

Exploring the genetic diversity of the *DRF1* gene in durum wheat and its wild relatives

Domenico Di Bianco^{1,2}, Karthikeyan Thiyagarajan^{1,2}, Arianna Latini^{1†}, Cristina Cantale¹, Fabio Felici¹ and Patrizia Galeffi^{1*}

¹ENEA Casaccia (UTAGRI-GEN), Rome, Italy and ²Scuola Superiore Sant'Anna, Pisa, Italy

Abstract

A drought-related gene belonging to the Dehydration Responsive Element Binding protein (DREB) family has been reported and characterized in durum wheat. Unlike other *DREB*-homologous genes, it consists of four exons and three introns and produces three transcripts by an alternative splicing mechanism. The gene sequence was analysed in a number of varieties/lines/accessions of durum wheat, triticale and in wheat genome donors, *Aegilops speltoides*, *A. tauschii* and *Triticum urartu*, in order to evaluate the variability and to detect other interesting molecular features. Herewith, some results are presented. In the exon 1, a single sequence repeat codifying for a stretch of alanine residues variable in length (from 3 to 7), was identified. A novel non-autonomous transposon was identified, encompassing the intron1–intron3 region and this was characterized in detail. Part of the exon 4, containing the APetala2 (AP2) domain, responsible for DNA recognition and binding, was isolated and sequenced in a collection of *Aegilops* species and *A. speltoides* accessions from the Fertile Crescent, a region characterized by a high wheat biodiversity. Although the 338 bp-long analysed sequences did not reveal substantial differences in the polymorphic patterns, using a geographic subdivision with three clusters (east, centre and west), they completely separated *Aegilops* from the *A. speltoides* genus.

Keywords: AP2 domain; biodiversity; DREB-family; single sequence repeat; transposon

Introduction

TdDRF1 (*T. durum* dehydration responsive factor 1), a *DREB2*-related gene that was isolated and characterized in durum wheat (Latini *et al.*, 2007), belongs to the AP2 gene family and is highly homologous to the barley *HvDRF1* gene (Xue and Loveridge, 2004) and to the bread wheat *wdreb2* gene (Egawa *et al.*, 2006). These genes share a complex gene structure, consisting of four exons and three introns, and produce three transcript variants by means of alternative splicing. In the present investigation, the different parts of the gene have been systematically analysed in several

accessions of durum wheat, triticale, wheat genome donors (*A. speltoides*, *A. tauschii* and *Triticum urartu*) as well as in other related plants.

Material and methods

Eight *T. durum* genotypes (total 67 accessions), a few accessions of *T. urartu*, *A. tauschii* and triticale, as well as 69 accessions of *A. speltoides v speltoides*, *A. speltoides v ligustica* and other *Aegilops* were grown in the greenhouse and used for DNA extraction.

The amplified fragments (PCR thermal cycle reaction: initial denaturation 94°C for 5 min, 94°C for 1 min, 55°C for 1 min, 72°C for 1 min and 30 s and final extension at 72°C for 7 min, then at 4°C for storage) were gel-purified, cloned (pCR®II-TOPO® vector by Invitrogen, USA) and sequenced (ABI 3730 DNA analyzer; Applied Biosystems, USA) following standard procedures.

*Corresponding author. E-mail: galeffi@enea.it

†Present address: Canadian Forest Service, Québec City, Québec, Canada.

D. D. Bianco and K. Thiyagarajan contributed equally to this work.

An exon 1 region of 150 bp at the beginning of protein CoDing Sequence (CDS), was sequenced in 100 genotypes (E1 For: 5'-AAGTCGACGCGCGCAA 3'; SSR1 Rev: 5'-CCG-GGATCTCGAAGGGTG 3').

An exon 4 part, from 400 bp to 1 kb long, including the whole AP2 domain coding region, was sequenced in 168 samples of *Aegilops* (E4 FOR: 5'-ATGATCCACAGGGT-GCAA 3'; E4 Rev: 5'-GGTCCACCATTGATCTTCATT 3').

Six full-length *DRF1* gene sequences, namely FJ858188, FJ858187, FJ843102, EU089819, EU197052 and GU017675, were submitted to Genbank and used for transposon identification and analysis.

Several computational tools were used for phylogenetic and molecular evolutionary analyses (Rozas *et al.*, 2003; Excoffier *et al.*, 2005; Huson and Bryant, 2006; Tamura *et al.*, 2007), sequence analyses as well as identification of repeats and palindromes (Kurz *et al.*, 2001; Kohany *et al.*, 2006), secondary structure prediction and modelling (for details see Supplementary Table S1, available online only at <http://journals.cambridge.org>).

Results and discussion

The various analyses provided interesting insights into the *DRF1* gene which are outlined below, separately, for each exon.

Exon 1

Simple sequence repeat (SSR) genetic polymorphism occurs due to variations in the number of repeated units, probably due, in turn, to slippage during DNA replication (Levinson and Gutman, 1987; Taylor *et al.*, 1999). An SSR was identified in the *DRF1* gene (Fig. 1). This SSR is located

at the N terminus of exon 1, position 10, and codes for a stretch of alanine residues variable in length (from 3 to 7). Another study on SSR loci in expressed sequences suggested the potential correlation between mutational events within SSR repeats and the evolutionary relationships across taxa (Rossetto *et al.*, 2002). It is worthwhile pointing out that the shortest SSRs belong mainly to the ancestor of *Triticum* and to *Brachypodium distachyon*.

Variability was also observed, resulting from a single nucleotide mutation (ALA to SER or THR), exclusively at the beginning or at the end of the stretch. The ALA stretch is present also in *Leymus chinensis* and *B. distachyon*, suggesting that this SSR represents a shared feature of the *Pooideae* subfamily (*Poaceae* family), and is not present in other subfamilies such as *Ehrhartoideae* (*Oryza sativa*) and *Panicoideae* (*Zea mays*). A double mutation changing ALA to LEU was observed in *L. chinensis*, while a single change, GLU to VAL, was found in *B. distachyon*.

Further investigations are needed to better understand the possible linkage of this SSR and its variability, both at genomic and transcriptomic level, in relationship with maps and phenotypic traits.

Exons 2–3

The six available full-length gene sequences from *T. durum* and *Aegilops* were analysed in order to identify possible transposable elements. Repeats, palindromes, tandem and inverted repeats were investigated. All the identified elements, namely a terminal inverted repeat (TIR-32 bp), a target site duplication (TSD-2 bp), an internal TSD (ITSD-4 bp), direct and reverse repeats and a variable number of tandem repeats revealed the presence of a transposable element in the *DRF1* gene which had never been described before.

E AAAAAA A P	34%	cDNA from many durum cultivars, as Ciccio, and some pseudogene forms from <i>A. speltoides</i>
E SAAAAA L P	21%	cDNA from pseudogene form from Ciccio durum cultivar
E AAAAAA A P		cDNA from <i>L. chinensis</i>
E AAAAAA T P	16%	cDNA from many durum cultivars, as Cannizzo, and some gDNA from <i>T. urartu</i> accessions
E AAAAA A P	26%	cDNA from many durum cultivars as Karalis, Duilio, Creso and Colosseo
E TAAAAA A P	4%	cDNA from Triticale Pollmer
E AAAA A P	4%	gDNA from <i>A. speltoides</i> and <i>A. tauschii</i> accessions
E AAAA A P	2%	gDNA from <i>A. speltoides</i>
E AAAT P	12%	gDNA from many <i>T. urartu</i> accessions
V AAAP		gDNA from <i>B. distachyon</i>

Fig. 1. Sequences and frequencies of the ALA-stretch in *T. durum*, *A. speltoides*, *T. urartu*, *A. tauschii* and in two related plants. (*Brachypodium d.*: Bd2:29505646.29515646, <http://www.brachybase.org/blast/>; *Leymus c.*: GenBank accession EU999998.1).

Thus, a novel transposable element, approximately 1.5 kb in length, was identified between intron 1 and intron 3, including exon 2 and exon 3. Since no sequence for transposase was found in the transposon identified, it was classified as a non-autonomous transposon. Furthermore, the presence of nested TIR-18bp was also observed, the sequences of which are highly homologous to the longest ones (Supplementary Fig. S1, available online only at <http://journals.cambridge.org>).

The new transposon is now included in Repbase (see Report at <http://www.girinst.org/2009/vol9/issue3/AsDRF1.html>; Thiyagarajan *et al.*, 2009).

All analysed genes showed the above observations, with minor mismatch/deletion, mainly located at 3' Terminal Inverted Repeat (TIR) that appears to be the most variable part of the transposable element.

The transposon is inserted into a poor CG (about 40%) region and its core (exon 2 + intron 2 + exon 3 and few bps of intron 1 and intron 3 at flanking sides) appears to be very similar (85% identity, 372 bp fragment, $E = 10^{-69}$) to a sequence from *B. distachyon*, suggesting that the transposon might have played a vital role in moving these exons during evolution of the *Pooideae* subfamily. Furthermore, it is tempting to hypothesize its possible involvement in the mechanism of the alternative splicing regulation of the *DRF1* gene.

Exon 4

Even if most of the variations intra- and inter- species are selectively neutral (Nei, 1987), there is an increasing interest in detecting genes, or genomic regions, that affect the fitness of the organisms (Nielsen, 2005). As the exon 4 sequence of *DRF1* gene contains the region codifying the AP2 domain, responsible for DNA recognition and binding, 168 sequences (137 from *A. speltoides* and 31 from other *Aegilops* species) were analysed to evaluate variability (Supplementary Fig. S2, available online only at <http://journals.cambridge.org>).

Overall, 279 sites were analyzed (excluding gaps/missing data) for all 168 sequences; 67 polymorphic sites were identified (25 singleton variable sites, two variants; 35 parsimony informative sites, two variants; and seven parsimony informative sites, three variants). Out of 67 polymorphic sites, five were located in the AP2 domain and their effect was investigated in a structural model. The AP2 local geometry was not affected (data not shown).

To investigate the evolutionary relationships among the 168 sequences, a minimum spanning tree was built. Two main groups were observed, lineage I and lineage II, separated by 21 mutation steps (Fig. 2). The neighbour-joining tree, estimated using a Kimura-2 parameter

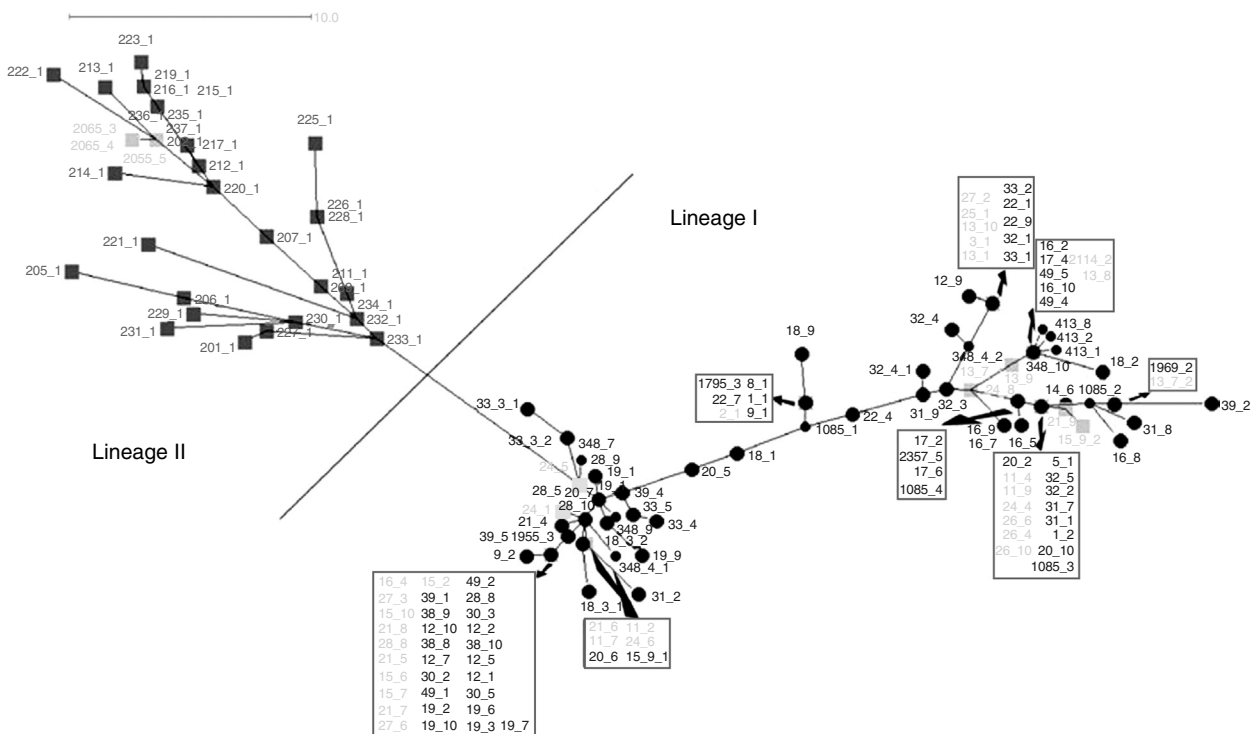


Fig. 2. Minimum spanning tree of 168 genotypes of *Aegilops* calculated by an exon 4 region of the *DRF1* gene. *A. speltoides* ssp *ligustica* is shown in light grey (■), *A. speltoides* ssp *speltoides* is shown in black (●), *A. speltoides* is shown in black (●) and other *Aegilops* are shown in dark grey (■).

model, clustered in the same way, a 70% bootstrap value (data not shown).

Lineage I is a complex network, with two main groups, 'star-like' pattern, including all except three *A. speltooides* accessions. These three genotypes, classified as *A. speltooides* (2065_3, 2065_4 and 2065_5), cluster in the lineage II, together with all *Aegilops* genotypes. Possibly, a reconsideration of the taxonomic assignment could be proposed, even if more loci have to be analysed before reaching a final conclusion.

A geographic subdivision was applied to investigate geographic frequency patterns, but the design, based on three clusters (east, centre and west) was unable to reveal substantial differences in polymorphic patterns, thus suggesting a neutral selection.

A design based on taxonomy was also tested; the variability observed in the analysed DNA region was unable to distinguish *ssp speltooides* from *ssp ligustica*, while it was able to distinguish *Aegilops* from *A. speltooides* (data not shown). Analysis of molecular variance (AMOVA) was carried out testing the genetic structure (two groups and four populations). The *F*-statistic was shown to be quite significant (*F*-statistic, i.e. specific population index (F_{st}) P value = 0.00) and a substantial variation between *A. speltooides* and *Aegilops* species was found, accounting for 62.34% of the total variation, even if not significant, $P > 0.1$. Due to missing data, locus-by-locus AMOVA provided a better estimate and, indeed, the data were significant (54.7%, $P = 0.00$).

In conclusion, the detailed analysis of the sequence of the *DRF1* gene and its variability intra- and inter-species highlighted some interesting features characterizing the complex structure of this gene, clearly contributing to our overall knowledge and thus opening new perspectives regarding the evolution and regulation of this gene.

Acknowledgements

Authors thank Dr. K. Ammar (CIMMYT) for valuable suggestions during this investigation and Mrs. Marian Boreham for the revision of English language.

References

- Egawa C, Kobayashi F, Ishibashi M, Nakamura T, Nakamura C and Takumi S (2006) Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes Genetics and Systems* 81: 77–91.
- Excoffier L, Laval G and Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47–50.
- Huson DH and Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.
- Kohany O, Gentles AJ, Hankus L and Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics* 7: 474.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J and Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29: 4633–4642.
- Latini A, Rasi C, Sperandei M, Cantale C, Iannetta M, Dettori M, Ammar K and Galeffi P (2007) Identification of a DREB-related gene in *Triticum durum* and its expression under water stress conditions. *Annals of Applied Biology* 150: 187–195.
- Levinson G and Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221.
- Nei M (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197–218.
- Rossetto M, McNally J and Henry RJ (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. *Theoretical and Applied Genetics* 104: 61–66.
- Rozas J, Sanchez-Delbarrio JC, Messeguer X and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
- Tamura K, Dudley J, Nei M and Kumar S (2007) *MEGA4*: Molecular Evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.
- Taylor JS, Durkin JMH and Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Molecular Biology and Evolution* 16: 567–572.
- Thiyagarajan K, Latini A, Galeffi P and Porceddu E (2009) A non-autonomous DNA transposon inserted in the first and third intron of the *AsDRF1* gene. *Rebase Reports* 9(3): 726, ISSN# 1534-830X.
- Xue G and Loveridge CW (2004) HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *The Plant Journal* 37: 326–339.