MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 3632–3644 doi: 10.1111/mec.13711

Adaptation to warmer climates by parallel functional evolution of *CBF* genes in *Arabidopsis thaliana*

J. GREY MONROE,*† CULLEN MCGOVERN,* JESSE R. LASKY,‡ KELSI GROGAN,* JAMES BECK§¶ and JOHN K. MCKAY*†

*Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523, USA, †Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523, USA, ‡Department of Biology, Pennsylvania State University, University Park, PA 16802, USA, §Department of Biological Sciences, Wichita State University, Wichita, KS 67260, USA, ¶Botanical Research Institute of Texas, Fort Worth, TX 76107, USA

Abstract

The evolutionary processes and genetics underlying local adaptation at a specieswide level are largely unknown. Recent work has indicated that a frameshift mutation in a member of a family of transcription factors, C-repeat binding factors or *CBF*s, underlies local adaptation and freezing tolerance divergence between two European populations of *Arabidopsis thaliana*. To ask whether the specieswide evolution of *CBF* genes in *Arabidopsis* is consistent with local adaptation, we surveyed *CBF* variation from 477 wild accessions collected across the species' range. We found that *CBF* sequence variation is strongly associated with winter temperature variables. Looking specifically at the minimum temperature experienced during the coldest month, we found that *Arabidopsis* from warmer climates exhibit a significant excess of nonsynonymous polymorphisms in *CBF* genes and revealed a *CBF* haplotype network whose structure points to multiple independent transitions to warmer climates. We also identified a number of newly described mutations of significant functional effect in *CBF* genes, similar to the frameshift mutation previously indicated to be locally adaptive in Italy, and find that they are significantly associated with warm winters. Lastly, we uncover relationships between climate and the position of significant functional effect mutations between and within *CBF* paralogs, suggesting variation in adaptive function of different mutations. Cumulatively, these findings support the hypothesis that disruption of *CBF* gene function is adaptive in warmer climates, and illustrate how parallel evolution in a transcription factor can underlie adaptation to climate.

Keywords: adaptation, climate, C-repeat binding factor, gene structure and function, landscape genetics, molecular evolution

Received 27 October 2015; accepted 20 May 2016

Introduction

Evidence for adaptation to local environments in nature is widespread (Leimu & Fischer 2008; Hereford 2009; Fournier-Level *et al*. 2011), and adaptive loci have been identified (Shapiro *et al*. 2004; Hoekstra *et al*. 2006; Uga *et al*. 2013; Des Marais *et al*. 2014; Pardo-Diaz *et al*. 2015). However, long-standing questions remain about the processes underlying adaptive evolution. What components of the environment are important selective drivers of divergent adaptive evolution? (Levins 1968; Endler 1986; Mullen & Hoekstra 2008; Lasky *et al*. 2012). At the species level, does molecular evolution proceed by the fixation of a single adaptive allele, or in parallel, through mutations appearing independently in different populations? (Maynard Smith & Haigh 1974; Pennings & Hermisson 2006; Ralph & Coop 2015; Remington 2015).

The flowering plant *Arabidopsis thaliana* (L.) Heynh. (hereafter referred to as *Arabidopsis*) is a well-established model for both functional and ecological genetics

Correspondence: John Grey Monroe, Fax: 970 491 3862; E-mail: monroejg@colostate.edu

(Mitchell-Olds 2001) and shows local adaptation across its range (Fournier-Level *et al.* 2011; Ågren & Schemske 2012). *Arabidopsis* populations, having expanded north following the end of the last Pleistocene glaciation (Beck *et al.* 2008), are found in environments that differ greatly in thermal regime, creating a gradient across which populations may have become locally adapted. For example, a reciprocal transplant between *Arabidopsis* accessions originating from the northern and southern portions of the species range, Sweden and Italy, respectively, revealed strong local adaptation and pointed to freezing tolerance as an important adaptive trait (Ågren & Schemske 2012). In a subsequent experiment, Ågren *et al.* (2013) carried out a long-term reciprocal transplant of a mapping population to uncover the genetic basis of local adaptation in the Italian and Swedish ecotypes. In this experiment, fitness (quantified by both survival and fruit production) was measured in an Italy × Sweden derived mapping population of recombinant inbred lines (RILs) planted in Italian and Swedish environments for three consecutive years. Genomic regions explaining differences in fitness between individuals in Italian and Swedish locations were identified by quantitative trait loci (QTL) mapping. They identified 15 QTL underlying variation in fitness in Italian and Swedish environments, one of which localized to a genomic region containing three transcription factors known to function in freezing tolerance called C-repeat binding factors (*CBF*s). Notably, this QTL exhibited a genetic trade-off; in Italy, the Italian allele in this region was adaptive while the Swedish allele was deleterious, with the opposite being true in Sweden. Cumulatively, these findings suggest that a fitness trade-off associated with freezing tolerance genes is driving local adaptation between *Arabidopsis* populations diverged along a thermal gradient (Ågren & Schemske 2012; Ågren *et al.* 2013).

Cold acclimation, the processes whereby plants respond to low but nonfreezing temperatures and achieve tolerance to future freezing conditions, has been extensively studied. Plants have evolved strategies to prepare for freezing conditions in order to prevent tissue damage caused by cellular freezing. In brief, a drop in temperature towards 0 °C cues plants to activate a suite of genes that make cellular osmotic adjustments conferring freezing tolerance to anticipated sub -0 °C temperatures (Zarka *et al.* 2003; Hannah *et al.* 2006). *Arabidopsis* has served as a model to elucidate the genetic and molecular basis of freezing tolerance. In particular, the role of *CBF*s in the regulation of genes involved in freezing tolerance has been well characterized (Thomashow 2001; Thomashow 2010; Park *et al.* 2015). The *CBF*s are a paralogous family of transcription factors, three of which are located in a tandem array on chromosome 4. Each encodes a protein containing an AP2 DNA binding domain that recognizes and binds CRT/DRE promoter regions of target genes (Stockinger *et al.* 1997; Gilmour *et al.* 1998), and a C-terminal activation domain responsible for the recruitment of transcriptional machinery (Stockinger *et al.* 1997; Wang *et al.* 2005; Kang *et al.* 2013). An increase in expression of the *CBF* genes occurs in a quantitative manner as the environment in which plants are growing approaches low but nonfreezing temperatures (Medina *et al.* 1999; Zarka *et al.* 2003). This increase in CBF expression results in the activation of over 100 gene targets, known as the *CBF* regulon (Thomashow 2001; Park *et al.* 2015).

Using the same mapping population used to find fitness QTL by Ågren *et al.* (2013), the genomic region containing the *CBF* genes was again detected by QTL mapping as explaining divergence in freezing tolerance between the Italian and Swedish ecotypes (Oakley *et al.* 2014). More recently, Gehan *et al.* (2015) analysed these Italian and Swedish *CBF* genes and discovered that the Italian allele contains a 13-bp deletion causing a frameshift mutation that disrupts the activation domain of *CBF2*. With transgenic experiments, they confirmed that this polymorphism disrupts CBF2 function and is responsible for divergence of freezing tolerance between the Italian and Swedish *Arabidopsis*. In the light of the results of Ågren *et al.* (2013), the findings of Gehan *et al.* (2015) suggest that the frameshift mutation in *CBF2* confers an adaptive loss of freezing tolerance in the Italian ecotype. The effect that this mutation has on freezing tolerance is consistent with knowledge of the central role that *CBF* genes play in cold response. Indeed, variation in *CBF* genes has been indicated in underlying divergence in freezing tolerance between other *Arabidopsis* accessions (Alonso-Blanco *et al.* 2005; Kang *et al.* 2013). More intriguing is the adaptive value this frameshift mutation in *CBF2* appears to have in Italy, which suggests that there is a fitness penalty for freezing tolerance in Italy that is avoided by a significant functional mutation in *CBF2*. While freezing tolerance has been shown to vary along a latitudinal gradient, decreasing in populations from lower latitudes (Hannah *et al.* 2006; Zhen & Ungerer 2008; Zuther *et al.* 2012), the adaptive significance of this pattern has been unclear. It is possible that unnecessary *CBF*-induced cold responses negatively impact phenotypes. Indeed, transgenic *Arabidopsis* lines overexpressing *CBF* perform poorly under above-freezing temperatures (Thomashow 2010 and references therein). The recent studies of locally adapted Italian and Swedish ecotypes (Ågren *et al.* 2013; Oakley *et al.* 2014; Gehan *et al.* 2015) experimentally connect increased fitness, measured in the field in Italy, and a frameshift mutation in a *CBF* gene that causes a significant reduction in freezing tolerance. This background

lays forth a framework and opportunity to investigate how locally adaptive molecular evolution proceeds at a continental scale, by considering a priori that a *CBF* frameshift mutation is selectively favoured in at least one southern environment. What remains unknown is the role of *CBF* genes in local adaptation in *Arabidopsis* populations elsewhere across the species range.

To address the hypothesis that adaptation to warmer climates is associated with variation in *CBF* across the range of *Arabidopsis,* we asked four questions: Is variation in *CBF* genes associated with a particular environmental gradient? Are *CBF* genes evolving neutrally along this gradient? If not, is there evidence for a hard selective sweep of a single adaptive allele or parallel evolution of multiple independently derived alleles? Is the geographical distribution of frameshift and premature stop codon alleles in *CBF* genes nonrandom with respect to the environment? Here, we generated or obtained sequences for *CBF1, CBF2* and *CBF3* from 477 accessions originating from across the species' range, comparing these sequence data with variables characterizing the environments of the accession's locations to assess the relationship between climate and molecular variation in *CBF* genes.

Materials and methods

Plant and genetic material

Extracted DNAs from 136 *Arabidopsis* accessions collected across Eurasia (Beck *et al.* 2008) (Table S1, Supporting information) were included. Additionally, we extracted DNA from ecotypes Castelnuovo (IT), Rödåsen (SW), Kas and Tsu-1 (McKay *et al.* 2008; Ågren & Schemske 2012; Ågren *et al.* 2013). We also downloaded sequence data for a subset of the *CBF* gene family from geo-referenced accessions of the MPICWang2013 collection made available through the *Arabidopsis* 1001 genomes project (see Acknowledgements and Heyndrickx *et al.* 2014). Specifically, we analysed *CBF1*, *CBF2* and *CBF3* coding region sequences (337 total), which are found in a tandem array on chromosome 4 and include the paralog containing a functional mutation in an Italian ecotype (*CBF2*, Ågren *et al.* 2013). We also downloaded the reference genome for *Arabidopsis lyrata* (L.) O'Kane & Al-Shehbaz (Hu *et al.* 2011).

CBF sequences

For accessions collected from the field (Beck *et al.* 2008) and IT, SW, Kas and Tsu-1, the coding regions of *CBF1*, *CBF2* and *CBF3* paralogs were PCR-amplified with the following primers: *CBF3*, forward primer sequence TTT TCC ACT CGT TTC TAC AAC A, *CBF3* reverse primer sequence CTA CTT AAA CCT TAT CCA GTT T, *CBF1* forward primer sequence TCA ATT TAA TTT ACA CTC GTT T, *CBF1* reverse primer sequence TTT CAG CAA ACC ATA CCA ACA, CBF2 forward primer sequence ACA TTC GTT TCT CAC AAC CAA and *CBF2* reverse primer sequence TCT CAT AAA CCT TAT CCA GTT T. All amplicons were Sanger-sequenced on an ABI 3130xL Genetic Analyzer using samples prepared with ABI's BigDye® TERMINATOR v3.1 kit. The *CBF* coding regions from the MPICWang2013 accessions were obtained digitally with a custom python script. This script is available for download at https://bitbucket.org/greymonroe/scripts_CBF. *CBF* sequences used in this study will be made available on Data Dryad.

Sequence alignment and polymorphism characterization

The coding regions of *CBF1*, *CBF2* and *CBF3* were concatenated into a single sequence. This concatenated sequence was aligned by MUSCLE using MEGA v6.06 (Tamura *et al.* 2013). A single nucleotide polymorphism (SNP) matrix was then generated by scoring individuals at each polymorphic locus as follows using custom R scripts made available for download at https://bitbucket.org/greymonroe/scripts_CBF. At each biallelic polymorphic site, an accession was scored '0' if the allelic state was shared by *A. lyrata* or '1' if the allelic state differed from *A. lyrata*. Sites containing more than one allelic state and indels were ignored in the generation of the SNP matrix. To analyse the role of significant functional mutations, frameshift and premature stop codon mutations were identified by examining translated protein sequences of each accession using Columbia gene models (TAIR 10).

Climate variables and analysis

We wanted to estimate the degree to which variation in SNPs in *CBF* genes was explained by different parameters of local environments. We compiled 116 environmental variables describing the known collection location for each accession (Table S1, Supporting information). Specifically, we obtained data from WORLDCLIM (Hijmans *et al.* 2005), CGIAR-CSI GLOBAL-ARIDITY database (Zomer *et al.* 2008), vapour pressure deficit using CLIMATE RESEARCH UNIT (CRU) data (New *et al.* 2002), NCEP reanalysis data (Kalnay *et al.* 1996), SRB data (https://eosweb.larc.nasa.gov/project/srb/srb_table), soil water capacity data (Dunne and Willmott, http://daac.ornl.gov/SOILS/guides/DunneSoil.html), a groundwater data set (Fan *et al.* 2013), a global soil aluminium

toxicity data set (Sanchez *et al.* 2003) and a soil pH data set (Nachtergaele & Batjes 2012). Details regarding these data sets are explained in Lasky *et al.* (2012). Additionally, we calculated the frequency of drought and favourable conditions using the vegetative health index (VHI) which uses historic normalized differential vegetative index (NDVI) and thermal condition index (TCI) to calculate drought (VHI <40) and favourable conditions (VHI >60) (Kogan *et al.* 2004). We chose such a large number of environmental variables so that our analysis could identify variables potentially causative of *CBF* variation. To assess the explanatory contribution of different aspects of climate on polymorphisms in *CBF* genes across landscapes, we used redundancy analysis, a multiple regression approach which analyses the relationship between independent and dependent variables that are each multivariate (Legendre & Legendre 1998). Here, those are the SNP matrix and climate data matrix, respectively. In our redundancy analysis, we excluded 53 of the 477 sequences that had incomplete climatic data (Table S1, Supporting information). We performed this analysis in R (R Core Team 2015) using the package 'VEGAN' (Okasanen *et al.* 2015). We calculated the explanatory contribution (P_x) of each climate variable as described in Lasky *et al.* (2012).

We wanted to determine what types of climate variables have the strongest association with CBF polymorphism. To do this, we tested for significant differences in the proportion of CBF polymorphism variation explained between groups of highly correlated climatic variables. First, we performed a hierarchical clustering of climate variables using absolute correlations to calculate Euclidean distance between variables. This clustering analysis revealed four distinct cluster groups of climate variables. Then, with P_x for each individual climate variable determined by the redundancy analysis, we compared the mean P_x between each of the four climate variable groups using Tukey-adjusted 95% confidence intervals.

Molecular evolution analysis

We identified a strong relationship between winter temperatures and CBF polymorphism through climate variable clustering and redundancy analysis. We wanted to further investigate the evolution of CBF genes along an ecologically significant climatic gradient related to winter temperature. Minimum temperature of the coldest month (Min.Tmp.Cld.M, Table S1, Supporting information) indicates the potential severity of cold-induced stress and was used for the remaining analyses. We performed a binomial logistical regression to assess the relationship between the minimum temperature of coldest month and the probability of an accession having a *CBF* allele containing a frameshift or premature stop codon. Additionally, to compare molecular evolution of *CBF* genes between different climates, we grouped accessions into 5 °C bins by the minimum temperature of the coldest month. We then performed a McDonald–Kreitman test on concatenated *CBF* sequences in each group using *A. lyrata CBF* sequences as an out-group. The test was carried out using the online tool developed by Egea *et al.* (2008, http://mkt.uab.es/mkt/). Here, Jukes–Cantor-corrected divergence values representing synonymous and nonsynonymous nucleotide divergence from *A. lyrata* fixed in the *A. thaliana* samples were compared with synonymous and nonsynonymous nucleotide polymorphisms within the samples in a 2 × 2 chi-square contingency analysis, testing for significant divergence from the neutral expectation that these ratios are equal.

Haplotype networks

We concatenated all three *CBF* gene sequences for the identification of haplotypes. Median-joining networks (Bandelt *et al.* 1999) of all haplotypes with an allele frequency >1% (at least four accessions) were generated with the software POPART (http://popart.otago.ac.nz). Because POPART cannot consider indels, we manually added two haplotypes that were distinguished by a single indel mutational event. The minimum temperature experienced in the coldest month was averaged across accessions with each haplotype in this network.

Results

Climate variable clusters

Four distinct clusters were identified in the dendrogram produced by hierarchically clustering climatic variables by their correlation with each other (Fig. S1, Supporting information). An inspection of the variables within each cluster reveals that they generally group both by type of variable (temperature/precipitation) and season of variable ('Winter'/'Summer') (Fig. S1 and Table S2, Supporting information). Thus, the four cluster groups contain variables related to or correlated with winter temperatures, summer temperatures, winter precipitation or summer precipitation.

Climate and CBF polymorphisms

To explore the relationship between climate and *CBF* variation, we performed a redundancy analysis to determine the contribution (P_x) of each climatic variable in explaining genetic variation in *CBF* genes across the range of *Arabidopsis*. We then compared mean P_x of

climate variables between each climate variable cluster. 'Winter Temperature' variables had significantly greater P_x values than all other variable clusters, and 'Summer Temperature' variables had significantly greater P_x values than 'Summer Precipitation' (Fig. 1). These results show that CBF polymorphism is not random with respect to climatic variables. Temperatures during cold months of the year are the most significant predictors of *CBF* molecular polymorphisms.

McDonald–Kreitman test

The results of the McDonald–Kreitman tests show that the *CBF* sequences in accessions that experience minimum temperatures between −5 °C to 0 °C ($X^2 = 7.789$, *P*-value = 0.005) and 0 °C to 5 °C ($X^2 = 4.279$, *P*-value = 0.038) are significantly enriched for nonsynonymous polymorphisms. *CBF* sequences in accessions in other minimum temperature bins did not deviate from neutral expectations of sequence polymorphism (Table 1).

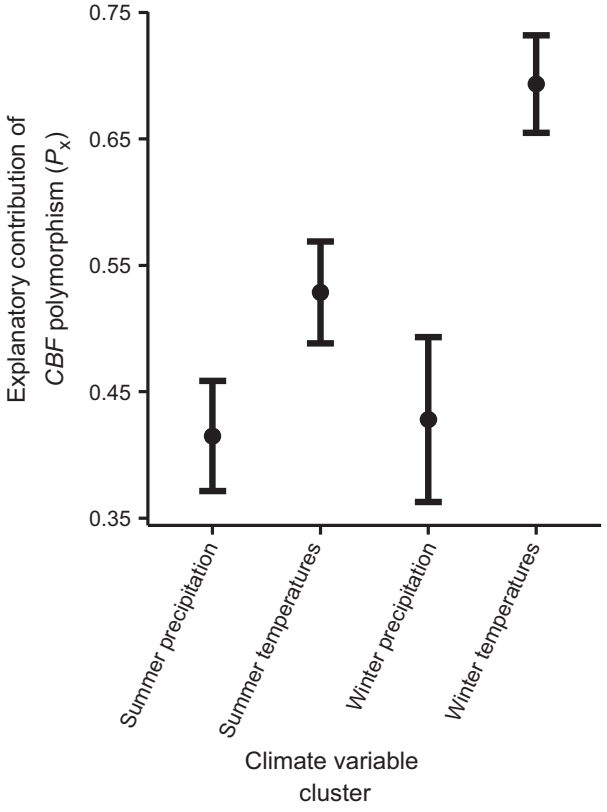

Fig. 1 Means and 95% confidence intervals for variance in CBF polymorphisms explained (P_x) by different clusters of climate variables. Variance explained by each variable was determined using redundancy analysis, and climate variable groups were formed by hierarchical clustering based on correlations between variables.

CBF haplotype network

We identified 177 unique haplotypes across the three concatenated *CBF* genes (Table S1, Supporting information). The network displaying haplotypes with a frequency >1% is displayed in Fig. 2. This network suggests multiple, evolutionarily independent haplotypes associated with warmer minimum winter temperatures, including haplotypes containing premature stop codon and frameshift mutations (Fig. 2).

CBF frameshift and premature stop codon mutations

A frameshift mutation causing a premature stop codon in *CBF2* appears to underlie a genetic trade-off between locally adapted populations, selectively favoured in Italy (Ågren *et al.* 2013; Oakley *et al.* 2014; Gehan *et al.* 2015). We found 10 additional frameshift mutations and three unique premature stop codon mutations (Fig. 3). These mutations, which presumably have a large effect on *CBF* function, were particularly interesting given the adaptive value that a similar mutation appears to have in Italy.

Several patterns emerge from the distribution of these mutations between and within *CBF* genes. Only one accession with a premature stop codon was observed in *CBF1*. In contrast, 14 accessions with five different mutations in *CBF3* and 12 accessions with seven mutations were observed in *CBF2*. Interestingly, we also found that the number of accessions containing these mutations that were downstream of the AP2/ERF DNA binding domain was 70% higher (17 vs. 10) than the number of accessions with these mutations affecting the protein sequence of this domain. The affected *CBF* proteins in accessions with mutations downstream of the DNA binding domain may retain the capacity to bind promoter regions of target genes but lack the domain responsible for recruitment of transcriptional machinery for gene activation (Park *et al.* 2015).

Climate and CBF frameshift and premature stop codon mutations

The results of the redundancy analysis identified variables describing winter temperatures as having a strong association with *CBF* allelic variation. We wanted to assess the role that winter temperatures, specifically minimum temperature of coldest month, may play in driving functional changes in *CBF* genes by analysing the climate distribution of frameshift or premature stop codon mutations. The map in Fig. 4 shows the geographical distribution in Eurasia of these mutations across landscapes differing in the minimum temperature of the coldest month. Accessions grouped by 5 °C

Table 1 McDonald–Kreitman test of molecular evolution *CBF1-3* in *Arabidopsis thaliana* accessions grouped by minimum temperature of the coldest month. *CBF* sequences from *Arabidopsis lyrata* were used to calculate between species molecular divergence. Polymorphisms reflect within-species variation in the *A. thaliana* samples. The asterisk indicates $P < 0.05$.

Minimum temperature coldest month	Synonymous polymorphism	Nonsynonymous polymorphism	Synonymous divergence (Jukes–Cantor adjusted)	Nonsynonymous divergence (Jukes–Cantor adjusted)	X^2	*P*
5 °C to 10 °C	17	19	77.74	89.73	0.007	0.93
0 °C to 5 °C	24	51	75.34	87.61	4.279	0.038*
−5 °C to 0 °C	33	69	80.53	81.08	7.789	0.005*
−10 °C to −5 °C	19	34	75.19	87.55	1.742	0.186
−15 °C to −10 °C	4	11	80.26	95.18	2.039	0.153
−20 °C to −15 °C	8	7	80.28	96.28	0.344	0.557
−25 °C to −20 °C	5	4	85.46	98.47	0.284	0.593

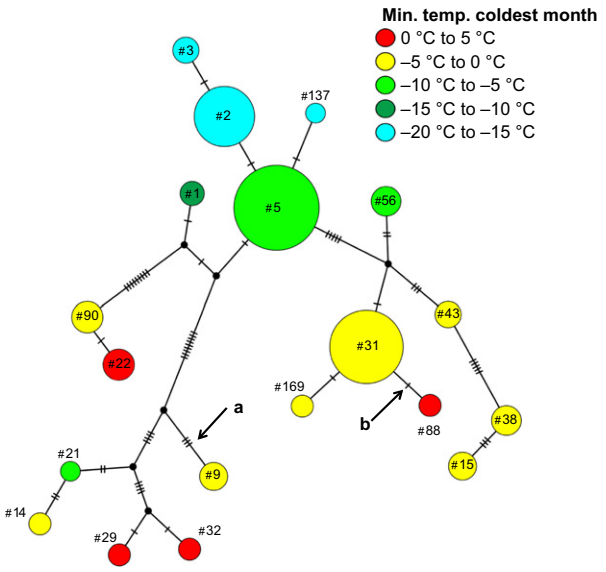

Fig. 2 Median-joining tree haplotype network of concatenated *CBF* genes showing haplotypes with an allele frequency >1%. The size of the circles is scaled to represent haplotype frequency. Each is coloured to indicate the minimum temperature of coldest month averaged across individuals exhibiting that haplotype. Solid black dots represent inferred but unsampled haplotypes. Tick marks along the lines connecting haplotype circles represent single nucleotide polymorphism (SNP) and indel mutations. All haplotypes appearing in this network are labelled by the haplotype numbers presented in Table S1 (Supporting information). Two notable mutations, unique to haplotypes also appearing in Fig. 3 and labelled on the map in Fig. 4, are highlighted here; (a) premature stop codon in *CBF3* (b) frameshift mutation in *CBF2*.

bins of minimum temperature of coldest month show an increasing relationship in the frequency of individuals within each bin having one of these mutations (Fig. 5). Additionally, a logistical regression reveals a significant positive relationship between minimum temperature of coldest month and the probability of an accession having a frameshift or premature stop codon mutation in one of its *CBF* genes ($\beta = 0.017721$, $t = 3.281$, *P*-value = 0.00103).

We tested for differences in climate experienced by accessions with frameshift and premature stop codon mutations differentially distributed between and within *CBF* paralogs. We compared the minimum temperature of coldest month between accessions with these mutations in *CBF2* and *CBF3* genes (Fig. 6). A Mann–Whitney test was performed because a Shapiro–Wilk test revealed non-normality in minimum temperatures of *CBF2* mutated accessions ($W = 0.82979$, *P*-value = 0.02085). The Mann–Whitney test results demonstrate that accessions with *CBF2* mutations (median = 1.44 °C) come from locations experiencing significantly warmer winters than those with mutations in *CBF3* (median = −0.592 °C) ($W = 134$, *P*-value = 0.005347). We also tested to see whether accessions containing frameshift or premature stop codon mutations affecting one vs. both functional domains of *CBF* genes differed significantly in their minimum temperature of coldest month. While accessions with retained AP2 DNA binding domain were on average from environments with warmer winters (mean = 0.724 °C) compared to accessions having both domains disrupted (mean = −0.367 °C), the difference was not significant (Student's *t*-test $t = 0.98882$, d.f. = 22.117, *P*-value = 0.1667).

Discussion

The environmental agents of selection

Here, we examined polymorphism across the species range in the *CBF* transcription factors, which have been shown to play a central role in freezing tolerance and appear to underlie local adaptation in *Arabidopsis* populations at the northern and southern range limits. We found that *CBF* variation is strongly associated with variables describing winter temperatures. Looking

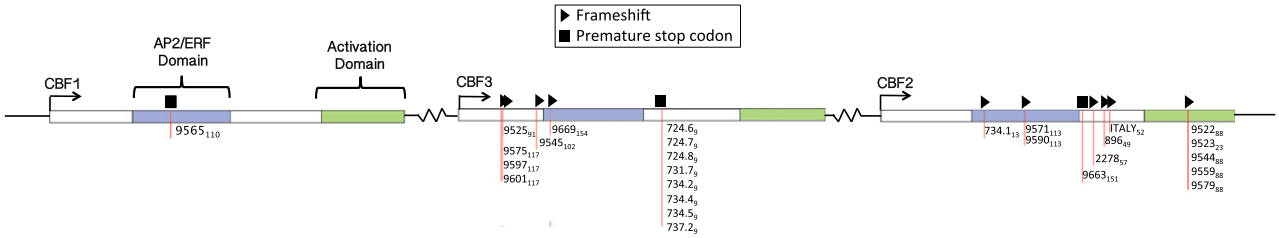

Fig. 3 Locations of frameshift and premature stop codon mutations in *CBF1*, *CBF2* and *CBF3*. Arrows show the direction of transcription. The AP2/ERF domain, which binds C-repeat promoter regions of target genes, is noted in blue; the activation domain which recruits transcriptional machinery is appears in green. The subscript in the accession name indicates the *CBF* haplotype which that accession contains (Table S1, Supporting information).

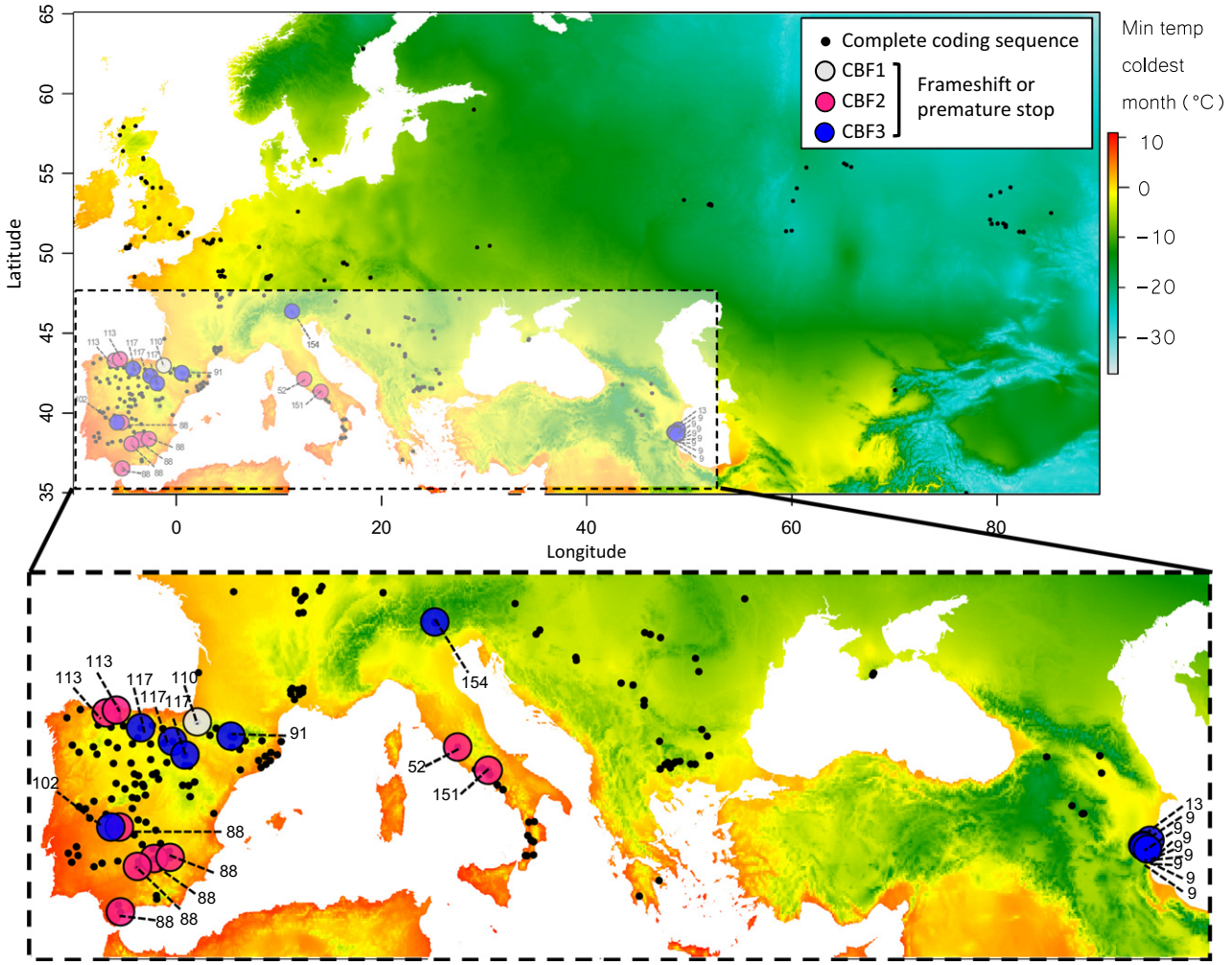

Fig. 4 Map showing locations of Eurasian *Arabidopsis* with complete *CBF* coding sequences and those with *CBF* alleles containing frameshift or premature stop codon mutations. Accessions with complete *CBF* coding sequenced are marked with black dots. Accessions with a frameshift or premature stop codon mutation (see Fig. 3) are coloured by the *CBF* paralog in which this mutation is found and labelled by their *CBF* haplotype number (Table S1, Supporting information). The landscape is coloured according to the minimum temperature of coldest month. For the sake of visualization, only Eurasia is mapped, excluding 23 accessions from North America and Eastern Asia.

specifically at minimum temperature of coldest month, we found that accessions from warmer climates exhibit a significant excess of nonsynonymous polymorphisms and reveal a *CBF* haplotype structure consistent with multiple independent transitions to warmer climates. Furthermore, we identified a number of newly

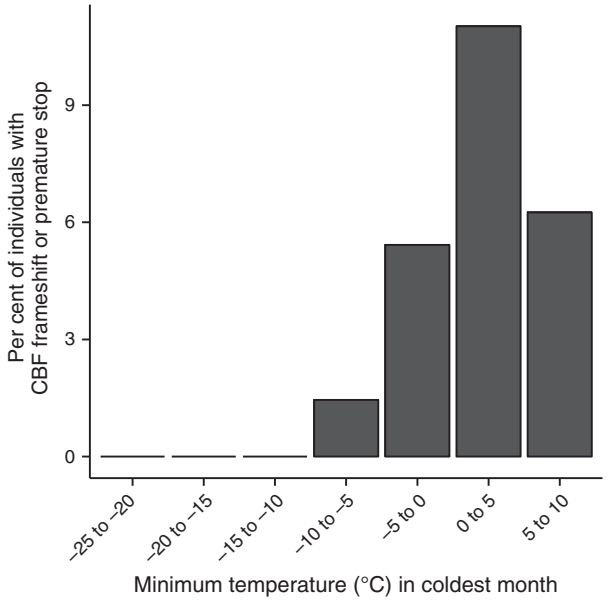

Fig. 5 Per cent of accessions (binned by minimum temperature of coldest month in 5 °C increments) that exhibit *CBF* sequences with frameshift or premature stop codon mutations. The frequency of such mutations is significantly positively correlated with the minimum temperature of coldest month (logistic regression, β = 0.017721, t = 3.281, P-value = 0.00103).

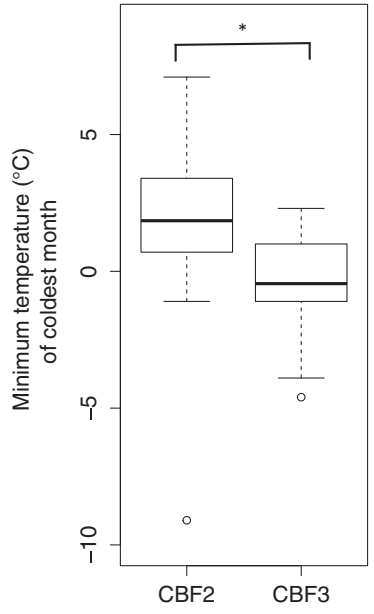

Fig. 6 Comparison of the minimum temperature of coldest month experienced by accessions with *CBF2* verses *CBF3* frameshift of premature stop codons. Mann–Whitney test (W = 134, P-value = 0.005347).

discovered *CBF* alleles containing mutations causing major protein modification, similar to the locally adaptive frameshift mutation in Italy. We found that these alleles are significantly associated with warmer minimum winter temperatures. Lastly, we uncovered relationships between climate and the position of large functional effect mutations between and within *CBF* paralogs suggesting variation in adaptive function of different mutations.

We found that variables describing winter temperatures are predictive of polymorphism in *CBF* genes. *Arabidopsis* plants exhibiting spring, summer or fall annual life cycles (i.e. completing life cycle within a single growing season) do not experience winter months in a vegetative state. However, a recent review of life history strategies observed in European *Arabidopsis* populations suggests that a winter annual life cycle in which plants overwinter as a vegetative rosette is the most common (Burghardt *et al.* 2015 and references therein). Accordingly, most *Arabidopsis* plants probably experience the coldest month of the year during the vegetative stage. The finding that winter temperatures are the strongest predictors of CBF variation is consistent with the hypothesis that winter temperatures are driving the evolution of *CBF* genes and logical given the known biological role of this gene in conferring freezing tolerance.

Parallel evolution and landscape heterogeneity

A central question in evolutionary biology is whether adaptation proceeds by the fixation of a single advantageous allele or in parallel by alleles with independent mutational origins (Maynard Smith & Haigh 1974; Pennings & Hermisson 2006; Chan *et al.* 2010; Haasl and Payseur 2015; Ralph & Coop 2015; Remington 2015). The McDonald–Kreitman tests reveal a significant excess of nonsynonymous polymorphisms in *CBF* genes in accessions from warmer climates (Table 1). This excess of nonsynonymous polymorphisms is consistent with local selection favouring loss of function alleles, leading to the retention of multiple adaptive alleles of independent evolutionary origin in warm climates. The haplotype network (Fig. 2) provides a visual demonstration that the transition to warm environments is accompanied by multiple independent trajectories of molecular evolution in *CBF* genes. These results indicate that *CBF* sequences in accessions from warmer climates are evolving non-neutrally and in parallel, supporting the hypothesis that mutations disrupting ancestral *CBF* function are locally favoured environments experiencing warmer minimum temperatures. This is consistent with the observed reduced freezing tolerance in *Arabidopsis* from southern populations (Zhen & Ungerer 2008) and other examples of the evolution of paralogous transcription factor families leading to diversification of morphology (Rensing 2014) and

stress responses within species (Des Marais *et al.* 2015; Lehti-Shiu *et al.* 2015). Although this study focuses only on coding sequences, a loss of function phenotype could also be achieved through mutation in regulatory regions (Alonso-Blanco *et al.* 2005).

Given that a functional polymorphism, specifically a frameshift mutation in *CBF2,* is contained in a QTL explaining fitness in at least one warm climate (Ågren *et al.* 2013) and responsible for loss of freezing tolerance (Gehan *et al.* 2015), we set out to address whether this allele shows evidence of moving to fixation in warm environments. However, the 13-bp deletion that was indicated in conferring a fitness advantage in Italy was found in no other accessions in this study. Instead, we found 12 additional alleles representing other functionally disruptive mutations such as frameshift and premature stop codons in *CBF* genes, and show that they tend to occur where minimum temperatures are warmer. This is consistent with parallel adaptive evolution in which multiple functionally disruptive mutations appear independently and are locally favoured by selection in warm climates.

Two important factors may influence the prevalence of hard selective sweep verses parallel evolutionary pathways: the genotype–phenotype map (Remington 2015) and the degree to which landscapes are heterogeneous with respect to selection pressures (Ralph & Coop 2015). For example, the molecular target size for adaptive mutations may differ such that an adaptive phenotype is produced only by a particular amino acid substitution, a mutational event that occurs at low frequency. Alternatively, an adaptive phenotype may be produced by any mutation that disrupts protein function, a mutational event that occurs at a much higher frequency (Pennings & Hermisson 2006). If the mutational target size is large, such as when an adaptive allele is generated anytime a mutation causes an amino acid change that produces a protein with disrupted function, the probability of a locally favoured allele being generated by mutation is greater. This effective increase in mutation rate in turn increases the probability of parallel molecular evolution in a heterogeneous landscape (Ralph & Coop 2015). Thus, patchy selection pressures in a heterogeneous environment may result in adaptive evolutionary trajectories involving parallel loss of function.

It is important to recall that the 13-bp deletion appearing to be adaptive in Italy exhibited a genetic trade-off; that is, it was maladaptive in a cold Swedish climate (Ågren *et al.* 2013). Geographic heterogeneity may limit the flow of warm-adaptive alleles between populations separated by landscape features such as mountain ranges where such alleles may be deleterious during cold alpine winters. Indeed, the natural range of *Arabidopsis* is considerably heterogeneous with respect to minimum temperature of coldest month, and different loss of function mutations appear locally restricted within warm patches (Fig. 4).

These findings have implications for techniques currently used to find loci under selection. Genomewide scans for selection (GWSS) continue to be a popular technique for detecting loci undergoing positive selection in nature (Haasl & Payseur 2015). Many of the methods used to scan for signatures of selection are based on models that consider the selective sweep of a single allele arising from one mutational event (Pennings & Hermisson 2006). We find support for the hypothesis that disruption of *CBF* gene function is advantageous for *Arabidopsis* accessions in warmer climates and that this is achieved through parallel evolution rather than a hard selective sweep. While it appears that *CBF* genes are experiencing selection in nature, they would be unlikely to be detected in a genomewide scan for selection based solely on outlier SNPs. However, sliding window approaches and association with climate determined by redundancy analysis may be able to detect evidence of selection in genes evolving similarly to *CBF* (Nielsen *et al.* 2005; Sasaki *et al.* 2015; Forester *et al.* 2016). From this perspective, we advise caution when interpreting negative results produced by some methods of genomic scans for selection.

CBF functional mutations

We found that the probability of an accession exhibiting a frameshift or premature stop codon in *CBF* genes increases significantly where minimum winter temperatures are warmer. Given previous reports indicating that a *CBF2* frameshift mutation is selectively advantageous in a warm Italian environment (Ågren *et al.* 2013; Gehan *et al.* 2015), the present findings may indicate that similar loss of function mutations are selectively advantageous in other warm environments within the species range. Thus, previously reported observations of major functional mutations in *CBF* genes from southern *Arabidopsis* populations (Zhen & Ungerer 2008; Kang *et al.* 2013) are corroborated here by a statistically significant link between climate and functional polymorphisms in *CBF* genes. We acknowledge the difficulty in distinguishing local or positive selection from relaxed purifying selection. Indeed, it is theoretically possible that the significant increase in *CBF* loss of function mutations in environments with warmer winters could be due to relaxed selection on *CBF* function or drift. However, field experiments suggest a significant fitness advantage conferred by a *CBF2* frameshift mutation in the field in Italy (Ågren *et al.* 2013; Gehan *et al.* 2015). Additionally, the results of the McDonald–Kreitman test performed

here show that molecular evolution of *CBF* genes in warmer climate deviates from the expectation if these genes are evolving neutrally. In the light of this, we feel it is probably that the significantly greater frequency of major functional *CBF* gene mutations in *Arabidopsis* from warmer climates revealed in the present study is driven in large part by climate mediated selective pressure for *CBF* loss of function alleles.

A more detailed look at the frequency of these mutations within climate bins may yield a more nuanced understanding of the evolutionary forces acting on this gene family. Accessions frequently experiencing temperatures low enough to induce *CBF* gene expression (Zarka *et al.* 2003), but rarely experiencing freezing temperatures could be most affected by any fitness penalties associated with unnecessary *CBF*-induced freezing tolerance. Indeed, when we binned the accessions by 5 °C increments and calculated the per cent of individuals with a frameshift or premature stop codon in *CBF* genes (Fig. 4), a notable peak appears in the 0 °C to 5 °C group. Additionally, the average (across years) minimum temperature experienced in the Italian site where a *CBF2* frameshift mutation is confirmed to be selectively advantageous falls within this range (2.7 °C, Table S1, Supporting information). Conversely, in the group of accessions experiencing minimum winter temperatures from 5 °C to 10 °C, we observe a decline in the trend towards increasing frequency of *CBF* frameshift and premature stop codon mutations. Temperatures above 5 °C may not induce *CBF* expression. Researchers studying *CBF* function have been long aware of the thermometer-like sensitivity of *CBF* expression to temperature, and temperatures below 5 °C are the methodological standard for inducing *CBF* gene expression and cold acclimation experimentally (e.g. Gilmour *et al.* 1998; Zarka *et al.* 2003; Kang *et al.* 2013). Because accessions where minimum temperatures are >5 °C infrequently face temperatures low enough to induce *CBF* expression and unnecessary *CBF*-induced freezing tolerance, they may experience a reduction in the selective pressure for functional disruption in *CBF* genes. This may also explain why accessions experiencing minimum winter temperatures from −5 °C to 5 °C show a significant excess of nonsynonymous polymorphisms whereas accessions within the 5 °C to 10 °C group do not (Table 1). These results suggest that the effects of loss of function on fitness may be nonlinear with respect to the minimum temperature of coldest month. Disruptive functional mutations in *CBF* genes may play the greatest role in adaptation to environments where *CBF* genes are frequently induced by temperatures just above freezing, but freezing stress is rare. This supports the hypothesis that there is a fitness penalty associated with unnecessary *CBF*-induced freezing tolerance, possibly due to a correlated response in an ecologically important trait such as growth rate.

The *CBF2* frameshift appearing to be adaptive in Italian *Arabidopsis* has a considerable effect on *CBF* function by eliminating its transactivation domain (Gehan *et al.* 2015). Here, we found other frameshift and premature stop codon mutations are associated with warmer winter temperatures suggesting that significant function altering mutations may underlie adaptation in these genes. We also found interesting patterns in the distribution of these functional mutations between and within *CBF* paralogs. First, in contrast to frameshift and premature stop codon mutations that cause disruption of both *CBF* functional domains, we found that considerably more accessions (70% more) contained such mutations that disrupted the transactivation domain alone (Fig. 3). Second, we found that the winter temperatures experienced by accessions with significant functional mutations in *CBF2* were significantly higher than those experienced by accessions with these mutations in *CBF3* (Fig. 6). Furthermore, in contrast to *CBF2* and *CBF3*, we found almost no frameshift or premature stop codons in *CBF1*. This suggests that mutations in these different *CBF* paralogs, which are often assumed to be functionally redundant, may not be functionally or selectively equivalent along a climatic temperature gradient. This highlights the important role that evolutionary studies can play in elucidating gene functional variation to guide future research. A potential explanation of our findings is that significant functional mutations in *CBF2* may be most adaptive in the warmest climates, followed by functional mutations in *CBF3*, and that disruption of *CBF1* may be maladaptive across the entire species range. Additionally, the observation that 70% more accessions with these functional mutations have intact DNA binding domains compared to accessions with loss of both domains could indicate that these mutations differentially impact phenotypes. Future empirical work will be needed to connect these findings to biology at the organismal level.

This work supports the hypothesis that divergent selection pressures in environments that differ in winter temperatures is driving adaptive functional changes in *CBF* genes. In particular, these results indicate that *CBF* genes in *Arabidopsis* from warmer winter temperatures have undergone parallel adaptive evolution involving disruption of function. This reveals how loss of function alleles generated via parallel molecular evolution may play an important role in local adaptation to rapidly changing climates.

Acknowledgements

We would like to thank C. Oakley, D. Schemske and M. Thomashow for insightful discussions on *CBF* evolution and the

researchers who generated the MPICWang2013 1001 genomes data set. These sequence data were produced by Monsanto Company and the Weigel laboratory at the Max Planck Institute for Developmental Biology. This research was supported by NSF grants DEB-1022196 and DEB-1556262 to JKM and USDA-NIFA National Needs Graduate Fellowship Program, Award no. 2014-38420-21801 to JGM.

References

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytologist*, **194**, 1112–1122.

Ågren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of local adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA*, **110**, 21077–21082.

Alonso-Blanco C, Gomez-Mena C, Llorente F, Koorneef M, Salinas J, Martìnez-Zapater J (2005) Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing tolerance quantitative trait locus in *Arabidopsis*. *Plant Physiology*, **139**, 1304–1312.

Bandelt H, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology*, **16**, 37–48.

Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Molecular Ecology*, **17**, 902–915.

Burghardt L, Metcalf C, Wilczek A, Schmitt J, Donohue K (2015) Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. *The American Naturalist*, **185**, 212–227.

Chan YF, Marks ME, Jones FC *et al.* (2010) Adaptive evolution o f pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*, **327**, 302–305.

Des Marais DL, Auchincloss LC, Sukamtoh E *et al.* (2014) Variation in *MPK12* affects water use efficiency in *Arabidopsis* and reveals a pleiotropic link between guard cell size and ABA response. *Proceedings of the National Academy of Science of the USA*, **111**, 2836–2841.

Des Marais DL, Skillern WD, Juenger TE (2015) Deeply diverged alleles in the *Arabidopsis* AREB1 transcription factor drive genome-wide differences in transcriptional response to the environment. *Molecular Biology and Evolution*, **32**, 956–969.

Dunne KA, Willmott CJ. Global distribution of plant-extractable water capacity of soil. Available from http://daac.ornl.gov/SOILS/guides/DunneSoil.html. **68.**

Egea R, Casillas S, Barbadilla A (2008) Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, **36**, W157–W162.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.

Fan Y, Li H, Miguez-Macho G (2013) Global patterns of groundwater table depth. *Science*, **339**, 940–943.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, **25**, 104–120.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–90.

Gehan MA, Park S, Gilmour SJ, An C, Lee C, Thomashow MF (2015) Natural variation in the C-Repeat Binding Factor (CBF) cold response pathway correlates with local adaptation of *Arabidopsis* ecotypes. *The Plant Journal*, **84**, 682–693.

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *The Plant Journal*, **16**, 433–442.

Haasl RJ, Payseur BA (2015) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, **25**, 5–23.

Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK (2006) Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiology*, **142**, 98–112.

Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, **173**, 579–588.

Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A functional and evolutionary perspective on transcription factor binding in *Arabidopsis thaliana*. *The Plant Cell*, **26**, 3894–3910.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.

Hu TT, Pattyn P, Bakker EG *et al.* (2011) The *Arabidopsis* lyrata genome sequence and the basis of rapid genome size change. *Nature Genetics*, **43**, 476–481.

Kalnay E, Kanamitsu M, Kistler R *et al.* (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.

Kang J, Zhang H, Sun T *et al.* (2013) Natural variation of C-repeat-binding factor (CBFs) genes is a major cause of divergence in freezing tolerance among a group of *Arabidopsis thaliana* populations along the Yangtze River in China. *New Phytologist*, **199**, 1069–1080.

Kogan F, Stark R, Gitelson A, Jargalsaikhan L, Dugrajav C, Tsooj S (2004) Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. *International Journal of Remote Sensing*, **25**, 2889–2896.

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd edn. Elsevier, New York.

Lehti-Shiu MD, Uygun S, Moghe GD *et al.* (2015) Molecular evidence for functional divergence and decay of a transcription factor derived from whole-genome duplication in *Arabidopsis thaliana*. *Plant Physiology*, **168**, 1717–1734.

Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS One*, **3**, e4010. doi: 10.1371/journal.pone.0004010.

Levins R (1968) *Evolution in Changing Environments*. Princeton University Press, Princeton, New Jersey.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.

McKay JK, Richards JH, Nemali KS *et al.* (2008) Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, KAS-1 x TSU-1. *Evolution*, **62**, 3014–3026.

Medina J, Bargues M, Terol J, Pérez-Alonso M, Salinas J (1999) The *Arabidopsis CBF* gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiology*, **119**, 463–470.

Mitchell-Olds T (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends in Ecology & Evolution*, **16**, 693–700.

Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–1570.

Nachtergaele F, Batjes N (2012) Harmonized World Soil Database. FAO, Rome. Available from: http://www.fao.org/nr/water/docs/harm-world-soil-dbv7cv.Pdf.

New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, **15**, 1566–1575.

Oakley CG, Ågren J, Atchinson RA, Schemske DW (2014) QTL mapping of freezing tolerance: links to fitness and adaptive trade-offs. *Molecular Ecology*, **23**, 4304–4315.

Okasanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: Community Ecology Package. R package version 2.3-0. Available from http://CRAN.R-project.org/package=vegan.

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, **6**, 445–464.

Park S, Lee C, Doherty CJ, Gilmour SJ, Kim Y, Thomashow MF (2015) Regulation of the *Arabidopsis CBF* regulon by a complex low-temperature regulatory network. *The Plant Journal*, **82**, 193–207.

Pennings PS, Hermisson J (2006) Soft sweeps II – molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, **23**, 1076–1084.

R Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.

Ralph PL, Coop G (2015) Convergent evolution during local adaptation to patchy landscapes. *PLOS Genetics*, **11**, e1005630. doi: 10.1371/journal.pgen.1005630.

Remington DL (2015) Alleles versus mutations: understanding the evolution of genetic architecture requires a molecular perspective on allelic origins. *Evolution*, **69**, 3025–3038.

Rensing S (2014) Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology*, **17**, 43–48.

Sanchez PA, Palm CA, Buol SW (2003) Fertility capability soil classification: a tool to help assess soil quality in the tropics. *Geoderma*, **114**, 157–185.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M (2015) "Missing" G x E variation controls flowering time in *Arabidopsis thaliana*. *PLOS Genetics*, **11**, e1005597. doi: 10.1371/journal.pgen.597.1005597.

Shapiro MD, Marks ME, Peichel CL *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the USA*, **94**, 1035–1040.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Thomashow MF (2001) So what's new in the field of plant cold acclimation? Lots! *Plant Physiology*, **125**, 89–93.

Thomashow MF (2010) Molecular basis of plant cold acclimation: insights gained from studying the *CBF* cold response pathway. *Plant Physiology*, **154**, 571–577.

Uga Y, Yamamoto E, Kanno N, Kawai S, Mizubayashi T, Fukuoka S (2013) A major QTL controlling deep rooting on rice chromosome 4. *Scientific Reports*, **3**, 3040. doi: 10.1038/srep 03040.

Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs in *Arabidopsis CBF1* COOH-terminus provide functional redundancy in trans-activation. *Plant Molecular Biology*, **58**, 543–559.

Zarka DG, Vogel JT, Cook D, Thomashow MF (2003) Cold induction of *Arabidopsis* CBF genes involves multiple ICE (inducer of CBF expression) promoter elements and a cold-regulatory circuit that is desensitized by low temperature. *Plant Physiology*, **133**, 910–918.

Zhen Y, Ungerer MC (2008) Relaxed selection on the *CBF*/DREB1 regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **25**, 2547–2555.

Zomer RJ, Trabucco A, Bossio DA, Verchot LV (2008) Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and re- forestation. *Agriculture, Ecosystems & Environment*, **126**, 67–80.

Zuther E, Schulz E, Childs LH, Hincha DK (2012) Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of *Arabidopsis thaliana* accessions. *Plant, Cell and Environment*, **35**, 1860–1878.

J.G.M. conducted data and statistical analyses, designed R code, created figures and drafted the manuscript. C.M. conducted data preparation and analysis. K.G. and K.A. carried out PCR amplifications and Sanger sequencing. J.R.L. obtained environmental data and assisted with data analyses, J.B. collected *A. thaliana* individuals and extracted DNA. J.K.M. assisted with study design and project coordination.

Data accessibility

CBF gene sequences are available on DRYAD, doi: 10.5061/dryad.96hr3. Scripts for analyses are available on bitbucket.com/greymonroe/scripts_CBF.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Dendrogram visualizing results of hierarchical clustering of climate variables based on absolute correlations between variables.

Table S1 *A. thaliana* accessions used in this study, and the climate variables that describe their location of origin.

Table S2 Results of redundancy analysis for each climate variable (explanatory contribution to *CBF* variation) and the cluster group to which each climate variable belonged.