Characterisation of puroindo-line genes in wild tetraploid and hexaploid wheats (*Triticum araraticum*, *T. timopheevii* and *T. zhukovskyi*)

Simeone MC[1], Sestili F[2], Laino P[2], Lafiandra D[2]

[1]Department of Agroforestry, University of Tuscia, Italy

[2]Department of Agrobiology & Agrochemistry, University of Tuscia, Italy

INTRODUCTION

Puroindolines are small (ca. 14 kDa) basic, cysteine-rich proteins found in the caryopses of many taxa of the *Triticeae* tribe. They belong to a protein super-family[1] that includes alpha-amylase/trypsin inhibitors, nonspecific lipid binding proteins and a mixture of puroindoline-like polypeptides (Grain Softness Proteins, GSPs). In recent years, puroindolines have gained a considerable interest among wheat geneticists and breeders, offering new perspectives in the genetic improvement of cereals, spanning from end-use applications[2] to disease resistance[3]. In this context, wheat (*Triticum aestivum*, AABBDD) constitutes the model species for biochemical and molecular studies of the puroindolines system, owing to the distinction of well suited market classes of end-uses based on grain texture. Several lines of evidence have repeatedly shown that the main sources of variation are costantly associated with the presence/absence or sequence polymorphism of puroindoline A and B, i.e. the products of the *Ha* locus, located on the short arm of chromosome 5D.

In order to assess the extent of such variation, several puroindoline alleles have recently been identified in analysing cultivated and wild wheats, including A, B, D, C and U diploids, barley, oat and rye[4]. Remarkably, AB tetraploid wheats have repeatedly shown to be devoid of puroindolines, whereas other major polyploid lineages of *Triticeae* have been little inspected and with contrasting results. In this paper we report the discovery and characterisation of puroindoline A and B genes in some wild tetraploid and hexaploid wheats [*Triticum araraticum* (AAGG), *T. timopheevii* (AAGG), and *T. zhukowsky* (AAAAGG)], where the absence of the *Ha* locus had been previously postulated due to unsuccesful isolation of the gene sequences.

MATERIALS AND METHODS

Triticum araraticum (genome composition AAGG), *T. timopheevii* (AAGG), and *T. zhukowsky* (AAAAGG) were investigated for the presence of puroindoline A and B. Total genomic DNAs were extracted from individual half kernels as described by Dellaporta *et al.*[5]. Puroindoline coding regions were PCR amplified with wheat specific primers as described by Gautier *et al.*[6]. Electrophoresis were performed on 1.4% agarose gels. Single-banded amplification products were purified with Nucleospin extract (Macherey-Nagel), ligated in pGemTeasy (Promega), and cloned in DH5α competent cells. Positive clones were sequenced in both directions with universal primers; Cycle Sequencing and the BigDye Terminator Ready Reaction Kit (Applied Biosystems) were used. Data were collected on ABI Prism 373A automated gel reader. Resulting electropherograms were further checked with the software CHROMAS 2.3 (www.technelysium.com.au), assembled and aligned for several standard descriptive parameters (including multiple alignments, protein translation, percentage pairwise distance) with the Sequence Analysis Software DNAMAN-2003 (Lynnon Biosoft). Puroindoline sequences of related taxa where derived from the GenBank and used for test comparisons.

RESULTS

Puroindoline A and B coding regions were discovered (Fig. 1, 2) and sequenced in wild tetraploid (*Triticum araraticum*) and cultivated form (*T. timopheevii*) and hexaploid (*T. zhukowskyi*) wheats. We failed in the gene detection of only one accession of *T. zhukowskyi*, where the cloned fragments amplified with both Pin A and Pin B primer pairs displayed unknown sequences, highly divergent (26%) from any other known puroindoline sequence. These fragments were considered as PCR artifacts.

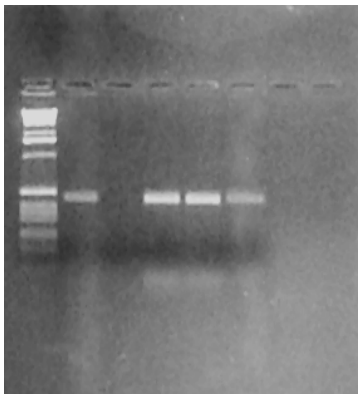

Fig. 1. PCR amplification of Pin A in bread wheat (Chinese Spring), durum wheat (Langdon), *T. timopheevii, T. araraticum* and *T. zhukowskyi*.

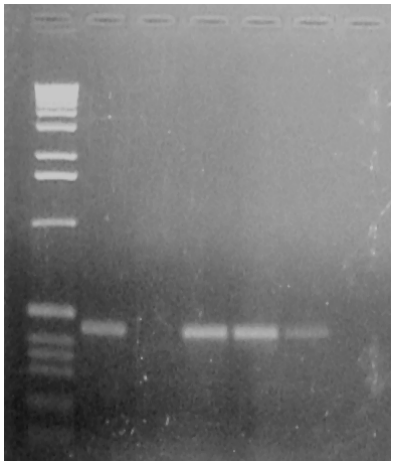

Fig. 2. PCR amplification of Pin B in bread wheat (Chinese Spring), durum wheat (Langdon), *T. timopheevi, T. araraticum* and *T. zhukowskyi*.

The four Pin A sequences displayed an overall 99.89% nucleotide identity: the three tetraploid accessions were 100% identical, whereas *T. zhukowskyi* showed two nucleotide substitutions. Identity with bread wheat (Chinese Spring) ranged between 98.2-98.7% (6 nucleotide substitutions with the tetraploids, 2 with *T. zhukowsky*). Identity with other wild wheats was 96.4% with *Ae. speltoides* (genome composition BB), 98.7% with *T. tauschii* (DD), and 99.8% with *T. urartu* (AA). Despite this variation (which coded 5 amino acid substitutions totally, compared to the wild type) the main puroindoline features (the W-rich domain, the 10 C-backbone), were all preserved. A Neighbor Joining analysis was run with default options and boostrapping (1000 replications); three main clades emerged from the root, corresponding to genome B (outgroup), genome D and genome A puroindolines (Fig. 3). The 3 wild taxa of the present study clustered together with *T. urartu* and *T. monococccum*.

The four Pin B sequences displayed an overall 99.61% nucleotide identity, corresponding to 7 substitutions: the two *timopheevii* accessions differed by 1-bp, the two tetraploid species showed 2 nucleotide substitutions.

Identity with Chinese Spring ranged between 93.7-94.2%. Identity with other wild wheats was 92.3-93.6% with *Ae. speltoides*, 95.7-96.2% with *T. tauschii*, 99.1-99.3% with *T. urartu*. Compared to the wild type, this variation coded 14 amino acid substitutions; the main puroindoline features were not preserved in *T. araraticum*, which showed a $^{C}/_{S}$ replacement, with the consequent breaking of the 10 C-backbone. This amino acid substitution is likely to compromise the protein function.

Congruently with data obtained for the Pin A sequences, the Neighbor Joining tree displayed the Pin B sequences of the studied germplasm grouping together with *T. urartu* and *T. monococcum*, thus confirming the closer relationships shared with the A genome groups (Fig. 4).

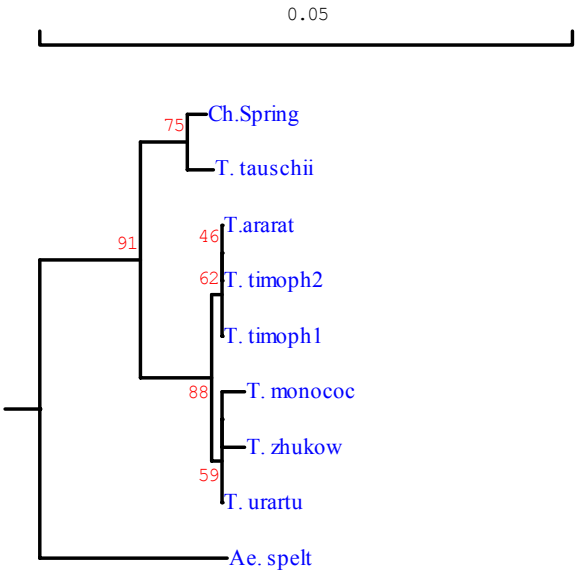

Fig. 3. Neighbor joining tree of Pin A sequences in wild and cultivated wheats. Bootstrap support indicated above branches.

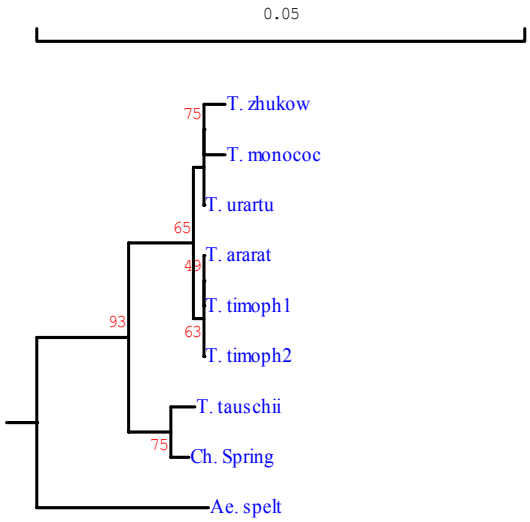

Fig. 4. Neighbor joining tree of Pin B sequences in wild and cultivated wheats. Bootstrap support indicated above branches.

DISCUSSION

Wild tetraploid and hexaploid wheats bearing genomes other than AB and ABD compositions showed to possess the coding sequences of both puroindoline types A and B. These genotypes are thus likely to have escaped the genomic rearrangements that shaped the Hardness locus in durum and bread wheats, under the

form of multiple insertions, deletions, duplications and insertions of transposable elements that have recently been explained[7,8]. Gautier *et al*.[9] firstly postulated the absence of these genes in tetraploid wheats, including *T. timopheevii*. However, in a recent study Pickering and Bhave[10] detected Puroindoline A in 11 accessions of *T. timopheevii* and in 1 accession of *T. zhukowskyi*, but failed to detect any purondoline B sequence in both taxa; as well, 6 accessions of *T. zhukowskyi* displayed no puroindolines. We isolated and characterised Puroindoline A and B in both these taxa, as well as in *T. araraticum*, here investigated for the first time. Our findings constitute the first evidence for the presence of the complete puroindoline molecular system in these genotypes. Therefore, we postulate that the wild germplasm of AG and AAG wheats is likely to possess a large puroindoline variation, including wild types, double-null and null puroindoline forms.

Nucleotide sequence comparisons displayed no insertions/deletions; we detected various degrees of point mutations compared with wild type from cultivated wheats and with alleles detected in other *Triticeae*.

Accordingly to recent data[4], variation in Pin B exceeded that in Pin A in every analysed genotype.

Despite this, the deduced main features of the proteins secondary structure were all mantained. Exception is one accession of *T. araraticum* which showed the lack of a Cysteine, for which we can assume disruption of the C-backbone with the consequent loss of functionality.

All the newly found genes in this study are likely to belong to the A genome groups, owing to the highest sequence identity shown with other genome A bearers (e.g. *T. urartu and T. monococcum*). The cloning strategy did not allow the identification of further sequences that could be ascribed to the G genome groups. In facts, sequencing of multiple clones from the same genotype showed no evidence for the amplification of extra-sequences resembling a puroindoline gene. This finding could be interpreted both as a consequence of primers specificity, which might have prevented the identification of diverging genes (at least in the priming sites), as well as a true lack of the genes on the G chromosome groups.

Southern analyses and characterisation of the friabilin fraction are currently ongoing to investigate gene copies and protein expression in the tetraploid and hexaploid wheat germplasm of the present study.

REFERENCES

1. Douliez, J.-P., Michon, T., Elmorhani, K., Marion, D. (2000) J. Cereal Sci. 32, 1–20.
2. Morris, C.F. (2002) Plant Mol. Biol. 48, 633-647.
3. Giroux, M.J., Sripo, T., Gerhardt, S., Sherwood, J. (2003) Biotech.Genet. Engin. Rev. 20, 277–290.
4. Bhave, M., Morris, C.F. (2008) Plant Mol. Biol. 66, 205-219.
5. Dellaporta, S.L., Wood, J., Hicks, J.B. (1983) Plant Mol. Biol. Rep. 1, 19-21.
6. Gautier, M.F., Aleman, M.E., Guirao, A., Marion, D., Joudrier, P. (1994) Plant Mol. Biol. 25, 43-57.
7. Turnbull, K.-M., Turner, M., Mukai, Y., Yamamoto, M., Morrell, M.K., Appels, R., Rahman, S. (2003) Genome 46, 330–338.
8. Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.F., Cattolico, L.,Beckert, M., Aubourg, S., Weissembach, J., Caboche, M., Bernard, M., Leroy, P., Chalhoud, B. (2005) Plant Cell 17, 1033–1045.
9. Gautier, M.-F., Cosson, P., Guirao, A., Alary, R., Joudrier, P. (2000) Plant Sci. 153, 81-91.
10. Pickering, P.A., Bhave, M. (2007) Plant Sci. 172, 371-379