2ploidy to estimate the ploidy and look for patterns of ploidy distribution in of all of the 96 trees sampled. Identification of ploidy variation is important for bioinformatics and statistical programs that are required in this and the subsequent chapters. Likewise, correct ploidy information is crucial for crop improvement and selective breeding strategies. With the increase in use of red alder as a crop species and in urban restoration projects, the data collected on ploidy can also benefit policy decisions and future restoration projects. Specifically, I have four primary goals in this chapter. 1) Identify red alder population structure using SNP's within the PNW. 2) Use this data to test whether ancient vicariance occurred in the red alder, or if

current distributions are the result of recent dispersal in the area. 3) Identify polyploidy in red alder. 4) If polyploidy is found, determine if it is common.

In chapter 3, to test for local adaptation, a total of 19 environmental variables covering a 30-year period were downloaded from ClimateWNA (Wang *et al*. 2012). I use the program BAYESCENV (Villemereuil & Gaggiotti, 2015) and a latent factor mixed model (LFMM; Frichot *et al*. 2013) to identify outlier SNPs that associate with environmental variables. To look for SNPs that associate with *Frankia* genotype, I use two different GWAS models (GLM, MLM) to identify SNPs associated with *Frankia* genotypes. *Frankia* associated with the alders were genotyped by amplifying the nifH gene with PCR. Sequences were aligned in MEGA, and a 97% similarity was selected as the threshold for genotype identification. The resulting *Frankia* genotype was input as a red alder phenotype in the GWAS models. Bonferroni corrections and false discovery rate (FDR) were calculated and applied to outlier SNPs and only SNPs with significant p values that fit within the Bonferroni and PDR thresholds were kept for sequence identification. Because there is no reference genome for red alder, for SNPs found to be associated with *Frankia* genotype or environmental variables I use methods proposed by Otto *et al*. (2017) for sequence identification. This method aligns the sequence surrounding the SNP in Blast against the NCBI plant genomes nucleotide collection of flowering plants. Also, Blastn is used against *Alnus glutinosa* the only member of the genus with a sequenced genome. These data were utilized for three primary goals in this chapter: 1) Identify candidate loci involved in local adaptations to environmental conditions in red alder. 2) Identify *Frankia* genotypes to be used and phenotypes in our GWAS analysis. 3) Identify candidate loci involved in the selection of specific *Frankia*

genotypes. In chapter 4 I tie all of the previous chapters together and make closing remarks.

Very little is known about red alder on a genomic level. The phylogeography of red alder is important in understanding how the species responded to past environmental conditions and how it may respond in the future as the current climate continues to change. This, as well as information on ploidy level and genetic diversity can help develop more informed future restoration and conservation policy decisions. Identification of alleles under selection in locally adapted populations, as well as any markers involved in the selection of *Frankia* strains will not only increase our overall understanding of these processes in red alder, but for plants on general. The benefits of the knowledge gained can also aid in targeted breeding and trait selection in red alder as a crop species. The identification of markers involved in the selection of *Frankia* is of particular interest. If the genetic nature of how alders select for specific *Frankia* genotypes can be identified, it is possible that this can be engineered in other plants. Accomplishment of this could reduce the reliance on chemical fertilizers in modern crops. This reduction would help reduce the problems with nutrient loading currently found in many freshwater systems.

References

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, *8*(2), 135–141. https://doi.org/10.1016/j.pbi.2005.01.001

Alberto, F. J., Aitken, S. N., Al\'\ia, R., González-Martinez, S. C., Hänninen, H., Kremer, A., … others. (2013). Potential for evolutionary responses to climate change--evidence from tree populations. *Global Change Biology*, *19*(6), 1645–1661.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-246

Avise, J. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics. Vol. 18*, (c), 489–522. https://doi.org/10.1146/annurev.ecolsys.18.1.489

Bagley, J. C., Heming, N. M., Gutiérrez, E. E., Devisetty, U. K., Mock, K. E., Eckert, A. J., & Strauss, S. H. (2020). Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (Populus tremuloides). *Ecology and Evolution*, *10*(11), 4609–4629. https://doi.org/10.1002/ece3.6214

Ballhorn, D. J., Elias, J. D., Balkan, M. A., Fordyce, R. F., & Kennedy, P. G. (2017). Colonization by nitrogen-fixing Frankia bacteria causes short-term increases in herbivore susceptibility in red alder (Alnus rubra) seedlings. *Oecologia*, *184*(2), 497–506.

Bird, K. A., An, H., Gazave, E., Gore, M. A., Pires, J. C., Robertson, L. D., & Labate, J. A. (2017). Population Structure and Phylogenetic Relationships in a Diverse Panel of Brassica rapa L. *Frontiers in Plant Science*, *8*, 321. https://doi.org/10.3389/fpls.2017.00321

Bormann, B. T. (1985). *Early wide spacing in red alder (Alnus rubra Bong.): effects on stem form and stem growth* (Vol. 423). US Dept. of Agriculture, Forest Service, Pacific Northwest Forest and Range

Bose, R., & Bartholomew, A. J. (2013). *Macroevolution in deep time*. Springer Science & Business Media.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse

samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bradshaw, R. (1991). Plant Community History. J. H. Tallis. Publisher Chapman and Hall, London 1991, £49.00 ISBN 0 412 30320 5. *Journal of Quaternary Science*, *6*(4), 327–328. https://doi.org/https://doi.org/10.1002/jqs.3390060408

Breen, A. L., Murray, D. F., & Olson, M. S. (2012). Genetic consequences of glacial survival: The late Quaternary history of balsam poplar (Populus balsamifera L.) in North America. *Journal of Biogeography*, *39*(5), 918–928. https://doi.org/10.1111/j.1365-2699.2011.02657.x

Brown, S. M. (1985). A study of reproductive biology of Alnus rubra along three elevational transects in Washington and Oregon. *On File with: USDA Forest Service, Pacificth Northwest Research Station, Olympia, WA*.

Brown, S. M. (1986). *Sexual Allocation Patterns in Red Alder (Alnus Rubra Bong.) Along Three Elevation Transects*. University of Washington.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). Comparative phylogeography of north- western North America : a synthesis. *Integrating Ecological and Evolutionary Processes in a Spatial Context*, 319–339.

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.*, *58*, 377–406.

Chen, W., Hou, L., Zhang, Z., Pang, X., & Li, Y. (2017). Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of Ziziphus jujuba Assessed with Genome-wide SNPs Developed by Genotyping-by-sequencing and SSR Markers. *Frontiers in Plant Science*, *8*, 575. https://doi.org/10.3389/fpls.2017.00575

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, *316*(5833), 1862–1866.

Eckert, A. J., Van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics*, *185*(3), 969–982. https://doi.org/10.1534/genetics.110.115543

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach

for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Furlow, J. J. (1979). The systematics of the American species of Alnus (Betulaceae). *Rhodora*, *81*(825), 1–121.

Furlow, J. J. (1990). The genera of Betulaceae in the southeastern United States. *Journal of the Arnold Arboretum*, *71*(1), 1–67. Retrieved from http://www.jstor.org/stable/43782230

Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, *17*(6), 1156–1167. https://doi.org/10.1111/1755-0998.12657

Harrington, C. A. (2006). Biology and ecology of red alder. *United States Department of Agriculture Forest Service General Technical Report PNW*, *669*, 21.

Harrington, C. A., & Curtis, R. O. (1986). *Height growth and site index curves for red alder*. 14. Retrieved from http://www.fs.fed.us/pnw/pubs/rp358.pdf

Harrington, C. A., Zasada, J. C., & Allen, E. A. (1994). Biology of red alder (Alnus rubra Bong.). *Biology and Management of Red Alder*, 3–22.

Haeussler, S. (1987). *Germination and first-year survival of red alder seedlings in the central Coast Range of Oregon*.

Haeussler, S., & Tappeiner II, J. C. (1993). Effect of the light environment on seed germination of red alder (Alnus rubra). *Canadian Journal of Forest Research*, *23*(7), 1487–1491.

Heuvel. (2011). Genomic and Breeding Resources. *Springer*. Retrieved from http://link.springer.com/content/pdf/10.1007/978-3-642-14387-8.

Huggins, T. D., Chen, M., Fjellstrom, R. G., Jackson, A. K., McClung, A. M., & Edwards, J. D. (2019). Association Analysis of Three Diverse Rice (Oryza sativa L.) Germplasm Collections for Loci Regulating Grain Quality Traits . *The Plant Genome*, *12*(1), 170085. https://doi.org/10.3835/plantgenome2017.09.0085

Hitchcock, C. L., Cronquist, A., Ownbey, M., Thompson, J. W., & others. (1964). Vascular plants of the Pacific Northwest. Part 2. Salicaceae to Saxifragaceae. *Vascular Plants of the Pacific Northwest. Part 2. Salicaceae to Saxifragaceae.*

Huisman, R., & Geurts, R. (2020). A roadmap toward engineered nitrogen-fixing nodule symbiosis. *Plant Communications*, *1*(1), 100019.

Jackson, S. T., Webb, R. S., Anderson, K. H., Overpeck, J. T., Webb III, T., Williams, J. W., & Hansen, B. C. S. (2000). Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, *19*(6), 489–508. https://doi.org/https://doi.org/10.1016/S0277-3791(99)00093-1

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15.

Jaramillo-Correa, J. P., Beaulieu, J., Khasa, D. P., & Bousquet, J. (2009). Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research*, *39*(2), 286–307.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., … Depamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–100. https://doi.org/10.1038/nature09916

Kennedy, P. G., Schouboe, J. L., Rogers, R. H., Weber, M. G., & Nadkarni, N. M. (2010). Frankia and Alnus rubra canopy roots: an assessment of genetic diversity, propagule availability, and effects on soil nitrogen. *Microbial Ecology*, *59*(2), 214–220.

Kennedy, P. G., Weber, M. G., & Bluhm, A. A. (2010). Frankia bacteria in Alnus rubra forests: genetic diversity and determinants of assemblage structure. *Plant and Soil*, *335*(1), 479–492.

Kovanda, M. (1984). Chromosome numbers in selected Angiosperms. *Preslia*, *56*(4), 289–301.

Lipus, A., & Kennedy, P. G. (2011). Frankia assemblages associated with Alnus rubra and Alnus viridis are strongly influenced by host species identity. *International Journal of Plant Sciences*, *172*(3), 403–410.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and

phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526–538. https://doi.org/10.1016/j.ympev.2011.12.007

McGee, A. B. (1988). *Vegetation response to right-of-way clearing procedures in coastal British Columbia*. University of British Columbia.

Melo, A. T. O., Bartaula, R., & Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics*, *17*(1), 1–15.

Mitton, J. B., Kreiser, B. R., & Latta, R. G. (2000). Glacial refugia of limber pine (Pinus flexilis James) inferred from the population structure of mitochondrial DNA. *Molecular Ecology*, *9*(1), 91–97. https://doi.org/10.1046/j.1365-294X.2000.00840.x

Oginuma, K., Futagawa, K., Hizume, M., Shibata, F., & Kondo, K. (2000). Chromosomal evolution of three genera in the Betulaceae concerning polyploidy revealed by in situ hybridization using 5S rDNA. *Chromosome Science*, *4*(2), 47–5

Otto, L. G., Mondal, P., Brassac, J., Preiss, S., Degenhardt, J., He, S., … Sharbel, T. F. (2017). Use of genotyping-by-sequencing to determine the genetic structure in the medicinal plant chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by genome-wide association mapping. *BMC Genomics*, *18*(1), 1–18. https://doi.org/10.1186/s12864-017-3991-0

Pawlowski, K., & Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma*, *249*(4), 967–979.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pereira-Dias, L., Vilanova, S., Fita, A., Prohens, J., & Rodríguez-Burruezo, A. (2019). Genetic diversity, population structure, and relationships in a collection of pepper (Capsicum spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). *Horticulture Research*, *6*(1). https://doi.org/10.1038/s41438-019-0132-8

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Rosen, D. E. (1978). Vicariant Patterns and Historical Explanation in Biogeography. *Systematic Biology*, *27*(2), 159–188. https://doi.org/10.2307/2412970

Ruffley, M., Smith, M. L., Espíndola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, *27*(4), 1012–1024. https://doi.org/10.1111/mec.14491

Siadjeu, C., Mayland-Quellhorst, E., & Albach, D. C. (2018). Genetic diversity and population structure of trifoliate yam (Dioscorea dumetorum Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biology*, *18*(1), 1–14. https://doi.org/10.1186/s12870-018-1593-x

Soltis, D. E., Buggs, R. J. A., Doyle, J. J., & Soltis, P. S. (2010). What we still don't know about polyploidy. *Taxon*, *59*(5), 1387–1403.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). *Systematics and Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America *. 206*, 353–373.

Soltis, D. E., Segovia-Salcedo, M. C., Jordon-Thaden, I., Majure, L., Miles, N. M., Mavrodiev, E. V, … Gitzendanner, M. A. (2014). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al*.(2011). *New Phytologist*, *202*(4), 1105–1117.

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I., & Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, *13*(2), 211–221. https://doi.org/10.1111/pbi.12249

Stettler, R. F. (1978). Biological aspects of red alder pertinent to potential breeeding programs. *USDA For Serv Gen Tech Rep PNW Pac Northwest For Rangeexp Stn*.

Strenge, D. D. (1994). *The Intraspecific Phylogeography of Polystichum Munitum and Alnus Rubra*. Retrieved from https://books.google.com/books?id=pWhlOAAACAAJ

Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., & Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (Cicer arietinum L.) using Genotyping-by-Sequencing (GBS). *Scientific Reports*, *5*(10531), 1–14. https://doi.org/10.1038/srep17512

Villemereuil, P. De, & Gaggiotti, O. E. (2015). *A new F ST -based method to uncover local adaptation using environmental variables*. 1248–1258. https://doi.org/10.1111/2041-210X.12418

Wang, X., Bao, K., Reddy, U. K., Bai, Y., Hammar, S. A., Jiao, C., … Fei, Z. (2018). The USDA cucumber (Cucumis sativus L.) collection: genetic diversity, population structure, genome-wide association studies, and core collection development. *Horticulture Research*, *5*(1). https://doi.org/10.1038/s41438-018-0080-8

Wang, T., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. (2012). ClimateWNA—high-resolution spatial climate data for western North America. *Journal of Applied Meteorology and Climatology*, *51*(1), 16–29.

Welsh, A. K., Dawson, J. O., Gottfried, G. J., & Hahn, D. (2009). Diversity of *Frankia* populations in root nodules of geographically isolated Arizona alder trees in central Arizona (United States). *Applied and Environmental Microbiology*, *75*(21), 6913–6918.

Wetzel, G. (1929). *Chromosomenstudien bei den Fagales*. Akademische Verlagsgesellschaft mbh.

White, C. M., & West, G. C. (1977). The annual lipid cycle and feeding behavior of Alaskan redpolls. *Oecologia*, *27*(3), 227–238.

Worthington, N. P., Ruth, R. H., Matson, E. E., & others. (1962). Red Alder [Alnus rubra]-its management and utilization. *Miscellaneous Publications. United States Department of Agriculture*, (881).

Zasada, J. C. (1992). The reproductive process in boreal forest trees. *A Systems Analysis of the Global Boreal Forest*, 211–233.

Chapter 2

Population Structure, Ancient Vicariance and Polyploidy in Red Alder

Abstract: Red alder (*Alnus rubra* Bong.) is the most common hardwood tree species in the Pacific Northwest. Commonly a riparian species, the shade intolerant red alder generates most of the leaf litter in the streams and rivers it occupies which has been shown to greatly influence the decomposer communities. This, in turn, has a cascading effect throughout the entire riparian ecosystem greatly influencing multiple levels of the food chain. When not occupying riparian habitat, red alder is a pioneer species that is well suited to occupy recently disturbed areas (e.g. timber harvesting, fire damage), or areas with poor soil conditions where essential nutrients may be limited. Red alder is able to do this because the symbiotic relationship it forms with the N-fixing Actinorhizal bacteria *Frankia alni*. *Frankia* fix nitrogen for red alder and in return they receive photosynthetic products produced by their symbiotic host. Red alders usually associate with specific strains of *Frankia*–which exhibit low diversity over large special scales– with the choice being controlled primarily by the host. Despite the invaluable ecosystem services and the high commercial value of red alder, very little is known about their genetics including limited information on ploidy. To help narrow the knowledge gap in red alder genetics we used a high-density single nucleotide polymorphism profile generated by genotype by sequencing to help fill in the gaps of knowledge in the population genetics and phylogeography of red alder. We use a suite of relatively recently developed bioinformatics programs to look for population structure in red alder, and then utilize these data to calculate dispersal dates into the disjunct populations in Idaho. A

combination of microscopy and statistical analysis was then used to look for polyploidy in red alder, and determine if polyploidy is common in the species. we were able to identify four sub-populations of red alder (2 coastal, 2 inland), and five separate dates in which we estimate red alder was able to disperse into the inland range. Diploid and tetraploid cytotypes were identified, and the distribution throughout the range was relatively even.

Introduction

Numerous plant species have widely distributed populations across large portions of the North American continent (Iverson *et al*. 2008). These species often occur over significant geological barriers (e.g. rivers, canyons, or mountain ranges) may encounter drastically different abiotic conditions such as precipitation gradients and variation in soil conditions (Avise *et al*. 1987). Phylogeography attempts to elucidate the past processes involved in shaping the current distributions of organisms and therefore increase our understanding about the species as a whole. In the history of the Pacific Northwest (PNW) of the United States two major events have been proposed to have shaped current species distributions. First, the uplift of the Cascade mountain range 2.5-5mya created a roughly 300km wide rain shadow in eastern Washington and Oregon which lead to xeric soil conditions and forced temperate forest species to either survive in suitable habitat to the west of the Cascades, or in a few isolated refugia in the Northern Rocky Mountain (NRM) range. Second, Quaternary/Pleistocene glaciation driven by 100,000 ky cycles of wobble in the Earth's orbit (Milanovitch cycles) which, in turn, further reduced the available habitat for temperate forest trees by burying as much as half of the PNW under

the cordilleran and alpine ice sheets (Delcourt & Delcourt 1993). During this time in western North America, most biota were forced to migrate to warmer areas in southern California (Roberts & Hamann 2015), or into other restricted glacial refugia (Peilou 1991; Roberts & Hamann 2015). Interestingly on the PNW coast, wet conditions likely prevailed even during peak glaciation allowing for conditions suitable for forest trees to survive (Heusser 1985; Ruffley *et al*. 2018). This suggests that the Olympic Peninsula, as well as the Vancouver and Queen Charlotte Islands likely represented a large refugium for temperate forest species to have persisted during glacial maximums (Roberts & Hamann 2015).

Most of the phylogeographic studies conducted so far have focused on animal species (see Avise 2000 for review), however the disjunct nature of the PNW has made this region a prime candidate for phylogeographical studies in plants (Carstens *et al*. 2004; Steele *et al*. 2005; Ruffley *et al*. 2018). At least 60 species of coastal plants (Brunsfeld 2001), as well as numerous fungi and animal species exist in the NRM (Johnson 1987) and are separated suitable habitat from the arid conditions of the eastern Washington and Oregon. Brunsfeld *et al*. (2001) proposed two main hypotheses to explain the presence of coastal species in the NRM. First, the ancient vicariance hypothesis (AV), explains cryptic divergence (i.e. divergence before the Pleistocene) seen across the Cascades disjunction. The AV hypothesis suggests that temperate coastal species had continuous distributions all the way through to the inland forests of the NRM. The uplift of the Cascades isolated the inland from the coastal populations restricting gene flow between the two areas. Throughout the Quaternary the NRM acted as its own glacial refugia from the advancing ice sheets allowing the populations to survive there

and evolve separately for the last 2.5-5my. Because of the extended time with no gene flow there should be evidence of cryptic diversity and possibly allopatric speciation. If this is the case, the inland and coastal populations would show evidence of genetic differentiation and population structuring. The second hypothesis, inland/recent dispersal (RD), suggests that the uplift of the Cascade mountain range extirpated the inland populations, and the current distributions are due to dispersal and recolonization as the glaciers retreated after the last glacial maximum (LGM) about 20 kya (Brunsfeld *et al*. 2001). If is the case, little time would have passed since colonization and there should be limited genetic differentiation between the isolated NRM and the inland populations. Interestingly, evidence exists for both hypothesis and it is species dependent (Brunsfeld *et al*. 2001, Ruffley *et al*. 2018).

One of the plant species that occurs in predominantly coastal habitats but also in isolated populations in the NRM is red alder (*Alnus rubra* Bong). Today, red alder is the dominant hardwood tree species in the PNW. It is a shade-intolerant, relatively short-lived (< 100 years) pioneer species that is common in riparian systems and in recently disturbed areas. The ability for red alder to form a symbiotic relationship with the N_2-fixing actinorhizal *Frankia* bacteria allows it to grow in and improve poor and degraded soils (Harrington 1994). Due to its quick growth and use of alder wood in furniture and other commercial products, red alder has become one of the most commercially valuable tree species in the PNW (Xie 2008). Red alder is commonly a lowland species with a current range from southeastern Alaska to southern California (Harrington 1994). It is generally found below 750m (mostly below 350m) in elevation and within 200 km of the

Pacific coast–although there are small disjunct populations in northern Idaho which are separated by arid conditions not suitable for temperate tree species (Figure 1.1).

Several studies have used genetic data to elucidate how red alder became established in the NRM. Strenge (1994), and Brumble (2008) both used cpDNA and their results suggest a recent dispersal of red alder into the inland NRM range. This is in concordance with a statistical model developed to infer the existence of ancient vicariance in the region (Espindola *et al*. 2016**)** which predicted with over 98% certainty that no vicariance should exist in red alder. Using cpDNA, Strenge (1994) identified a northern and southern genotype in the coastal populations with an area of mixed genotypes near the Oregon/Washington border, however the pattern has not been shown in somatic DNA, and no population structure was identified. Red alders in the NRM consisted of the southern coastal genotype (Strenge 1994; Brumble 2008) which suggests a recent dispersal by a southern route (Brunsfeld *et al*. 2001). In both Strenge (1994), and Brumble (2008), cpDNA was used and no similar patterns were found in genomic DNA which raised the possibility that these inferences were being driven by a single polymorphism (Ruffley *et al*. 2018). To circumvent this problem Ruffley *et al*. (2018) used ddRAD sequencing in hopes to generate thousands of SNPs to be used for analysis. Their study found solid evidence of separate inland and coastal sub-populations and suggested at least some recent gene flow between the NRM and coastal sites occurred as well. No population structuring was identified in the coastal samples in this case either. Unfortunately, Ruffley *et al*. (2018) chose to use relatively few (49) samples that were prepared herbarium specimens to extract DNA from which only generated about 10% of the SNPs they expected. However, even with the fewer than expected number of SNPs

they were able to test different models on possible explanations for the patterns seen in the Idaho population. Their results suggested that neither the AV or RD scenarios explained the current distributions of the Idaho populations but rather it was a combination of the two.

It is generally believed that all angiosperm species have had a genome duplication event at some time in their history (Jiao *et al*. 2011). It has been suggested that polyploidization can lead to instant speciation (Mayr 1963), and thus represents a significant force in plant evolution (see Soltis & Soltis 2009 and Baduel *et al*. 2018 for reviews). The frequency of polyploidy generally increases with latitude in the north of the equator (Stebbins 1950; Johnson & Packer 1965). The insight that this pattern correlated with glaciation and refugia restrictions rather than latitude was initially proposed by Stebbins (1985). Many polyploid plants, and even a few known polyploid insects are widespread in previously glaciated areas while diploids of the species are restricted to areas in proximity to glacial refugia (Ehrendorfer 1980; Stebbins 1984). When the ice sheets retreated at the end of the LGM, an enormous amount of newly available, highly disturbed terrain became available for polyploids to colonize. Polyploidy can alter a plants' physical characteristics (*e.g*. seed and stomata size), increase abiotic stress tolerance, and alter the ecology of the specific species in as little as one generation (Levin 2002) leading to individuals that can tolerate changing environmental conditions and exploit new niches (Leitch & Leitch 2008). Resultantly, polyploids are generally assumed to be more successful than diploids in stressful or recently disturbed environments (Stebbins 1985; Otto & Whitton 2000). While polyploidy has been studied in wild plants (Martínková *et al*. 2015; Mavrodiev *et al*. 2015), much of the knowledge we have of

polyploidization comes from crop research (Renny-Byfield & Wendel 2014). The identification of new genetic resources assists efforts to identify the molecular basis for the observed variation in phenotypes, and aids in the identification of effective marker-assisted breading strategies (Yoon *et al*. 2015). Having information on variation in ploidy is necessary for these types of genetic assessments. For example, whole genome duplications can increase heterozygosity, increase selfing, and increase genetic diversity at the molecular and population levels (Soltis & Soltis 2000). Also, accurate allele frequencies are often required in the calculation of many common population genetic metrics (e.g. F_{ST}) including genetic diversity (H_E).

The genus *Alnus* shows extensive polyploidy and exhibits high variation in chromosome counts (Heuvel, 2011). *Alnus glutinosa*, *A. incana*, and *A. viridis* have been shown to have a base chromosome number of x=14 (Kovanda 1984; Furlow 1990) – which is typically considered the base chromosome number in *Alnus* (Furlow 1990). However, *A. pendula, A. senlatoides, A. japonica,* and *A. sieboidinna* were found to have a base chromosome count of x=7 (Oginuma *et al*. 2000). Red alder has a diploid chromosome count of 2n=14 (Wetzel 1929) placing it in the x=7 category like *A. pendula, A. senlatoides, A. japonica.*, however polyploidy in red alder has yet to be identified.

At this point, despite cpDNA analyses suggesting distinct northern and southern genotypes, no somatic variation or population structure has been identified in the coastal red alders, and studies on inland and coastal differentiation have given conflicting results. We believe that our high-density SNP profile will allow us to identify structure related to the coastal north/south genotype partitioning and help determine if the NRM are really

uniformly distributed as suggested by Strenge (1994), or if there is discontinuity among the coastal and inland sites as suggested by Ruffley *et al*. (2018). Our SNP profile should also allow us to perform a rooted, time-measured phylogeny utilizing a Bayesian framework. While these methods can be used for constructing phylogenies, it also provides the framework for testing evolutionary hypotheses (Drummond *et al*. 2012) such as AV and RD. By calculating dates of divergence of the inland and coastal alders, we will estimate times of divergence and match them up to historical events such as warming events or past glacial maximums. This will also allow us to determine if inland dispersal events were common in the periods of warming between peak glaciation, or if the inland dispersal since the LGM was a unique event. Finally, our SNP profile–supplemented by microscopy–should also allow us to test for polyploidy using allelic ratios generated during the SNP calling procedure.

This study had four main objectives. 1) Identify population structure in coastal and inland red alders. 2) Use this data to test whether there is evidence of ancient vicariance in the NRM. 3) Identify if polyploidy occurs in red alder and 4) If polyploidy is found, identify if it is common, and look for any patterns in the distribution. The gained knowledge from these objectives will aid in calculating common population genetic metrics within the PNW and NRM allowing for the identification of areas of genetic diversity and polyploid cytotypes which can be utilized to help clarify gaps in red alder phylogeography, as well as providing useful information for timber crop improvement and conservation efforts.

Materials and methods

Study area and sampling design

A total of 96 red alders in 10 sampling areas were selected throughout Oregon, Washington and Idaho (Figure 1.2) encompassing most of the range of red alder in the PNW. In Idaho, the sampling was conducted in the Clearwater Natural Research Area with special permission from the Idaho Forest Service. In order to compensate for potential rare genotypes or missing data and to maximize SNP coverage we attempting to collect ~10 alders per site (Bagley *et al*. 2020). However, red alder was difficult to find in the DE site so only three trees were sampled in this area. The DE site is the most eastern among the coastal sites and most of the habitat in this area that is suitable for red alder was occupied by white alder (*Alnus rhombifolia*), possibly indicating the far eastern range of red alder distribution. The PO site–located near Portland, Oregon– was also underrepresented consisting of just two trees. Every other sampling location contained at least eight. In order to decrease the chance of sampling closely related trees, alders sampled were at least 100m apart. Fresh leaves and buds were collected and immediately put on ice to be returned to the lab at Portland State University within 24 hours. Leaves were then lyophilized and stored for later use for GBS.

DNA extraction and GBS library preparation

DNA was extracted from lyophilized leaf material using a modified CTAB procedure **(**Tel-Zur *et al*. 1999). Briefly, ~20mg of dried leaf tissue was added to sterile Eppendorf tubes. A tungsten bead was added to each tube and tissue was broken down in two 1-minute cycles at 30 Hz on a Qiagen TissueLyser II. 400µl of CTAB buffer was

added and the cycles were repeated at 20 Hz and then incubated at 65 °C for 1 hour. Samples were centrifuged for 10 minutes at 4000 rpm and 140 µl of the CTAB buffer was transferred to a new tube where it was mixed 1:1 with chloroform isoamyl alcohol. Samples were centrifuged for 10 minutes at 4000 rpm and the aqueous supernatant was removed and transferred to a new tube. Samples were then washed and resuspended overnight twice with ethanol. Lastly, the samples were centrifuged for 10 minutes at 4000 rpm, the ethanol was removed and samples were dried for 10 minutes in a Thermo Savant Speedvac SC210A and resuspended in TE buffer. Electrophoresis was performed on a 1% agarose gel to assess DNA quality, and DNA concentrations were determined using a Invitrogen Qubit 3.0 Fluorometer. DNA concentrations were normalized before GBS library preparation of the 96 multiplexed samples according to the GBS protocol developed by Elshire *et al*. (2011). The 5 base pair single cutting restriction enzyme ApeKI was chosen for paired-end sequencing to be conducted at the University of Wisconsin-Madison Biotechnology facility using an Illumina HiSeq 3000 flowcell.

Sequence analysis and SNP filtering

The bioinformatics pipeline GBS-SNP-CROP (Melo *et al*. 2016) was used for SNP processing. Briefly, paired end read files were matched according to their barcode and maintained for further processing. Reads that were lacking an identified barcode were discarded. Raw reads were trimmed using Trimmomatic v0.39 (Bolger *et al*. 2014). The retained reads were then demultiplexed into separate FASTQ files for both forward and reverse reads of each genotype. Paired-end reads were then merged into a single read

using PEAR v0.9.11 (Zhang *et al*., 2014). Redundant reads were discarded using VSEARCH v2.13.7 (Rognes *et al*. 2016). The centroids clustering algorithm was utilized to stitch together the merged reads into a mock genome for reference and SNP alignment. The BWA-mem algorithm (Li & Durbin, 2009) was used to align the reads, genotype-by-genotype, to the reference genome. SAMtools (Li *et al*. 2009) filtered the mapped reads, and converted and indexed the files to the reference for each genotype. Lastly, the SAMtools mpileup algorithm was used for identification of all potential SNPs based on the individual alignments of each separately selected genotype. Files were then parsed so only those data containing reads polymorphic to the reference sequence were retained with alignment information for potential variants extracted and organized for downstream processing. Initial filtering was conducted using the criteria suggested in the GBS-SNP-CROP manual (Melo *et al*. 2016). and files were converted to TASSEL formatting for more filtering. TASSEL v5.0 (Bradbury *et al*. 2007) was used to eliminate one of the samples from the DE sampling site because of extensive missing sequencing data. Only SNPs that were present in 90% of the samples were retained and all SNPs with an MAF<0.01 were discarded. Paired-end sequencing yielded 357,322,720 raw reads (178,661,360 for each paired-end) with an average GC content of 47.5%. Paired-end files were scanned for identifying barcodes and the reads without the appropriate barcode designation were discarded retaining ~90% of the raw reads. After final SNP calling and filtering on the remaining 95 samples, 5,454 high quality SNPs were retained for analysis of population structure and molecular variance.

Population structure analysis

Population structure was assessed using a Bayesian Markov Chain Monte Carlo model implemented in STRUCTURE v2.3.4 (Pritchard *et al*. 2000). STRUCTURE was chosen because it has been shown to be robust enough to handle data of polyploid and mixed ploidy SNP data (Merimans 2018). Three runs were conducted for each number estimated population (k) 1-8 in both the admixture and no admixture models. Burn-in time and replication number were set for 100,000 for each run. To determine the run with the most likely k value, the output of the STRUCTURE runs was assessed using CLUMPAK (Kopelman *et al*. 2015). To compliment STRUCTURE analysis, the adegenet package (Jombart & Ahmed, 2011) in Rstudio v3.5.0 was used to asses potential population clusters using discriminant analysis of principal components (DAPC; Jombart *et al*. 2010). Because DAPC is based on within-individual allele frequencies and does not make assumptions about Hardy-Weinberg equilibrium or panmixia, it is easily applied to polyploid and mixed ploidy data (Merimans 2018). DAPC initially transforms data using Principal components analysis (PCA), then it uses the cross-validation method to perform a discriminate analysis on the retained principal components. Based upon the results of DAPC analysis, pairwise genetic differentiation among the subpopulations was estimated with F_{ST}. Expected heterozygosity (H_E), observed heterozygosity (H_O), and analysis of molecular variance (AMOVA) were all calculated in Genalex v 6.503

(Peakall & Smouse 2007) with 999 permutations for testing variance components. Corresponding G statistics were also calculated in Genalex.

Estimated divergence

A timed phylogeny ran in BEAST v2.6.2 (Drummond *et al*. 2012) was used to estimate time of divergence among inland red alders. BEAST uses a Bayesian Markov Chain Monte Carlo model to estimate time-scaled phylogenies of the species being analyzed. Evidence is convincing for at least some recent dispersal into the NRM since the ice sheets retreated. Because of this, we used a range of 20,000-24,000 (Martinson *et al*. 1987) years as the time of the most recent divergence to coincide with the end of the LGM. If the RD hypothesis is accurate, all the inland samples should have diverged recently, and at roughly the same time. If there is ancient vicariance, we expect to see divergence times at more than 2mya. If there is vicariance and multiple dispersal events, we should be able to estimate the times of the divergences and attempt to find historical processes that may have driven these dispersals. A strict clock model was used and the program was run with chain lengths of 10,000,000 states with sampling every 1,000 states (Balkan *et al*. 2020). Maximum clade credibility trees were generated in FigTree v 1.4.4 (Rambaut 2010).

Ploidy analysis

To detect individual ploidy levels, we used the gbs2ploidy (Gompert & Mock 2017) package in Rstudio v3.5.0. This package was designed to estimate ploidy from allelic heterozygosity ratios produced in the GBS process. The heterozygosity ratios were

generated in the GBS-SNP-CROP output and formatted for use in R in Excel. The estprops function was used to estimate the ploidy for each individual. The results were plotted with the mean probability for allelic ratios on the y and the 3:1, 2:1, and 1:1 ratios on the x axis. This process was repeated with corresponding ratios for higher ploidy as well (up to 12:1, however this did not change any of the results. Gbs2ploidy was designed to be able to function at low read depth <15x and improves in accuracy as read depth increases, however our data did not suffer from a lack of read depth with samples generally containing read depth of ~200x which suggests ploidy assignment was accurate. Since multiple ploidy levels were detected, a small subset of 12 red alders were subjected to chromosome counts via florescent microscopy to confirm the gbs2ploidy results using 2n=14 (Wetzel 1929) as the diploid chromosome count, and 2n=4x=28 for tetraploids.

Results

Population structure analysis

Analysis of the STRUCTURE output there indicted a disagreement on the optimal number of clusters between the two methods used for K estimation. The ln Pr(X|*K*) (Pritchard *et al*. 2000) method showed peaks at both K=4 and K=6 (2.1A) and the Evanno Δ*K* method (Evanno *et al*. 2005) indicated 4 as the optimal number of clusters (figure. 2.1B). When STRUCTURE models data with populations that have undergone bottlenecks similar scenarios of drift, it is not uncommon for non-existent population admixture to be inferred (Lawson *et al*. 2018). If that was the case in our data, then the inferred populations 5 and 6 would only be found in the populations that would have

undergone the bottleneck (e.g. the inland populations). Indeed, that is exactly what we found with the two extra inferred clusters being fit in the Idaho admixed cluster only. For that reason, the peak at K=6 was disregarded. The pairwise F_{ST} and AMOVA results confirm the differentiation among the populations predicted by STRUCTURE and DAPC. Of the 95 alders sampled, 84 were assigned to one cluster which contained all of the coastal samples (CN; hereafter the coastal northern) with the exception of one (Figure 2.1; Figure 2.2). Sample Rt-2 was the only coastal sample assigned to a different Cluster (CS; hereafter the coastal southern), and Rt-1 and Rt-11 showed some admixture with the southern and northern clusters. A distinct Idaho cluster (IO; hereafter the Idaho cluster) was identified with Id-2, ID-6, Id-7, and Id-8 being placed entirely within this cluster and showing little or no admixture which suggests an ancient Idaho population with no gene flow from the coastal populations. The remaining Idaho samples indicate admixture (IA; hereafter Idaho admixed) with the Idaho cluster, and both the southern and northern sub-populations which suggests gene flow between the coastal and inland Idaho sub-population. The no-admixture model produced fairly similar results to the admixture model with the same 84 samples being assigned to the northern coastal cluster (figure 2.3). The difference in the two models was in the remaining 11 trees not assigned to the northern cluster. There was still a distinct Idaho cluster consisting of four trees and the rest of the Idaho samples were separated into two clusters. Rt-2 which was originally assigned to the coastal southern cluster in the previous model is now included with several trees originally assigned to the Idaho admixed. The rest of the Idaho trees were assigned to a cluster that was admixed with the coastal northern cluster in the previous model which suggests that the no-admix model assigned trees from the Idaho sites based

which coastal cluster migrated into the NRM. The STRUCTURE results are backed up by the high F_{ST} /G_{ST} between the Idaho population and the two coastal populations supports a distinct differentiation consistent with group that has been isolated for a long period of time (Table 2.1). The Idaho admixed cluster has a similar F_{ST} from both the Idaho cluster and the Coastal northern cluster providing evidence admixing between both coastal and the Idaho admixed populations.

DAPC analysis generally agreed with the STRUCTURE results also indicating the optimal number of clusters of K=4. Individuals were assigned to the identical coastal northern (CN), southern (CS), and Idaho (IO) clusters predicted by STRUCTURE (Figure 2.4) with the six Idaho samples that were classified as admixed by STRUCTURE being assigned their own cluster (IA). The inland populations (CN,CS) were clustered more closely together than the IO cluster with cluster IA occurring intermediate to the other clusters.

AMOVA was performed to determine how much of the total genetic variation could be attributed to the hierarchical levels of the total population. Genalex does not allow AMOVA to be calculated if the number of individuals in a population is one so population CS was not included in this calculation. With mixed ploidy populations the p statistic is most useful in AMOVA calculations (Meriman 2019), so they have been included as well. The AMOVA p value results confirm the existence of multiple populations (Table 2.2). Among population variation accounted for 55% of the total variation seen. Within individual variation accounted for 40%, and the remaining 5% of was attributed to variation among individuals.

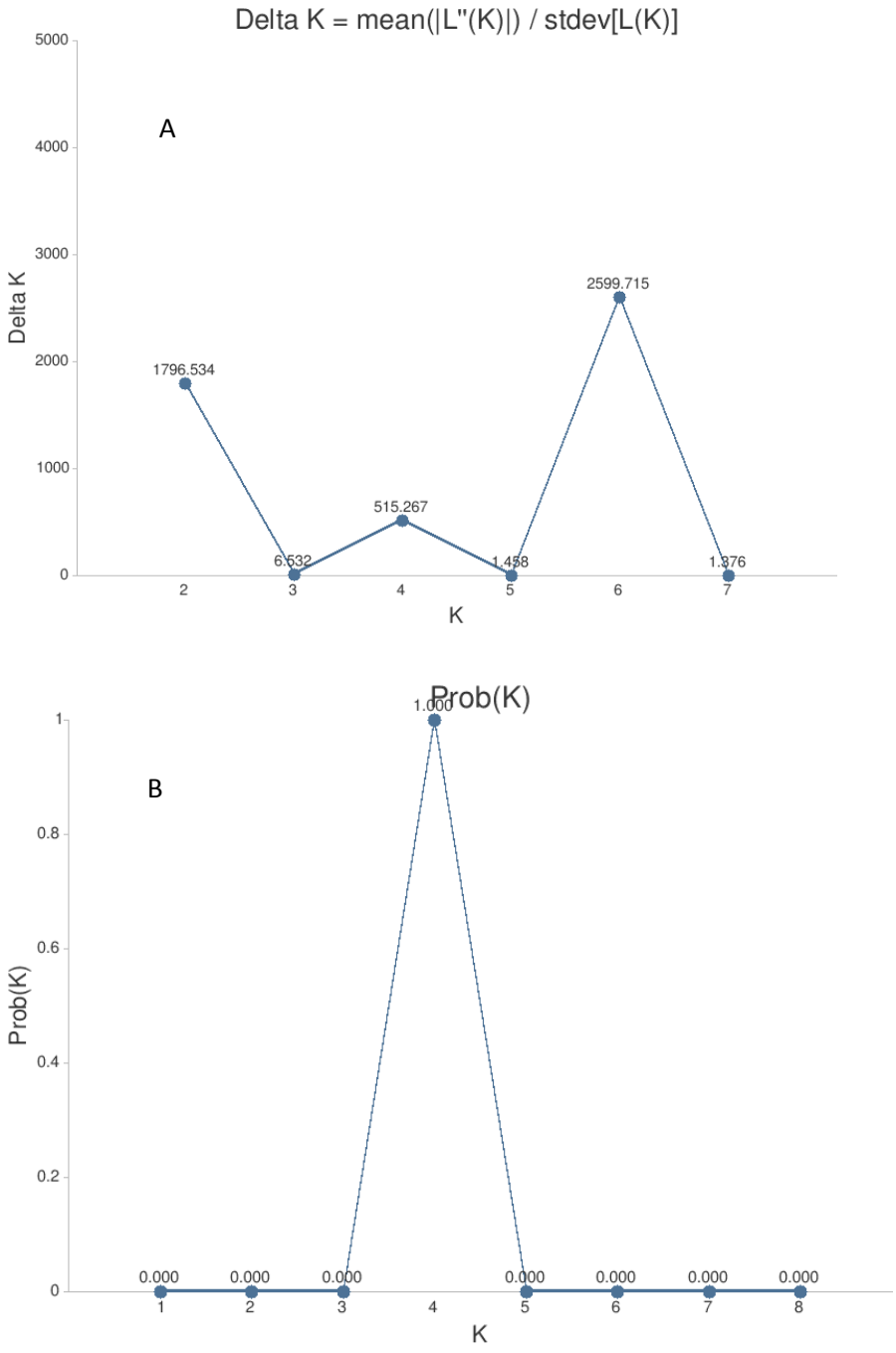


Figure 2.1. Number of Sub-Populations Identified in CLUMPAK. Two separate CLUMPK outputs are included identifying the correct number of sub-populations (K) based on STRUCTURE analysis of the high-density SNP profile. Four or six sub-populations were initially suggested (A), and a second analysis identified four as the correct number of K with a 100% probability.

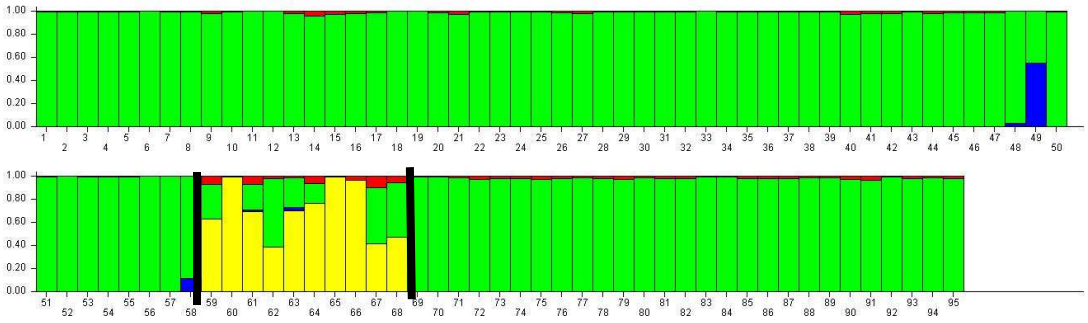


Figure 2.2. STRUCTURE Output Identifying Four Sub-Populations. Colors indicate the different assigned clusters: CN=green, CS=blue, IO=yellow, mixed color bars indicate admixture. Disjunct samples are contained within the black lines.

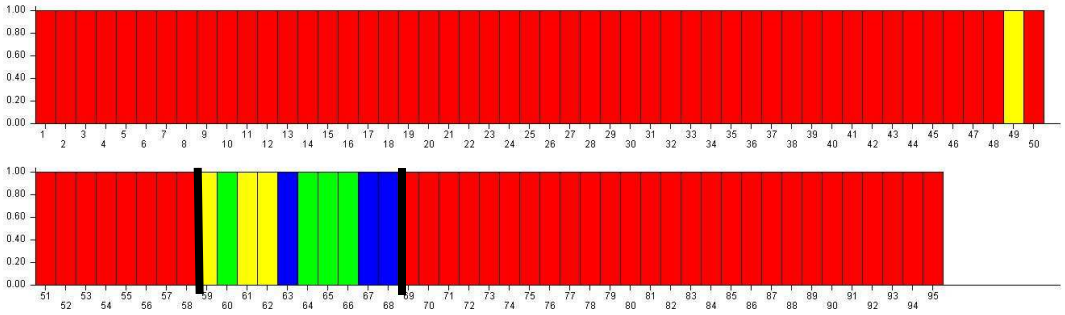


Figure 2.3. STRUCTURE Output of the No-Admixture Model. Colors indicate the different assigned cluster: CN=red, CS=yellow, IO=green, IA=blue. Disjunct samples are contained within the black lines.

Estimated divergence

The timed phylogeny produced by BEAST and viewed in FigTree indicated that 4 of the 10 inland samples diverged recently (ID-1, ID-2, ID-4 ID-9), and at roughly the same time (~20 kya) which suggests that some of the inland samples are a product of RD. The model also identifies a distinct Idaho individual (Id-8) that diverged and evolved

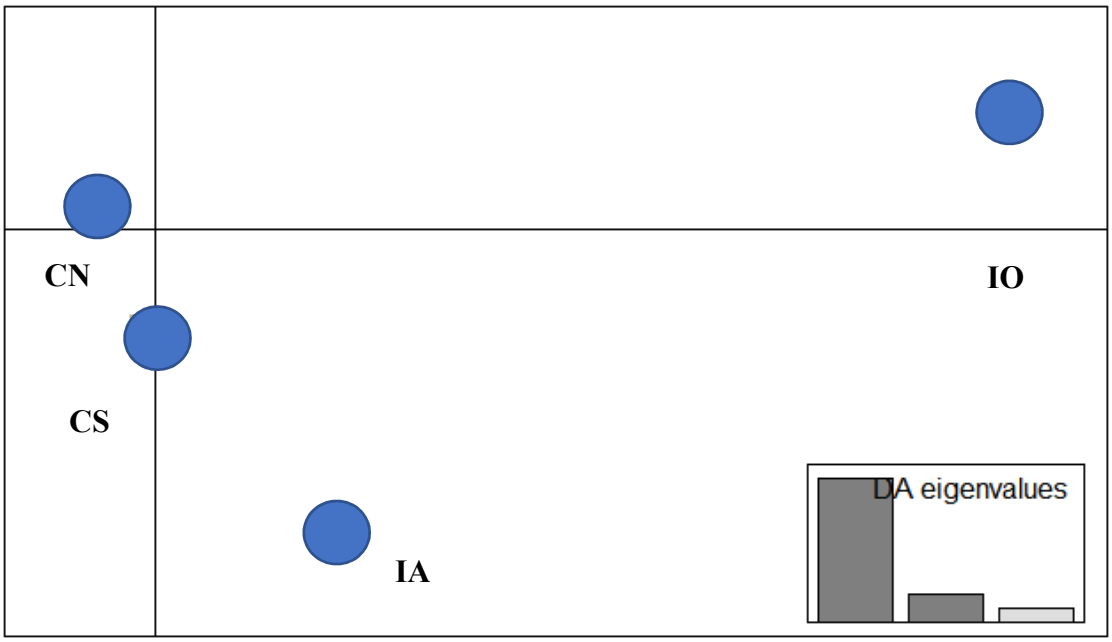


Figure 2.4. DAPC Indicating the Four Identified Sub-Populations and their relation in ordination space. Eigen values are shown for the first 100 principle components. Note the two coastal sub-populations (CN, CS) are clustered more closely to each other than the inland sites (IA, IO) suggesting they are more closely related. The IA sub-population is intermediate to the coastal and the IO sub-populations which is consistant with admixture among the coastal and inland sub-populations.

separately much earlier (Figure 2.5). The estimated date range for the divergence of this alder is 2.34my-2.8my which suggests AV. The remaining 5 inland alders showed a range of dates of divergence suggesting multiple inland dispersals throughout the Quaternary. Id-6 also showed divergence estimated to have happened 872k-1.16mya. Both Id-8 and Id-6 were assigned to the same Idaho only cluster in STRUCTURE and showed no evidence of admixture or gene flow between any of the other individuals sampled. The divergence estimates of the remaining admixed populations suggest separate dispersal events at 220k-260kya (Id-5), 55k-90kya (Id-3, Id-7, Id-10).

Interestingly, individuals from both the northern and southern coastal clusters contributed to the dispersal (Figure 2.5A; simplified 2.5B). This suggests that red alders were able to disperse inland from the coast through both the northern and southern route.

Table 2.1. F/G statistics from GENALEX among the four identified sub-populations. Pairwise F_{ST} values show the coastal sub-populations (CN, CS) are minimally differentiated (F_{ST}=0.004). The inland IO site was significantly differentiated (F_{ST}=0.302). Significant population structure was identified (Overall F_{ST}=0.349). Heterozygosity was much higher than expected as indicated by the highly negative F_{IS} (F_{IS}= -0.365).

	Fst pairwise				**F/G statistics**	
	CN	IO	CS	IA	Overall	
CN	0	0.301764	0.003552	0.157106	**Gis**	-0.282
IO	0.301764	0	0.148172	0.133241	**Gst**	0.317
CS	0.003552	0.148172	0	0.0229	**Fis**	-0.365
IA	0.157106	0.133241	0.0229	0	**Fst**	0.349

Genetic diversity

Genetic diversity was assessed based on the four groups identified in the DAPC and STRUCTURE analysis. Because of the small sample size of the southern coastal cluster, the allele frequencies are not known with high accuracy so analysis should be considered accordingly. Overall F-statistics and G-statistics both showed H_O consistently being higher than H_E which indicates excessive heterozygosity as shown with the negative F_{IS} (Table 2.1). An overall F_{ST} of 0.349 confirms that there is significant population structuring. H_O and H_E were also calculated for each population individually (Table 2.1) and all four of the populations were consistent with the overall trend with a negative F_{IS}. Pairwise F_{ST} values showed that the coastal northern cluster (CN) is only minimally differentiated from the coastal southern (CS) population which suggests weak population structure. The Idaho population (IO) shows strong differentiation from the

Table 2.2. Table Showing Results From AMOVA Analysis of Our High-Density SNP profile. Variance was tested among populations (55%), among individuals (5%), and within individuals (40%).

Source	df	SS	MS	Est. Var.	%
Among Pops	2	428.359	214.18	11.114	55%
Among Indiv	91	898.774	9.877	0.909	5%
Within Indiv	94	757.5	8.059	8.059	40%
Total	187	2084.633		20.081	100%

costal northern (F_{ST}=0.301) which suggests an extended period without gene flow. The Idaho population showed moderate differentiation between the Idaho admixed (IA) and the coastal northern (F_{ST}=0.148) and coastal southern (F_{ST}=0.133). The similarity in the F_{ST} is consistent with admixture among the populations.

Ploidy determination

Both diploid and tetraploid individuals were identified in the remaining 95 red alder samples analyzed by GBS2ploidy (Table 2.2). Due to the small size of red alder chromosomes and challenging cell preparation it was difficult to obtain countable metaphase plates, however we were able to confirm GBS2ploidy results with chromosome counts on 12 of the 95 red alders (Figure 2.6; appendix A). Both diploids and tetraploids were identified in all sampling location apart from the Portland (Po) site which only had tetraploids. Overall, the distributions of ploidy levels were evenly distributed with 53.7% of the samples being diploid and 46.3% being tetraploid. Only three of the sites that had more tetraploids than diploids and one of them was the Po site which only had two trees sampled. Of the remaining sites, only the Ws and Id had more tetraploids than diploids.

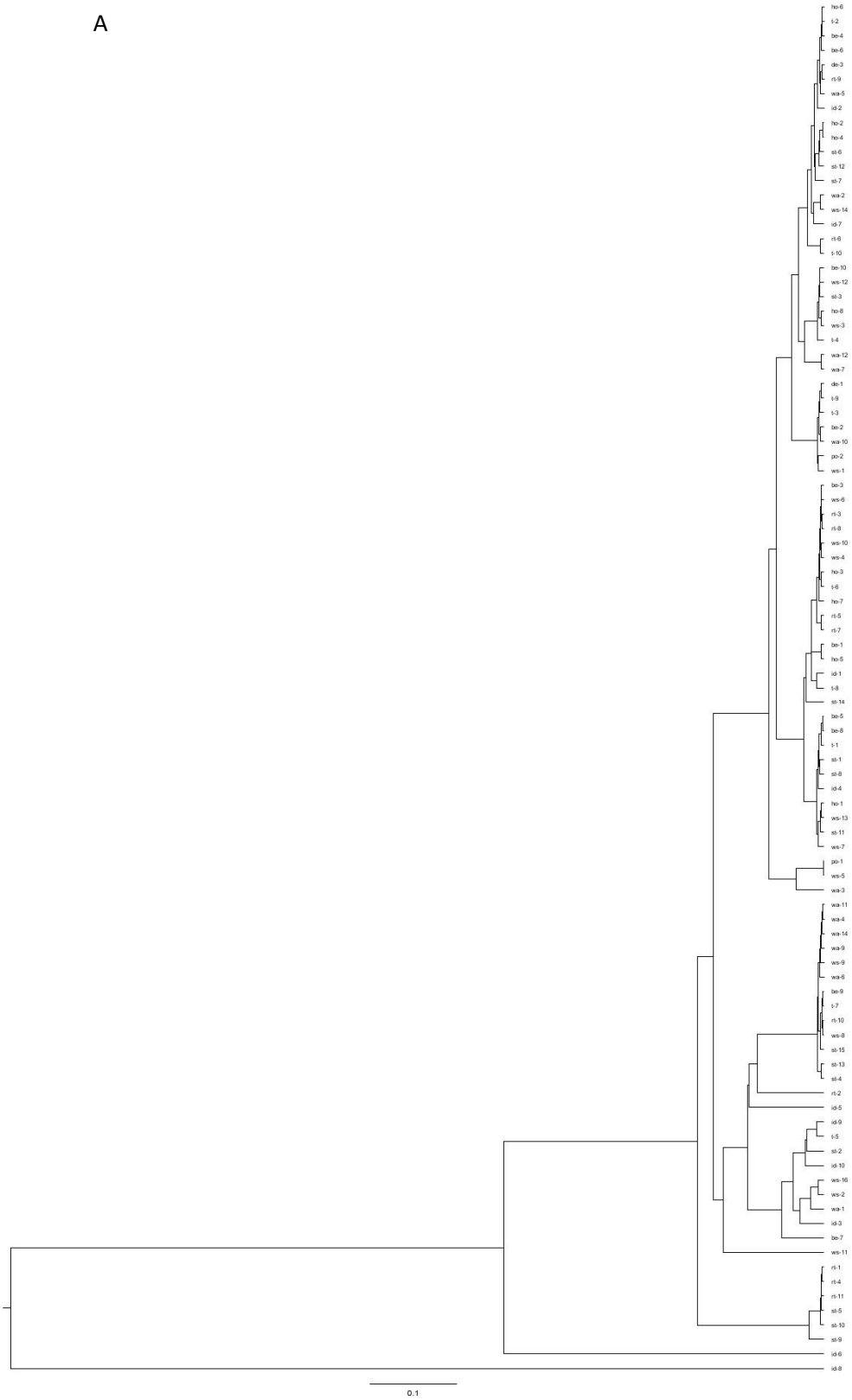
Table 2.3. Ploidy Estimates for 95 Alder Samples From Gbs2ploidy. The number and percent of sampled red alders total per site, plus overall totals for both diploid and tetraploid cytotypes sampled in the Pacific Northwest.

Site	N Diploid	% diploid	N Tetraploid	% Tetraploid
Be	6	60	4	40
De	1	50	1	50
Ho	5	62.5	3	37.5
Id	4	40	6	60
Po	0	0	2	100
Rt	6	54.5	5	45.5
St	8	53.3	7	46.7
Ti	8	80	2	20
Wa	8	66.6	4	33.3
Ws	5	33.3	10	66.6
Total	51	53.7	44	46.3

Discussion

Population structure for red alder was investigated using GBS from 95 red alders sampled throughout the PNW and the Idaho NRM. Analysis of our STRUCTURE results identified population structure in both coastal and inland sampling sites. This is the first-time structure has been identified in coastal red alder populations in the PNW, and the first time multiple sub-populations have been identified for red alder in the NRM. The observed heterozygosity in the northern and southern coastal sub-populations was lower than that of other members of the genus and particularly low for an outcrossing tree species; however, they were very similar to previous studies of red alder in the northern coastal area (Xie *et al*. 2002). Xie *et al*. (2002) also had similar levels of observed heterozygosity as seen in our coastal sites but were both lower than that identified in the

A


ho-6
li-2
be-4
be-6
de-3
ri-9
wa-5
id-2
ho-2
ho-4
st-6
st-12
st-7
wa-2
ws-14
id-7
ri-6
li-10
be-10
ws-12
st-3
ho-8
ws-3
li-4
wa-12
wa-7
de-1
li-9
li-3
be-2
wa-10
po-2
ws-1
be-3
ws-6
ri-3
ri-8
ws-10
ws-4
ho-3
li-6
ho-7
ri-5
ri-7
be-1
ho-5
id-1
li-8
st-14
be-5
be-8
li-1
st-1
st-8
id-4
ho-1
ws-13
st-11
ws-7
po-1
ws-5
wa-3
wa-11
wa-4
wa-14
wa-9
ws-9
wa-6
be-9
li-7
ri-10
ws-8
st-15
st-13
st-4
ri-2
id-5
id-9
li-5
st-2
id-10
ws-16
ws-2
wa-1
id-3
be-7
ws-11
ri-1
ri-4
ri-11
st-5
st-10
st-9
id-6
id-8
0.1

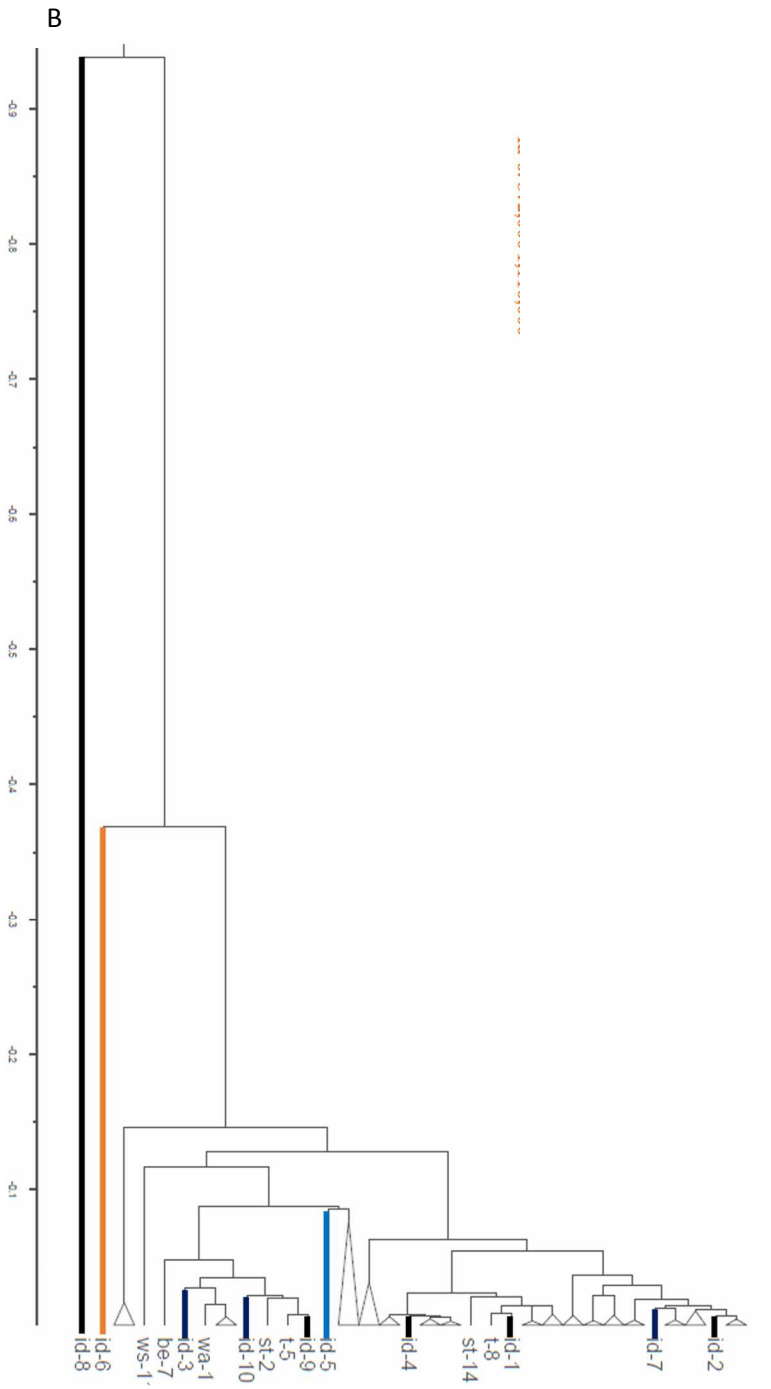


Figure 2.5. BEAST Output Showing Estimated Dates of Divergence (A) with a reduced version (B) so dates are more easily identified. Idaho samples and dates of divergence are identified by colored bars indicating five separate dispersal events into the inland sites. ID-2,ID-9,ID-1,ID-4 are estimated to have diverged ~20kya and represent the most recent divergence. ID-7, ID-10, ID-3 are estimated to have diverged ~60kya. Id-5 is estimated to have diverged ~220kya, Id-6 ~872kya, and Id-8 ~2.4mya suggesting ancient vicariance.

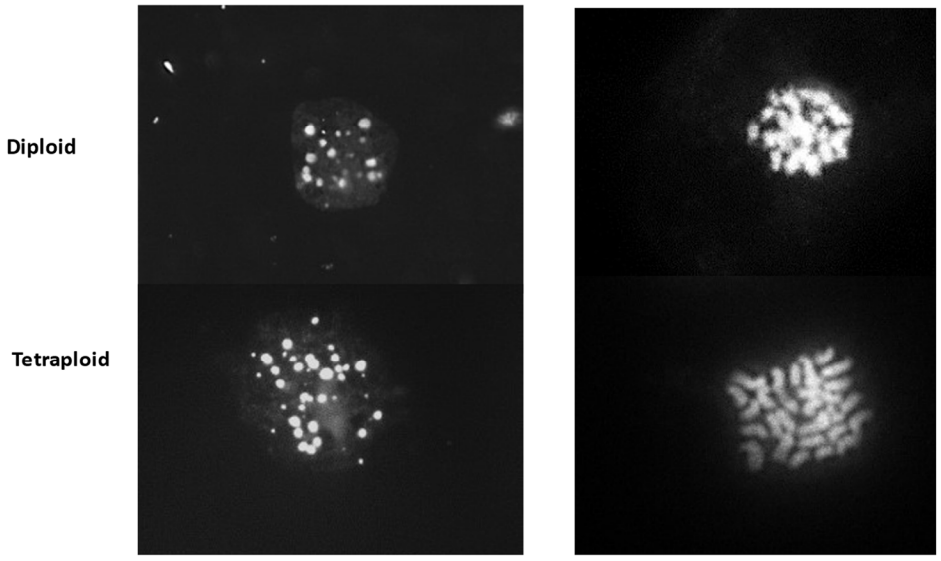


Figure 2.6. Chromosome Counts Visualized with Fluorescent Microscopy. Chromosome counts of metaphase plates of red alder using two different preparation methods: Chromosome squash (left), steam drop (right). Both Diploid (top), and tetraploid (bottom) cytotypes were identified.

Idaho, and Idaho admixed populations. This is likely due to multiple factors. First, pioneer species in general have lower rates of heterozygosity than late successional species (Hamrick *et al*. 1981). Second, pioneer species are overly influenced by physical factors (e.g. fresh disturbance, poor soils), that, coupled with their colonizing abilities, can reduce genetic variation (Hamrick *et al*. 1981; Xie *et al*. 2002). Last, the isolation within glacial refugia would have greatly reduced the effective population size leading to fixation of alleles via genetic drift. The Idaho sub-population having a higher observed heterozygosity than the coastal populations was surprising because it would have likely suffered from an extreme bottlenecking event once it became established. We suggest

several possible reasons for this observation. First, multiple migrations into the NRM can explain the much higher observed heterozygosity in the Idaho sub-population compared to the other three by being able to replenish the gene pool and provide alternate alleles for the ones that had already been fixed. Second, polyploid populations are expected to have and maintain higher levels of heterozygosity than their diploid progenitors due to polysomic inheritance (Haldane 1930; Moody *et al*. 1993; but see Soltis & Soltis 2000). Third, the inland populations could have increased selfing rates to help reduce the rate of fixation and increase heterozygosity. Selfing rates are thought to increase in colonizing species, particularly polyploids. Red alder is generally a highly outcrossing species with an outcrossing rate of ~0.85, however there can be significant variation. For example, outcrossing rates were consistently higher in red alder when comparing mainland to island populations with some of the island alder stands having outcrossing rates reduced to 0.56 (Xie *et al* 2002).

At least a dozen species of plants and animals that have been studied for vicariance in the PNW, all indicate either AV, or a recent RD into the NRM through a northern route with two exceptions. One amphibian showed AV but also indicated RD through a northern route (Brunsfeld 2001) and red alder which showed evidence of RD through a southern route (Strenge 1994; Brunsfeld 2001). Red alder is the only NRM species studied so far to indicate evidence of dispersal by a southern route, and it is also the only species to show evidence for multiple dispersal events. We suggest past environmental cataclysms can help explain why this is the case. Phylogeography studies in the PNW have focused mainly on habitat availability in terms of glacial refugia.

During the warming periods ending the 100ky glacial cycles the landscape in the PNW faced dynamic change. One aspect of this dynamic change was the cataclysmic flooding that occurred as the glaciers cyclically advance and retreat. Flooding from lake Missoula at the end of the LGM is well studied (Baker & Bunker 1985; Smith 1993). Periodically an ice dam would form on the Clark Fork River in northern Idaho creating the massive Lake Missoula. Occasionally the ice dam would rupture causing massive devastating flooding throughout much of eastern Washington and Oregon. The dam would then reform recreating lake Missoula and the cycle would repeat itself. It is suspected that 40-60 floods occurring in roughly 50-year increments occurred over a 2k period at the end of the LGM (Brunner *et al*. 1999). Flooding likely occurred from other sources throughout the last 2.5my as well. The Rocky Mountain Trench in Canada, and a massive subglacial reservoir in central British Columbia have been proposed as possible sources of flooding (Shaw 1999). Isotope records indicate a glacial-interglacial cycle of varying intensities over the last 2.6my, and conditions conducive to cataclysmic flooding would have occurred at least two thirds of the time (Mandole 1995). Much of the evidence for flooding before this time has been destroyed by subsequent flooding and erosion, however studies suggests that cataclysmic flooding events occurred 55-90kya (McDonald & Busacca 1992; Berger & Busacca 1995), 200-400kya, 780k-1.7mya, and there is even evidence suggesting cataclysmic flooding as far back as 2.6mya (Bjornstad *et al* 2001; Medley 2012). Interestingly, all of our estimated divergence times in the NRM correlates with the estimated times of these floods. We suggest that these cataclysmic flooding events could help describe the current distributions of red alder in the NRM by periodically creating viable corridors for inland red alder migration from both the

northern and southern refugia. Recurring flooding would produce massive amounts of freshly disturbed riparian areas throughout eastern WA and OR that would be prime habitat for red alder to colonize. It should be noted that the primary areas in the NRM where red alder is currently found are contained within the Missoula flood plain which lends support to the idea.

The periodic flooding and reforming of lake Missoula after the LGM lasted for a period of 2k years. Considering trees are sedentary and that most seeds fall within a few meters of the parent, 2k years may not have been enough time for individuals to disperse from the coast to the NRM. While there is a dearth of information on the dispersal ability of red alder, it is assumed to be similar to other members of Betulaceae (Zasada *et al*. 1992). Red alder seeds are winged nutlets, small, and are formed in pairs on the bracts of woody strobili that is small and cone-like making them ideal for long distance wind dispersal (Schopmeyer 1974). For comparison, the winged nutlet seeds of the related yellow birch (*Betula alleghaniensis*) have been shown to travel up to 400 m over crusted snow (Erdmann 1990). Korstian (1937) estimated that roughly 50% of the seeds from a 15 m tall yellow birch would disperse farther than 200 m. Considering red alders have been shown to produce more than 5 million seeds per individual (Brown 1985), if red alder followed a similar pattern as yellow birch, even if only 1% of red alder seeds reached the outside 400 m barrier it would still constitute ~50 k seeds. Black alder (*Alnus glutinosa*) lacks winged nutlets making it less efficient at wind dispersal than other alders, even so, in Europe black alder expanded and dispersed from glacial refugia extremely rapidly after the LGM with estimates from fossil data suggesting a dispersal

rate of 500-2000 m/year (Huntley & Birks 1983). This expansion was likely facilitated by water dispersal, which has been shown in red alder as well (Brown 1986). Because water would flow from the NRM towards the coast it is unlikely that water dispersal played a large role for red alder in our study, but it does show that alders can disperse rapidly when conditions are suitable. Dispersal by birds may have also facilitated the advance of red alder through the flood plain into the NRM. Birds can be passive or active seed dispersers by either disrupting catkins within the trees, seeds floating in water may attach to the feathers of waterfowl, or birds may distribute seeds by direct ingestion. Red alder seed have been shown to be an important food source for the American goldfinch (*Spinus tristis;* White & West 1977). The American goldfinch is a seasonal migrant that thrives in flood plains, meadows, and in riparian systems. This habitat preference continues during migration periods and both the PNW and the NRM are in in the goldfinch migration zone (Dougherty *et al*. 2018). Assuming previous flood cycle to be similar to the cycle ending in the LGM, there should have been plenty of time for red alder to have advanced to the NRM. Species distribution models by Ruffley *et al*. (2018) also suggest that areas east of the Cascades that are currently too arid for red alder due to the rain shadow may have been more suitable during peak glacial cycles. This would potentially allow for island red alder populations to persist in parts of eastern Washington and Oregon when they otherwise would have died off further facilitating the dispersal of the southern genotype into the NRM.

Analysis of the variation in ploidy in red alder identified both diploid and tetraploid cytotypes with a relatively even distribution throughout the Pacific Northwest

and in the NRM. Large mixed cytotype zones extending over extensive land areas tend to be rare (Castro *et al*. 2012). This is largely due to problems that polyploids encounter early on in their establishment. The establishment of a stable polyploid population is constrained by a low probability of early polyploids persisting among their diploid progenitors by both direct competition (Yamauchi *et al*. 2004), and frequency dependent selection (minority cytotype exclusion; Levin 1975). This states that barriers in mating and low fitness of triploid intermediate cytotypes makes it more likely that polyploids will go extinct before establishment (Levin 1975; Husband 2000). This is does not exclude polyploids from becoming established, and many large mixed cytotype populations have been identified (Halverson *et al*. 2008; Duchoslav *et al*. 2010). New research has shown several ways in which nature has been able to counter difficulties in early establishment and recruitment (see Baduel *et al*. 2018 for a review). Interestingly, Husband (2000) showed that the higher the percentage of tetraploids in a given population, the more fit the tetraploid cytotype becomes until tetraploids eventually its fitness is higher than that of diploids at ~67%. The percentage of tetraploids in our population was split closer to 50/50 which suggests that diploids should still hold a marginal fitness advantage over tetraploids, however life history may help overcome any fitness advantage diploid maintain. Pioneer species are frequently polyploid and are similar to invasive species in that polyploidy is considered making them better adapted to new habitats and disturbed areas (Hobbs & Huenneke 1992; Te Beest *et al*. 2012). The ability of polyploids to tolerate changing conditions as the glaciers retreated and newly available habitat became available would certainly be advantageous. Indeed, glaciation and the subsequent retreat during the LGM are also suggested to have played a role in the

distribution of tetraploid populations of black alder (*Alnus glutinosa)* in Europe (Mandák *et al*. 2016). Projections on the future distributions of red alder have suggested that climate change will only expand their range, particularly in the disjunct NRM (Cortini *et al*. 2012). If their projections are accurate, it seems likely that polyploidy in red alder is also likely to expand as the climate warms and disturbed habitat through development is continuously created.

References

Avise, J. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics. Vol. 18*, (c), 489–522. https://doi.org/10.1146/annurev.ecolsys.18.1.489

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard university press.

Baduel, P., Bray, S., Vallejo-Marin, M., Kolávr, F., & Yant, L. (2018). The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution*, *6*, 117.

Bagley, J. C., Heming, N. M., Gutiérrez, E. E., Devisetty, U. K., Mock, K. E., Eckert, A. J., & Strauss, S. H. (2020). Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (Populus tremuloides). *Ecology and Evolution*, *10*(11), 4609–4629. https://doi.org/10.1002/ece3.6214

Baker, V. R., & Bunker, R. C. (1985). Cataclysmic late Pleistocene flooding from glacial Lake Missoula: A review. *Quaternary Science Reviews*, *4*(1), 1–41.

Balkan, M. A., Stewart, N. U., Kauffman, E. S., Wolfe, E. R., & Ballhorn, D. J. (2020). Genotypic Diversity and Host-Specificity of Frankia Bacteria Associated with Sympatric Populations of Alnus rubra and Alnus rhombifolia in Oregon. *Northwest Science*, *93*(3–4), 244–252.

Berger, G. W., & Busacca, A. J. (1995). Thermoluminescence dating of late Pleistocene loess and tephra from eastern Washington and southern Oregon and implications for the eruptive history of Mount St. Helens. *Journal of Geophysical Research: Solid Earth*, *100*(B11), 22361–22374.

Bjornstad, B. N., Fecht, K. R., & Pluhar, C. J. (2001). Long history of pre-Wisconsin, ice age cataclysmic floods: evidence from southeastern Washington state. *The Journal of Geology*, *109*(6), 695–713.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brown, S. M. (1985). A study of reproductive biology of Alnus rubra along three elevational transects in Washington and Oregon. *On File with: USDA Forest Service, Pacificth Northwest Research Station, Olympia, WA*.

Brown, S. M. (1986). *Sexual Allocation Patterns in Red Alder (Alnus Rubra Bong.) Along Three Elevation Transects*. University of Washington.

Brunner, C. A., Normark, W. R., Zuffa, G. G., & Serra, F. (1999). Deep-sea sedimentary record of the late Wisconsin cataclysmic floods from the Columbia River. *Geology*, *27*(5), 463–466.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2000). Comparative phylogeography of north- western North America : a synthesis. *Integrating Ecological and Evolutionary Processes in a Spatial Context*, 319–339.

Carstens, B. C., Stevenson, A. L., Degenhardt, J. D., & Sullivan, J. (2004). Testing nested phylogenetic and phylogeographic hypotheses in the Plethodon vandykei species group. *Systematic Biology*, *53*(5), 781–792. https://doi.org/10.1080/10635150490522296

Castro, S., Loureiro, J., Procházka, T., & Münzbergová, Z. (2012). Cytotype distribution at a diploid--hexaploid contact zone in Aster amellus (Asteraceae). *Annals of Botany*, *110*(5), 1047–1055.

Cortini, F., Comeau, P. G., Wang, T., Hibbs, D. E., & Bluhm, A. (2012). Climate effects on red alder growth in the Pacific Northwest of North America. *Forest Ecology and Management*, *277*, 98–106.

Dougherty, P. J., & Wilson Jr, W. H. (2018). Evidence for a relationship between the movements of the Common Redpoll (Acanthis flammea) and the American Goldfinch (Spinus tristis). *The Open Ornithology Journal*, *11*(1).

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Duchoslav, M., Šafá\vrová, L., & Krahulec, F. (2010). Complex distribution patterns, ecology and coexistence of ploidy levels of Allium oleraceum (Alliaceae) in the Czech Republic. *Annals of Botany*, *105*(5), 719–735.

Ehrendorfer F. (1980) Polyploidy and Distribution. In: Lewis W.H. (eds) Polyploidy. Basic Life Sciences, vol 13. Springer, Boston, M. https://doi. org/10. 1007/97.-1-4613-3069-1_3.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Erdmann, G. G. (1990). Betula alleghaniensis Britton. yellow birch. *Silvics of North America*, *2*, 133–147.

Espíndola, A., Ruffley, M., Smith, M. L., Carstens, B. C., Tank, D. C., & Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1841). https://doi.org/10.1098/rspb.2016.1529

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Furlow, J. J. (1990). The genera of Betulaceae in the southeastern United States. *Journal of the Arnold Arboretum*, *71*(1), 1–67. Retrieved from http://www.jstor.org/stable/43782230

Gompert, Z., & Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Molecular Ecology Resources*, *17*(6), 1156–1167. https://doi.org/10.1111/1755-0998.12657

Haldane, J. B. S. (1930). Theoretical genetics of autopolyploids. *Journal of Genetics*, *22*(3), 359–372.

Halverson, K., Heard, S. B., Nason, J. D., & Stireman III, J. O. (2008). Origins, distribution, and local co-occurrence of polyploid cytotypes in Solidago altissima (Asteraceae). *American Journal of Botany*, *95*(1), 50–58.

Hamrick, J. L., Mitton, J. B., & Linhart, Y. B. (1981). *Levels of genetic variation in trees: influence of life history characteristics*.

Heusser, C. J., Heusser, L. E., & Peteet, D. M. (1985). Late-Quaternary climatic change on the American North Pacific Coast. *Nature*, *315*(6019), 485–487. https://doi.org/10.1038/315485a0

Heuvel. (n.d.). Genomic and Breeding Resources. *Springer*. Retrieved from http://link.springer.com/content/pdf/10.1007/978-3-642-14387-8.pdf

Hobbs, R. J., & Huenneke, L. F. (1992). Disturbance, diversity, and invasion: implications for conservation. *Conservation Biology*, *6*(3), 324–337.

Huntley, B., & Birks, H. J. B. (1983). *An Atlas of Past and Present Pollen Maps for Europe: 0-13000 Years Ago*. Cambridge University Press.

Husband, B. C. (2000). Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1440), 217–223.

Iverson, L. R., Prasad, A. M., Matthews, S. N., & Peters, M. (2008). Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management*, *254*(3), 390–406. https://doi.org/10.1016/j.foreco.2007.07.023

Harrington, C. A., Zasada, J. C., & Allen, E. A. (1994). Biology of red alder (Alnus rubra Bong.). *Biology and Management of Red Alder*, 3–22.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., … Depamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–100. https://doi.org/10.1038/nature09916

Johnson, P. J. (1987). *Larval taxonomy, biology, and biogeography of the genera of North American Byrrhidae*. MS thesis, University of Idaho, Moscow, ID.

Johnson, A. W., & Packer, J. G. (1965). Polyploidy and environment in arctic Alaska. *Science*, *148*(3667), 237–239.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1–15.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191.

Kovanda, M. (1984). Chromosome numbers in selected Angiosperms. *Preslia*, *56*(4), 289–301.

Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 1–11.

Levin, D. A. (1975). Minority cytotype exclusion in local plant populations. *Taxon*, *24*(1), 35–43.

Levin, D. A. (2002). *The role of chromosomal change in plant evolution*. Oxford University Press.

Leitch, A. R., & Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid plants. *Science*, *320*(5875), 481–483. https://doi.org/10.1126/science.1153585

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows--Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Madole, R. F. (1995). Spatial and temporal patterns of late Quaternary eolian deposition, eastern Colorado, USA. *Quaternary Science Reviews*, *14*(2), 155–177.

Mandák, B., V\'\it, P., Krak, K., Trávniček, P., Havrdová, A., Hadincová, V., … Douda, J. (2016). Flow cytometry, microsatellites and niche models reveal the origins and geographical structure of Alnus glutinosa populations in Europe. *Annals of Botany*, *117*(1), 107–120.

Martinková, J., Klimešová, J., Doležal, J., & Kolá\vr, F. (2015). Root sprouting in Knautia arvensis (Dipsacaceae): effects of polyploidy, soil origin and nutrient availability. *Plant Ecology*, *216*(6), 901–911.

Martinson, D. G., Pisias, N. G., Hays, J. D., Imbrie, J., Moore, T. C., & Shackleton, N. J. (1987). Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quaternary Research*, *27*(1), 1–29. https://doi.org/10.1016/0033-5894(87)90046-9

Mavrodiev, M. M., Chester, V. N., Suárez-Santiago, C., Visger, A., Susanna, R., Rodriguez, R., … Soltis, D. E. (2015). Parentage, multiple origins, and distribution of the allotetraploid Tragopogon castellanus Levier sl (Asteraceae). *New Phytol*, *206*, 1172–1183.

Mayr, E. (2013). *Animal species and evolution*. Harvard University Press.

McDonald, E. V, & Busacca, A. J. (1992). Late Quaternary stratigraphy of loess in the Channeled Scabland and Palouse regions of Washington State. *Quaternary Research*, *38*(2), 141–156.

Medley, E. (2012). *Ancient Cataclysmic Floods in the Pacific Northwest: Ancestors to the Missoula Floods*.

Melo, A. T. O., Bartaula, R., & Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics*, *17*(1), 1–15.

Meirmans, P. G., Liu, S., & van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *Journal of Heredity*, *109*(3), 283–296.

Moody, M. E., Mueller, L. D., & Soltis, D. E. (1993). Genetic variation and random drift in autotetraploid populations. *Genetics*, *134*(2), 649–657.

Oginuma, K., Futagawa, K., Hizume, M., SHIBATA, F., & KONDO, K. (2000). Chromosomal evolution of three genera in the Betulaceae concerning polyploidy revealed by in situ hybridization using 5S rDNA. *Chromosome Science*, *4*(2), 47–55.

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics 34, 401-437*

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pielou, E.C., 2008. After the ice age: the return of life to glaciated North America. University of Chicago Press.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Rambaut, A. (2009). FigTree v1. 3.1. *Http://Tree. Bio. Ed. Ac. Uk/Software/Figtree/*

Renny-Byfield, S., & Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany*, *101*(10), 1711–1725.

Roberts, D. R., & Hamann, A. (2016). Climate refugia and migration requirements in complex landscapes. *Ecography*, *39*(12), 1238–1246. https://doi.org/10.1111/ecog.01998

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

Ruffley, M., Smith, M. L., Espíndola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, *27*(4), 1012–1024. https://doi.org/10.1111/mec.14491

Schopmeyer, C. S. (1974). *Seeds of woody plants in the United States*. Forest Service, US Department of Agriculture.

Shaw, J., Munro-Stasiuk, M., Sawyer, B., Beaney, C., Lesemann, J.-E., Musacchio, A., … Young, R. R. (1999). The channeled scabland: back to Bretz? *Geology*, *27*(7), 605–608.

Smith, G. A. (1993). Missoula flood dynamics and magnitudes inferred from sedimentology of slack-water deposits on the Columbia Plateau, Washington. *Geological Society of America Bulletin*, *105*(1), 77–100.

Soltis, P. S., & Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences*, *97*(13), 7051–7057.

Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, *60*, 561–588.

Stebbins Jr., C. L. (1950). *Variation and evolution in plants.* Retrieved from https://www.cabdirect.org/cabdirect/abstract/19510302133

Stebbins, G.L., 1984. Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. Botanica helvetica, 94(1), P. .-13. (n.d.). *No Title*.

Stebbins G.L. (1985). Polyploidy , Hybridization , and the Invasion of New Habitats Author (s): G . Ledyard Stebbins Source : Annals of the Missouri Botanical Garden , Vol . 72 , No . 4 , (1985), pp . 824-832 Published by : Missouri Botanical Garden Press Stable URL : http. *Garden*, *72*(4), 824–832.

Steele, C. A., Carstens, B. C., Storfer, A., & Sullivan, J. (2005). Testing hypotheses of speciation timing in Dicamptodon copei and Dicamptodon aterrimus (Caudata: Dicamptodontidae). *Molecular Phylogenetics and Evolution*, *36*(1), 90–100. https://doi.org/10.1016/j.ympev.2004.12.001

Strenge, D. D. (1994). *The Intraspecific Phylogeography of Polystichum Munitum and Alnus Rubra*. Retrieved from https://books.google.com/books?id=pWhlOAAACAAJ

Te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubešová, M., & Pyšek, P. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany*, *109*(1), 19–45.

Tel-Zur, N., Abbo, S., Myslabodski, D., & Mizrahi, Y. (1999). Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae). *Plant Molecular Biology Reporter*, *17*(3), 249–254.

Wetzel, G. (1929). *Chromosomenstudien bei den Fagales*. Akademische Verlagsgesellschaft mbh.

White, C. M., & West, G. C. (1977). The annual lipid cycle and feeding behavior of Alaskan redpolls. *Oecologia*, *27*(3), 227–238.

Xie, C.-Y. (2008). Ten-year results from red alder (Alnus rubra Bong.) provenance-progeny testing and their implications for genetic improvement. *New Forests*, *36*(3), 273–284. https://doi.org/10.1007/s11056-008-9098-3

Xie, C.-Y., El-Kassaby, Y. A., & Ying, C. C. (2002). Genetics of red alder (Alnus rubra Bong.) populations in British Columbia and its implications for gene resources management. *New Forests*, *24*(2), 97–112.

Yamauchi, A., Hosokawa, A., Nagata, H., & Shimoda, M. (2004). Triploid bridge and role of parthenogenesis in the evolution of autopolyploidy. *The American Naturalist*, *164*(1), 101–112.

Yoon, J. B., Kwon, S.-W., Ham, T.-H., Kim, S., Thomson, M., Hechanova, S. Lou, … Park, Y. (2015). Marker-assisted breeding. In *Current Technologies in Plant Molecular Breeding* (pp. 95–144). Springer.

Zasada, J. C. (1992). The reproductive process in boreal forest trees. *A Systems Analysis of the Global Boreal Forest*, 211–233.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620.

Chapter 3

Association analysis

Abstract: Commonly a riparian species, Red alder (*Alnus rubra* Bong.) is the dominant hardwood tree species in the Pacific Northwest. Red alder is a pioneer species that is well suited to occupy recently disturbed areas (e.g. timber harvesting, fire damage), or areas with poor soil conditions where essential nutrients may be limited. Red alder is primarily a coastal species, however there are disjunct populations in the Northern Idaho Rocky Mountains. The disjunct populations encounter different abiotic conditions than those on the coast. This, as well as the lack of regular gene flow into the disjunct inland populations provide a good system for studying local adaptation via association studies in red alder. Red alder also forms a symbiotic relationship with the N_2-fixing Actinorhizal bacteria *Frankia alni*. *Frankia* fix nitrogen for red alder and in return they receive photosynthetic products produced by their symbiotic host. Red alders usually associate with specific strains of *Frankia*–which exhibit low diversity over large special scales–with the choice being controlled primarily by the host. Utilizing a high-density SNP profile produced by genotype by sequencing we looked for signatures of local adaptation related to environmental variables. We then used the same SNP profile to run a pair of genome-wide association studies where we used the genotype of the *Frankia* strain associated with specific red alders as a red alder phenotype. we found 5 candidate loci for local adaptation that associated with temperature related environmental variables. We were also able to identify 2 candidate loci for selection of specific *Frankia* strains.

Introduction

Local adaptation is common in nature but the genetic basic for these adaptations are poorly understood (Leimu & Fischer 2008; Fraser *et al*. 2011). Understanding how a species responds to environmental change is becoming increasingly important considering the current rate of global climatic change (Kremer *et al*. 2012). Abiotic stresses can drive local adaptation and lead to evolutionary divergence of populations (Nosil *et al*. 2009). The ability to rapidly adapt to local climate may be particularly important in plants due to their sedentary nature (Alberto *et al*. 2013). To avoid extirpation or extinction in changing conditions, a species must either alter their historical distributions (e.g. move north, migrate to a higher elevation) or adapt to their new environmental conditions. Some adaptations can be induced due to phenotypic plasticity, others occur across several generations by local adaptation (Aitken *et al*. 2008). Natural selection is key in assuring the best fitted alleles to a specific environment persist and become prevalent through positive selection, the primary driving force in adaptive evolution in plants (Bose & Bartholomew 2013). Identification of beneficial alleles under selection is key for understanding plant adaptive evolution. Information on local adaptation and the identification of specific beneficial phenotypes provides insight into how a species may adapt to changing conditions, while also providing useful information on conservation, restoration, and crop improvements. Understanding the genetic basis for local adaptation in natural populations, as well as the ecological conditions driving natural selection and genetic differentiation may also provide insight into how that species may respond to future environmental variation due to deforestation and climate change (Fisichelli *et al*. 2014).

We explore how different environmental conditions have driven the evolution of natural Red alder (*Alnus rubra* Bong.) utilizing Genotype by sequencing (GBS). Red alder is fast-growing shade-intolerant pioneer species that, thrives in recently disturbed areas, and riparian systems. Due to its ability to form a symbiotic relationship with the N_2-fixing bacteria *Frankia alni,* red alder is able to grow in poor soil conditions which, in turn, increases usable nitrogen in the soil for subsequent plant species to utilize. Because of this, red alders are important in stream bank stabilization and have a disproportional impact on soil chemistry and plant community composition (Mejnartowicz 2008). In the Pacific Northwest (PNW) red alder is the dominant *Alnus* species and is generally found at elevations below 750 m, and within 200 km of the Pacific coast, however a few disjunct populations exist in the northern rocky mountains (NRM) in Idaho. Initially dismissed as a nuisance species, the beautiful wood and quick growth rate of red alders has made them one of the most commercially valuable trees in the PNW (Thomas & Turner 2006). Recent evidence has emerged indicating that some red alders have been isolated in the NRM since the ice sheets advanced southward at the onset of the Quaternary ~2.5 mya – subsequent vacillations in the ice sheets allowed for periodic dispersal events into the NRM (Chapter 2). Red alder is a coastal species and individuals in the NRM are subject to different environmental conditions than that of their coastal relatives. Specifically, the NRM range where the alders were sampled received less precipitation, had a shorter growing season, and were subject to both colder and warmer temperature extremes than their inland counterparts (Wang *et al*. 2012; appendix B). The existence of alders and the variation in environment experienced in the NRM provides an ideal system to study local adaptation. How a species responds adaptively to variation in

local environments is relevant for conservation and management of this important forest tree species in natural settings, as well as providing valuable insight for resource exploitation as a crop.

Frankia is a genus of Gram-positive aerobic filamentous soil bacteria with a high gc composition that grow branching and tip extensions similar to *Streptomyces sp*. *Frankia sp.* are found in extremely variable soil types on all continents except Antarctica and are saprophytic as well as symbiotic which explains how *Frankia* can survive and exist in abundance in soils where no host plants exist (Gauthier *et al*. 2000). *Frankia sp.* fix N for their host, a common limiting agent for terrestrial plants. In return the host provides *Frankia* sp. with sugars and other photoassimilates. Mutualisms like this can be costly to the host. For example, *rhizobia* have been shown to utilize up to 40% of the host's net photosynthetic C production (Chapin *et al.* 1987). Alders are generally only infected by specific *Frankia* strains (Lipus & Kennedy 2011) with the specificity being determined by the host (Lipus & Kennedy 2011; Polme *et al*. 2014). Alder-associated *Frankia* have been shown to exhibit low diversity over large spatial scales with as few as two genotypes constituting 96% of all identified samples (Kennedy *et al*. 2010a; Kennedy *et al*. 2010b). Currently there is a limited understanding of what factors are involved in *Frankia*-alder specificity in natural systems, however as more research elucidates finer details of rhizobia-legume and actinorhizal symbioses, there is an increased interest in the potential to engineer N-fixing ability to non-legume crops (Beatty & Good 2011; Geurts *et al*. 2012). While the legume-*Rhizobia* symbiosis is understood in much more detail, certain factors involved in actinorhizal symbioses

suggest that they might be more suited to be adapted to infect crop plants. For example, *Frankia* can nodulate with symbiotic hosts from eight angiosperm families whereas all rhizobia nodulation takes place in legumes with *Parasponia sp.* Being the only exception. Also, in both *Parasponia sp.* and 75% of actinorhizal host genera, the infection thread enters the root hairs intercellularly (Baker 1992; Wall 2000; Pawlowski & Sprent 2007). In contrast, in 75% of legumes the rhizobia infection thread enters the root hairs intracellularly. This is important because the intercellular infection utilized by actinorhizal symbioses evolved earlier and is less sophisticated than the intracellular method utilized in rhizobia (Madsen *et al*. 2010).

Next generation sequencing

The use of molecular markers for whole genome genetics studies is becoming more common (Ruffley *et al*. 2018; Chapel *et al* 2019 Bagley *et al*. 2020). Single nucleotide polymorphisms (SNPs) are the most frequent type of sequence differentiation in the genomes of plants (Edwards & Batley 2010). SNPs are suitable local adaptation and population structure studies, and essentially any plant genomic or breeding applications that require large numbers of molecular markers covering the entire genome (Kumar *et al*. 2012). The advent of next generation sequencing has provided a fast, cost-effective molecular tool that is already being used in plant genetics. Specifically, genotype by sequencing (GBS; Elshire *et al*. 2011) can be used for synchronous marker discovery and genotyping–and because many samples can be multiplexed, this further reduces costs (He *et al*. 2014). Association studies generally assume that loci involved in adaptation show stronger differentiation among populations coupled with lower diversity

within a population compared to neutral genomic regions (Storz 2005). These loci are considered outliers and can be identified among populations by calculating F_{ST} (Excoffier *et al*. 2009). Because of this, F_{ST} can determine signatures of selection evolving shifting abiotic factors caused by adaptive differentiation and speciation in plants (Schluter 2001; Funk *et al*. 2006). Utilizing GBS to detect signatures of local adaptation due to climate have gained popularity (Goicoechea *et al*. 2019; Cayuela *et al*. 2020; Stuart *et al*. 2021), however GBS in prone to missing data and the process frequently identifies false positives suggesting identification of an outlier may not be truly represent selection (Narum *et al*. 2013; Savolainen *et al*. 2013). Several ways have been introduced to help account for false positives (e.g. p value correction and calculation of the false discovery rate; FDR). In this study we addressed the issue of missing data and FDR by balancing the incorporation of replication into our sampling strategy. First by collecting multiple samples per site (Bagley *et al*. 2020). Next we utilized two different programs (BAYESCENV, Villemereuil & Gaggiotti, 2015; LFMM Frichot *et al*. 2013) for environmental associations (Crossley *et al*. 2017; Andrew *et al*. 2018; Jeffery *et al*. 2018; Chavarria-Pizarro *et al*. 2019; Goicoechea *et al*. 2019; Shu 2020), and two models (GLM, MLM) for our genome genome-wide association studies (GWAS). BAYESCENV and LFMM both utilize a Bayesian models with a Markov chain Monte Carlo (MCMC) sampling strategy with various means to deal with false positives built into the algorithm (Villemereuil & Gaggiotti, 2015, Frichot *et al*. 2013). FDR and p value corrections were necessary for GLM and MLM analyses. GWAS studies scan the SNP profile and test their effect on the phenotype of interest. In our study we utilized the genotype of the *Frankia* strain infecting individual red alders as our phenotype of interest. To the best of

our knowledge this is the first time this approach has been attempted. Our study has three main objectives: 1) Identify candidate loci involved in local adaptations to environmental conditions in red alder. 2) Identify *Frankia* genotypes to be used and phenotypes in our GWAS analysis. 3) Identify candidate loci involved in the selection of specific *Frankia* genotypes.

Materials and methods

Study area and sampling design

A total of 96 red alders in 10 sampling sites were selected throughout Oregon, Washington and Idaho encompassing much of the range of red alder in the PNW (Figure 1.2). In Idaho, the sampling was conducted in the Clearwater Natural Research Area with special permission from the Idaho Forest Service. To compensate for potential rare genotypes or missing data and to maximize SNP coverage we harvested leaves from ~10 trees per collection site (Bagley *et al*. 2020). In three of the Oregon sites (Po, Ho, De) less than ten individuals were collected but overall the average number of alders per site was 9.6. In order to decrease the chance of sampling closely related trees, alders sampled were at least 100m apart. Fresh leaves were collected and immediately put on ice to be returned to the lab at Portland State University within 24 hrs. Leaves were then lyophilized to be used for GBS. For each tree sampled at least 10 *Frankia* nodules were collected. Nodules were collected within top 15 cm of soil within 50 cm of the base of the tree sampled. Samples were immediately put on ice and returned to the lab within 24 hrs.

GBS

DNA extraction and GBS library preparation

DNA was extracted from lyophilized leaf material using a modified CTAB procedure **(**Tel-Zur *et al*. 1999). Briefly, ~20mg of dried leaf tissue was added to sterile Eppendorf tubes. A tungsten bead was added to each tube and tissue was broken down in two 1-minute cycles at 30 Hz on a Qiagen TissueLyser. 400µl of CTAB buffer was added and the cycles were repeated at 20 Hz and then incubated at 65 °C for 1 hour. Samples were centrifuged for 10 minutes at 4000 rpm and 140 µl of the CTAB buffer was transferred to a new tube where it was mixed 1:1 with chloroform isoamyl alcohol. Samples were centrifuged for 10 minutes at 4000 rpm and the aqueous supernatant was removed and transferred to a new tube. Samples were then washed and resuspended overnight twice with ethanol. Lastly the samples were centrifuged for 10 minutes at 4000 rpm, the ethanol was removed and samples were dried for 10 minutes in a Thermo Savant Speedvac SC210A and resuspended in TE buffer. Electrophoresis was performed on a 1% agarose gel to assess DNA quality, and DNA concentrations were determined using an Invitrogen Qubit 3.0 Fluorometer. DNA concentrations were normalized before GBS library preparation of the 96 multiplexed samples according to the GBS protocol developed by Elshire *et al*. (2011). The 5 base pair single cutting restriction enzyme ApeKI was chosen for paired-end sequencing to be conducted at the University of Wisconsin-Madison Biotechnology facility using an Illumina HiSeq 3000 flowcell.

Sequence analysis and SNP filtering

The bioinformatics pipeline GBS-SNP-CROP (Melo *et al*. 2016) was used for SNP processing. Briefly, paired end read files were matched according to their barcode and maintained for further processing. Reads that were lacking an identified barcode were discarded. Raw reads were trimmed, demultiplexed, and paired-end reads were merged into a single read. The centroids clustering algorithm was utilized to stitch together the merged reads into a mock genome for reference and SNP alignment. Reads were then aligned to the mock genome, then filtered and indexed genotype by genotype. Files were then parsed so only those data containing reads polymorphic to the reference sequence were retained with alignment information for potential variants extracted and organized for downstream processing. Initial filtering was conducted using the criteria suggested by Melo *et al*. (2016). Initial paired-end sequencing yielded 357,322,720 raw reads (178,661,360 for each paired-end) with an average GC content of 47.5%. After all alignment and filtering was concluded, 2631 SNPs were retained for further analysis.

Climate variables

Long-term weather observations covering a 30-year span were retrieved from CLIMATE-WNA (Wang *et al*. 2012). Gps coordinates collected from each sampling location were used to download data from each of 20 environmental variables (Table 3.1). Red alders grow in wet humid environments with high annual precipitation with soil moisture during the growing season appearing to be the main condition influencing where the red alder grows (Harrington 1994). Because of this, we are particularly interested in

environmental variables MSP (May to September Precipitation), AHM (annual heat-moisture index), SHM (summer heat-moisture index), Eref (Hargreaves reference evaporation), and CMD (Hargreaves climatic moisture deficit). They can also tolerate a large range in temperatures ranging from -30^0C in Alaska and Idaho, to 46^0C in parts of southern California (Harrington 1994). None of the sites we sampled were subject to these types of extreme temperatures but most of the temperature variation found our study was from the Idaho samples.

Environmental association

Association analyses tend to have a high false positive rate. To help compensate for this we used two different models to test environmental associations. First, BAYESCENV (Villemereuil & Gaggiotti, 2015) was used to test for the sensitivity of each locus to environmental differentiation. BAYESCENV detects signatures of selection by identifying loci that show large positive FST (outside of the neutral model FST distribution) that are significantly correlated with environmental variables. Environmental variables were standardized and BAYESCENV was run individually for each variable according to manual specifications. After 20 pilot runs of 2,000 iterations and a burn□in of 50,000 iterations, 5,000 MCMC samples were taken with 10 steps between each sample. We retained all significant SNP–climate associations and used an FDR of $q \leq 0.05$ for significance. Next, LFMM (Frichot *et al*. 2013) uses a hierarchical Bayesian mixed model based on the residuals of PCA to take into account population genetic structure. LFMM requires priori knowledge of genetic structure. Whenever possible it is suggested to use pre-existing estimates from independent dataset (Frichot *et*

al. 2013). A less conservative SNP matrix was used to test for population structure in chapter 2 which identified K=4 as the proper number of sub-populations. To confirm the K for this dataset, the PEGAS package (Paradis 2010) in Rstudio v3.5.0 was used to asses potential population clusters using Principal Components Analysis (PCA). PCA analysis determined the optimal number of cluster (k=4; supplemental) needed for running the LFMM analyses. The Gibbs sampler was run for 10,000 cycles after a burn-in of 5,000 for each environmental variable. After all runs were complete, we computed the genomic inflation factor (λ) according to the manual by taking the median zscores across the samples and using the following equation λ = median(z 2)/0.456 which identified λ=0.991 as the inflation factor. Choosing values of K for which the estimate of λ is close to or slightly below 1.0 warrants that the FDR can be controlled efficiently (Frichot *et al*. 2013).

Annotation and protein identification

The sequences surrounding the SNPs found to be significant in both models were extracted from the cluster's output produced by GBS-SNP-CROP for downstream processing. These sequences were aligned and compared against the Nation Center for Biotechnology Information (NCBI) database first using BLASTN against the entire NCBI database, and against the genome of *Alnus glutinosa*, the only member of the genus with a sequenced genome.. To assign function to any sequence identified, we compared the aligned sequences to the NCBI protein database by means of BLASTX (Altschul *et al*. 1990). When more than one potential

Table 3.1. Climate Variables and Their Abbreviations Downloaded From CLIMATE-WNA. These 20 variables were used to run climate association analysis utilizing a high-density SNP profile generated by genotype by sequencing.

Abbreviation	Variable
MWMT	Mean warmest month temp
MCMT	Mean coldest month temp
TD	Temperature differential (MWMT/MCMT)
MAP	Mean annual precipitation
MSP	May-Sept precipitation
AHM	Annual heat-moisture index
SHM	Summer heat-moisture index
DD<0	Degree days<0
DD>5	Degree days>5
DD<18	Degree-days<18
DD>18	Degree-days>18
NFFD	Frost-free days
FFP	Frost-free period
BFFP	Day FFP begins
EFFP	Day FFP ends
PAS	Precipitation as snow
EMT	Extreme minimum temperature over 30 yrs
EXT	Extreme maximum temperature over 30 yrs
EREF	Hargreaves reference evaporation
CMD	Hargreaves moisture deficit

protein function was identified, identified proteins were given chosen over hypothetical protein sequences.

Frankia

Surface Sterilization

Collected nodules were washed with deionized water to remove soil and organic matter. Nodules were excised from roots with a sterile scalpel and placed into individual sterile scintillation vials. Each vial was then filled to making sure the nodules were

completely covered 10% bleach solution. Vials were placed in a shaker at room temperature and agitated for 2 minutes. The bleach was then decanted, washed with deionized water, and the bleaching procedure was repeated. After the second bleach wash, the process was repeated three more times with deionized water to wash the nodules of bleach.

DNA Extraction

DNA was extracted from single nodule lobes using the Sigma Tissue Extract-N-Amp Kit (Sigma-Aldrich, St. Louis, MO). Forty µL of extraction buffer was added to a 1.5 mL microcentrifuge tube containing a single nodule which was then crushed using a sterile micropestle. The resulting homogenate was centrifuged for 1 minute at 4,000 rpm, and 20 µL of supernatant from each tube transferred into an individual 0.2 mL PCR strip-tube. Tubes containing extractions were moved to a thermal cycler and incubated at 65 °C for 10 minutes, followed by 95 °C for 10 minutes. 25 µL of neutralization buffer was pipetted into each sample, briefly vortexed and stored at 4 °C.

PCR and Sequencing

A 606 bp (base pair) portion of the nifH gene was amplified by PCR using *Frankia*-specific primers nifHf1 (5′-GGC AAG TCC ACC ACC CAG C-3′) and nifHr (5′-CTC GAT GAC CGT CAT CCG GC3′). PCR reactions were set up in 24 µL volumes containing 8.45 µL PCR water, 12.5 µL GoTaq Master Mix, 1.25 µL BSA (1 mg/1 mL), 0.4 µL of each primer (10 µM), and 1 µL of 1:10 diluted template. Reaction mixtures were then subjected to the following thermal cycling conditions: 96 °C for 5 minutes, 35

cycles at 95 °C for 30 seconds, 60 °C for 30 seconds, 72 °C for 45 seconds, and a final extension at 72 °C for 7 minutes. Following PCR, all reaction products were visualized by gel electrophoresis. Sanger Sequencing was performed in both directions using the same primers used in PCR on an ABI 3730xl (Applied Biosystems, Foster City, CA) at Functional Biosystems, Inc. Forward and backward reads were visually inspected and the read with the best quality was reversed if necessary and trimmed to 580 bp (Kennedy *et al*. 2010). The Neighbor-Joining method was used in MEGA v10.1.8 to create a phylogeny of the *Frankia* sampled. Operational taxonomic units (OTUs) were generated using a 97% sequence similarity threshold. This similarity cutoff was chosen because previous work has shown that *Frankia* genotypes are accurately placed in appropriate genomic groups at this cutoff (Lipus & Kennedy 2011, Higgins & Kennedy 2012, Rodriguez *et al*. 2016).

GWAS

The OTUs identified in MEGA were considered as a phenotype and used to test the association of *Frankia* OTUs in TASSEL v5.0 (Bradbury *et al*. 2007). The kinship matrix (K) and Principal Component Analysis (PCA) were estimated and both general linear models (GLM) including a Q-matrix, and mixed linear models (MLM) with Q-matrix and a K-matrix. The two matrices were used to account for corrections in population structure (Q), and genetic relatedness (K). The uniform Bonferroni threshold for significance was calculated using $P \leq 1/N$ (a=0.05) which gave a $-\log_{10} = 4.7$. We then set a FDR at 0.05 using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Any sequences identified as significant by both models were aligned

and annotated in BLAST in the same manner as the significant SNPs identified in the environmental analysis.

Results

Environmental associations

To identify molecular signals of environmental related selection we used two genome environmental association models (BAYESCENV and LFMM) to find candidate SNPs for local adaptation. LFMM identified 29 SNPs that associated to one or more environmental variables while BAYESCENV identified 96. A total of five SNPs were shown to have significant association with environmental variable in both models (table 3.2). Three of the five SNPs were associated with more than one environmental variable with SNP 382 associating with five environmental variables (SNP 382=<18, mat, td, ffp, nffd) and SNP 2012 associating four (SNP 2012= dd5, <18, mat, dd0). SNP 775 associated with three (dd0, emt, mcmt), and Two SNPs (30, 1017) were both associated with a single environmental variable (dd5). Interestingly all the environmental variables that showed significant associations in both models were related to temperature and frost conditions (table 3.2).

BLAST

Of the five SNPs that were found to have significant associations in both models, no protein hits were identified aligning against the *Alnus glutinosa* reference genome. BLASTN results identified other members of Betulaceae as the closest hit except for SNP 1017 which the closest hit was for *Ligustrum quihoui* (Table 3.2). Interestingly, BLASTN

results indicated that SNPs 30 and 1017 were both mitochondria sequences with SNP 30 identifying mitochondria from the closely related *Betula pendula* and 1017 identifying mitochondria from *Ligustrum quihoui,* which is not closely related or even in the same order as *Alnus rubra*. BLASTX produced no viable protein hits when aligned with the annotated genome of *Alnus glutinosa.* When aligned against the full database, no proteins were identified for SNPs 30, 1017, and 2012. SNP 775 was identified as a chloroplast F-box protein and SNP 382 yielded protein sequences from the yeasts Zygosaccharomyces, Cryptococcus, Kluyveromyces, all of which were yeasts with identical BLAST scores and are represented by cryptococcus in our table (table 3.2).

Table 3.2. BLAST Results of the SNP Sequences that Associated With Environmental Variables. The SNP ID, significant variable, and organelle or protein sequence identified by BLAST.

SNP ID	Environmental variable	BLAST
30, 1017	dd5	Mitochondria
382	<18, mat, td, ffp, nffd	Cryptococcus 40s r-protein S17-3
775	dd0, emt, mcmt	Chloroplast F-box protein
2012	dd0, dd5, <18, mat	No protein hits

Frankia

We sequenced a portion of the nifH gene from *Frankia* nodules associated with the roots of 95 *Alnus rubra* individuals to assess the diversity and identify specific genotypes to be utilized as phenotype in GWAS analysis. Of the 95 nodules processed we successfully generated nifH sequences for 85, which were comprised of nine OTUs

(herein referred to as genotypes; figure 3.1). One genotype (DJB-F1) was dominant consisting of 78.8% (67) of the total and was found in all ten sampling locations (table 3.3.). Of the 18 remaining *Frankia* with genotypes, ~78% (14 of the 18) were in just three sites (Id, St, Wa). Genotypes Kl-2, Kl-2.6 F-016 consisted of a single individual each and were only found in the Wa and St sites. No location had more than two *Frankia* with nifH sequences that did not amplify.

GWAS

Two SNPs (2230, 2297) were found to have significant associations in both the GLM and MLM. BLASTN identified *Corylus avellana* as the top hit with BLASTX identifying a glycerol kinase from *Pistacia vera* for SNP 2230 and a cytochrome P450 CYP 736 family from *Prunus yedoensis var. nudiflora* for SNP 2297 (table 3.4).

Table 3.3. *Frankia* OUT's Identified in MEGA. *Frankia* strain, number (N) percent total and accession number of *Frankia* OUT's identified from the amplified nifH region of *Frankia* extracted from root nodules from red alder. One strain (DJB-F1) was the dominant strain found in 78.8% of all samples. A second OUT (BK-3) was the second most common strain with all other OUT's being rare.

Frankia strain	N	% of total	Accession	
DJB-F1	67	78.8	MK529665.1	All
BK-3	5	5.9	MK105603.1	Ti,Ho,Id,Wa
BK-1	2	2.4	MK105601.1	Wa only
DJB-F7	3	3.5	MK529667.1	St only
BK-5	2	2.4	MK105605.1	De, Rt
F074	3	3.5	JN088561.1	St,Id
kl-2	1	1.2	GU810474.1	Wa
kl-2.6	1	1.2	HM031957.1	Wa
F-016	1	1.2	JN088552.1	St

Discussion

We used multiple complementary approaches for detecting candidate loci involved in selection associated with environmental variables in 10 *A. rubra* stands in the PNW. The analysis identified significant signatures of selection in five SNPs. Interestingly, All five significant SNPs were associated with variables related to temperature. It has been reported that moisture during the growing season appears to be one of the main conditions that influence where the alder grows and generally, it does not grow in areas where drought is common and in Idaho and southern California it seldom grows on south facing slopes. Red alder compensates for water deficiencies and higher temperatures by only growing on the border of lakes and streams in California and the NRM (Harrington 1994) which explains why SNPs associated with variables related to temperature rather than moisture. Red alders experience more extreme temperature variations in the NRM compared to the coast. For example, temperature variation in coastal climates is generally between 10-15 ^{o}C where inland climates can vary as much as 40 ^{o}C. Coastal climates also tend to have wet winters and dry summers while inland climates tend to have humid summers and dryer winters. When looking at temperature variations among the sites, the Idaho sites had the lowest mean annual temperature and was the only area to have any mean monthly temperature below freezing (appendix b). The mean warmest month temperature in the Idaho sites was closer to average and there was less variation in the warmer temperatures.

The whole genome sequence of *A. glutinosa* provides a useful database for searching for candidate genes associated with environmental variables. Interestingly only one

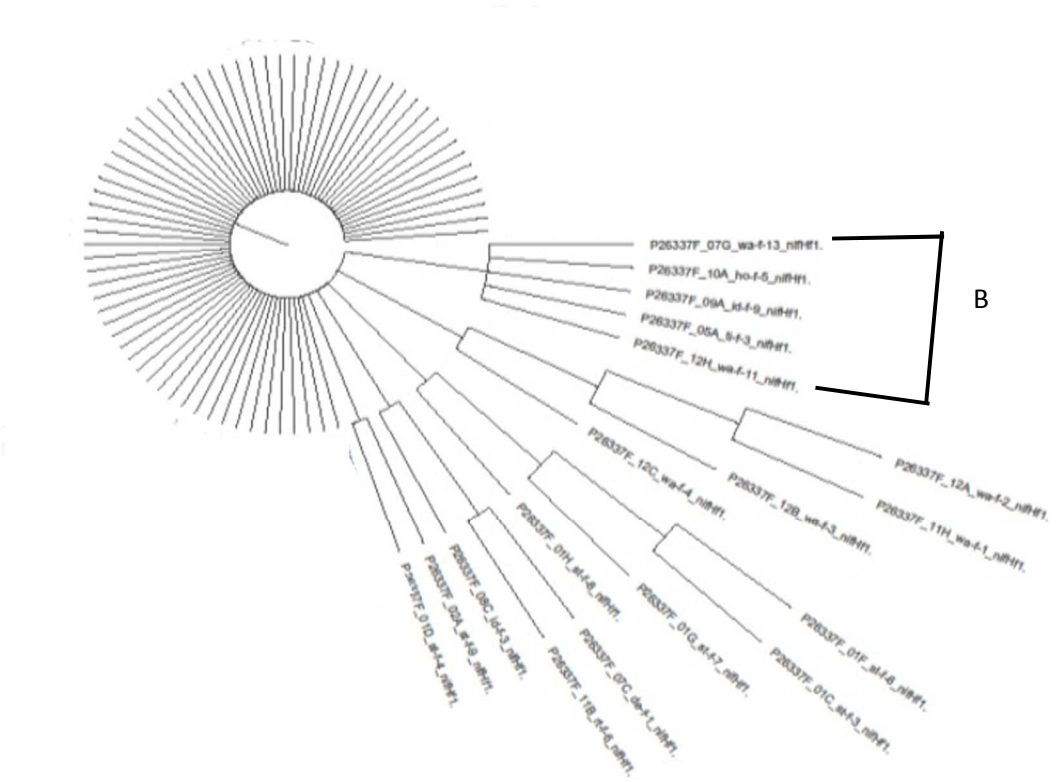


Figure 3.1. Phylogenetic Tree Produced in MEGA Identifying *Frankia* OTUs. Phylogenetic tree identifying nine OTU's of *Frankia* strains that associate with red alder in the Pacific Northwest. Strain DJB F-1 (A) was the dominant OTU making up 78.8% of all amplified sequences. BK-3 (B) consisted of five individuals with the rest of the OTU's being rare.

SNP associated with a sequence from somatic DNA in *A. glutinosa* and BLAST revealed no significant protein hits for the SNP. All of the other SNPs either associated with an organelle (2 mitochondria 1 chloroplast) or an endophyte of red alder. This is likely possible due to the lack of a reference genome for SNP calling. GBS-SNP-CROP utilizes the reads from GBS to make a mock-reference genome to align and call SNPs. This

Table 3.4. BLAST Results of 2 SNPs Associated With *Frankia* selection. Table includes the number of the SNP that associated with *Frankia* selection, the protein identified by BLAST, and the accession number associated with the protein.

SNP ID	Protein	Accession #
2230	Glycerol kinase	XP_031250009.1
2297	Cytochrome P450 CYP7364	PQM39580.1

allows for SNPs to be called from sequences that would normally be excluded if a reference genome was available. For example, Sonah *et al*. (2015) conducted an analysis of unmapped reads in soybean and discovered that chloroplast and mitochondria contribute to 4.6% and 1.8% of the total population reads. Mitochondria in plants are key for cellular ATP production as well as a major source of C-skeletons for biosynthesis making them crucial for plant survival. Respiration rate is one of the first metabolic processes disrupted by temperature stress (Atkin *et al*. 2015). Stress induced by low temperatures initiates rapid changes in mitochondrial respiration rates and respiratory homeostasis (RH) - RH is the tendency of the plant to achieve certain rates of respiration regardless of growth conditions which is associated with cell performance under poor conditions (Yamori *et al*. 2009; Atkin *et al*. 2015)–with the timing and severity of their response is generally plastic depending on length and intensity of low temperature exposure (Heidarvand *et al*. 2017). Variations in the respiration rates can be both positive and negative and can be caused by short-term (shock) exposure to low temperatures, or acclimation after long-term exposure (Heidarvand *et al*. 2017). Generally, initial exposure to low temperatures causes a decrease in the rate of respiration (Bravo *et al*. 2007, Matos *et al*. 2007), however there are exceptions to this (Mizuno *et al*. 2008). Long-term cold exposure can lead to a major reprogramming of metabolism through

structural and biochemical alterations in the cell (Winfield *et al*. 2010; Genga *et al*. 2011). Cold temperatures are likely causal in the association of SNP 775 as well. SNP 775 associated with 3 environmental variables, dd0 (degree days <0˚C), emt (extreme minimum temperature), mcmt (mean coldest month temperature).

BLASTX revealed SNP 775 to be a chloroplast protein related to flowering time. Flowering time is of key importance for plant survival and reproduction for several reasons. First, the timing and duration of flower emergence determine reproductive ability of the plant; second, timing of flowering relative to abiotic constraints is key for seed propagation; and third, synchronization with insects and other pollinators are crucial for fitness and proliferation (Jagadish *et al*. 2016). All of these functions are influenced by temperature conditions therefore flowering time is one of the major factors that determining plant adaptation to changing climatic conditions (Jagadish *et al*. 2016). For example, flowering time for *Syringa vulgaris* has shown a near linear relationship between temperature increase and early onset of flowering time (Cayan *et al*. 2001). Globally, in response to recent climate warming, first flowering dates have advanced several days per decade with the potential to alter whole plant distributions (Crimmins *et al*. 2011; Hulme 2011).

The final SNP associating with a ribosomal protein from a yeast was also unexpected and we suggest that the sequence identified was from a fungal endophyte of the red alder. Endophytic yeasts have been reported to form non-pathogenic associations with a variety of host plants (Vieira *et al*. 2012; Rhoden *et al*. 2013). Microsymbionts are ubiquitous in terrestrial plant species and play a key role plant fitness, development, and

persistence (Rho *et al.* 2018; Dellagi *et al.* 2020). These symbionts can form mutualisms in plant root systems like arbuscular mycorrhizal fungi which aid in nutrient uptake in more than 80% of terrestrial plant families (Brachmann & Parniske 2006), and endophytes which are restricted to growth in plant tissues (Rodriguez *et al*. 2009). Recent studies suggest that endophytes and bacterial symbionts initiate diverse metabolic processes that help plant survival when encountering severe environmental conditions including stress related to cold (Dighton & White 2017; Hill *et al.* 2019).

Endosymbionts have been shown to aid in phenotypic plasticity (Goh *et al.* 2013), nutrient absorption (Hill *et al.* 2019) osmotic regulation (Azad & Kaminskyj 2016), and herbivore defense (Gonzáles Teuber 2016). Previous studies have shown that both bacterial symbionts and fungal endophytes are capable of stabilizing physiological and metabolic processes at low temperatures (Theocharis *et al.* 2012; Thalmann & Santelia 2017). We suggest that red alders in the NRM select for endophytes that are better able to stabilize their metabolic activities in the more extreme temperature conditions experienced in the inland sites.

Frankia

Similar to previous studies (Kennedy *et al*. 2010a; Kennedy *et al*. 2010b), we identified a relatively limited number of *Frankia* genotypes that associate with red alder. The lack of diversity is similar to other alnus-*Frankia* symbioses as well including *A. tenuifolia* and *A. viridis* (Anderson *et al*. 2009). While it is not yet known why *Frankia* assemblages have such low diversity, it has been suggested that it is reflective of selective pressure on the host to maintain beneficial genotypes. This is supported by a

recent study where an evaluation of over 400 *Frankia*-Alnus symbioses covering 22 Alnus species found that host genetics were mainly responsible for selection (Polme *et al*. 2014). Using the different *Frankia* genotyped as phenotype in our GWAS study identified two significant SNPs associated with *Frankia* selection. The protein sequence of first of the two SNPs identified in BLASTX in our *Frankia* GWAS was that of a glycerol kinase. Glycerol kinases catalyzes the transfer of a phosphate from ATP to glycerol forming glycerol-3-phosphate (G3P). G3P is crucial for the induction of systemic acquired resistance (SAR) in plants (Mandal *et al*. 2011; Shine *et al*. 2019). Systemic immunity is a form of immunity that provides resistance against a range of plant pathogens (Durrant *et al*. 2004). Infection with a pathogen often generates a mobile signal that permeates tissues surrounding the infection and activates a broad-spectrum resistance throughout the plant called SAR (Chanda *et al*. 2011). There is a dearth of information on actinorhizal symbioses, however the legume-rhizobia symbiosis is better understood. *Rhizobia* are initially able to suppress their host's defenses long enough to become established. After initial infection, specific plant immunity responses are required for nodule development and defense (Samac & Graham 2007). Recently, Shine *et al*. (2019) looked into the role of G3P in rhizobia-legume symbioses in the soybean (*Glycine max*). They found that when the host interacts with an incompatible rhizobium, a yet unknown signal is produced which induces the accumulation of G3P subsequently activating foliar resistance to pathogens. Accumulated G3P is then transported to the root infection site which enables the exclusion of incompatible or unwanted rhizobia in a pathway that significantly overlaps with that of SAR. More studies are needed to help elucidate the specific intermediates and steps involved in this process, however what little

information is known is intriguing as it relates to our study. The results from our GWAS analysis suggest the possibility that the hosts SAR response is involved in the process of excluding unwanted *Frankia* from becoming established in the host.

The protein sequence of second of the two SNPs identified in BLASTX was a Cytochrome p450 (CYP) 7364 family protein. CYP 7364 proteins have been shown to be crucial in plant communication with pathogenic microorganisms (Guengerich *et al*. 2010), or in the host-symbiont communication in the legume *Glycine max* (Neelakandan *et al*. 2010). Plants produce a diverse group of compounds that are involved in structural support, defense and cell communication termed secondary metabolites because they are not important to basal cell functions such as photosynthesis, respiration, or protein synthesis (Fraser & Chapple 2011). The precursors for a variety of secondary metabolites are provided by the shikimate pathway which is the point for phenylpropanoids in plants (Weaver *et al*. 1997). The ability of secondary metabolism to be involved in the production of hundreds of thousands of natural plants compounds is based on a variety of genes (Mizutani 2012), of which the Cytochrome p450 (CYP) are the most important (Chapple 1998). CYP proteins are widely distributed in archaea, bacteria, eukaryotes, and even viruses (Mizutani & Ohta 2010). In plants they form the largest superfamily of enzymes which comprise roughly 1% of total gene products (Mizutani and Ohta 2010; Kawai *et al*. 2014). Interestingly, despite their obvious importance in plants as indicated above, diversification of p450s is such that it leads to the production of specific metabolites for different species so much so that most p450s studied so far produce a unique plant-specific protein (Zhao 2014). CYPs are heme-containing enzymes that are

integral in biosynthesis of plant secondary metabolites such as alkaloids, fatty acids, hormones, and phenylpropanoids such as flavonoids and isoflavonoids (Chapple 1998). For plants it has been demonstrated that flavonoids are essential signal molecules for establishment of root symbioses (Reddy *et al*. 2007) and we suggest it is possible the CYP protein that showed association in our study was related to flavonoid production.

Legumes release a variety of flavonoids and isoflavonoids which induce symbiotic genes in bacterial symbionts (See Cohn *et al*. 1998; Perret *et al*. 2000 for reviews). Recognition by the bacterial symbiont to a specific flavonoid produced by the host is key to partner choice because only the correct signals induce gene expression in the bacteria (Simms & Taylor 2002). Host symbiont communication in actinorhizal symbiosis is poorly understood but flavonoids are the prime candidates as signaling molecules in their associations as well (Abdel-Lateif *et al*. 2012). For example, Benoît & Berry (1997) showed enhanced nodulation of *Frankia* when introduced to flavonoid containing seed washes from A. *rubra*. Similar results were seen in the enhanced nodulation of *Alnus glutinosa* when introduced to flavonoids (Hughes *et al*. 1999). It is also likely that flavonoids are necessary for early host/symbiont communication arbuscular mycorrhizal associations (Akiyama *et al*. 2002), however the identification of strigolactones as important molecules in in these associations (Akiyama *et al*. 2005) suggests that flavonoids are not exclusive in AM signaling (Steinkellner *et al*. 2007).

References

Abdel-Lateif, K., Bogusz, D., & Hocher, V. (2012). The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and *Frankia* bacteria. *Plant Signaling & Behavior*, *7*(6), 636–641.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95–111.

Akiyama, K., Matsuoka, H., & Hayashi, H. (2002). Isolation and identification of a phosphate deficiency-induced C-glycosylflavonoid that stimulates arbuscular mycorrhiza formation in melon roots. *Molecular Plant-Microbe Interactions*, *15*(4), 334–340.

Alberto, F. J., Aitken, S. N., Alia, R., González-Mart\'\inez, S. C., Hänninen, H., Kremer, A., … others. (2013). Potential for evolutionary responses to climate change--evidence from tree populations. *Global Change Biology*, *19*(6), 1645–1661.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Anderson, M. D., Ruess, R. W., Myrold, D. D., & Taylor, D. L. (2009). Host species and habitat affect nodulation by specific Frankia genotypes in two species of Alnus in interior Alaska. *Oecologia*, *160*(4), 619–630.

Andrew, S. C., Jensen, H., Hagen, I. J., Lundregan, S., & Griffith, S. C. (2018). Signatures of genetic adaptation to extremely varied Australian environments in introduced European house sparrows. *Molecular Ecology*, *27*(22), 4542–4555.

Atkin, O. K., Bloomfield, K. J., Reich, P. B., Tjoelker, M. G., Asner, G. P., Bonal, D., … others. (2015). Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *New Phytologist*, *206*(2), 614–636.

Azad, K., & Kaminskyj, S. (2016). A fungal endophyte strategy for mitigating the effect of salt and drought stress on plant growth. *Symbiosis*, *68*(1–3), 73–78.

Bagley, J. C., Heming, N. M., Gutiérrez, E. E., Devisetty, U. K., Mock, K. E., Eckert, A. J., & Strauss, S. H. (2020). Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (Populus tremuloides). *Ecology and Evolution*, *10*(11), 4609–4629. https://doi.org/10.1002/ece3.6214

Baker, D. D. (1992). Actinorhizal Symbioses Dwight D. Baker and Beth C. Mullin. *Biological Nitrogen Fixation*, 259.

Beatty, P. H., & Good, A. G. (2011). Future prospects for cereals that fix nitrogen. *Science*, *333*(6041), 416–417. https://doi.org/10.1126/science.1209467

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Benoit, L. F., & Berry, A. M. (1997). Flavonoid-like compounds from seeds of red alder (Alnus rubra) influence host nodulation by Frankia (Actinomycetales). *Physiologia Plantarum*, *99*(4), 588–593.

Bose, R., & Bartholomew, A. J. (2013). *Macroevolution in deep time*. Springer Science & Business Media.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brachmann, A., & Parniske, M. (2006). The most important symbiosis on earth. *Soil Biol*, *4*, 239.

Bravo, L. A., Saavedra-Mella, F. A., Vera, F., Guerra, A., Cavieres, L. A., Ivanov, A. G., … Corcuera, L. J. (2007). Effect of cold acclimation on the photosynthetic performance of two ecotypes of Colobanthus quitensis (Kunth) Bartl. *Journal of Experimental Botany*, *58*(13), 3581–3590.

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the western United States. *Bulletin of the American Meteorological Society*, *82*(3), 399–416.

Cayuela, H., Rougemont, Q., Laporte, M., Mérot, C., Normandeau, E., Dorant, Y., … others. (2020). Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Molecular Ecology*, *29*(13), 2379–2398.

Chanda, B., Xia, Y., Mandal, M. K., Yu, K., Sekine, K.-T., Gao, Q., … others. (2011). Glycerol-3-phosphate is a critical mobile inducer of systemic immunity in plants. *Nature Genetics*, *43*(5), 421–427.

Chapel, D. B., Patil, S. A., Plagov, A., Puranik, R., Mendybaeva, A., Steinhardt, G., … others. (2019). Quantitative next-generation sequencing-based analysis indicates progressive accumulation of microsatellite instability between atypical hyperplasia/endometrial intraepithelial neoplasia and paired endometrioid endometrial carcinoma. *Modern Pathology*, *32*(10), 1508–1520.

Chapin, F. S., Bloom, A. J., Field, C. B., & Waring, R. H. (1987). Plant responses to multiple environmental factors. *Bioscience*, *37*(1), 49–57.

Chapple, C. (1998). Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annual Review of Plant Biology*, *49*(1), 311–343.

Chavarria-Pizarro, T., Gomez, J. P., Ungvari-Martin, J., Bay, R., Miyamoto, M. M., & Kimball, R. (2019). Strong phenotypic divergence in spite of low genetic structure in the endemic Mangrove Warbler subspecies (Setophaga petechia xanthotera) of Costa Rica. *Ecology and Evolution*, *9*(24), 13902–13918.

Christian, N., Whitaker, B. K., Clay, K., Dighton, J., & White, J. F. (2017). Chapter—a novel framework for decoding fungal endophyte diversity. *The Fungal Community: Its Organization and Role in the Ecosystem, Fourth Edition. Boca Raton: CRC Press Taylor & Francis Group*, 65–78.

Cohn, J., Day, R. B., & Stacey, G. (1998). Legume nodule organogenesis. *Trends in Plant Science*, *3*(3), 105–110.

Crimmins, T. M., Crimmins, M. A., & Bertelsen, C. D. (2011). Onset of summer flowering in a 'Sky Island'is driven by monsoon moisture. *New Phytologist*, *191*(2), 468–479.

Crossley, M. S., Chen, Y. H., Groves, R. L., & Schoville, S. D. (2017). Landscape genomics of Colorado potato beetle provides evidence of polygenic adaptation to insecticides. *Molecular Ecology*, *26*(22), 6284–6300.

Dellagi, A., Quillere, I., & Hirel, B. (2020). Beneficial soil-borne bacteria and fungi: a promising way to improve plant nitrogen acquisition. *Journal of Experimental Botany*, *71*(15), 4469–4479.

Durrant, W. E., & Dong, X. (2004). Systemic acquired resistance. *Annu. Rev. Phytopathol.*, *42*, 185–209.

Edwards, D., & Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal*, *8*(1), 2–9.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285–298.

Fisichelli, N., Wright, A., Rice, K., Mau, A., Buschena, C., & Reich, P. B. (2014). First-year seedlings and climate change: species-specific responses of 15 North American tree species. *Oikos*, *123*(11), 1331–1340.

Fraser, C. M., & Chapple, C. (2011). The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book/American Society of Plant Biologists*, *9*.

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. (2011). Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, *106*(3), 404–420.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences*, *103*(9), 3209–3213.

Gauthier, D., Jaffré, T., & Prin, Y. (2000). Abundance of Frankia from Gymnostoma spp. in the rhizosphere of Alphitonia neocaledonica, a non-nodulated Rhamnaceae endemicto New Caledonia. *European Journal of Soil Biology*, *36*(3–4), 169–175.

Genga, A., Mattana, M., Coraggio, I., Locatelli, F., Piffanelli, P., & Consonni, R. (2011). Plant Metabolomics: A characterisation of plant responses to abiotic stresses. *Abiotic Stress in Plants-Mechanisms and Adaptations. Rijeka: InTech*, 309–350.

Geurts, R., Lillo, A., & Bisseling, T. (2012). Exploiting an ancient signalling machinery to enjoy a nitrogen fixing symbiosis. *Current Opinion in Plant Biology*, *15*(4), 438–443.

Goicoechea, P. G., Guillardin, L., Fernández-Ibarrodo, L., Valbuena-Carabaña, M., González-Mart\’\inez, S. C., Alia, R., & Kremer, A. (2019). Adaptive introgression promotes fast adaptation in oaks marginal populations. *BioRxiv*, 731919.

González-Teuber, M. (2016). The defensive role of foliar endophytic fungi for a South American tree. *AoB Plants*, *8*.

Guengerich, F. P., Tang, Z., Salamanca-Pinzón, S. G., & Cheng, Q. (2010). Characterizing proteins of unknown function: orphan cytochrome p450 enzymes as a paradigm. *Molecular Interventions*, *10*(3), 153.

Harrington, C. A., Zasada, J. C., & Allen, E. A. (1994). Biology of red alder (Alnus rubra Bong.). *Biology and Management of Red Alder*, 3–22.

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*, 484.

Heidarvand, L., Millar, A. H., & Taylor, N. L. (2017). Responses of the mitochondrial respiratory system to low temperature in plants. *Critical Reviews in Plant Sciences*, *36*(4), 217–240.

Higgins, L. M., & Kennedy, P. G. (2012). Symbiotic Frankia bacteria in Alnus forests in Mexico and the United States of America: is geographic location a good predictor of assemblage structure? *Botany*, *90*(6), 423–431.

Hill, P. W., Broughton, R., Bougoure, J., Havelange, W., Newsham, K. K., Grant, H., … others. (2019). Angiosperm symbioses with non-mycorrhizal fungal partners enhance N acquisition from ancient organic matter in a warming maritime Antarctic. *Ecology Letters*, *22*(12), 2111–2119.

Hughes, M., Donnelly, C., Crozier, A., & Wheeler, C. T. (1999). Effects of the exposure of roots of Alnus glutinosa to light on flavonoids and nodulation. *Canadian Journal of Botany*, *77*(9), 1311–1315.

Hulme, P. E. (2011). Contrasting impacts of climate-driven flowering phenology on changes in alien and native plant species distributions. *New Phytologist*, *189*(1), 272–281.

Jagadish, S. V, Bahuguna, R. N., Djanaguiraman, M., Gamuyao, R., Prasad, P. V, & Craufurd, P. Q. (2016). Implications of high temperature and elevated CO2 on flowering time in plants. *Frontiers in Plant Science*, *7*, 913.

Kawai, Y., Ono, E., & Mizutani, M. (2014). Evolution and diversity of the 2--oxoglutarate-dependent dioxygenase superfamily in plants. *The Plant Journal*, *78*(2), 328–343.

Kennedy, P. G., Schouboe, J. L., Rogers, R. H., Weber, M. G., & Nadkarni, N. M. (2010). Frankia and Alnus rubra canopy roots: an assessment of genetic diversity, propagule availability, and effects on soil nitrogen. *Microbial Ecology*, *59*(2), 214–220.

Kennedy, P. G., Weber, M. G., & Bluhm, A. A. (2010). Frankia bacteria in Alnus rubra forests: genetic diversity and determinants of assemblage structure. *Plant and Soil*, *335*(1), 479–492.

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., … others. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, *15*(4), 378–392.

Kumar, S., Banks, T. W., & Cloutier, S. (2012). SNP discovery through next-generation sequencing and its applications. *International Journal of Plant Genomics*, *2012*.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PloS One*, *3*(12), e4010.

Lipus, A., & Kennedy, P. G. (2011). Frankia assemblages associated with Alnus rubra and Alnus viridis are strongly influenced by host species identity. *International Journal of Plant Sciences*, *172*(3), 403–410.

Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S., … Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume Lotus japonicus. *Nature Communications*, *1*(1), 1–12.

Mandal, M. K., Chanda, B., Xia, Y., Yu, K., Sekine, K., Gao, Q., … Kachroo, P. (2011). Glycerol-3-phosphate and systemic immunity. *Plant Signaling & Behavior*, *6*(11), 1871–1874.

Matos, A. R., Hourton-Cabassa, C., Ciçek, D., Rezé, N., Arrabaça, J. D., Zachowski, A., & Moreau, F. (2007). *Alternative oxidase involvement in cold stress response of Arabidopsis thaliana fad2 and FAD3+ cell suspensions altered in membrane lipid composition*.

Mejnartowicz, L. (2008). Genetic variation within and among naturally regenerating populations of alder [Alnus glutinosa]. *Acta Societatis Botanicorum Poloniae*, *77*(2).

Melo, A. T. O., Bartaula, R., & Hale, I. (2016). GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics*, *17*(1), 1–15.

Mizutani, M. (2012). Impacts of diversification of cytochrome P450 on plant metabolism. *Biological and Pharmaceutical Bulletin*, *35*(6), 824–832.

Mizutani, M., & Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology*, *61*, 291–315.

Mizuno, N., Sugie, A., Kobayashi, F., & Takumi, S. (2008). Mitochondrial alternative pathway is associated with development of freezing tolerance in common wheat. *Journal of Plant Physiology*, *165*(4), 462–467.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841.

Neelakandan, A. K., Nguyen, H. T. M., Kumar, R., Tran, L.-S. P., Guttikonda, S. K., Quach, T. N., … Nguyen, H. T. (2010). Molecular characterization and functional analysis of Glycine max sterol methyl transferase 2 genes involved in plant membrane sterol biosynthesis. *Plant Molecular Biology*, *74*(4–5), 503–518.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*(3), 145–156.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated--modular approach. *Bioinformatics*, *26*(3), 419–420.

Pawlowski, K., & Sprent, J. I. (2007). Comparison between actinorhizal and legume symbiosis. In *Nitrogen-fixing actinorhizal symbioses* (pp. 261–288). Springer.

Perret, X., Staehelin, C., & Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews*, *64*(1), 180–201.

Põlme, S., Bahram, M., Kõljalg, U., & Tedersoo, L. (2014). Global biogeography of Alnus-associated Frankia actinobacteria. *New Phytologist*, *204*(4), 979–988. https://doi.org/10.1111/nph.12962

Reddy, A. M., Reddy, V. S., Scheffler, B. E., Wienand, U., & Reddy, A. R. (2007). Novel transgenic rice overexpressing anthocyanidin synthase accumulates a mixture of flavonoids leading to an increased antioxidant potential. *Metabolic Engineering*, *9*(1), 95–111.

Rho, H., Hsieh, M., Kandel, S. L., Cantillo, J., Doty, S. L., & Kim, S.-H. (2018). Do endophytes promote growth of host plants under stress? A meta-analysis on plant stress mitigation by endophytes. *Microbial Ecology*, *75*(2), 407–418.

Rhoden, S. A., Garcia, A., Azevedo, J. L., Pamphile, J. A., & others. (2013). In silico analysis of diverse endophytic fungi by using ITS1-5, 8S-ITS2 sequences with isolates from various plant families in Brazil. *Genet. Mol. Res*, *12*, 935–950.

Rodriguez, D., Guerra, T. M., Forstner, M. R. J., & Hahn, D. (2016). Diversity of Frankia in soil assessed by Illumina sequencing of nifH gene fragments. *Systematic and Applied Microbiology*, *39*(6), 391–397.

Rodriguez, R. J., White Jr, J. F., Arnold, A. E., & Redman, a R. S. (2009). Fungal endophytes: diversity and functional roles. *New Phytologist*, *182*(2), 314–330.

Ruffley, M., Smith, M. L., Espíndola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, *27*(4), 1012–1024. https://doi.org/10.1111/mec.14491

Samac, D. A., & Graham, M. A. (2007). Recent advances in legume-microbe interactions: recognition, defense response, and symbiosis from a genomic perspective. *Plant Physiology*, *144*(2), 582–587.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, *16*(7), 372–380.

Shine, M. B., Xiao, X., Kachroo, P., & Kachroo, A. (2019). Signaling mechanisms underlying systemic acquired resistance to microbial pathogens. *Plant Science*, *279*, 81–86.

Shu, M. (2020). *Association genetics of drought tolerance in ponderosa pine (Pinus ponderosa)*. UC Merced.

Simms, E. L., & Taylor, D. L. (2002). Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative Biology*, *42*(2), 369–380.

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I., & Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, *13*(2), 211–221. https://doi.org/10.1111/pbi.12249

Steinkellner, S., Lendzemo, V., Langer, I., Schweiger, P., Khaosaad, T., Toussaint, J.-P., & Vierheilig, H. (2007). Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules*, *12*(7), 1290–1306.

Storz, J. F. (2005). Nonrandom dispersal and local adaptation. *Heredity*, *95*(1), 3–4.

Stuart, K. C., Cardilini, A. P. A., Cassey, P., Richardson, M. F., Sherwin, W. B., Rollins, L. A., & Sherman, C. D. H. (2021). Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology*, *30*(6), 1419–1434.

Tel-Zur, N., Abbo, S., Myslabodski, D., & Mizrahi, Y. (1999). Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae). *Plant Molecular Biology Reporter*, *17*(3), 249–254.

Thalmann, M., & Santelia, D. (2017). Starch as a determinant of plant fitness under abiotic stress. *New Phytologist*, *214*(3), 943–951.

Theocharis, A., Bordiec, S., Fernandez, O., Paquis, S., Dhondt-Cordelier, S., Baillieul, F., … Barka, E. A. (2012). Burkholderia phytofirmans PsJN primes Vitis vinifera L. and confers a better tolerance to low nonfreezing temperatures. *Molecular Plant-Microbe Interactions*, *25*(2), 241–249.

Thomas, K., & Turner, J. (2006). Red Alder Stand Establishment Decision Aid. *Journal of Ecosystems and Management*, *7*(2).

Vieira, M. L. A., Hughes, A. F. S., Gil, V. B., Vaz, A. B. M., Alves, T. M. A., Zani, C. L., … Rosa, L. H. (2012). Diversity and antimicrobial activities of the fungal endophyte community associated with the traditional Brazilian medicinal plant Solanum cernuum Vell.(Solanaceae). *Canadian Journal of Microbiology*, *58*(1), 54–66.

Villemereuil, P. De, & Gaggiotti, O. E. (2015). *A new F ST -based method to uncover local adaptation using environmental variables*. 1248–1258. https://doi.org/10.1111/2041-210X.12418

Wang, T., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. (2012). ClimateWNA—high-resolution spatial climate data for western North America. *Journal of Applied Meteorology and Climatology*, *51*(1), 16–29.

Wall, L. G. (2000). The actinorhizal symbiosis. *Journal of Plant Growth Regulation*, *19*(2), 167–182.

Weaver, L. M., & Herrmann, K. M. (1997). Dynamics of the shikimate pathway in plants. *Trends in Plant Science*, *2*(9), 346–351.

Winfield, M. O., Lu, C., Wilson, I. D., Coghill, J. A., & Edwards, K. J. (2010). Plant responses to cold: transcriptome analysis of wheat. *Plant Biotechnology Journal*, *8*(7), 749–771.

Yamori, W. (2016). Photosynthetic response to fluctuating environments and photoprotective strategies under abiotic stress. *Journal of Plant Research*, *129*(3), 379–395.

Zhao, Y.-J., Cheng, Q.-Q., Su, P., Chen, X., Wang, X.-J., Gao, W., & Huang, L.-Q. (2014). Research progress relating to the role of cytochrome P450 in the biosynthesis of terpenoids in medicinal plants. *Applied Microbiology and Biotechnology*, *98*(6), 2371–2383.

Chapter 4

Discussion

This study on the genetics of the commercially and ecologically important red alder has shown the advantage of using high-density SNP profiles in the elucidation of genetic processes involved at the population level, as well as identifying signatures of selection in local environments and symbiont selection. In chapter 2 I showed that the uplift of the Cascade mountain range and the onset of cyclic glacial cycles during the Quaternary had a significant impact on the distributions of red alder during the last ~3my. The repeated advance and retreat of ice sheets created new habitat allowing alders to advance from their refugia into the freshly disturbed areas, only to be forced back into refugia as the ice sheets returned. Drift associated with reduced population size and differing abiotic stresses in these refugia likely lead to the patterns of population structuring seen in the coastal populations. In both coastal populations genetic diversity was comparatively low for a highly outcrossing species but similar to other woody species in the Pacific Northwest (PNW; see Xie 2002). The isolation of red alders in glacial refugia would have restricted gene flow leading to reduction of heterozygosity through genetic drift. In the Northern Idaho Rocky Mountains (NRM) alders would have likely experienced a bottleneck and could have increased the rate of selfing to help counter keep alleles from drifting to fixation (Xie 2002). It is likely that the Idaho admixed has the highest H_O of all the populations because it shows evidence of repeated admixture with the other identified clusters. These areas are important for conservation efforts because they represent areas of genetic diversity for the species. The lack of uniformity in the disjunct populations and our timed phylogeny indicate ancient vicariance within the alders in the

NRM (Brunsfeld 2001), however it is unclear if this vicariance is due to alders that have survived in the NRM since before the uplift of the Cascade Mt. range, or if the alders were originally extirpated in the NRM and their current distribution was due to an ancient dispersal event. My data indicates multiple dispersal events into the NRM beginning as early as 2.6mya. So far red alder is the first temperate species in the disjunct NRM to have shown more than one dispersal event into the inland ranges (Brunfeld 2001; but see Ruffley *et al*. (2018), however as NGS studies of the region become more common, I believe that more species will be identified to have had multiple dispersal events in the future.

The advance and retreat of ice sheets occurred in 100ky intervals with melting and reforming cycles lasting roughly 10ky. Massive cataclysmic flooding occurred with many of these cycles (McDonald & Busacca 1992; Berger & Busacca 1995; Bjornstad *et al* 2001). and the estimated time of these floods match up with the estimated dispersal times we predicted in our analysis. These cataclysmic events caused flooding on a level that is difficult to comprehend when compared to contemporary natural phenomenon. The resulting disturbance could have provided an avenue for red alder to disperse into the NRM more readily which would explain why there were multiple dispersal events from both northern and southern genotypes. Red alder seeds are tailor made for dispersal over long distances and the dispersal could have been aided by animal vectors such as the goldfinch, of which many migrate along a corridor that follows the ancient floodplains. The use of flood plains as a dispersal corridor is also supported by the fact that all the known disjunct red alder populations in the NRM, as well as alders from both the

northern and southern coastal clusters have been identified in proximity to the ancient flood plains. As mentioned before, I believe more species will be identified to have had multiple dispersals into the NRM, however more research is needed to determine if the multiple dispersal events from the coast into the NRM identified in our study were unique and made possible by the excellent dispersal ability of alder seeds and status as a pioneer species, or if it was a common occurrence for other species as well throughout the Quaternary.

Because polyploidy is common in pioneer species, and polyploids more frequently occupy recently glaciated areas, polyploids would have likely been more abundant in these disturbed areas than in the refugia. The pattern of melting and then reforming of the ice sheets would have been repeated over and over in roughly 100ky cycles and it is likely that these repeated cycles would have produced enough polyploids to overcome barriers in reproductive fitness (Levin 1975) that might have otherwise kept polyploid alders from becoming established. The pattern of polyploid radiation from refugia that would have occurred during each previous cycle would have repeated itself. Because red alder likely survived in at least two separate refugia, the subsequent dispersal of polyploids from these refugia after the LGM likely led to the large mixed cytotype zone identified in our study. It also seems likely that anthropogenic disturbance seen throughout red alder's range has provided new habitat for polyploids to invade which may also help facilitate the persistence of polyploid red alder populations in their contemporary distributions.

In chapter 3 we utilized our high-density SNP profile to look for signatures of local adaption to environmental variables and selection of specific *Frankia* strains. The NRM red alder sites encounter significantly different abiotic conditions than those found in the coastal range. Because of the difference in these environmental conditions; the four sub-populations identified in chapter 2–along with our BEAST analysis–identified areas of genetic diversity and limited gene flow into the NRM providing an excellent system to look for genetic markers involved in selection. Five SNPs were found to associate with temperature related environmental variables. Interestingly, only one of these markers was from red alder somatic DNA, and this SNP yielded no protein hits in BLAST. The rest of the significant SNPs were from organelles (2 mitochondria, 1 chloroplast), or from an endophyte of red alder. This suggests that tolerance to environmental stress is more complicated than just how the host plants genetics can adapt, but rather that organelles and endophytes are selected to help mediate abiotic stresses in different environments in conjunction with the host plant genetics. Evidence for mitochondria selection to mediate temperature stresses has been shown in other plants (see Atkin *et al*. 2005 for a review), and evidence indicates that this type of selection is more common in fast-growing pioneer species (Atkin *et al*. 2005). Interestingly this can be quantitatively inherited in the plant lineage (De Santis *et al*. 1999). Temperature effects mediation isn't limited to mitochondria, temperature stress tolerance as it relates to endophytes is well known, particularly as it pertains to crop plants (see Miliute *et al*. 2015 for a review). Resultantly, commercialization of endophytes and their products is integral to the future of crop improvements (Chitnis *et al*. 2020).

One more interesting aspect of the environmental association analysis in this study is the SNPs associating with non-somatic DNA. This is likely only possible because of the mock reference genome utilized in GBS-SNP-CROP. If red alder had a reference genome these associations would likely not have been identified because they would not have been included in the alignment aspect of the SNP calling. This opens up potential use of GBS for mitochondrial, chloroplast, or endophyte study in future NGS studies.

For the second part of chapter 3 we used our high-density SNP profile to look for signatures of selection to specific *Frankia* strains to help elucidate the genetics involved in host-symbiont communication in this ecologically important N-fixing symbiosis. To accomplish this we developed the novel approach of using *Frankia* genotype as a red alder phenotype in a GWAS study. The *Frankia* in our study exhibited the low diversity usually seen in the species (Kennedy *et al*. 2010) with one genotype being found in all sampling locations consisting of nearly 80% of all *Frankia* we identified. A second genotype consisted of five red alders with all remaining alders consisting of rare genotypes only found in a few select locations. Using this information in our GWAS identified two SNPs that showed significant associations. One of these SNPs was shown to be a protein that is important in systemic acquired resistance in plants. Systemic acquired resistance is important in the sequestering of non-target species in legume-*rhizobia* symbioses and offers an intriguing line of inquiry for future studies. The second SNP associated with a cytochrome p450 family of enzymes that have been identified to be instrumental in bacteria plant communications in both pathogenic response and

legume-*rhizobia* symbioses as well. The potential for this information to be utilized in future crop improvement strategies has only just began to be understood.

Conclusion

Utilizing our high-density SNP profile, we were able to identify four sub-populations in red alder in the PNW, two on the coast and two in the NRM. The two sub-population in the NRM are areas of unique genetic diversity in which have been shaped by range reductions related to the uplift of the Cascades, and vacillations in ice sheets during the last ~2.5 my. We found evidence of AV and for multiple dispersal events from the coast into the NRM which may have been aided by massive flooding in time of dynamic change in the PNW, and biological vectors. Polyploidy was also identified and found to be common throughout our sampling range. This is likely due to glaciation and the life history of red alder as pioneer species are frequently polyploid, and polyploidy is common in previously glaciated areas following the LGM. Polyploidy, multiple dispersals and possibly increased selfing due to limited population size in the NRM is likely responsible for the higher-than-expected heterozygosity observed in the NRM compared to the coast. The disjunct populations in the NRM provide an avenue of research in population genetics, inland migrations, and potential speciation. We believe that as NGS sequencing is utilized in more genetic studies on disjunct populations in the NRM, patterns of multiple dispersals will be identified in other species as well.

For our association studies, initially we expected our associations to be related to precipitation and humidity, however our results do make sense. Red alder is usually found within 200 km of the Pacific coast and coastal areas tend to see smaller

temperature variation than similar latitudes inland. This is due to the high specific heat of water and it is exactly what we found in the Idaho sites. In Idaho, the number of DD0 were nearly 7 times the average of the rest of the sites, and there were significant associations on the other extreme dd<18. Idaho sites were the only ones to have a mean coldest month temperature below 0, had 50 fewer frost-free days. The temperature difference from the mean warmest and coldest months was also much higher in the Idaho sites (ave=5x) compared to the rest of the sites, all of which indicate that red alder in Idaho have to deal with a much higher fluctuation in temperatures, not only daily but annually as well. We suggest that alder may have survived in the wider cold swings by selecting for mitochondria that function more efficiently, and endophytes that are able to help regulate metabolic activities when it faces extreme temperatures. Even if the advantage is minimal, red alder is shade intolerant and the competition for light is key to their survival. Even a slight advantage could be the difference between growing taller than neighboring plants and flourishing or falling behind and getting shaded out. Using *Frankia* genotype as a phenotype in our GWAS we identified to candidate marker loci involved in *Frankia* selection. Both SNPs yielded protein hits that provide key insights into the genetics involved in the selection of specific *Frankia* genotypes in this valuable N-fixing symbiosis. This information is invaluable for future crop improvement strategies not only for red alder, but potentially for many more species in the future if the genetics involved in the symbioses can be further elucidated.

Future direction

Our research is key first step in understanding the genetics of red alder which will be useful for crop improvement and selection of individuals better suited for habitat restoration. Specifically, the NRM populations are a source of genetic variation that does not exist in the coastal range. The disjunct nature and the existence of ancient vicariance in the NRM provides a fertile avenue of research on many topics including speciation and local adaptation. Future research should look at the whole of the range of red alder including Canada, and into southern California. It would be interesting to see if more population structuring exists in the coastal populations, and if it could be attributed to past environmental events as well. Of particular interest would be to look for evidence of more migration events into the NRM. The ease in which red alder and other plants dispersed into the NRM at the end of the LGM makes it seem unlikely that the five dispersal events we identified were the only ones in the last 2.5 my. The pumice plain on Mt. St. Helens would be an interesting place to look more in depth at polyploidy in red alder. We did find some evidence that there could be octoploid or higher cytotypes may exist in this narrow sampling area. It would be interesting to see if this was actually the case, and to what extent the massive devastation by the eruption had on ploidy levels of those alders that were first able to recolonize the pumice plain after the eruption. Perhaps the most interesting find of the entire project was the SNPs that associated with specific *Frankia* strains. In the future, a much larger SNP array consisting of thousands of SNPs from hundreds of red alders throughout the entire range could help identify more SNPs involved in this symbiosis. The potential gain in understanding the genetics involved in this process is almost without measure if in the future we were able to engineer crops that could nodulate with N_2 fixing bacteria and fix their own nitrogen. The impact of

fertilizers on freshwater ecosystems leading to eutrophication is devastating to local communities. The ability to reduce our reliance on fertilizers would have cascading effects throughout freshwater ecosystems and may one day help us eliminate the dead zones that exist of the coasts where our biggest rivers meet the oceans.

References

Atkin, O. K., Bruhn, D., Hurry, V. M., & Tjoelker, M. G. (2005). Evans Review No. 2 - The hot and the cold: Unravelling the variable response of plant respiration to temperature. *Functional Plant Biology*, *32*(2), 87–105. https://doi.org/10.1071/FP03176

Berger, G. W., & Busacca, A. J. (1995). Thermoluminescence dating of late Pleistocene loess and tephra from eastern Washington and southern Oregon and implications for the eruptive history of Mount St. Helens. *Journal of Geophysical Research: Solid Earth*, *100*(B11), 22361–22374.

Bjornstad, B. N., Fecht, K. R., & Pluhar, C. J. (2001). Long history of pre-Wisconsin, ice age cataclysmic floods: evidence from southeastern Washington state. *The Journal of Geology*, *109*(6), 695–713.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2000). Comparative phylogeography of north- western North America : a synthesis. *Integrating Ecological and Evolutionary Processes in a Spatial Context*, 319–339.

Chitnis, V. R., Suryanarayanan, T. S., Nataraja, K. N., Prasad, S. R., Oelmüller, R., & Shaanker, R. U. (2020). Fungal Endophyte-Mediated Crop Improvement: The Way Ahead. *Frontiers in Plant Science*, *11*.

De Santis, A., Landi, P., & Genchi, G. (1999). Changes of mitochondrial properties in maize seedlings associated with selection for germination at low temperature. Fatty acid composition, cytochrome c oxidase and adenine nucleotide translocase activities. *Plant Physiology*, *119*(2), 743–754. https://doi.org/10.1104/pp.119.2.743

Kennedy, P. G., Schouboe, J. L., Rogers, R. H., Weber, M. G., & Nadkarni, N. M. (2010). Frankia and Alnus rubra canopy roots: an assessment of genetic diversity, propagule availability, and effects on soil nitrogen. *Microbial Ecology*, *59*(2), 214–220.

Levin, D. A. (1975). Minority cytotype exclusion in local plant populations. *Taxon*, *24*(1), 35–43.

McDonald, E. V, & Busacca, A. J. (1992). Late Quaternary stratigraphy of loess in the Channeled Scabland and Palouse regions of Washington State. *Quaternary Research*, *38*(2), 141–156.

Miliute, I., Buzaite, O., Baniulis, D., & Stanys, V. (2015). Bakterinių endofitų reikšmė žemės ūkio augalų atsparumui stresui: Apžvalga. *Zemdirbyste*, *102*(4), 465–478. https://doi.org/10.13080/z-a.2015.102.060

Ruffley, M., Smith, M. L., Espíndola, A., Carstens, B. C., Sullivan, J., & Tank, D. C. (2018). Combining allele frequency and tree-based approaches improves phylogeographic inference from natural history collections. *Molecular Ecology*, *27*(4), 1012–1024. https://doi.org/10.1111/mec.14491

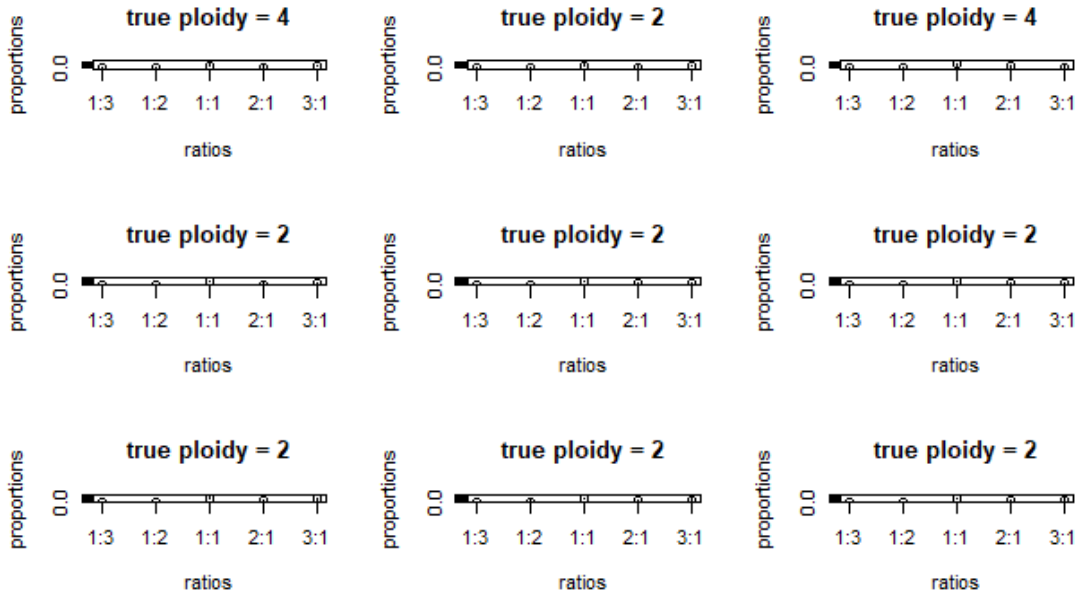
Xie, C.-Y., El-Kassaby, Y. A., & Ying, C. C. (2002). Genetics of red alder (Alnus rubra Bong.) populations in British Columbia and its implications for gene resources management. *New Forests*, *24*(2), 97–112.

Appendix A. Chapter 2 Supplemental

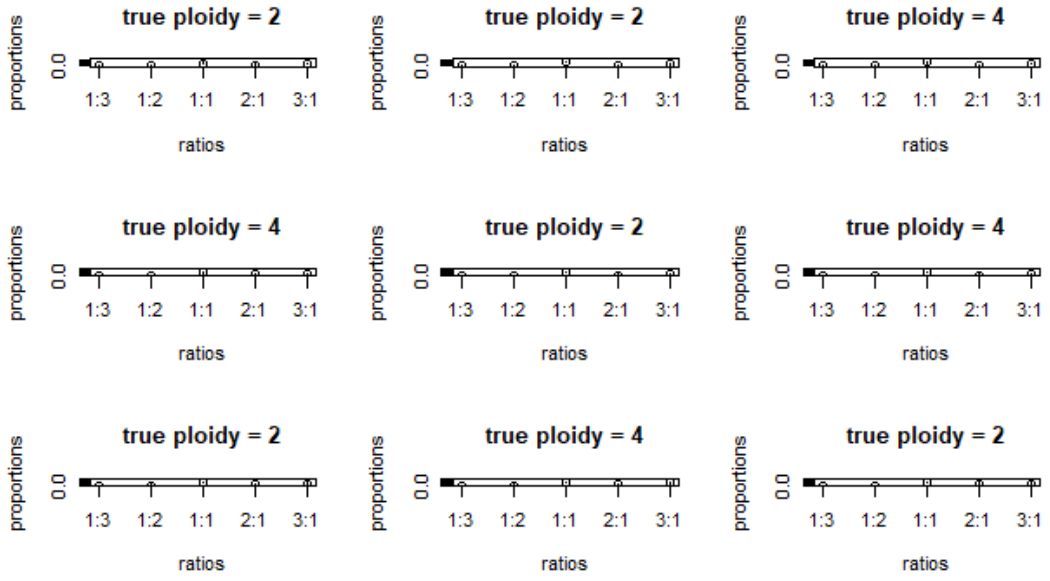
List of numerical designation for red alder samples.

1	t-1	34	be-5	67	id-9
2	t-2	35	be-6	68	id-10
3	t-3	36	be-7	69	ws-1
4	t-4	37	be-8	70	ws-2
5	t-5	38	be-9	71	ws-3
6	t-6	39	be-10	72	ws-4
7	t-7	40	ho-1	73	ws-5
8	t-8	41	ho-2	74	ws-6
9	t-9	42	ho-3	75	ws-7
10	t-10	43	ho-4	76	ws-8
11	st-1	44	ho-5	77	ws-9
12	st-2	45	ho-6	78	ws-10
13	st-3	46	ho-7	79	ws-11
14	st-4	47	ho-8	80	ws-12
15	st-5	48	rt-1	81	ws-13
16	st-6	49	rt-2	82	ws-14
17	st-7	50	rt-3	83	ws-15
18	st-8	51	rt-4	84	wa-1
19	st-9	52	rt-5	85	wa-2
20	st-10	53	rt-6	86	wa-3
21	st-11	54	rt-7	87	wa-4
22	st-12	55	rt-8	88	wa-5
23	st-13	56	rt-9	89	wa-6
24	st-14	57	rt-10	90	wa-7
25	st-15	58	rt-11	91	wa-9
26	po-1	59	id-1	92	wa-10
27	po-2	60	id-2	93	wa-11
28	de-1	61	id-3	94	wa-12
29	de-3	62	id-4	95	wa-14
30	be-1	63	id-5		
31	be-2	64	id-6		
32	be-3	65	id-7		
33	be-4	66	id-8		

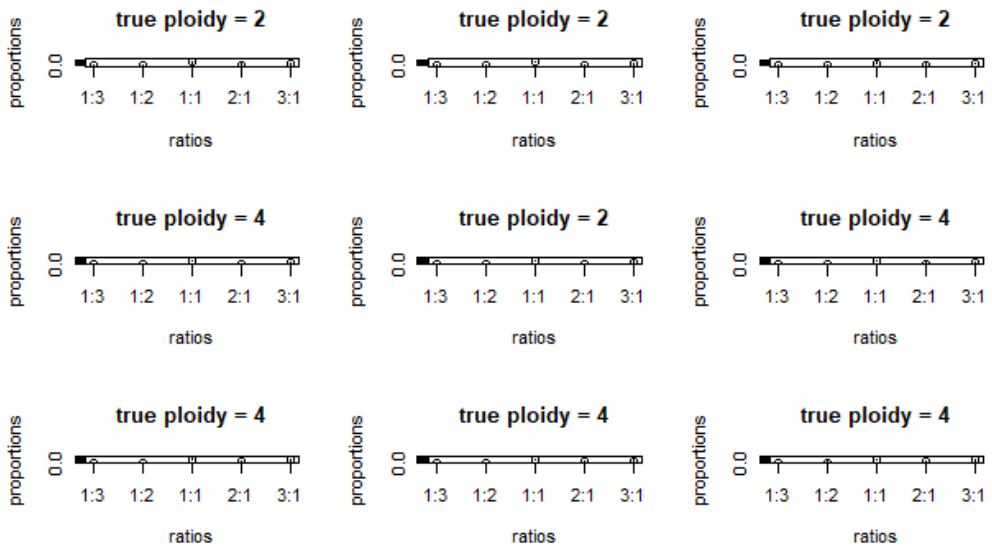
1-9


true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 2
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

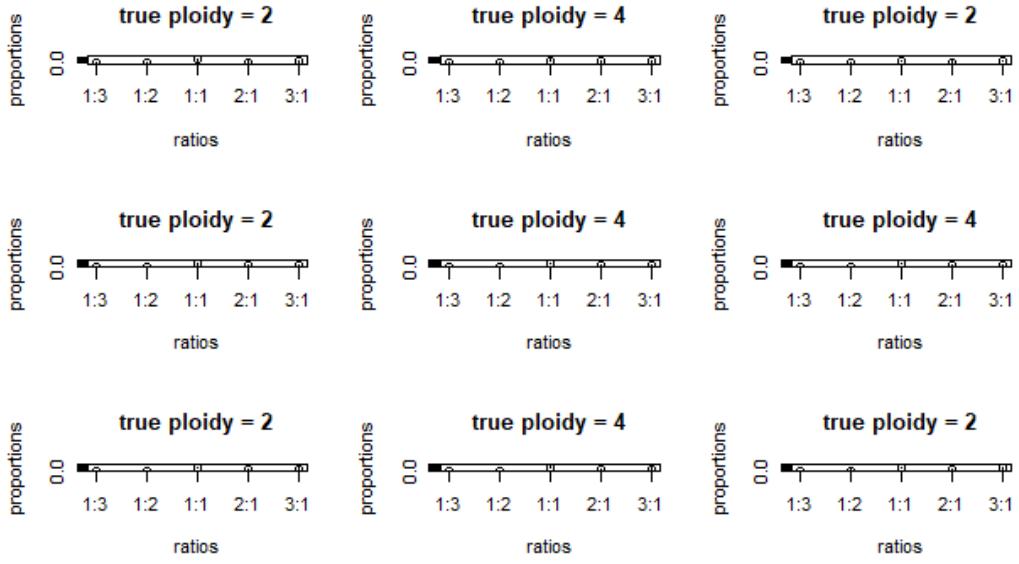
10-18


true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 2
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

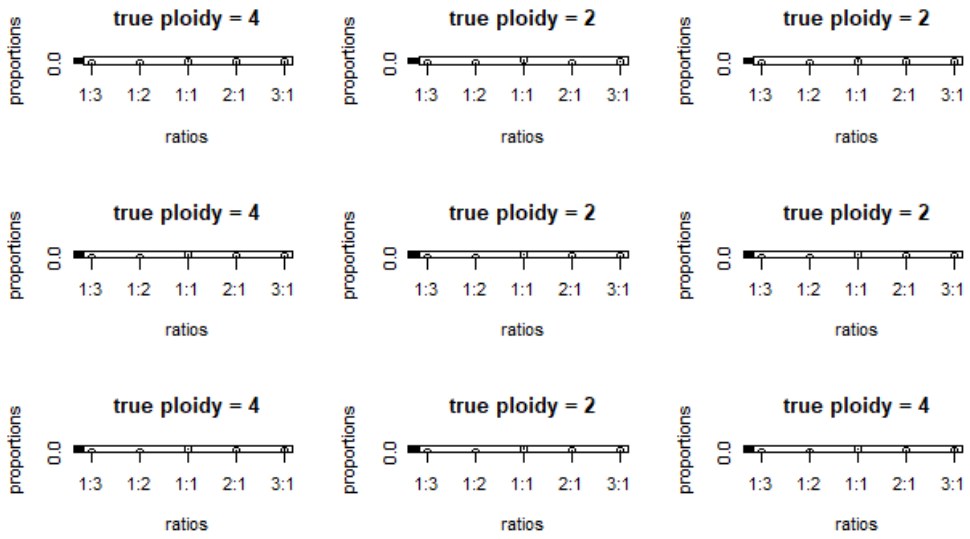
19-27


true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 4
true ploidy = 4
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

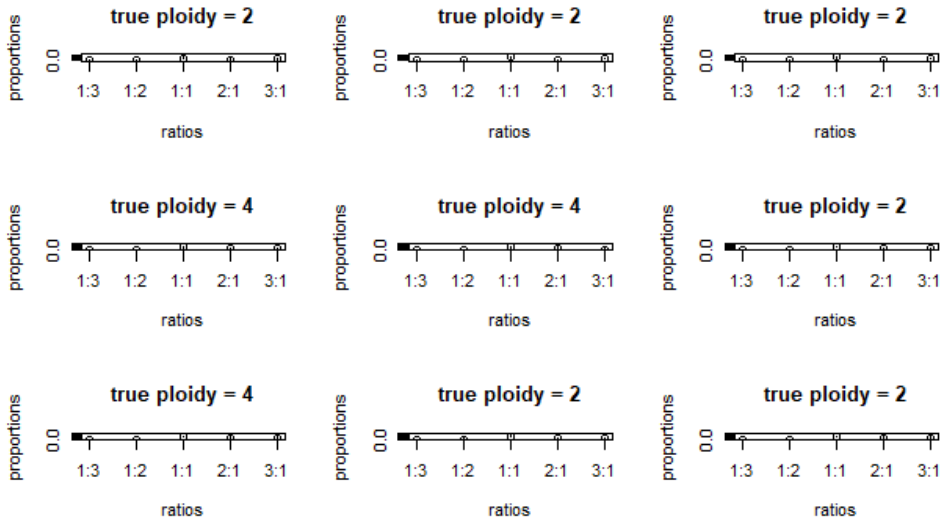
28-36


true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 2
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

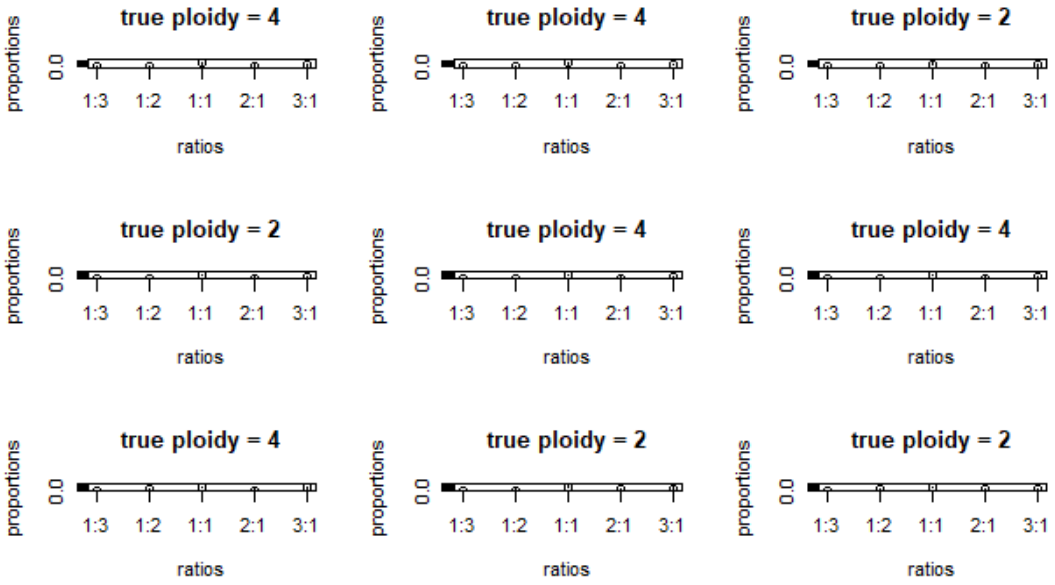
37-45


true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 4
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

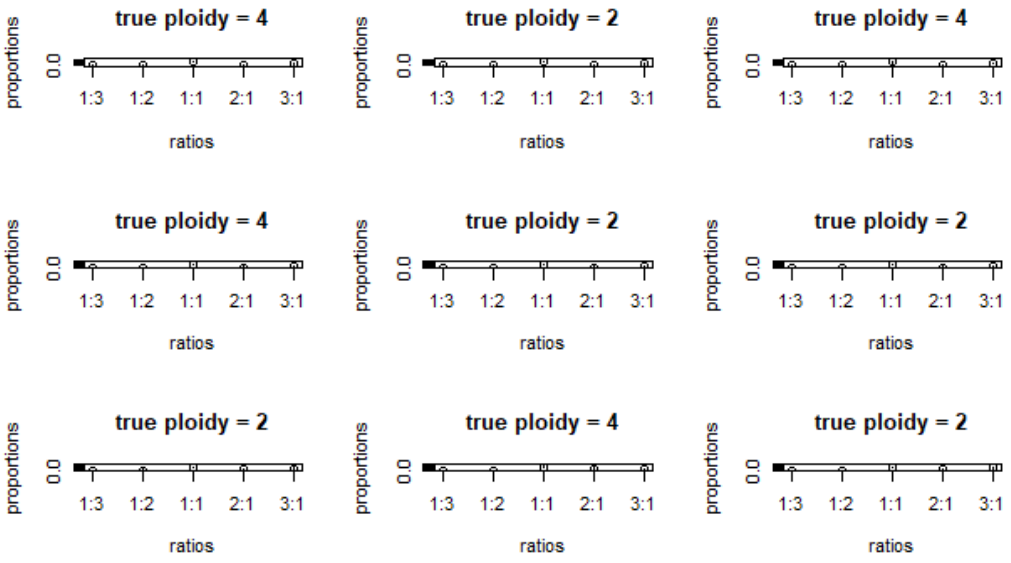
46-54


true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 2
true ploidy = 2
proportions
0.0
1:3
1:2
1:1
2:1
3:1
ratios

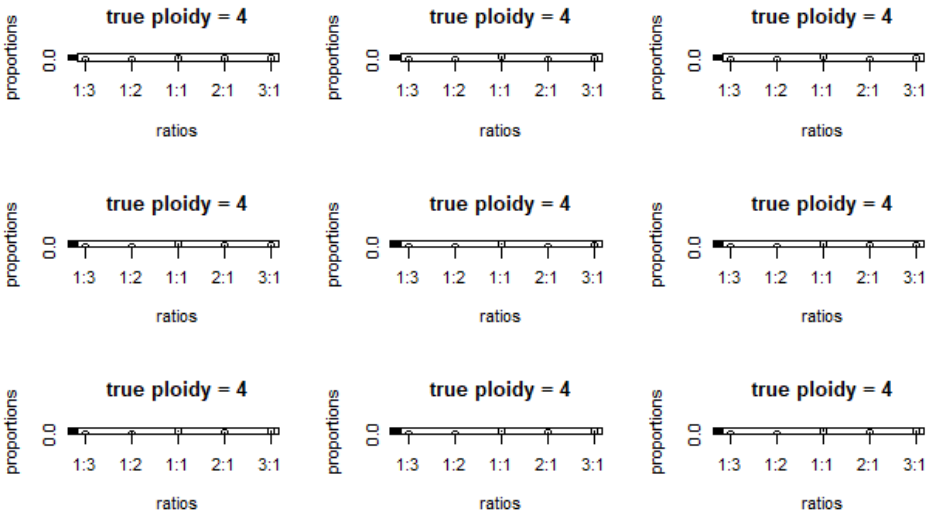
55-63


true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 2
proportions
0.0
1:3 1:2 1:1 2:1 3:1
ratios

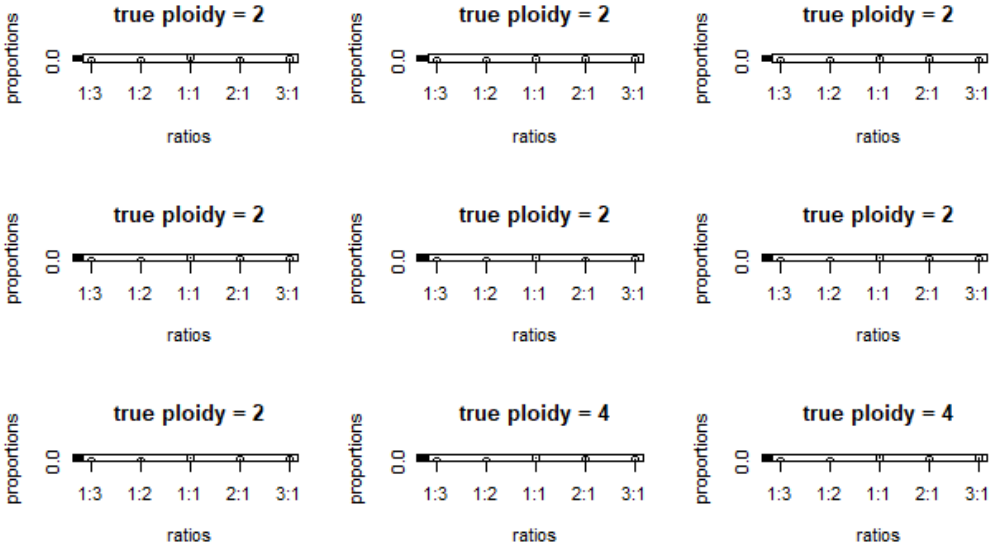
64-72


true ploidy = 4
true ploidy = 2
true ploidy = 4
true ploidy = 4
true ploidy = 2
true ploidy = 2
true ploidy = 2
true ploidy = 4
true ploidy = 2
proportions
0.0
1:3 1:2 1:1 2:1 3:1
ratios

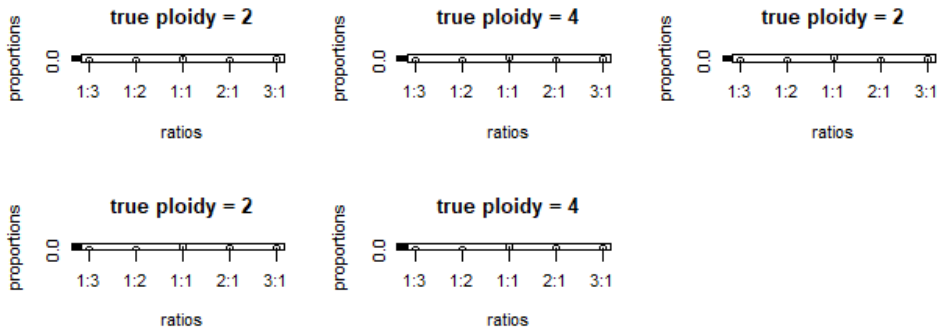
73-81


true ploidy = 4
proportions
0.0
1:3 1:2 1:1 2:1 3:1
ratios

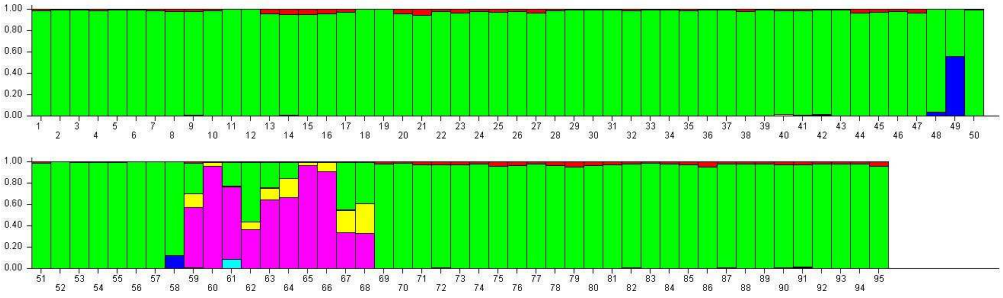
82-90


true ploidy = 2
true ploidy = 4
proportions
0.0
1:3 1:2 1:1 2:1 3:1
ratios

91-95



List and figures of ploidy results from GBS2ploidy on 95 red alder samples from the Pacific Northwest



Structure output showing sub-populations 5 and 6 fitted in the Idaho sites.

Appendix B. Chapter 3 Supplemental

Table showing 30 year averages for 20 environmental variables.

Area	MAT	MWMT	MCMT	TD	MAP	MSP
Be	10.6	14.9	6.6	8.3	1844	287
De	11.2	18.9	4.7	14.2	1310	223
Ho	10.2	18.5	2.9	15.6	1680	259
Id	7.9	19	-2.6	21.6	855	273
Po	11.9	19.8	4.5	15.3	1207	203
Rt	11.8	19.6	5	14.6	1037	140
St	9.2	17.8	1.4	16.4	2343	344
Ti	10.7	16.7	5.4	11.3	2967	434
Wa	9.4	17.1	2.4	14.7	1681	313
Ws	10	17.2	3.7	13.5	2501	380

Area	DD<18	DD>18	NFFD	FFP	bFFP	eFFP
Be	2783	11	328	252	74	326
De	2590	172	309	222	93	315
Ho	2928	150	287	191	112	303
Id	3772	167	202	130	139	269
Po	2388	241	323	258	69	327
Rt	2422	214	313	222	94	316
St	3254	99	275	193	111	304
Ti	2737	66	314	228	95	323
Wa	3212	76	277	181	118	299
Ws	2964	88	294	197	109	306

Area	AHM	SHM	DD0	EMT	PAS	Eref
Be	11.2	51.8	6	-11.3	23	715
De	16.2	84.6	21	-14.9	27	899
Ho	12	71.5	56	-18.2	76	854
Id	20.9	69.7	320	-29.3	140	872
Po	18.1	97.6	23	-14.1	24	833
Rt	21	140.1	17	-14.7	19	963
St	8.2	51.8	111	-20	176	759
Ti	7	38.5	13	-13.3	57	790
Wa	11.5	54.6	70	-19.1	82	766
Ws	8	45.3	37	-16.5	80	777