Tree Genetics & Genomes (2018) 14: 10
https://doi.org/10.1007/s11295-017-1225-x

ORIGINAL ARTICLE

Linking phenotype, genotype and environment to unravel genetic components underlying cold hardiness in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*)

Carl Vangestel[1,2,3] · Andrew J. Eckert[4] · Jill L. Wegrzyn[5] · J. Bradley St. Clair[6] · David B. Neale[3]

Received: 14 June 2017 / Revised: 22 November 2017 / Accepted: 28 December 2017 / Published online: 10 January 2018

Abstract

Global climate change may detrimentally affect future generations of numerous forest tree species, hampering their long-term sustainability if appropriate evolutionary responses remain lacking. To face these novel threats, conservation biologists are in need of a thorough understanding and identification of adaptive variation in key fitness traits. We here provide an elaborate synthesis of pre-existing and novel analyses of an association mapping, genecological and landscape genomic study integrating genotypic, environmental and phenotypic data to gain insights into the genetic basis of cold-hardiness adaptation in coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*). Data were collected across part of the natural range for a total of 643 individuals. A landscape genomic approach revealed 28 putative non-neutral genes, although a variance partitioning analysis indicated only moderate power of this gene set in explaining cold-hardiness-related phenotypic variation, and suggests many important genes await discovery. Integrating these results within the entire phenotype-genotype-environment spectrum allowed us to delineate the six most promising candidate genes under selection. By combining genomic, phenotypic and environmental data, this study attempts to gain insights in the genetic basis of key adaptations, which may ultimately aid forestry managers to establish resilient ecosystems in face of future climate change.

Keywords Coastal Douglas-fir · *Pseudotsuga menziesii* · Cold-hardiness · Adaptation · Landscape genomics

Introduction

Characterizing adaptive genetic variation and the concomitant evolutionary potential of a species is a major goal in conservation and evolutionary biology (Harrisson et al. 2014). More than two centuries of provenance trials in forestry have clearly showed strong adaptation of trees to their local environment (Langlet 1971). Climate change, however, may impose new phenotypic optima upon populations, while the lack of allelic diversity at loci underlying these key traits will hamper evolutionary responses thus creating local patterns of maladaptation. Trees are extremely vulnerable to rapid environmental change as predictions of climate-induced environmental shifts appear to be too fast and extensive to be countered by migration, plasticity, or adaptation (Aitken et al. 2008; Breed et al. 2011; Schoville et al. 2012; Alberto et al. 2013). Among other triggers, cold hardening in plants is initiated by cool temperatures at fall (Beck et al. 2004). Average temperatures are expected to increase in the near future and concomitantly could delay cold hardening (Guak et al. 1998), yet climate

Communicated by S. C. González-Martínez

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-017-1225-x) contains supplementary material, which is available to authorized users.

✉ Carl Vangestel
cvangestel@naturalsciences.be

[1] O.D. Taxonomy & Phylogeny, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000 Brussels, Belgium

[2] Terrestrial Ecology Unit, Ghent University, 9000 Ghent, Belgium

[3] Department of Plant Sciences, University of California, Davis, CA 95616, USA

[4] Department of Biology, Virginia Commonwealth University, Richmond, VA 23284, USA

[5] Department Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

[6] USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA

variability is predicted to increase as well and extreme frost spells will continue to occur (IPCC 2013). Hence, as a result of these climatological shifts, conifers may actually become more vulnerable to frost injuries (Bansal et al. 2015) and negatively impact sustainability of conifer forests. In addition, assisted migration has been put forward as a tool to mitigate detrimental effects of climate change, as Douglas-fir trees tolerate moderate distances between seed sources and target populations without suffering extensive cold damage (Bansal et al. 2015). As the predicted elevational shift in optimal growing conditions for Douglas-fir in 2060 (Rehfeldt et al. 2014; Bansal et al. 2015) might exceed contemporary tolerated seed transfer distances, careful consideration needs to be taken into account to what extent short and long term differences in cold tolerance are acceptable when planting tree seeds in the near future (Bansal et al. 2015). Hence, in the face of rapid climate change, knowledge of the spatial distribution of adaptive genetic variation on cold-hardiness becomes increasingly important and may improve reforestation guidelines through identification of appropriate seed stands, thereby facilitating the long-term persistence of tree species (Neale and Kremer 2011).

Environmental heterogeneity often creates strong local selective forces that shift allele frequencies for genes underlying fitness-related traits to new equilibria. Researchers have taken advantage of this tight link between the different components of the phenotype-genotype-environment triad to characterize local adaptation. Notwithstanding various logistic constraints (Sork et al. 2013), common garden experiments have successfully tackled the phenotype-environment axis by uncovering genetic clines and environmental gradients reflecting adaptation across a variety of tree species (Kremer et al. 2012). Seedlings originating from environmentally distinct sources are typically grown under similar conditions and fitness-related phenotypic traits, such as survival, growth, phenology, and stress tolerance, are related to sets of climatological conditions ((St Clair et al. 2005; González-Martinez et al. 2006). Landscape genomics (genotype-environment axis) and association/quantitative trait locus mapping (genotype-phenotype axis) aim to study adaptation in an explicit molecular framework. The latter approach links phenotypes to allelic variants of either pre-selected or a genome-wide set of genes (Sork et al. 2013; Pardo-Diaz et al. 2015). In the former, environmental gradients, are related to allele frequency distributions while correcting for demographic processes (Schoville et al. 2012). Ideally, to enhance the probability of including the appropriate environmental gradient while avoiding data dredging, environmental gradients could be brought forward as those identified as putative drivers of selective forces by previously conducted environment-phenotype association studies. While studies with an explicit focus on genotype-environment or genotype-phenotype associations have proven their value (Wheeler et al. 2005; Eckert et al. 2009a, 2010; Wegrzyn et al. 2010), each may differ in the sensitivity of picking up adaptive signals (Le Corre and Kremer 2012) causing incongruence between different approaches. In addition, most analyses studying these components of the phenotype-genotype-environment triad often require a multitude of statistical tests, urging researchers to deal with concomitant multiple testing issues and to adhere some cautiousness when interpreting results (Benjamini and Hochberg 1995). Hence, many have advocated integrating all facets of the phenotype-genotype-environment spectrum into a single exhaustive study to delineate a conservative, yet robust set of genes underlying phenotypic traits (Barrett and Hoekstra 2011; Fournier-Level et al. 2011; Hancock et al. 2011; Sork et al. 2013; Berg and Coop 2014; Lepais and Bacles 2014; Bragg et al. 2015; Yeaman et al. 2016). Empirical examples exploring these triad components have been described for a number of species such as *Alnus glutinosa* (De Kort et al. 2014), *Pinus lambertiana* (Eckert et al. 2015), and spruce (Holliday et al. 2010). These studies exemplify how a holistic approach across the phenotype-genotype-environment spectrum may aid in identifying key genes underlying phenotypic adaptation to local climate conditions.

A long-term study on coastal Douglas-fir (*Pseudotsuga menziesii var. menziesii*) provides a unique opportunity to illustrate how covering the entire phenotype-genotype-environment spectrum may enhance our understanding of genes underlying cold-hardiness associated traits. Coastal Douglas-fir is one of the most ecologically and economically valuable tree species in western North America. Its natural range covers the North American Pacific Coast, ranging from California, Oregon, Washington to British Columbia (St Clair et al. 2005). The range of coastal Douglas-fir includes a wide variety of habitats that differ in both temperature and moisture. Temperature appears to be one of the strongest abiotic selective forces and environmental constraints for this species, as populations along these gradients show marked differences in susceptibility to cold damage (St Clair 2006). Genecological studies of coastal Douglas-fir using a multi-year common garden approach have addressed the geographical, topographical and climatological patterns in a set of a priori selected cold-hardiness related traits (St Clair et al. 2005; St Clair 2006). The main results indicated winter temperatures and frost dates as key features to local adaptation of coastal Douglas-fir throughout its natural range (phenotype-environment associations). By complementing earlier common garden experiments on quantitative trait loci (Jermstad et al. 2001a; Jermstad et al. 2001b, 2003; Wheeler et al. 2005) and using the exact same phenotypic data as the genecological study, Eckert et al. (2009a) were able to highlight a set of 12 genes showing significant associations with 10 cold-hardiness related traits in an association mapping study (phenotype-genotype associations). Gene effects were typically small (1.9–

3.6% variance explained), suggesting a polygenic architecture of cold-hardiness-related traits.

The aim of this study is to expand our current view on adaptive genetic variation underlying cold-hardiness related traits in coastal Douglas-fir. We apply an exhaustive strategy of screening the entire phenotype-genotype-environment spectrum in order to delineate a robust set of candidate genes under selection. First, we focus on the environment-genotype associations by conducting environmental association analyses between allele variation of a priori selected target genes and a set of cold related environmental variables ('bottom-up' strategy sensu Sork et al. (2013)). Second, building upon these results we assess the relative contribution of this putative adaptive genetic component to cold-hardiness phenotypic variation. Thirdly, we combine the novel results of our genotype-environment analysis with the outcomes of genotype-phenotype (Eckert et al. 2009a) and environment-phenotype (St Clair et al. 2005) analyses that were previously obtained using the same sampling scheme, and assign those genes that show consistent relationships across the entire phenotype-genotype-environment spectrum as main candidates underlying adaptation to cold-hardiness.

Material and methods

Sample collection

A sample of 643 coastal Douglas-fir parent trees in Oregon and Washington was selected from previously published datasets on genecology and association mapping (St Clair et al. 2005, St Clair 2006, Eckert et al. 2009a, Eckert et al. 2009b). Samples east of the Cascade crest were not included in this study as it was previously shown that these samples inflated the number of significant phenotype-genotype associations (Eckert et al. 2009a). Samples were collected from natural stands throughout Washington and western Oregon, which covers part of the natural range of coastal Douglas-fir and encompasses a suite of environmental gradients (Fig. 1).

Environmental data

For each sampled tree, we collected both geographical (latitude, longitude, and elevation) and climate data. As seeds of trees were collected between 1966 and 1995, elevation adjusted climate data were obtained from the gridded 30-s arc ClimateWNA database (approximately 1 km resolution) using the reference period 1961–1990 (Wang et al. 2012). A total of 28 climate variables were assessed. Annual and seasonal ('sm' = summer, 'at' = autumn, 'wt' = winter, 'sp' = spring) climate data encapsulated mean coldest temperature (Tmin), amount of snow (PAS), number of frost-free days (NFFD), degree-days below 0 °C (DD0), and degree-days above 5 °C (DD5). Annual variables further comprised data on differences between coldest and warmest month (continentality) (TD), begin (bFFP) and end (eFFP) of frost-free period, duration of frost-free period (FFP) and the extreme minimum temperature over 30 years (EMT). Correlations between environmental variables were graphically explored using the software package Corrgram (Friendly 2002) in the R environment (R Core Team 2013). A principal component analysis was performed to reduce the high dimensionality of the environmental data using the PRINCOMP procedure in SAS 9.3. (SAS Institute 2008, Cary, NC, USA). Principal components (PC) characterized by eigenvalues (λ) > 1 were retained.

Candidate gene selection and SNP genotyping

SNP data were obtained from a previous phenotype-genotype association study (Eckert et al. 2009a, b). In brief, candidate genes putatively involved in cold-tolerance mechanisms were a priori selected by Eckert et al. (2009a, b) based on a threefold criterion: (i) collocation of genes with QTLs for cold-hardiness (Wheeler et al. 2005), (ii) genes involved in physiological metabolic pathways of a cold tolerance response and (iii) genes differentially expressed in microarray studies in *A. thaliana* (Lee et al. 2005). Available coastal Douglas-fir expressed sequence tag (EST) libraries were screened using a standard BLAST algorithm and a threshold of 1E-10 was set to call putative orthologs. Of the original 384 SNPs that Eckert et al. (2009a, b) selected, 228 SNPs that met the predefined quality thresholds (0.35 and 0.85 for respectively the indices $GenCall_{50}$ and call rate) were retained by the authors. These SNPs were located in 117 distinct genes. As no maternal needle tissue was available for these preselected trees, DNA was extracted from haploid megagametophyte tissue of 10 seeds to infer the diploid maternal genotype. We opted to conduct an environmental association analysis based on maternal genotypes rather than progeny genotypes, as the geographical source of the paternal contribution to the offspring's genotype remains unknown, which may hamper the identification of patterns of adaptive genetic variation. In contrast, maternal analyses harbor a direct link between geography and genotype and are therefore not flawed by such deficiencies in study design. Genotyping was carried out on an Illumina GoldenGate genotyping platform at the DNA Technologies Core Facility located at the Genome Center of the University of California, Davis. Gene annotations are described in Eckert et al. (2009a).

Patterns in neutral genetic variation

Spatial patterns of neutral genetic structure were assessed by performing a principal coordinates analysis (PCoA) on an individual-by-individual genetic distance matrix across all SNPs. Squared genetic distances d^2 for a locus with alleles i and j were defined as $d^2(ii, ii) = 0$, $d^2(ij, ij) = 0$, $d^2(ii, ij) = 1$ and

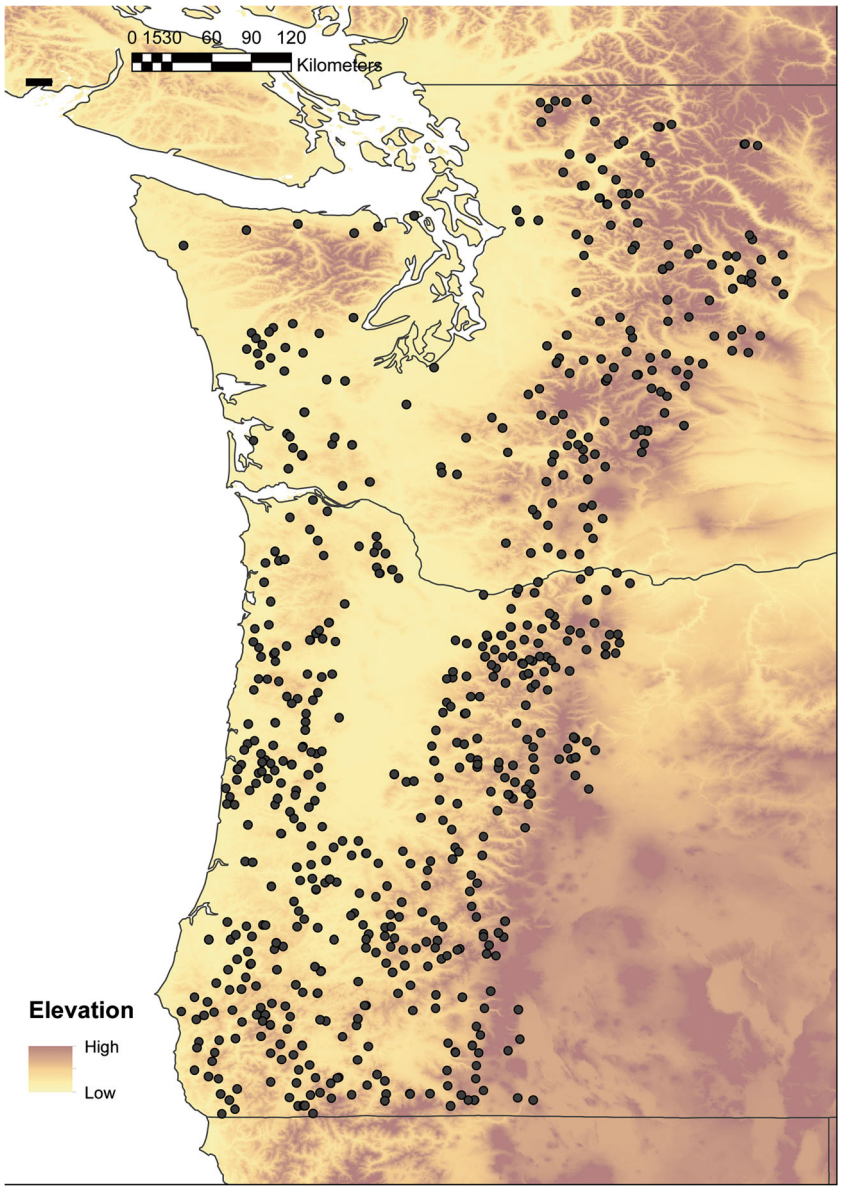

Fig. 1 Source locations of 643 sampled coastal Douglas-fir trees. Elevation data was downloaded from the WorldClim database (Hijmans et al. 2005)

$d^2(ii, jj) = 4$. We further explored whether genetic similarity between individuals decreased with geographical distance ("isolation-by-distance") by conducting a spatial autocorrelation analysis (Peakall et al. 2003; Vekemans and Hardy 2004). Correlation coefficients between individuals were depicted as a function of increasing inter-individual geographical distance and evidence of positive genetic structure was considered when confidence intervals (CIs), based on 9999 bootstrap replications across SNPs, did not contain the null value. All analyses were performed in GenAlEx 6.501 (Peakall and Smouse 2006).

Characterization of adaptive genetic variation

To identify loci that exhibit strong deviations from patterns of neutral genetic differentiation, we made use of two complementary analytical methods. Firstly, we applied an individual-based latent fixed mixed model (LFMM) to associate SNP frequencies with environmental variation while accounting for neutral population structure (Frichot et al. 2013). The latter is accomplished by modeling an a priori defined number of latent factors (K) resembling shared ancestral or demographic history. As previous studies did not indicate strong genetic population structure across the sampled area, we opted to perform a sensitivity analysis by allowing the number of latent factors to vary from 2 to 15, and assessed the influence of K on outlier identification (Eckert et al. 2009a; Krutovsky et al. 2009). For each value of K, we ran the model five times, using 1000 sweeps for burn-in and 10,000 additional sweeps as run-length. We obtained the median Z-value of the five replicated runs for each locus separately, and SNPs were considered as non-neutral when characterized by a median $|Z|$-value > 3. Secondly, for each SNP, we regressed genotypic frequencies

against environmental factors using a multinomial logistic regression model (MLR). Loadings along the five most important genetic principal coordinate axes (see above) were included to account for the neutral covariance structure across genotype frequencies. Notably, the number of PCoA axes included as covariates had no substantial effect on the outcome of the analyses, as test statistics (χ^2) for models including 1 to 8 PCoA axes were highly correlated (Pearson's correlation coefficients, r: 0.75–0.94). Multinomial logistic regression models were fitted using the LOGISTIC procedure in SAS 9.3. (SAS Institute 2008, Cary, NC, USA). To account for the multitude of tests being conducted, we specified the false discovery rate at 0.01 using the qvalue package implemented in Bioconductor (Storey 2002; Storey and Tibshirani 2003). In both LFMM and MLR models, genetic variation was associated with the most important composite environmental principal components variables (i.e., $\lambda > 1$). For reasons outlined below, we additionally included important univariate environmental variables that previously showed strong associations with cold-hardiness-related phenotypic traits ($|r| > 0.4$) (St Clair et al. 2005; St Clair 2006).

Phenotype-genotype-environment spectrum

Phenotypes are the result of their genetic makeup, environmental influences and the interaction between them. To estimate the relative contribution of genes and environment to cold-hardiness-related phenotypic variation, we performed a multivariate canonical redundancy analysis (RDA) of outlier genetic, geographic and environmental data on phenotypic variables. Phenotypic data of the same maternal trees as the ones used in this study were obtained from a previously published association mapping (Eckert et al. 2009a) and genecological study (St Clair et al. 2005; St Clair 2006). These authors gathered phenotypic data on cold-hardiness-related traits by means of a common garden experiment. Twenty progeny of each maternal tree were grown for 2 years, while data on emergence, cold injury, phenology, growth and resource partitioning were recorded (Supporting Information Table S1). Rates of emergence were estimated based on the cumulative number of seedlings out of 20 that emerged (EMEAN). Bud set was measured at the end of the first (BS1) and second (BS2) growing season, while bud burst was quantified at the beginning of the second growing season (BB2). Cold-hardiness on bud (BCI), needle (NCI) and stem (SCI) was assessed using temperature-controlled freezers following the protocol outlined in Aitken and Adams (1996). After harvesting, seedlings were measured for root length (RTLG), root-to-shoot ratio (RTSH) and diameter to height ratio (TAPER). Breeding values were calculated as family means, while genetic checklots were used to standardize year-to-year variation. While numerous other cold-hardiness-related traits were measured in the original studies, we restricted ourselves to these phenotypic traits that were characterized by either a significant phenotype-genotype (Eckert et al. 2009a) or phenotype-environment (St Clair et al. 2005) association. Environmental variables comprised the most important environmental PC axes, while latitude and longitude constituted the geographical component, and the genetic dataset contained allelic variation of all outlier SNPs as indicated by the LFMM analysis or MLR. A RDA-based variance partitioning analysis was performed by specifying full and partial models using the vegan package in R (Oksanen et al. 2017). Prior to analysis, both dependent and explanatory variables were standardized.

Finally, to gain a deeper insight into the genomic regions that are tightly linked to cold-hardiness-related traits, and hence putative important contributors of the phenotype, we combined results from our landscape genomic study (genotype-environment) with those from the previously published association mapping (genotype-phenotype) and genecological (environment-phenotype) studies (St Clair et al. 2005; St Clair 2006; Eckert et al. 2009a). We assessed to what extent results remained consistent across the phenotype-genotype-environment spectrum (i.e., whether both components of the phenotype-genotype association are linked to the same environmental variable). To make our results more comparable with the genecological study, we limited our focus to the same environmental variables as the ones used in St Clair et al. (2005), namely latitude, elevation, frost related variables, minimum temperature in winter and the difference between the coldest and warmest month. We excluded environmental composite measures as these were lacking in the study of St Clair et al. (2005, 2006).

Results

Environmental analysis

Environmental variables showed strong correlations among all pairwise comparisons resulting in two clusters of environmental variables, where within-cluster associations were positive and between-cluster associations were negative. The first cluster comprised Tmin, DD5, MCMT, EMT, NFFD and eFFP, while the other cluster encapsulated PAS, DD0, TD and bFFP (Fig. S1). In concordance with the strong correlations among environmental variables, the first PC explained 77.4% of the total variance ($\lambda_1 = 24.0$). We further retained the second, third and fourth PCs as composite environmental variables, which additionally captured 7.2% ($\lambda_2 = 2.2$), 6.6% ($\lambda_3 = 2.0$) and 3.7% ($\lambda_4 = 1.1$) of the overall variance, respectively. All seasonal variables showed strong positive correlations with PC1 with the exception of DD0 and PAS, which were negatively correlated ($0.75 < |r| < 0.97$ with median value of 0.93). Similar

trends were observed for the annual variables with positive correlations for MCMT, DD5, NFFD, eFFP, FFP and EMT and negative ones for TD, DD0, bFFP and PAS ($0.65 < |r| < 0.98$ with median value of 0.93). The first PC showed a strong negative association with longitude ($r = -0.73$) and elevation ($r = -0.75$) reflecting the mountainous region at the east side of the coastal Douglas-fir distribution (Supporting information Fig. S2 and S3). Trees at these locations faced harsh winter conditions (i.e., low winter temperatures, high amount of snow fall and extensive frost periods). Scores for PC2 were positively correlated with PAS (annual and seasonal) ($0.39 < |r| < 0.45$) and negatively with TD, bFFP and DD5_sm ($0.27 < |r| < 0.36$). Furthermore, PC2 showed a positive correlation with latitude and a negative one with elevation, contrasting high elevation in the south against lower elevations in the north. PC3 mainly contrasted regions characterized by warm summers and cold winters against those enduring less extreme differences between both seasons. It showed positive associations with NFFD_sm, Tmin_sm, DD5_sm and TD ($0.36 < |r| < 0.61$), negative correlations with NFFD_wt, Tmin_wt and DD5_wt ($0.16 < |r| < 0.32$) and was oriented from high elevation regions at the southwest towards low elevation ones in the northeast. Most variables were positively correlated to PC4, with the exception of especially frost-free-related variables, and this axis showed a strong negative association with latitude (Figs. S2 and S3). A summary of the interpretation of each PC is given in Table S2 (Supporting information).

Neutral population structure

The PCoA indicated the absence of strong genetic population structure. Groups of geographically proximal individuals did not cluster together, while the cumulative variation explained by the three most important PCoA axes only accounted for 6.8% of the total genetic variation present in the data (Supporting Information Fig. S4). These findings were corroborated by the autocorrelation analysis. Confidence intervals around correlation coefficients, however, did not include the null value at smaller scales, but all correlation coefficients were only marginally larger than zero (Fig. 2). Both analyses denote substantial amount of gene flow across the sampled area and/or recent divergence among populations. To ascertain that the inclusion of putative adaptive SNPs did not confound estimates of neutral population structure, we reran both analyses after excluding 35 outlier SNPs (see below). Scores along the first two and most important PCoA axes in both data sets were highly correlated ($r > 0.95$), while correlation coefficients between scores along the 3rd–5th PCoA axes ranged from 0.36 to 0.76. Removal of outlying genetic variation did not affect the scale at which confidence intervals around spatial autocorrelation coefficients included the null value (i.e., 500 km) nor did it affect the pattern of the autocorrelogram

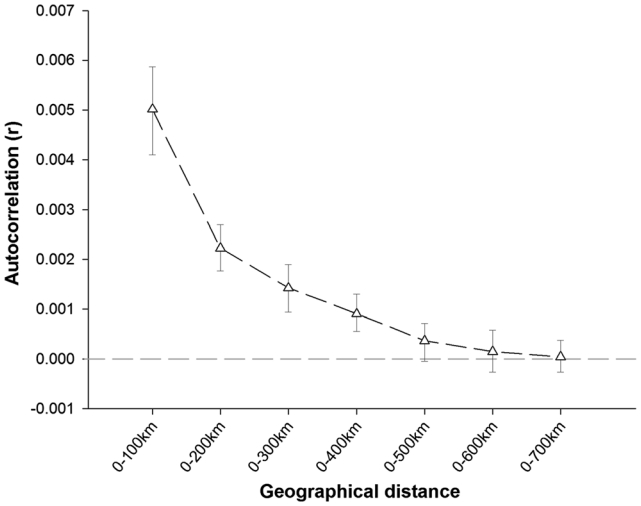

Fig. 2 Spatial autocorrelation plot illustrating the gradual decay of positive, albeit very small, genetic structure with increasing geographical distance. Error bars depict 95% CI as determined by 9999 bootstraps across SNPs

itself. As such, quantification of neutral population structure did not seem to be influenced by outlier SNPs.

Adaptive genetic variation

The LFMM sensitivity analysis indicated that the value of K did not influence the outcome of the analysis, which was not unexpected given the lack of strong population structure. For each environmental factor, Z values of individual SNPs remained consistent across $K = 2$ to $K = 15$ (r: 0.94–0.99 with a median value of 0.97). Hence, we opted to report median Z-values across all K values for each SNP-environment association. Considering composite and univariate environmental variables, the LFMM analysis classified 19 SNPs located in 19 genes as putatively non-neutral (Table 1). MLR models revealed 16 putative selective SNPs distributed over 14 genes (Table 1, Table S3 Supporting Information). Interestingly, the outcome of both analyses showed only a moderate overlap, as five genes were consistently assigned as putative non-neutral in both LFMM and MLR. These included a dehydration-induced protein (ERD15), two unknown hypothetical proteins (ES420757.1, ES421311.1), 4-coumarate:CoA ligase 1 (4CL1) and a SOUL heme-binding protein (Pm_CL783Contig1) (Table 1). While univariate environmental associations corroborated most of the significant associations between genes and environmental PC axes, some genes were exclusively linked to environmental PCs. For example, 60S ribosomal protein L31a (60s RPL31a) showed strong affinity to PC3, a MYB-like transcription factor (CN637306.1) was strongly associated to PC3 and PC4, while genetic variation at a cysteine proteinase (Pm_CL135Contig1) was linked to PC3. The correlative nature among environmental variables was again reflected by the results of the MLR and the LFMM analyses, as 31% (MLR) and

Table 1 Outlier SNPs showing significant associations with environmental variables using LFMM ($|z| > 3$) and MLR (FDR < 0.01). SNP position refers to the base pair position within each gene. See text for abbreviations of each environmental variable

Locus ID	SNP position	Gene product	Environmental variable	
			LFMM	MLR
4CL1	34	4-coumarate: CoA ligase 1	bFFP/eFFP/FFP/TD/Tmin_wt	PC2/latitude
4CL1	520	4-coumarate: CoA ligase 1		PC3
60s RPL31a	295	60s ribosomal protein L31a		PC3/longitude
60s RPL31a	418	60s ribosomal protein L31a		PC3/longitude
60s RPL31a	55	60s ribosomal protein L31a		PC3/longitude
aba	609	Abscisic acid inducible protein		latitude
apx	738	Ascorbate peroxidase	Latitude	
CN637306.1	381	MYB-like transcription factor		PC3/PC4
CN637339.1	337	Unknown hypothetical protein	bFFP/eFFP/FFP/TD/Tmin_wt	
CN639087.1	366	LRR receptor-like protein kinase	Longitude/PC3/PC4	
CN639480.1	430	2-hydroxyacid dehydrogenase		Latitude
CN640485.1	250	HNH endonuclease domain containing protein		Longitude
CN640521.1	479	DNA-binding bromodomain-containing protein	bFFP/eFFP/FFP/Tmin_wt	
erd15	174	Early response to dehydration protein	Latitude	
erd15	327	Early response to dehydration protein		Latitude
erd15	635	Early response to dehydration protein		PC2/PC3/latitude/longitude
LEA-EMB11	372	Late embryogenesis abundant EMB11 like protein	Tmin_wt	
LEA-EMB11	263	Late embryogenesis abundant EMB11 like protein		latitude
ES420757.1	86	Unknown hypothetical protein		Latitude
ES420757.1	120	Unknown hypothetical protein	bFFP/eFFP/FFP/Tmin_wt/TD/PC3	Latitude
ES421311.1	369	Unknown hypothetical protein	Latitude	PC2
Pm_CL135Contig1	715	Cysteine proteinase		PC3
Pm_CL1811Contig1	258	Chromatin remodeling ATPase	bFFP/eFFP/FFP/Tmin_wt	
Pm_CL783Contig1	117	SOUL heme-binding family protein	Longitude/PC3/PC4	
PM_CL783Contig1	212	SOUL heme-binding family protein		PCA1/Tmin_wt/eFFP/FFP
ES420603.1	56	Dehydrin-like protein	Latitude	
ES419223.1	449	Phytosulfokine precursor	eFFP/FFP/Tmin_wt	
ES419242.1	416	Response regular protein	Longitude/TD/Tmin_wt/PC3/PC4	
CN639236.1	518	Guanine nucleotide = binding beta subunit protein		Latitude
CN636795.1	990	Xyloglucan:xyloglucosyl transferase	TD	
sSPcDFD040B03103	274	MADS-box transcription factor		PC1/PC2/Tmin_wt/eFFP
sSPcDFE025C06206	485	Purple acid phosphatase	Latitude	
sSPcDFE028B10110	198	β-amylase	bFFP	
sSPcDFE049E11411	44	Pentatricopeptide (PPR) containing protein	bFFP	
ubq	814	Polyubiquitin	TD/Tmin_wt/FFP/PC3	

53% (LFMM) of the SNPs were associated with more than one environmental variable. For example, genetic variation at 4-coumarate:CoA ligase 1 (4CL1) was associated to bFFP, TD, eFFP, FFP and Tmin_wt, where the former two environmental variables were positively correlated with one another, but negatively with the latter three variables. Combined, these two methods identified 35 SNPs distributed over 28 genes.

The phenotype-genotype-environment triad

Environmental PCs, outlier genetic variation and geography (i.e., latitude and longitude) together explained 44.3% of cold-hardiness-related phenotypic variability. Of this amount, 22% was exclusively due to the genetic component, while the joint action of outlier loci and environment

and/or geography accounted for another 18% of the phenotypic variation. Only 16 and 2% could be exclusively attributed to environmental and geographical variables respectively, while a large proportion of phenotypic variability (41%) was explained by the joint action of geography and environment alone (Table 2).

Finally, we integrated results of our landscape genomic study into those of a previously published association mapping (Eckert et al. 2009a) and genecological study (St Clair et al. 2005; St Clair 2006). Hence, outcomes of the genotype-environment association analysis (this study) allowed us to complement previously published genotype-phenotype and environment-phenotype association studies, and made it possible to narrow our focus to those genes that are part of an entirely closed circular statistical network (i.e., genes associated to a specific phenotype and environment, which on their turn were related to each other as well). When considering only univariate environmental variables, LFMM and/or MLR were able to identify 22 genes related to 7 environmental variables, while 32% of those genes were previously shown to be associated to cold-hardiness-related traits (Eckert et al. 2009a) (Fig. 3). Genotype frequencies of late embryogenesis abundant protein (LEA-EMB11) were related to latitude, taper and timing of the first green needles from the terminal bud. The latter two phenotypic traits were strongly related to latitude thereby closing the phenotype-genotype-environmental statistical network. Similarly, genetic variation at the SOUL heme-binding protein (Pm_CL783Contig1) was associated to the duration of the frost, date of last frost and minimum temperature at winter, while bud set showed a strong link with all environmental variables and the SOUL heme-binding protein. Genetic variation at a guanine nucleotide-binding beta subunit protein (CN639236.1) showed strong association to latitude and rate of emergence, while latitude influenced the rate of emergence.

Table 2 Variance partitioning analysis associating cold-hardiness-related phenotypic variation to environmental outlier SNPs, spatial and environmental variation. Total variation explained by the model was further decomposed in genetic, environmental and geographic components using multiple partial RDAs

	Inertia	Proportion
Total phenotypic variation	9.21	1.00
Total variation explained by model	4.09	0.44
Total variation explained by model	4.09	1.00
Pure genotype	0.91	0.22
Pure environment	0.65	0.16
Pure geography	0.09	0.02
Joint genotype/environment	0.10	0.02
Joint genotype/geography	−0.02	0.00
Joint geography/environment	1.62	0.41
Joint genotype/geography/environment	0.67	0.16

Genotype frequencies at 4-coumarate:CoA ligase 1 (4CL1) were associated to latitude and levels of needle cold injury, while the level of needle damage after frost treatment was correlated to latitude. Genetic variation at an unknown hypothetical protein (CN637339.1) was tightly linked to minimum temperature at winter, begin and end of the frost-free period and stem cold injury, while the latter was also correlated to all these environmental variation. Finally, genotypes of sSPcDFD040B03103 were associated with minimum temperature at winter, end of the frost-free period and cold injury at stem and needles. Damage at these tissues was likewise strongly affected by minimum temperature at winter and end of frost-free period (Figs. 3, S4).

Discussion

We provided a comprehensive synthesis of an association mapping, genecological and landscape genomic study integrating genotypic, environmental and phenotypic data to gain insights into the genetic basis of cold-hardiness adaptation in coastal Douglas-fir. Assembling the main findings of all three studies resulted in an interlinked phenotype-genotype-environment network highlighting a set of genes as putative key targets of natural selection.

Landscape genomics

Landscape genetics is most effective when based on a combination of complementary analytical methods (Jones et al. 2013). We found substantial discordance among results of different gene scan methods, although such discrepancy was not entirely unexpected as each method relies on different premises (Prunier et al. 2012; de Villemereuil et al. 2014; Nadeau et al. 2016). Of the 117 screened genes, both environment-based methods combined identified 28 environmentally associated genes, yet only five were corroborated by both methods. Although both analyses apply different methods to account for neutral population structure, these dissimilarities most likely do not account for the lack of congruence between methods, as this tree species exhibits little population structure across the sampled range, rendering the outcome rather insensitive to demographic processes as substantiated by the sensitivity analysis. An important difference between both methods, however, is that LFMM models allele frequencies instead of genotype frequencies, which may, depending on the underlying evolutionary model, affect the sensitivity of association tests. In addition, LFMM and MLR assume that fitness effects of each allele at a dialleleic locus are highest at opposit ends of an environmental gradient. Yet, it remains plausible that allele-associated fitness effects have local optima and are characterized by bell shaped curves at specific distinct environmental values (Prunier et al. 2012).

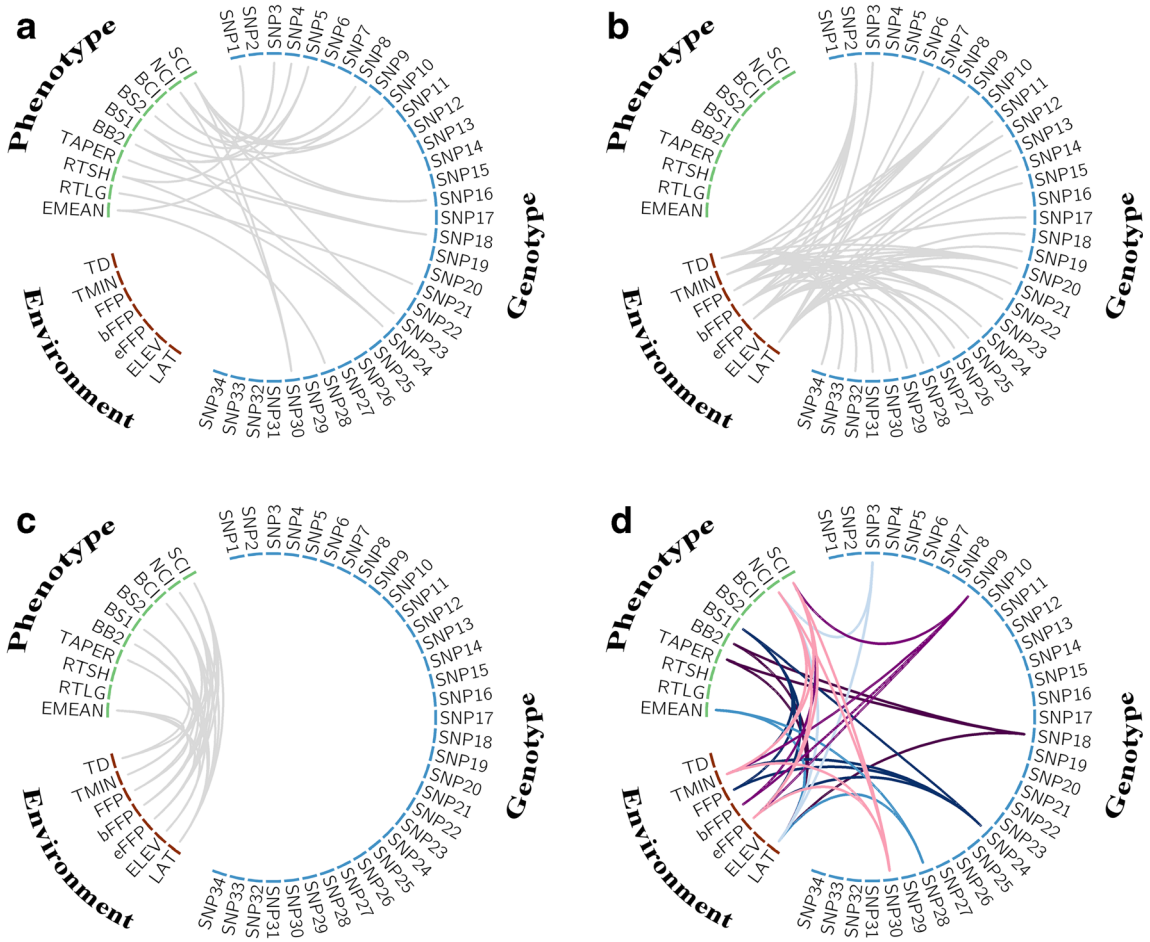

Fig. 3 Summary of the main findings among cold-hardiness-related traits, genes and environment. Significant associations are depicted as a connection between corresponding phenotype, genotype or environmental variables. **a** Phenotype-genotype associations. **b** Genotype-environmental associations. **c** Phenotype-environmental associations. **d** Integrating phenotype-genotype-environmental associations for six selected genes. See text for abbreviation of phenotypic and environmental variables, see Table S4 Supporting Information for link between SNP index and gene product)

Such scenarios are not characterized by a smooth shift from one allelic variant to the other along an environmental cline, and may hamper and differentially affect the detection rate of adaptive SNPs via LFMM and MLR (Narum and Hess 2011; Prunier et al. 2012). Here, estimating standardized transformed allele frequencies, for which the effect of population structure has been removed, and subsequent fitting of non-linear models on those modified allele frequencies might allow for the detection of non-linear trends (Günther and Coop 2013).

Integrating phenotypic, genetic and environmental variation

It is well known that neutral population structure may confound genotype-environment associations, especially when local gene flow creates an isolation-by-distance cline that coincides with environmental gradients. In such instances, it becomes notoriously difficult to dissect effects of demography from local adaptation (Nadeau et al. 2016). However, genetic population structure was weak in this study, and it did not correlate with environmental gradients. Although we only had access to a set of a priori selected SNPs, a previous study using putative neutral microsatellites confirmed the absence of strong genetic population structure in this species (Krutovsky et al. 2009), rendering it unlikely that population structure exerted strong effects on signatures of local adaptation. Moreover, in the event spatial patterns of adaptive and neutral variation did coincide, association analyses corrected for neutral population structure would result in a conservative estimation of the number of adaptive SNPs rather than an overestimation.

Allele surfing, a process in which novel neutral mutations may thrive and spread along a wave of an expanding population may generate allele frequency gradients as well (Edmonds et al. 2004; Klopfstein et al. 2006; Travis et al. 2007). New mutations that arise at the edge of population expansions can rise to high frequencies through the concerted

efforts of founder effects and genetic drift. When the directionality of such an expansion wave coincides with an environmental gradient, clines of neutral allele frequencies can be strongly associated to environmental variation, thereby blurring the distinction between effects of selection and demography (Klopfstein et al. 2006; Travis et al. 2007; Nadeau et al. 2016). A combinatorial approach as the one applied here may prevent researchers to erroneously perceive neutral genes as adaptive ones. Common garden experiments may complement landscape genomics in dissecting neutral alleles from adaptive ones. However, while the strategy to focus on genes that form complete phenotype-genotype-environmental networks may be most suitable to differentiate the most promising candidate genes from false positives ones, some caution is warranted. Missing links may arise frequently for several reasons. Statistical power of common garden experiments strongly hinges on the genetic variation present in experimental samples, the number of experimental environments and the extent to which phenotypic plasticity blurs genetically determined phenotypic differences (Neale and Kremer 2011; De Kort et al. 2014).

Polygenic adaptation

Gene-based analytical approaches as presented here focus on single-locus effects while cold-hardiness, like many other phenotypic traits in trees, is assumed to be polygenic (Jermstad et al. 2001b; Eckert et al. 2009a; Holliday et al. 2012). Under such circumstances selection may affect hundreds or even thousands of loci simultaneously, but only moderately alter the allele frequency at a given single locus (Le Corre and Kremer 2003). Although the RDA indicated that the outlying genetic component already accounted for 40% of the phenotypic variation that could be explained, we may have missed many important genes underlying these cold-hardiness-related traits given that conifers are often strongly adapted to local environmental conditions (Neale and Kremer 2011),. Adaptation may be driven by covariance among key loci underlying the adaptation rather than strong differentiation at the individual loci themselves, increasing the probability that small-effect single locus signatures remain undetected from their neutral genomic background (Le Corre and Kremer 2012; Berg and Coop 2014; de Villemereuil et al. 2014; Rajora et al. 2016; Lind et al. 2017). This has rendered the identification of adaptive loci a challenging task and, hence, many authors have advocated the use of multivariate statistics or random forests to unravel the complexity underlying adaptive genetic polymorphisms (Holliday et al. 2012; Le Corre and Kremer 2012; Sork et al. 2013; Forester et al. 2017), which were successfully applied in Sitka spruce (Holliday et al. 2012) and black alder (De Kort et al. 2014). Recently, Berg and Coop (2014) outlined an analytical framework to unravel genetic signatures of polygenic adaptation based on genome wide association studies (GWAS). They developed a set of robust tests on genetic values (i.e., the combinatorial effect of allele frequencies and effect sizes of GWAS SNPs). By exploring the level of over-dispersion of transformed genetic values, and thus taking into account an arbitrary relatedness structure among populations (Coop et al. 2010; Günther and Coop 2013), they were able to identify a concerted shift in allele frequencies across many a priori selected candidate loci. These methods require a large set of control SNPs to reliably estimate variance-covariance matrices of allele frequencies and null distributions of test statistics. Such data set was unfortunately not at our disposal. Nevertheless, such approaches should be considered in the future to further complement single-locus analyses as presented here. In this study, we aim to weigh the relative importance of a single specific trait-associated SNP, thereby focussing on genes exerting large effects, and moreover allowing us to integrate previously published single-SNP analyses into our own analytical framework.

Maternal breeding values

In contrast to other studies (e.g., De Kort et al. 2014), we were able to take advantage of the fact that maternal phenotypes were estimated as breeding values (St Clair et al. 2005, 2006), a traditional metric in quantitative forest genetics. Breeding values provide an estimate of the additive genetic component of the target phenotype, the genetic source that natural selection can act upon, and is evaluated based on the performance of multiple progeny (Falconer and Mackay 1996; Neale and Kremer 2011). Breeding values thus heavily rely on the accurate and precise estimation of family means. Accuracy, the difference in estimated and true breeding value, strongly hinges on the number of test sites and progeny per parent included in the analysis. Moreover, differences in accuracy between estimates based on high and low number of offspring are exacerbated by low trait heritability (White et al. 2007). In coastal Douglas-fir, narrow-sense heritability estimates vary between traits and seasons, ranging from low in fall, for budset and needle cold injury, to high in spring for budburst (Aitken and Adams 1996; O'Neill et al. 2000; Bigras and Colombo 2001). Such variation in heritability would urge caution when interpreting maternal breeding values for some traits if based on small family sizes, yet phenotypic values of 20 offspring were considered in this study (St Clair et al. 2005, 2006). Such sample sizes warrant accurate breeding value estimates for all traits and similar benefits from large family sizes on the precision of estimates of maternal breeding values are expected. Our sample size falls well within the range of 10 to 20 offspring per parent that has been suggested as sufficient for reliable estimates (Fins et al. 1992).

Identifying phenotype-genotype-environment networks

Previous association mapping and genecological studies (St Clair et al. 2005; Eckert et al. 2009a) revealed a two-way interaction between 11 genes, environment and cold-hardiness related phenotypes. Of those 11 candidate genes, our landscape genomic study further confirmed 6 genes, coding for late embryogenesis abundant protein (LEA-EMB11), (4-coumarate:CoA ligase 1 (4CL1), guanine nucleotide-binding beta subunit protein (CN639236.1), an unknown hypothetical protein (CN637339.1), a SOUL heme-binding family protein (Pm_CL783Contig1) and MADS-box transcription factor (sSPcDFD040B03103) to be part of a closed statistical circuit (i.e., when both components of the phenotype-genotype association are linked to the same environmental variable). Most of these genes show direct links with cold-induced physiological pathways.

Frequencies for SNPs in the LEA protein were highly dependent on latitude and linked to both timing of needle appearance (bud burst) and the ratio of diameter to height of 2nd year seedlings (taper). Assigning a direct causative relationship between allelic variation at the LEA protein and both traits remains elusive, however, even if the extent of linkage disequilibrium in natural conifer populations is expected to be extremely small and target SNPs could therefore very likely be situated within a gene (Neale and Savolainen 2004). Conifers could adapt to extreme freezing conditions by both avoiding freezing and increase their resilience to frost damage. At higher latitude, needles at the terminal bud appear to emerge later and taper to be reduced. These traits resemble a trade-off between postponing the initiation of a growing season and reducing the probability of extensive frost damage during early spring. However, exposure to low temperatures often induces biochemical changes such as an up-regulation of LEA proteins in a variety of plants (Hannah et al. 2005; Lorenz et al. 2006). These hydrophilic proteins purportedly mitigate the disruptive effects on lipid bilayers during frost through membrane stabilization and prevention of cellular crystallization (Garay-Arroyo et al. 2000).

As all trait responses may be interrelated it remains problematic to assign the true phenotypic effect of allelic variation observed at a locus in observational studies. Albeit the functional role of SOUL heme-binding proteins in cold tolerance remains rather elusive, they were characterized by a marked upregulation during cold stress in *Thellungiella halophile* (Gao et al. 2009). Guanine nucleotide binding proteins act as important elements of various cell-signaling cascades, including stress-induced responses (Colaneri and Jones 2014). Cold stress treatment have shown to induce strong alterations in transcription rates of this gene, where cyclic periods of upregulation were alternated with a period of downregulation (Yadav et al. 2014). To counter frost-induced damage to cell membranes, plants may change the lignin content and composition as a defense mechanism (Moura et al. 2010). Previous research confirms that cold treatments triggers the expression of 4-coumarate:CoA ligase 1, a key enzyme in the lignin biosynthesis, and that its gene products accumulated as response to cold stress (Moura et al. 2010; Chowdhury et al. 2013). MADS-box transcription factors are involved in various aspects of plant developmental processes. MADS-box genes constitute key determinants of the responsiveness to vernalization by regulating a cascade of other genes that initiate flowering (Trevaskis et al. 2003; Deng et al. 2015; Li et al. 2015). In addition, several members of this gene family target genes that modulate freezing tolerance and hence showed differential expression in response to cold stressors (Arora et al. 2007; Deng et al. 2015; Li et al. 2015).

Conclusion

To conclude, this study exemplifies a detailed approach to identify candidate genes under selection by integrating multiple disciplines. Merging landscape genomics, common garden experiments and association mapping allowed us to delineate a robust set of single-locus associations throughout the phenotype-genotype-environment spectrum, strongly suggesting these genomic regions are involved in adaptation to cold-hardiness. Ultimately, insights in the genetic basis of cold-hardiness adaptation and other important traits will allow forestry managers to select the most pre-adapted seeds and establish resilient ecosystems in face of future climate change. However, this list should not be considered as exhaustive but rather as an important starting point that warrants further research, as outlier loci only explained a moderate fraction of observed phenotypic variation. Such an outcome underpins once again the need for conducting large-scale genomic studies, which will soon become feasible as sequencing costs keep decreasing. Yet, we believe that only within a holistic and integrative framework such as the one shown here, will large-scale sequencing efforts pave the road for elucidating the genetic mechanisms underlying important traits such as cold-hardiness.

Acknowledgements We are indebted to John Liechty for his invaluable bioinformatics assistance. Funding for this project was made available through a U.S. Department of Agriculture National Research Initiative Plant Genome grant (04-712-0084). A postdoctoral fellowship from the Belgian American Educational Foundation (BAEF) and a travel grant from the Research Foundation–Flanders (FWO) was awarded to CV.

Data archiving statement Genotypic, phenotypic and environmental data has been deposited in the Figshare Database and has been made accessible using the provisional DOI https://doi.org/10.6084/m9.figshare.5167657.v1.

References

Aitken SN, Adams WT (1996) Genetics of fall and winter cold hardiness of coastal Douglas-fir in Oregon. Can J For Res 26(10):1828–1837. https://doi.org/10.1139/x26-208

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1(1):95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, Savolainen O (2013) Potential for evolutionary responses to climate change—evidence from tree populations. Glob Chang Biol 19(6):1645–1661. https://doi.org/10.1111/gcb.12181

Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, Kapoor S (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8(1):242–242. https://doi.org/10.1186/1471-2164-8-242

Bansal S, St Clair JB, Harrington CA, Gould PJ (2015) Impact of climate change on cold hardiness of Douglas-fir (*Pseudotsuga menziesii*): environmental and genetic considerations. Glob Chang Biol 21(10):3814–3826. https://doi.org/10.1111/gcb.12958

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12(11):767–780. https://doi.org/10.1038/nrg3015

Beck EH, Heim R, Hansen J (2004) Plant resistance to cold stress: mechanisms and environmental signals triggering frost hardening and dehardening. J Biosci 29(4):449–459. https://doi.org/10.1007/BF02712118

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B-Methodol 57:289–300

Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genet 10(8):e1004412. https://doi.org/10.1371/journal.pgen.1004412

Bigras FJ, S Colombo 2001 Conifer cold hardiness, Dordrecht, DOI: https://doi.org/10.1007/978-94-015-9650-3

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. New Phytol 207(4):953–967. https://doi.org/10.1111/nph.13410

Breed MF, Ottewell KM, Gardner MG, Lowe AJ (2011) Clarifying climate change adaptation responses for scattered trees in modified landscapes. J Appl Ecol 48(3):637–641. https://doi.org/10.1111/j.1365-2664.2011.01969.x

Chowdhury MEK, Choi B, Cho BK, Kim JB, Park SU, Natarajan S, Lim HS, Bae H (2013) Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. Plant OMICS 6:254–262

Colaneri AC, Jones AM (2014) The wiring diagram for plant G signaling. Curr Opin Plant Biol 22:56–64. https://doi.org/10.1016/j.pbi.2014.09.004

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. Genetics 185(4):1411–1423. https://doi.org/10.1534/genetics.110.114819

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J (2014) Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. Mol Ecol 23(19):4709–4721. https://doi.org/10.1111/mec.12813

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? Mol Ecol 23(8):2006–2019. https://doi.org/10.1111/mec.12705

Deng W, Casao MC, Wang P, Sato K, Hayes PM, Finnegan EJ, Trevaskis B (2015) Direct links between the vernalization response and other key traits of cereal crops. Nat Commun 6:5882. https://doi.org/10.1038/ncomms6882

Eckert A, Maloney P, Vogler D, Jensen C, Mix A, Neale D (2015) Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). Tree Genet Genomes 11:1–17

Eckert AJ, Bower AD, González-Martinez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol 19(17):3789–3805. https://doi.org/10.1111/j.1365-294X.2010.04698.x

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, Krutovsky KV, Clair JBS, Neale DB (2009a) Asssociation genetics of coastal Douglas fir (*Pseudotsuga menziesu* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. Genetics 182(4):1289–1302. https://doi.org/10.1534/genetics.109.102350

Eckert AJ, Wegrzyn JL, Pande B, Jermstad KD, Lee JM, Liechty JD, Tearse BR, Krutovsky KV, Neale DB (2009b) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*). Genetics 183(1):289–298. https://doi.org/10.1534/genetics.109.103895

Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci U S A 101(4):975–979. https://doi.org/10.1073/pnas.0308064100

Falconer DS, TFC Mackay 1996 Introduction to quantitative genetics

Fins L, ST Friedman, JV Brotschol 1992 Handbook of quantitative forest genetics, Dordrecht, DOI: https://doi.org/10.1007/978-94-015-7987-2

Forester BR, JR Lasky, HH Wagner, DL Urban 2017 Using genotype environment associations to identify multilocus local adaptation. bioRxiv:1–24. doi: https://doi.org/10.1101/129460

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. Science 334(6052):86–89. https://doi.org/10.1126/science.1209271

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol 30(7):1687–1699. https://doi.org/10.1093/molbev/mst063

Friendly M (2002) Corrgrams: exploratory displays for correlation matrices. Am Stat 56(4):316–324. https://doi.org/10.1198/000313002533

Gao F, Zhou Y, Zhu W, Li X, Fan L, Zhang G (2009) Proteomic analysis of cold stress-responsive proteins in *Thellungiella* rosette leaves. Planta 230(5):1033–1046. https://doi.org/10.1007/s00425-009-1003-6

Garay-Arroyo A, Colmenero-Flores JM, Garciarrubio A, Covarrubias AA (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. J Biol Chem 275(8):5668–5674. https://doi.org/10.1074/jbc.275.8.5668

González-Martinez SC, Krutovsky KV, Neale DB (2006) Forest-tree population genomics and adaptive evolution. New Phytol 170(2):227–238. https://doi.org/10.1111/j.1469-8137.2006.01686.x

Guak S, Olsyzk DM, Fuchigami LH, Tingey DT (1998) Effects of elevated CO(2) and temperature on cold hardiness and spring bud burst and growth in Douglas-fir (Pseudotsuga menziesii). Tree Physiol 18(10):671–679. https://doi.org/10.1093/treephys/18.10.671

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. Genetics 195(1):205–220. https://doi.org/10.1534/genetics.113.152462

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. Science 334(6052):83–86. https://doi.org/10.1126/science.1209244

Hannah MA, Heyer AG, Hincha DK (2005) A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. PLoS Genet 1(2):e26. https://doi.org/10.1371/journal.pgen.0010026

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. Evol Appl 7(9):1008–1025. https://doi.org/10.1111/eva.12149

Hijmans RJ, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25(15):1965–1978. https://doi.org/10.1002/joc.1276

Holliday JA, Ritland K, Aitken SN (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (Picea sitchensis). New Phytol 188(2):501–514. https://doi.org/10.1111/j.1469-8137.2010.03380.x

Holliday JA, Wang T, Aitken S (2012) Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. G3: Genes, Genomes, Genetics 2:1085–1093

IPCC (2013) Fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, New York, USA

Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB (2003) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics 165(3):1489–1506

Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001a) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. Theor Appl Genet 102(8):1142–1151. https://doi.org/10.1007/s001220000505

Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, Neale DB (2001b) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. Theor Appl Genet 102(8):1152–1158. https://doi.org/10.1007/s001220000506

Jones MR, Forester BR, Teufel AI, Adams RV, Anstett DN, Goodrich BA, Landguth EL, Joost S, Manel S (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. Evolution 67(12):3455–3468. https://doi.org/10.1111/evo.12237

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol 23(3):482–490. https://doi.org/10.1093/molbev/msj057

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecol Lett 15(4):378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Krutovsky KV, Clair JBS, Saich R, Hipkins VD, Neale DB (2009) Estimation of population structure in coastal Douglas-fir *Pseudotsuga menziesii* (Mirb.) *Franco var. menziesii* using allozyme and microsatellite markers. Tree Genet Genomes 5(4):641–658. https://doi.org/10.1007/s11295-009-0216-y

Langlet O (1971) Two hundred years of genecology. Taxon 20(5/6):653–722. https://doi.org/10.2307/1218596

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164(3):1205–1219

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol 21(7):1548–1566. https://doi.org/10.1111/j.1365-294X.2012.05479.x

Lee BH, Henderson DA, Zhu J-K (2005) The *Arabidopsis* cold-responsive transcriptome and its regulation by ICE1. Plant Cell 17(11):3155–3175. https://doi.org/10.1105/tpc.105.035568

Lepais O, Bacles CFE (2014) Two are better than one: combining landscape genomics and common gardens for detecting local adaptation in forest trees. Mol Ecol 23(19):4671–4673. https://doi.org/10.1111/mec.12906

Li B, Ning L, Zhang J, Bao M, Zhang W (2015) Transcriptional profiling of *Petunia* seedlings reveals candidate regulators of the cold stress response. Front Plant Sci 6:118

Lind BM, Friedline CJ, Wegrzyn J, Maloney PE, Vogler DR, Neale DB, Eckert A (2017) Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Englm.) across fine spatial scales of the Lake Tahoe Basin, USA. Mol Ecol 26(12):3168–3185. https://doi.org/10.1111/mec.14106

Lorenz WW, Sun F, Liang C, Kolychev D, Wang H, Zhao X, Cordonnier-Pratt MM, Pratt LH, Dean JF (2006) Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. Tree Physiol 26(1):1–16. https://doi.org/10.1093/treephys/26.1.1

Moura JC, Bonine CA, de Oliveira Fernandes Viana J, Dornelas MC, Mazzafera P (2010) Abiotic and biotic stresses and changes in the lignin content and composition in plants. J Integr Plant Biol 52(4):360–376. https://doi.org/10.1111/j.1744-7909.2010.00892.x

Nadeau S, Meirmans PG, Aitken SN, 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. Ecol Evol 6(24):8649–8664. https://doi.org/10.1002/ece3.2550

Narum SR, Hess JE (2011) Comparison of F(ST) outlier tests for SNP loci under selection. Mol Ecol Resour 11:184–194. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12(2):111–122. https://doi.org/10.1038/nrg2931

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9(7):325–330. https://doi.org/10.1016/j.tplants.2004.05.006

O'Neill GA, Aitken SN, Adams WT (2000) Genetic selection for cold hardiness in coastal Douglas-fir seedlings and saplings. Can J For Res 30(11):1799–1807. https://doi.org/10.1139/x00-114

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin RP, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner APH 2017 Vegan: community ecology package. R package version 2.4–2

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. Methods Ecol Evol 6(4):445–464. https://doi.org/10.1111/2041-210X.12324

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes. Evolution 57(5):1182–1195. https://doi.org/10.1111/j.0014-3820.2003.tb00327.x

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6(1):288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Prunier J, Gerardi S, Laroche J, Beaulieu J, Bousquet J (2012) Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. Mol Ecol 21(17):4270–4286. https://doi.org/10.1111/j.1365-294X.2012.05691.x

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rajora OP, Eckert AJ, Zinck JWR (2016) Single-locus versus multilocus patterns of local adaptation to climate in eastern white pine (Pinus strobus, Pinaceae). PLoS One 11(7):e0158691. https://doi.org/10.1371/journal.pone.0158691

Rehfeldt GE, Jaquish BC, Sáenz-Romero C, Joyce DG, Leites LP, Bradley St Clair J, López-Upton J (2014) Comparative genetic responses to climate in the varieties of Pinus ponderosa and Pseudotsuga menziesii: reforestation. For Ecol Manag 324:147–157. https://doi.org/10.1016/j.foreco.2014.02.040

SAS Institute (2008) SAS/STAT User's guide, version 9.2. Cary, NC: SAS Institute Inc.

Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S (2012) Adaptive genetic variation on the landscape: methods and cases. Annu Rev Ecol Evol Syst 43(1):23–43. https://doi.org/10.1146/annurev-ecolsys-110411-160248

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet Genomes 9(4):901–911. https://doi.org/10.1007/s11295-013-0596-x

St Clair JB (2006) Genetic variation in fall cold hardiness in coastal Douglas-fir in western Oregon and Washington. Can J Bot Rev Can Bot 84:1110–1121

St Clair JB, Mandel NL, Vance-Boland KW (2005) Genecology of Douglas fir in western Oregon and Washington. Ann Bot 96(7):1199–1214. https://doi.org/10.1093/aob/mci278

Storey JD (2002) A direct approach to false discovery rates. J Roy Stat Soc Ser B Stat Methodol 64(3):479–498. https://doi.org/10.1111/1467-9868.00346

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100(16):9440–9445. https://doi.org/10.1073/pnas.1530509100

Travis JMJ, Münkemüller T, Burton OJ, Best A, Dytham C, Johst K (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. Mol Biol Evol 24(10):2334–2343. https://doi.org/10.1093/molbev/msm167

Trevaskis, B., D. J. Bagnall, M. H. Ellis, W. J. Peacock, and E. S. Dennis. 2003. MADS box genes control vernalization-induced flowering in cereals. Proceedings of the National Academy of Sciences 100: 13099–13104

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13(4):921–935. https://doi.org/10.1046/j.1365-294X.2004.02076.x

Wang T, Hamann A, Spittlehouse D, Murdock TN (2012) ClimateWNA – high-resolution spatial climate data for western North America. J Appl Meteorol Climatol 61:16–29

Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai CJ, Neale DB (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol 188(2):515–532. https://doi.org/10.1111/j.1469-8137.2010.03415.x

Wheeler NC, Jermstad KD, Krutovsky K, Aitken SN, Howe GT, Krakowski J, Neale DB (2005) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. IV. Cold-hardiness QTL verification and candidate gene mapping. Mol Breed 15(2):145–156. https://doi.org/10.1007/s11032-004-3978-9

White TL, WT Adams, DB Neale 2007 Forest genetics. Cambridge, DOI: https://doi.org/10.1079/9781845932855.0000

Yadav DK, Shukla D, Tuteja N (2014) Isolation, in silico characterization, localization and expression analysis of abiotic stress-responsive rice G-protein β subunit (RGB1). Plant Signal Behav 9(5):e28890. https://doi.org/10.4161/psb.28890

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, Liepe KJ, Hamann A, Holliday JA, Whitlock MC, Rieseberg LH, Aitken SN (2016) Convergent local adaptation to climate in distantly related conifers. Science 353(6306):1431–1433. https://doi.org/10.1126/science.aaf7812