Looking for the needle in a downsized haystack: Whole-exome sequencing unravels how selection and gene flow have shaped climatic adaptation in Douglas-fir (*Pseudotsuga menziesii*)

Jan-Peter George^{1,2}, Silvio Schueler³, Michael Grabner⁴, Sandra Karanitsch-Ackerl⁴, Konrad Mayer⁴, Michael Stierschneider⁵, Lambert Weissenbacher² & Marcela van Loo^{2*}

¹ University of Tartu, Faculty of Science & Technology, Tartu Observatory, Observatooriumi 1, 61602 Tartu maakond (Estonia); ² Austrian Research Centre for Forests, Dptm. of Forest Genetics, Seckendorff-Gudent-Weg 8, 1131 Vienna (Austria); ³ Austrian Research Centre for Forests, Dptm. of Forest Growth & Silviculture, Seckendorff-Gudent-Weg 8, 1131 Vienna (Austria); ⁴ University of Natural Resources and Life Sciences (BOKU), Institute of Wood Science and Technology, Konrad-Lorenz-Straße 24, 3430 Tulln an der Donau (Austria); ⁵ Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, 3430 Tulln (Austria);

*contact for correspondence: marcela.vanloo@bfw.gv.at; Austrian Research Centre for Forests, Dptm. of Forest Genetics, Seckendorff-Gudent-Weg 8, 1131 Vienna (Austria)

Abstract

The widespread Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) occurs along a steep gradient of diverse climates throughout its natural range, which is expected to result in spatially varying selection to local climate conditions. However, phenotypic signals of climatic adaptation can often be confounded, because unraveled clines covary with signals caused by neutral evolutionary processes such as gene flow and genetic drift. Here, we present phenotypic and genotypic data from a common garden experiment showing a putative signal of adaptation to climate after trees have been growing for 40 years in a common environment. Sixteen Douglas-fir provenances originating from a North-to-South gradient of approx. 1,000 km were analyzed and genomic information was obtained from exome capture, which resulted in an initial genomic dataset of >90,000 single nucleotide polymorphisms. We used a restrictive and conservative filtering approach which permitted us to include only SNPs and individuals in environmental association analysis (EAA) that were free of potentially confounding effects (LD, relatedness among trees, heterozygosity deficiency and deviations from Hardy-Weinberg proportions). We used four conceptually different genome scan methods based on F_{ST} outlier detection and gene-environment association in order to disentangle truly adaptive SNPs from neutral SNPs and found that a relatively small proportion of the exome showed a truly adaptive signal (0.01-0.17%) when population substructuring and multiple testing was accounted for. Nevertheless, the unraveled SNP candidates showed significant relationship with climate at provenance origins which strongly suggests that they have most likely featured adaption in Douglas-fir across a steep climatic gradient. Two SNPs were independently found by three of the employed algorithms and one could be assigned with high probability to a *Picea abies* homolog gene involved in circadian clock control as was also found in *Populus balsamifera*.

Keywords: Environmental association analysis; EAA; Douglas-fir; climatic adaptation; exome capture; common garden experiment

43 1. Introduction

44 Conifers have successfully occupied a large number of habitats and different climates
 45 throughout their past-glacial histories, which allowed them to survive and establish in harsh
 46 environments but also to colonize ecoregions with high productivity and optimal growing
 47 conditions (Farjon, 2010). As such, Douglas-fir (*Pseudotsuga menziesii*) constitutes one of
 48 the most widespread conifers in western North America with a distribution from Southern
 49 California up to Northern British Columbia (Gugger et al. 2010). Within its current range of
 50 occurrence two distinct varieties are known which appear genetically and phenotypically
 51 different: the coastal variety (also called *P. menziesii* var. *menziesii*) and the interior variety
 52 (*P. menziesii* var. *glauca*). While the coastal variety is mainly found from central California to
 53 coastal British Columbia along the Pacific coast, the interior variety has its main south-to-
 54 north expansion from Wyoming/Montana over Alberta up to central British Columbia
 55 (Hermann & Lavender 1990; Martinez, 1949). Both of these macro-geographic regions differ
 56 substantially in climate with cool and moist climate conditions in the Pacific habitats towards
 57 more continental and dry growing conditions in the interior (Gugger et al. 2010). Both
 58 habitats are separated by the Great Basin and Columbia Plateau, respectively. Accordingly,
 59 several common garden experiments found intra-specific trait differences among Douglas-fir
 60 provenances that were related to growth, physiology, and phenology and which showed
 61 associations with differences in source climate among provenance locations (e.g. Anekonda
 62 et al. 2004; Fielder & Owens 1989; Kleiber et al. 2017; Pharis & Ferrel 1966). While such
 63 trait-climate associations can indeed suggest patterns of local adaptation due to spatially
 64 varying selection (Leinonen et al. 2013), their interpretation can still be doubtful, because the
 65 climatic clines at which provenances occur often show parallelism with re-colonization routes
 66 after population contraction and expansion (Nadeau et al. 2016). Consequently, the putative
 67 adaptive signal obtained from such trait analyses can be confounded by those that were left
 68 behind by neutral processes such as genetic drift and gene flow, which have often resulted in
 69 strong population substructuring (Günther & Coop, 2013).

70 Incorporating information from genetic markers has significantly improved our
 71 understanding of the role of neutral processes in population structuring in many plant species
 72 and has shed light into their postglacial migration histories (Hewitt, 1999). However, the
 73 main challenge still remained, which is to distinguish truly adaptive regions in the genome
 74 from those which show strong divergence just as a result of neutral genetic processes.
 75 Identifying true environmental associations and adaptive regions in the genome (e.g. from

single nucleotide polymorphisms or SNPs) is of utmost importance, since only adaptive markers have the potential to assist tree breeding of more resistant genotypes and selection of climatically adapted genotypes for forest management under climate change (e.g. Grattapaglia et al. 2018; Harfouche et al. 2012; Neale & Kremer, 2011). Yet, one of the strongest limitations in disentangling adaptive from neutral genomic locations is to acquire a marker set which is large and dense enough to cover the genome in a representative and unbiased fashion. This is particularly difficult in conifers such as Douglas-fir, which harbour huge genomes in the magnitude of several gigabases (Neale et al. 2017; Nystedt et al. 2013; De La Torre et al. 2014). Nevertheless, recent improvement in massive parallel sequencing and bioinformatics have made it possible to reduce the complexity of such genomes by limiting it to the protein coding region, which is called the *exome* (e.g. Neves et al. 2013). Exome capture in conifers results in massive complexity reduction which allows to rapidly generate markers at high number and relatively low costs in order to investigate genomic regions of interest thoroughly (see for instance Capblancq et al. 2019; Suren et al. 2016; Thistlethwaite et al. 2017 for examples on conifers). Even though the exome represents only a small fraction of the total genome, this data is particularly suited for environmental-association-analyses (EAA), since adaptive candidate SNPs can be further interrogated when a gene model or annotated reference genome is available. Moreover, by using a subset of SNPs in the exome which show no significant deviation from neutrality expectation (e.g. F_{st}), the dataset can be used for estimating population substructuring as well.

In this study, we investigate a 40-year old common garden experiment with Douglas-fir provenances from 16 accessions across its distribution by combining dendroclimatological methods with modern sequencing technique which allowed us to generate a dense set of single nucleotide polymorphisms throughout the entire exome. Based on the initial finding that provenances showed a strong association between growth traits (response of annual increment to summer temperature) and seed source climate, we used classical F_{ST} -outlier approaches, bayesian inference methods, and latent factor mixed modelling (LFMM) in order to identify SNPs with a „truly“ adaptive signal for climate adaptation. We hypothesize that spatially varying selection has shaped actual patterns of trait differentiation in Douglas-fir despite a high amount of neutral covariance caused by genetic drift and gene flow.

2. Materials & Methods

2.1 Common garden experiment

The analyzed trees originate from a block-replicated common garden experiment that was established in 1976 in eastern Austria with a total of 49 native Douglas-fir seed sources. The mean annual temperature of the trial site is 7.4°C and mean annual precipitation is 650mm (average over the period from 1961-1990; source: Austrian Central Station for Meteorology). Trees have been planted as two-year-old plants in a 2x2m spacing and each provenance was replicated three times in a random fashion within the trial. In 2012, two cores per tree were taken from a subset of 16 provenances originating from a wide part of the natural distribution encompassing Oregon, Washington and British Columbia (see Fig. 1, Tab. 1 and more information below). In brief, tree cores were taken at breast height for a total of 178 trees (9-15 per provenance and evenly sampled across the three blocks) and stored in plastic tubes until further processing in the lab. Cores were then cut into approx. 1.4mm wide cross sections with a double-blade circular saw, placed on microfilms and exposed to a 10 kV (24mA) X-ray source for 25 minutes. Thereafter, the obtained microfilms were analyzed using the software WinDENDRO 2009 (Regent Instrument, Quebec, Canada) by measuring annual increments to the nearest 0.001 mm for early- and latewood, respectively (see also George et al. 2019, 2017, 2015 for more information on the methodology). The 50% percentile of the density distribution between minimum and maximum density of each ring was used for the identification of earlywood-latewood boundaries (Fries & Ericsson 2009; illustrated in Supporting Information Fig. S1). The radial increment data from this common garden experiment, which is in detail described in George et al. (2019) was re-analyzed for this study and therefore additional needle samples for DNA analyses were taken from the same 178 trees in May 2019. Needles were shot from the lower part of the tree crowns by using a shot gun and stored in silica gel until DNA extraction.

2.2 DNA analysis, probe design & SNP calling

For each sample, genomic DNA was extracted from lyophilized needle tissue using a CTAB protocol (van der Beek et al. 1992) with minor modifications made for the processing of 96-well plates. DNA concentration and quality was assessed with Qubit® QuantIT dsDNA BR Assay kit and a Qubit1 Fluorometer (Thermofisher, Waltham, MA, USA). A total volume of 50 µL genomic DNA for each sample (average concentration: 24 ng/ µL) was sent to the genotyping service provider RAPiD Genomics LLC (Gainesville, FL, USA) for sequencing.

Exome capture probes were designed as described in Neves et al. (2013): briefly, DNA was sheared to a mean fragment length of 400bp, fragments were end-repaired, followed by incorporation of unique dual-indexed Illumina adapters and PCR enrichment. Intron-exon boundaries were designed by mapping transcripts from the published Douglas-fir reference transcriptome from Howe et al. (2013) (<https://www.ncbi.nlm.nih.gov/Traces/wgs/?val=GAEK01>) to the Douglas-fir reference genome assembly *Dougfir* 1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCA_001517045.1/#/st) from Neale et al. (2017). Only probes that hit once to the genome and which showed sufficient GC content were included. This resulted in 20K probes that were found in 12,272 scaffolds with an average number of 1.6 probes per scaffold and a maximum number of 18 probes per scaffold. Sequence capture was performed using RAPiD Genomics proprietary chemistry and workflows. Samples were pooled equimolar and sequenced using a HiSeq 2x150.

After sequencing, *Trimmomatic* (Bolger et al. 2014) was used to remove sequencing adapters and the trimmed reads were mapped with BWA (Li et al. 2009) using the default settings to the Douglas-fir reference genome. SNPs within probes were identified with *FreeBayes* (Garrison & Marth, 2012, available at <https://github.com/ekg/freebayes>). *VCFtools* (Danacek et al. 2011) was used to store the identified SNPs for further downstream analysis in VCF format by applying the following filter: accepted mean depth across all individuals=750, minimum total depth per individual=3, Percent of individuals allowed missing to retain a site=0.4, minimum SNP quality=10.

2.3 Consecutive SNP filtering and exclusion of trees with confounding effects

In order to create a genomic dataset that is free of sources which could potentially affect the false positive discovery rate of adaptive SNPs, we used a restrictive and conservative filtering approach: as such, we removed all SNPs with a minor allele frequency < 0.05 as well as those SNPs that showed significant deviation from Hardy-Weinberg expectation (threshold: 10^{-6}). Furthermore, we performed pairwise LD pruning between markers with a LD threshold of 0.2 in order to retain only unlinked SNPs for environmental association analysis. For this, we used the *snpGdsLDpruning* function implemented in the *SNPrelate* package in R (R Development Core Team 2003).

Furthermore, kinship among individuals was estimated by calculating the identity-by-descent (IBD) methods of moments between all pairs of individuals using a subset of SNPs that was

already corrected for linkage disequilibrium (see section above). We retained only those trees that showed a kinship coefficient <0.25 to any other tree in the dataset. Trees that showed signals of heterozygote deficit (calculated as $1 - (\text{Het}_{\text{obs}}/\text{Het}_{\text{Exp}})$ over all loci) were removed prior to analysis with a threshold of 0.1. Finally, we included only trees with an overall SNP call rate of 0.95 for subsequent analyses. All filtering steps were performed in R by using the packages *gdsfmt*, *SNPrelate* (Zheng et al. 2012), and *vcfR* (Knaus & Grünwald, 2017).

2.4 Climate data & trait analyses

Information on provenance climate was obtained from the ClimateNA database (Wang et al 2016) under <http://www.climatewna.com/>. Latitude/longitude and elevation information for provenance origins were used to retrieve a total of 247 annual, seasonal, and monthly climate variables (reference period: 1961-1990). The full list of climate variables can be found in the **Supplementary Information S1**. In order to reduce the complexity of this data, the 247 variables were transformed into principal components and the first four PCs were standardized (i.e. expressed as standard deviations from the mean). These standardized climate PCs (hereafter called environmental PCs) were used for subsequent analyses of environmental association analysis.

Growth traits for this dataset were derived from the study in George et al. (2019). Briefly, tree-ring series were standardized using a 15-year cubic smoothing spline in order to remove the biological age trend. The standardized series were aggregated to chronologies for each provenance by calculating the year-to-year biweight robust mean among trees in each provenance (Bunn 2008). For the purposes of this study we used the bootstrapped response functions that were applied to ring width chronologies of the same Douglas-fir provenances in George et al. (2019). Response functions and in particular response coefficients between a growth trait (in our case ring width) and trial site climate provide insights of the relative importance of an environmental variable for a trait (e.g. Levesque et al. 2014; Zang & Biondi 2013). However, when calculated at provenance level, bootstrapped response coefficients can be related to provenance source climate in order to unravel genecological clines (that is: some provenances may react more sensitive to an environmental factor than others as a result of local adaptation). For environmental association analysis in this study we used the response of earlywood increment to July temperature at trial site because of two reasons: first, the data in George et al. (2019) showed significant differentiation among Douglas-fir provenances for

this trait and, second, growth response in earlywood has been shown to be significantly involved in drought response and subsequent survival in Douglas-fir (e.g. Martinez-Meier et al. 2008) which suggests that this trait most likely harbours signals of spatially varying selection among provenances. Multi-year correlations between standardized earlywood time series and July temperature at trial site between 1979 and 2011 were bootstrapped by using the R packages *bootRes* (Zang & Biondi, 2013) and *dplR* (Bunn 2008).

2.5 Outlier detection & environmental association analysis (EAA)

The aim of this study is to distinguish SNP markers throughout the exome which have a putative adaptive signal for an environmental factor from non-adaptive SNPs. For this, we used four different algorithms which are either based on F_{ST} outlier detection methods (Arlequin & BayScEnv) or on correlations between allele frequencies and environmental variables (BayEnv2 & LFMM2). Regardless of the algorithm that is used for doing this, controlling for variation in allele frequencies that is simply caused by neutral evolutionary processes such as genetic drift or gene flow is a crucial step for identifying adaptive candidate SNPs (Rellstab et al. 2015). Since the fraction of such adaptive SNPs is usually small compared to all analyzed SNPs, adjusting the results carefully for false positive discovery rates is of utmost importance. We therefore picked four up-to-date population genomic programs (Arlequin, BayScEnv, BayEnv2, LFMM2) which have implemented various testing strategies in order to control for confounding effects and false positive errors.

The coalescent approach implemented in *Arlequin* 3.5 (Excoffier & Lischer 2010) compares locus-specific patterns of population differentiation to a null distribution which is simulated under the assumption of a hierarchical island model (Excoffier et al. 2009). P-values are estimated locus-wise from the joint distribution of heterozygosity and F_{ST} . If the locus-specific F_{ST} exceeds the confidence boundaries of the global observed F_{ST} the respective locus is considered to be an outlier and likely to be under selection. Since we observed significant population substructure (see results below) in our dataset, we chose the hierarchical island model instead of the finite island model in order to reduce the number of false positive candidate SNPs, since confidence intervals around a given F_{ST} value become narrower the more populations are sampled (see Excoffier et al. 2009). Groups for the hierarchical island models were defined based on the first two principal components obtained from 1,500 randomly chosen SNPs. We used 200,000 simulations in *Arlequin* vers. 3.5 with 10 simulated

groups and 100 demes in each group. SNPs were considered to be under selection when their observed F_{ST} exceeded the 5% upper quantile of the simulated null distribution.

BayScEnv (Villemereuil & Gaggiotti, 2015) is a F_{ST} -based genome-scan method that incorporates also environmental differentiation. The algorithm tests a model of local adaptation with environmental differentiation (i.e. the parameter $\mathbf{g}$ in Villemereuil & Gaggiotti, 2015) against two alternative models: i) the neutral model consisting largely of demographic effects and ii) the locus-specific model which includes locus-specific effects other than selection due to environmental differentiation. Posterior error probabilities for choosing the model of local adaptation in favour of the two alternative models are calculated by means of MCMC simulations and results are also adjusted for false discovery rates by providing q-values for each locus. We used 20 pilot runs each with a length of 2000 and a burn-in of 50,000 with a thinning interval of 10. Furthermore, a prior probability of 0.1 was used for the parameter π (assuming every 10th locus harbours a non-neutral signal), while a prior probability of 0.5 was used for the parameter p (assuming around 50% of all loci are associated with an environmental variable). Acceptance rates and convergence of the MCMC chain were inspected after each model run and for each of the four environmental variables (i.e. the standardized climate PCs) by using the *coda* package in R.

In contrast to the F_{ST} -based outlier detection method *Bayenv2.0* (Günther & Coop 2013) calculates standardized allele frequencies and tests for correlations between environmental variables and those frequencies. Neutral and spurious correlations arising from shared population history and gene flow are controlled for by including a covariance matrix obtained from a subset of unlinked and putatively neutral SNPs. We used a subset of 1,500 randomly chosen unlinked SNPs for matrix estimation and ran *Bayenv2.0* with 100,000 iterations. Subsequently, all identified SNPs for each of the four environmental variables with a Bayes factor > 30 (Jeffreys 1961) were considered as candidates for environmental association analysis. Since the algorithm in *Bayenv2.0* is based on MCMC simulations, each run was repeated several times in order to ensure that results remain robust. We additionally calculated spearman rank correlation coefficients in addition to Bayes factors for each SNP in order to ensure that detected candidates were correctly identified and not confounded by outliers (see recommendation in Günther & Coop, 2013).

Finally, we used a regression model combining fixed (i.e. environmental) and latent model effects (i.e. neutral population structure). The model has the form:

$$Y = XB^T + W + E \quad (1)$$

With Y being the response genotype matrix, XB^T the matrix of fixed effects (climate PCs), E the matrix of residual errors, and W the „latent matrix“ (see Caye et al. 2019). W determines the confounding effect of neutral population structure and is determined by k latent factors, where k defines the number of ancestral populations as obtained from a subset of putative neutral loci. As an approximation of k , we used the screeplot obtained from principal component analysis performed on the 1,500 randomly chosen SNPs described above. Calculations were carried out using the *lmm2* function in the LEA package in R (Frichot & Francois, 2015).

2.6 Adaptive signal of outlier SNPs and functional interrogation

In order to further interrogate the identified candidate SNPs, we first defined a set of consensus variants. Consensus SNPs were defined as those SNPs which were independently detected by at least two of the four chosen methods according to thresholds in Tab. 2. From those SNPs 500 bp long fragments were extracted (250 downstream and 250 bp upstream of the variant) and blasted against the Norway spruce (*Picea abies*) genome assembly (*Picea abies* 1.0 (complete) available under www.congenie.org) since this is so far the closest relative to Douglas-fir having an functionally annotated genome assembly. For every homolog contig that was found E-values, average identity, and putative gene function (if known) were reported. We always selected the homolog with the lowest E-value (i.e. number of hits one can expect by chance alone) and -if more than one gene was known to be located on that homolog- all putative gene functions were reported.

Finally and in order to prove whether the unraveled consensus SNPs have the ability to discriminate between provenances adapted to different climates, the first and second eigenvectors for each tree were calculated based on consensus SNPs and, for comparison, based on an equal number of randomly chosen neutral SNPs, respectively. For this, the *snpgdsPCA* function from the SNPRelate package in R was employed.

3. Results

3.1 SNP dataset

Within the 20K probes which were designed for exome capture, a total of 90,979 SNPs were successfully called with *Freebayes* and passed the applied quality filter as described in section 2.2. Since a large proportion of these 90,979 SNPs had either a minor allele frequency

<0.05 or a call rate <0.95, 41,009 SNPs were consequently removed and not considered for downstream analysis. 4,269 markers showed significant deviations from Hardy-Weinberg proportion and were excluded from analyses accordingly. Finally, testing for linkage disequilibrium among markers revealed 17,489 SNPs which were physically unlinked and only these SNPs were included in all subsequent analysis steps.

From the 178 trees, 5 showed significant heterozygosity deficiency probably as a result of inbreeding and were hence excluded from subsequent analyses. There was no significant family structure among trees within provenances as revealed by IBD methods of moments so that all remaining 173 trees were included in environmental association analysis.

3.2 Climatic groups, population structure & growth response functions

The first two principal components of the 250 long-term climatic variables together explained 79.9% of variation (Fig. 1). The first principal component (hereafter called climate PC1) was strongly related to temperature variables (correlation to mean annual temp. -0.99 and to minimum temperature of coldest month: -0.97) and length of growing season (correlation to number of growing degree days>5°C=0.88). The second component (climate PC2) was strongly correlated with precipitation regime at seed origin ($r_{\text{MAP}}=0.94$; $r_{\text{summer:heat moisture index}}=0.81$). The third and fourth PCs explained less variation and were related to solar radiation and annual snow precipitation, respectively. As expected, provenances could be roughly dividend into two main groups when plotted against the first two climate PCs, that is inland and coastal provenances (Fig. 1d). Provenances Cle Elum (WA) and Pine Grove (OR) clustered climatically more closely with inland varieties, although they belong geographically still to the coastal zone.

Based on the subset of 1,500 randomly chosen SNPs five population clusters were unraveled: i) provenances from northern inland British Columbia (Fort St. James, Clemina, Adams Lake), ii) the three inland varieties from interior Washington and Southern British Columbia (Newport, Spokane (both WA) and Nelson (BC)), iii) coastal varieties from British Columbia, Washington and Oregon (Alberni, Matlock, Cle Elum, Pine Grove, Randle, Darrington, Abiqua Basin & Cascadia), iv) a single provenance cluster Snowqualmie Pass (WA) and v) another single provenance cluster including provenance D'Arcy in the transition zone between coastal and inland British Columbia (Fig. 1c).

Response of earlywood growth to July temperature at the trial site varied significantly among provenances. In general, inland provenances from British Columbia and Washington had lower response coefficients than their coastal counterparts (Fig. 2). Response coefficients varied between 0.17 (Newport WA) to 0.42 (Cascadia OR). Differences among provenances were statistically significant as suggested by 95% confidence intervals of the bootstrapped coefficients. When regressed against climate PC1 and PC2 the relationship was significant with a stronger positive earlywood response towards warmer provenance origins ($r=0.57$; Fig 2a) as well as towards wetter provenance origins ($r=0.48$; Fig 2b). There were no significant relationships to climate PC3 and climate PC4 and consequently only the first two climate PCs were considered for environmental association analysis.

3.3 Outlier SNPs and environmental associations

The total number of outlier SNPs and those which showed associations with climate varied significantly among detection methods. Arlequin and LFMM2 found a total of 1148 and 1082 SNPs which showed signals of selection and association to climate PC1, respectively. However, when corrected for multiple comparisons, only 29 and 2 SNPs, respectively, passed the adjusted p-value threshold ($-\log_{10}$ corrected=5.54). BayEnv2 revealed a total of 10 SNPs associated with climate PC1 and BayScEnv exhibited 4 SNPs (Fig.3, Fig.4a & Tab.2). 194 SNPs were commonly found by Arlequin and LFMM2, 2 SNPs by LFMM2 and BayEnv2, each 1 between Arlequin and BayEnv2 and LFMM2 and BayScEnv and all four SNPs detected with BayScEnv were also listed as outliers in Arlequin. As the most promising candidates SNPs #15099 and #78509 appeared independently as outliers associated with temperature in three of the four employed programs (Fig. 5a).

For climate PC2 LFMM2 detected 908 SNPs (2 SNPs still significant after Bonferroni correction), BayEnv2 18 SNPs while no significant SNPs were detected by BayScEnv. 46 SNPs which appeared as outliers in Arlequin were also shortlisted in LFMM2 and SNPs #51115 and #69292 were commonly detected by LFMM2 and BayEnv2. However, for climate PC2 no SNP appeared to be shortlisted more than twice.

3.4 Interrogation of putative adaptive SNPs

A total of 18 consensus SNPs (11 for Climate PC1 and 7 for climate PC2) were selected for functional interrogation and spatial frequency analysis. All 18 SNPs were at least found independently by two or three of the used algorithms when the thresholds described in Table 2 were applied. However, since a large number of common SNPs were found between LFMM2 and Arlequin (194 and 46, respectively) we selected only the top five candidates ranked by the $-\log_{10}$ LFMM p-value for both climate PCs, respectively.

For climate PC1 the analysed flanking sequences matched to *Picea abies* homologs with expect values ranging between 0 and 5.13×10^{-6} (note that the lower the E-value the better the match). 8 out of the eleven found homologs had entries for annotated genes and for 5 of these genes putative functions were found in the congene database (Table 3). Among those were transcription factors and proteins involved in DNA and RNA binding (e.g. TF VOZ1, Heterogeneous nuclear ribonucleo R-like isoform X), but also universally conserved proteins with still unknown function such as HemK methyltransferase family member 2 (HemK protein). Most interestingly, for SNP #78509 which was independently detected three times as outlier and significantly associated with temperature, a circadian clock protein involved in photoperiod sensing confirmed in *Arabidopsis thaliana* and *Populus trichocarpa* was found to be coded by a gene located on the *P. abies* homolog contig MA_6946339 (Tab. 3).

For climate PC2 five of the seven sequences showed homolog contigs with E-values between 2.23×10^{-159} to 3.17×10^{-4} and for four homologs genes with putative protein functions were listed. As such, we found proteins involved in carbohydrate metabolism (Bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1) as well as transcription factors known to be involved in stomata formation in *Nicotiana tomentosiformis* (Transcription factor FAMA isoform X2).

Finally, our two candidate SNPs #15099 and #78509, which were three times independently identified as truly adaptive outliers showed significant differentiation in space, since the alternative allele increased in frequency for both SNPs towards the inland and also towards colder provenance locations (Pearson correlation coefficients between allele frequencies and mean warmest month temperature were -0.31 and -0.32, respectively; Fig. 6).

When comparing the first two eigenvectors based on the 18 consensus SNPs to eigenvectors obtained from an equal number of randomly chosen neutral SNPs, provenances could be clearly separated into inland and coastal provenances whereas no clear differentiation was observed for the neutral subset (Fig. 7).

390

391 4. Discussion

392 In this study we combined observations from a common garden experiment with whole-
393 genome information in order to unravel polymorphisms which could have caused the strong
394 genecological cline provenances exhibited after 40 years of growing. While such old
395 common garden experiments are rare for most tree species, their retrospective analysis by
396 using dendroclimatology can shed light in evolutionary history and adaptation patterns,
397 because a high number of repeated observations is naturally archived in tree-rings. Although
398 this study is not the first attempt that linked molecular genomics with tree-ring data (e.g.
399 Trujillo-Moya et al. 2018; Housset et al. 2018; Heer et al. 2018), it is one of the very first
400 environmental association analyses for the economically important Douglas-fir, which is
401 recently attracting attention outside its natural range due to its ability to cope with warming
402 and drought (Eilmann & Rigling 2012). Hence, we will discuss our findings also in the light
403 of adaptive forest managment.

404 4.1 Growth response of trees after four decades in the common garden

405 The sixteen analyzed provenances showed a strong phenotypic cline with temperature and
406 precipitation at seed origin. Provenances with earlywood growth strongly responding to July
407 temperature at trial site originated mainly from the warmer and wetter coastal sites in British
408 Columbia, Washington and Oregon, whereas less responsive provenances came from colder
409 and drier inland sites. Many other studies have revealed such a differentiation pattern
410 between the coastal variety *menziesii* and inland variety *glauca* either in terms of productivity
411 (Eilmann et al. 2013), physiology (Anekonda et al. 2004), and even stress response (Kleiber
412 et al. 2017). Our findings could indeed add evidence that coastal varieties are generally more
413 productive, since earlywood response to summer temperature is a strong indicator for overall
414 productivity. Earlywood cells in conifers mainly ensure that water demand in the crown will
415 be sufficiently covered during phases of high evapotranspiration and trees with larger
416 earlywood fractions can consequently allocate more carbon (Björklund et al. 2017). The
417 climate at our trial site already represented the dry margin of the natural distribution of
418 Douglas-fir (George et al. 2019) and we can assume that trees have been growing under
419 stressful conditions in most of the years. The study by George et al. (2019) in which the same
420 trees have been analyzed also demonstrated that provenances from warmer locations were the
421 most tolerant ones during years with extreme water deficit, whereas colder inland

provenances suffered most. However, it has yet to be confirmed whether the higher drought tolerance of coastal provenances (expressed as ratio between growth during the drought year compared to a pre-drought period) really mirrors the ability of trees to withstand drought or simply exhibits conservative growth patterns which could come at the cost of lower recovery or higher mortality after drought periods (e.g. Montwe et al. 2015).

4.2 Outlier SNPs and the role of selection across a steep environmental cline

One of the main goals of this study was to identify polymorphisms which show true signals of adaptation to climate and which are not confounded by neutral processes such as gene flow, drift, and migration. The analysed provenances showed strong genetic substructuring which partly concided with phenotypic differentiation. This makes disentangling adaptive from non-adaptive signals particularly challenging (Ahrens et al. 2018; Nadeau et al. 2016; Rellstab et al. 2015). However, we employed four conceptually different algorithms which have implemented various methodologies for accounting for neutral processes and each of them identified a certain number of outliers. The genome scan method in Arlequin identified the highest number of SNPs, which is not surprising, since Arlequin does neither specifically associate polymorphisms to climatic factors nor does it take environmental differentiation into account. Hence, the identified SNPs are probably loaded with adaptive signals from other environmental drivers than temperature and precipitation. Surprisingly, after correcting for multiple testing only a very small number of SNPs was still standing out as truly adaptive and represented only between 0.01% (LFMM) to 0.17% (Arlequin) of the analyzed SNPs. Although direct comparison with other studies is complicated due to varying experimental settings and sample size (see for example Ahrens et al. 2018), adaptive SNPs in other genera such as *Fagus* (Pluess et al. 2015), *Populus* (Fahrenkrog et al. 2017), *Alnus* (DeKort et al. 2014), and *Pinus* (Ruiz Daniels et al. 2019) comprised between 1.6% (Fahrenkrog) to 11% (Pluess et al. 2015) of all SNPs that were analyzed. Moreover, when the data was further aggregated to consensus SNPs which were simultaneously detected by more than one method, only 11 candidates out of 17489 (0.06%) were detected for temperature and 7 (0.04%) for precipitation regimes. Since the analyzed provenances exhibit a strong pattern of population substructuring into five clearly delimited clusters, we strongly presume that this neutral variation is the most important cause for this finding. In particular, provenances D'Arcy (BC) and Snowqualmie Pass (WA) represented strongly isolated subpopulations based on their neutral genetic pattern (Fig. 1c), although their phenotypic signal fit well into the observed environmental cline with both climate PCs (Fig 2).

The 18 consensus outlier SNPs that were either associated with temperature or precipitation were capable of discriminating between two clusters of provenances which could be clearly assigned to one coastal and one inland group (Fig. 7). This underpins that these SNPs have most likely featured phenotypic and genetic differentiation in Douglas-fir beyond neutral processes such as isolation-by-distance. Interestingly, provenances D'Arcy (BC) and Snowqualmie Pass (WA) appeared not any longer as single subpopulation clusters when the outlier dataset was used for clustering (compare to Fig. 1c with the neutral dataset of 1,500 SNPs). One possible explanation could be that isolation of these provenances occurred relatively late compared to climatic adaptation. This finding would be corroborated by the fact that both populations showed a rather expected phenotypic signal which was similar as for the other coastal provenances (Fig. 2). Coastal and interior Douglas-fir varieties have most likely diverged during orogeny of the Cascade range, which have led to xerification of the Great Basin during the Pliocene around 2 Ma ago (Gugger et al. 2010; Brunsfeld et al. 2001). However, population contraction and expansion around the last LGM 18ka ago could have caused local spots with limited gene exchange, in particular in areas with high mountain barriers (Gugger & Sugita, 2010).

Nevertheless and despite the overall small number of SNPs that was found, we were able to extract two hot candidate SNPs associated with temperature regime at seed origin. Minor allele frequencies of both SNPs showed significant correlation with mean warmest month temperature in space, which is highly correlated with the initial response climate variable from the common garden (correlation of July temperature and MWMT=0.99). In both cases, the frequency of the minor C allele increased towards inland areas with lower temperature. In light of the strict filtering applied in this study and given the number of employed algorithms, the evidence strongly suggests that these SNPs could have featured adaptation to temperature regimes in Douglas-fir across the steep gradient that was analyzed here. Although the correlation between allele frequencies and temperature regime was moderate and characterized by a few unexpected outliers, it is likely that the rather small sample size of analyzed trees within provenances can be responsible for that. We hence see this result as a starting point for further investigations including also landscape genomics in order to corroborate these findings with more data.

4.3 Putative biological functions of candidate SNPs

Many of the consensus SNPs had homolog contigs at the annotated *Picea abies* reference genome from Nystedt et al. (2013) and we found some contigs containing genes with an annotated known biological function (Tab. 3). While not all of them refer directly to functions related to temperature response or water supply, their further interrogation could be though valuable, since some of the transcription factors and proteins are, for instance, involved in carbohydrate metabolism. For example, Bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1 are important catalysator proteins which are involved in the process of cell wall biosynthesis and cell wall organization in stem tissue of higher plants (see for more information the Uniprot entry Q42605 under <https://www.uniprot.org>) which in turn is an important trait in softwood species conferring drought resistance (Isaac-Renton et al. 2018).

Most interestingly, the triple-found SNP #78509 had a highly similar homolog contig which contained one annotated gene coding a WD repeat-containing LWD1 protein. This protein belongs to the circadian clock protein family and is regulating circadian period length and photoperiodic flowering in *Arabidopsis spec.* (Wang et al. 2011; Wu et al. 2008). Moreover, a closely related protein of this family, named GIGANTEA-5, is involved in adaptation to temperature regimes (Cao et al. 2005) and was also found to be under strong selection in *Populus balsamifera* (Keller et al. 2012; Fitzpatrick & Keller 2015). In the study by Fitzpatrick & Keller (2013), turnover in frequency of SNPs located in this gene were best explained by differences in temperature regimes among provenance origins, which strongly corroborates the findings of this study. We see this finding hence as a putative needle in the haystack, which could be the starting point for further investigations at landscape level.

4.4 Douglas-fir genomic resources for adaptive forest management

Douglas-fir is currently discussed as a promising substitute species outside its natural range given its putatively higher drought tolerance and yield (Chakraborty et al. 2016; Eilmann & Rigling, 2012). Our results could be pertinent for future studies and applications which are aiming at identifying adapted seed sources for Douglas-fir. For instance, SNP information can be used in order to genotype large arrays of trees from progeny trials or provenance tests in order to decide whether the selected material is generally site-adapted or likely to be maladapted to temperature and drought regime. Additionally, our dataset can be used as reference dataset for genomic assignment of individuals or stands given that many Douglas-

fir seed stands in Europe are still without confirmed geographic origin. Therefore, we provide detailed SNP and phenotypic information such as genomic positions, flanking sequences, and rank scores for more than 17,000 SNPs obtained from the various programs in the Supplementary Material S2.

5. Conclusions

Based on a steep phenotypic cline observed in a common garden experiment we were able to disentangle adaptive signals in Douglas-fir from those that were simply caused by neutral demographic processes. We showed that combining dendroclimatological data with whole-exome information can lead to valuable insights into the adaptation history of a widespread conifer. Although a very small fraction of the analyzed polymorphisms stood out as adaptive candidates, their functional interrogation strongly suggests that SNPs could have indeed featured climatic adaptation in Douglas-fir. Thereby, we shed new light into the adaptive history of another conifer with high economic and ecological importance.

6. Acknowledgements

This study is part of the Translational Research Program 122 ('Softwood for the Future') and was funded by the Austrian Science Fund (FWF). Dr. Marcela van Loo was funded by the Austrian Science Fund (FWF) project „Verhalten der `Europäischen´ Douglasie“ (P26504). We also would like to thank the Austrian Central Institute for Meteorology and Geodynamics (ZAMG) for providing weather data as well as Michael Eberhardt for help during the field work. We are also grateful to Thorsten Günther (Uppsala University) for advice with Bayenv2.

References

- Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, 27(6), 1342-1356.
- Anekonda, T., Jones, C., Smith, B. N., & Hansen, L. D. (2004). Differences in physiology and growth between coastal and inland varieties of Douglas-fir seedlings in a common garden. *Thermochimica acta*, 422(1-2), 75-79

- 549 Björklund, J., Seftigen, K., Schweingruber, F., Fonti, P., von Arx, G., Bryukhanova, M. V., ...
550 & Frank, D. C. (2017). Cell size and wall dimensions drive distinct variability of earlywood
551 and latewood density in Northern Hemisphere conifers. *New Phytologist*, 216(3), 728-740.
552
- 553 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
554 sequence data. *Bioinformatics*, 30(15), 2114-2120.
555
- 556 Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001). Comparative phylogeography of
557 northwestern North America: a synthesis. In: Integrating Ecology and Evolution in a Spatial
558 Context (eds Silvertown J, Antonovics J), pp. 319–339. Blackwell Publishing, Williston,
559 Vermont.
560
- 561 Bunn, A. G. (2008). A dendrochronology program library in R (dplR). *Dendrochronologia*,
562 26(2), 115-124.
563
- 564 Cao, S., Ye, M., & Jiang, S. (2005). Involvement of GIGANTEA gene in the regulation of
565 the cold stress response in Arabidopsis. *Plant cell reports*, 24(11), 683-690.
566
- 567 Capblancq, T., Butnor, J. R., Deyoung, S., Thibault, E., Munson, H., Nelson, D. M., ... &
568 Keller, S. R. (2020). Whole-genome sequencing reveals a long-term decline in effective
569 population size of red spruce (*Picea rubens*). *Evolutionary Applications*.
570
- 571 Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate
572 inference of gene-environment associations in genome-wide studies. *Molecular biology and
573 evolution*, 36(4), 852-860.
574
- 575 Chakraborty, D., Wang, T., Andre, K., Konner, M., Lexer, M. J., Matulla, C., ... & Schueler,
576 S. (2016). Adapting Douglas-fir forestry in Central Europe: evaluation, application, and
577 uncertainty analysis of a genetically based model. *European Journal of Forest Research*,
578 135(5), 919-936.
579
- 580 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... &
581 McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-
582 2158.
583
- 584 De Kort, H., Vandepitte, K., Bruun, H. H., Closset-Kopp, D., Honnay, O., & Mergeay, J.
585 (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to
586 temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology*, 23(19),
587 4709-4721.
588
- 589 De La Torre, AR., Birol, I., Bousquet, J., Ingvarsson, P. K., Jansson, S., Jones, S. J., ... &
590 Street, N. (2014). Insights into conifer giga-genomes. *Plant Physiology*, 166(4), 1724-1732
- 591 De Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local
592 adaptation using environmental variables. *Methods in Ecology and Evolution*, 6(11), 1248-
593 1258.
594
- 595 Eilmann, B., de Vries, S. M., den Ouden, J., Mohren, G. M., Sauren, P., & Sass-Klaassen, U.
596 (2013). Origin matters! Difference in drought tolerance and productivity of coastal Douglas-

- 597 fir (*Pseudotsuga menziesii* (Mirb.)) provenances. *Forest Ecology and Management*, 302, 133-
598 143.
- 599
- 600 Eilmann, B., & Rigling, A. (2012). Tree-growth analyses to estimate tree species' drought
601 tolerance. *Tree Physiology*, 32(2), 178-187.
- 602
- 603 Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically
604 structured population. *Heredity*, 103(4), 285-298.
- 605
- 606 Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to
607 perform population genetics analyses under Linux and Windows. *Molecular ecology*
608 *resources*, 10(3), 564-567.
- 609
- 610 Fahrenkrog, A. M., Neves, L. G., Resende Jr, M. F., Dervinis, C., Davenport, R., Barbazuk,
611 W. B., & Kirst, M. (2017). Population genomics of the eastern cottonwood (*Populus*
612 *deltoides*). *Ecology and evolution*, 7(22), 9426-9440.
- 613
- 614 Farjon, A. (2010). *A Handbook of the World's Conifers* (2 vols.) (Vol. 1). Brill.
- 615 Fielder, P., & Owens, J. N. (1989). A comparative study of shoot and root development of
616 interior and coastal Douglas-fir seedlings. *Canadian Journal of Forest Research*, 19(5), 539-
617 549
- 618 Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level
619 modelling of biodiversity: Mapping the genomic landscape of current and future
620 environmental adaptation. *Ecology Letters*, 18(1), 1-16.
- 621
- 622 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological
623 association studies. *Methods in Ecology and Evolution*, 6(8), 925-929.
- 624
- 625 Fries, A., & Ericsson, T. (2009). Genetic parameters for earlywood and latewood densities
626 and development with increasing age in Scots pine. *Annals of Forest Science*, 66(4), 1-8.
- 627
- 628 Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read
629 sequencing. *arXiv preprint arXiv:1207.3907*.
- 630
- 631 George, J. P., Grabner, M., Campelo, F., Karanitsch-Ackerl, S., Mayer, K., Klumpp, R. T., &
632 Schüler, S. (2019). Intra-specific variation in growth and wood density traits under water-
633 limited conditions: long-term-, short-term-, and sudden responses of four conifer tree species.
634 *Science of The Total Environment*, 660, 631-643.
- 635
- 636 Gugger, P. F., & Sugita, S. (2010). Glacial populations and postglacial migration of Douglas-
637 fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews*, 29(17-18),
638 2052-2070.
- 639
- 640 Gugger, P. F., Sugita, S., & CAVENDER-BARES, J. (2010). Phylogeography of
641 Douglas fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses
642 from the fossil record. *Molecular Ecology*, 19(9), 1877-1897.

- Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1), 205-220.
- Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S., Tan, B., ... & El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science*, 9, 1693.
- Harfouche, A., Meilan, R., Kirst, M., Morgante, M., Boerjan, W., Sabatti, M., & Mugnozza, G. S. (2012). Accelerating the domestication of forest trees in a changing world. *Trends in Plant Science*, 17(2), 64-72.
- Heer, K., Behringer, D., Piermattei, A., Bässler, C., Brandl, R., Fady, B., ... & Vendramin, G. (2018). Linking dendroecology and association genetics in natural populations: Stress responses archived in tree rings associate with SNP genotypes in silver fir (*Abies alba* Mill.). *Molecular Ecology*, 27(6), 1428-1438.
- Hermann, R. K., & Lavender, D. P. (1990). Douglas-fir. *Silvics of North America*, 1, 527-540
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society*, 68(1-2), 87-112.
- Housset, J. M., Nadeau, S., Isabel, N., Depardieu, C., Duchesne, I., Lenz, P., & Girardin, M. P. (2018). Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. *New Phytologist*, 218(2), 630-645.
- Howe, G. T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., ... & Dean, J. F. (2013). A SNP resource for Douglas-fir: de novotranscriptome assembly and SNP detection and validation. *BMC genomics*, 14(1), 137.
- Isaac-Renton, M., Montwé, D., Hamann, A., Spiecker, H., Cherubini, P., & Treydte, K. (2018). Northern forest tree populations are physiologically maladapted to drought. *Nature communications*, 9(1), 1-9.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford Classic Texts in the Physical Sciences. Oxford Univ. Press, Oxford.
- Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution*, 29(10), 3143-3152.
- Kleiber, A., Duan, Q., Jansen, K., Verena Junker, L., Kammerer, B., Rennenberg, H., ... & Kreuzwieser, J. (2017). Drought effects on root and needle terpenoid content of a coastal and an interior Douglas fir provenance. *Tree Physiology*, 37(12), 1648-1658.
- Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44-53.
- Leinonen, T., McCairns, R. S., O'hara, R. B., & Merilä, J. (2013). Q ST–F ST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, 14(3), 179-190.

- Lévesque, M., Rigling, A., Bugmann, H., Weber, P., & Brang, P. (2014). Growth response of five co-occurring conifers to drought across a wide climatic gradient in Central Europe. *Agricultural and Forest Meteorology*, 197, 1-12.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Martinez M. 1949. Las Pseudotsugas de Mexico. *Anales del Instituto de Biología* 20: 129–184
- Martinez-Meier, A., Sanchez, L., Pastorino, M., Gallo, L., & Rozenberg, P. (2008). What is hot in tree rings? The wood density of surviving Douglas-firs to the 2003 drought and heat wave. *Forest Ecology and Management*, 256(4), 837-843.
- Montwé, D., Spiecker, H., & Hamann, A. (2015). Five decades of growth in a genetic field trial of Douglas-fir reveal trade-offs between productivity and drought tolerance. *Tree Genetics & Genomes*, 11(2), 29.
- Nadeau, S., Meirmans, P. G., Aitken, S. N., Ritland, K., & Isabel, N. (2016). The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecology and Evolution*, 6(24), 8649-8664.
- Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., ... & Casola, C. (2017). The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. *G3: Genes, Genomes, Genetics*, 7(9), 3157-3167.
- Neale, D. B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics*, 12(2), 111-122.
- Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal*, 75(1), 146-156.
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., ... & Vicedomini, R. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579-584.
- Pharis, R. P., & Ferrell, W. K. (1966). Differences in drought resistance between coastal and inland sources of Douglas fir. *Canadian Journal of Botany*, 44(12), 1651-1659.
- Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G., & Oddou-Muratorio, S. (2016). Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist*, 210(2), 589-601.
- Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24(17), 4348-4370.
- Ruiz Daniels, R., Taylor, R. S., González-Martínez, S. C., Vendramin, G. G., Fady, B., Oddou-Muratorio, S., ... & Beaumont, M. A. (2019). Looking for local adaptation: convergent microevolution in Aleppo pine (*Pinus halepensis*). *Genes*, 10(9), 673.

- Suren, H., Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Smets, P., Rieseberg, L. H., ... & Holliday, J. A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, 16(5), 1136-1146.
- Thistlethwaite, F. R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoeck, M. U., & El-Kassaby, Y. A. (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC genomics*, 18(1), 1-16.
- Trujillo-Moya, C., George, J. P., Fluch, S., Geburek, T., Grabner, M., Karanitsch-Ackerl, S., ... & Schueler, S. (2018). Drought sensitivity of Norway spruce at the species' warmest fringe: quantitative and molecular analysis reveals high genetic variation among and within provenances. *G3: Genes, Genomes, Genetics*, 8(4), 1225-1245.
- Van der Beek, J. G., Verkerk, R., Zabel, P., & Lindhout, P. (1992). Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics*, 84(1-2), 106-112.
- Venables, W. N., & Smith, D. M. (2003). The R development core team. *An Introduction to R, Version, 1(0)*.
- Wang, T., Hamann, A., Spittlehouse, D., & Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PloS one*, 11(6), e0156720.
- Wang, Y., Wu, J. F., Nakamichi, N., Sakakibara, H., Nam, H. G., & Wu, S. H. (2011). LIGHT-REGULATED WD1 and PSEUDO-RESPONSE REGULATOR9 form a positive feedback regulatory loop in the Arabidopsis circadian clock. *The Plant Cell*, 23(2), 486-498.
- Wu, J. F., Wang, Y., & Wu, S. H. (2008). Two new clock proteins, LWD1 and LWD2, regulate Arabidopsis photoperiodic flowering. *Plant physiology*, 148(2), 948-959.
- Zang, C., & Biondi, F. (2013). Dendroclimatic calibration in R: the bootRes package for response and correlation function analysis. *Dendrochronologia*, 31(1), 68-74.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326-3328.

Tab.1: Overview of provenance origin and climate

Provenance ID	Name	Region	Latitude	Longitude	MAT	MAP	N
DA	D'Arcy	British Columbia transition	50.5567	-122.5	8.3	516	13
NS	Nelson	British Columbia inland	49.5	-117.2667	6.5	941	13
AB	Alberni	British Columbia coastal	49.325	-124.85	9.2	2161	10
NP	Newport	Washington inland	48.2	-117.05	7.3	736	11
SK	Spokane	Washington inland	47.7833	-117.2	8.4	518	10
ML	Matlock	Washington coastal	47.25	-123.4167	10.1	2353	11
CE	Cle Elum	Washington coastal	47.2167	-121.1167	7	992	9
PG	Pine Grove	Oregon coastal	45.1	-121.3833	8.5	453	10
FJ	Fort St. James	British Columbia inland	54.4833	-124.25	2	590	10
CM	Clemina	British Columbia inland	52.5833	-119.3167	3.9	854	9
RD	Randle	Washington coastal	46.4847	-121.9491	9.3	1604	13
DR	Darrington	Washington coastal	48.1601	-121.4968	6	3623	15
AL	Adams Lake	British Columbia inland	50.9651	-119.7056	6.5	528	11
AQ	Abiqua Basin	Oregon coastal	44.8802	-122.506	8.3	2191	11
CC	Cascadia	Oregon coastal	44.439	-122.41	10.2	1947	10
SP	Snoqualmie Pass	Washington coastal	47.4129	-121.4411	6.1	2551	12
						total	178

806 Tab. 2: Utilized algorithms and chosen thresholds for SNP identification

Algorithm	Method for accounting for neutral processes	Climate data included in calculations	threshold	Total Number of outlier SNPs	Climate PC1	Climate PC2
Arlequin	F_{ST} hierarchical island model	No	$-\log_{10} (1-F_{ST} \text{ quantile}) > 1.3$	1148 (29)		
BayScEnv	F_{ST} null model	Yes	$-\log_{10} (q \text{ value}) > 1.3$	4	4	0
BayEnv2	Matrix estimation from neutral SNPs	Yes	$BF > 30 ; \rho > 0.4$	28	10	18
LFMM2	K latent factors from PCA	Yes	$-\log_{10} (p\text{-value}) > 1.3$	1921 (4)	1082(2)	908 (2)

817 **Tab. 3:** Genomic positions, homolog contigs at the *P. abies* 1.0 (complete) genome assembly, and functionality of adaptive candidate SNPs. SNPs in bold
818 were detected by three of the four algorithms.

819

SNP-ID	F _{ST}	Environmental factor (Climate PC)	Contig	Position	Ref	Alt	<i>Picea abies</i> homolog contig	E-value	Identity [%]	Function
15099	0.297343	Temperature	LPNX010200061.1	392557	T	C	MA_10430533	8.12E-116	88.15	no entry
45651	0.176026	Temperature	LPNX010610246.1	234874	A	T	MA_31242	8.68E-11	87.84	Nuclear factor related to kappa-B-binding
78509	0.412301	Temperature	LPNX011078866.1	83694	T	C	MA_6946339	3.39E-155	89.28	WD repeat-containing LWD1
20020	0.311698	Temperature	LPNX010261061.1	122684	G	T	MA_10426649	5.13E-06	90.00	no entry
13432	0.318517	Temperature	LPNX010175959.1	196183	G	T	MA_16875	2.23E-122	86.44	unknown
71701	0.388778	Temperature	LPNX010985912.1	451465	C	T	MA_10427644	8.64E-11	92.59	Transcription factor VOZ1; Glycosyl transferase family 2;Heterogeneous nuclear ribonucleo R-like isoform X1
12985	0.364605	Temperature	LPNX010167952.1	204395	T	C	MA_14317	1.83E-33	84.10	60S ribosomal
30680	0.285955	Temperature	LPNX010408536.1	201415	C	A	MA_248133	7.32E-16	90.03	unknown
70743	0.359548	Temperature	LPNX010971660.1	402162	A	G	MA_10428327	0.00E+00	96.10	HemK methyltransferase family member 2
69642	0.303341	Temperature	LPNX010958373.1	129773	T	G	MA_10434143	3.48E-84	90.42	unnamed product
89806	0.30836	Temperature	LPNX011224705.1	67931	A	T	MA_206051	2.10E-08	91.23	no entry
51115	0.205466	Precipitation	LPNX010683611.1	224568	C	A	MA_10428611	7.48E-36	94.23	Bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1
69292	0.281179	Precipitation	LPNX010953873.1	35635	G	C	MA_2566561	1.70E-27	87.69	no entry
58318	0.339572	Precipitation	LPNX010792684.1	264816	A	G	MA_57244	2.23E-159	88.48	Transcription factor FAMA isoform X2
31046	0.39123	Precipitation	LPNX010411427.1	510083	T	C	MA_10430971	1.31E-23	88.72	unknown
57319	0.265295	Precipitation	LPNX010776984.1	676889	G	A	MA_113422	1.78E-141	88.84	no entry
17435	0.261599	Precipitation	LPNX010232549.1	270839	T	A	MA_70676	6.53E-61	85.91	Chloroplastic-like
60539	0.379081	Precipitation	LPNX010822297.1	383798	G	C	MA_10430923	3.17E-04	92.31	Exosome complex component RRP41-like

group

British Columbia coastal	British Columbia transition	Washington coastal
British Columbia inland	Oregon coastal	Washington inland

Climate PC1

Climate PC2

Climate PC1

Climate PC2

Provenance

- AB
- AL
- AQ
- CC
- CE
- CM
- DA
- DR
- FJ
- ML
- NP
- NS
- PG
- RD
- SK
- SP

Outlier SNPs

Neutral SNPs