DOI: 10.1007/s10535-015-0497-y BIOLOGIA PLANTARUM 59 (2): 245-252, 2015

ICE genes in *Arabidopsis thaliana*: clinal variation in DNA polymorphism and sequence diversification

A. KURBIDAEVA*, M. NOVOKRESHCHENOVA, and T. EZHOVA

Department of Genetics, Faculty of Biology, Lomonosov Moscow State University,
1-12 Leninskie Gory, 119991 Moscow, Russia

Abstract

Natural accessions of *Arabidopsis thaliana* exhibit a clinal variation in freezing tolerance following temperature changes across the natural habitat. Here we performed molecular evolution and population genetic analyses of homologous *INDUCER OF CBF EXPRESSION1* (*ICE1*) and *ICE2* genes, the master regulators of plant cold response. A study of *ICE* genes polymorphism was performed using 60 *A. thaliana* ecotypes grouped according to their geographic origin. The genetic diversity of *ICE2* was characterized by a high number of haplotypes and an overall high diversity. The levels of nonsynonymous nucleotide polymorphism increased from a northern group southward. On the contrary, the *ICE1* gene sequence was less diverse and there was no clinal variation in the sequence polymorphism. Thus, different selection forces acting on the *ICE2* gene might be one of the reasons of clinal variation in freezing tolerance. This clinal variation also indicates that *ICE2* is more important for a cold response than *ICE1*. The study of the ratio of numbers of nonsynonymous to synonymous substitutions (K_a/K_s) between *A. thaliana* paralogs shows that the sequence diversification and emergence of two new ICE2-specific motifs could contribute to the functional diversification of the duplicates. The K_a/K_s for *ICE2* of *A. thaliana* and *A. lyrata* orthologs was an order of magnitude greater than that for the *ICE1* orthologs, which suggests that the protein sequence of *ICE2,* an early duplicate, evolved under a weaker selective constraint. A relaxed selection on *ICE2* in southern populations and more stringent in northern populations also confirmed its role in a cold resistance. The selection pressure on *ICE1* might be caused by its role in the control of more essential than cold response functions.

Additional key words: cold acclimation, natural selection, transcription factor.

Introduction

Understanding the mechanisms of adaptive phenotypic variation represents a central aim in evolutionary biology. The extent and pattern of DNA sequence variation in natural populations can help to uncover the evolutionary forces acting on a species. One of the ways in which those forces can act on a species are climatic conditions, *e.g.*, temperature. Traits exhibiting patterns of clinal variation represent excellent models for studies of molecular adaptation, especially when variation in phenotype can be linked to fitness in different environments. To understand the genetic basis of adaptive differences, one should link variation in a phenotype to a functional polymorphism at individual genes (Stinchcombe *et al.* 2004).

Arabidopsis thaliana provides an excellent model for studying such mechanisms. Environmental conditions have been linked to variations in certain genes controlling flowering time (Le Corre *et al.* 2002) and cold response (Zhen and Ungerer 2008). These studies have shown that adaptive phenotypic differences arise from various local selection pressures, and a naturally occurring phenotypic variation in freezing tolerance could be linked to polymorphism present in the genes. Moreover, the study of natural genetic variation is an alternative means

---

Submitted 3 June 2014, *last revision* 31 October 2014, *accepted* 10 November 2014.

Abbreviations: bHLH - basic helix-loop-helix; *ICE1 - INDUCER OF CBF EXPRESSION1*; *ICE2 - INDUCER OF CBF EXPRESSION2*; InDel - insertion or deletion; K_a - number of nonsynonymous substitutions per non-synonymous site; K_s - number of synonymous substitutions per synonymous site; MYA - million years ago; NLS - nuclear localization signal; PEST - proline (P), glutamic acid (E), serine (S), and threonine (T)-rich sequence; ZIP - leucine zipper domain.

Acknowledgments: We thank the anonymous reviewers for their many insightful comments and suggestions. This work was supported by grants RFBR №11-04-01306-a and 12-04-31535-mol_a.

* Corresponding author; fax: (+7) 495 9393512, e-mail: amina.kur@gmail.com

to elucidate the functional role of certain genes in a given process (Mitchell-Olds and Schmitt 2006).

Low temperature is one of the important stresses affecting plant growth and development and a strong agent of natural selection. Natural *A. thaliana* accessions are distributed over a broad geographic range where selection pressures for a low temperature tolerance are diverse (Koornneef *et al.* 2004). Previous studies have shown that these accessions differ significantly in their ability to tolerate low temperature with and without a previous cold acclimation by determining either electrolyte leakage after freezing of detached leaves (Hannah *et al.* 2006, Zuther *et al.* 2012) or scoring for rosette tissue damage after whole-plant freezing (McKhann *et al.* 2008, Zhen and Ungerer 2008).

Cold stress induces resistance mechanisms in plants by activating sets of genes known as regulons. These genes code for transcriptional factors as well as cryoprotective proteins and enzymes. After exposure to low non-freezing temperatures, plant ability to survive freezing temperatures increases (the process is named a cold acclimation). C-repeat binding factor (CBF) proteins, which belong to the AP2/EREBP gene family including CBF1, CBF2, and CBF3, are essential transcriptional factors for this process. They regulate the expression of numerous cold responsive (*COR*) genes by binding to C-repeat/dehydration-responsive (CRT/DRE) elements in the promoter regions of these genes (Yamaguchi-Shinozaki and Shinozaki 1994, Thomashow 1999). The CBF-mediated pathway is considered as one of the major signaling pathways involved in plant cold acclimation (Medina *et al.* 2011). *INDUCER OF CBF EXPRESSION 1* and *2* (*ICE1* and *ICE2*) encoding basic helix-loop-helix (bHLH) transcription factors are the main regulators of the *CBF* genes. *ICE1* and *ICE2* genes were found to be involved in cold stress response through the cold-mediated activation of *CBF3* (Chinnusamy *et al.* 2003) and *CBF1* (Fursova *et al.* 2009) genes, respectively. The *ICE2* gene shows a high degree of homology with the *ICE1* (Fursova *et al.* 2009). Molecular clock analyses suggested that the *ICE2* gene originates approximately 17.9 MYA. *ICE1* and *ICE2* are found within a duplicated segment of the *A. thaliana* genome and are most likely not the product of the last whole-genome duplication that is speculated to have occurred between 24 and 48 MYA (Blanc *et al.* 2003, Bowers *et al.* 2003, Ermolaeva *et al.* 2003), but of a later segmental duplication (Kurbidaeva *et al.* 2014). Functional redundancy between duplicated genes is predicted to be transitory, as one gene either loses its function or gains a new function in a process known as sub-neofunctionalization (He and Zhang 2005). The genes have diverged functionally: *ICE2* provides cold tolerance of apical meristems in contrast to *ICE1* which acts on the whole plant level (Kurbidaeva *et al.* 2014).

The *ICE* genes are thus good objects for evolutionary genetic analyses in the context of geographically structured variation in freezing tolerance because they act early in the CBF cold response pathway. We aimed to study the molecular evolution and population genetic patterns of the nucleotide variation of the *ICE* transcriptional factor genes in the context of geographical variation in freezing tolerance in *A. thaliana*. We addressed the question of whether selection acted variously on the nucleotide sequences of the *ICE* genes in different regions using a divergence-based approach and neutrality tests.

Materials and methods

Sixty ecotypes of *Arabidopsis thaliana* (L.) Heynh. were used in this study (Table 1 Suppl**.**). The sequences of the *ICE2* gene were obtained from the *1001* genome database (http://1001genomes.org/). The *ICE1* and *ICE2* genes of Cvi-0 ecotype were sequenced. The seeds were chilled at 4 °C in the dark for 2 d before being grown in soil. Total DNA was isolated from young rosettes using the Dellaporta *et al.* (1983) protocol. The *ICE1* and *ICE2* genes were amplified as two overlapping segments. The pairs of primers used were ICE1-1F (5’-TGTGCAAATGTTTTGTCTGTCTT-3’), ICE1-1R (5’-GGTGGTTTTGTTGGGAACAG-3’), ICE1-2F (5’-CCTCCTCCTCCAATCTGAAC-3’), ICE1-R (5’-TGACGGTGAGAGAGAGAGAGA-3’), ICE2-1F (5’-TTTGATTATTAACGGTCGGATT-3’) and ICE2-1R (5’-TCGTCAATCTCTCTCTCGTAGC-3’), and ICE2-2F (5'-GCTCGCAGCCAACTCTGTT-3') and ICE2-2R (5’-AGCCTTTTCTCATGACAGACTC-3’). PCR products were purified using a *Cleanup* standard kit (*Evrogen*, Moscow, Russia). Two biological replicates were used. DNA sequencing was carried out by *Genome-centre* (Moscow, Russia). Sequence polymorphisms were rechecked visually from chromatograms and confirmed by comparing forward and reverse reads of the same region.

Sequences were assembled using *BioEdit v. 7.2.0* (Hall 1999) and aligned with *ClustalW* (Thompson *et al.* 1994). The calculation of nucleotide diversity, Tajima's D values (Tajima 1989), and sliding window analyses were carried out using *DnaSP 5.2* (Rozas *et al.* 2003). The percentage of sequence identity was calculated using *UGENE v. 1.11.5* (Okonechnikov *et al.* 2012). Phylogenetic relationships between accessions were reconstructed based on full-length nucleotide sequences using a neighbor-joining (NJ) (Saitou and Nei 1987) procedure with *MEGA 5.2* (Tamura *et al.* 2011, http://www.megasoftware.net). The NJ analysis was done using p-distance methods, a pairwise deletion of gaps, and default assumptions that substitution patterns among lineages, and substitution rates among sites were homogeneous. Support for each node was tested with 1 000 bootstrap replicates. Tests of heterogeneity in polymorphism-to-divergence ratios across each gene were performed using *DNA Slider* (McDonald 1998). For each test, 1 000 simulations at *R* values of 1, 2, 4, 8, 16, and 32 were run, and the highest *P* value was reported.

Potential PEST domains were predicted using *Emboss epestfind* (Rogers *et al.* 1986, http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind). Potential nuclear localization signals (NLS) were predicted using *NLStradamus* (Ba *et al.* 2009, http://www.moseslab.csb.utoronto.ca/NLStradamus/). Polyadeny-lation sites were predicted using *Poly(A) Signal Miner* (Liu *et al.* 2005).

Sequence data from this article can be found in the GenBank nucleotide sequence database under following accession numbers: KM593299 (the *ICE1* gene sequence of the Cvi-0 ecotype), KM210288 (the *ICE2* gene sequence of the Cvi-0 ecotype).

Results

For this study, we selected 60 accessions from different regions and sorted them based on their latitudinal origins. Northern accessions were designated as those from latitudes at or above 50 °N, temperate region accessions from 43 till 50 °N and southern as those from latitudes at or below 43 °N (Table 1 Suppl.). The sequences of the genes were obtained for all the 60 *A. thaliana* accessions either from the *1001* genome database, or from DNA sequencing. A total of 2 258 nt of the *ICE2* gene were analyzed. This included 405 nt in the 5'-flanking region, four exons and three introns, and 218 nt in the 3'-flanking region. We analyzed 2 419 nt of the *ICE1* gene: 294 nt in the 5'-flanking region, four exons and three introns, and 263 nt in the 3'-flanking region. Nucleotide diversity in the *ICE* genes loci is summarized in Tables 1, 2, and 3 Suppl. Genetic diversity of the *ICE2* was characterized by a high number of haplotypes (Table 1 Suppl.). Forty-six different single nucleotide polymorphism (SNP) haplotypes and 10 insertion or deletion (InDel) haplotypes were found among the 60 ecotypes studied, the *ICE1* being less diverse: 24 different SNP haplotypes and 7 InDel haplotypes were found (Table 1). There was a majority of parsimony informative polymorphic sites in both the genes (Table 1 Suppl.).

A neighbor-joining tree was constructed based on the *ICE1* and *ICE2* nucleotide sequences. Groups of certain sequences formed separated clusters with a bootstrap value of 99 %, the remaining sequences also formed several subgroups (Fig. 1 Suppl.). However, these groups comprised ecotypes from all the climatic regions, and we could not find a relationship between the tree structure and the geographical origin of the ecotypes. It seems that the genetic variability in *A. thaliana* is influenced more by its worldwide expansion and/or by selection pressures than by present patterns of migration and isolation by distance.

When considering all the ecotypes (Table 3 Suppl.), the nucleotide diversity of *ICE2* (π = 0.0109) was relatively high compared to several *A. thaliana* genes studied previously (0.0031 - 0.0104 reviewed in Aguadé 2001 and Yoshida *et al.* 2003), but lower than the overall level of genome-wide polymorphism in *A. thaliana* estimated by values between 0.06 and 0.07 (Nordborg *et al.* 2005, Schmid *et al.* 2005). This nucleotide diversity is characteristic of neutrally evolving genes in *A. thaliana* (Aguadé 2001). As it is in the case of most genes studied, the coding region of *ICE2* was much less diverse (0.007) than the non-coding region (0.0167). Interestingly, the 3'-flanking region was highly diverse (0.2749), as opposed to the low diversity of the 5'-flanking region (0.00104). The *ICE1* gene sequence was overall less diverse (π = 0.00678). The coding, noncoding, and

Table 1. Nucleotide variation in *ICE1* and *ICE2* regions of *A. thaliana*. S - a number of variable sites, h - a number of haplotypes, π - nucleotide diversity per site, S(i) - a total number of InDel sites, h(i) - a number of InDel haplotypes, π(i) - InDel diversity per site.

Gene	Group	S Singleton variable sites	Parsimony informative sites	h	π	S(i)	h(i)	π(i)	Number of accessions with stop codons in the sequence
ICE1	northern	16	32	13	0.00570	8	3	0.00042	0
	temperate	8	44	11	0.00776	8	6	0.00105	0
	southern	23	26	11	0.00457	6	4	0.00044	0
	all	18	45	24	0.00678	7	9	0.00063	0
ICE2	northern	6	37	17	0.00653	3	4	0.00036	0
	temperate	13	55	15	0.01139	8	9	0.00130	0
	southern	13	56	20	0.01075	7	8	0.00073	4
	all	20	64	46	0.0109	8	10	0.00064	4

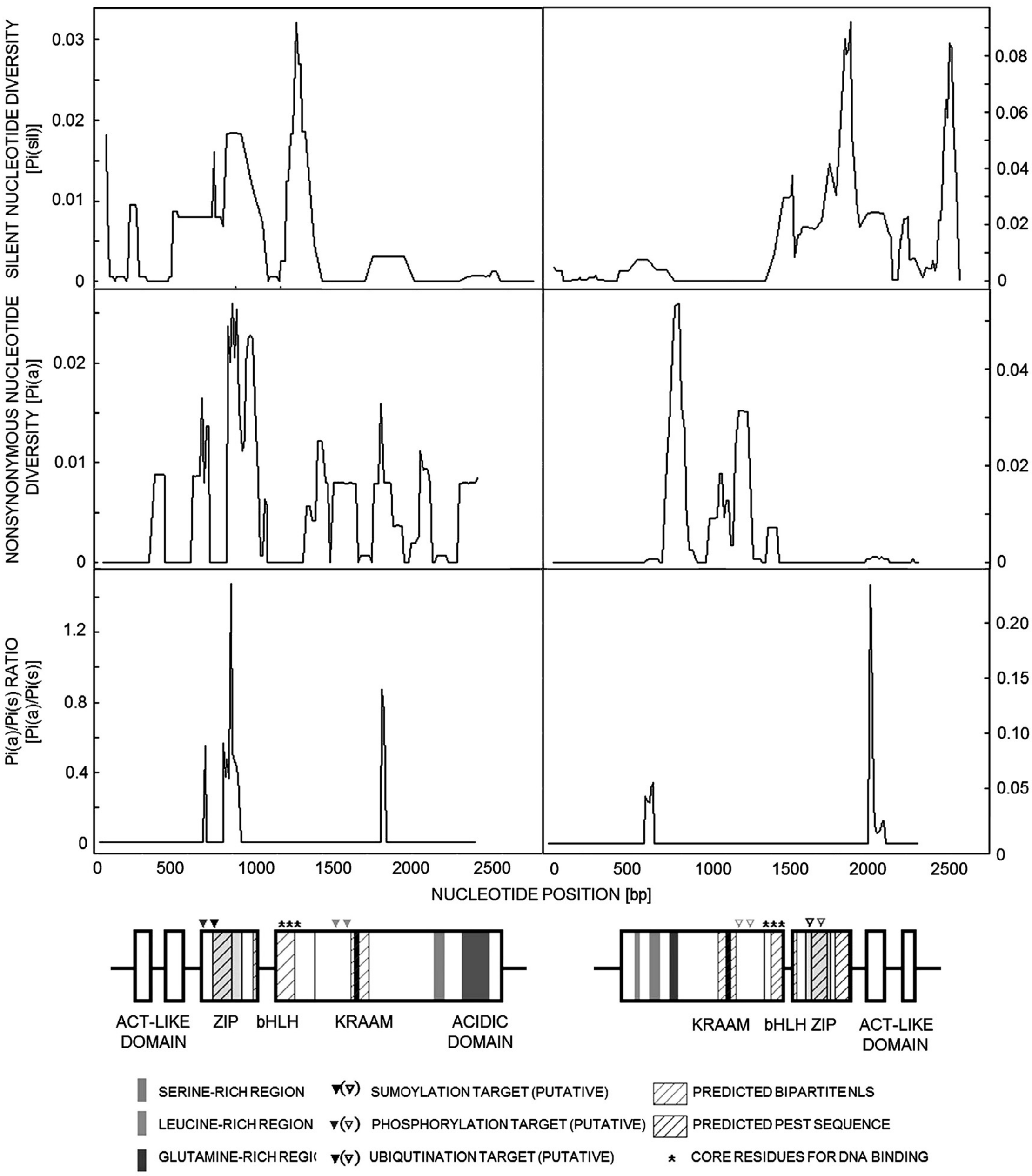

Fig. 1. A sliding-window analysis of *ICE1* (*left column*) and *ICE2* (*right column*) nucleotide diversity within *A. thaliana*. A window size is 50 bp, and a step size is 10 bp. The positions of exons are represented by boxes, introns by lines below the plot.

3'-regions of *ICE1* were characterized by almost equal levels of polymorphism, whereas the 5'-region was less diverse (Table 2 Suppl.).

Next, we assessed nucleotide diversity separately for the ecotypes from the southern, temperate, and northern ranges (Tables 2 Suppl. and 3 Suppl.). The southern and temperate accessions exhibited contrasting patterns of nucleotide polymorphism in *ICE2* compared to those from the northern regions. Nucleotide diversity values for the temperate ecotypes were 0.01075 and 0.01139 for the southern, and 0.00653 for northern ecotypes. Nucleotide polymorphism at nonsynonymous sites was 4.5-fold higher in the southern accessions *versus* the northern accessions and 1.6-fold higher than in the temperate ones. At synonymous sites, the levels of nucleotide polymorphism were also higher in the southern accessions though to a lesser extent (0.6-fold higher in the southern accessions *versus* the northern accessions and 0.8-fold higher than in the temperate ones). The northern accessions were also characterized by the lowest InDel polymorphism compared to all the other groups (Table 1). Four accessions with the *ICE2* sequence containing stop codons were found among the southern accessions (Table 1). These results indicate that the selection forces acting on the *ICE2* gene sequence were relaxed in the *A. thaliana* southern range, more stringent in the temperate, and especially strict in northern range. On the contrary, the populations from the northern, temperate, and southern regions had almost the same nucleotide polymorphism of the *ICE1* gene sequence, with the southern group being less diverse (Table 2 Suppl.). These levels, however, were about two times lower than those of *ICE2*, which indicates that selection acted stronger on the *ICE1* gene.

A sliding-window analysis was conducted to better examine the distribution of silent (*i.e.*, noncoding and synonymous) and nonsynonymous gene variations along the *ICE1* and *ICE2* gene regions (Fig. 1). In *ICE2*, the peaks of silent diversity were present in the 3'-flanking region, the first intron, and the ACT-like domain. Interestingly, the highest peak was observed in the second intron. Alternative polyadenylation sites were recently reported in the second intron for grape *ICE2*, and in the first intron of grape *ICE1* (Rahman *et al.* 2014). The search for alternative poly(A) sites in the *A. thaliana ICE* genes using *Poly(A) Signal Miner* (Liu *et al.* 2005) resulted in zero hits in intron sequences, thus, the observed peak could not be annotated to any specific site. The peaks of silent diversity in the *ICE1* gene sequence were observed in the 3'-flanking region, the sumoylation site and bHLH, and the NLS domains (Fig. 1). A nonsynonymous *ICE2* variation resided mainly in the first exon, where two peaks were observed: at the glutamine-rich domain and at the presumed phosphorylation site. The bHLH domain and exons two to four were almost not affected (Fig. 1). In *ICE1*, nonsynonymous variation peaks were spread across the whole sequence. The maximum peak was detected in the region corresponding to the sumoylation site (Fig. 1). Overall, the biggest exon was characterized by a considerably high diversity in *ICE2* and by a medium diversity in *ICE1*. The bHLH domain was characterized by a minimum nonsynonymous diversity and by a maximum silent diversity at the within-species level. Thus, it was constrained against amino acid replacements. This might be signature of an ancient history and a high importance of this domain.

At the within-species level, the ratio of nonsynonymous to synonymous diversity [Pi(a)/Pi(s)] for *ICE1* was 2.1, and 0.79 for *ICE2*. A sliding-window analysis of the populations across all the ranges studied shows that the region responsible for this high *ICE1* Pi(a)/Pi(s) ratio resided in the leucine zipper domain (ZIP) and predicted proline (P), glutamic acid (E), serine (S), and threonine (T)-rich sequence (PEST) domains (Fig. 1). The analysis of the Pi(a)/Pi(s) ratios in the three groups of accessions revealed elevated Pi(a)/Pi(s) ratios in the ZIP domain in all the ranges (data not shown). Next, we plotted sliding windows of *ICE2* Pi(a)/Pi(s) for groups of the ecotypes of the different ranges (data not shown). This ratio is significantly less than one for all the groups throughout the sequence, excluding the southern accessions, where a peak of Pi(a)/Pi(s) ratio was observed in the NLS domain. The needlessness of adaptation to cold treatment in the southern range allowed the accumulation of mutations in this gene domain essential for a transcription factor.

We examined the patterns of protein evolution between orthologs and paralogs of the *ICE1* and *ICE2* genes to determine if there were relaxed and/or selective divergences in the amino acid sequences. The ratio of K_a/K_s indicates protein evolution, where the values > 1, = 1, and < 1 indicate a positive, neutral, and purifying selection, respectively. As summarized in Table 2, the resulting K_a/K_s ratio for the *ICE2* orthologs (0.506) was an order of magnitude greater than for the *ICE1* orthologs (0.053), which suggests that the protein sequence of *ICE2* had evolved under a weaker selective constraint. This is consistent with a notion that *ICE2*, as a recent duplicate, was subject to rapid evolutionary changes (Kurbidaeva *et al.* 2014). Both the loci, however, had K_a/K_s values lower than 1 indicating that amino acid replacements between orthologs were functionally constrained.

We did not observe much difference in K_a/K_s ratios between pairs of paralogs. These ratios reflect that the average value across the entire coding sequence was less than one, indicating a purifying selection (Table 2). It is still possible that certain functional regions exhibit a molecular signature of a diversifying selection. To test this possibility, a sliding window analysis of K_a/K_s values was performed (Fig. 2). Between *ICE* paralogs, the K_a/K_s ratio exceeded one in the aspartate kinase, chorismate mutase, TyrA (ACT)-like domain, and in the NLS region, with the KRAAM motif. The highest K_a/K_s values were detected in the leucine-rich region which is specific only

Table 2. Sequence distances between orthologs and between paralogs. Ks - an average distance between groups at synonymous sites, Ka - an average distance between groups at nonsynonymous sites.

		K_s	K_a	K_a/ K_s
Orthologs	*A. thaliana ICE1* vs. *A. lyrata ICE1*	0.1829	0.0097	0.053
	A. thaliana ICE2 vs. *A. lyrata ICE2*	0.0888	0.0449	0.506
Paralogs	*A. thaliana ICE1* vs. *A. thaliana ICE2*	0.5697	0.1868	0.328
	A. lyrata ICE1 vs. *A. lyrata ICE2*	0.5997	0.1576	0.263

to the ICE2 protein and to the region where a 19-amino acid sequence common to ICE2 *Brassicaceae* proteins and a 24-amino acid sequence common to the ICE1 were previously found (Kurbidaeva *et al.* 2014). The ICE1-specific motif was neutrally charged (in total), whereas the total charge of the ICE2-motif was negative.

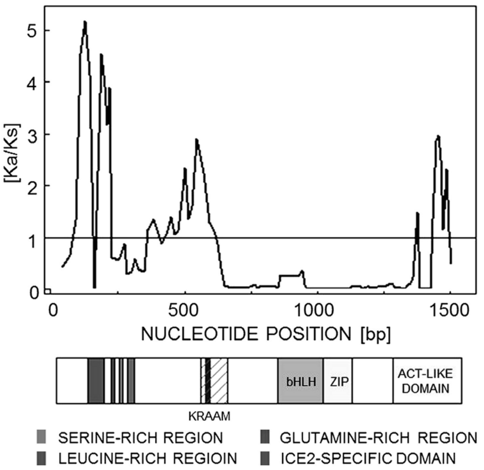

Fig. 2. A sliding-window analysis of Ka/Ks for coding regions of *A. thaliana* paralogous *ICE1* and *ICE2* genes. A window size is 50 bp, and a step size is 10 bp. The positions of protein domains are indicated below the plot.

Table 3. The application of the McDonald-Kreitman test to sequence polymorphism and divergence data. G_{mean} - mean sliding G statistic, G_{max} - maximum sliding G statistic, K_R - runs statistic, D_{KS} - Kolmogorov-Smirnov statistic.

Gene	G_{mean}	G_{max}	K_R	D_{KS}
ICE1	0.160	0.160	0.960	0.230
ICE2	0.177	0.118	0.063	0.076

The emergence of these specific domains might contribute to the functional diversification of the *ICE* genes.

Heterogeneity in nucleotide variation patterns across a DNA sequence can result from a non-neutral evolution (McDonald 1996). A statistical test for heterogeneity, based on the ratio of polymorphisms to fixed differences (between species) (McDonald 1996, 1998), was applied to the *ICE1* and *ICE2* regions using the software *DNA Slider* (McDonald 1998). This test analyses a sliding window of polymorphism levels relative to fixed differences between species across the gene of interest. Here we used an ortholog from *A. lyrata* for the interspecific divergence comparison. The neutral equilibrium model predicts a constant ratio of polymorphism to divergence across the gene. This test utilizes several different test statistics to access the heterogeneity. The test for *ICE1* and *ICE2* resulted in insignificant heterogeneity ($P > 0.05$, Table 3). We also performed a neutrality test by calculating Tajima's D values to evaluate the allele frequency spectrum and quantify the excess of rare alleles (rare alleles generate negative D values) and, therefore, get a more insight into the evolutionary history of the genes. Tajima's D was marginally positive in most regions of *ICE2* and *ICE1* (Tables 2 and 3 Suppl.). Negative Tajima’s D values were observed in the 5’ and 3’-flanking regions of both paralogs and in the noncoding region of *ICE1* indicating the presence of an excess of low-frequency polymorphism. Consistent with these tests, we could not detect a trace of positive or balancing selection. Next, we scanned different regions of the *ICE1* and *ICE2* genes using sliding windows of the Tajima's D test to identify regions which deviated from neutral expectations. Only in two regions, D statistic was significantly negative (data not shown): one corresponded to the 5’-flanking region, the other to the highly conservative bHLH domain.

Discussion

Our analyses provide further insights into the molecular evolution and nucleotide diversity of the cold response pathway regulators. To date, only the *CBF* genes evolution and diversity have been investigated (McKhann *et al.* 2008, Zhen and Ungerer 2008). We demonstrated that different selection forces had been acting on the different parts of the homologous *ICE* genes. Genetic diversity of *ICE2* was characterized by a higher number of haplotypes and a higher nucleotide diversity compared to *ICE1*. Moreover, we observed differentiation of the diversity at the *ICE2* locus but not at the *ICE1* across the ranges studied. As evidenced from the phylogenetic analysis, the genetic variability of the genes studied in *A. thaliana* was influenced more by selection pressures. An elevated polymorphism in the southern and temperate accessions suggests that the purifying selection might be relaxed in the regions with warmer climates. We presume that the relaxed selection at the *ICE2* locus have occurred in a recent evolutionary history of *A. thaliana*, as evidenced by a higher nonsynonymous polymorphism in the southern accessions compared to the northern accessions, but still lower than a synonymous polymorphism in the southern accessions (Table 3 Suppl.). A significantly lower rate of nonsynonymous than synonymous variation indicates that the variation was driven mainly by a purifying selection. The cold acclimation pathway is expensive for plants as it involves changes in gene expression and protein and other metabolite synthesis (Hannah *et al.* 2006). In the southern regions, where plants are unlikely to undergo a cold stress, mutations hindering the cold response pathway may not be lethal and even may be beneficial, as they can provide important systems of a plant with resources otherwise involved in a cold acclimation. *A. thaliana* is thought to originate in the Caucasus (Beck *et al.* 2008) and is native to Europe and Central Asia (Al-Shehbaz and O'Kane 2002, Koornneef *et al.* 2004). Hence, its current habitat which includes Mediterranean and subtropical regions is a result of later expansion. Our data shows, following an initial range expansion into warmer climates, that the relaxed purifying selection on the *ICE2* gene allowed multiple mutations to arise. These mutations have possibly resulted in an impaired freezing tolerance of the southern, and an elevated tolerance of northern populations, and hence, in a clinal variation in freezing tolerance reported repeatedly (Hannah *et al.* 2006, Zhen and Ungerer 2008, Zuther *et al.* 2012).

The nonsynonymous *ICE1* polymorphism was approximately two times higher than that of *ICE2* and than the synonymous polymorphism of *ICE1*. The nonsynonymous *ICE1* polymorphism was almost equal in the southern, temperate, and northern accessions. These data may imply that *ICE1* had a more ancient history than *ICE2*. Indeed, our recent study showed that *ICE2* originated 17.9 MYA from duplication of the presumed *ICE1* ancestor (Kurbidaeva *et al.* 2014). *ICE2* homologs are found only in *Brassicaceae* species. Thus, after the duplication event, the new *ICE2* gene underwent a range expansion into warmer climates together with the range expansion of the *Brassicaceae* family and was exposed to a weaker selection pressure than *ICE1*. A more recent history of *ICE2* could lead to a higher sequence diversification and to the observed patterns of polymorphism. Recent duplicates often undergo a process of rapid accumulation of mutations, which leads to sub- or neo-functionalization (He *et al.* 2005). Consistent with this notion, the K_a/K_s ratio for the *ICE2* orthologs in *A. thaliana* and *A. lyrata* is an order of magnitude greater than for the *ICE1* orthologs, which suggests that the protein sequence of *ICE2* evolved under a weaker selective constraint in these two species. Although at the between-species level, the protein evolution of both genes is functionally constrained as evidenced by K_a/K_s ratios lower than the neutral expectation of one in some regions of these paralogs, the K_a/K_s ratio indicates that the sequence diversification and emergence of two new ICE2-specific motifs could contribute to the functional diversification of the *ICE* genes. The clinal variation observed for the *ICE2* gene polymorphism confirms the previously suggested role of the gene in a cold response regulation (Fursova *et al.* 2009, Kurbidaeva *et al.* 2014). The absence of a clinal variation in the *ICE1* polymorphism is surprising since this gene was also shown to participate in a cold tolerance (Chinnusamy *et al.* 2003). A higher selection pressure acting on *ICE1* can be caused by its role in more essential processes, such as stomata development (Kanaoka *et al.* 2008) or growth cessation through the *CBF* regulon genes (Achard *et al.* 2008).

Several neutrality tests that we applied show that both the *ICE* genes have been evolving neutrally since the duplication event. Elevated Pi(a)/Pi(s) ratios may indicate that a positive selection have been acting at the within-species level on the *ICE1* gene sequence. However, the overall low polymorphism suggests, that the observed excess of nonsynonymous variation might be random and insignificant. In *ICE2*, only in the southern group of ecotypes, we detected elevated Pi(a)/Pi(s) ratios in the NLS domain. A nonsynonymous variation is located almost exclusively to the first exon in *ICE2*, whereas it is spread across the whole *ICE1* gene. The most diverse at within- and between-species levels is the first *ICE2* exon, and the 5'-flanking region and the bHLH domain appear to be the most conservative parts. It could be the evidence of a purifying selection acting on these essential gene regions, as evidenced by the Tajima's D statistic. The 5'-flanking region may be important for a gene expression regulation, and the bHLH domain is necessary for DNA-binding (Chinnusamy *et al.* 2003).

This study adds new data about the microevolution of the duplicate genes in the light of adaptive variations to climate and emphasizes the potential role of regulatory genes in underlying adaptive phenotypic variation in nature.

References

Aguadé, M.: Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. - Mol. Biol. Evol. **18**: 1-9, 2001.

Al-Shehbaz, I.A., O'Kane, S.L.: Taxonomy and phylogeny of *Arabidopsis* (*Brassicaceae*). - In: Somerville, C.R., Meyerowitz, E.M. (ed.): The *Arabidopsis* Book [Internet; doi: 10.1199/tab.0001]. Amer. Soc. Plant Biol., Rockville 2002.

Achard, P., Gong, F., Cheminant, S., Alioua, M., Hedden, P., Genschik, P.: The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins *via* its effect on gibberellin metabolism. - Plant Cell. **20**: 2117-2129, 2008.

Ba, A.N.N., Pogoutse, A., Provart, N., Moses, A.M.: NLStradamus: a simple hidden Markov model for nuclear localization signal prediction. - BMC Bioinformatics **10**: 202, 2009.

Beck, J.B., Schmuths, H., Schaal, B.A.: Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. - Mol. Ecol. **17**: 902-915, 2008.

Blanc, G., Hokamp, K., Wolfe, K.H.: A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. - Genome Res. **13**: 137-144, 2003.

Bowers, J. E., Chapman, B. A., Rong, J., Paterson, A. H.: Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. - Nature **422**:

433-438, 2003.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.H., Hong, X., Agarwal, M., Zhu, J.K.: *ICE1*: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. - Genes Dev. **17**: 1043-1054, 2003.

Dellaporta, S.L., Wood, J., Hicks, J.B.: A plant DNA minipreparation: version II. - Plant mol. Biol. Rep. **1**: 19-21, 1983.

Ermolaeva, M.D., Wu, M., Eisen, J.A., Salzberg, S.L.: The age of the *Arabidopsis thaliana* genome duplication. - Plant mol. Biol. **51**: 859-866, 2003.

Fursova, O.V., Pogorelko, G.V., Tarasov, V.A.: Identification of *ICE2*, a gene involved in cold acclimation which determines freezing tolerance in *Arabidopsis thaliana*. - Gene **429**: 98-103, 2009.

Hall, T.A. *BioEdit*: a user-friendly biological sequence alignment editor and analysis program for *Windows 95/98/NT*. - Nucl. Acids Symp. Ser. **41**: 95-98, 1999.

Hannah, M.A., Wiese, D., Freund, S., Fiehn, O., Heyer, A.G., Hincha, D.K.: Natural genetic variation of freezing tolerance in *Arabidopsis*. - Plant Physiol. **142**: 98-112, 2006.

He, X., Zhang, J.: Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. - Genetics **169**: 1157-1164, 2005.

Kanaoka, M.M., Pillitteri, L.J., Fujii, H., Yoshida, Y., Bogenschutz, N.L., Takabayashi, J., Zhu, J-K., Torii, K.U.: SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. - Plant Cell **20**: 1775-1785, 2008.

Koornneef, M., Alonso-Blanco, C., Vreugdenhil, D.: Naturally occurring genetic variation in *Arabidopsis thaliana*. - Annu. Rev. Plant Biol. **55**: 141-172, 2004.

Kurbidaeva, A., Ezhova, T., Novokreshchenova, M.: *Arabidopsis thaliana ICE2* gene: phylogeny, structural evolution and functional diversification from *ICE1*. - Plant Sci. **229**: 10-22, 2014.

Le Corre, V., Roux, F., Reboud, X.: DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. - Mol. Biol. Evol. **19**: 1261-1271, 2002.

Liu, H., Han, H., Li, J., Wong, L.: *DNAFSMiner*: a web-based software tool box to recognize two types of functional sites in DNA sequences. - Bioinformatics **21**: 671-673, 2005.

McDonald, J.H.: Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. - Mol. Biol. Evol. **13**: 253-260, 1996.

McDonald, J.H.: Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. - Mol. Biol. Evol. **15**: 377-384, 1998.

McKhann, H.I., Gery, C, Bérard, A, Lévêque, S, Zuther, E, Hincha, D.K., De Mita, S., Brunel, D., Téoulé, E.: Natural variation in *CBF* gene sequence, gene expression and freezing tolerance in the Versailles core collection of *Arabidopsis thaliana*. - BMC Plant Biol. **8**: 105, 2008.

Medina, J., Catalá, R., Salinas, J.: The CBFs: three arabidopsis transcription factors to cold acclimate. - Plant Sci. **180**: 3-11, 2011.

Mitchell-Olds, T., Schmitt, J.: Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. -. Nature **441**: 947-952, 2006.

Nordborg, M., Hu, T.T., Ishino, Y., Jinal, J., Christopher, T., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., Bergelson, J.: The pattern of polymorphism in *Arabidopsis thaliana*. - Plos Biol. **3**: 1289-1299, 2005.

Okonechnikov, K., Golosova, O., Fursov, M.: *Unipro UGENE*: a unified bioinformatics toolkit. - Bioinformatics **28**: 1166-1167, 2012.

Rahman, M.A., Moody, M.A., Nassuth, A.: Grape contains 4 *ICE* genes whose expression includes alternative polyadenylation, leading to transcripts encoding at least 7 different ICE proteins. - Environ. exp. Bot. **106**: 70-78. 2014.

Rogers, S.W., Wells, R., Rechsteiner M.: Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. - Science **234**: 364-368, 1986.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R.: DnaSP, DNA polymorphism analyses by the coalescent and other methods. - Bioinformatics **19**: 2496-2497, 2003.

Saitou, N, Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. - Mol. Biol. Evol. **4**: 406-425, 1987.

Schmid, K.J., Ramos-Onsins, S., Ringys-Becktein, H., Weisshaar, B., Mitchell-Olds, T.: A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. - Genetics **169**: 1601-1615, 2005.

Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D., Schmitt, J.: A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. - Proc. nat. Acad. Sci. USA **101**: 4712-4717, 2004.

Tajima, F.: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. - Genetics **123**: 585-595, 1989.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S.: *MEGA5*: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum Parsimony methods. - Mol. Biol. Evol. **28**: 2731-2739, 2011.

Thomashow, M.F.: Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. – Annu. Rev. Plant. Physiol. Plant. mol. Biol. **50**: 571-599, 1999.

Thompson, J.D., Higgins, D.G., Gibson, T.J.: *CLUSTAL W*: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. - Nucl. Acids Res. **22**: 4673-4680. 1994.

Yamaguchi-Shinozaki, K., Shinozaki, K.: A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. - Plant Cell **6**: 251-264, 1994.

Yoshida, K., Kamiya, T., Kawabe, A., Miyashita, N.T.: DNA polymorphism at the *ACAULIS5* locus of the wild plant *Arabidopsis thaliana*. - Genes. Genet. Syst. **78**: 11-21, 2003.

Zhen, Y., Ungerer, M.C.: Clinal variation in freezing tolerance among natural accessions of *Arabidopsis thaliana*. - New Phytol. **177**: 419-427, 2008.

Zuther, E., Schulz, E., Childs, L.H., Hincha, D.K.: Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of *Arabidopsis thaliana* accessions. - Plant Cell Environ. **35**: 1860-1878, 2012.