ELSEVIER

Journal of Cereal Science 44 (2006) 75–85

Journal of CEREAL SCIENCE

www.elsevier.com/locate/jnlabr/yjcrs

Transcriptome analysis reveals differentially expressed storage protein transcripts in seeds of *Aegilops* and wheat

Yunchao Kan [a,b], Yongfang Wan [a], Frederic Beaudoin [a], David J. Leader [a,1], Keith Edwards [c], Rebecca Poole [c], Daowen Wang [b], Rowan A.C. Mitchell [a], Peter R. Shewry [a,*]

[a] *CPI Division, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*
[b] *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*
[c] *Department of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK*

Received 4 January 2006; received in revised form 3 April 2006; accepted 5 April 2006

Abstract

Transcriptional profiles of developing grain of wheat and three *Aegilops* species (*Ae. caudata*, *Ae. cylindrica*, *Ae. tauschii*) were compared using an EST-based array procedure. This identified 22 genes which were over-expressed in all three *Aegilops* species, two of which were 'weakly similar' to avenin storage proteins of oats. Sequencing of the corresponding transcripts demonstrated that each corresponded to a small family of proteins called avenin-like a and b. Sequence comparisons demonstrated that these proteins belong to the 'prolamin superfamily' of plant proteins, with the closest relatives being the γ-gliadins and LMW subunits of wheat and avenins of oats. Furthermore, the type a and b proteins differ from each other in that the latter contain a duplicated sequence of about 120 residues. The a-type proteins clearly correspond to the LMW gliadins characterised previously. In contrast, proteins corresponding to the b-type sequences have not been previously characterized but may form part of the glutenin fraction and hence contribute to processing quality. The higher expression levels of the avenin-like proteins in *Aegilops* species and variation in the amino acid sequences of the b-type proteins between the species suggest that they could provide a source of variation for wheat improvement.

Keywords: Avenin-like proteins; Transcriptome; Grain development; Processing quality

1. Introduction

Cultivated wheat (*Triticum aestivum*) is a hexaploid species with three homoeologous (i.e. homologous but non-pairing) genomes (A, B, D), each comprising seven pairs of chromosomes (Feldman et al., 1995). It is thought to have arisen some 8000 years ago by natural hybridisation between tetraploid *T. turgidum* (possibly cultivated emmer, *T. turgidum* var *dicoccum*), which has the A and B genomes, with the wild grass *Ae. tauschii* which has the D genome (Feldman and Sears, 1981). The A genome of bread wheat is closely related to those of A genome diploids, cultivated and wild forms of einkorn (*T. monococcum*) and wild *T. urartu*. In contrast, the origin of the B genome is still not clear but it is thought to be related to the S genome present in a number of diploid *Aegilops* species (Feldman and Sears, 1981).

The close relationship between the A, B and D genomes of cultivated wheat and the genomes of wild and cultivated species of *Aegilops* and *Triticum* means that the latter can be exploited as sources of useful genes for wheat improvement, using either crossing and introgression (Feldman and Sears, 1981; Knott, 1987) or genetic engineering technology (Jones, 2005). This is particularly straightforward for diploid species with the D genome as these species can be crossed with tetraploid *T. turgidum* to give synthetic hexaploids, which can then be incorporated into breeding programmes for bread wheat (Feldman and Sears, 1981).

We have therefore been studying species of the *Aegilops* sections *Vertebrata* and *Cylindropyrum*, which have the related C and D genomes (Dvorak et al., 1998; Linc et al., 1999; Miller, 1987), to identify novel types of gluten protein which can be exploited to improve the functional properties of bread wheat (Wan et al., 2000, 2005). In the present study we use transcriptome analysis to identify novel storage protein transcripts, which are highly expressed in three species of

* Corresponding author. Tel.: +44 1582 763133; fax: +44 1582 763010.
E-mail address: peter.shewry@bbsrc.ac.uk (P.R. Shewry).
[1] Present address: Scottish Crops Research Institute, Invergowrie, Dundee DD2 5DA, UK.

0733-5210/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2006.04.004

Aegilops compared with wheat and speculate that the proteins encoded by these transcripts may be exploited to improve wheat quality.

2. Materials and methods

Seed of *Aegilops caudata* (CC) (accessions Ae14, Ae15, Y511) were provided by the Triticeae Research Institute, Sichuan Agricultural University (Dujiangyan 611830, China) and of *Ae. cylindrica* ($C^cC^CD^CD^C$) (accessions RM0109, RM0115, RM0111) and *Ae. tauschii* (DD) (accession RM0198) by Dr L Li, Institute of Crop Germplasm Resources of the Chinese Academy of Agricultural Sciences (Beijing, China). Seeds of *Ae tauschii* (accession 222005) were obtained from the John Innes Centre (Norwich, UK).

2.1. RNA isolation

Whole grains were isolated from the middle spikelets of ears at 14 days after anthesis in the eight *Aegilops* accessions and wheat (cv Cadenza). Total RNA fractions were extracted in triplicate from each grain sample as previously described (Chang et al., 1993) and the integrity of the RNA was checked with the Agilent Bioanalyser.

2.2. Arrays

Indirect labelling of cDNA samples, hybridisation to arrays and scanning were performed as previously described (Wilson et al., 2004). To compensate for dye bias, samples were divided in two and labelled with Alexa Fluor 555 and 647 for dye-swap comparisons. Each of the eight *Aegilops* accessions were hybridised in turn to the array, with a wheat sample labelled with the appropriate dye, totalling six replicate arrays (three RNA isolation replicates × reverse dye) for each comparison.

Image analysis was performed using GenePix Pro 5.0 software (Molecular Devices Corporation). Background-corrected median intensities for the two emission wavelengths were imported into the GeneSpring 7.1 (Agilent Technologies) analysis package. Data were transformed such that the signal ratio was always calculated as *Aegilops*/wheat. Data were normalised using Lowess regression to remove any intensity-dependent bias and ensure the median log ratio of signals was zero for each slide. A set of genes was identified whose relative expression was at least 1.4 fold different (i.e. >1.4 or <0.71) in *Aegilops* compared to wheat. This set of genes was then further filtered to give those whose relative expression was significantly different at $P<0.05$ as estimated by a *t*-test with a Benjamini–Hochberg multiple testing correction. Annotations were assigned to ESTs from the top BLASTX hit against the NCBI/EBI/DDBJ Uniprot database, or where this was uninformative, from the top BLASTN hit within the TIGR wheat gene index (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=wheat). Blast analyses for annotation and EST counts were performed with the BLASTALL programme (Altschul et al., 1997) and custom Perl scripts.

2.3. Cloning of transcripts

Similar wheat ESTs present in the public dbEST database (http://www.ncbi.nlm.nih.gov/projects/dbEST) to the array elements were identified by BLAST and assembled into contigs. The consensus sequences from these were used to design primers to amplify cDNA templates from the samples used on the arrays. PCR reactions and product isolation were as previously described (Wan et al., 2005). PCR products were ligated into pGEM-T easy vector (Promega) which was used to transform NovaBlue Singles competent cells (Novagen) using manufacturers' protocols. Positive clones were identified and minipreps of DNA made with QIAprep kit (QIAGEN). Sequencing was carried out using BigDye reactions by the Sequencing Service of Oxford University, UK.

3. Results and discussion

3.1. Identification of novel transcripts in Aegilops by microarray

Wilson et al. (2004) described the construction of a transcriptome resource for functional genomic studies of wheat. This comprises a high density microarray of 8966 unique ESTs selected from 35 cDNA libraries prepared from wheat (http://www.cerealsb.uk.net) with approximately two thirds of the elements on the array being from libraries derived from developing grain tissues. We decided to use this resource to identify novel transcripts expressed in developing seeds of *Aegilops* spp. To do this we focused on three species, *Ae. tauschii* (section *Vertebrata*) which is a diploid species which shares the D genome of bread wheat, *Ae. caudata* (section *Cylindropyrum*) which is a diploid with the related C genome and *Ae. cylindrica* (section *Cylindropyrum*) which is a tetraploid with the C^c and D^c genomes. Seeds from two accessions of *Ae. tauschii* and three accessions each of *Ae. caudatai* and *Ae. cylindrica* were planted in the glasshouse and developing whole seeds harvested at 14 days after flowering. Developing seeds were also harvested from the UK spring wheat cv Cadenza at 14 days after flowering for comparison.

Approximately, half the genes represented on the array (4199) were significantly differently expressed in wheat when compared to at least one of the *Aegilops* accessions. This set of genes was used to generate a tree showing similarities of relative gene expression among accessions (Fig. 1). The *A. caudata* accessions Ae14 and Ae15 are most similar, but other accessions did not cluster together by species, indicating substantial differences between accessions.

To identify genes differentially expressed between wheat and all *Aegilops* species, those common to all the significantly down-regulated sets and all the significantly up-regulated sets for the eight *Aegilops* accessions were identified. This resulted in a set of five genes with lower relative expression and a set of 22 genes with higher relative expression for every *Aegilops* accession (Table 1). The source library of each EST is given as a guide to possible differences in tissue abundance between *Aegilops* and wheat. Since RNA was isolated from whole