Theor Appl Genet (2012) 124:809–819
DOI 10.1007/s00122-011-1735-0

ORIGINAL PAPER

Intron-length polymorphism identifies a Y_2K_4 dehydrin variant linked to superior freezing tolerance in alfalfa

Yves Castonguay · Marie-Pier Dubé · Jean Cloutier · Réal Michaud · Annick Bertrand · Serge Laberge

Received: 30 June 2011 / Accepted: 14 October 2011 / Published online: 9 November 2011
© Her Majesty the Queen in Rights of Canada, as represented by the Minister of Agriculture and Agri-Food Canada 2011

Abstract Breeding alfalfa (*Medicago sativa* L.) with superior freezing tolerance could be accelerated by the identification of molecular markers associated to that trait. Dehydrins are a group of highly hydrophilic proteins that have been related to low temperature tolerance. We previously identified a dehydrin restriction fragment length polymorphism (RFLP) among populations recurrently selected for superior tolerance to freezing (TF). Analysis of crosses between genotypes with (D+) or without (D−) that RFLP revealed a significant impact on freezing tolerance. In this study, we sought to develop a PCR marker for freezing tolerance based on prior evidence of a relationship between size variation in Y_2K_4 dehydrins and the RFLP. Results confirm the enrichment of Y_2K_4 sequences of intermediate size (G2 group) in response to recurrent selection and in the D+ progeny. Analysis of genomic sequences revealed significant intron-length polymorphism (ILP) within the G2 group. G2 sequences with a characteristic short intron were more frequently found in D+ genotypes. Amplification using sequence-characterized amplified region (SCAR) primers bordering the intron confirmed an increase in the number of fragments with small introns in the D+ progeny and in the ATF5 population obtained after five cycles of recurrent selection for superior TF within the cultivar Apica (ATF0). Conversely, there was a reduction in the number of fragments with long introns in the D+ progeny and in ATF5 as compared to ATF0. Recurrent selection for superior tolerance to freezing in combination with ILP identified a sequence variant of Y_2K_4 dehydrins associated to the phenotypic response to selection.

Communicated by T. Close.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1735-0) contains supplementary material, which is available to authorized users.

Y. Castonguay (✉) · M.-P. Dubé · J. Cloutier · R. Michaud · A. Bertrand · S. Laberge
Crops and Soils Research and Development Center, Agriculture and Agri-Food Canada, 2560 Hochelaga Blvd, Quebec, QC G1V 2J3, Canada
e-mail: yves.castonguay@agr.gc.ca

Introduction

The development of marker-assisted selection (MAS) approaches for the improvement of complex traits in open-pollinated populations of polyploids is a daunting task. The challenge is to sort out and to identify functional genetic markers among a myriad of allelic possibilities (Ponting et al. 2007). Sequence-based allele mining of candidate genes among plant siblings with contrasted phenotype would give tremendous impetus to precision breeding in crop plants (Kumar et al. 2010). In that respect, analysis of populations of autotetraploid ($2n = 4x = 32$) alfalfa recurrently selected for superior tolerance to freezing (TF) (Castonguay et al. 2009) could facilitate the identification of sequences associated with superior winter survival.

Evidence from genetic and physiological studies support a close relationship between the accumulation of dehydrins and low temperature tolerance in plants (Brini et al. 2007; Houde et al. 2004; Ismail et al. 1999; Marian et al. 2004; Patton et al. 2007; Yin et al. 2006). Dehydrin refers to a large family of plant proteins induced under conditions leading to extensive cellular desiccation including drought, salinity and extracellular freezing (Beck et al. 2007). These proteins are structurally characterized by conserved domains known as the Y-segment (V/TDEYGNP) located at the N-terminus, the S-segment (stretches of Ser) and the

Lys-rich K-segment (EKKGIMDKIKEKLPG) that are separated by Φ segments that contain repeated stretches of Gly and polar amino acids (Close 1996). Dehydrins are categorized into five sub-classes identified as Y_nSK_n, SK_n, K_n, Y_nK_n and K_nS based on the combination and number of the Y-, S- and K-segments (Rorat 2006).

Using bulk segregant analysis (BSA) of alfalfa populations recurrently selected for superior tolerance to freezing, Rémus-Borel et al. (2010) uncovered a Y_2K_4 dehydrin restriction fragment length polymorphism (RFLP) that intensifies in response to selection. Analysis of progenies from crosses between genotypes with (D+) or without (D−) that dehydrin RFLP showed a significant impact of that polymorphism on freezing tolerance. In that same study, amplification of the K-segment region of the Y_2K_4 dehydrin revealed fragment size variations that were assigned to three major groups of amplicons based on the number of nucleotides: the G1, G2 and G3 groups. A more intense amplification of the G2 group was observed in pools of genotypes with the dehydrin RFLP and in recurrent selection with a superior freezing tolerance phenotype.

In the current study, we sought to determine whether Y_2K_4 dehydrins of intermediate size (G2 group) can be used as markers for the presence of the dehydrin RFLP associated with superior tolerance to freezing. Our objective was to develop a robust marker to screen broad-base populations and facilitate the introgression of cold tolerance in alfalfa germplasms from diverse origins.

Materials and methods

Plant materials and DNA extraction

Genotypes of alfalfa (*Medicago sativa* spp. *sativa*), cultivar Apica (ATF0) and populations (ATF2, ATF4, ATF5 and ATF6) derived from this cultivar through, respectively, 2, 4, 5, and 6 cycles of recurrent selection for superior tolerance to freezing (Castonguay et al. 2009) were grown as described in Rémus-Borel et al. (2010). Progenies from crosses obtained by inter crossing ten ATF0 genotypes with (D+ cross) and ten ATF0 genotypes without (D− cross) a polymorphic variant of the cold-inducible *msa*CIG gene (EU652338) encoding a Y_2K_4 dehydrin were grown under the same environmentally controlled conditions.

Genomic DNA was extracted from 1 g of leaves of individual and pooled genotypes (~45) from each of the cultivar Apica (ATF0), populations ATF2, ATF4, ATF5 and ATF6 recurrently selected for superior tolerance to freezing and individual and pooled genotypes (~20) of the D− and D+ crosses. Total genomic DNA was extracted using the CTAB procedure of Rogers and Bendich (1988) and subsequently used for PCR amplification.

Primer design, PCR amplification and gel electrophoresis of amplified products

The cDNA sequence of the cold-inducible *msa*CIG gene (GenBank accession no. EU652338) was used to design primers (Fig. 1; Table 1) with the Oligo Explorer software, version 1.1.0 (T. Kuulasma, University of Kuopio, Kuopio, Finland). PCR reactions were performed in a total volume of 25 µl in 0.2 ml PCR strips containing 2.5 µl of 10× PCR buffer, 1 µl each of 5 µM primers, 0.5 µl of 10 mM dNTP (Roche Diagnostics, Laval, QC, Canada), 0.5 µl of 5 Prime *Taq* polymerase 5U/µl (Inter Medico, Markham, ON, Canada) and 5 µl of 10 ng/µl genomic DNA (50 ng). The conditions for PCR were as follows: an initial denaturing step was performed at 94°C for 3 min followed by 35 cycles at 94°C for 1 min, a 1 min annealing at the temperature indicated in Table 1 for each primer pair, an extension at 72°C for 1 min and a final extension at 72°C for 7 min. Annealing temperature of the 164c-299r primer pair was raised to 70°C to increase specificity. All the reactions were performed on an Eppendorf Mastercycler ep System (Eppendorf Canada, Mississauga, ON, Canada). Upon completing the PCR cycles, 20 µl of each reaction were run for 2–3 h at 70 V on a 2% agarose gel stained with ethidium bromide. DNA fragments were visualized using a UVP BioDoc-It system (UVP, Upland, CA).

Cloning and sequencing of amplified fragments

DNA fragments were recovered from agarose gels using the QIAquick gel extraction kit (QIAGEN Inc., Mississauga, ON, Canada) according to the manufacturer's recommendations. Purified DNA was cloned into the pGEM®-T Easy Vector (Promega, Madison, WI) according to the manufacturer's instructions. Positive transformants were recovered, grown in liquid medium and plasmids purified using the QIAprep Spin Miniprep Kit (QIAGEN Inc., Mississauga, ON, Canada). Plasmid preparations were sent for sequencing using M13 forward and reverse primers for bi-directional sequencing.

Bioinformatic analyses

BLASTn and BLASTp sequence identity searches were performed at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). Chromatograms from the sequencing of the different fragments were edited and assembled using BioEdit software version 7.0.9.0 (Hall 1999). Amino acid sequences were aligned using the CLUSTALW tool included in BioEdit and refined manually. VecScreen, a online contaminant vector screening software developed by NCBI (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) was used to

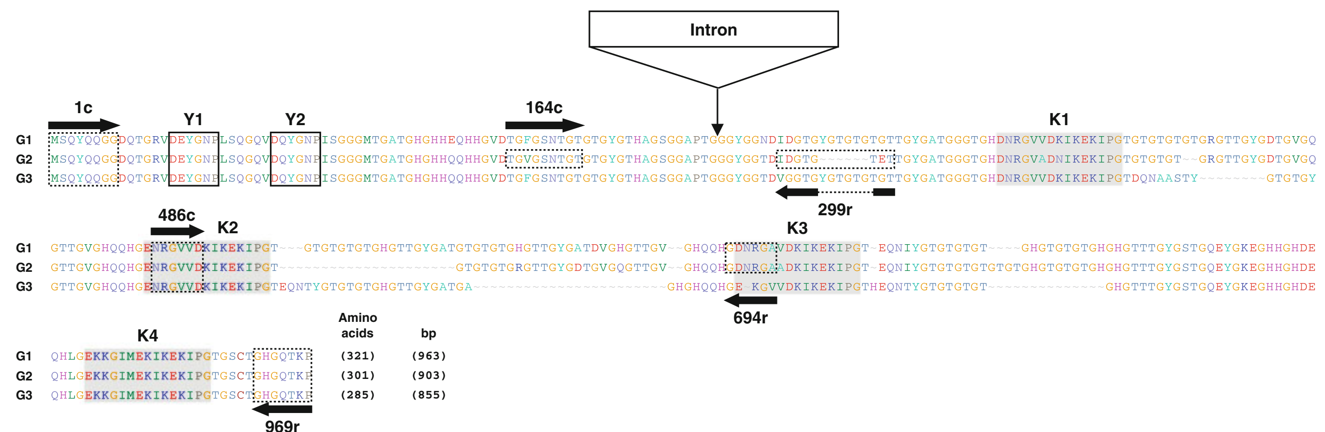


Fig. 1 Alignment of representative full-length coding sequences for the G1, G2 and G3 groups of size variants of Y_2K_4 dehydrins previously defined by Rémus-Borel et al. (2010) from partial sequences. These sequences are deposited in BenBank under the accession numbers JN226736 to JN226738. The location of a single intron is indicated. Conserved Y motifs are identified in *boxes with solid lines* while the K-segments are shown in *shaded areas*. Primers used for the amplification of internal regions of dehydrins are indicated. Nucleotide variations at annealing sites between the three groups of sequences are highlighted by *boxes with dotted lines*. The number of amino acids and nucleotides (bp) for the open reading frame is indicated at the end of the sequences

Table 1 PCR primers used for cloning, characterizing and sequencing Y_2K_4 dehydrin variants from *Medicago sativa* cv. *Apica* with their corresponding Tm

Primers	Description	Sequences (5′→3′)	Tm (°C)
1c	5′ forward primer for full length msaCIG	ATGTCTCAATATCAACAAGGTGG	58.0
969r	3′ reverse end for full length msaCIG	TGGTTTAGTTTGTCCATGTCC	57.0
486c	Forward primer for discrimination of msaCIG group sizes	AAATAGAGGAGTTGTGGACAA	54.0
694r	Reverse primer for discrimination of msaCIG group sizes	CAGCTCCTCTATTATCTCC	49.0
164c	Forward primer for intron amplification	CAGGTGTTGGGAGCAACACAGGTA	67.0
299r	Reverse primer for intron amplification	GTTTCGGTTCCGGTTCCATCAA	67.0
M13F	Universal M13 forward primer for sequencing in pGEM T	TGTAAAACGACGGCCAGT	79.0
M13R	Universal M13 reverse primer for sequencing in pGEM T	CAGGAAACAGCTATGAC	69.4

The primers are named based on their position on msaCIG open reading frame (GenBank accession no. EU652338)

remove vector sequence. Representative full-length sequences of Y_2K_4 dehydrins presented in this article have been deposited in the GenBank data library under the accession numbers JN226736 to JN226743.

Results

PCR discrimination of Y_2K_4 dehydrins size variants

Based on sequencing a portion of Y_2K_4 dehydrins in alfalfa, Rémus-Borel et al. (2010) previously identified three size groups identified as G1, G2 and G3 that were differentiated by deletions between the K-segments. In the current study, analysis of full-length coding sequences obtained from a pool of ATF5 genotypes was performed to determine whether additional regions of Y_2K_4 dehydrins could be used to differentiate G2 from the other groups of size variants. Results confirmed the presence of three broadly defined group sizes that vary between ~960, ~905 and ~860 bp for the G1, G2 and G3 groups, respectively. Representative genomic sequences from each group are illustrated in Fig. 1 and were deposited in GenBank under the accession numbers JN226736 to JN226738. These genomic sequences reveal that alfalfa Y_2K_4 dehydrins are highly homologous and are characterized by the presence of two exons interrupted by a single intron located between the Y- and K-segments near the 5′ end of the sequence. They also show that the most significant variations between the G1, G2 and G3 sequences occur in the K-segments region.

A set of primers (486c-694r; Table 1) was designed to discriminate between the three groups of Y_2K_4 dehydrins variants. These primers allow the detection of fragments of ~200 and ~150 bp for the G1 and G2 groups, respectively. The presence of a deletion at the putative annealing site of the 694r primer precluded amplification of the G3 group (Fig. 1). These primers were then used to assess variations in the amplification of the G1 and G2

groups in populations recurrently selected for superior tolerance to freezing (Fig. 2a). Size variation was noticeable within the G1 group whereas the length of the G2 fragments was more uniform. An intensification of the G2 signal with increasing number of selection cycles supported prior evidence by Rémus-Borel et al. (2010) that the G2 group of dehydrins is enriched in response to recurrent selection for superior freezing tolerance.

Enhancement in the intensity of the G2 group was also noted in progenies issued from crosses between genotypes

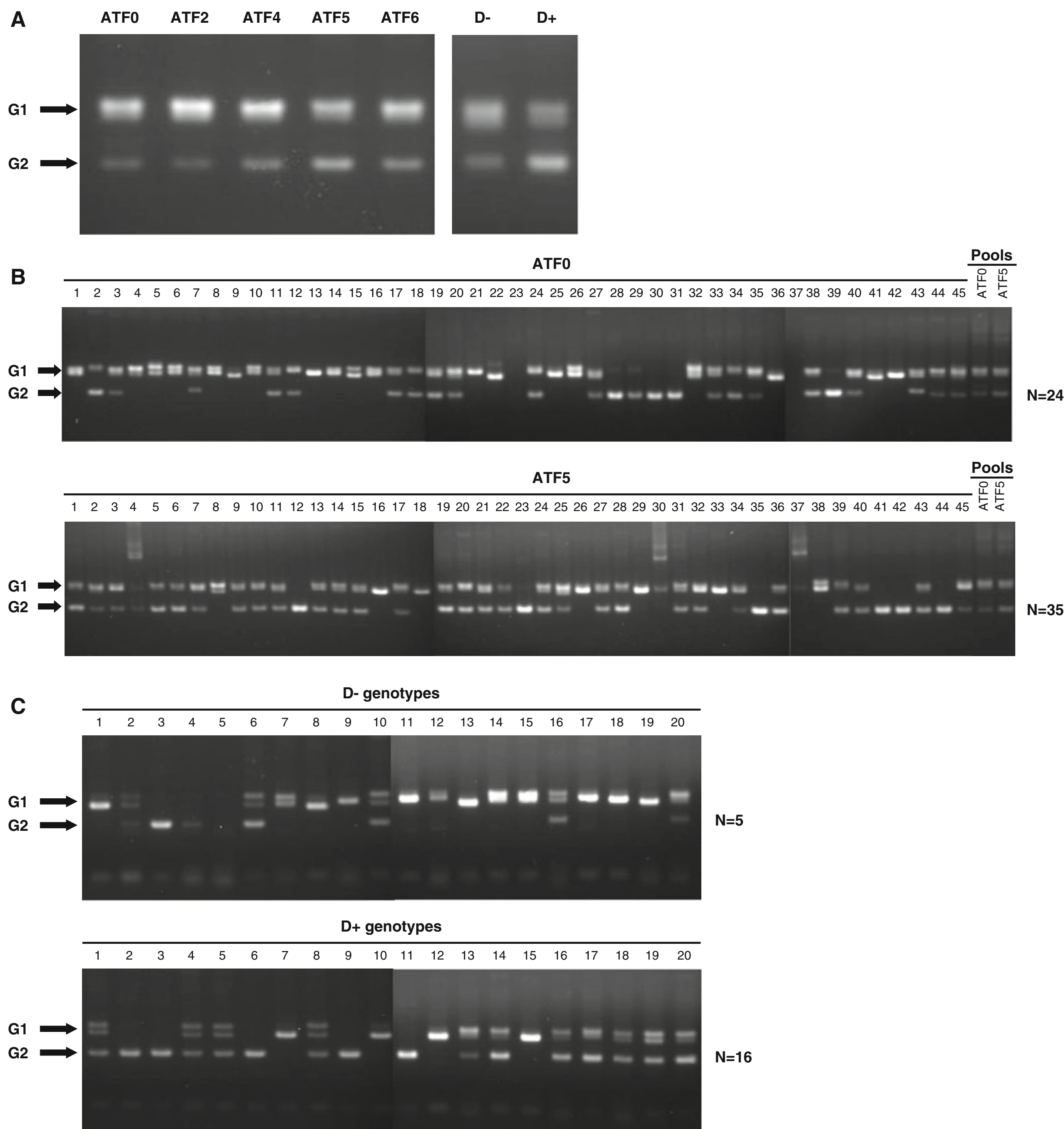


Fig. 2 Amplification of the Y_2K_4 group size variants using the 486c-694r primer pair with: **a** pooled DNA samples from populations recurrently selected for improved tolerance to freezing (ATF0 to ATF6) and pooled samples from crosses of ATF0 genotypes with (D+) or without (D−) a dehydrin polymorphism; **b** 45 genotypes within each of the ATF0 and ATF5 populations; **c** 20 genotypes derived from the D+ and D− crosses. *Arrows* indicate fragments derived from the G1(∼200 bp) and G2 (∼150 bp) groups. *N* indicates the number of genotypes with a positive amplification for the G2-derived fragment

with a dehydrin RFLP linked to the improvement of freezing tolerance (D+) as compared to progenies obtained by crossing genotypes without the RFLP (D−). Subsequent screening of 45 genotypes within each of the ATF0 and the ATF5 populations with the 486c-694r primers revealed an increase in the number of genotypes with G2 amplicons from 24 in ATF0 to 35 in ATF5 (Fig. 2b). Genotypic assessment also showed an enrichment of G2 amplicons from 5 in the D− cross to 16 in the D+ cross within a set of 20 genotypes randomly sampled within each progeny (Fig. 2c).

The similarity of the Δ of percent increase of G2 amplicons and of the dehydrin RFLP between the ATF0 and ATF5 populations suggests a relationship between these two observations (Table 2). Although PCR amplifications support the enrichment of the G2 group in response to selection, the high frequency of their occurrence in ATF0 (53%) as compared to the RFLP (15%) indicates that only a subset of variants is associated to the dehydrin RFLP. Eighty percent of D+ progenies yielded positive G2 amplifications confirming an enrichment of these sequences as compared to D− progenies and population ATF0. However, the observation that twenty-five percent of the D− progenies give a positive amplification further indicates that G2 sequences unrelated to the RFLP polymorphism are present within the Apica (ATF0) genetic background.

Sequence analysis of the G2 group of Y_2K_4 dehydrins

In order to determine whether or not a subset of the G2 group of Y_2K_4 dehydrins may be associated with the dehydrin RFLP, we amplified full-length sequences from ATF5 genotypes either positive (D+) or negative (D−) for

Table 2 Frequency of PCR amplifications of G2 amplicons in initial genetic background (ATFO) and population ATF5 obtained after five cycles of recurrent selection for superior tolerance to freezing

Populations	PCR amplifications		RFLP polymorphism[a]	
	Frequency (%)	Δ (%)	Frequency (%)	Δ (%)
ATFO	53	25	15	23
ATF5	78		38	
D−	25			
D+	80			

Frequency of G2 amplicons was also assessed in progenies from populations D+ and D− obtained by inter crossing ATFO genotypes with (+) or without (−) a dehydrin RFLP previously identified by Rémus-Borel et al. (2010). Comparative frequency of this RFLP in populations ATFO and ATF5 is also provided. Difference of frequency (Δ) is indicated

[a] Source: Rémus-Borel et al. (2010)

the dehydrin RFLP (4 in each group) using the 1c-969r primer pair (Fig. 1; Table 1). The D+ genotypes 1, 6, 22 and 41 and the D− genotypes 9, 10, 11 and 40 were selected within a group of 45 ATF5 genotypes previously scored for the dehydrin RFLP by Rémus-Borel et al. (2010). These numbers correspond to ATF5 genotypes used in amplifications presented in Figs. 4a, 5 of this study.

Fragments were size-selected and validated as being G2 sequences using the 486c-694r primer pair. Sequence analysis of cloned fragments revealed that the length of the open-reading frame of the G2 sequences varied from 867 to 921 bp (Fig. 3a). The shortest sequences resulted from a deletion before the intron while their K-segments region was typical of the G2 group including the presence of the annealing site for primer 694r (Fig. 1). The most noticeable variations among the exons of G2 sequences were found to occur in the sequences bordering the intron. Substitution of a valine to an alanine within the K3 segment and the occurrence of four amino acid substitutions before the K3 segment are also noticeable features of the I_2 group of sequences in Fig. 3a.

For ease of comparison, the introns were translated into notional peptides. Based on variations in the introns and their flanking regions, sequences were divided into five distinct groups identified as I_1–I_5 (Fig. 3a and supplementary Fig. 1). Intron polymorphism arose mainly from length variation ranging from 189 up to 447 bp (Table 3). A representative sequence from each group was deposited in GenBank under accession numbers JN226739 to JN226743. Single nucleotide polymorphisms accounted for the other variations. The presence of multiple allelic forms for each intron variant in alfalfa genotypes indicates the presence of paralogs within the G2 group of Y_2K_4 dehydrins (Table 3). Sequences A to D within the intron variants group I_1, share a highly homologous short intron that was only found in full-length sequences of D+ genotypes. Intron variant I_1-A occurred at a much higher frequency than the other three variants within that group and was the only one observed in the four D+ genotypes. Intron sequences were more variable in the D− genotypes particularly with regard to their length (Table 3). Also, no single variant was shared by all D− genotypes. Intron variant I_2 with noticeable exonic variations in the K-segments region was the only other intron variant cloned from a D+ genotype.

Analysis of exonic variations within subgroup I_1-A (Fig. 3b) revealed the presence of 9 allelic forms among the 16 sequences available from the 4 D+ genotypes (Table 3). Allelic variation was associated with minor changes resulting from amino acid substitutions either found outside or within the K-segments. Peptide sequence variant no. 1 in Fig. 3b was present in all D+ genotypes.

A

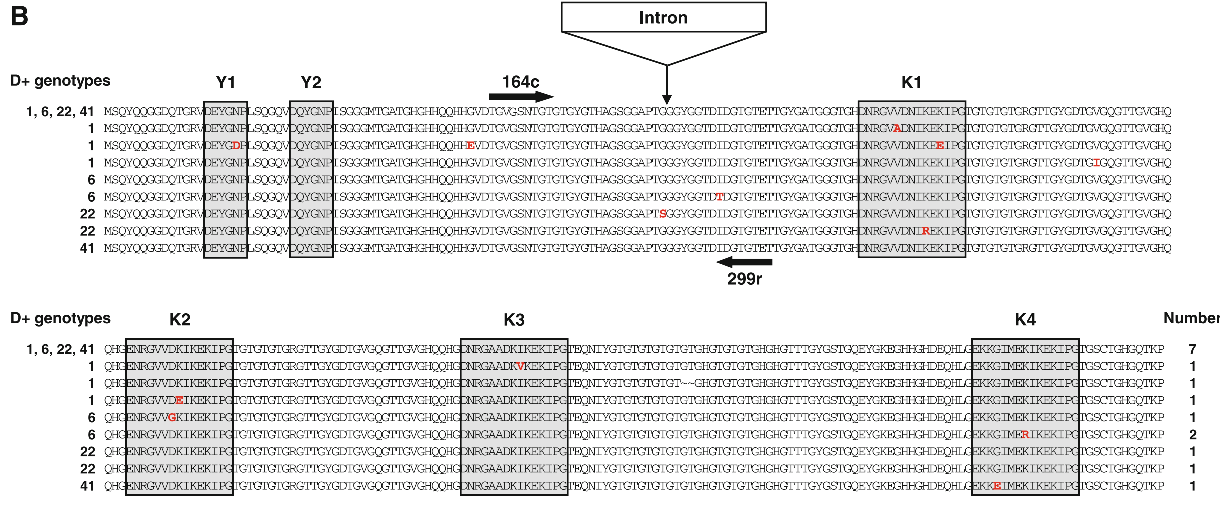
Y1 Y2 164c Intron

I1 A B C D
I2 A B C D E
I3 A B C
I4 A B C D
I5 A B C D

Intron K1 K2

299r

K2 K3 K4 Amino acids bp



◀ **Fig. 3** **a** Full-length alignment of exonic and intronic sequences of the G2 group of Y_2K_4 dehydrins amplified with the 1c and 969r primer pair. The Y- and K-segments and the intron are indicated in *shadowed boxes*. Introns were translated into notional peptides for ease of comparison. Five groups (I_1–I_5) were defined by the presence of InDels and sequence variations in the intron and their flanking regions. Variations in amino acids in the exons are indicated in *bold letters*. The orientation and anchoring sites of primers 164c and 299r are indicated with *black arrows*. The number of amino acids and nucleotides (bp) for the open reading frame is indicated in *brackets* at the end of each sequence. A representative sequence from each group was deposited in GenBank under accession numbers JN226739 to JN226743. **b** Full-length alignment of coding sequences of variant A of the intron group I_1. The Y- and K-segments are indicated in *shadowed boxes* and the location of the intron is shown. The orientation and anchoring sites of primers 164c and 299r are indicated with *black arrows*. Variations in amino acids are indicated in *bold letters*. Identification of D+ genotypes in which sequences were found is provided in the *left margin*. The number of observed sequences in D+ genotypes is indicated at the end of each peptide sequence

Table 3 Number of intronic variant sequences in positive (D+) or negative (D−) genotypes

Intron variant		Frequency in		Length (bp)	Genotypes
		D+	D−		
I_1	A	16		217	1^+, 6^+, 22^+, 41^+
	B	1		217	6^+
	C	1		217	6^+
	D	1		217	41^+
I_2	A	1		215	41^+
	B		1	215	40^-
	C		5	215	40^-
	D		1	215	40^-
	E		1	189	9^-
I_3	A		1	223	10^-
	B		2	223	10^-, 11^-
	C		1	224	9^-
I_4	A		3	446	9^-
	B		1	445	9^-
	C		1	444	9^-
	D		1	445	9^-
I_5	A		1	446	11^-
	B		1	447	11^-
	C		1	446	11^-

The length of the intron and the presence in genotypes positive (+) or negative (−) for the dehydrin RFLP is indicated

Development of SCAR markers

SCAR primers (164c-299r) were developed based on the sequences flanking the intron (Fig. 3a; supplementary Fig. 1). Differential amplification resulting from polymorphisms at the anchoring sites and variations in the length of the introns allowed us to differentiate G2 intron variants. Intron group I_2 and I_3 with short introns and group I_4 with a long intron gave no amplification products due to the presence of InDels or single nucleotide changes at primers annealing sites. Therefore a single group of sequences with a short intron (I_1) and a single group with a long intron (I_5) were amplified. A nucleotide mutation within the 164c annealing site of the I_1-C sequence also prevented its amplification. The presence of InDels at the annealing sites of primers also precluded the amplification of G1 and G3 dehydrins (Fig. 1).

Amplification with the 164c-299r primer pair clearly distinguished the small and long introns (Fig. 4a). The presence of a short fragment that was more strongly amplified in D+ than in D− genotypes, was particularly striking. Conversely, a weaker amplification signal for this short fragment and the presence of additional bands of higher and slightly lower molecular weights were noticed in D− genotypes. The reduction in the number of sequences with a long intron with the concomitant intensification of sequences with small introns was equally observed between pools of genotypes from the D− and D+ crosses and between pools of genotypes from population ATF0 and population ATF5 obtained after five cycles of selection for superior tolerance to freezing (Fig. 5, pools). We also observed that the strong amplification of a short fragment co-segregates with the polymorphic dehydrin initially detected by RFLP analysis (Figs. 4b, 5). Confirmation that this short fragment correspond to intron variant I_1 was achieved by sequence analysis from four D+ genotypes (ATF5 genotypes 4, 7, 25 and 34 in Fig. 5) and in four D− genotypes (ATF5 genotypes 26, 27, 35 and 36 in Fig. 5). Results showed the presence of the intron variant I_1 in all genotypes regardless of the strength of the amplification signal or the relationship with the dehydrin RFLP (Supplementary Fig. 2). Taken together, these results suggest that the dehydrin RFLP is associated with an enrichment of group I_1 of intron variants.

Discussion

Analysis of recurrent selections is an effective strategy for the discovery of DNA polymorphisms associated with variations of quantitative traits (Falke et al. 2007). In the current study, we sought further insight into the allelic form of a Y_2K_4 dehydrin size variant previously found to be associated with a dehydrin RFLP with a demonstrated impact on freezing tolerance in alfalfa (Rémus-Borel et al. 2010). Our objective was to develop a robust marker to screen alfalfa genetic backgrounds for the presence of a cold adaptive locus. Limited data on full-length cDNAs or genomic sequences of dehydrins in alfalfa and in other

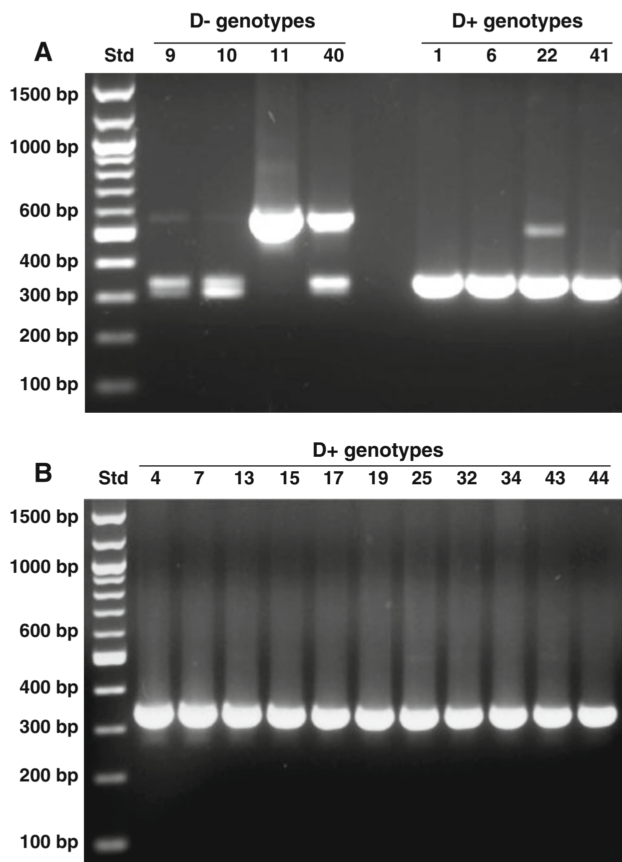


Fig. 4 **a** SCAR amplification of the intronic region of the G2 group of Y_2K_4 dehydrins in four ATF5 genotypes without (D−) and four genotypes with (D+) a dehydrin RFLP associated to superior tolerance to freezing using the 164c-299r primer pair. **b** SCAR amplification of 11 ATF5 genotypes with (D+) a dehydrin RFLP associated to superior tolerance to freezing, using the 164c-299r primer pair. Annealing temperature of SCAR primers was raised to 70°C to increase specificity. Molecular weight standards (Std) are indicated (100 bp ladder, New England Biolabs)

plant species severely limit the conduct of studies on their functional roles and the development of marker-assisted applications. To fill that gap, we investigated the variability of Y_2K_4 dehydrin size variants in *Medicago sativa* spp *sativa*. PCR primers designed to discriminate the three groups of size variants identified by Rémus-Borel et al. (2010) confirmed the enrichment of G2 sequences in response to recurrent selection for superior tolerance to freezing. A unexpectedly high frequency of G2 sequences in the initial genetic background indicated that only a subset of G2 sequences is related to the dehydrin RFLP.

Intron polymorphism identifies a dehydrin variant linked to freezing tolerance

The presence of a single intron is a feature typical of dehydrins that has been documented in Y_nK_n (Welling et al. 2004) and in Y_nSK_n (Mehta et al. 2009; Natali et al. 2003; Xiao and Nassuth 2006) genes. Pennycooke et al. (2008) reported the presence of a short intron of 190 bp in a Y_2K_4 dehydrin from *Medicago sativa* spp. *falcata* and of a larger intron of 698 bp in a homologous sequence from related *Medicago truncatula*. Analysis of full-length sequences of the G2 group of Y_2K_4 dehydrins in alfalfa revealed variable intron structures among otherwise highly homologous G2 sequences (92–100% identity). Natali et al. (2003) similarly observed a higher allelic diversity in the intron of a *Helianthus annuus* dehydrin gene than in the coding and the 3′-UTR regions. The identification of allelic variants within individual genotypes for each of the five groups of intron variants (I_1–I_5) indicated the presence of paralogs within the G2 group of Y_2K_4 dehydrins. Variations with regard to the consensus K motif (Close 1996) were observed in Y_2K_4 dehydrins of alfalfa. The frequent replacement of the highly conserved lysine (K) residue by asparagine (N) in position 2 and 8 of the K motif is particularly noteworthy as it may affect the secondary structure of the polypeptide. Analysis of intron sequences among genotypes with (D+) or without (D−) the dehydrin RFLP associated with superior freezing tolerance showed a noticeably lower intron variability in D+ genotypes (5 distinct sequences in two groups) than in the D− genotypes (16 distinct sequences in four groups). A group of sequences that share a characteristic short intron (variant I_1 in Fig. 3a; Table 3) was present at a high frequency (19 out of 20 sequences) in D+ genotypes; the other groups of variants (I_2–I_5) had a lower occurrence and were almost exclusively found in the D− genotypes. We thus conclude that intron polymorphism can be used to track the full-length sequence of a dehydrin variant previously linked to superior tolerance to freezing using RFLP analysis of recurrent selections (Rémus-Borel et al. 2010).

The conservative location of an intron of variable length within the Y_2K_4 dehydrins gave us the opportunity to pursue intron-length polymorphism (ILP) screening of G2 group of size variants. Using this approach, we confirmed the enrichment in population ATF5 as compared to ATF0, of a unique form of a small intron (I_1) which was accompanied by a concomitant decrease in the frequency of highly homologous Y_2K_4 dehydrin variants with a longer intron (I_5). Similar ILPs were observed between crosses of genotypes with (D+) and without (D−) the dehydrin RFLP providing strong evidence for a convergence between these two polymorphisms (RFLP and ILP). Whether the intron variability observed in our study is specific to the Y_2K_4 family of dehydrins or if it reflects a more general behavior of intron sequences in alfalfa is unknown. Giordani et al. (2011) recently analyzed intron polymorphisms in eight candidate genes cloned from maize inbred lines of contrasted level of drought tolerance. Although intron variability was low and in the same range than that of coding regions for seven of these genes, there was in contrast a

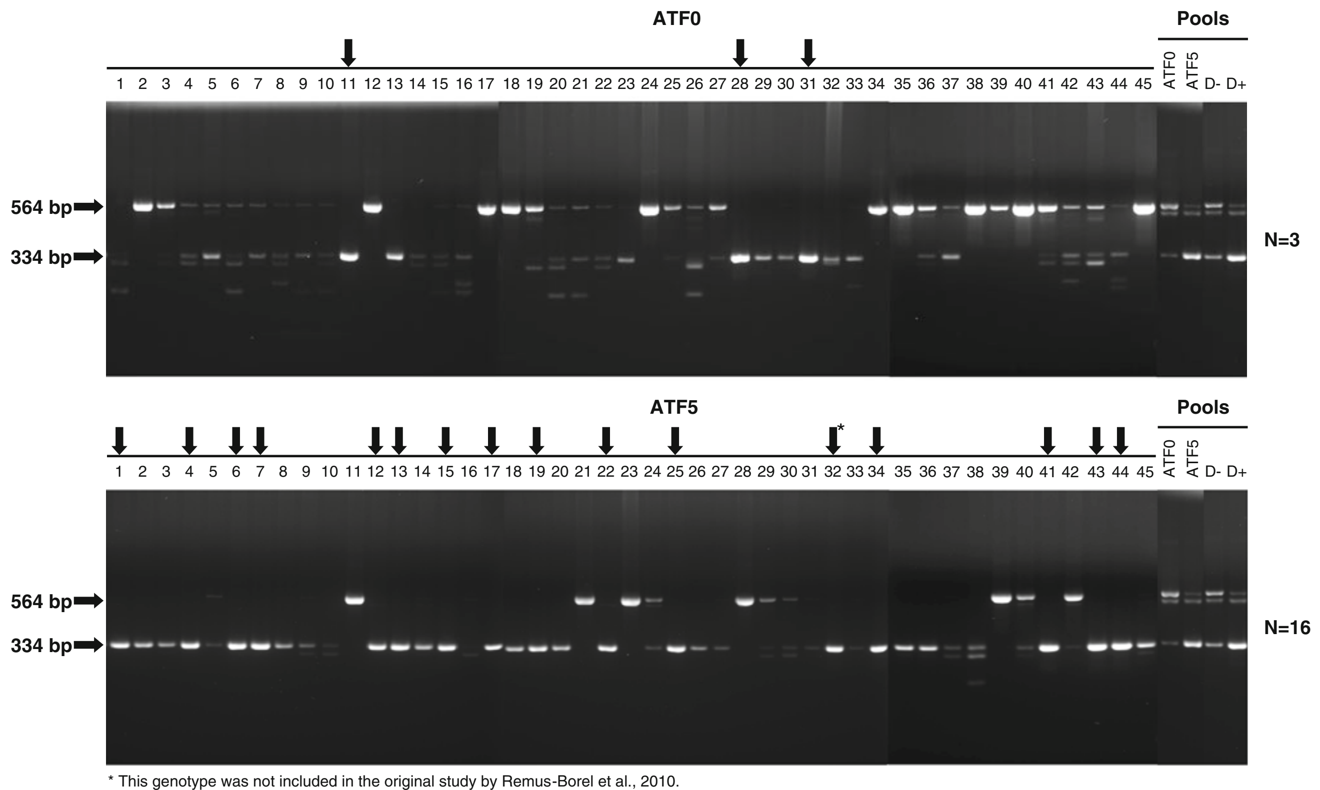


Fig. 5 SCAR amplification with primer pair 164c-299r using 45 genotypes within the ATF0 population and forty-five other genotypes within the ATF5 population. Amplification was also performed with DNA samples of pooled genotypes from the ATF0 and ATF5 populations and the D− and D+ crosses. Annealing temperature of SCAR primers was raised to 70°C to increase specificity. *Black arrows* indicate genotypes with intense amplification signal for the intron group I_1. Size of amplicons are indicated in the *left* margin while the number (*N*) of genotypes yielding a strong amplification signal for the I_1 intron group is indicated in the *right* margin

large nucleotide diversity for the intron of a dehydrin homolog. It is noteworthy that some of the introns in alfalfa Y_2K_4 dehydrins were comparatively larger (400–500 bp) than introns (80–260 bp) present in other dehydrins (Fan and Wang 2006; Mehta et al. 2009; Xiao and Nassuth 2006). The impact of intron length variation, if any, in alfalfa Y_2K_4 dehydrin is unknown. There is evidence that ILP may have an effect on phenotypic diversity in eukaryotes (Kumar et al. 2010; Sureshkumar et al. 2009). Whether or not ILP in Y_2K_4 dehydrins affects their expression or the phenotype are questions that should be addressed in the future. At this point, ILP can only be interpreted as a linkage association with the trait since functional aspects may lie in other mutations within translated or untranslated regions of the dehydrin variant or even in another gene that co-segregates with that allele.

SCAR markers can track ILP in Y_2K_4 dehydrins

Sequence-characterized amplified regions (SCAR) markers have been successfully used for the simple, low-cost selection of desirable traits in various plant species (Fondevilla et al. 2008; Gutierrez et al. 2007; He et al. 2009; Sakthivel et al. 2009). Based on the sequence of the full-length genomic sequences of Y_2K_4 dehydrins, we developed “intron-flanking” SCAR primers to discriminate intron variants. Our results showed that SCAR primers targeting the intron region can differentiate among highly homologous Y_2K_4 paralogs. A strong amplification of the short I_1 intron was shown to be co-inherited with a dehydrin RFLP associated to superior tolerance to freezing. Interestingly, intron sequence analyses revealed the presence of the short I_1 intron in both D+ and D− genotypes; That specific intron variant was however, more strongly amplified in D+ genotypes. The underlying causes of intensification of the short intron signal in genotypes with the dehydrin RFLP remain unknown. The possibility that the intensification of the I_1 intron signal observed in several genotypes from populations recurrently selected for superior freezing tolerance may be the result of copy number variation (CNV) is certainly worth investigating. Debolt (2010) recently observed an exceptional rate of CNV in *Arabidopsis* progeny obtained after only five generations of exposure to temperature stress and salicylic acid treatment. Whether a specific allele or the entire Y_2K_4 paralog defined by intron variant I_1 is associated to the RFLP remains to be

determined. Genotyping based on high-throughput sequencing of PCR products or high resolution melting (HRM) analysis of single nucleotide polymorphisms will help clarify this question in future association studies.

Conclusions

Using populations recurrently selected for superior tolerance to freezing and ILP, a SCAR marker targeting a Y_2K_4 dehydrin variant associated to cold adaptation was developed. The selection of germplasm has been identified as the greatest challenge for the identification of functional alleles (Kumar et al. 2010). Our results show that recurrent selections are rich reservoirs of rare alleles that can facilitate the identification of trait-related sequences among highly homologous variants arising from allelic forms within paralogs. Our results concur with the recent observation by Herrmann et al. (2010) that a candidate gene approach is a relevant strategy for the discovery of marker-trait associations in synthetic populations of autotetraploid alfalfa. Future studies will assess the impact of selection based on this SCAR marker on alfalfa tolerance to exposure to subfreezing temperatures and other abiotic stresses as well as the adaptive value of the dehydrin variant targeted by this marker.

Acknowledgments The authors want to acknowledge the invaluable contribution of Mrs. Josée Michaud and Mr David Gagné for their technical assistance. The participation of Marie-Pier Dubé to this project was supported by a NSERC visiting fellowship in Canadian federal government laboratories.

References

Beck EH, Fettig S, Knake C, Hartig K, Bhattarai T (2007) Specific and unspecific responses of plants to cold and drought stress. J Bio 32:501–510

Brini F, Hanin M, Lumbreras V, Amara I, Khoudi H, Hassairi A, Pages M, Masmoudi K (2007) Overexpression of wheat dehydrin DHN-5 enhances tolerance to salt and osmotic stress in *Arabidopsis thaliana*. Plant Cell Rep 26:2017–2026

Castonguay Y, Michaud R, Nadeau P, Bertrand A (2009) An indoor screening method for improvement of freezing tolerance in alfalfa. Crop Sci 49:809–818

Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plant 97:795–803

DeBolt S (2010) Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. Genome Biol Evol 2:441–453

Falke K, Flachenecker C, Melchinger A, Piepho H, Maurer H, Frisch M (2007) Temporal changes in allele frequencies in two European F_2 flint maize populations under modified recurrent full-sib selection. Theor Appl Genet 114:765–776

Fan Z, Wang X (2006) Isolation and characterization of a novel dehydrin gene from *Capsella bursa-pastoris*. Mol Biol 40:43–50

Fondevilla S, Rubiales D, Moreno M, Torres A (2008) Identification and validation of RAPD and SCAR markers linked to the gene *Er3* conferring resistance to *Erysiphe pisi* DC in pea. Mol Breed 22:193–200

Giordani T, Buti M, Natali L, Pugliesi C, Cattonaro F, Morgante M, Cavallini A (2011) An analysis of sequence variability in eight genes putatively involved in drought response in sunflower (*Helianthus annuus* L.). Theor Appl Genet 122:1039–1049

Gutierrez N, Avila C, Rodriguez-Suarez C, Moreno M, Torres A (2007) Development of SCAR markers linked to a gene controlling absence of tannins in faba bean. Mol Breed 19:305–314

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41

He YH, Ning GG, Sun YL, Qi YC, Bao MZ (2009) Identification of a SCAR marker linked to a recessive male sterile gene (*Tems*) and its application in breeding of marigold (*Tagetes erecta*). Plant Breed 128:92–96

Herrmann D, Barre P, Santoni S, Julier B (2010) Association of a constans-like gene to flowering and height in autotetraploid alfalfa. Theor Appl Genet 121:865–876

Houde M, Dallaire S, N'Dong D, Sarhan F (2004) Overexpression of the acidic dehydrin WCOR410 improves freezing tolerance in transgenic strawberry leaves. Plant Biotech J 2:381–387

Ismail AM, Hall AE, Close TJ (1999) Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Nat Acad Sci USA 96:13566–13570

Kumar GR, Sakthivel K, Sundaram RM, Neeraja CN, Balachandran SM, Rani NS, Viraktamath BC, Madhav MS (2010) Allele mining in crops: prospects and potentials. Biotech Adv 28:451–461

Marian CO, Krebs SL, Arora R (2004) Dehydrin variability among rhododendron species: a 25-kDa dehydrin is conserved and associated with cold acclimation across diverse species. New Phytol 161:773–780

Mehta PA, Rebala KC, Venkataraman G, Parida A (2009) A diurnally regulated dehydrin from *Avicennia marina* that shows nucleo-cytoplasmic localization and is phosphorylated by casein kinase II in vitro. Plant Physiol Biochem 47:701–709

Natali L, Giordani T, Cavallini A (2003) Sequence variability of a dehydrin gene within *Helianthus annuus*. Theor Appl Genet 106:811–818

Patton AJ, Cunningham SM, Volenec JJ, Reicher ZJ (2007) Differences in freeze tolerance of Zoysiagrasses: I. Role of proteins. Crop Sci 47:2162–2169

Pennycooke JC, Cheng H, Stockinger EJ (2008) Comparative genomic sequence and expression analyses of *Medicago truncatula* and alfalfa subspecies *falcata* cold-acclimation-specific genes. Plant Physiol 146:1242–1254

Ponting RC, Drayton MC, Cogan NOI, Dobrowolski MP, Spangenberg GC, Smith KF, Forster JW (2007) SNP discovery, validation, haplotype structure and linkage disequilibrium in full-length herbage nutritive quality genes of perennial ryegrass (*Lolium perenne* L.). Mol Genet Genom 278:585–597

Rémus-Borel W, Castonguay Y, Cloutier J, Michaud R, Bertrand A, Desgagnés R, Laberge S (2010) Dehydrin variants associated with superior freezing tolerance in alfalfa (*Medicago sativa* L.). Theor Appl Genet 120:1163–1174

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. In: Gelvin S, Schilperoort RA (eds) Plant Molecular Biology Manual A6. Kluwer Academic Publ, Dordrecht, The Netherlands, pp 1–10

Rorat T (2006) Plant dehydrins—tissue location, structure and function. Cell Mol Biol Lett 11:536–556

Sakthivel K, Shobha Rani N, Pandey M, Sivaranjani A, Neeraja C, Balachandran S, Sheshu Madhav M, Viraktamath B, Prasad G,

Sundaram R (2009) Development of a simple functional marker for fragrance in rice and its validation in Indian Basmati and non-Basmati fragrant rice varieties. Mol Breed 24:185–190

Sureshkumar S, Todesco M, Schneeberger K, Harilal R, Balasubramanian S, Weigel D (2009) A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*. Science 323:1060–1063

Welling A, Rinne P, Vihera AA, Kontunen SS, Heino P, Palva ET (2004) Photoperiod and temperature differentially regulate the expression of two dehydrin genes during overwintering of birch (*Betula pubescens* Ehrh.). J Exp Bot 55:507–516

Xiao H, Nassuth A (2006) Stress- and development-induced expression of spliced and unspliced transcripts from two highly similar dehydrin 1 genes in *V-riparia* and *V-vinifera*. Plant Cell Rep 25:968–977

Yin Z, Rorat T, Szabala BM, Ziokowska A, Malepszy S (2006) Expression of a *Solanum sogarandinum* SK_3-type dehydrin enhances cold tolerance in transgenic cucumber seedlings. Plant Sci 170:1164–1172