

SNPs in a Large Genomic Scaffold Are Strongly Associated with *Cr1^R*, Major Gene for Resistance to White Pine Blister Rust in Range-Wide Samples of Sugar Pine (*Pinus lambertiana*)

Jessica W. Wright,^{1,†} Kristian A. Stevens,² Paul Hodgskiss,¹ and Charles H. Langley³

¹ U.S. Department of Agriculture Forest Service, Pacific Southwest Research Station, Davis, CA 95618

² Departments of Computer Science and Evolution and Ecology, University of California Davis, Davis, CA 95616

³ Department of Evolution and Ecology, University of California Davis, Davis, CA 95616

Abstract

Sugar pine, *Pinus lambertiana* Douglas, is a keystone species of montane forests from Baja California to southern Oregon. Like other North American white pines, populations of sugar pine have been greatly reduced by the disease white pine blister rust (WPBR) caused by a fungal pathogen, *Cronartium ribicola*, that was introduced into North America early in the twentieth century. Major gene resistance to WPBR segregating in natural populations has been documented in sugar pine. Indeed, the dominant resistance gene in this species, *Cr1*, was genetically mapped, although not precisely. Genomic single nucleotide polymorphisms (SNPs) placed in a large scaffold were reported to be associated with the allele for this major gene resistance (*Cr1^R*). Forest restoration efforts often include sugar pine seed derived from the rare resistant individuals (typically *Cr1^R/Cr1^r*) identified through an expensive 2-year phenotypic testing program. To validate and geographically characterize the variation in this association and

investigate its potential to expedite genetic improvement in forest restoration, we developed a simple PCR-based, diploid genotyping of DNA from needle tissue. By applying this to range-wide samples of susceptible and resistant (*Cr1^R*) trees, we show that the SNPs exhibit a strong, though not complete, association with *Cr1^R*. Paralleling earlier studies of the geographic distribution of *Cr1^R* and the inferred demographic history of sugar pine, the resistance-associated SNPs are marginally more common in southern populations, as is the frequency of *Cr1^R*. Although the strength of the association of the SNPs with *Cr1^R* and thus, their predictive value, also varies with geography, the potential value of this new tool in quickly and efficiently identifying candidate WPBR-resistant seed trees is clear.

Keywords: CAPS, *Cr1*, major gene resistance, sugar pine, white pine blister rust

Sugar pine, *Pinus lambertiana* Douglas, is an economically and ecologically important white pine species found in the forests of Western North America (Lanner 2007). It occurs naturally from Baja California, north into Oregon. In California, it grows primarily in the Sierra Nevada as well as the Coast and Cascade ranges. The distribution of two distinct chloroplast genome haplotypes suggests that populations in these two regions have a partially separated demographic history (Liston et al. 2007). An apparent contact zone in northeastern California is evident. Limited studies of nuclear genomic variation have found agreement with the conclusions of Liston et al. (2007), and show genetic differentiation between northern and southern populations (Vangestel et al. 2016).

In the last century, white pine blister rust (WPBR) was introduced into North America, devastating populations of sugar pine and other white pines. WPBR is a tree disease caused by the fungal pathogen *Cronartium ribicola* (J.C. Fisch.) that infects many members of the white pine subgenus, often leading to mortality. It causes cankers in the stems of sugar pines that eventually result in tree death, either through girdling or breakage of the main stem. WPBR was first observed in western North America on Vancouver Island, British Columbia, Canada (Kinloch 2003). The history and epidemiology of

the pathogen has been extensively reviewed elsewhere (Geils et al. 2010; King et al. 2010; Kinloch 2003). The pathogen has a complex life cycle, with an alternate host, a *Ribes* spp., required for transmission to sugar pine. Kinloch et al. (1970) discovered major gene resistance to the pathogen, *Cr1^R*. The variant is a dominant allele, with both homozygotes and heterozygotes equally resistant to the pathogen (King et al. 2010). Resistant seedlings infected with the pathogen show a characteristic hypersensitive response, but the disease does not progress further (King et al. 2010). Kinloch (1992) showed *Cr1^R* is always rare, ranging from 0.08 in the southern Sierras populations to essentially 0.0 in the North Coast portion of its California range (Kinloch 1992).

Forest restoration projects often involve planting seedlings. A tree improvement program in California that started in the 1950s, continuing today, involves the screening of potentially resistant trees from across California for major gene resistance to WPBR (King et al. 2010). Forest managers in the U.S. Department of Agriculture Forest Service (USDA-FS) Region 5 (California) often plant seedlings derived from such identified *Cr1^R* parent trees, with the expectation that they will increase the success of planting operations in impacted areas. Nearly all sugar pine seedlings ordered by the USDA-FS from the Region 5 Nursery are requested to be from resistant *Cr1^R* parent trees (Gary Cline, personal communication).

To determine the presence of the *Cr1^R* allele, seedlings of a maternal tree are challenged with exposure to *C. ribicola* basidiospores, which will cause a hypersensitive-like reaction on the resistant genotypes, and cause death in the susceptible ones. Given the low frequency of *Cr1^R* in natural populations, most resistant trees are heterozygous, and most pollen carry the wild-type (nonresistant) *Cr1^r* allele. This results in an expectation of 50% of wind-pollinated seed collected from a resistant seed tree having the *Cr1^R* allele. Therefore, screenings must include enough tested seedlings to ensure that *Cr1^R/Cr1^r* heterozygous trees are efficiently and correctly identified. Generally, the screening by the USDA-FS in Region 5, California, involves 100 seedlings per seed tree to determine if the tree carries the *Cr1^R* allele (Samman and Kitzmiller 1996). This process takes about 2 years

†Corresponding author: J. W. Wright; jessica.w.wright@usda.gov

Funding: This work was supported by the U.S. Department of Agriculture Forest Service.

*The e-Xtra logo stands for “electronic extra” and indicates that there are supplementary materials published online.

The author(s) declare no conflict of interest.

Accepted for publication 3 December 2021.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2022.

(seed collection, processing, germination, inoculation, and then finally phenotyping) and a great deal of investment in staff time and space resources. Fewer than 10% of screened seed trees are expected to be resistant, given the results of Kinloch (1992). The replacement cost of a tree carrying *CrI^R* was estimated by the USDA FS R5, and ranges from \$2,215.24 to \$23,550 depending on the seed zone (USDA Forest Service 2018). The potential benefit of a more efficient and effective molecular genotyping approach is obvious.

The 2016 sequencing of the complete genome of *P. lambertiana* (Stevens et al. 2016) brought new resources for marker development. An apparently biallelic locus, *CrI* had been mapped in several *P. lambertiana* families (Devey et al. 1995; Harkins et al. 1998; Jermstad et al. 2011). Stevens et al. (2016) used the draft whole genome shotgun assembly to anchor cloned randomly amplified polymorphic DNA sequences (Jermstad et al. 2011) in the draft assembly and identify five physically proximal genes. They annotated single nucleotide polymorphisms (SNPs) by sequencing a PCR amplicon designed in the *CrI* genomic region. Further exploration of sequence variation in the *CrI* genomic region identified five SNPs with nearly complete association with the *CrI^R* allele in a small sample of trees. The genomic reference sequence and the *CrI^R*-associated variation also provided for the development of cleaved amplified polymorphic sequence (CAPS) assays (Konieczny and Ausubel 1993) that are suitable for the diploid genotyping of many samples in this huge genome. The CAPS marker applied here is based on the detection of one of these five SNPs, all contained in a single PCR amplicon, the *CrI* Associated Marker 1 (*CrIAMI*).

Given the ecological and economic importance of *P. lambertiana*, the increased rate of wildland fires in California and the time- and cost-intensive methods available to screen seedlings for the *CrI^R* allele, molecular genotyping of SNPs associated with *CrI^R* could greatly expedite the identification of candidate *CrI^R* allele carriers, and hence make reforestation with WPBR-resistant seedling more efficient. Direct genotyping of seedlings could further increase the impact of sugar pine WPBR-resistant reforestation.

In this article we:

1. demonstrate that the genomic segment containing the *CrI^R*-associated SNPs can be amplified in diploid needle tissue and genotyped using the CAPS marker approach;
2. confirm that the *CrI^R*-associated SNPs identified in Stevens et al. (2016) are indeed in linkage disequilibrium with *CrI* throughout the species' range;
3. describe geographic variation in the *CrIAMI* marker;
4. investigate variation in the association of *CrIAMI* marker with *CrI^R*; and
5. discuss the potential uses of this new tool is the selection of seed sources and seedlings for use in reforestation.

Materials and Methods

Fresh needles were collected in the summer and fall of 2016 from approximately 100 tested resistant trees (the resistant sample), and a comparable number of observed susceptible trees (the susceptible sample; see Supplementary Tables S1 and S2). The sampled trees are derived from seeds sourced from across the species range in California and are now growing in two common gardens. The resistant trees were collected from the USDA-FS Foresthill Seed Orchard, outside of the town of Foresthill, CA (39.084°, -120.731°). The trees there are grafted using scions from tested trees that have been extensively genotyped and known to be resistant. They are mostly *CrI^R/CrI^I* heterozygotes, but some homozygous trees are present as well. The seed orchard contains trees sourced from across the National Forests of California – most of the geographic range of the species. The seed from this orchard will be used in postfire reforestation efforts on National Forest lands in California.

Susceptible material was collected from the Harrel provenance test on the Eldorado National Forest (38.658°, -120.466°). Planted by the USDA-FS Pacific Southwest Research Station, Institute of Forest Genetics in 1992, it contains 124 sources of trees from across

the entire species range (planting records on file at the Institute of Forest Genetics, Placerville, CA). One individual from each of 103 California sources was sampled. Sources are bulk collections from seed zones and elevation bands (Supplementary Table S1). The Harrel provenance test has a greater range of geographic variability than the Foresthill seed orchard and included sources from Baja California and the Transverse range (sources from Oregon were not tested). Susceptibility to WPBR was determined visually by the presence of a bole canker or other symptom on the living tree. The assumption was made that only a susceptible *CrI^I/CrI^I* tree would be infected and show evidence of the pathogen (Geils et al. 2010). As expected from previous reports (J.W.W., personal observation), infected trees were very common at the test; indeed, many trees had already died (of the 16,000 trees planted, approximately 6,000 were still alive in 2010).

The fresh needle tissue was cut into 100-mg aliquots, placed into labeled 1.5-ml microcentrifuge tubes and stored in a -80°C freezer until DNA extraction.

DNA extraction was conducted using the Macherey–Nagel Nucleo-spin Plant II DNA extraction kit, following the manufacturer instructions. The 100 mg of needle tissue was ground in liquid nitrogen, and lysed using the PL1 solution, a cetrimonium bromide-based extraction buffer. To assess the quality of the eluted DNA, 1 µl of DNA was mixed with 1 µl of loading dye and 4 µl of water and loaded into 1.0% FBM (Faster Better Media, Cockeysville, MD) agarose gel in FBM LB buffer, with the UV fluorescent DNA stain, GelRed (2 µl of stain to 50 ml of agarose gel). Electrophoresis of the DNA proceeded at 80 V for 5 min, then at 100 V for 40 min.

Generating the *CrIAMI* marker requires a nested PCR amplification of the BC_432 region (Jermstad et al. 2011), then digestion with *MseI*. Primer sequences are given in Supplementary Table S3. The first PCR reaction was in 25 µl and was as follows: New England Biolabs (NEB, Ipswich, MA) Standard Taq buffer, 1.5 mM of MgCl₂, 200 µM of dNTPs (NEB), 0.2 µM of cr11B_F1 primer, 0.2 µM of cr11B_R1primer (Stevens et al. 2016; Supplementary Table S3), <1,000 ng of template DNA, and 1.25 U of Taq DNA polymerase (NEB). Thermocycling of the PCR reaction was 94°C for 5 min, then 30 cycles of 30 s at 94°C, 30 s at 57°C, and 60 s at 68°C. It was followed by 68°C for 5 min, then held at 4°C.

The second PCR reaction was in 50 µl. The reaction was the same as the first PCR, except adjusted to a larger volume and the template consisted of 1 µl of the product of the first PCR reaction and 0.2 µM of the nested primers cr11B_nF1b and cr11B_nR1b (Stevens et al. 2016; Supplementary Table S3). Thermocycling of the 2nd PCR reaction is 94°C for 5 min, then 30 cycles of 30 s at 94°C, 30 s at 56°C, and 60 s at 68°C. This was followed by 68°C for 5 min, then held at 4°C.

The products of the second PCR were then digested with NEB *MseI* enzyme. Each sample was digested for 20 min at 37°C, followed by 20 min at 65°C. The 50-µl reaction was as follows: 34 µl of nuclease free water, 5 µl of NEB *MseI* buffer solution, 10 U of NEB *MseI* enzyme, and 10 µl of the second PCR reaction.

The DNA products of the second PCR, and corresponding *MseI* enzyme digest, were visualized in a 1.75% FBM agarose gel with FBM LB buffer and GelRed UV fluorescent DNA stain after 80 V for 5 min, then 100 V for 40 min. Gels were scored manually based on the expected banding pattern. The extent of amplification of heterozygous individuals was not equal between haplotypes, and care was needed to obtain the correct score. Samples with unclear results were run a second time.

Samples were selected for sequencing based on *MseI* digest fragment patterns, with the goal to sequence two individuals from each banding pattern. Undigested DNA from the second PCR amplification was purified according to manufacturer's protocols of the AMPure XP magnetic beads system (Beckman Coulter, Pasadena, CA). DNA sequencing of the undigested second PCR product was completed by the UC Davis DNA sequencing facility (<https://dnaseq.ucdavis.edu/>). Sequences were aligned using the program BioEdit (<https://bioedit.software.informer.com/>; Hall 1999).

The statistical fits of genotype frequencies to Hardy–Weinberg (H–W) proportions were inferred based on exact calculations of probabilities (Graffelman and Weir 2016) using the HWTriExact function in the R package HardyWeinberg (R Core Team 2021; <https://www.r-project.org/>; Graffelman et al. 2021). Statistical inference of heterogeneity in genotype frequencies were based on Fisher’s exact test implemented in the `fisher.test` in the R package `stats` (R Core Team 2021).

Results

Development of CAPS genotyping assay of *CrIAM1*. The target genomic region for the *CrIAM1* assay was successfully PCR-amplified in both resistant and susceptible samples. Based on the DNA sequence data from a representative sample of noncloned PCR products from diploid leaf tissue (Supplementary Fig. S1), we found six *MseI* cut sites (highlighted in yellow on Supplementary Fig. S1) that resulted in three banding patterns we could reliably score both as homozygotes and heterozygotes on an agarose gel matrix (Fig. 1). Several bands were too close to distinguish, but discrimination of those bands was not necessary to assign each individual’s genotype. *MseI* cut sites 1, 2, and 6 were monomorphic, and alone made up the homozygous *A/A* banding pattern (see Fig. 1). Cut sites 3 and 4 were rare, creating the second heterozygous banding pattern (scored as a *B/A*) that was also associated with susceptibility, presumed *CrI^R/CrI^r*. One of the five SNPs associated with *CrI^R* in Stevens et al. (2016) is in the middle of cut site 5, and hence its association with *CrI^R*. The remaining four *CrI^R*-associated SNPs exhibit the expected pattern of polymorphism across the sequenced resistant and susceptible trees, highlighted in pink in Supplementary Figure S1, but were not associated with an *MseI* cut site. The diploid genotypes at *CrIAM1* of all samples could be inferred from the CAPS assay results and are shown in Supplementary Tables S1 and S2.

Confirmation of association of *CrI^R* with SNPs in large genomic region in a range-wide sample from needles. While both the susceptible and resistant samples were collected from across the range of sugar pine, geographic differentiation in allele frequencies could distort the genotype frequencies from H–W proportions (Wahlund 1928). However, the distribution of genotypes in the range-wide sample of putatively susceptible individuals fits H–W expectations. Table 1 shows the counts of *CrIAM1* genotypes, among randomly selected individuals from throughout the range (presumed to be the common susceptible genotype *CrI^r/CrI^r*), and these genotypic frequencies do meet H–W expectations. On the other hand, the observed *CrIAM1* genotypes in a range-wide sample of trees identified by progeny testing to segregate the major gene for WPBR resistance, thus the *CrI^R/CrI^r* genotype do not meet the H–W expectations (Table 1). The full data are in Supplementary Tables S1 and S2, respectively. The frequency of *CrIAM1* genotypes carrying at least one *C* haplotype among selected resistant trees is 0.82 (76/93), while among randomly

sampled (presumed *CrI^r/CrI^r*) susceptible trees the frequency is only 0.15 (15/103). The probability of this large or larger difference occurring by chance is 9.5×10^{-9} . Thus, we conclude that the SNP at cut site 5 is strongly associated with the major gene resistance to WPBR, *CrI* in populations of *P. lambertiana*.

Geographic associations of *CrIAM1*. Given the demonstrated changes in frequency of *CrI^R* across the species range (Kinloch 1992), the results of Liston et al. (2007) demonstrating northern and southern chloroplast haplotypes and the geographic differentiation analysis in 2020 of a collection of SNPs (Weiss et al. 2020) in sugar pine, we explored the possibility that the *CrIAM1^C* haplotype had a similar change in frequency across the range. Figure 2 shows the observed numbers of individuals with and without a *CrIAM1^C* haplotype among putatively *CrI^R* individuals sampled from the three geographic regions (northwestern CA) Coast Range, (eastern CA) Sierra Nevada, and (northeastern CA) Cascades, as defined by Liston et al. (2007). These regions are similar, but not identical, to the areas where the North and South chloroplast haplotypes are found, as well as the contact zone where they are both found. In addition, they correspond roughly to two of the three genetic clusters identified by Weiss et al. (2020). No statistically significant heterogeneity in the frequencies of individuals with at least one *CrIAM1^C* haplotype among the three regions is observed ($P < 0.07$). There is the suggestion that, like *CrI^R* itself, *CrIAM1^C* is more common in the southern (Sierra Nevada) portion of the range among individuals genotyped as *CrI^R/CrI^r*. Indeed, the contrast in frequencies between the southern and the pooled two northern regions has a $P < 0.02$. Supplementary Tables S4 and S5 show the complete results both by geographic zone (Supplementary Table S4) and chloroplast haplotype (Supplementary Table S5).

Given the geographic (latitudinal) differentiation in the frequency of *CrI^R* (Kinloch 1992), it was reasonable to look for association of *CrI^R* with altitude, because many ecological factors change systematically in both dimensions. Figure 3 shows that the frequency of the

Table 1. Fit of observed (*O*) diploid genotype frequencies to expected (*E*) Hardy–Weinberg proportions in a random sample of putatively susceptible, *CrI^r* trees and in a selected sample of resistant, *CrI^R* trees^a

Genotype	Random sample		Selected sample	
	<i>O</i>	<i>E</i>	<i>O</i>	<i>E</i>
<i>A/A</i>	75	76.0	16	24.8
<i>A/B</i>	12	12.0	0	4.7
<i>A/C</i>	15	12.9	64	41.8
<i>B/B</i>	1	0.5	1	0.2
<i>B/C</i>	0	1.0	7	3.9
<i>C/C</i>	0	0.6	5	17.6
Total	103	103.0 NS	93	93.0*

^a NS, not significant ($P < 0.45$); * $P < 9.5 \times 10^{-9}$.

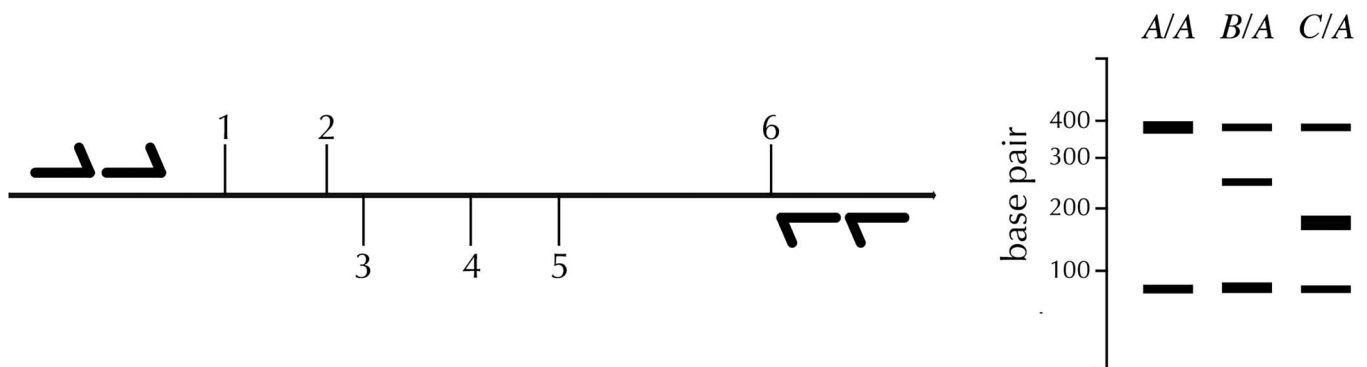


Fig. 1. Cleaved amplified polymorphic sequence assay of the *Cr1* Associated Marker 1, *CrIAM1*. On the left is a diagram of the 490-bp PCR target and the approximate flanking nested primers. The positions of the *MseI* restriction sites are also shown (top, monomorphic; below, polymorphic). On the right is a depiction of the distributions of fragment sizes for the three common genotypes comprised of *A* haplotype (restriction sites 1, 2, and 6): 81 and 359 bp; *B* haplotype (1, 2, and 6 plus 3 and 4): 81, 30, 86, and 243 bp; and *C* haplotype (1, 2, and 6 plus 5): 81, 187, and 172 bp.

Cr1AM1^C haplotype is consistent between trees at 6,000 to 7,000 ft and those either above or below, although the resistance-associated haplotype appears to decrease with increasing altitude. This result is similar to that for *Cr1^R* (Kinloch 1992; Supplementary Table S6).

Discussion

Since its introduction to North America at the beginning of the last century, WPBR has devastated populations of sugar pine and the other species of the *Strobos* subgenus endemic to the western hemisphere (Geils et al. 2010). The identification of inherited resistance in these species has held out hope for their survival and eventual recovery (King et al. 2010; Kinloch 2003). Reforestation efforts that leverage genetic information may expedite recovery, at least locally. WPBR major gene resistance in sugar pine is conferred by the *Cr1* locus. Complexly inherited (multifactor) resistance to WPBR is well documented in white pine (Kinloch 2003; Weiss et al. 2020). However, without a molecular marker associated with the locus, screenings for resistance are based on the phenotypic responses of seedlings to challenge by *C. ribicola* spores in a greenhouse.

The determination of the draft *P. lambertiana* genome sequence (Crepeau et al. 2017; Stevens et al. 2016) has facilitated the deployment of genome-wide SNP genotyping and the foundation of genomic selection for WPBR resistance and other traits (Weiss et al. 2020). Although this approach requires large resources and commitment, it promises steady improvement and shelter from emergence of new virulence.

The mapping and identification of the SNPs in high linkage disequilibrium with *Cr1^R* creates a substantial and more immediate

opportunity to change the impact of the disease in sugar pine. For the last several decades, reforestation of sugar pine has leveraged the ability to identify trees carrying the *Cr1^R* allele through greenhouse trials (Kinloch 2003; Samman and Kitzmiller 1996). Although finding these rare resistant trees has been costly and time-consuming, seedlings from such trees have been widely used in reforestation simply because the disease would otherwise effectively undermine the effort. The questions raised by the results of genetic mapping of *Cr1^R* (Jermstad et al. 2011) and the determination of a full genome, draft reference sequence (Crepeau et al. 2017; Stevens et al. 2016) is how robust and useful will this toehold prove to be? Can it be quickly adapted to greatly facilitate reforestation with disease-resistant seedlings? Can it open the door for research into a deeper understanding of the biology of WPBR which, in turn, will lead to more informed management? As a basis upon which to address these prospects, let us first discuss the results reported here:

1. Practically relevant to both goals is our demonstration that the genomic segment containing the *Cr1^R*-associated SNPs, *CrAM1*, can be reliably PCR-amplified from DNA isolated from readily available, field-collected diploid needle tissue. Furthermore, these amplicons support robust genotyping – in this case, a simple (but not scalable) CAPS assay.
2. Using this genotyping approach, we confirm that the *Cr1^R*-associated SNPs, *CrAM1*, identified in Stevens et al. (2016), are indeed in strong but not complete linkage disequilibrium with *Cr1^R*.
3. Importantly, we establish that despite the substantial geographic differentiation of the frequency of *Cr1^R* across the sugar pine range, the association with *CrAM1* remains high.

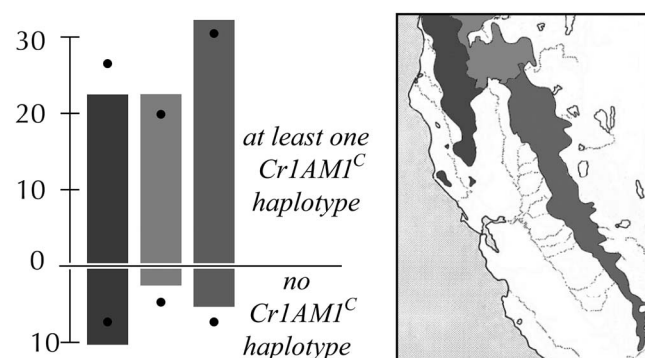


Fig. 2. Distribution of *Cr1* Associated Marker 1, *Cr1AM1* in samples of WPBR-resistant (presumed *Cr1^R/Cr1^I*) from the (northwestern California) Coast Range (dark gray), (eastern California) Sierra Nevada (medium gray), and (northeastern California) Cascades (light gray; see text and Supplementary Table S4). The y axis shows the number of sampled individuals in each category. Homogeneity among geographic regions in the frequencies of *Cr1AM1^C* cannot be rejected, $P < 0.07$; the dots are the expected values. The pattern suggests that (similar to *Cr1^R* itself) *Cr1AM1^C* is more common in the southern part of the range.

Confirmation of association of *Cr1^R* with SNPs in a large genomic region in a range-wide sample from needles. While results in Stevens et al. (2016) stand on their own, it is noteworthy that their linkage results used haploid material (DNA preparations of the actual megagametophytes of rescued seedlings) from the original mapping experiment (Jermstad et al. 2011). Thus, their mapping results confirm that the new SNPs in the large genomic scaffolds are indeed genetically linked to *Cr1^R* among the same progeny of that single maternal tree. Their independent detection of association between these SNPs and *Cr1^R* (Stevens et al. 2016) was statistically significant but based on a small sample of haploid individuals, neither representative of a single population nor of the species range. An independent rigorous validation of this initial observation of linkage disequilibrium (and thus linkage) is our primary goal. To effectively and independently confirm the observed linkage disequilibrium between SNPs in specific genomic scaffolds and *Cr1^R* (Stevens et al. 2016) required a reliable genotyping protocol in readily available tissue (e.g., needles) from large and representative samples of trees for which the *Cr1* genotype has already been inferred based on phenotype. Using genomic DNA extracted from the needles of 196 samples as

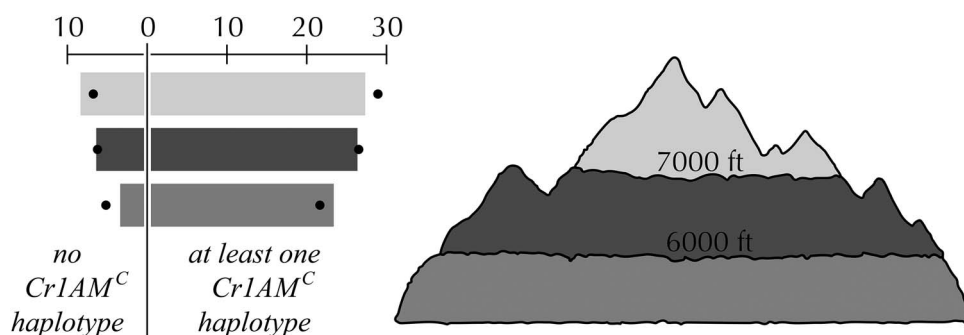


Fig. 3. The counts of individuals carrying at least one *Cr1AM^C* haplotype among putatively *Cr1^R* individuals sampled in three different altitudinal zones. Dots are the expected values. The x axis shows the number of sampled individuals in each category. Although statistically significant heterogeneity is rejected ($P < 0.53$), the frequency appears to decrease with altitude (see data in Supplementary Table S6).

described above, we found three haplotypes, *CrIAM1^A*, *CrIAM1^B*, and *CrIAM1^C*.

Geographic associations of *CrIAM1*. Despite relative uniformity in morphology and ecology over its range (Mirov 1967), geographic differentiation in frequency of *CrI^R* between northern and southern populations of sugar pine has been reported (Kinloch 1992) and a range-wide study using SNPs found differentiation between northern and southern populations (Vangestel et al. 2016). Similarly, Liston et al. (2007) reported the apparently recent (Pleistocene) introgression and replacement by a *P. albicaulis*-related chloroplast genome in the ancestral population of present-day sugar pine in the northern half of its range. The narrow geographic zone of secondary contact is marked by stark polymorphism for these quite diverged chloroplasts (Liston et al. 2007). Consistent with this scenario, the authors note that the distribution of estimated frequencies of *CrI^R* (Kinloch 1992) drops precipitously in parallel with increase in frequency of this introgressed chloroplast genome. In contrast, Weiss et al. (2020) found no differences among similar genetic clusters for quantitative resistance to WPBR (as opposed to the qualitative resistances focused on here). Thus, it is of interest to examine the frequencies of *CrIAM1^C* in the northern populations fixed for the introgressed chloroplast, in the southern populations fixed for the ancestral chloroplast lineage and in the contact zone. Not surprising, there is no evidence of geographic differentiation of the frequencies *CrIAM1^C* among the unselected samples ($P < 0.97$; see Supplementary Table S5).

The immediate impact of these results will be the potential transformation of the ongoing efforts to putatively identify *CrI^R* seed sources suitable for reforestation. Extensive collection of needles from candidate trees across seed zones followed by *CrAM1* genotyping can be expected to greatly increase the efficiency of the ongoing progeny testing-based *CrI* genotyping. It is worth noting that routine field collection of needles from candidate trees involves much less effort than the collection of cones with suitable seeds and can occur nearly any time of year. Indeed, the cost of collecting needles and genotyping is more comparable to that of initially identifying candidate trees. This approach of identifying potential candidate seed source trees followed by seed collection and testing of seedling for WPBR resistance can be expected to yield as few as one tree in every 100 candidates with the *CrI^R/CrI^r* genotype. If only individuals with a *C* haplotype were included in the greenhouse screening, the efficiency of testing would be expected to increase. In the results presented here, only 15% of susceptible trees had a *C* haplotype, while 82% of resistant trees carried one.

If this proposal proves valid, one can consider a radically different strategy in which reforestation simply involves seeds from candidate seed sources that carry the *CrIAM1^C* haplotype. While this would involve many fewer truly resistant seedlings, the cost per resistant seedling might be substantially less. In this context it is worth noting that this approach utilizes seedlings from *CrI^R/CrI^r* mother trees. But these seeds are derived from random, i.e., mostly *CrI^r* pollination. Thus, a full half of the seedlings are expected to be susceptible. If only mother trees with a *C* haplotype are used, we would expect ~82% of those trees to be *CrI^R/CrI^r*. Again, with most wind pollen being *CrI^r*, we would expect 42.5% of seedlings to carry the *CrI^R* allele. A more leveraged application of the *CrIAM1* genotyping could be the genotyping of seedlings before planting. The prospects of this approach depend on the perceived potential value of higher achieved proportions of *CrI^R* carrier seedlings. One can expect that if genotyping is conducted at a large scale with technologically advanced methods that additional cost of genotyping seedlings could be small compared with the total expense. Although exact impact on ongoing sugar pine reforestation is not clear, it seems certain that improvements based on *CrIAM1* genotyping are to be anticipated.

All these calculations are based on the total sample. However, our results also suggest differences in the association between the *C* haplotype and the *CrI^R/CrI^r* genotype across the species range, with higher efficiencies (fewer cases of the “wrong” genotype) in the southern populations, and lower efficiencies in more northern populations (Supplementary Tables S4 and S5), though these results are

not significant (Fig. 2). This is particularly interesting in light of other studies showing genetic differences between northern and southern populations (Liston et al. 2007; Vangestel et al. 2016; Weiss et al. 2020) as well as differences in the frequency of *CrI^R* between northern and southern populations (Kinloch 1992). Unfortunately, the northern populations have the lowest frequency of *CrI^R* as well as the weakest association between *CrI^R* and the *C* haplotype.

The demonstration that SNPs in a large scaffold of the sugar pine genome are in strong linkage disequilibrium opens avenues of research that could greatly enhance the application of genetics in reforestation. The first path is simply to investigate genomic variation in the 6.4-Mbp scaffold now available in the updated V1.5 of the Sugar pine reference genome (available at the TreeGenes database, ID = 17231; <https://treegenesdb.org>). The first question would be “in which direction from the marker is *CrI^R*?” This could be addressed by assessing whether SNPs at one of the two ends of the scaffold are more tightly associated with *CrI^R*. If, as expected, SNPs at the other end of a large scaffold are more weakly associated with *CrI^R*, we could infer the genomic direction toward the gene. Similarly, can we attempt to find markers that are completely associated, i.e., individuals with that SNP are always resistant? There are also the annotated genes in both the original scaffold (Stevens et al. 2016) and the updated assembly (Crepeau et al. 2017). Are there good candidates for *CrI^R*? Obviously finding the *CrI^R* gene itself would yield the best marker to use in the practical applications (such as above). It would also open up structure/function studies of the mechanisms involved in resistance as well as a rich window into evolutionary origins of *CrI^R* (Kinloch 1992), including the comparative genomics with WPBR resistance in other pines. The *P. lambertiana* *CrI* gene may be unique among identified WPBR resistance genes. While WPBR resistance genes with similar properties are known (Kinloch et al. 1999) and progress in molecular mapping has been reported (Liu et al. 2020), the genomic synteny that would support possible homology has not been demonstrated.

Acknowledgments

All relevant data for this article is contained in the manuscript, and in Stevens et al. (2016). The full sugar pine genome is available at <https://treegenesdb.org>, ID = 17231. We acknowledge the native peoples of California and Mexico as the traditional caretakers of the ecosystems sampled for this project. The authors thank Tom Blush, the staff at the Foresthill Seed Orchard, and the staff at the Institute of Forest Genetics who planted the Harrel test in 1992. We thank Anna Schoettle, Melissa Jenkins, and two anonymous reviewers for their comments on an earlier draft of this manuscript. Any use of product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This research was supported in part by the U.S. Department of Agriculture, Forest Service. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official U.S. Department of Agriculture or U.S. Government determination or policy.

Literature Cited

- Crepeau, M. W., Langley, C. H., and Stevens, K. A. 2017. From pine cones to read clouds: Rescaffolding the megagenome of sugar pine (*Pinus lambertiana*). *Genes Genomes Genetics* 7:1563-1568.
- Devey, M. E., Delfino Mix, A., Kinloch, B. B., and Neale, D. B. 1995. Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white pine blister rust in sugar pine. *Proc. Natl. Acad. Sci. USA* 92:2066-2070.
- Geils, B. W., Hummer, K. E., and Hunt, R. S. 2010. White pines, *Ribes*, and blister rust: A review and synthesis. *For. Pathol.* 40:147-185.
- Graffelman, J., Chang, C., Puig, X., Wigginton, J., Ortoleva, L., and Engels, W. R. 2021. HardyWeinberg: Statistical Tests and Graphics for Hardy-Weinberg Equilibrium R package V. 1.7.2. <https://cran.r-project.org/web/packages/HardyWeinberg/HardyWeinberg.pdf>
- Graffelman, J., and Weir, B. S. 2016. Testing for Hardy-Weinberg equilibrium at bi-allelic genetic markers on the X chromosome. *Heredity* 116:558-568.
- Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41: 95-98.
- Harkins, D. M., Johnson, G. N., Skaggs, P. A., Delfino Mix, A., Dupper, G. E., Devey, M. E., Kinloch, Jr., B. B., Neale, D. B., and Johnson, G. N. 1998. Saturation mapping of a major gene for resistance to white pine blister rust in sugar pine. *Theor. Appl. Genet.* 97:1355-1360.
- Jermstad, K. D., Eckert, A. J., Wegrzyn, J. L., Delfino Mix, A., Davis, D. A., Burton, D. C., and Neale, D. B. 2011. Comparative mapping in *Pinus*: Sugar pine (*Pinus lambertiana* Dougl.) and loblolly pine (*Pinus taeda* L.). *Tree Genet. Genomes* 7:457-468.

- King, J. N., David, A., Noshad, D., and Smith, J. 2010. A review of genetic approaches to the management of blister rust in white pines. *For. Pathol.* 40: 292-313.
- Kinloch, B. B. 1992. Distribution and frequency of a gene for resistance to white pine blister rust in natural populations of sugar pine. *Can. J. Botany-Rev. Can. De Bot.* 70:1319-1323.
- Kinloch, B. B. 2003. White pine blister rust in North America: Past and prognosis. *Phytopathology* 93:1044-1047.
- Kinloch, B. B., Parks, G. K., and Fowler, C. W. 1970. White pine blister rust: Simply inherited resistance in sugar pine. *Science* 167:193-195.
- Kinloch, B. B., Jr., Snieszko, R. A., Barnes, G. D., and Greathouse, T. E. 1999. A major gene for resistance to white pine blister rust in western white pine from the Western Cascade Range. *Phytopathology* 89:861-867.
- Konieczny, A., and Ausubel, F. M. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4:403-410.
- Lanner, R. M. 2007. *Conifers of California*. Cachuma Press, Los Olivos, CA.
- Liston, A., Parker-Defeniks, M., Syring, J. V., Willyard, A., and Cronn, R. 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: A case study in *Pinus lambertiana*. *Mol. Ecol.* 16:3926-3937.
- Liu, J. J., Williams, H., Zaman, A., Li, X. R., Gellner, S., and Snieszko, R. A. 2020. Development and application of marker-assisted selection (MAS) tools for breeding of western white pine (*Pinus monticola* Douglas ex D. Don) resistance to blister rust (*Cronartium ribicola* JC Fisch.) in British Columbia. *Can. J. Plant Pathol.* 42:250-259.
- Mirov, N. T. 1967. *The Genus Pinus*. Ronald Press, New York.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Samman, S., and Kitzmiller, J. H. 1996. The sugar pine program for development of resistance to blister rust in the Pacific Southwest Region. Pages 162-170 in: *Sugar Pine: Status, Values, and Roles in Ecosystems*. B. B. Kinloch, Jr., M. Marosy, and M. E. Huddleston, eds. University of California, Division of Agriculture and Natural Resources, Davis, CA.
- Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., Paul, R., Gonzalez-Ibeas, D., Koriabine, M., Holtz-Morris, A. E., Martínez-García, P. J., Sezen, U. U., Marçais, G., Jermstad, K., McGuire, P. E., Loopstra, C. A., Davis, J. M., Eckert, A., de Jong, P., Yorke, J. A., Salzberg, S. L., Neale, D. B., and Langley, C. H. 2016. Sequence of the sugar pine megagenome. *Genetics* 204:1613-1626.
- USDA Forest Service. 2018. *Forest Service Handbook Pacific Southwest Region (R5) Vallejo, California. FSH 2409.18 - Timber Sale Preparation. Chapter 50 - Final Contract Preparation (gate 4)*. U.S. Department of Agriculture Forest Service, <https://www.fs.fed.us/im/directives/field/r5/fsh/2409.18/r5-2409-18-50-2018-1-508%20Verified.pdf>
- Vangestel, C., Vázquez-Lobo, A., Martínez-García, P. J., Calic, I., Wegrzyn, J. L., and Neale, D. B. 2016. Patterns of neutral and adaptive genetic diversity across the natural range of sugar pine (*Pinus lambertiana* Dougl.). *Tree Genet. Genomes* 12:51.
- Wahlund, S. 1928. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11:65-106.
- Weiss, M., Snieszko, R. A., Puiu, D., Crepeau, M. W., Stevens, K., Salzberg, S. L., Langley, C. H., Neale, D. B., and De La Torre, A. R. 2020. Genomic basis of white pine blister rust quantitative disease resistance and its relationship with qualitative resistance. *Plant J.* 104:365-376.