NRC Research Press

ARTICLE

Balancing selection maintains diversity in a cold tolerance gene in broadly distributed live oaks[1]

Jose Eduardo Meireles, Anne Beulke, Daniel S. Borkowski, Jeanne Romero-Severson, and Jeannine Cavender-Bares

Abstract: Cold poses major physiological challenges to plants, especially long-lived trees. In trees occurring along variable temperature clines, the expected direction and consequences of selection on cold acclimation ability and freezing tolerance are not straightforward. Here we estimated selection in cold acclimation genes at two evolutionary timescales in all seven species of the American live oaks (*Quercus* subsection *Virentes*). Two cold response candidate genes were chosen: *ICE1*, a key gene in the cold acclimation pathway, and *HOS1*, which modulates cold response by negatively regulating *ICE1*. Two housekeeping genes, *GAPDB* and *CHR11*, were also analyzed. At the shallow evolutionary timescale, we demonstrate that *HOS1* experienced recent balancing selection in the two most broadly distributed species, *Q. virginiana* and *Q. oleoides*. At a deeper evolutionary scale, a codon-based model of evolution revealed the signature of negative selection in *ICE1*. In contrast, three positively selected codons have been identified in *HOS1*, possibly a signature of the diversification of *Virentes* into warmer climates from a freezing adapted lineage of oaks. Our findings indicate that evolution has favored diversity in cold tolerance modulation through balancing selection in *HOS1* while maintaining core cold acclimation ability, as evidenced by purifying selection in *ICE1*.

Key words: balancing selection, *Quercus*, cold tolerance, *ICE1*, *HOS1*.

Résumé : Le froid pose des défis physiologiques importants chez les plantes, particulièrement chez les arbres qui vivent longtemps. Chez les arbres qu'on retrouve dans des régions affectées par un gradient de température, la direction attendue et les conséquences de la sélection pour le potentiel d'acclimatation au froid et la tolérance au gel ne sont pas simples. Dans ce travail, les auteurs ont estimé la sélection dans deux gènes d'acclimatation au froid sur deux échelles de temps chez les sept espèces américaines de chênes persistants (*Quercus* sous-section *Virentes*). Deux gènes candidats impliqués dans la réponse au froid ont été choisis : *ICE1*, un gène clé dans le sentier de l'acclimatation au froid, et *HOS1*, lequel module la réponse au froid en régulant l'expression de *ICE1*. Deux gènes stables, *GAPDB* et *CHR11*, ont également été analysés. À l'échelle de temps proche, les auteurs ont démontré que le gène *HOS1* avait connu un épisode récent de sélection équilibrante chez deux des espèces les plus répandues, *Q. virginiana* et *Q. oleoides*. À une échelle évolutive plus lointaine, un modèle d'évolution basé sur les codons a révélé une signature de sélection négative au sein de *ICE1*. Au contraire, trois codons ayant fait l'objet d'une sélection positive ont été identifiés chez *HOS1*, possiblement une signature de la diversification des *Virentes* dans des climats plus chauds à partir d'une lignée de chêne adaptée au froid. Les résultats indiquent que l'évolution a favorisé la diversité dans la modulation de la tolérance au froid à travers une sélection équilibrante chez *HOS1* tout en maintenant l'aptitude centrale d'acclimatation au froid, tel que démontré par la sélection purificatrice chez *ICE1*. [Traduit par la Rédaction]

Mots-clés : sélection équilibrante, *Quercus*, tolérance au froid, *ICE1*, *HOS1*.

Introduction

Cold poses major challenges to plant performance and survival and is known to be an important driver of distribution ranges of species and clades (MacArthur 1972; Donoghue 2008; Zanne et al. 2013). Chilling temperatures (15–0 °C, Guy et al. 2008) decrease enzymatic activity, metabolic rates, and impair cellular transport due to reduced membrane fluidity (Thomashow 1998; Chinnusamy et al. 2007; Knight and Knight 2012). Freezing may cause the intracellular formation of ice crystals that perforate membranes and kill the cell. Extracellular ice formation, while not lethal to cells, leads to dehydration stress since the water potential of ice is more negative than that of liquid water (Cavender-Bares 2005; Chinnusamy et al. 2007).

Many plants are able to acquire cold tolerance by going through a series of physiological changes in cell wall and membrane structure and accumulating cryoprotective molecules, in a process known as cold acclimation (Cavender-Bares 2005; Janska et al. 2010). These changes have been long hypothesized to incur physiological costs to the plant, ultimately leading to a trade-off between the degree of cold tolerance on one hand and growth and reproductive rates on the other (MacArthur 1972; Loehle 1998). It has been empirically demonstrated that freezing tolerance is negatively correlated with growth rates, and that the degree of cold tolerance often correlates with the temperatures experienced by different populations (Fig. 1A; Darychuk et al. 2012; Koehler et al. 2012). However, the mechanisms underlying this trade-off may be

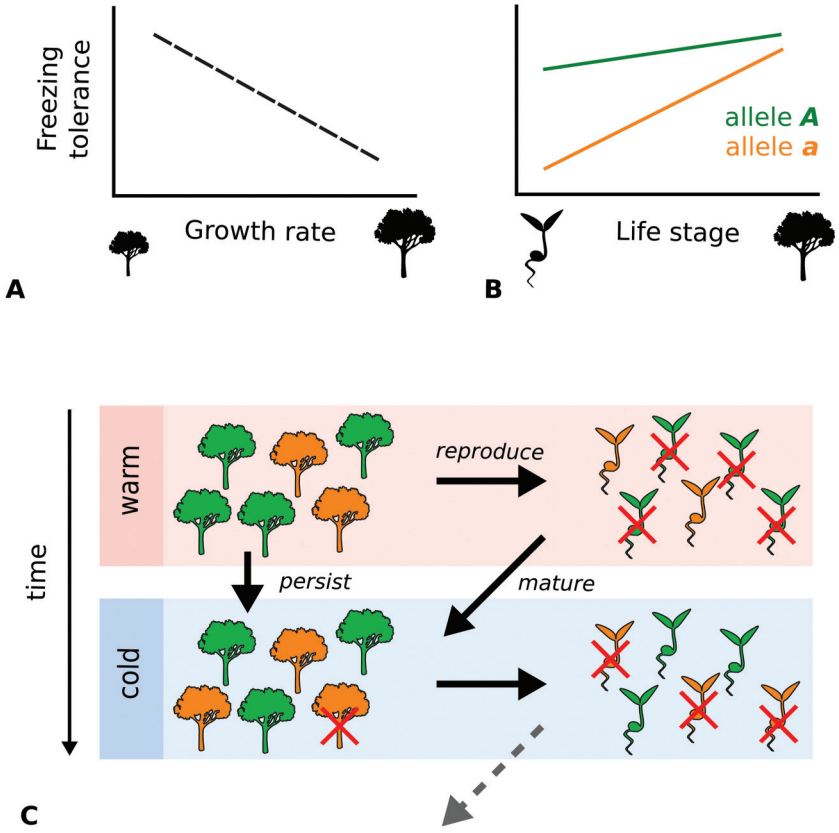
Received 23 November 2016. Accepted 4 April 2017.

Corresponding Editor: Andrew L. Hipp.

J.E. Meireles, A. Beulke, and J. Cavender-Bares. Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA.
D.S. Borkowski and J. Romero-Severson. Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA.
Corresponding author: Jose Eduardo Meireles (email: meireles@umn.edu).

[1]This paper is part of a Special Issue entitled The Evolution of Tree Diversity.

Fig. 1. Conceptual model of how (A) the cold tolerance/growth rate tradeoff may lead to balancing selection provided that (B) alleles confer different tolerances and that seedlings are more susceptible to cold damage than adults. (C) A warm year selects against seedlings carrying the green allele (high freezing tolerance, slow growth) more strongly than the adults. However, the green allele persists from a warm year to a cold year in the adults, when it has an advantage over low freezing tolerance (orange allele) individuals. This process results in the maintenance of both alleles in the population.



more complex than simple differential resource allocation (Savage and Cavender-Bares 2013).

Species and populations that are broadly distributed, span steep temperature gradients, or inhabit locations of highly inconstant climate will be exposed to variable selective environments in space and time (Hedrick 2006; Savolainen et al. 2007; Zhen and Ungerer 2008). For example, the same degree of cold tolerance would not be beneficial across the entire range of broadly distributed species because of temperature differences across geographic regions (Kelly 2006; Darychuk et al. 2012; Koehler et al. 2012). Similarly, the same population may experience different climates at varying temporal scales, such that different levels of cold tolerance may be favored in different years or decades (Fig. 1; Hedrick 2006; Vitasse et al. 2014). Furthermore, selection associated with yearly and decadal temperature variation may have stronger effects in seedlings given that they are more susceptible to freezing stress (Fig. 1B; Boorse et al. 1998).

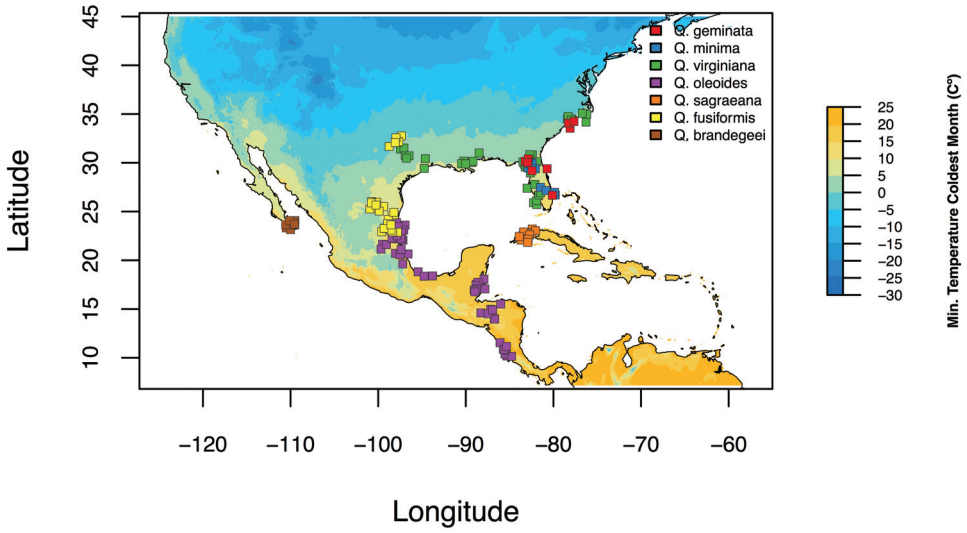
Given the stark consequences of cold damage for survival and fitness and the potential trade-offs between cold tolerance and growth, it is expected that genetic variation underlying cold acclimation should be under selection (Zhen and Ungerer 2008; Zhen et al. 2011; Preston and Sandve 2013). The specific form of selection depends on many factors, including the degree of climatic variability experienced, degree of gene flow between populations, the role of specific genes in response to cold and freezing stress and particularly on the evolutionary timescale in question.

The outcome of selection for cold acclimation at shallow evolutionary scales, related to population level processes, depends on its interplay with climatic variability (Zhen and Ungerer 2008; Zhen et al. 2011). In predictable climatic conditions, selection is generally expected to purge suboptimal alleles, favoring locally adapted populations where the mean cold response matches the home site conditions (Hedrick 2006; Savolainen et al. 2007). Alternatively, balancing selection may favor the maintenance of diversity in cold tolerance in populations that experience strong climatic fluctuations, if different levels of cold tolerance are advantageous in different years (Turelli and Barton 2004; Hedrick 2006; Vitasse et al. 2014). Signatures of selective dynamics at this scale are generally inferred from population models that use allele frequency and genetic structure information (Nielsen 2005; Keller et al. 2012; Vitti et al. 2013).

Molecular evolution analyses at deeper timescales help to uncover the history of long-term selective pressures that have acted on cold tolerance genes. It is known, for example, that codons of functionally indispensable genes tend to be constrained and show stronger signatures of purifying (negative) selection whereas genes involved in regulation or non-essential functions are generally less constrained (Anisimova and Kosiol 2009; Vitti et al. 2013). Additionally, footprints of processes such as shifts in climatic niche may in some cases be recovered from molecular models. For instance, the evolution of cold-intolerant alleles from a cold-tolerant clade may have been favored as a clade invades warmer climates, leaving a signature of positive selection in certain codons.

The live oaks (*Quercus* subsection *Virentes*) provide a unique opportunity to investigate selection on genes associated with cold tolerance. *Virentes* spans the wide temperature gradient across the Tropical–Temperate divide of North America and Mesoamerica (Fig. 2, Cavender-Bares et al. 2015). Three of its species, *Q. oleoides*, *Q. virginiana*, and *Q. fusiformis*, are broadly distributed, and their pop-

Fig. 2. Geographic location of populations sampled in this study, broadly covering the ranges of all seven species of live oaks (*Virentes*). Map colors show the minimum temperature in the coldest month and highlights the temperature variation experienced by distributed species. The population delimitation for each species is provided in Table S1[2].



ulations face strikingly different climatic conditions (Cavender-Bares et al. 2015). Broadly distributed *Virentes* also experience strong seasonal and inter-annual climatic variability such that different tolerance levels may be favored in different years. Furthermore, Koehler et al. (2012) have identified variation in freezing tolerance both among and within species in *Virentes*, demonstrating that population home temperatures were negatively correlated with and significantly predict freezing tolerance after cold acclimation. Interestingly, some species within the *Virentes* that only rarely experience chilling or freezing have retained some ability to express freezing tolerance.

Cold acclimation in plants is governed by a complex transcriptional pathway with two key upstream genes, *ICE1* and *HOS1* (Chinnusamy et al. 2007; Knight and Knight 2012). The *ICE1* (*inducer of CBF expression*) gene encodes a transcription factor essential to initiate the cold acclimation cascade. It induces the transcription of *CBF* genes (Chinnusamy et al. 2007; Miura and Furumoto 2013), promoting the expression of *COR* (*cold responsive*) genes that are responsible for many cold responses (Hannah et al. 2005), for instance, membrane restructuring and stabilization (e.g., *COR15a* and *DHN1*, Steponkus et al. 1998; Koag et al. 2003), enzyme stabilization during freezing (*COR85*, Kazuoka and Oeda 1994), and accumulation of intracellular polysaccharides (Miura et al. 2012). On the other hand, the *HOS1* (*high expression of osmotically responsive*) gene negatively regulates cold response expression by the ubiquitination of *ICE1* transcription factor (Dong et al. 2006; Miura and Furumoto 2013).

Our major goals were to estimate the direction of selection in two candidate genes associated with cold tolerance (*ICE1* and *HOS1*) at two different evolutionary timescales within seven live oak species. We first evaluated whether cold acclimation genes have mostly experienced recent diversifying selection or, alternatively, whether allelic diversity within and between populations is maintained by balancing selection. Second, we tested whether SNP variation was associated with the climatic clines experienced by populations of different species. Finally, we estimated the long-term selection patterns of the two candidate cold tolerance genes using phylogenetically informed codon-based models of molecular evolution. Our inferences of selection at both evolutionary scales were compared to results from two putative housekeeping genes, *GAPDB* and *CHR11*.

Methods

Sampling and freezing tolerance measurement

All seven live oak (*Quercus* subsection *Virentes*) species were sampled: *Q. virginiana*, *Q. oleoides*, *Q. geminata*, *Q. minima*, *Q. fusiformis*, *Q. sangreana*, and *Q. brandegeei*. The number of sampled specimens from each species varied from 7 for *Q. geminata* to 51 for *Q. oleoides*. Supplementary data Table S1[2] summarizes our sampling and includes population assignment, coordinates of the collection site, and number of individuals sampled. Populations have been delimited based on previous genetic studies using microsatellite and chloroplast data (Cavender-Bares et al. 2011; Gugger and Cavender-Bares 2013; Deacon and Cavender-Bares 2015; Cavender-Bares et al. 2015) as well as on geography, and the choice of populations to sample was designed to cover the geographic and climatic range of each species (Cavender-Bares et al. 2015). Leaves were collected either from the field or from trees grown in the greenhouse at the University of Minnesota for a previous study (Koehler et al. 2012). Leaf tissue was stored at −80 °C whether collected in the field or from the greenhouse.

Stem freezing tolerance data was taken from Koehler et al. (2012) and Cavender-Bares et al. (2015). Freezing tolerance was estimated with the electrolyte leakage method, which uses variation in electrical conductivity to estimate the degree of cell lysis in response to different freezing treatments (Murray et al. 1989). More details on how the measurements were conducted are given in Koehler et al. (2012).

Two key cold tolerance genes were sampled, *ICE1* and *HOS1*. As stated before, *ICE1* is a fundamental upstream gene in the *CBF* expression cascade that induces several cold acclimation responses, while *HOS1* negatively regulates *ICE1* transcripts, acting to decrease cold tolerance (reviews in Chinnusamy et al. 2007; Miura and Furumoto 2013). Two putative housekeeping genes were chosen for the purposes of comparison: *GAPDB* (*glyceraldehyde-3-phosphate dehydrogenase subunit B*), essential for glycolysis, and *CHR11* (*chromatin remodeling protein*), an imitation switch (ISWI) pro-

[2]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2016-0208.

tein in a family of ATP-dependent chromatin remodeling factors. *CHR11* is needed for the regulation of cellular identity in the vegetative phase and for nuclear proliferation during female gametogenesis (Li et al. 2012).

DNA extraction, amplification, and sequencing

Regions in four genes (*HOS1*, *ICE1*, *GAPDB*, and *CHR11*) were amplified and sequenced to explore the genetic variation in cold tolerance. The frozen leaf tissue was ground and used for genomic DNA extractions following the standard CTAB extraction protocol (Doyle and Doyle 1987). Primers were designed to start and end in an exon while spanning an intron region of each candidate gene, using the whole-genome databases for *Prunus persica* and *Vitis vinifera* (Table S2[2]). They were tested in *Quercus* by PCR, sequencing, and finally BLASTing the results simply to confirm that the expected genomic regions were amplified. The primer annealing temperatures were then optimized for binding to the live oak genomic DNA by running gradient PCRs. The 20 μL PCR reactions included the following: 1 μL of template DNA (at about 10 ng/μL), 14.5 μL ddH_2O, 2 μL of 10× PCR buffer containing MgCl2, 0.4 μL dNTPs (40 mM total, each dNTP at 10 mM), 1 μL forward primer (at 10 μM), 1 μL reverse primer (at 10 μM), and 0.1 μL *Taq* DNA polymerase (5 U/μL). The thermal cycler amplification program performed 40 cycles of denaturing for 30 s at 94 °C, annealing for 45 s at 58–63 °C, and extension for 2 min at 72 °C, flanked by an initial 2 min denaturing step at 94 °C and a final 10 min extension step at 72 °C. PCR products were then prepared for sequencing at University of Notre Dame and Beckman Coulter Genomics. Prior to Sanger sequencing, Notre Dame PCR products were purified and quantified with a NanoDrop 2000, whereas Beckman Coutler samples were quantified with the Qubit dsDNA HS Assay. Sequencher (v 4.10.1) was used to visualize chromatograms, assess sequence quality and base calls, and identify heterozygous sites, which were labeled with their corresponding IUPAC ambiguity code. The processed sequences were finally exported in FASTA format.

Identifying coding regions, sequence alignment, and SNP extraction

Coding regions were identified by BLASTing a representative *Virentes* sequence of each gene against GenBank's protein database using the blastX algorithm. First, we ensured that the resulting blastX proteins matched the queried gene. The top 10 blastX hits were downloaded as DNA and their coding sequences (CDS) were added to the *Virentes* fasta. Subsequently, multiple alignment was performed in the augmented dataset using the FFT–NS–i algorithm in MAFFT v7 (Katoh and Standley 2013), allowing us to identify the coding sequences in each gene and annotate them as such. Finally, the CDSs from the blastX results were removed from the alignment. For each gene, both coding-sequence-only and SNP-only alignments were generated by subsetting the full alignment. The *HOS1* sequences included exon 2 and part of its flanking introns (749 bp, Table S2[2]), *ICE1* sequences encompassed exon 2 and its 3′ intron (458 bp, Table S2[2]), *GAPDB* sequences included exon 3 and part of its flanking introns (578 bp, Table S2[2]), and *CHR11* sequences encompassed three small exons (1, 2, and 3) along with their neighboring introns (685 bp, Table S2[2]).

Analyses

Allelic diversity (*A*) and expected heterozygosity (H_e) is provided for each gene for all seven species of *Virentes*. For each locus, H_e was calculated using $H_e = 1 - \Sigma_{i=1}^{k} p_i^2$, where p_i is the frequency of the *i*th of *k* alleles. The analysis of recent selection was conducted with a Bayesian multinomial dirichlet model implemented in Bayescan (Foll and Gaggiotti 2008). In this model, population allele frequency differentiation from a common gene pool is measured with the F_{ST} coefficient, which is decomposed with a logistic regression into a population-specific parameter β and a locus-specific component α. Significantly positive values of α suggest diversifying selection, whereas significantly negative α indicates balancing or purifying selection. Statistical significance for the model with selection was tested using Bayes Factors and evaluated in light of Jeffrey's interpretation (Kass and Raftery 1995). Input files for Bayescan were produced from the SNP alignments using a custom R script. Each of the seven species was analyzed separately, and populations were delimited as shown in Table S1[2]. The Bayescan MCMC chain was first tuned by running 20 independent pilot runs of length 5000, and then sampled for 100 000 iterations. The first 50 000 samples were discarded as burn-in and the remaining samples were thinned to every 10th iteration. The default Bayescan prior odds for the neutral model, with no selection, is 10. We have set prior odds 20 instead to minimize the chance of false positives.

The association between SNPs and freezing tolerance and temperature cline was investigated through generalized linear models (GLM) performed in R. For each individual, the minimum temperature at the coldest month (bioclim6) was extracted from the bioclim raster (Hijmans et al. 2005) at the 30arcsec resolution. Freezing tolerances and minimum temperature values were taken to be the dependent variable to be explained by each SNP, which entered the model as the independent variables. Population latitude was used as a covariate to minimize the chance of false positives, given that neutral SNPs may covary with environment because of population structure (Yu et al. 2006; van Heerwaarden et al. 2015), especially north–south population differentiation (Eckert et al. 2010). Model selection using stepwise AIC was performed and *t*-values associated with each SNP were recorded and *p*-values lower than 0.05 were considered statistically significant.

Selective dynamics at deeper evolutionary time was inferred at the gene level for all species jointly using a phylogenetic model of codon evolution implemented in FUBAR (Fast Unconstrained Bayesian Approximation) developed by Murrell et al. (2013). It infers the rates of non-synonymous (*dN*) and synonymous (*dS*) substitutions for each codon, which are expected to be the same under neutral evolution. Positive selection acts to increase *dN*, leading to a positive *dN–dS*, whereas negative selection increases synonymous rates (*dS*), yielding negative *dN–dS* values. The coding sequences of each candidate gene were subjected to a FUBAR analysis performed with HyPhy in the datamonkey server (http://www.datamonkey.org). FUBAR's default neighbor joining trees used in the analyses were calculated with a HKY model of DNA evolution. We have also subjected our datasets to the Branch-Site REL model (Kosakovsky Pond et al. 2011) which, unlike FUBAR, also identifies which lineages account for the inferred selection patterns. The Branch-Site REL was also run with HyPhy in the datamonkey server.

Results

Recent selection estimated from allele frequencies

Allelic diversity patterns were strikingly different among the four sampled genes. Across all species, we found 17 alleles of *ICE1*, 32 alleles of *GAPDB*, 55 alleles of *CHR11*, and 110 alleles of *HOS1*. The disparity in allelic diversity between *ICE1* and *HOS1* was also found within the two most widespread species: *Q. virginiana* had 7 alleles of *ICE1* and 24 of *HOS1* (Table S3[2]), while 6 alleles of *ICE1* and 29 alleles of *HOS1* were found in *Q. oleoides* (Table S3[2]). It should be noted that polymorphisms in non-coding regions of the genes were also considered when identifying alleles.

Bayescan detected very strong evidence for recent balancing selection in *HOS1* in the two most widely distributed species: *Q. oleoides* had an α of –2.191 and $\log_{10}$ Bayes Factor (logBF) of 3.3 while *Q. virginiana's* estimated α was –1.935 and logBF of 2.0 (Fig. 3; Table 1). No evidence for selection was found in any other species, including the broadly distributed *Q. fusiformis*. Bayescan did not detect significant evidence for selection in the *ICE1* gene nor in any of the housekeeping genes (Fig. 3; Table 1).

Fig. 3. Bayescan results pooled across the independent analyses for each of the seven species within *Virentes*. Negative alpha estimates (*y* axis) indicate balancing selection whereas positive values would suggest positive selection. Log_{10} Bayes factors (logBF) greater than 0.5 (grey box) indicates a significant deviation from a null model with no selection. See Table S1[2] for details on population delimitation for each species.

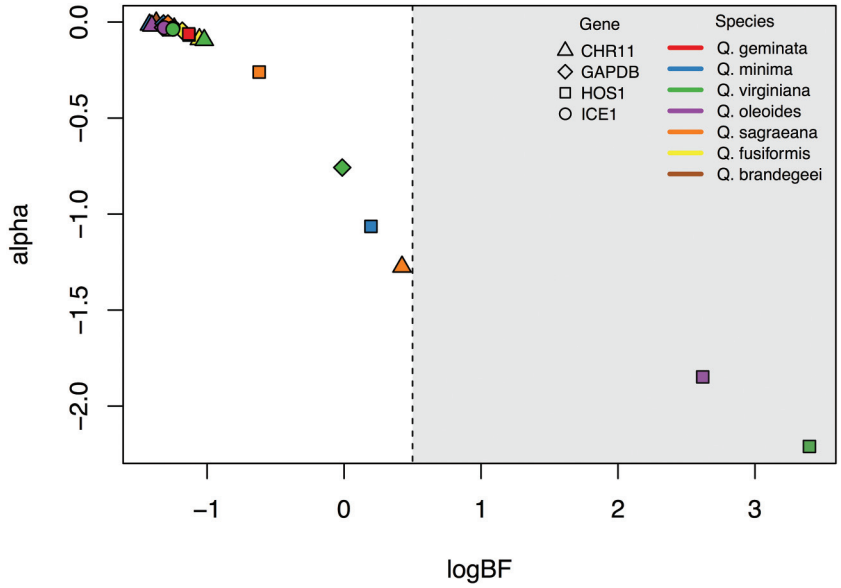


Table 1. Bayescan results for two housekeeping and two candidate genes in the three most broadly distributed live oaks.

Gene	Species	LogBF	alpha
CHR11	*Quercus fusiformis*	−1.03	−0.103
CHR11	*Quercus oleoides*	−1.11	−0.057
CHR11	*Quercus virginiana*	−0.93	−0.118
GAPDB	*Quercus fusiformis*	−1.18	−0.042
GAPDB	*Quercus oleoides*	−1.28	−0.031
GAPDB	*Quercus virginiana*	0.53	−1.367
HOS1	*Quercus fusiformis*	−1.17	−0.041
HOS1	*Quercus oleoides*	3.3	−2.191
HOS1	*Quercus virginiana*	2	−1.935
ICE1	*Quercus fusiformis*	−1.26	−0.037
ICE1	*Quercus oleoides*	−1.3	−0.028
ICE1	*Quercus virginiana*	−1.19	−0.037

Note: Negative values of the selection parameter α indicate balancing selection. Models with $\log_{10}(BF) > 1.0$ were considered significant and are highlighted in grey.

Gene level selection at phylogenetic timescales

We considered codons to significantly deviate from neutral evolution in the FUBAR (Murrell et al. 2013) analyses whenever the posterior probability for either positive or negative selection were higher than 0.85 and empirical Bayes factors (logBF) were at least substantial (Kass and Raftery 1995). Three *HOS1* codons, 31, 145, and 146, showed significant (logBF 0.95, 1.21, and 1.58) signatures of positive selection as indicated by positive *dN–dS* values, while three others, 30, 101, and 103, showed negative selection (Fig. 4, logBF 2.17, 0.93, and 0.82). In contrast, FUBAR only estimated significantly negative *dN–dS* values in *ICE1*, specifically on codons 10, 65, and 133 (logBF 1.16, 1.1, and 0.99). FUBAR inferred pervasive negative selection (10 codons with negative *dN–dS*) in *GAPDB* (Fig. 4), but found weak evidence for selection in the putative housekeeping gene *CHR11*, revealing one codon under positive and another under negative selection. The Branch-Site REL model (Kosakovsky Pond et al. 2011) did not identify significant differences in *dN*/*dS* rates across lineages in any of the genes according to a stepwise likelihood ratio test and Bonferroni corrected *p*-values.

Associations between coding region variation and freezing tolerance

The generalized linear models (GLMs) uncovered significant associations between *HOS1* polymorphisms (SNPs) and both individual freezing tolerance (Fig. 5; Tables S4[2] and S5[2]) and minimum temperature at the collection locations (Tables S6[2] and S7[2]) in *Q. oleoides* and *Q. virginiana*. The effect of a SNP on freezing tolerance was never significant in both species. For example, the *A* allele in SNP 282 was significantly (*p*-value 0.007) associated with less freezing tolerance in *Q. oleoides* but not in *Q. virginiana*. The only *ICE1* polymorphism significantly associated with the freezing tolerance and minimum temperatures was SNP 402 in *Q. oleoides* (*p*-value 0.002).

Discussion

We have found evidence that the diversity in *HOS1*, a key regulatory gene of cold acclimation, has been maintained through balancing selection (Table 1; Fig. 3). This finding indicates that evolution has favored high standing genetic diversity in *Virentes*, which are distributed broadly across the tropical temperate divide and experience high climatic variability. These results are compatible with empirical observations of wide variation in cold tolerance within and among those species (Koehler et al. 2012; Cavender-Bares et al. 2015). Our analysis of codon evolution demonstrates that *HOS1* has experienced positive selection in the past (Fig. 4), possibly a signature of the diversification of *Virentes* into warmer climates from a freezing adapted lineage of oaks.

Environmental variability and the maintenance of genetic diversity through balancing selection

Our results suggest that variation in the cold tolerance gene *HOS1* has been maintained by recent balancing selection in *Q. virginiana* and *Q. oleoides* (Table 1; Fig. 3). Interestingly, they are the two most broadly distributed of the seven species of *Virentes* and span major variation in climate (Cavender-Bares et al. 2015). We did not detect a significant signature of selection in *HOS1* in *Q. fusiformis*, another broadly distributed species, possibly because our sample size was limited for this species.

In principle, negative α parameters estimated from Bayescan may reflect two different evolutionary processes, balancing or possibly purifying selection, since both may act to decrease pop-

Fig. 4. Codon selection profiles across the candidate genes *HOS1* and *ICE1* (first column) and putative housekeeping genes *GAPDB* and *CHR11* (second column) inferred from a FUBAR analysis in HyPhy. Sites that experienced significant negative selection ($dN–dS < 0$) are highlighted in orange whereas sites sites that experienced significant positive selection ($dN–dS > 0$) are shown in blue. *HOS1* codons significantly associated with freezing tolerance or minimum temperature (Tables S4–S7[2]) are marked with red asterisks.

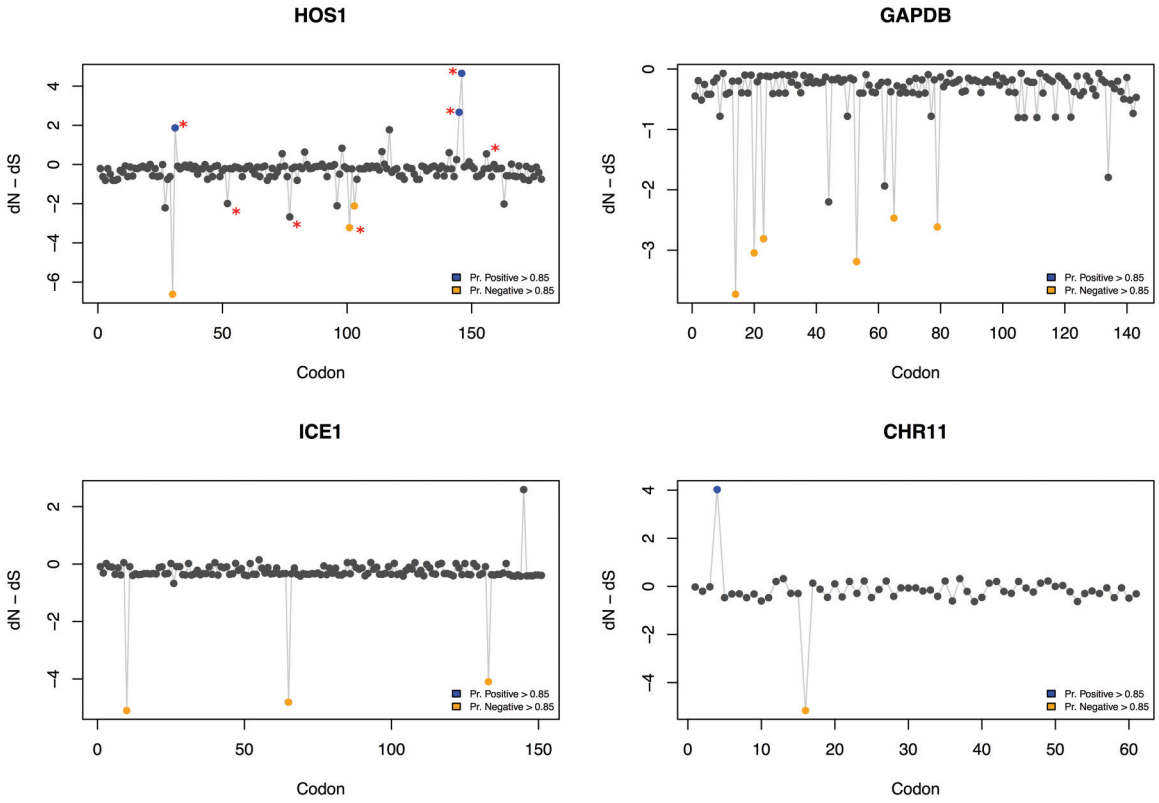
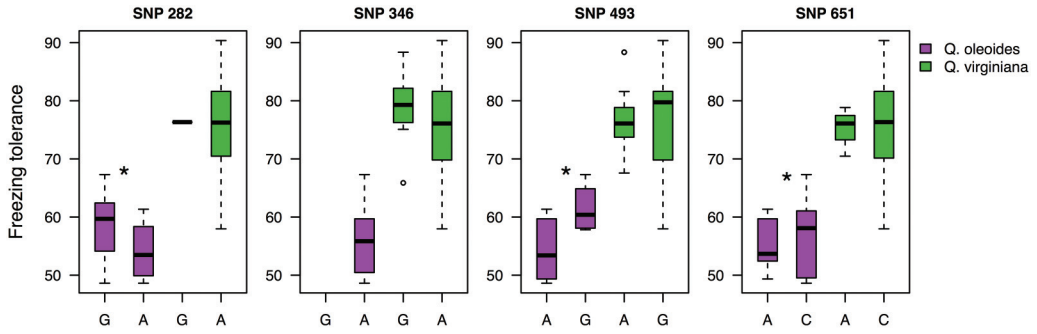


Fig. 5. Examples of the association between different *HOS1* SNPs in two broadly distributed live oak species, *Quercus virginiana* (green) and *Q. oleoides* (purple) and freezing tolerance measured with the electrolyte leakage method. Significant effects in intraspecific comparisons are marked with an asterisk. Statistical details from the GLMs are shown in Tables S4[2] and S5[2].



ulation differentiation (Foll and Gaggiotti 2008; Foll et al. 2010). However, because purifying selection acts to reduce genetic variability (Mitchell-Olds et al. 2007), the high allelic diversity found in *HOS1* strongly supports the interpretation that *HOS1* has experienced balancing selection.

The inference of balancing selection in *HOS1* provides a possible explanation for the observed variability in freezing tolerance in *Virentes*. Studies based on either experimental manipulations of saplings grown in controlled environments (Cavender-Bares 2007; Koehler et al. 2012) or under field conditions (other evergreen oaks, Cavender-Bares et al. 2005; Gimeno et al. 2008) have documented high variability in freezing tolerance within and between populations.

Balancing selection associated with environmental variability has been inferred in other plant and animal systems. For instance, Weinig et al. (2003) conducted a transplant experiment in *Arabdopsis thaliana* and suggested that allelic diversity observed in several QTLs were maintained by environmental heterogeneity and epistatic selection. Schmidt et al. (2000) have shown that the maintenance of diversity in the *MPI* gene in the northern acorn barnacle was associated with varying levels of environmental stress. Balancing selection has also been detected in genetic variation underlying responses to biological stress. For example, that has been suggested that balancing selection has been responsible for the maintenance of polymorphisms in *R*-genes, related to pathogen resistance in *Arabdopsis* (Tian et al. 2002; Mitchell-Olds and Schmitt 2006). Similarly, a long-term study of immune system genes (*MHC*) in sheep (Charbonnel and Pemberton 2005) demonstrated that *MHC* was under balancing selection, likely in response to variation in parasite load.

Molecular selection at deeper evolutionary scales

In contrast to allele-frequency based methods, which capture recent evolution, codon-based models of molecular evolution such as FUBAR (Fast Unconstrained Bayesian Approximation, Murrell et al. 2013) uncover the long-term selective dynamics that acted on coding regions of a gene (Anisimova and Kosiol 2009; Arenas 2015). Although our FUBAR analysis detected codons under selection in all four genes (Fig. 4), the Branch-Site REL model was not able to identify which lineages accounted for the inferred selection. This result is not surprising, however, since the Branch-Site REL model is parameter-rich and less likely to perform well in small datasets (Kosakovsky Pond et al. 2011).

Our FUBAR molecular evolution analysis of the *ICE1* gene identified three codons under strong purifying selection (significantly negative, *dN–dS*, Fig. 4), but no positive selection. Additionally, we found the allelic diversity of *ICE1* to be very low when compared to *HOS1*. These findings suggest that *ICE1* has been functionally constrained by selection over long evolutionary timescales, which was expected given the pivotal role that *ICE1* plays in cold temperature sensing and initiation of the acclimation cascade in plants (Chinnusamy et al. 2007; Miura and Furumoto 2013). Deleterious mutations in *ICE1* have been shown to impair the expression of *CBF*s and lead to significant decreases in acclimation potential and cold tolerance in *Arabidopsis* (Chinnusamy et al. 2003). The observation that *ICE1*'s function is conserved across flowering plants, from rice (*Oryza sativa*) to tomatoes (*Solanum lycopersicum*) and aspen (*Populus tremula*), further supports the idea that it has been under purifying selection on a deep evolutionary scale (Miura et al. 2012; Miura and Furumoto 2013).

On the other hand, we identified three *HOS1* codons that experienced past positive selection (positive, *dN–dS*, Fig. 4) and three other codons that have undergone negative selection (negative, *dN–dS*, Fig. 4). This indicates that different *HOS1* alleles have experienced directional selection at some point during the evolution of *Virentes*, possibly related to differences in cold tolerance. *HOS1* negatively regulates cold response by physically interacting with *ICE1* transcription factors and mediating their ubiquitination and degradation. Therefore, cold sensitivity and acclimation responses can be modulated by *HOS1*, whether through differential affinity to *ICE1* by different *HOS1* alleles or differential *HOS1* expression. In any case, it is important to note that deleterious mutations in *HOS1* lead to less stark consequences than the loss of *ICE1* function. Dong et al. (2006) and others have shown that under-expression or loss of function in *HOS1* leads to the accumulation of *ICE1* products and consequently enhance cold acclimation responses (Chinnusamy et al. 2007). In contrast, over-expression of *HOS1* leads to a reduction in cold acclimation and freezing tolerance, since *ICE1* transcripts become degraded at higher rates (Dong et al. 2006; Chinnusamy et al. 2007; Miura and Furumoto 2013).

Certain *HOS1* SNPs are associated with stem freezing tolerance, minimum temperature at the sample location, or both, even though population latitude was the most important explanatory covariate in all models (Tables S4–S7[2]). This result indicates that climatic variation has been a factor influencing genetic variation (Fig. 5). Interestingly, the SNPs associated with freezing tolerance generally match codons showing a strong signature of selection (Fig. 4), providing additional evidence that the molecular evolution results are biologically meaningful and not false positives. Interpreting the direction of the effect of SNPs on freezing tolerance is not straightforward and out of the scope of this study. The freezing tolerance phenotype is the result of the interaction of different *HOS1* mutations, *ICE1* alleles, and many other components not investigated here.

Our data cannot directly address which pressures the positively selected *HOS1* sites were responding to during evolution. Nevertheless, we hypothesize that the positive selection in *HOS1* is a consequence of the expansion of *Virentes* into warmer climates. The ancestral *HOS1* allele was likely optimized to confer cold acclimation ability since live oaks likely evolved from freezing tolerant lineages (Cavender-Bares et al. 2015; Hipp et al. 2014) and the colonization of tropical and subtropical regions by *Q. oleoides* and *Q. virginiana* (Fig. 2) is relatively recent (Cavender-Bares et al. 2015). The intrinsic trade-off between cold tolerance and growth rates in *Virentes* (Koehler et al. 2012) may have favored phenotypes less responsive to cold but with faster growth in species inhabiting warmer climates. As a consequence, *HOS1* alleles with higher binding affinity to *ICE1* might have been positively selected.

Long-lived trees that span broad geographic ranges must cope with highly variable climatic conditions, such that selective pressures to enhance growth or stress tolerance vary through space and time. Our findings suggest that in broadly distributed live oak species, a resilient response to climate change occurs on a population level, by maintaining the standing genetic variation in key regulatory genes. The capacity for flexibility ensures survival of at least some individuals during episodic extreme events (e.g., severe cold spells or late spring freezes), buffering the species against extinction.

Acknowledgements

This work was supported by the National Science Foundation grants IOS-0843665 (to J.C.-B.), IOS-1025974 (to J.R.-S.), and DEB-1146380 (to J.C.-B. and J.R.-S.) as well as a Micromorph NSF grant to A.B. We thank the Notre Dame Genomics Core facility and the University of Minnesota Genomics Center for assistance with amplicon sequencing and Michael Pfrender for helpful discussions.

References

Anisimova, M., and Kosiol, C. 2009. Investigating protein-coding sequence evolution with probabilistic codon substitution models. Mol. Biol. Evol. **26**(2): 255–271. doi:10.1093/molbev/msn232.

Arenas, M. 2015. Trends in substitution models of molecular evolution. Front. Genet. **6**(OCT). doi:10.3389/fgene.2015.00319.

Boorse, G.C., Ewers, F.W., and Davis, S.D. 1998. Response of chaparral shrubs to below-freezing temperatures: acclimation, ecotypes, seedlings vs. adults. Am. J. Bot. **85**(9): 1224–1230. doi:10.2307/2446631.

Cavender-Bares, J. 2005. Impacts of freezing on long-distance transport in woody plants. *In* Vascular transport in plants. *Edited by* N.M. Holbrook and M.A. Zwieniecki. pp. 401–424. doi:10.1016/B978-012088457-5/50021-6.

Cavender-Bares, J. 2007. Chilling and freezing stress in live oaks (*Quercus* section *Virentes*): Intra- and inter-specific variation in PS II sensitivity corresponds to latitude of origin. Photosynth. Res. **94**(2–3): 437–453. doi:10.1007/s11120-007-9215-8. PMID:17805986.

Cavender-Bares, J., Cortes, P., Rambal, S., Joffre, R., Miles, B., and Rocheteau, A. 2005. Summer and winter sensitivity of leaves and xylem to minimum freezing temperatures: a comparison of co-occurring Mediterranean oaks that differ in leaf lifespan. New Phytol. **168**(3): 597–612. doi:10.1111/j.1469-8137.2005.01555.x. PMID:16313643.

Cavender-Bares, J., Gonzalez-Rodriguez, A., Pahlich, A., Koehler, K., and Deacon, N. 2011. Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. J. Biogeogr. **38**(5): 962–981. doi:10.1111/j.1365-2699.2010.02451.x.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D.A.R., Hipp, A.A.L., Beulke, A., and Manos, P.S. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Mol. Ecol. **24**(14): 3668–3687. doi:10.1111/mec.13269. PMID:26095958.

Charbonnel, N., and Pemberton, J. 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity, **95**(5): 377–388. doi:10.1038/sj.hdy.6800735. PMID:16118661.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.-H., Hong, X., Agarwal, M., and Zhu, J.-K. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev. **17**(8): 1043–1054. doi:10.1101/gad.1077503. PMID:12672693.

Chinnusamy, V., Zhu, J., and Zhu, J.-K. 2007. Cold stress regulation of gene expression in plants. Trends Plant Sci. **12**(10): 444–451. doi:10.1016/j.tplants.2007.07.002. PMID:17855156.

Darychuk, N., Hawkins, B.J., and Stoehr, M. 2012. Trade-offs between growth and cold and drought hardiness in submaritime Douglas-fir. Can. J. For. Res. **42**(8): 1530–1541. doi:10.1139/x2012-092.

Deacon, N.J., and Cavender-Bares, J. 2015. Limited pollen dispersal contributes to population genetic structure but not local adaptation in *Quercus oleoides* forests of Costa Rica. PLoS ONE, **10**(9): e0138783. doi:10.1371/journal.pone.0138783. PMID:26407244.

Dong, C.-H., Agarwal, M., Zhang, Y., Xie, Q., and Zhu, J.-K. 2006. The negative

regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. Proc. Natl. Acad. Sci. **103**(21): 8281–8286. doi:10.1073/pnas.0602874103. PMID:16702557.

Donoghue, M.J. 2008. A phylogenetic perspective on the distribution of plant diversity. Proc. Natl. Acad. Sci. **105**(S1): 11549–11555. doi:10.1073/pnas.0801962105. PMID:18695216.

Doyle, J.J., and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**: 11–15.

Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L., Nelson, C.D., Ross-Ibarra, J., Gonzalez-Martinez, S.C., and Neale, D.B. 2010. Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). Genetics, **185**(3): 969–982. doi:10.1534/genetics.110.115543. PMID:20439779.

Foll, M., and Gaggiotti, O.E. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, **180**(2): 977–993. doi:10.1534/genetics.108.092221. PMID:18780740.

Foll, M., Fischer, M.C., Heckel, G., and Excoffier, L. 2010. Estimating population structure from AFLP amplification intensity. Mol. Ecol. **19**(21): 4638–4647. doi:10.1111/j.1365-294X.2010.04820.x. PMID:20874760.

Gimeno, T.E., Pias, B., Lemos-Filho, J.P., and Valladares, F. 2008. Plasticity and stress tolerance override local adaptation in the responses of Mediterranean holm oak seedlings to drought and cold. Tree Physiol. **29**(1): 87–98. doi:10.1093/treephys/tpn007. PMID:19203935.

Gugger, P.F., and Cavender-Bares, J. 2013. Molecular and morphological support for a Florida origin of the Cuban oak. J. Biogeogr. **40**(4): 632–645. doi:10.1111/j.1365-2699.2011.02610.x.

Guy, C., Kaplan, F., Kopka, J., Selbig, J., and Hincha, D.K. 2008. Metabolomics of temperature stress. Physiol. Planta. **132**(2): 220–235. doi:10.1111/j.1399-3054.2007.00999.x.

Hannah, M.A., Heyer, A.G., and Hincha, D.K. 2005. A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. PLOS Genet. **1**(2): 893–902. doi:10.1371/JOURNAL.PGEN.0010026.

Hedrick, P.W. 2006. Genetic polymorphism in heterogeneous environments: the age of genomics. Annu. Rev. Ecol. Evol. Syst. **37**(1): 67–93. doi:10.1146/annurev.ecolsys.37.091305.110132.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**(15): 1965–1978. doi:10.1002/joc.1276.

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., and Manos, P.S. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS ONE, **9**(4): e93975. doi:10.1371/journal.pone.0093975. PMID:24705617.

Janska, A, Marsik, P., Zelenkova, S., and Ovesna, J. 2010. Cold stress and acclimation — what is important for metabolic adjustment? Plant Biol. **12**(3): 395–405. doi:10.1111/j.1438-8677.2009.00299.x.

Kass, R.E., and Raftery, A.E. 1995. Bayes factors. J. Am. Stat. Assoc. **90**(430): 773–795. doi:10.1080/01621459.1995.10476572.

Katoh, K., and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. **30**(4): 772–780. doi:10.1093/molbev/mst010. PMID:23329690.

Kazuoka, T., and Oeda, K. 1994. Purification and characterization of COR85-oligomeric complex from cold-acclimated spinach. Oxford University Press. Available from http://pcp.oxfordjournals.org/content/35/4/601.full.pdf+html.

Keller, S., Levsen, N., Olson, M., and Tiffin, P. 2012. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. Mol. Biol. Evol. **29**(10): 3143–3152. doi:10.1093/molbev/mss121. PMID:22513286.

Kelly, J.K. 2006. Geographical variation in selection, from phenotypes to molecules. Am. Nat. **167**(4): 481–495. doi:10.1086/501167. PMID:16670992.

Knight, M.R., and Knight, H. 2012. Low-temperature perception leading to gene expression and cold tolerance in higher plants. New Phytol. **195**(4): 737–751. doi:10.1111/j.1469-8137.2012.04239.x. PMID:22816520.

Koag, M.-C., Fenton, R.D., Wilkens, S., and Close, T.J. 2003. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. Plant Physiol. **131**(1): 309–316. doi:10.1104/pp.011171. PMID:12529538.

Koehler, K., Center, A., and Cavender-Bares, J. 2012. Evidence for a freezing tolerance-growth rate trade-off in the live oaks (*Quercus* series *Virentes*) across the tropical–temperate divide. New Phytol. **193**(3): 730–744. doi:10.1111/j.1469-8137.2011.03992.x. PMID:22171967.

Kosakovsky Pond, S., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., and Scheffler, K. 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. **28**(11): 3033–3043. doi:10.1093/molbev/msr125. PMID:21670087.

Li, G., Zhang, J., Li, J., Yang, Z., Huang, H., and Xu, L. 2012. Imitation Switch chromatin remodeling factors and their interacting RINGLET proteins act together in controlling the plant vegetative phase in *Arabidopsis*. Plant J. **72**(2): 261–270. doi:10.1111/j.1365-313X.2012.05074.x. PMID:22694359.

Loehle, C. 1998. Height growth rate tradeoffs determine northern and southern range limits for trees. J. Biogeogr. **25**(4): 735–742. doi:10.1046/j.1365-2699.1998.2540735.x.

MacArthur, R. 1972. Geographical ecology: patterns in the distribution of species. Harper, New York.

Mitchell-Olds, T., and Schmitt, J. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature, **441**(7096): 947–952. doi:10.1038/nature04878. PMID:16791187.

Mitchell-Olds, T., Willis, J.H., and Goldstein, D.B. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat. Rev. Genet. **8**(11): 845–856. doi:10.1038/nrg2207. PMID:17943192.

Miura, K., and Furumoto, T. 2013. Cold signaling and cold response in plants. Int. J. Mol. Sci. **14**(3): 5312–5337. doi:10.3390/ijms14035312. PMID:23466881.

Miura, K., Shiba, H., Ohta, M., Kang, S.W., Sato, A., Yuasa, T., et al. 2012. *SlICE1* encoding a MYC-type transcription factor controls cold tolerance in tomato, *Solanum lycopersicum*. Plant Biotechnol. **29**(3): 253–260. doi:10.5511/plantbiotechnology.12.0303a.

Murray, M.B, Cape, J.N., and Fowler, D. 1989. Quantification of frost damage in plant tissues by rates of electrolyte leakage. New Phytol. **113**: 307–311. doi:10.1111/j.1469-8137.1989.tb02408.x.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. Mol. Biol. Evol. **30**(5): 1196–1205. doi:10.1093/molbev/mst030. PMID:23420840.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. **39**(1): 197–218. doi:10.1146/annurev.genet.39.073003.112420. PMID:16285858.

Preston, J.C., and Sandve, S.R. 2013. Adaptation to seasonality and the winter freeze. Front. Plant Sci. **4**(12): 3–11. doi:10.3389/fpls.2013.00167.

Savage, J.A., and Cavender-Bares, J. 2013. Phenological cues drive an apparent trade-off between freezing tolerance and growth in the family Salicaceae. Ecology, **94**(8): 1708–1717. doi:10.1890/12-1779.1. PMID:24015515.

Savolainen, O., Pyhäjärvi, T., and Knürr, T. 2007. Gene flow and local adaptation in trees. Annu. Rev. Ecol. Evol. Syst. **38**(1): 595–619. doi:10.1146/annurev.ecolsys.38.091206.095646.

Schmidt, P.S., Bertness, M.D., and Rand, D.M. 2000. Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. Proc. R. Soc. B Biol. Sci. **267**(1441): 379–384. doi:10.1098/rspb.2000.1012.

Steponkus, P.L., Uemura, M., Joseph, R.A., Gilmour, S.J., and Thomashow, M.F. 1998. Mode of action of the *COR15a* gene on the freezing tolerance of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. **95**(24): 14570–14575. doi:10.1073/pnas.95.24.14570. PMID:9826741.

Thomashow, M.F. 1998. Role of cold-responsive genes in plant freezing tolerance. Plant Physiol. **118**(1): 1–8. doi:10.1104/pp.118.1.1. PMID:9733520.

Tian, D., Araki, H., Stahl, E., Bergelson, J., and Kreitman, M. 2002. Signature of balancing selection in *Arabidopsis*. Proc. Natl. Acad. Sci. **99**(17): 11525–11530. doi:10.1073/pnas.172203599. PMID:12172007.

Turelli, M., and Barton, N.H. 2004. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G × E interactions. Genetics, **166**(2): 1053–1079. doi:10.1534/genetics.166.2.1053. PMID:15020487.

van Heerwaarden, J., van Zanten, M., and Kruijer, W. 2015. Genome-wide association analysis of adaptation using environmentally predicted traits. PLoS Genet. **11**(10): e1005594. doi:10.1371/journal.pgen.1005594. PMID:26496492.

Vitasse, Y., Lenz, A., and Korner, C. 2014. The interaction between freezing tolerance and phenology in temperate deciduous trees. Front. Plant Sci. **5**(October): 541. doi:10.3389/fpls.2014.00541. PMID:25346748.

Vitti, J.J., Grossman, S.R., and Sabeti, P.C. 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. **47**(1): 97–120. doi:10.1146/annurev-genet-111212-133526. PMID:24274750.

Weinig, C., Dorn, L.A., Kane, N.C., German, Z.M., Halldorsdottir, S.S., Ungerer, M.C., et al. 2003. Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. Genetics, **165**(1): 321–329. PMID:14504239.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**(2): 203–208. doi:10.1038/ng1702. PMID:16380716.

Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., et al. 2013. Three keys to the radiation of angiosperms into freezing environments. Nature, **506**(7486): 89–92. doi:10.1038/nature12872. PMID:24362564.

Zhen, Y., and Ungerer, M.C. 2008. Relaxed selection on the *CBF/DREB1* regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. Mol. Biol. Evol. **25**(12): 2547–2555. doi:10.1093/molbev/msn196. PMID:18775899.

Zhen, Y., Dhakal, P., and Ungerer, M.C. 2011. Fitness benefits and costs of cold acclimation in *Arabidopsis thaliana*. Am. Nat. **178**(1): 44–52. doi:10.1086/660282. PMID:21670576.