1-1-2010

Phylogeographic Patterns and Intervarietal Relationships within Lupinus lepidus: Morphological Differences, Genetic Similarities

Kevin Allen Weitemier
Portland State University

Phylogeographic Patterns and Intervarietal Relationships within *Lupinus lepidus*: Morphological Differences, Genetic Similarities

by

Kevin Allen Weitemier

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

Master of Science
in
Biology

Thesis Committee:
Mark Fishbein, chair
Mitchell Cruzan
Susan Masta

Portland State University

Abstract

Lupinus lepidus (Fabaceae) contains many morphologically divergent varieties and was restricted in its range during the last period of glaciation. A combination of phylogenetic (with the *trnDT* and *LEGCYC1A* loci) and population genetics approaches (with microsatellites and *LEGCYC1A*) are used here to characterize intervarietal relationships and examine hypotheses of recolonization of areas in the Pacific Northwest affected by glaciation. Sequenced loci are not found to form a clade exclusive to *L. lepidus*, nor are any of the varieties found to form clades. Population genetics analyses reveal only negligible genetic structure within *L. lepidus*, with the majority of variation being found within populations. Isolation-by-distance analysis reveals some correlation between population genetic distances and geographic distance. Microsatellite and sequence results are consistent with a scenario whereby the Oregon and Washington regions were rapidly colonized from the south, with independent invasions along the eastern and western sides of the Cascade Mountains. A predicted disjunction between northern and southern populations is found within the microsatellite data but not the sequence data, suggesting that northern populations were recolonized via a process involving the spread of novel microsatellite mutations, perhaps through the persistence of a glacial refuge isolated from southern populations. Varieties are not shown to be genetically isolated, and are interpreted as representing ecotypes, with local selection outpacing the effects of migration.

Acknowledgments

This work would not have been possible without the support of the Portland State University Biology Department faculty and support staff. I would particularly like to thank my thesis advisor Mark Fishbein and committee members Mitch Cruzan and Susan Masta: their advice and comments improved every stage of this work. Dr. Cruzan deserves special thanks for making me welcome in his lab during this project's final year. Thanks also go to Diane Bland for assistance handling herbarium specimens, John Bishop for providing critical samples from Mt. Baker and Mt. Lassen, Chris Drummond and Veronica Morris for providing advance copies of their work, and James Bryne for assistance with the intra-population sequencing. Financial support for this work was provided by the Society of Systematic Biologists, the American Society of Plant Taxonomists, the Native Plant Society of Oregon, the Washington Native Plant Society, and the Portland State University Forbes-Lea Fund.

Table of Contents

Abstract i

Acknowledgments ii

List of Tables iv

List of Figures v

Chapter 1: Introduction 1

Chapter 2: Phylogenetic and population genetic analyses of DNA sequence data 14

Chapter 3: Population genetic analyses of microsatellite data 50

Chapter 4: Conclusion 67

References 78

List of Tables

Table 1: Sample collection information.....34

Table 2: Primers used for amplification, sequencing, and microsatellites 37

Table 3: Parameter estimates for nucleotide substitution models.....39

Table 4: LEGCYC1A-TCP variety summary statistics.....45

Table 5: LEGCYC1A-TCP population summary statistics.....46

Table 6: LEGCYC1A-TCP variety pairwise F_{ST} values.....47

Table 7: LEGCYC1A-TCP population pairwise F_{ST} values.....48

Table 8: LEGCYC1A-TCP AMOVA.....49

Table 9: Microsatellite variety summary statistics 63

Table 10: Microsatellite variety AMOVA.....63

Table 11: North-South structuring of genetic diversity.....64

Table 12: Net nucleotide distance between Structure clusters.....65

List of Figures

Figure 1: Schematic of LEGCYC1A 33

Figure 2: *Lupinus lepidus* sampling locations 36

Figure 3: Examples of two mock haplotypes.....38

Figure 4: LEGCYC1A-TCP ML tree 40

Figure 5: LEGCYC1A-TCP tree from Bayesian analysis 41

Figure 6: LEGCYC1A-TCP haplotype network.....42

Figure 7: *trnDT* ML tree 44

Figure 8: Population microsatellite pairwise F_{ST} values.....65

Figure 9: Microsatellite Structure analysis 66

Chapter 1

Introduction

To some extent, all species exist within limited geographic, environmental, and ecological ranges. Following significant environmental change, species may relocate, adapt, or face extinction. As a consequence, local biotas change through reassembly of communities and the origin and extinction of species. Geographic movements and range shifts can act as a potent evolutionary force: by splitting and creating new populations that can then evolve independently, and by bringing formerly separated populations into secondary contact. Comparative biogeographic studies of co-distributed species can illuminate historical factors that have contributed to the development of biotas and influenced evolutionary trajectories. Such studies are based on detailed investigations of the evolution and distribution of focal groups, especially of historical and contemporary factors that affect population genetic structure and phylogeography. Such studies also provide crucial information on what effects future alterations -- such as global climate change -- may have on whole ecosystems. Despite the increase in phylogeographic studies, many regions of the globe have received little or no study into the processes shaping their biogeographic history.

Pacific Northwest biogeography

The Cascade mountain range of the northwestern United States is a region that has received little biogeographic attention. This is true in spite of increased attention to the broader Pacific Northwest (PNW) region as a whole. Several studies examining the coastal PNW (Cascade and Coast ranges) and inland PNW (Northern Rocky Mountains

in Idaho) have found a genetic disjunction between the two areas. Causes for this disjunction may include an ancient vicariance between the two areas, as in the case of many amphibians (*Ascaphus*: Nielson et al. 2001; *Dicampton* and *Pleththo*don: Carstens et al. 2005), or movement from coastal to inland populations along a Northern (*Microtus, Salix*, and *Pinus*; Carstens et al. 2005) or Southern (*Alnus*, Strenge 1994) route, as seen in some mammalian and plant species.

Other studies have focused on a broad north-south axis within the PNW. These studies have generally focused on areas west of and including the Cascade mountain range. Soltis et al. (1997) reviewed analyses of cpDNA restriction site variation in seven plant species (*Tolmiea menzeisii, Tellima grandiflora, Tiarella trifoliata, Alnus rubra, Ribes bracteosum, Polysticum munitum,* and *Picea sitchensis*) occurring along the Cascade axis. Six of these species showed strong genetic disjunctions between northern and southern populations, with the contact area between the two groups generally falling within Oregon. Additionally, the southern populations displayed high genetic diversity, while the northern populations displayed very little. The exception, *Polysticum munitum*, also displayed northern and southern genotypes but had a much greater range of overlap and contained some southern genotypes in far northern locations.

To account for the prevalence of north-south genetic disjunctions, Soltis et al. proposed two hypotheses. The "north-south recolonization hypothesis" states that during the last period of glaciation (~20-12 kya), populations became relegated to the large southern portion of their ranges and retained one or more small refuge populations in their northern ranges. After glacial retreat, the refuge populations recolonized the region. The isolation of the northern and southern populations resulted in the disjunction between

northern and southern genotypes. The small northern refuges created a bottleneck effect, resulting in limited diversity of northern genotypes, and extended isolation from southern populations allowed the spread of novel mutations. Soltis et al. suggested potential locations of northern refuges, including north-central Idaho, Vancouver Island, the Queen Charlotte Islands, and unglaciated areas of Alaska.

The "leading edge hypothesis" also predicts a large southern refuge, but does not predict the existence of any northern refuges. In this model, high genetic diversity is retained in the southern population, which begins to recolonize areas after glacial retreat. A few long distance dispersals from the southern refuge create the “leading edge” of the expanding population. Subsequent dispersals, likely from the most northern subpopulations, expand the leading edge farther north. Many of these dispersals will represent bottleneck events, reducing genetic diversity in northern subpopulations.

Soltis et al. (1997) did not favor either hypothesis over the other, and subsequent studies of species demonstrating a north-south distribution within the PNW have failed to reject either hypothesis. A study of the red tree vole, *Phenacomys longicaudus*, may lend support for the north-south recolonization hypothesis, although it contains some confounding factors (Miller et al. 2006). While *P. longicaudus* exhibits a genetic disjunction between northern and southern populations, levels of genetic diversity are comparable between the two. Additionally, the range of *P. longicaudis* is very small, located almost entirely in Oregon and only west of the Cascades. Because of this, it is not thought that the northern areas of its range were directly glaciated. Instead, climatic conditions may have fragmented the forest habitat of *P. longicaudus*, creating an east-west disjunction in northern populations. This disjunction could mimic the secondary

contact that would be expected from the expansion of populations out of multiple northern refuges, following the north-south recolonization hypothesis.

Studies of the rough-skinned newt, *Taricha granulosa*, utilizing mtDNA and allozyme data support a range expansion from northern California north through Canada and into southern Alaska (Kuchta and Tan 2005). Unlike most species, *T. granulosa* is thought to have expanded its range during the last glacial maximum, expanding from California into Washington along the unglaciated Pacific coast. After glacial retreat, rapid expansion into Alaska took place. While this example of leading edge expansion is an original range expansion rather than a retreat and recolonization, it exhibits both a disjunction between northern and southern populations, and decreased diversity in the north.

The Pacific giant salamander, *Dicamptodon tenebrosus*, has a wider range than *P. longicaudus* and may show support for the north-south recolonization hypothesis. *D. tenebrosus* ranges from northern California to northern Washington west of the Cascades. In a phylogeographic study of *D. tenebrosus*, Steele and Storfer (2006) showed that northern and southern populations form monophyletic groups and were able to reject hypotheses involving single refuges. They were also able to reject a dual refuge hypothesis that placed the separation between clades during the last glacial maximum, as predicted by the recolonization hypotheses. Instead, Steel and Storfer were unable to reject dual refuge hypotheses that split the clades at a much earlier point (either mid-Pleistocene 800 kya, or early Pleistocene 1.7 mya). The refuges they proposed for *D. tenebrosus* were the Klamath-Siskiyou Mountains in the south, and the Columbia River Valley in the north.

A study of garter snakes, *Thamnophis sirtalis*, ranging from southern California to southern British Columbia may be consistent with the north-south recolonization hypothesis. Janzen et al. (2002) sequenced regions of the snakes' mtDNA and found support for three distinct clades: a California clade, a northwest clade (west of the Cascades), and a northeast clade (Oregon north, east of the Cascades). Relationships among these clades were unresolved. The authors conclude that these results generally reject hypotheses involving a single northern or southern refuge, in favor of an explanation invoking both refuges; however, the lack of resolution in this study makes this conclusion somewhat tenuous. Additionally, the range of *Thamnophis sirtalis* is much larger than that sampled by Janzen et al., extending to the east coast of North America. While the western populations are thought to form a single lineage, this is not certain and may confound comparisons with other phylogeographic analyses in the PNW.

With the exception of Soltis et al. (1997), biogeographic studies of plants distributed along the Cascade axis have been lacking. Those studies that do involve organisms along this axis strongly emphasize low elevation species and species limited to areas west of the Cascades. Studies on species occurring at high elevations within the Cascades and among the heterogeneous environments found on either side of the Cascades can offer fresh perspectives on Cascade biogeography including diversification along the range, locations of potential refuges, and patterns of recolonization after glacial events.

***Lupinus lepidus* species complex**

Lupinus lepidus Douglas ex Lindley (Fabaceae) is a perennial lupine ranging from the central Sierra Nevada mountains north into British Columbia and from Oregon east to the Rocky Mountains. This species contains several varieties, with the highest varietal diversity occurring in Oregon. Seven of the species' ten varieties occur in Oregon with the additional three varieties endemic to California (Oregon Flora Project, Hickman 1993, Isely 1998, Barneby 1989).

The wide morphological variation within this species complex has caused, and continues to cause, taxonomic difficulties. Among other treatments, the species has been referred to as the *L. lepidus – L. caespitosus* complex and divided into at least ten separate species, some of which contained several varieties (Cox 1973, who included ten species, two subspecies, and six varieties). The morphological variation is so high that "undue emphasis on minor races leads inexorably to the quicksands that ultimately engulfed Charles Piper Smith," (Barneby 1989) who identified half a dozen novel species in Idaho alone (Smith 1946). A consensus of regional floras delimits this group as a single species, *L. lepidus* Douglas ex Lindl., containing ten varieties: var. *aridus* (Douglas ex Lindl.) Jeps., var. *ashlandensis* (Cox) Isely, var. *confertus* (Kellogg) C.P.Sm., var. *culbertsonii* (Greene) C.P.Sm., var. *cusickii* (S.Wats.) C.L.Hitch., var. *lepidus*, var. *lobbii* (A.Gray ex S.Wats.) C.L.Hitch., var. *ramosus* Jeps., var. *sellulus* (Kellogg) Barneby, and var. *utahensis* (S.Wats.) C.L.Hitch. (Oregon Flora Project, Isely 1998, Hitchcock and Cronquist 1996, Hickman 1993, Barneby 1989). Two of these varieties, *ashlandensis* and *cusickii*, are endemic to very small regions within Oregon,

and variety *lepidus* has declined severely across its historic range in the Willamette Valley.

Lupinus lepidus varieties are characterized by a combination of several features. Plant heights range from very tall (*confertus*, >30 cm) to very short (*utahensis*, <5 cm). The exertion of the raceme relative to the basal cluster of foliage ranges from completely exerted (*lepidus, sellulus*) to sessile (*utahensis*). Leaflet size (very small in *lobbii* and *utahensis*), branching above the base (found in *confertus*, *ramosus*, and *cusickii*), and leaflets along the length of the stem (found in *confertus* and *ramosus*) can also be important characters. Floral characters are not as divergent between varieties as vegetative characters, but flower size, banner size and shape, and characters of the raceme are used to distinguish some varieties (*ashlandensis, cusickii, culbertsonii*).

Phylogenetic work has been performed on the *Lupinus* genus as a whole (Drummond 2008, Hughes and Eastwood 2006, Aïnouche 2004). These studies show *L. lepidus* to belong to a clade consisting of western North American perennial lupines that diverged relatively recently (0.7-2.1 mya) from other New World lupines (Drummond 2008). However, *L. lepidus* has received no specific attention using molecular sequence data or quantitative analyses. Cox (1973) could only draw tentative conclusions based on morphological and protein-based characteristics: generally, that the "*lepidus-caespitosus* complex" originated in the southern Sierra Nevada range, and that the morphological diversity arose through local adaptation compounded by very low to nonexistent gene flow. (With heavy seeds showing no overt adaptations for dispersal, *L. lepidus* has a weak ability for local dispersal, with most seedlings growing within 2 m of their maternal parent [Wood and del Moral 1987]. Long-distance dispersal ability is unknown.)

Clearly, the morphological diversity within *L. lepidus* can provide a rich arena for taxonomic delimitation and the study of local adaptation. In addition, three of its varieties are ideally located to inform biogeographic studies of the Cascade mountain range. Variety *lepidus* occurs at low elevations from southern British Columbia to northwest Oregon west of the Cascades. Variety *lobbii* occurs at higher elevations (subalpine to alpine) along the Cascade crest and upper slopes on both sides of the range. Populations of var. *lobbii* are also found in the Coast range and at higher elevations in eastern Oregon and Idaho. Variety *sellulus* occurs in northern California and southern Oregon, extending north and east of the Cascades nearly to the Columbia River. Should these varieties be found to form a single lineage, the biogeographic conclusions that could be drawn from their study would be especially great.

There are some similarities between *Lupinus* and the other systems previously studied. For example, both the *Thamnophis* and *Lupinus* systems are spread across a wide geographic range, particularly in latitude; both species occur on both sides of the Cascade and Sierra Nevada mountains; and both contain several morphologically distinct subspecific taxa. However, comparisons between *Lupinus* and these other systems are necessarily limited due to the major taxonomic and life history differences.

Nevertheless, *L. lepidus* provides an opportunity to study biogeographic patterns in this region from a novel perspective: *L. lepidus* varieties occur at a range of altitudes, they are found on either side of the mountain range, and along the entire length of the range. This provides the potential for *L. lepidus* to be highly informative regarding patterns of diversification, responses to glaciation of high elevation organisms, and postglacial recolonization of a variety of habitats.

Molecular techniques

Patterns of diversification can be measured through the observation of molecular DNA markers found in the organisms of interest. Information can be gathered about gene lineages by resolving the relationships between alleles at a locus, and historical patterns of divergence and gene flow among populations can be inferred through population genetic approaches analyzing sets of alleles within populations (often including several loci). Two of these techniques are especially germane to observations at the infraspecific level: DNA sequencing and microsatellites.

Genetic sequencing

Genetic sequencing is a powerful tool that has a wide variety of applications. Sequences at different regions in the genome evolve at different rates, allowing genetic sequencing to be effective over a broad range of taxonomic levels. While intraspecific resolution of phylogenetic patterns is toward the fine end of the utility of sequencing, useful and sometimes striking relationships can be shown (Liston et al. 2007, Brunsfeld and Sullivan 2005, Janzen et al. 2002). Regions within plastid, mitochondrial, or nuclear genomes can be targeted through sequencing, providing data from not only independent gene lineages, but from loci inherited through different systems (biparental or uniparental).

Determination of current and historical relationships between populations within a species presents several challenges. First, there must be sufficient isolation between populations to be able to measure differences between them. Even low levels of gene flow can obscure differences between populations, and if levels are sufficiently high the very concept of separate populations with independent histories begins to break down.

For example, one reproductive migrant per generation is theoretically enough to prevent population differentiation in the absence of selection (Wright 1931).

Even after the isolation of separate populations, sufficient time may not have passed for mutations to accumulate in the alleles found within those populations. For many loci there will be no variation within a species, and use of fast-evolving loci is needed.

If allelic variation is found within a locus, that variation still may not have arisen after the isolation of the present-day populations. The ancestral population that gave rise to the current, or daughter, populations may have contained multiple alleles at that locus, and when the ancestral population split, one or more of those alleles would have filtered into the daughter populations. Therefore, populations that have recently become isolated may contain very similar sets of alleles at a locus, obscuring the signal that isolation has occurred. Over time, a single allele (or alleles from a single lineage) will become fixed within each population; however, because this fixation process is random, the relationships between alleles in different populations may not reflect the true relationships of those populations. This phenomenon, known as incomplete lineage sorting (ILS), increases in probability with larger (effective) population sizes, and shorter times since divergence (Maddison 1997). The recent divergence of the western perennial clade of lupines (0.7-2.1 mya) and the *L. lepidus* complex, makes it likely that ILS is common.

Microsatellites

Microsatellites are short (2 to 9 nucleotide) tandem repeats within a region of DNA. These short repeats occur within noncoding regions of the genome and, due to the

nature of DNA polymerase, are especially prone to extensions or deletions of the number of repeats that occur (Avise 2004). Because of these two factors, microsatellites are highly variable and very useful for population level observations. Microsatellite markers are codominant, meaning that if an individual is heterozygous at a locus, both alleles can be identified using appropriate techniques. Microsatellites are generally assumed to be in non-coding regions of the genome and so evolve in a neutral manner.

Microsatellite analyses can be used to sample multiple individuals within a population, providing a sample of the alleles present in that population. Allele frequencies are expected to change after the first generation following population separation, so techniques that measure a sample of alleles from multiple populations can provide an early indication of population divergence (Hartl and Clark 1997). And, because microsatellite analyses (usually) incorporate several loci and many samples per population, they are less likely to be hampered by the effects of ILS: The more loci and samples that are used, the lower the likelihood that recently split populations (i.e. "sister" populations) will be less similar to each other than either population to an earlier diverging population. Furthermore, the rapid evolution of microsatellite loci allows unique alleles to evolve relatively quickly in isolated populations. The use of multiple loci in microsatellite analyses increases the power of these measurements and can even allow levels of gene flow between populations to be assessed.

Approaches that use a combination of sequencing and microsatellite techniques can take advantage of the benefits of both analyses, gathering information from both the sets of alleles sampled from populations and the relationships between alleles. Recent population divergences that may be difficult to resolve using sequencing techniques

could be measurable using microsatellite analyses. Sequencing techniques, in turn, can measure the phylogenetic relationships between alleles found in those populations, and have the potential to reveal unexpected results. Sequencing has the additional advantage that identical haplotypes found in multiple populations are much more likely to be the result of shared ancestry rather than homoplasy, which is more likely to be the case with identical microsatellite alleles.

Goals

In this study I used DNA sequencing of a nuclear and chloroplast locus and microsatellite analyses to study the Cascade biogeography and the phylogeny of the *L. lepidus* species complex. My work addresses the following aims:

1) Determine the phylogenetic history of the *L. lepidus* complex and whether recognition of separate taxa is supported.

A population-level phylogeny of the complex is reconstructed using chloroplast and nuclear sequence data.

2) Elucidate the biogeographic patterns of the Cascade mountain range. I will answer the following questions through interpretation of microsatellite data in light of a reconstructed phylogeny for the group (see Aim 1).

A) Do the three Cascade varieties of *L. lepidus* demonstrate a north-south disjunction like that found by Soltis et al. (1997) for the western Cascades and Coast?

B) If so, do they favor the leading edge, north-south recolonization, or a novel hypothesis?

C) Do the results point to specific locations of glacial refuges utilized by *L. lepidus*?

D) By what process did postglacial colonization by the three varieties take place? (For example, did alpine populations jump from mountaintop to mountaintop or independently arise from lower elevation populations?)

E) Can these patterns and processes be used to predict future range expansions and contractions or extinction as climate stability deteriorates?

The results of this work attempt to resolve phylogenetic relationships between members of this group and infer patterns of diversification. This work will enhance understanding of biogeographic processes, particularly those relating to range shifts in response to the opening of areas to recolonization and the development of phenotypic diversity.

In addition to the scientific understanding gained by this project, the work will be beneficial within a conservation context. Knowledge of the biogeographic history of the *L. lepidus* complex and how it previously responded to climate shifts will not only benefit conservation efforts related to this group, it will potentially benefit large communities in the Pacific Northwest through the identification of glacial refuges. Furthermore, characterization of intervarietal and interpopulation relationships within *L. lepidus* will inform conservation efforts related to narrowly endemic varieties.

Chapter 2

Phylogenetic and population genetic analyses of DNA sequence data

In order to assess historical relationships among the varieties of *Lupinus lepidus* and to infer previous demographic changes among their populations, DNA loci from multiple regions of the genome were sequenced. Phylogenetic analyses of these sequences can estimate the branching order of the gene lineages and provide information regarding the ancestral relationships of the individuals that contain them. By sequencing these regions in multiple individuals within a population, population genetic approaches can be used to compliment the phylogenetic information when inferring phylogeographic events.

The nuclear floral symmetry gene *CYCLOIDEA* (CYC, characterized in the snapdragon *Antirrhinum majus*) has putative homologs within the legumes (LEGCYC) (Citerne et al. 2003). These genes belong to the TCP family of transcription factors and are involved in floral organ identity and symmetry (Cubas et al. 1999, Ree et al. 2004). LEGCYC has been shown to be more rapidly evolving and able to provide more resolved phylogenetic hypotheses than the commonly employed nuclear ITS region (Ree et al. 2004) and has been used in phylogenetic studies of the Fabaceae, generally, and *Lupinus*, specifically (Citerne et al. 2003, Ree et al. 2004, Hughes and Eastwood 2006). Several LEGCYC paralogs exist within the legumes: LEGCYC1A, LEGCYC1B, and LEGCYC2 (Citerne et al. 2003). Primer sets have been developed to amplify these paralogs separately (Fig. 1).

The use of chloroplast loci in plant phylogenetics has become commonplace due to their haploid state, high copy number, inheritance mode (often exclusively maternal, as

has been shown for *Lupinus perennis*), and rates of evolution useful for many applications (Corriveau and Coleman 1988). In a screen of 21 non-coding chloroplast regions, Shaw et al. (2005) found that the *trnD-trnY-trnE-trnT* intergenic spacers (*trnDT*) contained the most potentially informative characters. A later study analyzing 34 loci ranked *trnDT* tenth in potentially informative characters (Shaw et al. 2007).

Multiple members of several *Lupinus lepidus* varieties were sequenced at the LEGCYC1A and *trnDT* loci to estimate phylogenetic relationships within the species and among varieties. Additional within-population sampling was performed with LEGCYC1A and analyzed with population genetic techniques in order to assess levels of diversity across populations and provide an alternate estimation of historical relationships among them.

Materials and Methods

Collections and Sampling

Sequences were obtained from new collections, herbarium material, and previous GenBank accessions. At each newly collected population at least one individual was pressed and retained as a voucher in the Portland State University Herbarium (HPSU). Leaf material was obtained from at least nine other individuals of each population (where possible) and was immediately placed in silica desiccant.

Populations of *Lupinus lepidus* were sampled ranging from Mt. Baker in northern Washington state, south along the Cascade Mountains into the northern Sierra Nevada Mountains at Mt. Lassen. Populations were sampled west of the Cascades in the Olympic Mountains, the Coast Range, the Puget Trough, and the Willamette Valley. Populations

were sampled along the eastern slope of the Cascades in Washington and Oregon, the Blue Mountains and Steens Mountain of Oregon, northern and western Nevada, and southern Utah (Table 1, Fig. 2). This sampling strategy densely covers populations along the Cascade Range, and more sparsely, the full range of variety *lobbii* at high elevations. Identification of varieties was based on a consensus from several standard keys (Isely 1998, Hitchcock and Cronquist 1996, Broich and Morrison 1995, Hickman 1993, Barneby 1989). The species sister to *L. lepidus* is unknown (Drummond 2008). Several members of the western perennial clade of *Lupinus* were collected to test the monophyly of *L. lepidus*. Two annual lupines, *L. microcarpus* and *L. nanus*, were included in the analyses, with *L. microcarpus* serving as an outgroup in both (*L. nanus* was unavailable for the *trnDT* locus; Drummond 2008).

Character Sampling

Dried leaf material was ground in liquid nitrogen and genomic DNA was isolated following the manufacturer's protocol using the Wizard Genomic DNA Purification Kit (Promega, Madison, Wisconsin) or DNeasy 96 Plant Kit (Qiagen, Valencia, California). Primer sites flanking the LEGCYC1 gene allow amplification of both LEGCYC1A and LEGCYC1B, whereas a region in the middle of LEGCYC1 provides primer sites specific to each paralog (Citerne 2005). In order to obtain the entire sequence of a single paralog, amplification and sequencing must take place in two parts (the TCP and the R region, respectively, as per Fig. 1). In order to include as many individuals as possible, sequencing efforts emphasized the TCP region of the LEGCYC1A paralog (LEGCYC1A-TCP). LEGCYC1A primers are listed in Table 2. The *trnDT* locus contained a large mononucleotide repeat that often required the locus to be amplified in

two parts. This was accomplished through the use of universal internal primers (*trnE* and *trnY*) and newly developed primers flanking the repeat (*LupF* and *LupR*, Table 2).

Loci were amplified by PCR using a 50 μL reaction containing 0.5 U *Taq* DNA polymerase (Promega), 1x reaction buffer supplied by the *Taq* manufacturer, 1.75 mM $MgCl_2$, 1 μM forward and reverse primers (Table 2), 5% DMSO, and ~5 ng template DNA. Templates that were difficult to amplify were amplified in similar reactions using Hotmaster™ *Taq* DNA polymerase (5 Prime, Gaithersburg, Maryland) and Hotmaster™ reaction buffer, and omitting $MgCl_2$ and DMSO. Thermal cycling was performed on an iCycler® (Bio-Rad Laboratories, Hercules, California) and included an initial denaturing step of 1 min @ 95°C; 35 cycles of 1 min @ 95°, 1 min @ 57°, and 3 min @ 65°; an extension step of 8 min @ 65°, and a final hold at 4°. The amplified product was cleaned either by column filtration using the Wizard SV Gel and PCR Cleanup System (Promega) following the manufacturer's protocol, or by chemical digestion using ExoSAP-IT (USB, Cleveland, Ohio) following the manufacturer's protocol.

Direct cycle sequencing was performed using the ABI Prism® BigDye® Terminator v3.1 (Applied Biosystems, Foster City, California) with reactions containing 40 ng template DNA, 0.4-0.5 μM primer, 2 mM $MgCl_2$, 80 mM Tris-Cl (pH 9.0), 2 μL BigDye® Terminator, and water to 10 μL. Thermal cycling included 25 cycles of 10 s @ 96°, 5 s @ 50°, and 4 min @ 60°; with a final hold at 4°. Unincorporated dye terminators were removed by column filtration with Sephadex™ G-50 Fine (GE Healthcare Bio-Sciences, Piscataway, New Jersey). The ABI Prism® 3100 Genetic Analyzer (Applied Biosystems) at the Oregon Health Sciences University Sequencing Core was used for visualization and analysis of dye-labeled fragments.

Phylogenetic Analyses

Sequence fragments were assembled and edited with the SeqMan™II module of Lasergene ver. 6.1 (DNASTAR, Madison, Wisconcin). During assembly of LEGCYC1A-TCP, sites with clean double peaks approximately half the height of the surrounding single peaks were considered putative polymorphisms (i.e. individuals were heterozygous at that site) and were coded with the appropriate IUPAC ambiguity code. Sequences were aligned by hand in MacClade ver. 4.08 (Maddison and Maddison 2005). All sequences of both LEGCYC1A-TCP and *trnDT* could be unambiguously aligned, with any gaps easily interpreted as either insertion or deletion (indel) events.

Phylogenetic trees were inferred using the maximum likelihood (ML) criterion (*trnDT* and LEGCYC1A-TCP) and Bayesian inference (LEGCYC1A-TCP). The optimal model of nucleotide substitution was chosen using hierarchical likelihood ratio tests (hLRT), Akaike information criterion (AIC), and Bayesian information criterion (BIC) tests as implemented in ModelTest ver. 3.7 (Posada and Crandall 1998) and jModelTest ver. 0.1.1 (Posada 2008, Guindon and Gascuel 2003).

The LEGCYC1A-TCP and *trnDT* ML trees were estimated using the program RAxML ver. 7.0.4 under the GTR+Γ (GTRGAMMA) model of nucleotide substitution (the RAxML author, Alexandros Stamatakis, feels that using GTR+Γ instead of GTR+I+Γ can avoid some problems of independently optimizing α and the proportion of invariable sites, while still effectively incorporating rate heterogeneity; 2006a, 2006b, Stamatakis et al. 2008). Support values were obtained via 10^4 bootstrap replicates utilizing the rapid bootstrap algorithm of RAxML (with the default of 25 GTRCAT rate

categories). RAxML for the LEGCYC1A-TCP locus was implemented via the online CIPRES Portal ver. 2.2 (Miller et al.).

Bayesian analysis was performed in MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) as an alternate approach to estimating the LEGCYC1A-TCP phylogeny. The Metropolis-coupled Markov chain Monte Carlo simulations were run with four linked chains (one cold and three heated, default temperature of 0.2) and default priors for all parameters. Two runs of more than 15×10^6 generations were compared to assess whether convergence to a stationary distribution of parameter values had been reached by measuring the average standard deviations of split frequencies between the two runs. A cut-off of 0.01 standard deviations was used as a guideline to assess convergence of runs. After a burn in period of 3×10^6 generations, parameter values were sampled every 100 generations to calculate posterior probabilities of parameters.

In order to gain a broader understanding of the frequency of LEGCYC1A-TCP genotypes and the potential relationships between them, a haplotype (genotype) network was constructed among the members of the western perennial clade, using the program TCS ver 1.21 (Clement et al. 2000). For this analysis, three non-overlapping indels were deleted and recoded as additional characters concatenated to the nucleotide sequence. The order in which sequences are added to the analysis can affect the haplotype calculations when sequences differ only by ambiguities, such that the sequences AT, AG, and AR can be called one haplotype (if AR is entered first) or two haplotypes (with AR being assigned to the first sequence entered). For this analysis, sequences were ordered such that sequences with zero polymorphic (“ambiguous”) sites, and no missing data, were

placed first, followed by sequences with one, then two polymorphic sites. Sequences with missing data were placed last, in order of increasing number of missing sites. The parsimony connection limit was estimated (95% limit).

Population genetic statistics

The LEGCYC1A-TCP locus was sequenced for multiple individuals of seven populations. For each population, standard diversity indices were calculated in Arlequin ver. 3.5.1.2 (Excoffier and Lischer 2010) including nucleotide diversity (π_n), θ based on the number of segregating sites (θS), and θ based on the average number of pairwise differences ($\theta\pi$). All individuals sampled for LEGCYC1A-TCP were grouped by variety and included in an analysis of hierarchical genetic structure (AMOVA, 10^3 permutations). F_{ST} values were calculated for each pair of varieties and of the seven populations (significance tested with 100 permutations). Finally, a test of population differentiation was carried out between all pairs of varieties and all pairs of populations utilizing an extension of the Fisher test for R x C contingency tables that incorporates a Markov chain to accommodate large data sets (the "exact test," using a 10^5 step Markov chain with a burnin of 10^4 steps; Excoffier and Lischer 2010, Raymond and Rousset 1995).

For all analyses requiring a distance measure between haplotypes, the Tamura and Nei (1993) measure with gamma correction factor for among site rate variation was utilized (TN+ Γ, see above for model selection). The alpha shape parameter used for the gamma correction was 0.444, calculated as the most likely from a preliminary ML analysis utilizing a ratchet algorithm (Nixon 1999, Morrison 2007). (This alpha parameter is very close to that calculated under a GTR+I+ Γ model using the reported

ML tree, α=0.496. An analysis was run using the alpha parameter estimated by Bayesian analysis that returned similar results, with no p-values changing significance, that is not reported).

Because direct sequences were obtained, nuclear haplotypes were not available and sequence data were coded with an unknown gametic phase. To do this, two mock haplotypes were created for each individual from that individual's LEGCYC1A-TCP sequence. Ambiguous sites (indicative of polymorphism) were coded as one or the other of the possible nucleotides for each mock haplotype, respectively. For example, a site in an individual with an R ambiguity code would be coded as an A in one haplotype, and a G in the other. This was performed for all polymorphic sites (Fig. 3). The placement of a nucleotide into one mock haplotype or the other is arbitrary, so long as each possible nucleotide is represented. Arlequin does not assume that mock haplotypes represent the true haplotypes when the gametic phase is unknown.

Results

Sequence Characteristics

The TCP region of LEGCYC1A was successfully sequenced in 62 *Lupinus lepidus* individuals, and 7 outgroup individuals; six sequences were obtained from Genbank (3 *L. lepidus* and 3 outgroups). Total aligned length was 577 base pairs (unaligned, 549-568), 83 of which were variable. The *trnDT* locus was successfully sequenced in 19 *Lupinus lepidus* individuals and six outgroup individuals. Total aligned length was 1235 base pairs (unaligned, 1021-1218), 31 of which were variable.

For LEGCYC1A-TCP, the path through the series of hLRTs, calculated by Modeltest, chose the HKY+I+Γ model (unlike the popular program MrModeltest, Modeltest performs only one series of hLRTs, but tests more models). Analysis using AIC as tested by Modeltest and jModelTest resulted in the TVM+I+Γ model followed by GTR+I+Γ (delta=1.86 in ModelTest, 1.52 in jModeltest) and HKY+I+Γ (delta=7.19 in ModelTest, 14.96 in jModelTest). Analysis using BIC in jModelTest resulted in TPM2uf+I+ Γ (delta of HKY+I+Γ=3.82, delta of GTR+I+Γ=7.75). The series of hLRTs did not compare the HKY to the GTR model. An LRT performed comparing the HKY to the GTR model rejected the HKY model (P=0.0097) and a comparison of HKY+I+Γ to GTR+I+Γ rejected the HKY+I+Γ model (P=0.0003). Of the models of evolution available in MrBayes, the GTR+I+Γ model was selected because it was selected by the AIC and direct LRTs between HKY and GTR models selected the GTR models. For *trnDT*, the series of hLRTs and analysis using AIC and BIC resulted in selection of the F81+I model. Of the models of evolution utilized by Arlequin, the Tamura and Nei model with gamma correction for variation among sites (TN+ Γ) had the lowest values in both the AIC and BIC tests. Table 3 presents the model parameters for both of these regions; the parameters for the LEGCYC1A-TCP and *trnDT* ML trees were re-estimated under the GTR+I+Γ and F81+I models, respectively, using the ML trees found by RAxML (PAUP* ver. 4.0b10; Swofford 2001).

Phylogeny of LEGCYC1A-TCP

The ML tree found by via RAxML had a log likelihood of -1895.130720 (-1857.68045 when recalculated in PAUP* under the GTR+I+Γ model). ML and Bayesian analyses produced tree topologies that were not incongruent, but provided

differing levels of resolution (Figs. 4-5). In neither analysis was the monophyly of *L. lepidus* supported. The analyses agree in placing *L. arboreus* within a clade containing the majority of *L. lepidus* individuals and in supporting a clade consisting of *L. caudatus* and one *L. lepidus* var. *sellulus* individual from California. The ML tree placed two *L. l. sellulus* individuals within a clade of *L. polyphyllus, L. albifrons,* and *L. nanus* (Fig. 4), but this relationship was not resolved in the Bayesian analysis (Fig. 5). Neither analysis found that any of the *L. lepidus* varieties formed monophyletic groups, although the Bayesian analysis was uninformative regarding the monophyly of varieties *ashlandensis* or *utahensis* and *lepidus* (Figs. 4-5). Of the populations from which multiple individuals were sampled, none were supported as monophyletic groups in the ML analysis (Fig. 4). The Bayesian analysis was uninformative on the monophyly of populations from Mt. Hood, Mt. Adams, and Mt. Ashland, but rejected the monophyly of populations from Mt. St. Helens, the Olympic Mountains, Mt. Rainier, and Grizzly Peak (posterior probabilities [PP] rejecting monophyly of 0.83, 0.80, 0.54, and 0.54, respectively).

Both analyses resolved a clade consisting of the majority of *L. lepidus* samples (at least 54 individuals) plus both *L. arboreus* samples (clade A). Two individuals (one of var. *lobbii* from Santiam Pass and one of var. *ashlandensis*) were placed within this clade in the Bayesian analysis (PP=0.66) but were placed in a polytomy at the subtending node in the ML analysis. Except for these two individuals, the remaining five sampled *L. lepidus* individuals that diverge from clade A at an earlier point in the tree come from locations south of Oregon.

Within clade A the ML tree contains two major clades (clades B and C), although these were not resolved in the Bayesian analysis. Clades B and C both contain members

of varieties *aridus, lobbii,* and *sellulus*, while variety *lepidus* is exclusive to clade B and variety *utahensis* is exclusive to clade C. Both clades contain individuals from the northern populations of *L. lepidus*, both contain individuals from southern populations (southern Oregon or farther south), and both contain individuals from the eastern reaches of *L. lepidus*. Representatives from the Mt. Rainier, Mt. Adams, and Grizzly Peak populations are found within both clades, while Olympic Mountain individuals are exclusive to clade B, and Mt. Hood and Mt. St. Helens individuals are exclusive to clade C.

The most heavily sampled southern population, Grizzly Peak, contains individuals with genotypes falling at three locations across clades B and C. At each of these locations there are identical genotypes found from the Mt. Rainier population and closely related genotypes from the other northern populations (Olympic Mountains, Mt. Adams, Mt. St. Helens, and Mt. Hood; Fig. 4).

While these relationships reflect the most likely tree analyzed, it should be noted that the ML tree overall has low bootstrap (BS) support (Fig. 4). Results supported with greater than 50% bootstrap support include the paraphyly of *L. lepidus* (BS=68); the paraphyly of varieties *lobbii,* and *sellulus* (BS=61 and 68, respectively); and the paraphyly of the Western Perennial lupines with *L. nanus* (BS=70). Low bootstrap support could be caused by several factors, including low sequence variation, recombination within the LEGCYC1A-TCP locus, and multiple possible resolutions of polymorphic sites coded as ambiguities. The fact that the haplotype network constructed from these data (see below, Fig. 6) contains individuals from different clades in the ML

tree sharing a haplotype, implies that alternative resolutions are a strong factor in this phenomenon.

LEGCYC1A-TCP haplotype network

Of 73 sequences, 48 non-compatible haplotypes were observed (i.e. haplotypes that differed by more than ambiguities, Fig. 6). The parsimony connection limit with greater than 95% probability was estimated to be 10 steps. Only three sequences were exactly identical to each other (all from the Mt. Hood Timberline population), meaning that other sequences with identical haplotypes actually differed at polymorphic sites. One haplotype was represented by eight individuals, two were represented by five individuals, two were represented by four individuals, and four were represented by two individuals. Reticulation is widespread in the network, with no clear ancestral relationships.

As in the ML tree, *L. lepidus* is "paraphyletic" in this network, with haplotype mutation pathways between some outgroups necessarily passing through haplotypes found in *L. lepidus* (Fig. 6). However, all outgroups are located on the periphery of the network, so that all *L. lepidus* haplotypes are connected by mutation pathways that do not contain outgroup taxa. No patterns are apparent with respect to the placement of *L. lepidus* varieties. No populations with multiple samples contain only a single haplotype, but often haplotypes very far from each other in the network (Grizzly Peak, Mt. Rainier, Mt. Hood, Mt. St. Helens, and Mt. Adams). As in the ML tree, individuals from northern populations often have identical haplotypes with southern individuals or are only a few steps from a southern individual: The Grizzly Peak population shares multiple haplotypes with northern populations, and some far southern individuals are placed closer to

northern individuals (e.g. Mt. Rainier and the Olympic Mtns.) than to any other individual.

Some individuals sharing the same haplotype in the TCS network are not shown to be monophyletic in the ML phylogeny. This is likely an effect of alternate resolutions of ambiguities. Alternate resolutions may also be responsible for the high level of reticulation seen in the network, as well as the lack of bootstrap and posterior probability support in the LEGCYC1A-TCP phylogenetic trees.

Phylogeny of trnDT

The *trnDT* ML tree found by via RAxML had a log likelihood of -1891.172148 (-1900.06634 when recalculated in PAUP* under the F81+I model; Fig. 7). As in the LEGCYC1A-TCP phylogeny, the *trnDT* data do not support the monophyly of *Lupinus lepidus*. None of the *L. lepidus* varieties are monophyletic, however these data are uninformative for varieties *ashlandensis* and *confertus* (although one *confertus* individual shares a haplotype that is also shared with varieties *aridus*, *lepidus*, and *sellulus*). None of the clades resolved in the *trnDT* phylogeny correspond to those in the LEGCYC1A-TCP phylogeny. Low bootstrap support in this tree is likely due to low sequence variation since recombination and a large number of ambiguous sites are not a factor.

Population statistics

Nucleotide diversity (π_n, the probability that two randomly chosen homologous nucleotides within a group are different) across all *L. lepidus* samples was 0.0174, the mean number of pairwise differences, π, was 8.486. Diversity statistics for each *L. lepidus* variety and for each population for which more than one individual was sampled are provided in tables 4 and 5, respectively. There is some variation in diversity measures

between varieties, whereas populations tend to be more clustered, except for the population from Mt. Hood which is less diverse. Pairwise F_{ST} values between varieties and populations are shown in tables 6 and 7, respectively. Only 6 of 21 pairwise F_{ST} values between varieties were significantly different from zero, whereas 14 of 21 comparisons between sampled populations were significant. Of those varietal comparisons that were significant, 5 were comparisons between variety *lobbii* and another variety, while one was between varieties *sellulus* and *utahensis*. Differences in population sample sizes may be causing instances where some non-significant F_{ST} values are greater than some significant values, and may indicate a lack of sufficient power for some comparisons.

Analysis of molecular variance (AMOVA) suggests that genetic variation within *L. lepidus* is apportioned non-randomly, with 55% of the variance among individuals within a population, 41.5% among populations within a variety, and only 3.6% among varieties (Table 8). Somewhat at odds with these values, exact tests of differentiation failed to reject a hypothesis of panmixia either among populations (P=0.72) or among groups (P=0.20). Likewise, non-differentiation could not be rejected for any pair of populations or groups (data not shown).

Discussion

Analyses of nuclear and chloroplast sequence data from *Lupinus lepidus* do not support monophyly of any of the morphologically defined varieties or the species as a whole. Surprisingly, *L. lepidus* is resolved as paraphyletic with every outgroup species included (and uninformative with respect to the root species *L. microcarpus*). This

includes not only other members of the western perennial clade, but the annual species *L. nanus*. These results are probably a signature of incomplete lineage sorting since *L. lepidus* is unlikely to have hybridized with *L. nanus*, or even with the western perennial *L. arboreus* which grows on coastal dunes. While no individuals were observed during this study that appeared to be hybrids with other western perennial lupines, some do grow sympatrically with *L. lepidus*, making the possibility of hybridization more plausible. It should be noted that ILS is less likely to explain the paraphyly of *L. lepidus* in the *trnDT* analysis, since this locus has half the effective population size of LEGCYC1A. Ancient hybridization may in fact explain the clade formed by *L. caudatus* and the individual of var. *aridus* from Idaho, as these share an identical 169 bp deletion in the *trnDT* locus.

Analyses of both LEGCYC1A-TCP and *trnDT* do not show any obvious patterns with regard to variety, population membership, or geographic origin, with the exception of a tendency in the LEGCYC1A-TCP phylogeny for earlier diverging *L. lepidus* populations to come from southern localities (Figs. 4-7). Most of the variation in LEGCYC1A-TCP sequences is found either between individuals in a population or between populations, with very little found between varieties (Table 8). More pairwise comparisons between populations had F_{ST} values significantly different from zero than comparisons between varieties.

These results are consistent with a hypothesis that the varieties of *L. lepidus* are either still exchanging genes or are undergoing incipient divergence, not that they have been isolated for a long period of time. The proportion of total variation that can be attributed to differences between varieties is small (3.6%) but is significantly different from zero (Table 8, P=0.032). Lack of monophyly of varieties for either the

LEGCYC1A-TCP or *trnDT* locus is not unexpected under this hypothesis because it is unlikely in the early stages of divergence that sufficient time has passed for the fixation of ancestral alleles in daughter taxa.

In contrast to the significant F_{ST} values, exact tests of population differentiation failed to reject hypotheses of panmixia between populations and varieties. This may be due to the use of genetic data of unknown gametic phase. Under this circumstance the exact tests of differentiation consider the frequencies of genotypes, not haplotypes, in group comparisons (Excoffier and Lischer 2010, Raymond and Rousset 1995). In effect, the measurement being tested is not the frequency of a particular allele (or set of alleles) in a population, but the frequency with which two particular alleles that are paired in an individual in one population are found, also paired in an individual, in another population. In populations meeting Hardy-Weinberg assumptions, and with very large sample sizes, these measures will converge. However, differences are expected to arise in finite populations, especially because a small sample size decreases the probability of sampling identical genotypes even if the frequency of each haplotype is properly represented.

Phylogeographic patterns

Under hypotheses of recolonization of the Cascade Mountains following recent periods of glaciation, northern populations should exhibit reduced genetic diversity compared to southern populations that were unaffected by glaciation (Soltis et al. 1997). In hypotheses involving northern refuges, this effect is caused by a population bottleneck in the northern refuge, with recolonization occurring from that genetically depauperate population. In hypotheses involving a recolonization from southern populations, migrants

from the south are assumed to be rare, but with the ability to establish new populations, effectively sampling only a small portion of the genetic diversity present in the south. In contrast to these expectations, northern populations of *L. lepidus* do not seem to have reduced genetic diversity compared to southern populations. Indeed, for all three measures of diversity, northern populations may have values greater or less than those of southern populations (Table 5).

Furthermore, there is substantial among population variation in the northern region. Of ten pairwise comparisons, 8 have F_{ST} values that differ significantly from zero (Table 7). This high divergence among populations is unlikely if they were all colonized from a single northern population, especially one hypothesized to have reduced genetic diversity.

For genetic diversity to remain comparable in northern populations, either the populations must have been present for a sufficiently long time to generate new diversity or receive additional southern migrants, or the recolonization of this region occurred in such a way that high diversity was maintained on the leading edge of a colonization front. Considering the glacial history of the region, it is unlikely that northern populations have been extant for a long period of time. This implies that genetic variation present in northern populations has been gained, with very little reduction, from southern populations, possibly from colonization events not affected by founder effects or from ongoing, high migration rates.

This conclusion is corroborated by the ML phylogeny and haplotype network of LEGCYC1A-TCP in three ways (Figs. 4, 6). First, identical genotypes are found in both southern and northern populations. Four genotypes were sampled in the Grizzly Peak and

Mt. Rainier populations. Three of these genotypes are identical (more precisely, ambiguities caused by polymorphic sites can be resolved to make them identical), while the fourth genotype from each population differs from a member of the other population by only a few substitutions and is part of a polytomy that can be resolved to make it sister to that genotype. One of these genotypes was also sampled from Mt. Hood.

Second, genotypes from southern localities are well mixed within clades containing northern samples. Within clade B there are samples from Mt. McLoughlin in Oregon, Medicine Lake in California, Nevada, and southern Utah. Within clade C there are representatives from Mt. Ashland and Grizzly Peak in Oregon, Nevada, and southern Utah.

Finally, *L. lepidus* samples that fall outside of clade A and diverge earlier in the tree are from southern localities. In some cases these are resolved as sisters to outgroup representatives. This pattern may indicate that northern populations did suffer a reduction in diversity when compared to those in the far southern range of *L. lepidus*. Southern regions that maintain a high level of diversity would be more likely to retain alleles originating before the divergence of *L. lepidus* and possibly before diversification of the western perennial clade of *Lupinus*.

Conclusion

The monophyly of *Lupinus lepidus* is not supported from analyses of the LEGCYC1A-TCP and *trnDT* loci, nor is the monophyly of any variety of *L. lepidus*. Lack of monophyly is likely due to the very recent divergence of *L. lepidus* and its varieties, preventing the complete sorting of ancestral alleles. Apportionment of some

genetic variation among the varieties of *L. lepidus* indicates that these may be at an early stage of divergence. Unlike other plants with a distribution along the Cascades, DNA sequencing does not show *L. lepidus* to have a strong genetic disjunction between northern and southern populations. Strong differentiation among northern populations and the sharing of genotypes with southern populations is inconsistent with a north-south disjuction, indicating the distribution of these loci was relatively unaffected by processes of glacial recolonization.

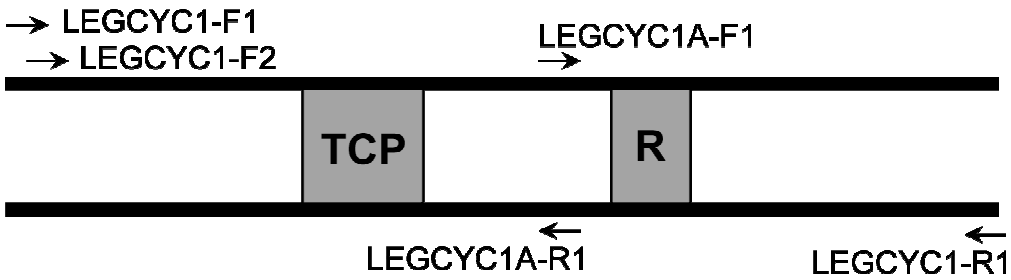


Figure 1: Schematic of LEGCYC1A adapted from Citerne (2005). Location of internal and external primers is indicated. Dark areas are regions highly conserved across taxa.

Table 1: Collection information. Sequences obtained from GenBank are noted under Description. Elevation given in meters.

Species	Variety	State	County	Description	Latitude	Longitude	Elev.	Collector	Herb.
albifrons		OR	Josephine	Rough and Ready				Fishbein6100	HPSU
arboreus		WA	Jefferson	Port Townsend	48.13530	-122.76200		KW08-02	HPSU
arboreus				GenBank DQ529972					
argenteus				GenBank DQ529970					
caudatus		OR	Harney	Steens Mtn.	42.69828	-118.59975		KW08-46	HPSU
lepidus	*aridus*	ID	Nez Perce		46.0841	-116.8395	1615	Mancusco1142	SRP
lepidus	*aridus*	WA	Klickitat	North Dalles	45.62114	-121.14580	61	KW08-18	HPSU
lepidus	*aridus*	OR	Wasco	Dufur	45.45793	-121.12349	427	KW08-07	HPSU
lepidus	*aridus*	OR	Wasco	Wapanita Hwy.	45.17202	-121.16944	552	KW08-05	HPSU
lepidus	*aridus*	NV	Humboldt	Sheldon Natl. Wildlife Refuge	41.87000	-119.03000	1554	Tiehm12144	UTC
lepidus	*aridus*	UT	Washington		37.4968	-113.8358	1782	L.Higgins19732	RM2
lepidus	*aridus*	OR	Morrow	GenBank DQ529939				Davis	USDA
lepidus	*ashlandensis*	OR	Jackson	Mt. Ashland	42.08130	-122.61929	2275	KW08-42	HPSU
lepidus	*confertus*	CA	Lassen		40.3167	-120.5417	1293	Halse6670	UTC
lepidus	*confertus*	NV	Mineral		38.5376	-118.8219	2712	Tiehm13006	UTC
lepidus	*lepidus*	WA	Thurston	Scatter Creek Rest Area	46.83830	-122.98538	65	KW08-22	HPSU
lepidus	*lepidus*	WA	Clark	The Quarry, Vancouver	45.62110	-122.49070	89	KW08-16	HPSU
lepidus	*lepidus*	OR	Linn	Jordan	44.72739	-122.69853	188	KW08-24	HPSU
lepidus	*lobbii*	WA	Whatcom	Mt. Baker	48.82000	-121.84000		Bishop	NA
lepidus	*lobbii*	WA	Clallam	Blue Mtn., Olympic Natl. Park	47.95448	-123.25946	1821	KW08-70	HPSU
lepidus	*lobbii*	WA	Pierce	Mt. Rainier	46.91470	-121.64002	1940	KW08-69	HPSU
lepidus	*lobbii*	WA	Skamania	Mt. St. Helens	46.27639	-122.22990	1198	KW08-76	HPSU
lepidus	*lobbii*	WA	Skamania	Mt. Adams	46.25103	-121.53378	1853	KW08-71	HPSU
lepidus	*lobbii*	OR	Clackamas	Mt. Hood, Zig Zag	45.3375	-121.7287	1790	Fishbein5661	HPSU
lepidus	*lobbii*	OR	Clackamas	Mt. Hood, Timberline Lodge	45.33326	-121.71203	1855	KW08-26	HPSU
lepidus	*lobbii*	OR	Wallowa	Mt. Howard	45.26031	-117.17894	2511	KW08-52	HPSU
lepidus	*lobbii*	OR	Wasco	Mt. Hood, Hwy. 26	45.16858	-121.66572	1085	KW08-27	HPSU

Species	Variety	State	County	Description	Latitude	Longitude	Elev.	Collector	Herb.
lepidus	*lobbii*	OR	Benton	Marys Peak	44.50416	-123.55287	1252	KW08-53	HPSU
lepidus	*lobbii*	OR	Linn	Santiam Pass (Hoodoo)	44.39062	-121.86083	1429	KW08-36	HPSU
lepidus	*lobbii*	OR	Grant	Baldy Mtn.	44.34733	-118.81207	2089	KW08-51	HPSU
lepidus	*lobbii*	OR	Deschutes	Mt. Bachelor	43.99951	-121.66365	1941	KW08-37	HPSU
lepidus	*lobbii*	OR	Douglas	Crater Lake entrance	43.08945	-122.11182	1777	KW08-39	HPSU
lepidus	*lobbii*	OR	Harney	Steens Mtn.	42.69828	-118.59975	2549	KW08-43	HPSU
lepidus	*lobbii*	NV	Washoe	Rose Mtn.	41.3197	-119.8941	2560	Pinzl12495	SRP
lepidus	*sellulus*	OR	Lane	Rattlesnake Butte	44.26279	-123.32456	320	KW08-23	HPSU
lepidus	*sellulus*	OR	Klamath	South of La Pine	43.61663	-121.54846	1301	KW08-38	HPSU
lepidus	*sellulus*	OR	Jackson	Mt. McLoughlin, base	42.49088	-122.38856	1053	KW08-40	HPSU
lepidus	*sellulus*	OR	Jackson	Grizzly Peak	42.26670	-122.61929	1795	KW08-41	HPSU
lepidus	*sellulus*	CA	Siskiyou	Medicine Lake	41.59167	-121.5975	2050	Fishbein5875	HPSU
lepidus	*sellulus*	CA	Shasta	Lost Creek Pass, Mt. Lassen	40.51000	-121.47000		Bishop	NA
lepidus	*sellulus*	CA	Placer	Donner Lake	39.33330	-120.29167	2134	Helmkamp5026	UTC
lepidus	*sellulus*	CA	Placer	Squaw Valley Ski Area, GenBank DQ529940	39.19872	-120.26692	2364	Bailey254	FHO
lepidus	*sellulus*	NV	Douglas		38.8869	-119.4978	2590	Tiehm12246	UTC
lepidus	*utahensis*	NV	Elko		41.6163	-114.7294	1903	Tiehm14310	UTC
lepidus	*utahensis*	UT	Garfield		38.1183	-111.5103	3383	Madsen, Merkley, Lovell1701	SRP
lepidus	*utahensis*	NV	Washoe	Rose Mtn., GenBank AY338907				Lavin	MONT
microcarpus nanus		CA		GenBank AY382156				Bishop	NA
polyphyllus	*polyphyllus*	WA	Skamania	Mt. Adams	46.25103	-121.53378		KW08-75	HPSU
polyphyllus	*saxosus* (2)	WA	Kittitas	Thorp	47.08807	-120.78480		KW08-21	HPSU
polyphyllus	*saxosus* (1)	OR	Wasco	Foreman Point	45.09410	-121.41908		Fishbein6085	HPSU

Table 1 continued.

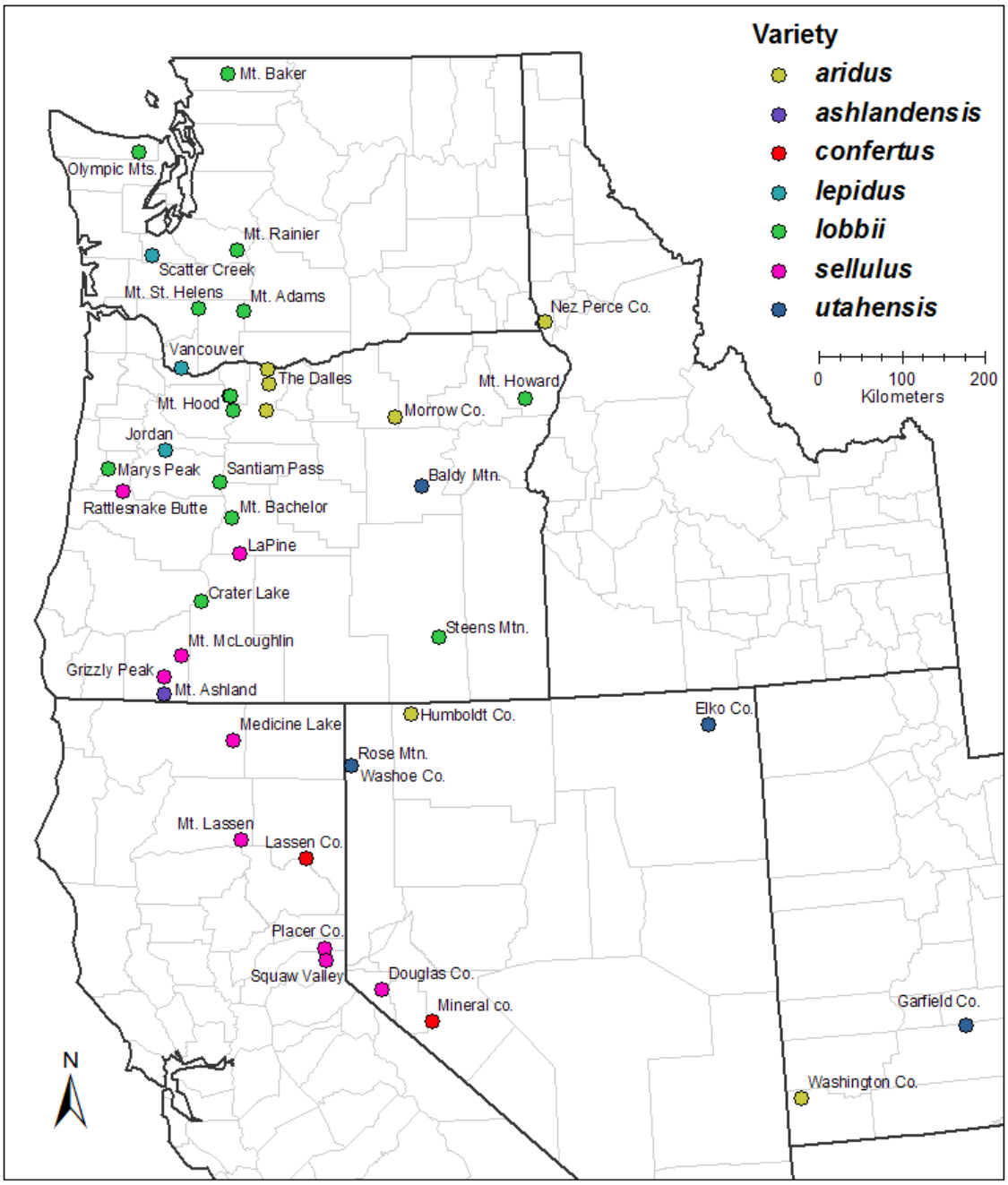


Figure 2: *Lupinus lepidus* sampling locations.

Table 2: Primers used for amplification, sequencing, and microsatellite analysis. Microsatellite forward primers (including LEGCYC1-F2, when used for microsatellite analysis) had the M13F sequence appended to their 5' ends.

Primer	Sequence 5'-3'	Source
LEGCYC1-F1	CTT CTA CTT ACA YWT CYT CAG GC	Citerne 2005
LEGCYC1-F2	(M13F) + CTT TCY TTA ACC CTG AAA ATG CTT C	Citerne 2005
LEGCYC1A-R1	CTA CYA CTA CCC CTT CTG G	Citerne 2005
LEGCYC1A-F1	CCA GAA GGG GTA GTR GTA G	Citerne 2005
LEGCYC1-R1	CAC TCY TCC CAR GAY TTT CC	Citerne 2005
trnD-F	ACC AAT TGA ACT ACA ATC CC	Shaw et al. 2005
trnY	CCG AGC TGG ATT TGA ACC A	Shaw et al. 2005
trnE	AGG ACA TCT CTC TTT CAA GGA G	Shaw et al. 2005
trnDT-Lup904R	CCT TCT TAC CCC GAT TCC CAG	This study
trnDT-Lup983F	GGT TGA GGG GCA GGA CAA ATG G	This study
trnT	CTA CCA CTG AGT TAA AAG GG	Shaw et al. 2005
AG81F	ATT TTC CAA CTC GAA TTG ACC	Peakall et al. 1998
AG81R	TCA TCA ATC TCG ACA AAG AAT G	Peakall et al. 1998
AG55-20-22F	M13F + TAG CAG AAC ACC CTT GCC TC	Drummond and Hamilton 2005
AG55-20-22R	GCA ATT CCT GAA CTA AAA CCA G	Drummond and Hamilton 2005
AG55-26-16F	M13F + CAA TCA CCC ATC AAG CCC TG	Drummond and Hamilton 2005
AG55-26-16R	AGC ACA GGT TCA TCA TTC ACC	Drummond and Hamilton 2005
CAC60-13-6F	M13F + TCA ACC TCA GTC TCA GAA GG	Drummond and Hamilton 2005
CAC60-13-6R	TTC ATC ATC AGA CAC CAT CC	Drummond and Hamilton 2005
Luna10F	M13F + CGT TGA TGC AAT GTA GGT ATC	Morris 2009
Luna10R	CGA AAC ATT GCG ATT CAT CT	Morris 2009
Luna14F	M13F + AAA CAA GCA TTA GAA GCA GT	Morris 2009
Luna14R	CAT CCT CTT ATG GCG GTG AT	Morris 2009
Luna16F	M13F + CGG ATC TTT CAA AGG GAA AT	Morris 2009
Luna16R	GAT CAC GGA TGT TGG GAG TC	Morris 2009
M13F	AGG GTT TTC CCA GTC ACG ACG TT	

Figure 3: Example resolution of ambiguous sequence data into two mock haplotypes. Note that the placement of a nucleotide into the first or second mock haplotype is arbitrary, as long as both nucleotides are represented.

Original Sequence	Mock Haplotypes
ATR**CG**Y	**1: AT**A**CG**C **2: AT**G**CG**T

Table 3: Parameter estimates of substitution models. Values are either maximum likelihood estimates recalculated in PAUP* from the RAxML tree, or mean values obtained from all Bayesian MCMC samples (after burnin). Substitution rates of the Bayesian values have been scaled so that $r_{GT} = 1$ from an original value of 0.0632. Note that in the F81+I model substitution rates and the α shape parameter are not estimated but are fixed at 1 and infinity, respectively.

Locus	Estimation method	Model	r_{AC}	r_{AG}	r_{AT}	r_{CG}	r_{CT}	r_{GT}	π_A	π_C	π_G	π_T	p_{inv}	α
LEGCYC1A-TCP	ML	GTR+I+Γ	1.2017	5.3643	2.1366	0.2515	4.0788	1	0.2940	0.2172	0.1628	0.3260	0.7288	0.4960
LEGCYC1A-TCP	Bayesian analysis	GTR+I+Γ	1.2033	6.3651	2.3162	0.3162	4.5841	1	0.2904	0.2156	0.1681	0.3259	0.7303	0.1822
trnDT	ML	F81+I	1	1	1	1	1	1	0.3285	0.1702	0.1716	0.3297	0.8825	∞

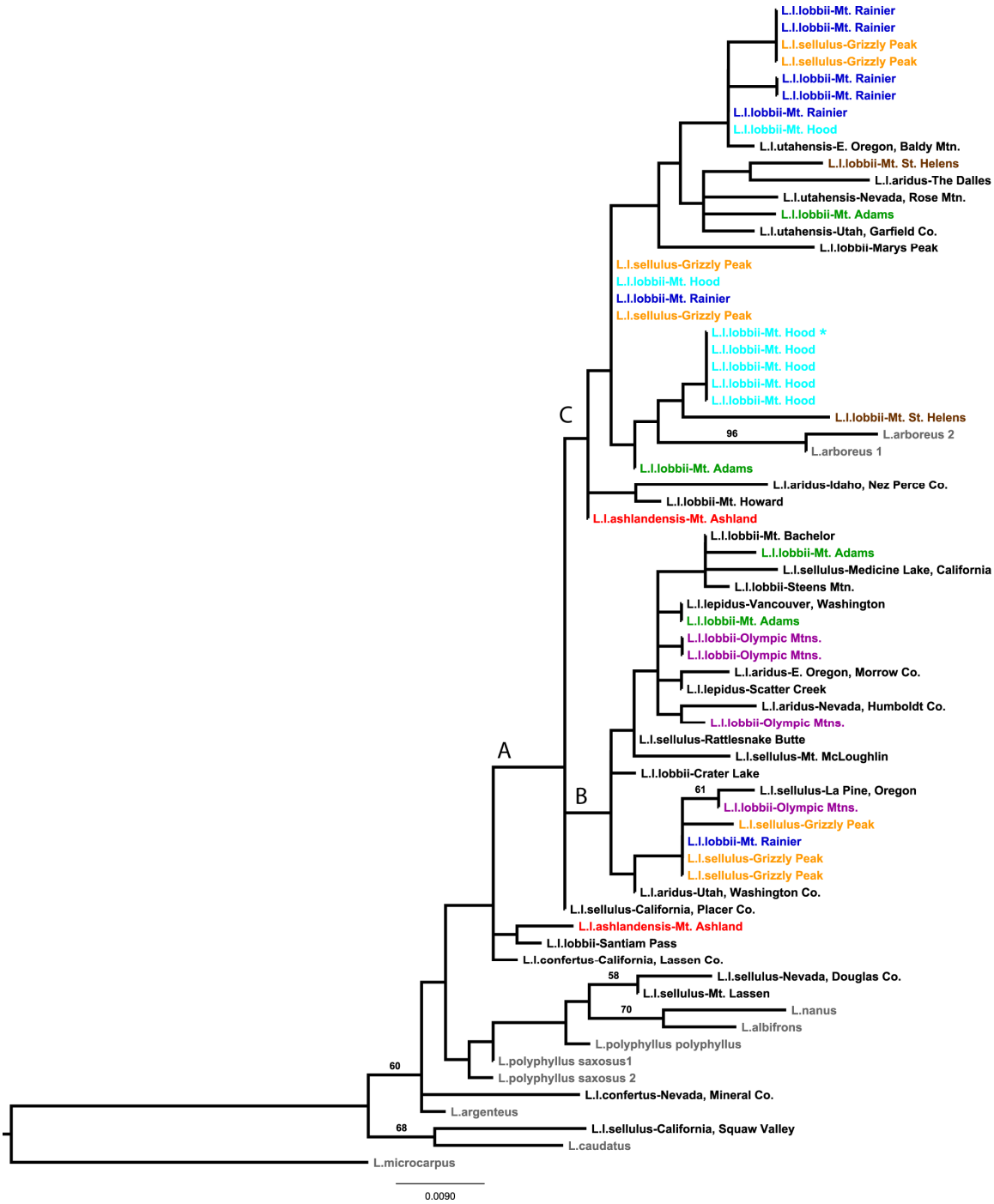


Figure 4: LEGCYC1A-TCP ML tree. Populations sampled for >1 individual are distinguished by color: Mt. Rainier dark blue, Mt. Hood light blue, Mt. Adams green, Mt. St. Helens brown, Olympic Mts. purple, Mt. Ashland red, Grizzly Peak orange. *L. lepidus* populations sampled for 1 individual are black and outgroup taxa are gray. Bootstrap values >50 are given for branches of non-trivial length. Scale is in substitutions per site. *Three individuals from the Mt. Hood Timberline population shared this sequence.

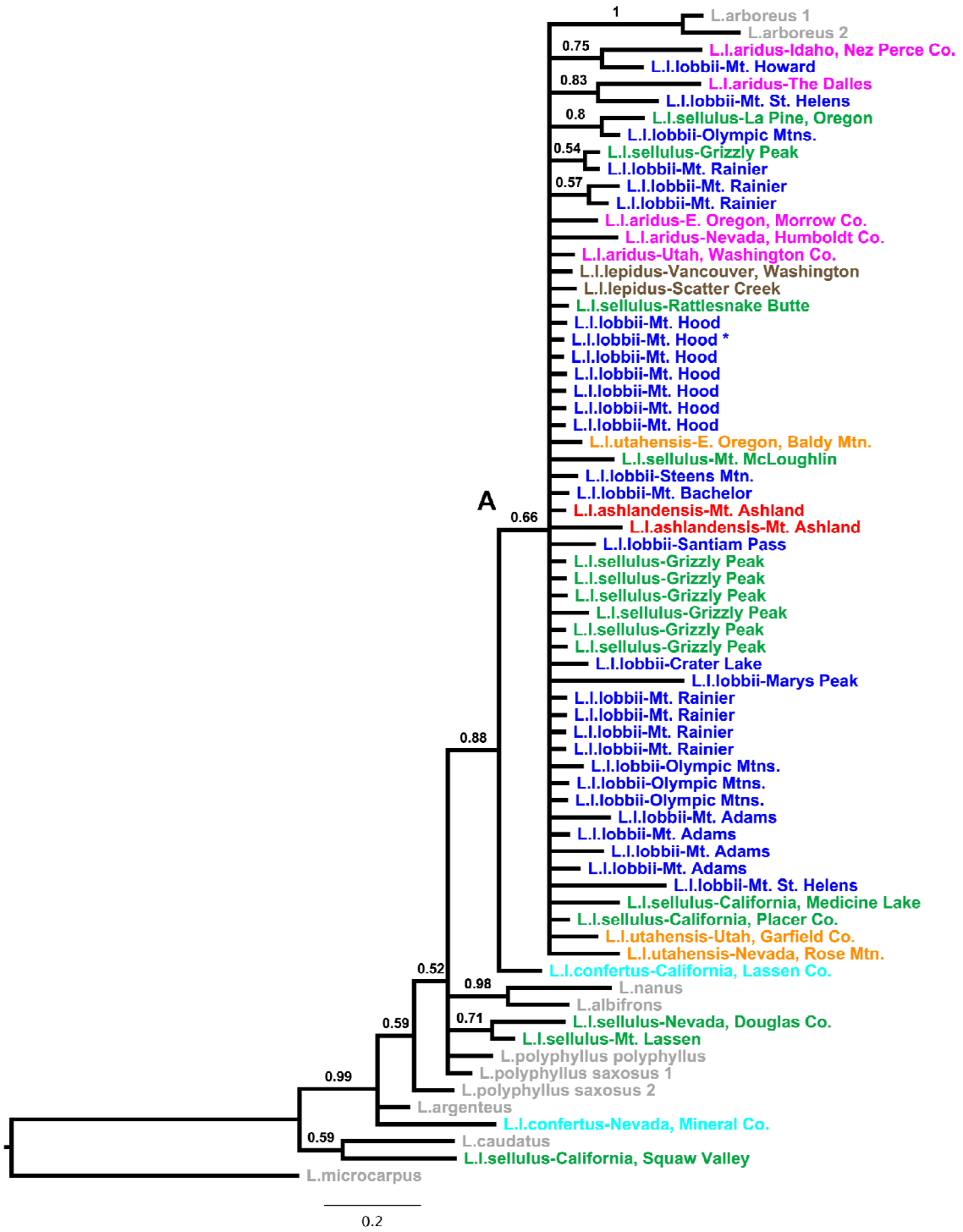
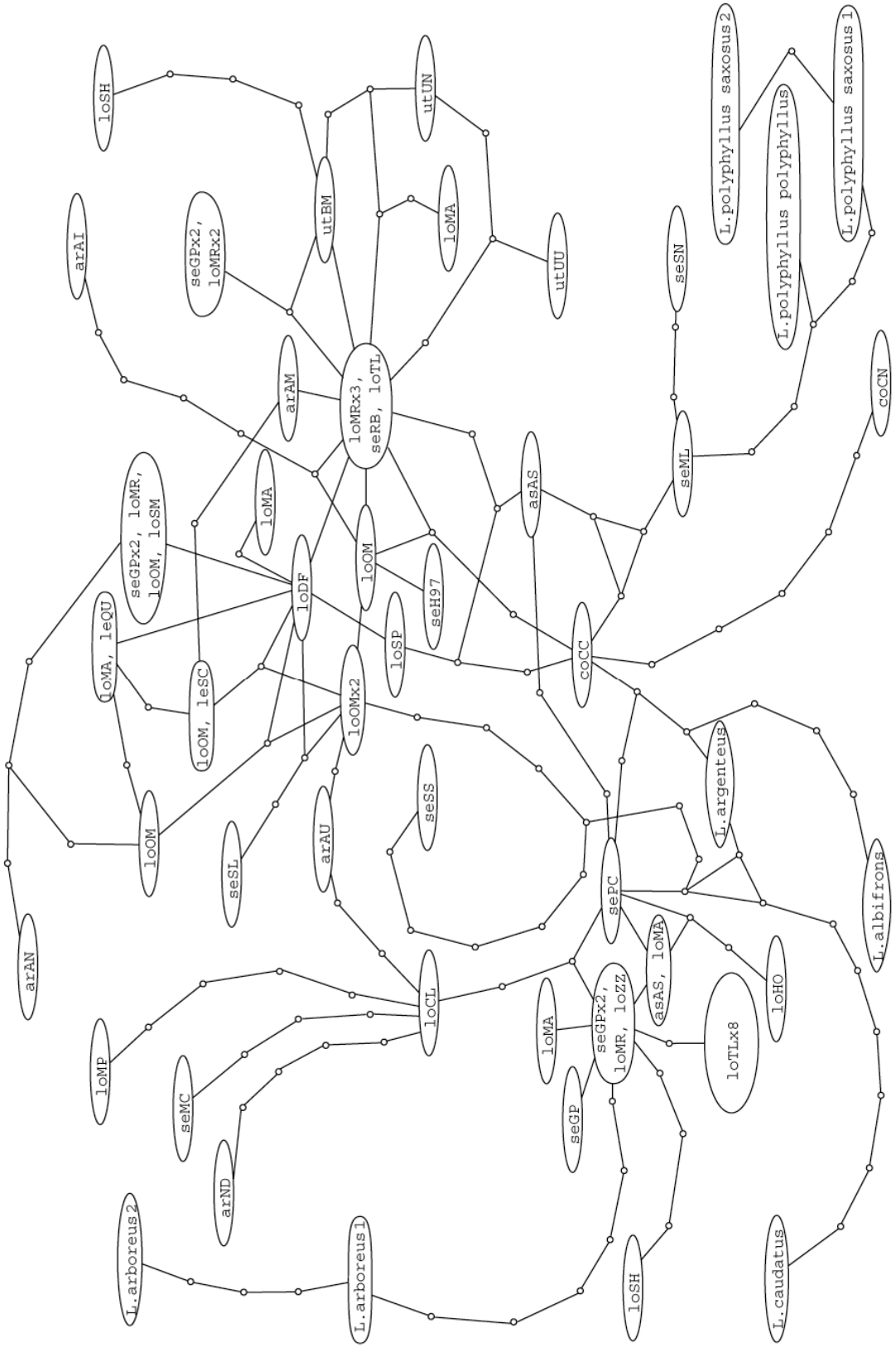


Figure 5: LEGCYC1A-TCP 50% majority rule consensus of trees sampled in the stationary phase of Bayesian analysis. *L. lepidus* varieties are highlighted by colors as follows: *lobbii* blue, *sellulus* green, *ashlandensis* red, *lepidus* brown, *aridus* magenta, *utahensis* orange, *confertus* light blue. Outgroup taxa are gray. Branch labels are posterior probabilities. Scale is in substitutions per site. *Three individuals from the Mt. Hood Timberline population shared this sequence.


loSH
utUN
arAI
seGPx2, loMRx2
utBM
loMA
utUU
seSN
L.polyphyllus saxosus2
L.polyphyllus polyphyllus
L.polyphyllus saxosus1
arAM
loMRx3, seRB, loTL
coCN
seML
seGPx2, loMR, loOM, loSM
loMA
asAS
loDF
loOM
seH97
loMA, leQU
loSP
coCC
loOM, leSC
loOMx2
seSS
L.argenteus
loOM
seSL
arAU
sePC
L.albifrons
arAN
loCL
asAS, loMA
loMA
seGPx2, loMR, loZZ
loHO
loMP
loTLx8
seMC
seGP
arND
L.arboreus2
L.arboreus1
loSH
L.caudatus

Preceding page: Figure 6: LEGCYC1A-TCP haplotype network. Each step represents a point mutation. Small circles represent unsampled haplotypes. Multiple individuals from the same population are indicated (e.g. x2). *L. lepidus* individuals are abbreviated to represent the variety and population of origin (given below).

Variety	Abbreviation	Population
aridus	arAI	Idaho, Nez Perce Co.
	arAM	Eastern Oregon, Morrow Co.
	arAN	Nevada, Humboldt Co.
	arAU	Utah, Washington Co.
	arND	The Dalles, Oregon
ashlandensis	asAS	Mt. Ashland, Oregon
confertus	coCC	California, Lassen Co.
	coCN	Nevada, Mineral Co.
lepidus	leQU	Vancouver, Washington
	leSC	Scatter Creek, Washington
lobbii	loCL	Crater Lake, Oregon
	loDF	Mt. Bachelor, Oregon
	loHO	Mt. Howard, Oregon
	loMA	Mt. Adams, Washington
	loMP	Marys Peak, Oregon
	loMR	Mt. Rainier, Washington
	loOM	Olympic Mtns., Washington
	loSH	Mt. St. Helens, Washington
	loSM	Steens Mtn., Oregon
	loSP	Santiam Pass, Oregon
	loTL	Mt. Hood, Timberline Lodge, Oregon
	loZZ	Mt. Hood, Zig Zag Canyon, Oregon
sellulus	seGP	Grizzly Peak, Oregon
	seH97	La Pine, Oregon
	seMC	Mt. McLoughlin, Oregon
	seML	Mt. Lassen, California
	sePC	California, Placer Co.
	seRB	Rattlesnake Butte, Oregon
	seSL	Medicine Lake, California
	seSN	Nevada, Douglas Co.
	seSS	California, Squaw Valley
utahensis	utBM	Baldy Mtn., Oregon
	utUN	Nevada, Rose Mtn.
	utUU	Utah, Garfield Co.

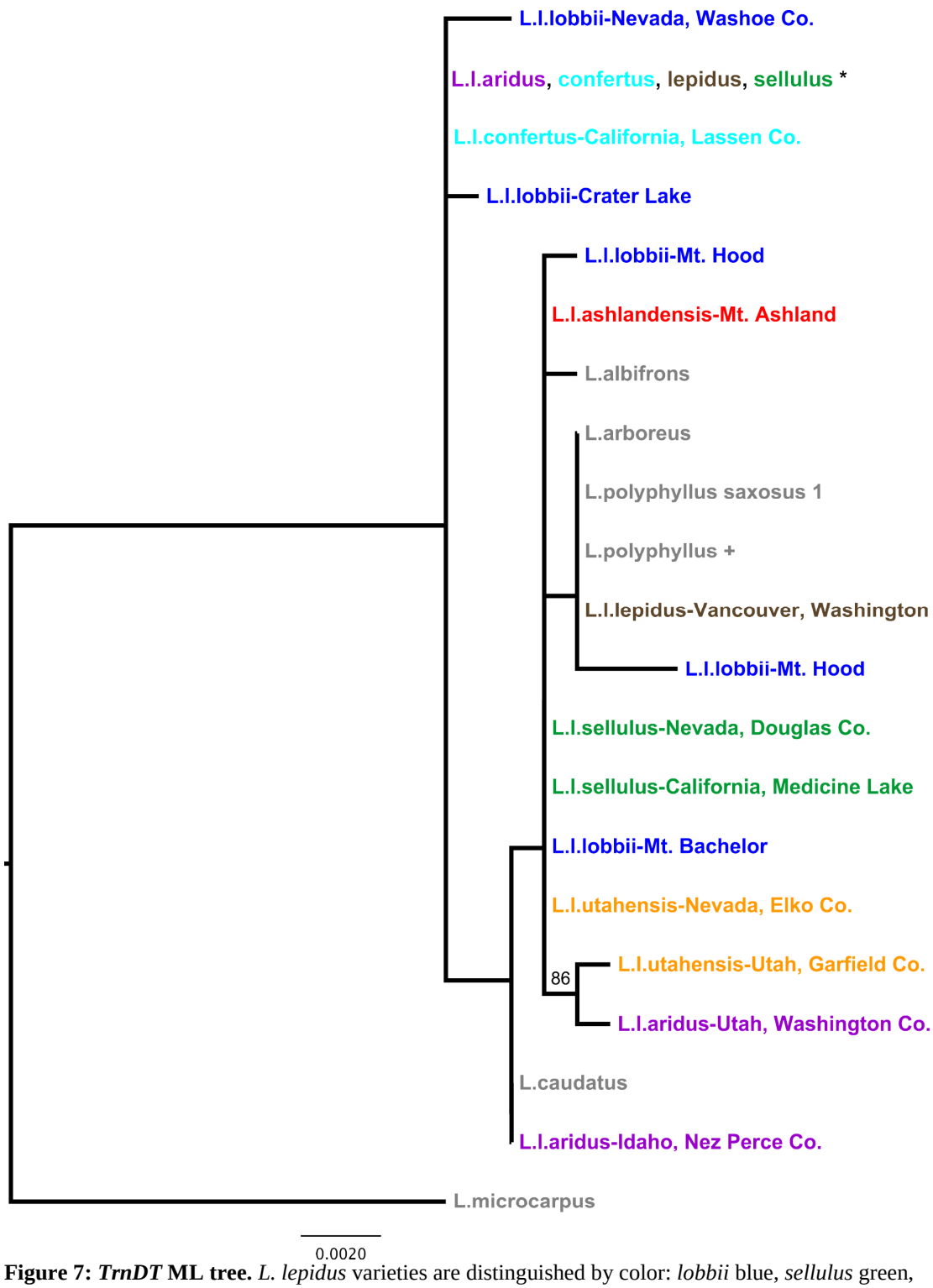


Figure 7: *TrnDT* ML tree. *L. lepidus* varieties are distinguished by color: *lobbii* blue, *sellulus* green, *ashlandensis* red, *lepidus* brown, *aridus* purple, *utahensis* orange, *confertus* light blue. Outgroup taxa are gray. Bootstrap values >50 are included for branches of non-trivial length. Scale is in substitutions per site. *Four individuals with identical sequences: *L. l.* var. *aridus*-The Dalles; var. *confertus*-Nevada, Mineral Co.; var. *lepidus*-Vancouver, Washington; var. *sellulus*-Grizzly Peak. +Two individuals with identical sequences: *L. polyphyllus* var. *polyphyllus* and var. *saxosus* 2.

Table 4: Summary statistics of genetic variation in *L. lepidus* varieties based on LEGCYC1A-TCP data. Genotypes are given as a range from the minimum number after resolving ambiguities to the maximum number of unique sequences.

Variety	N (individuals)	Genotypes	Nucleotide diversity, π_n (s.d.)	θS (s.d.)	θπ (s.d.)
aridus	5	5	0.0218 (0.0122)	10.6046 (4.6140)	11.6292 (6.5170)
confertus	2	2	0.0152 (0.0107)	7.0909 (4.1528)	7.8346 (5.5224)
lepidus	2	2	0.0056 (0.0043)	2.7273 (1.7896)	3.0649 (2.3882)
sellulus	15	12 - 15	0.0191 (0.0100)	10.6016 (3.5903)	9.9830 (5.2257)
utahensis	3	3	0.0149 (0.0093)	7.4453 (3.8336)	7.6760 (4.8190)
ashlandensis	2	2	0.0114 (0.0083)	4.3636 (2.6800)	5.2611 (3.8356)
lobbii	36	18 - 34	0.0159 (0.0083)	8.4590 (2.5054)	7.7122 (4.0300)
Mean	9.29	6.29 – 9.00	0.0148 (0.0090)	7.3275 (3.3094)	7.5944 (4.6197)

Table 5: Summary statistics of *L. lepidus* populations sampled for >1 individual, based on LEGCYC1A-TCP data. Genotypes are given as a range from the minimum number after resolving ambiguities to the maximum number of unique sequences.

Population	N (individuals)	Genotypes	Nucleotide diversity, π_n (s.d.)	θS (s.d.)	θπ (s.d.)
Olympic Mtns.	6	5 - 6	0.0141 (0.0079)	7.6162 (3.2585)	7.7762 (4.3874)
Mt. Rainier	7	4 - 7	0.0136 (0.0076)	6.2890 (2.6513)	7.4945 (4.1821)
Mt. Adams	5	5	0.0163 (0.0095)	6.7162 (3.0446)	6.8581 (3.9896)
Mt. St. Helens	2	2	0.0110 (0.0079)	4.9091 (2.9753)	6.0524 (4.3549)
Mt. Hood, Timberline	9	2 - 7	0.0062 (0.0038)	3.4888 (1.5279)	2.9391 (1.8053)
Mt Ashland	2	2	0.0114 (0.0083)	4.3636 (2.6800)	5.2611 (3.8356)
Grizzly Peak	7	4 - 7	0.0127 (0.0072)	5.6601 (2.4182)	6.4224 (3.6328)
Mean	5.43	3.43 - 5.14	0.0122 (0.0075)	5.5776 (2.6508)	6.1148 (3.7411)

Table 6: Pairwise F_{ST} values between *L. lepidus* varieties based on LEGCYC1A-TCP data. F_{ST} estimates are below the diagonal, P-values are above. Numbers in bold indicate F_{ST} estimates that are significantly greater than zero (P<0.05).

	aridus	***confertus***	***lepidus***	***sellulus***	***utahensis***	***ashlandensis***	***lobbii***
aridus	0	0.108	0.486	0.117	0.144	0.423	**0.036**
confertus	0.29803	0	0.369	0.081	0.081	0.288	**0.000**
lepidus	0.17760	0.61345	0	0.090	0.126	0.324	**0.018**
sellulus	0.07903	0.24808	0.21969	0	**0.027**	0.450	**0.045**
utahensis	0.16546	0.26899	0.49668	**0.16307**	0	0.189	**0.045**
ashlandensis	0.20807	0.33373	0.58339	0.06909	0.30653	0	0.225
lobbii	**0.08990**	**0.30770**	**0.24862**	**0.03631**	**0.14423**	0.12016	0

Table 7: Pairwise F_{ST} values between *L. lepidus* populations sampled for >1 individual, from LEGCYC1A-TCP data. F_{ST} estimates are below the diagonal and P-values are above. Blue values compare northern populations to northern populations, yellow values compare southern to northern, and green values compare southern to southern. Numbers in bold indicate F_{ST} estimates that are significantly greater than zero (P<0.05).

	Olympic Mtns.	Mt. Rainier	Mt. Adams	Mt. St. Helens	Mt. Hood, Timberline	Mt Ashland	Grizzly Peak
Olympic Mtns.	0	**0.000**	**0.000**	**0.036**	**0.000**	0.054	**0.000**
Mt. Rainier	**0.22491**	0	**0.000**	0.054	**0.000**	**0.036**	0.333
Mt. Adams	**0.22843**	**0.21486**	0	0.117	**0.009**	0.279	**0.000**
Mt. St. Helens	**0.49538**	0.34173	0.33601	0	**0.036**	0.315	0.090
Mt. Hood, Timberline	**0.45509**	**0.30172**	**0.26220**	**0.43664**	0	**0.036**	**0.000**
Mt Ashland	0.27980	**0.25777**	0.12846	0.41312	**0.22147**	0	**0.000**
Grizzly Peak	**0.22899**	0.02180	**0.22948**	0.38383	**0.08884**	**0.29009**	0

Table 8: Results of hierarchical analysis of molecular variance (AMOVA) from LEGCYC1A-TCP data.

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation	*P*-value
Among varieties	6	68.39	0.15861	3.60	0.03226
Among populations within varieties	27	246.736	1.82932	41.51	<0.00001
Within populations	96	232.205	2.41880	54.89	<0.00001
Total	129	547.331	4.40674		

Chapter 3

Population genetic analyses of microsatellite data

Reproductive isolation between populations will leave genetic signatures due to the fixation of different sets of alleles and the accumulation of novel alleles in separate populations. Utilizing several rapidly evolving loci in population analyses allows the detection of these signatures after relatively short periods of isolation. Microsatellites are a popular tool for population genetic studies due to their relatively high mutation rate (~10^3 – 10^4 substitutions per generation), their codominant nature, and, once primers have been developed, the relative ease of screening many individuals at many loci (Avise 2004). Microsatellite analyses were carried out in several populations of *Lupinus lepidus* in order to measure the differentiation of morphologically defined varieties and to test hypotheses regarding the colonization of populations in the northern portion of the Cascade Mountains.

Materials and Methods

Collections and Character Sampling

Collections of *Lupinus lepidus* and total genomic DNA isolation are described in Chapter 2. Microsatellite primers were previously developed for *Glycine max, Lupinus* group *Microcarpi*, and *Lupinus nanus* (Table 2; Peakall et al. 1998, Drummond and Hammilton 2005, Morris 2009). Included in analyses as one of the microsatellite loci was a 15 bp indel found within LEGCYC1A-TCP (i.e. 1 or 2 repeats of a 15 bp motif) using the primers LEGCYC1-F2 and LEGCYC1A-R1. To allow the pool-plexing of multiple loci the M13F primer sequence was added as an extension to the 5’ end of the forward

primers (O'Quinn and Fishbein 2009, Schuelke 2000). A separate M13F primer, 5' labeled with a FAM or HEX fluorescent dye, was added to the reaction mix. During thermal cycling the labeled M13F primer can anneal to a previously copied locus fragment, creating a fluorescently labeled fragment after the next extension step.

PCR reactions were carried out with 0.25 U Biolase *Taq* DNA polymerase (Bioline, Taunton, MA), 1 X Biolase reaction buffer, 200 μM each dNTP, 400 nM each forward and reverse primer, 100 nM fluorescently labeled M13F primer, 2.5 mM $MgCl_2$, and water to 10 μL. Thermal cycling for loci from Peakall et al. (1998) and Drummond and Hamilton (2005) included an initial step of 5 min @ 95°C; 35 cycles of 30 s @ 94°, 60 s @ 55°, and 30 s @ 72°; a final extension at 72° for 10 min, and a final hold at 4°. Thermal cycling for the LEGCYC1A-TCP locus and the loci from Morris (2009) included an initial step for 4 min @ 95°; 2 cycles of 60 s @ 94°, 60 s @ 60°, and 35 s @ 70°; 18 cycles of 45 s @ 93°, 45 s @ 59°, dropping 0.5° per cycle, and 45 s @ 70°; 20 cycles of 30 s @ 92°, 30 s @ 50°, and 60 s @ 70°; a final extension at 72° for 5 min, and a final hold at 4°.

Labeled PCR product from up to 4 loci were pooled and run together with the internal size standard ROX-500 on an ABI 3730 capillary sequencer (Applied Biosystems) at the Oklahoma State University Recombinant DNA/Protein Core Facility. Individuals were genotyped using the program STRand 2.3.106 (Toonen and Hughes 2001).

Population genetic statistics

Within-variety average gene diversity was computed across loci using Arlequin 3.5.1.2 (Excoffier and Lischer 2010). *F*-statistics and *R*-statistics were computed for

pairwise comparisons of populations and calculation of apportionment of genetic variation among varieties, among populations within varieties, and within populations (AMOVA). Significance of AMOVA statistics and pairwise comparisons between varieties were calculated from a distribution of $2x10^4$ and $2x10^3$ permutations, respectively. Loci with greater than 10% missing data within a variety or population were excluded from summary statistics or pairwise comparisons of that group, loci with greater than 10% missing data across all individuals were excluded from AMOVA analyses. Within each variety, each locus was tested for divergence from Hardy-Weinberg equilibrium (HWE) using a Markov process (10^6 steps, discarding 10^5 as burnin).

In addition to grouping populations by variety, analyses were performed grouping populations geographically into a northern and a southern group. Five analyses were performed with the southern group comprising all populations south of Mt. Rainier, Scatter Creek, Mt. Hood, Santiam Pass, and Mt. Bachelor, respectively. A sixth analysis included all populations south of Mt. Bachelor, but also placed populations in eastern Oregon (Baldy Mtn. and Mt. Howard) with the southern group. This range of locations was chosen because is spans the latitudes of north-south disjunctions found in other regional plant groups, and accounts for geographic clustering of peaks along the Cascades (i.e. the close grouping of Mounts Hood, St. Helens, and Adams). Groups were analyzed for group gene diversity, mean population gene diversity, pairwise F_{ST}, and the percent of variation apportioned among groups.

Two additional analyses were performed to test alternative hypotheses of geographic structuring. To test for an effect of elevation, populations were grouped into either high elevation (on a peak or the crest of the Cascades, N=17) or low elevation

(valley or at the foot of a mountain, N=8) categories. To test for a longitudinal effect possibly created by the east-west barrier of the Cascades, populations were grouped into either an eastern (N=7), western (N=5), or central category (containing populations within the Cascades, N=13). For each analysis, percent of variation apportioned among groups was measured.

The influence of the stepwise mutation process on measures of genetic distance was calculated by permutation of allele size (repeat number) among alleles within a locus using the program SPAGeDi 1.3 (Hardy and Vekemans 2002) with $2x10^4$ permutations. This program was also used to test the correlation of genetic distance with geographic distance. Permutation of population location among populations provided a test for significance ($2x10^4$ permutations; H_0, slope of regression=0), and jackknifing over loci provided an estimate of confidence intervals for the slope of the regression.

Microsatellite data were also analyzed using the program Structure ver. 2.3.3 (Pritchard et al. 2000, Falush et al 2003). Instead of analyzing an *a priori* grouping of populations, Structure assigns individuals to one or more of K genetic clusters (the Structure literature refers to these as 'populations,' whereas I am using 'population' to refer to the individuals of a particular site, so I will continue to use the term 'cluster' to refer to the groupings Structure creates). This can allow the user to uncover unexpected sources of genetic structuring. Individuals that show signs of genetic admixture can be jointly assigned to more than one cluster (i.e. some proportion of an individual's genome came from a particular cluster, Pritchard et al. 2000). In the analysis performed here, the posterior cluster probabilities of each individual in a population are pooled and displayed as the proportion of each cluster contributing to that population.

Structure analysis was performed using the admixture model with correlated allele frequencies among clusters. Default values were used for all other parameters including the degree of admixture (α) and allele frequencies (λ). One locus, CAC60-13-6, had missing data non-randomly distributed across populations and was coded as containing a recessive, null allele (Falush et al. 2007). The number of clusters, K, is set by the user and is not known *a priori* in this case. The optimal value of K was estimated using the method of Evanno et al. (2005), measured using 10 independent runs of each K value from 1-10 (using Markov chains of $3x10^4$ steps, discarding 10^4 as burnin). The optimal K value was then run independently 5 times, each with 10^5 steps and discarding $3x10^4$ steps as burnin. Cluster assignments from the 5 runs were averaged using the program CLUMPP (Jakobsson and Rosenberg 2007).

Results

All of the 8 loci studied were polymorphic across varieties. One, CAC60-13-6, had greater than 10% missing data. Gene diversity (the probability that two randomly chosen haplotypes are different), observed and expected levels of heterozygosity (H_O and H_E, averaged across loci), and the number of loci significantly diverging from HWE are given for each variety in Table 9. Each variety had a lower average observed than expected level of heterozygosity, with varieties *aridus*, *lobbii*, *sellulus*, and *utahensis* having F_{IS} values significantly greater than zero ($P<0.01$ in all cases). The mean F_{IS} value was significantly greater than zero ($F_{IS}=0.128$, $P<0.00001$) with 9 of 25 populations having a F_{IS} value significantly greater than zero.

Global *R*-statistics were not significantly different than *F*-statistics after permutation of allele size among alleles within each locus (P-values: *Rit*=0.36, *Ris*=0.95, *Rst*=0.22), indicating that stepwise mutation does not significantly influence genetic structure. Because *R*-statistics generally have greater sampling variances than the corresponding *F*-statistics, only *F*-statistics are further considered (Hardy et al. 2003).

AMOVA analysis indicates that little genetic structuring is present in microsatellite loci of *Lupinus lepidus*, with 75.06% of variation present between individuals within a population, 20.56% among populations, and only 4.38% between varieties (Table 10). Permutation tests indicate that apportionment of variation between varieties is significantly greater than zero (P=0.033).

Total gene diversity is similar between northern and southern groupings of populations, and average population gene diversity is similar between northern and southern groups, except when the northern group is highly restricted. This indicates that the most northern populations are less diverse on a per-population level, but the group of far northern populations is not significantly genetically depauperate compared to the southern group (Table 11).

As the northern grouping is expanded to include more southern populations, the pairwise F_{ST} value between the groups decreases. Likewise, the percent of genetic variation apportioned between groups decreases until it is no longer significantly different from zero (Table 11). When populations at or farther north than Mt. Hood are considered, grouping populations by northern or southern identity accounts for more of the genetic variation than grouping by variety (12.33%, 7.10% and 5.25% for north-south

groupings vs. 4.38% for varietal grouping). In all grouping schemes, the percent of variation apportioned among and within populations is similar to grouping by variety.

Tests of elevational and longitudinal structuring indicate that the percent of variation among groups (Va) is not significantly different from zero (elevation: Va=1.54, P=0.094; longitude: Va=1.82, P=0.078). Tests of geographic distance on genetic structure indicate that spatial distance does correlate with genetic distance (slope > 0, P=0.0075) and can explain nearly 14% of the variation in population pairwise F_{ST} values (R^2=0.1377, using Slatkin's linearized F_{ST} value, $F_{ST}/(1- F_{ST})$). A histogram of all population pairwise comparisons is given in Figure 8 and displays a unimodal shape.

Structure analysis indicates that the microsatellite data is optimally grouped into 4 clusters (K=4, mean Ln probability of the data = -4012.49, Fig. 9). One cluster, cluster B, is found almost exclusively within the three northernmost populations: Mt. Rainier, Blue Mtn., and Mt. Baker. These populations, in turn, are represented almost exclusively by cluster B (91.7% - 94.9%). Cluster A is found mainly on the eastern side of the Cascade Mountains, although is strongly represented within the Cascades at Mt. Bachelor and Mt. Adams, and in the northern Sierra Nevadas at Mt. Lassen. A majority of genes are also representative of cluster A in the Vancouver, Washington population (56.1%, N=2). Cluster C is found primarily in populations in eastern Oregon (Mt. Howard and Steens Mtn.) and in the central Cascades (Santiam Pass, Mt. Hood, and Mt. St. Helens). Cluster D is heavily represented in populations west of the Cascades and in the central and southern Cascades (from La Pine to Mt. Ashland). While clusters A, C, and D are often found at lower proportions in populations outside of their primary regions (outlined above), Cluster B is only rarely represented outside of the northernmost populations, and

then usually at a very low proportion: Jordan Creek, 25.4%; Mt. Hood (Hwy. 26), 15.7%; Mt. Ashland, 14.7%; Scatter Creek, 14.4%, Mt. St. Helens, 11.2%: and Mt. Howard, 9.0%.

Gene diversity (expected heterozygosity) is high within each of the four clusters, ranging from 0.377 in cluster B to 0.584 in cluster A. F_{ST} values (as calculated in Falush et al. 2003) for each cluster range from moderately low (A=0.062, C=0.074), to medium (D=0.142), to high (B=0.249). Pairwise net nucleotide distances are very low between clusters A and C, and high between B and D (other comparisons are moderate, Table 12). Cluster B contains eight alleles across three loci that are at a very low frequency (<1%) in the other clusters or are nearly twice as frequent in cluster B; three of these alleles are both at a frequency >10% in cluster B and <1% in the other clusters.

Discussion

Populations of *Lupinus lepidus* are characterized by high gene diversity, with the great majority of genetic variation across the species found at the intra-population level. Differences between populations account for a minority of overall genetic variation (20%). However, some of the population differences (14%) can be correlated to geographic distance. A very small portion of the total genetic variation is explained by variety identity (4.38%).

High varietal F_{IS} values may be due to cryptic structuring within varieties (except *utahensis* which contained one population; i.e. the Wahlund effect; Hartl and Clark 1997). The ability of *L. lepidus* to reproduce via autogamy may be causing F_{IS} values in some populations to be greater than zero. The prevalence of outcrossing is thought to be

variable and related to pollinator abundance, and the overall F_{IS} value reported here (0.128) is much lower than that previously reported in *L. lepidus* (0.49 in populations on Mt. St. Helens unaffected by the eruption; Bishop and Dyer 1996). (The F_{IS} value currently measured at the Mt. St. Helens population is not significantly different from zero, but is comparable to previously measured values in populations recolonizing after the eruption [0.16 vs. 0.19, respectively; Bishop and Dyer 1996]).

Varieties of *L. lepidus* are spatially segregated, complicating measurements of the relative influences of variety identity and geographic distance. Still, the very low apportionment of genetic variation across varieties implies that geographically close populations of different varieties can be more similar than distant populations of the same variety. This can be seen, for example, in the Crater Lake (var. *lobbii*) population, which has a similar genetic makeup to the Mt. McLoughlin (var. *sellulus*) population, but is dissimilar to other var. *lobbii* populations on Mt. Hood or in the northern populations (Fig. 9).

Hypotheses of recolonization of northern areas in the Pacific Northwest predict that northern population groups will have reduced genetic diversity compared to the group of southern populations, and will have alleles that are unique to the north or are at low frequency in the south (Soltis et al. 1997). The data here support these predictions only for the three northernmost populations sampled. The northernmost populations (Mt. Rainier, Scatter Creek, Olympic Mtns., and Mt. Baker) have lower genetic diversity values than southern populations, although measures of total gene diversity are not significantly different between groups of northern and southern populations (Table 11). This could be due to a recolonization from a northern refuge, or a recolonization event

from the south involving high migration rates or multiple introductions, followed by a later restriction of migration. The pairwise F_{ST} values and the percent of variation explained by a northern and southern grouping remain moderate for most of the north-south groupings, but are much higher when comparing the most northern populations against the remaining populations.

The microsatellite data exhibit a strong break between the three northernmost populations and the remaining populations. These three populations are almost entirely represented by cluster B, which has several alleles that are nearly unique to that cluster. These microsatellite data are consistent with an observation of a northern disjunction, and meet the expectations presented by Soltis et al. (1997). Unlike the north-south disjunctions presented by Soltis et al. (1997), which are all located south of the Columbia River, the break shown here is located much farther north at Mt. Rainier.

Cluster B is not entirely restricted to the three northern populations. It is present at the Jordan Creek site (which is represented by a single individual), one of the populations on Mt. Hood, the Scatter Creek population, and to a lesser extent at Mt. St. Helens and Mt. Howard. Oddly, it is also moderately represented (14.7%) at the Mt. Ashland population several hundred kilometers away from Mt. Rainier. It is likely that many of these outlying locations of cluster B are the result of gene flow out of the northern populations. However, it is possible that as new areas became open to recolonization the northern locations were seeded from a small number of migrants that carried cluster B alleles, perhaps from Mt. Hood, Mt. St. Helens, or Scatter Creek. Indeed, in the wake of the volcanic eruption on Mt. St. Helens, cleared areas on the pumice plains were

colonized by a single individual of *L. lepidus* that then reproduced to form a very large population (>16,000 individuals; Bishop and Dyer 1996).

Among the populations south of the microsatellite disjunction, there may be slight geographic structuring. As the grouping of northern populations is expanded to include more southern populations, pairwise F_{ST} values and the percent of variation among groups tend to decrease (Table 11). This may indicate that these populations have sets of alleles that are more similar to the larger group of southern populations. Conversely, the decrease in these values when eastern populations are grouped with the south implies that these populations tend to share alleles with the northern group. These conclusions are supported by the Structure analysis, showing the central populations of Santiam Pass and Mt. Bachelor strongly represented by clusters A and D, which are common in the south; and the eastern populations of Mt. Howard and Baldy Mtn. strongly represented by cluster C, which is at low frequency in the south (Fig. 9).

The distribution across the landscape of the genetic clusters inferred by Structure presents possible scenarios for the expansion of *Lupinus lepidus* north from California (Fig. 9). First, clusters A and C are separated by a very small net nucleotide distance of 0.031 (and may not represent distinct clusters: the Structure model used may overestimate K in some cases, and results when K=3 tend to combine these clusters into one; Falush et al. 2003). Both of these clusters are moderately distant from clusters B and D (ranging from 0.065-0.084, Table 12), and net nucleotide distance between B and D is high (0.151). This is consistent with a scenario in which an early population in northern California (such as at Mt. Lassen) expanded northward on both sides of the Cascades. In populations of the southern Cascades and farther west allele frequencies would move

toward those now identified as cluster D, populations of the north-central Cascades and east would move toward clusters A+C. As migrants from clusters A+C moved into the far north, they would further diverge into cluster B (either due to the migration itself as in the leading edge hypothesis, or after a period of isolation as in a scenario involving a northern refuge). This scenario accounts for the strong divergence between clusters B and D, but moderate divergence of both clusters away from A+C.

Alternatively, populations in Oregon may have originally been homogenous, and only split into clusters D and A+C during glaciation of the Cascade Range. In either case, migration may be common throughout most of the region, as seen in the mixing between clusters observed along the crest of the Cascades. Additional movement between the east and west sides of the Cascades may be possible via the Columbia River Gorge, as seen in the similarity between the Vancouver, Washington population and those of the eastern regions (Fig. 9).

A rapid expansion into the Oregon and Washington region is additionally supported by the distribution of pairwise F_{ST} values between populations (Fig. 8). The unimodal shape of this distribution is consistent with a rapid increase in population size or a rapid spatial expansion coupled with high migration rates (Excoffier and Lischer 2010, Slatkin and Hudson 1991). However, it should be noted that predictions of a unimodal distribution were made for mismatch distributions between individuals, whereas here it is populations that are being compared.

Conclusions

Analysis of 8 microsatellite loci across 25 populations of *Lupinus lepidus* has found only negligible genetic structure. Most of the genetic variation is found within populations. Differences between populations account for a minority of genetic variation and differences between any tested grouping of populations account for even less. These results do not support a hypothesis that varieties of *L. lepidus* have been reproductively isolated for a long period of time. The microsatellite data do show evidence that the three northernmost populations examined are genetically distinct from the southern populations. The data are consistent with a hypothesis of colonization that involved a rapid expansion from southern populations, possibly independently on either side of the Cascades, coupled with an additional divergence that formed the northernmost populations. Maintenance of strong morphological differences in spite of low divergence between varieties will be discussed in the Conclusion, but likely involves differential selection between habitats that is strong enough to overcome ongoing gene flow.

Table 9: Mean microsatellite gene diversity of *Lupinus lepidus* varieties. Unanalyzed loci for each variety have >10% missing data, and are excluded from the gene diversity calculation. $L \neq H_E$: number of usable loci that significantly differ from HWE.

Variety	N	Polymorphic loci analyzed	Gene Diversity	(s.d.)	H_O	H_E	$L \neq H_E$
sellulus	46	6	0.506	(0.295)	0.338	0.510	3
aridus	23	7	0.606	(0.340)	0.466	0.569	2
lepidus	11	4	0.391	(0.241)	0.528	0.714	0
ashlandensis	6	4	0.439	(0.294)	0.456	0.614	1
utahensis	10	5	0.397	(0.239)	0.500	0.635	2
lobbii	125	8	0.503	(0.279)	0.360	0.526	6

Table 10: Results of hierarchical analysis of molecular variance (AMOVA) from microsatellite data, testing the genetic structure for *L. lepidus* varieties.

Source of Variation	*df*	Sum of squares	Variance components	Percentage of variation	*P*-value
Among varieties	5	59.00	0.07869	4.38	0.03311
Among populations within varieties	19	151.34	0.36927	20.56	<0.00001
Within populations	417	562.30	1.34845	75.06	<0.00001
Total	441	772.64	1.79641		

Table 11: North-south structuring of genetic diversity. Group gene diversity, mean population gene diversity, pairwise F_{ST}, and percent of variation apportioned among groups. Bold values are significantly different (P<0.05) between northern and southern groups (mean population genetic diversity) or 0 (F_{ST} and percent variation).

Latitudinal cutoff	Populations per group north / south	Group gene diversity north : south	s.d. north : s.d. south	Mean population gene diversity north : south	Pairwise F_{ST}	% variation between groups
South of Mt. Rainier	3/22	0.350 : 0.503	0.208 : 0.284	**0.246** : **0.449**	**0.154**	**12.33**
South of Scatter Creek	4/21	0.394 : 0.508	0.229 : 0.286	**0.273** : **0.454**	**0.099**	**7.10**
South of Mt. Hood	13/12	0.545 : 0.486	0.304 : 0.276	0.423 : 0.428	**0.069**	**5.25**
South of Santiam Pass	15/10	0.506 : 0.528	0.280 : 0.304	0.419 : 0.434	**0.053**	**3.50**
South of Mt. Bachelor	18/7	0.510 : 0.502	0.282 : 0.285	0.423 : 0.431	**0.037**	1.48
South and east of Mt. Bachelor (see text)	16/9	0.507 : 0.560	0.281 : 0.320	0.418 : 0.437	**0.029**	0.94

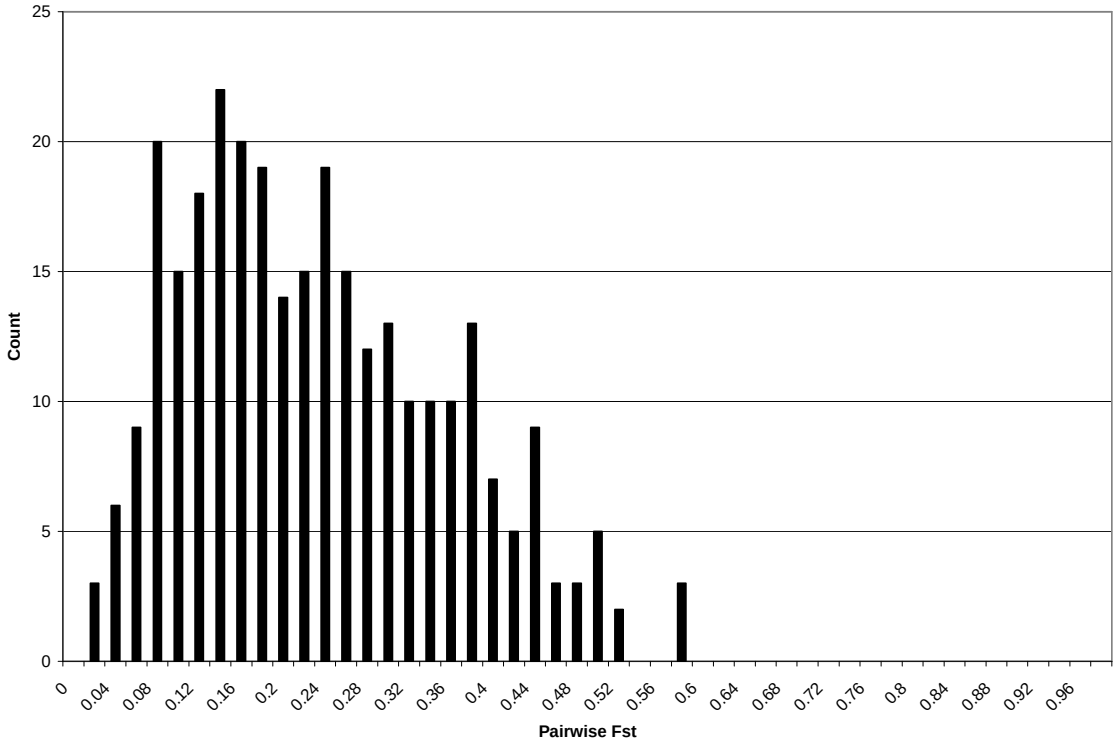


Figure 8: Pairwise F_{ST} values for microsatellite data between all populations.

Table 12: Net nucleotide distances between clusters inferred from Structure.

Cluster	A	B	C
B	0.084		
C	0.031	0.069	
D	0.065	0.151	0.069

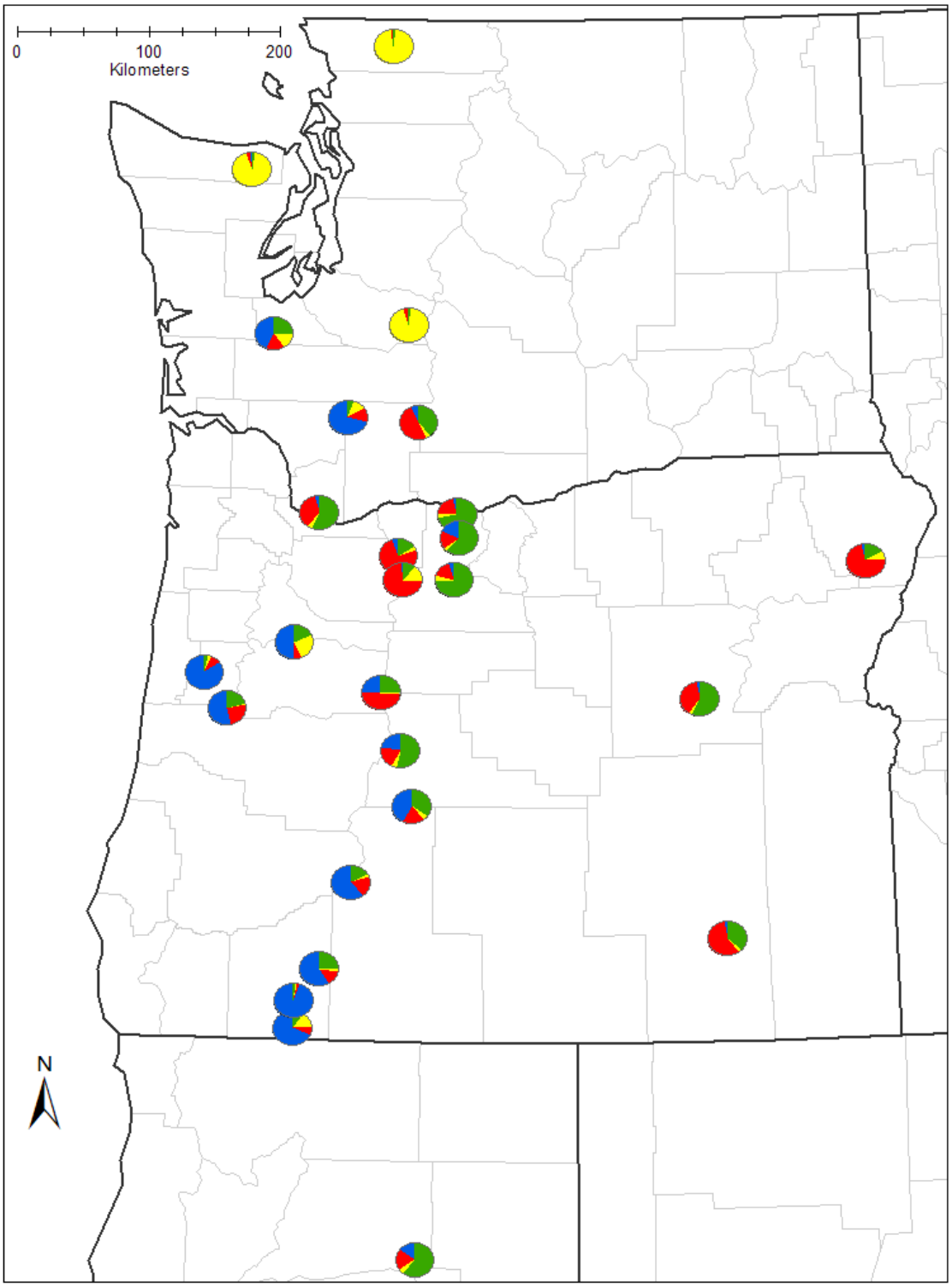


Figure 9: Structure analysis of *L. lepidus* microsatellite data. Each pie represents the proportion of each genetic cluster contributing to that population. Cluster A green; B yellow; C red; D blue.

Chapter 4

Conclusion

Microsatellite analyses and sequence data from multiple loci describe alternate patterns within *Lupinus lepidus*. In both analyses the varieties of *L. lepidus* are not shown to form monophyletic groups and there is very little genetic differentiation between them. Phylogeographic patterns expected from a recolonization of the northern portions of the Cascade Mountains are not found in the sequence data, but are present in the microsatellite data. There are additional indications that colonization of the Oregon and Washington regions occurred from a rapid expansion from the south that may have independently spread along each side of the Cascades. In light of strong morphological differentiation found within the species coupled with negligible genetic structure, it is likely that the varieties of *Lupinus lepidus* are subject to different suites of selective forces and are undergoing rapid adaptation to local ecologies.

Lupinus lepidus **is not monophyletic, nor are its varieties.**

Sequence data from nuclear and chloroplast loci do not show *L. lepidus* forming a monophyletic group. *Lupinus lepidus* is consistently shown to fall within several separate clades, with several samples being resolved as sister to outgroup taxa (Figs. 4-5, 7). This does not necessarily indicate that *L. lepidus* is not a "good species," with cryptic taxa that truly share more recent common ancestors with outgroup taxa than with other members of *L. lepidus*, rather, these findings can reasonably be considered the result of a very recent divergence of *L. lepidus* from other members of the western perennial clade of *Lupinus*.

Taxa that have recently diverged are more likely to share alleles that differentiated at a point earlier than the separation of the two taxa (Maddison 1997). Phylogenetic analyses of these alleles, without consideration that they represent gene lineages within species lineages, can lead to the incorrect inferences concerning species monophyly and relationships. Indeed, simulations have shown that for some combinations of branch lengths and effective population sizes, the most likely gene tree is incongruent with the true species tree, although this is rarely demonstrated in empirical data (Huang and Knowles 2009, Degnan and Rosenberg 2006). As a result, multiple alleles sampled within a species may be resolved as paraphyletic to homologous alleles from an outgroup taxon (incomplete lineage sorting, ILS).

In light of the recent diversification of the western perennial *Lupinus* (0.7-2.1 mya), the distinctive morphology of *L. lepidus* (its caespitose habit), and the previous difficulty in defining relationships within the western perennial clade (likely due in part to ILS), the lack of monophyly inferred from the LEGCYC1A-TCP and *trnDT* loci should not be considered evidence that *L. lepidus* is truly a paraphyletic group (Huang and Friar 2009, Drummond 2008). It is additionally possible that hybridization has occurred in the past between members of *L. lepidus* and other western perennial lupines, causing introgression of these alleles into *L. lepidus*. However, no populations or individuals observed for the present study exhibited morphological signs of hybridization with other western perennial lupines, including in at least three populations where they were growing sympatrically with *L. lepidus*.

Similarly, LEGCYC1A-TCP and *trnDT* do not show any of the varieties of *L. lepidus* to be clades. Again, the recent origin of these varieties (<0.7-2.1 mya) makes this

a probable finding, even under a hypothesis of complete isolation between varieties. Therefore, while a finding of reciprocal monophyly among varieties could have supported such a hypothesis, the lack of such a finding, in two loci that themselves are poorly resolved, cannot distinguish between hypotheses ranging from complete isolation to panmixia.

The use of multiple fast-evolving loci such as microsatellites should be able to measure the effect of genetic isolation much earlier than a few, slowly evolving loci. However, analysis of microsatellite data in *L. lepidus* shows only negligible genetic structure, and actually apportions a greater amount of genetic variation to the within population level than analysis of sequence data (75% and 55%, respectively, Tables 10 and 8). While both analyses attribute some genetic variation to varietal identity, both values are low (<5%), arguing against a history of extended isolation. Instead, these values indicate that varieties have either become isolated only very recently or that gene flow still occurs among them. High within-population genetic variance paired with lower among-population variance implies that reproductive barriers are low between populations (i.e. that migration is not uncommon) and favors the later explanation.

Sequence and microsatellite data conflict regarding a northern disjunction

Previous hypotheses regarding the recolonization of northern regions of the Cascades predict a genetic disjunction between northern and southern populations. This disjunction is predicted to be manifested in two ways: 1) Populations in the north will have lower overall genetic diversity, and 2) The alleles found in northern populations will be different from those found in the south (or at a very low frequency in the south). This

is hypothesized to occur from either a northern refuge population, which diverges from the south via a genetic bottleneck before recolonizing previously uninhabitable regions in the north; or from a series of founder events decreasing diversity along the leading edge of a migration front arising from the south (Soltis et al. 1997).

Evidence for a north-south disjunction is not shown in the LEGCYC1A-TCP data. Mt. Rainier and the Olympic mountains have among the highest levels of diversity at the LEGCYC1A-TCP locus (Table 5; Mt. Baker and Scatter Creek were not measured for LEGCYC1A-TCP diversity). Many of the LEGCYC1A-TCP genotypes found in northern populations are identical to or very close to those found in southern populations. Clades that contain members of northern populations also contain southern individuals from across the southern range of *L. lepidus*, refuting the prediction that northern populations contain only a small sample of southern genes (Figs. 4, 6).

Microsatellite data, on the other hand, distinctly exhibit a northern disjunction. The three northernmost populations (Mt. Rainier, the Olympic Mtns., and Mt. Baker) have significantly lower levels of diversity than southern populations (Table 11). The pairwise F_{ST} value between this northern grouping and the south is much higher than any other grouping tested (0.154), and north/south pooling in this manner accounts for a moderately high percentage of variation between groups (12.33%). Structure analysis strikingly confirms this disjunction, with these northern populations being almost exclusively represented by cluster B, and cluster B being found almost solely within these populations (Fig. 9).

What type of scenario can account for the signal of a northern disjunction to be present in microsatellite data but not sequence data? A scenario that invokes only

processes that involve allele frequency change (drift, founder effects, or bottlenecks) cannot account for this pattern, as all loci of the same ploidy should be subject to these effects equally (reduced diversity, in this case). However, microsatellite loci are expected to have much higher mutation rates than coding loci. A process that includes mutation could account for this pattern, particularly if the time period were long enough to allow microsatellite mutation, but short enough that coding mutations were rare.

A hypothesis involving a long leading edge, with multiple migration steps northward, is unlikely to create the observed pattern. The series of migration events would act mainly as a series of founder effects, reducing the diversity of northern populations at both the coding and microsatellite loci. The north-south recolonization hypothesis, however, involves effects produced via a bottleneck, but also involves separation of a northern refuge away from southern populations, allowing novel mutations to develop and spread within the northern refuge. *Lupinus lepidus* may have been isolated in a northern refuge during the last glacial cycle for a long enough time to develop novel microsatellite alleles, but no or very few novel coding alleles. Several such microsatellite alleles are found within cluster B.

It is unclear where a potential refuge for northern populations may have been located. Mt. Baker was well within the Cordilleran ice sheet, and Mt. Rainier, Blue Mtn. (of the Olympic Mtns.), and Glacier Peak were near its edge and subjected to alpine glaciation (Waitt and Thorson 1983). However, var. *lobbii* may have been able to tolerate these or nearby conditions (it currently grows at higher elevations than some glaciers on Mt. Baker). Other potential refuges include other areas of the Olympic Mtns., areas on or among Mts. Hood, Adams, and St. Helens, or lower elevation areas of the Puget Trough

or Columbia River Gorge. Current populations at Scatter Creek, Mt. St. Helens, and Mt. Hood are partially represented by cluster B, and could have provided material to seed the far northern populations. In particular, there are areas between Mt. Hood and Mts. Adams and St. Helens that were free from glaciation, and the Columbia River Gorge contains several plant species with otherwise boreal distributions (Porter et al. 1983, Detling 1966). (I will avoid the semantic issue of whether a refuge on the southern border of the northern disjunction qualifies as a northern refuge or an extension of the southern refuge.)

Comparison of sequence and microsatellite data can inform biogeographic hypotheses of the southern populations as well. Although most of the genetic variation within *L. lepidus* is found within-populations, among population variation accounts for a significant portion (20.6% and 41.5% in microsatellite and LEGCYC1A-TCP data, respectively). Much of this among population variation (13.8%) can be explained by the spatial distance between populations. Spatial structuring is not likely to be found in a panmictic population, and the presence of spatial structuring can indicate that, although gene exchange between populations may be high, alleles are more likely to migrate to neighboring populations than to distant areas.

Maintenance of high sequence diversity in northern groups and shared genotypes between northern and southern populations can be explained by a hypothesis where colonization of the Oregon and Washington regions arose from the south but gene flow between populations remained high. This would effectively move a larger portion of southern genes to the north, and agrees with the analyses of genetic structure indicating high within-population variation paired with spatial effects on population differentiation. A rapid southern expansion is also implied by the distribution of F_{ST} values between

populations (Fig. 8). The unimodal shape of this distribution is predicted from a population that has undergone either a recent demographic expansion or a spatial expansion paired with high levels of gene flow (Excoffier and Lischer 2010, Slatkin and Hudson 1991).

These results are compatible with patterns suggested by the Structure analysis wherein colonization from the south took place independently (or with little mixing) along either side of the Cascade Mountains. Populations of the southern Cascades and west became dominated by alleles representing cluster D, while populations of the north-central Cascades and east contain alleles representing clusters A+C. It is possible (perhaps likely) that this colonization took place well before the last glacial period. Under this scenario the disjunct northern populations represent descendents of a later split from the A+C cluster that became isolated during periods of heavy glaciation.

***Lupinus lepidus* varieties are undergoing rapid ecological divergence.**

The genetic similarities among the varieties of *Lupinus lepidus*, paired with the strong morphological differences among them, imply that selective pressures within local habitats are much greater than the homogenizing effects of migration. Unfortunately, formal common garden and transplant studies have not been performed for *L. lepidus* varieties. However, the strong morphological differences among varieties in seemingly similar habitats (e.g. varieties *ashlandensis*, *lobbii,* and *sellulus* in alpine areas within 300 km of each other; varieties *aridus*, *cusickii,* and *utahensis* within the arid Great Basin) imply that morphological differences are not merely due to phenotypic plasticity.

Additionally, plants of variety *lobbii* grown from seed under greenhouse conditions maintain their alpine habit (John Bishop, personal communication).

Observations of morphological divergence with little genetic divergence caused by high migration may cause difficulties similar to those faced by theories of sympatric speciation: namely, a reliance on mechanisms involving disruptive selection (e.g. assortative mating, frequency-dependent selection; Diekmann et al. 2004). However, it should be pointed out that populations of *L. lepidus* are not at all sympatric, being separated by tens to hundreds of kilometers, and distinct members of separate varieties have not been found together in a single population. While gene flow may be possible between populations, separate sites are still subject to differing environments and possibly very different selection regimes.

The varieties of *L. lepidus* can be viewed as divergent ecotypes, in which morphological differences may correlate with habitat differences, coupled with few barriers to reproduction between taxa (Clausen et al. 1939). Indeed, one of the organisms with which Clausen would later work, *Achillea millefolium*, has recently been subject to modern genetic analyses (Ramsey et al. 2008). The similarities between the results presented here and those found within *Achillea* are striking: both studies reconstructed sequence-based phylogenies, finding a large within-species polytomy; both studies included rapidly evolving loci (AFLPs in *Achillea*), finding unimodal distributions of pairwise differences; and in both studies analyses of population structure found most genetic variation within populations, a moderate amount among populations, and a very small amount among varieties or groups of populations (Table 5 in Ramsey et al. 2008).

Ramsey et al. hypothesized that the rapid adaptive divergence in *Achillea millefolium* was caused by a competitive release from congenerics in Europe and Asia, coupled with the highly heterogeneous environments of western North America. *L. lepidus* probably did not experience a competitive release since western North America is a center of diversity for *Lupinus*. However, its ability to inhabit disturbed areas and areas with poor soils (because it can fix nitrogen via *Rhizobium* bacteria) may have provided an advantage over ecological competitors, and may have aided the diversification of western perennial lupines as a whole. Meanwhile, it's possible its caespitose habit provided an unknown advantage relative to other *Lupinus*.

Taxonomic and conservation implications

Genetic differentiation between varieties of *L. lepidus* is very low, providing convincing evidence that they are not reproductively isolated. Despite this, striking morphological differences suggest that selective forces are strong enough to maintain these varieties over time. Given this, it is possible that reproductive barriers will eventually be erected between varieties. For example, in variety *cusickii* it is thought that properly timed rainfall is crucial to seed production (Massatti et al. 2009). Future phenological adaptation to differing rainfall conditions could be a driver of reproductive isolation among varieties. Clausen et al. (1939) accepted that ecotypes should be named as separate taxonomic entities and similarly it is recommended here that varieties of *Lupinus lepidus* retain recognition at the varietal rank.

While some varieties of *L. lepidus* are widespread with large populations, some are endemic to very narrow areas. Within Oregon, variety *ashlandensis* is restricted to the

summit of Mt. Ashland, and variety *cusickii* is restricted to a few small populations in Baker County. Variety *lepidus*, which was historically present throughout much of the Willamette Valley, has been extirpated from much of its range and is probably now only present in Oregon at a handful of sites. Continued study and monitoring of these varieties is necessary to preserve the morphological diversity and ecological adaptations that have developed among them. Because these morphological differences likely have a strong genetic component, if population reintroductions or supplementation become necessary, it is strongly recommended that the introduced plants come from convarietal populations.

Conclusion

Lupinus lepidus has been shown to consist of varieties best described as ecotypes. Low genetic differentiation among varieties paired with high within population variation indicates that gene transfer between populations can be high, and that reproductive barriers between varieties either do not exist or have formed so recently as to not be detectable.

Differing signals of a north-south disjunction between coding and microsatellite data imply that such a disjunction is real, and was created under a scenario where allele frequencies were modified by the spread of novel mutations rather than solely via drift, bottlenecks, or founder events. This favors the presence of a northern refuge population isolated for some time from southern populations. Overall high gene diversity across regions, coupled with shared alleles between northern and southern regions, supports a colonization of the Oregon and Washington areas characterized by a rapid expansion from the south capable of maintaining much of the southern diversity in northern

populations. Microsatellite data suggest this occurred separately on either side of the Cascades.

Future studies should quantify the environmental and genetic influences on varietal morphology. This system would also benefit from a characterization of mechanisms of and barriers to migration. Further work on biogeographic patterns within the Pacific Northwest and along the Cascade Mountains may elucidate circumstances where a recolonization or expansion event is more or less likely to create a pattern of genetic disjunction, and further work on species exhibiting ecotypes may quantify the competing influences of migration maintaining homogeneity and selection driving diversification.

References

Aïnouche A, Bayer RJ, Misset M. 2004. Molecular phylogeny, diversification and character evolution in *Lupinus* (Fabaceae) with special attention to mediterranean and african lupines. Plant Syst Evol 246: 211-22.

Avise JC. 2004. *Molecular Markers, Natural History, and Evolution*. 2nd edition, Sunderland, MA: Sinauer Associates, Inc.

Barneby RC. 1989. *Fabales*. Vol. 3, Part B of *Intermountain Flora: Vascular Plants of the Intermountain West, U.S.A.*, A Cronquist, AH Holmgren, NH Holmgren, JL Revel, PK Holmgren. Bronx, New York: The New York Botanical Garden.

Bishop JG, RJ Dyer. 1996. The formation and loss of genetic variance among newly-founded lupine populations on Mount St. Helens. Unpublished manuscript. http://www.vancouver.wsu.edu/fac/bishop/Assets/Bishop%20and%20Dyer%20lupin%20population%20genetics%20.pdf.

Broich SL, LA Morrison. 1995. The taxonomic status of *Lupinus cusickii* (Fabaceae). Madroño 42: 490-500.

Brunsfeld S and J Sullivan. 2005. A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conserv Genet 6: 895-904.

Cartens BC, SJ Brunsfeld, JR Demboski, JM Good, J Sullivan. 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. Evolution 59: 1639-52.

Clausen J, DD Keck, WM Hiesey. 1939. The concept of species based on experiment. Amer J Bot 26: 103-106.

Clement M, D Posada, KA Crandall. 2000. TCS: A computer program to estimate gene genealogies. Mol Ecol 9: 1657-60.

Citerne HL. 2005. A primer set for specific amplification of two CYCLOIDEA-like genes in the Genistoid clade of Leguminosae subfam. Papilionoideae. Edinburgh J Bot 62: 119-126.

Citerne HL, D Luo, T Pennington, E Coen, and QCB Cronk. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. Plant Physiol 131: 1042-53.

Corriveau JL, AW Coleman. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Amer J Bot 75: 1443-58.

Cox BJ. 1973. Protein relationships among perennial caespitose lupines. Bull Torrey Bot Club 100: 153-158.

Cubas P, N Lauter, J Doebley, E Coen. 1999. The TCP domain: a motif found in proteins regulating plant growth and development. Plant J 18: 215-222.

Degnan JH, NA Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet 2(5):e68.

Detling LE. 1966. The flora of the Columbia River Gorge. Northwest Science 40: 133-137.

Diekmann U, M Doebeli, JAJ Metz, D Tautz, eds. 2004. *Adaptive Speciation*. New York: Cambridge University Press.

Drummond CS. 2008. Diversification of *Lupinus* (Leguminosae) in the western New World: Derived evolution of perennial life history and colonization of montane habitats. Mol Phylogenet Evol 48: 408-421.

Drummond CS, MB Hamilton. 2005. Isolation and characterization of nuclear microsatellite loci in *Lupinus* group *Microcarpi* (Leguminosae) Mol Ecol Notes 5: 510–513.

Evanno G, S Regnaut, J Goudet. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. Mol Ecol 14: 2611-2620.

Excoffier L, HEL Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resources 10: 564-567.

Falush D, M Stephens, JK Pritchard. 2003. Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Falush D, M Stephens, JK Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7: 574-578.

Guindon S, O Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

Hardy OJ, N Charbonnel, H Fréville, M Heuertz. 2003. Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. Genetics 163: 1467-1482.

Hardy OJ, X Vekemans. 2002. SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes 2: 618-620.

Hartl DL, AG Clark. 1997. *Principles of Population Genetics*. 3rd edition, Sunderland, MA: Sinauer Associates, Inc.

Hickman JC, ed. 1993. *The Jepson manual: Higher plants of California*. Berkeley, CA: University of California Press.

Hitchcock CL, A Cronquist. 1996. *Flora of the Pacific Northwest*. Seattle, WA: University of Washington Press.

Huang D, EA Friar. 2009. Perennial *Lupinus* of western North America: Challenges in defining species boundaries in a recent radiation. Paper presented at Botany 2009, July 25-29, in Snowbird, UT.

Huang H, LL Knowles. 2009. What is the danger of the anomaly zone for empirical phylogenetics? Sys Biol 58: 527-536.

Huelsenbeck JP, F Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

Hughes C, R Eastwood. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. Proc Natl Acad Sci USA 103: 10334-10339.

Isely, D. 1998. *Native and Naturalized Leguminosae (Fabaceae) of the United States (exclusive of Alaska and Hawaii).* Provo, UT: Brigham Young University Monte L. Bean Life Science Museum.

Jakobsson M, NA Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Janzen FJ, JG Krenz, TS Haselkorn, ED Brodie Jr., ED Brodie III. 2002. Molecular phylogeography of common garter snakes (*Thamnophis sirtalis*) in western North America: implications for regional historical forces. Mol Ecol 11: 1739-1751.

Kuchta SR, A Tan. 2005. Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. Mol Ecol 14: 225-44.

Liston A, M Parker-Defeniks, JV Syring, A Willyard, R Cronn. 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: A case study in *Pinus lambertiana*. Mol Ecol 16: 3926-37.

Maddison DR, WP Maddison. 2005 MacClade: analysis of phylogeny and character evolution, version 4.08. Sunderland, MA: Sinauer Associates, Inc.

Maddison WP. 1997. Gene trees in species trees. Syst. Biol. 46: 523-536.

Massatti RT, AS Thorpe, TN Kaye. 2009. *Lupinus lepidus* var. *cusickii* monitoring in Denny Flat, Baker County, Oregon. Institute for Applied Ecology, Corvallis, Oregon, and USDI Bureau of Land Management, Vale District. iv + 16 pp.

Miller MP, MR Bellinger, ED Forsman, SM Haig. 2006. Effects of historical climate change, habitat connectivity, and vicariance on genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*) in the Pacific Northwestern United States. Mol Ecol 15: 145-59.

Miller MA, MT Holder, R Vos, PE Midford, T Liebowitz, L Chan, P Hoover, T Warnow. The CIPRES Portals. CIPRES. http://www.phylo.org/portal2/.

Morris VRF. 2009. Nurture over nature: Summer germinating *Lupinus nanus* are a result of anthropogenic germination cues and are not an independently evolving population. PhD Diss., University of California at Berkeley. Berkeley, CA: University Press.

Morrison DA. 2007. Increasing the efficiency of searches for the maximum likelihood tree in a phylogenetic analysis of up to 150 nucleotide sequences. Syst Biol 56: 988-1010.

Nielson M, K Lohman, J Sullivan. 2001. Phylogeography of the tailed frog (*Ascaphus truei*): Implications for the biogeography of the Pacific Northwest. Evolution 55: 147-60.

Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15: 407-414.

O'Quinn RL, M Fishbein. 2009. Isolation, characterization, and cross-species amplification of polymorphic microsatellite loci in *Asclepias* (Apocynaceae). Conserv Genet 10: 1437-1440.

Oregon Flora Project. Oregon Plant Atlas. http://www.oregonflora.org/atlas.php.

Peakall R, S Gilmore, W Keys, M Morgante, A Rafalski. 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. Mol Biol Evol 15: 1275-1287.

Porter SC, KL Pierce, TD Hamilton. 1983. Late Wisconsin Mountain Glaciation in the Western United States. In *The Late Pleistocene*, ed. SC Porter, 71-111. Vol. 1 of *Late-Quaternary Environments of the United States*. ed. HE Wright Jr. Minneapolis, MN: University of Minnesota Press.

Posada D. 2008. jModeltest: Phylogenetic model averaging. Mol Bio Evol 25: 1253-56.

Posada D, KA Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Pritchard JK, M Stephens, P Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Ramsey J, A Robertson, B Husband. 2008. Rapid adaptive divergence in New World *Achillea*, and autopolyploid complex of ecological races. Evolution 62-3: 639-653.

Raymond M, F Rousset. 1995. An exact test for population differentiation. Evolution 49: 1280-83.

Ree RH, HL Citerne, M Lavin, QCB Cronk. 2004. Heterogeneous selection on *LEGCYC* paralogs in relation to flower morphology and phylogeny of *Lupinus* (Leguminosae). Mol Biol Evol 21: 321-331.

Ronquist F, JP Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnol 18: 233-234.

Shaw J, EB Lickey, JT Beck, SB Farmer, W Liu, J Miller, KC Siripun, CT Winder, EE Schilling, RL Small. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Amer J Bot 92: 142-166.

Shaw J, EB Lickey, EE Schilling, RL Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. Amer J Bot 94: 275-288.

Slatkin M, RR Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129: 555-562.

Smith CP 1946. *Species Lupinorum*. sig. 32. Paper 57.

Soltis DE, MA Gitzendanner, DD Strenge, PS Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst Evol 206: 353-73.

Stamatakis A. 2006a. RAxML-VI-HPC: maximum likelihood-based phylogenetic analysis with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

Stamataks A. 2006b. Phylogenetic models of rate heterogeneity: A high performance computing perspective. In Proc. IPDPS 2006, Rhodos, Greece.

Stamatakis A, P Hoover, J Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web Servers. Syst Biol 57: 758-771.

Steele CA, A Storfer. 2006. Coalescent-based hypothesis testing supports multiple pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (*Dicamptodon tenebrosus*). Mol Ecol 15: 2477-87.

Strenge DD. 1994. The intraspecific phylogeography of *Polystichum munitum* and *Alnus rubra*. Masters thesis. Washington State University.

Swofford D. 2001. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods. Version 4.0b10. Sunderland, MA: Sinauer Associates.

Tamura K, M Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512-526.

Toonen RJ, S Hughes. 2001. Increased throughput for fragment analysis on ABI Prism 377 automated sequencer using a membrane comb and STRand software. Biotechniques 31: 1320-24.

Waitt RB Jr., RM Thorson. 1983. The Cordilleran Ice Sheet in Washington, Idaho, and Montana. In *The Late Pleistocene,* ed. SC Porter, 53-70. Vol. 1 of *Late-Quaternary Environments of the United States*. ed. HE Wright Jr. Minneapolis, MN: University of Minnesota Press.

Wood DM, R del Moral. 1987. Mechanisms of early primary succession in subalpine habitats on Mt. St. Helens. Ecology 68: 780-790.

Wright S. 1931. Evolution in Mendelian Populations. Genetics 16: 97-159.