

Genome-level identification of cell wall invertase genes in wheat for the study of drought tolerance

Hollie Webster^A, Gabriel Keeble^B, Bernard Dell^A, John Fosu-Nyarko^C, Y. Mukai^D, Paula Moolhuijzen^B, Matthew Bellgard^B, Jizeng Jia^E, Xiuying Kong^E, Catherine Feuillet^F, Frédéric Choulet^F, International Wheat Genome Sequencing Consortium^G and Rudi Appels^{B,H}

^ASchool of Biological Sciences and Biotechnology, Murdoch University, Perth, WA, Australia.

^BCentre for Comparative Genomics, Murdoch University, Perth, WA, Australia.

^CState Agricultural Biotechnology Centre, Murdoch University, Perth, WA, Australia.

^DLaboratory of Plant Molecular Genetics, Division of Natural Science, Osaka Kyoiku University, Kashiwara, Osaka 582-8582, Japan.

^EKey Laboratory of Crop Gene Resources and Germplasm Enhancement, Ministry of Agriculture, National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China.

^FINRA UMR 1095, Génétique Diversité et Ecophysiologie des Céréales, 63100 Clermont-Ferrand, France.

^GInternational Wheat Genome Sequencing Consortium, Executive Director Kellye Eversole.
 Email: eversole@eversoleassociates.com

^HCorresponding author. Email: rappels@ccg.murdoch.edu.au

Abstract. In wheat (*Triticum aestivum* L.) drought-induced pollen sterility is a major contributor to grain yield loss and is caused by the downregulation of the cell wall invertase gene IVR1. The IVR1 gene catalyses the irreversible hydrolysis of sucrose to glucose and fructose, the essential energy substrates which support pollen development. Downregulation of IVR1 in response to drought is isoform specific and shows variation in temporal and tissue-specific expression. IVR1 is now prompting interest as a candidate gene for molecular marker development to screen wheat germplasm for improved drought tolerance. The aim of this study was to define the family of IVR1 genes to enable: (1) individual isoforms to be assayed in gene expression studies; and (2) greater accuracy in IVR1 mapping to the wheat genetic map and drought tolerance QTL analysis. Using a cell wall invertase-specific motif as a probe, wheat genomics platforms were screened for the presence of unidentified IVR1 isoforms. Wheat genomics platforms screened included the IWGSC wheat survey sequence, the wheat D genome donor sequence from *Aegilops tauschii* Coss, and the CCG wheat chromosome 3B assembly: contig506. Chromosome-specific sequences homologous to the query motif were isolated and characterised. Sequence annotation results showed five previously unidentified IVR1 isoforms exist on multiple chromosome arms and on all three genomes (A, B and D): IVR1-3A, IVR1-4A, IVR1-5B, IVR1.2-3B and IVR1-5D. Including three previously characterised IVR1 isoforms (IVR1.1-1A, IVR1.2-1A and IVR1.1-3B), the total number of isoform gene family members is eight. The IVR1 isoforms contain two motifs common to cell wall invertase (NDPN and WECPDF) and a high degree of conservation in exon 4, suggesting conservation of functionality. Sequence divergence at a primary structure level in other regions of the gene was evident amongst the isoforms, which likely contributes to variation in gene regulation and expression in response to water deficit within this subfamily of IVR1 isoforms in wheat.

Additional keywords: gene family, grain yield, pollen sterility, sequencing, sucrose supply, water stress.

Received 6 April 2012, accepted 17 May 2012, published online 12 July 2012

Introduction

In wheat (*Triticum aestivum* L.), grain number per ear is the main determinant of yield and is significantly reduced by pollen sterility when drought is incurred during pollen meiosis. Expression of the cell wall invertase (CW-INV) gene IVR1 in

tapetal tissues that surround pollen during its formation is essential in maintaining pollen fertility. IVR1 functions to catalyse the irreversible hydrolysis of sucrose to glucose and fructose, the essential energy substrates that support pollen development. IVR1 is downregulated by water stress, which

leads to pollen being starved of its energy source, resulting in sterility, reduced grain number and ultimately reduced yield (Koonjul *et al.* 2005; Parish *et al.* 2012; Powell *et al.* 2012).

In view of this evidence and the increasing frequency of drought events in Australia over the last decade, there has been growing demand for wheat pre-breeding research to elucidate the molecular determinants of IVR1 gene regulation and expression under limited and non-limited water conditions (Ji *et al.* 2010; Powell *et al.* 2012). Underpinning the interest for improved understanding of sugar metabolism during head development is the objective of breeding wheat cultivars capable of maintaining pollen fertility and thus grain set under drought conditions. Achieving this objective would better enable the genetic potential of existing wheat germplasm, capable of high yields in non-limited water environments, to be fully exploited in regions where water is frequently limited during the period in the growing season that coincides with head development and grain fill (Reynolds *et al.* 2005; Ji *et al.* 2010).

Evidence supporting molecular approaches to breeding wheat for stable pollen fertility was seen in recent studies by Ji *et al.* (2010), where wheat varietal-specific expression of IVR1 under water deficit conditions was examined. Results from this study were consistent with similar studies by Koonjul *et al.* (2005) and demonstrated a positive correlation between IVR1 expression and pollen fertility. However, the degree of IVR1 downregulation in response to water deficit showed significant differences between varieties, in that certain varieties maintained gene expression levels under water stress, whereas others did not. This result is consistent with IVR1 being a strong candidate gene for drought tolerance quantitative trait loci (QTL) analysis (Ji *et al.* 2010).

Plant invertases (β -D-fructofuranosidase EC 3.2.1.26) constitute a family of enzymes that hydrolyse sucrose to glucose and fructose. In plants, three types of invertases have been identified: cell wall (CW-INV), vacuolar (V-INV) and cytoplasmic (C-INV). Invertases have roles in several plant physiological processes related to long-distance nutrient allocation as well as regulating developmental processes, hormone responses and biotic and abiotic interactions (Lalonde *et al.* 1997; Tymowska-Lalanne and Kreis 1998; Roitsch and Gonzalez 2004; Ruan *et al.* 2010). In order to provide a source of energy, sucrose is transported from leaves to the stem, reproductive structures and roots (Ruan *et al.* 2010). The sucrose molecule is cleaved at the α 1- β 2-glycosidic bond, either by invertases or sucrose synthase (SST). Invertases catalyse the irreversible hydrolysis of sucrose to glucose and fructose and SST catalyses reversible cleavage of sucrose to UDPglucose and fructose (Roitsch and Gonzalez 2004). Upon cleavage of sucrose, glucose and fructose are transported into the cells via hexose transporters located in the plasma membrane (Bush 1999).

The invertase gene family consists of many members within each of the three broad functional groups (i.e. CW-INV, V-INV and C-INV) (Tymowska-Lalanne and Kreis 1998). In the present study, the gene nomenclature used defines IVR1 isoforms as belonging to the cell wall invertase (CW-INV) gene subfamily in wheat. The term isoform is used to designate different proteins derived from related genes. A widely accepted explanation of multiple invertase isoform existence within each of the functional

groups is that this allows for greater flexibility in the control of sucrose metabolism, translocation and storage under different internal and external stimuli and at various developmental stages or tissues, or at the whole-plant level (Tymowska-Lalanne and Kreis 1998; Akhunov *et al.* 2003). Koonjul *et al.* (2005) found that individual IVR1 and V-INV isoforms were expressed in a highly specific manner in response to water availability, with respect to both tissue type and developmental stage. Similar results were recently reported by Crismani *et al.* (2011), who showed that in rice, closely related IVR1 isoforms within anther-specific transcripts exhibited clear temporal variation under non-limiting water conditions.

Studies of invertase transcription regulation conducted in several different crop species have found that transcription is influenced by a range of signals including; low oxygen, plant hormones, sugars and environmental and pathogenic cues (Long *et al.* 2002; Koch 2004; Proels and Roitsch 2009). In studies specifically examining the role of sucrose in regulating CW-INV expression, varied results were reported; some CW-INV appear to have expression reduced by sucrose accumulation, whereas others isoforms are upregulated (Kaufman *et al.* 1973; Sturn and Chrispeels 1990). Tymowska-Lalanne and Kreis (1998) suggest that variable CW-INV transcription under the control of sucrose signalling may be related to different isoforms being regulated in a highly specific manner, as a reflection of subtle but important differences in their function.

The aim of this paper was to define IVR1 isoforms at the genome level in wheat, in order to provide a basis for greater accuracy in discriminating between the isoforms that are expressed in developing wheat heads under normal and water deficit conditions. Evidence from this study suggests that characterisation of the IVR1 isoforms is essential to further gene expression studies as it will enable definitive assays of single isoforms. IVR1 isoform-specific assays will permit variation in spatial and temporal expression to be analysed in the context of improving understanding of IVR1 gene regulation and functionality.

Materials and methods

Wheat BAC sequencing

The assembly of BACs from flow-sorted chromosome 3B DNA was as described by Paux *et al.* (2008) and utilised a modified HICF. SNaPshot protocol was in summary, five restriction endonucleases were used to digest DNA samples. Anchoring of the BAC contigs to a genetic map was conducted using established SSR markers as well as probes for insertion-based polymorphisms, using a mapping population derived from a cross between wheat varieties Chinese Spring \times Renan and genetic lines carrying chromosomal deletions (Paux *et al.* 2008). For contig506 (ctg506), the minimum tiling path was established using the FPC software at a high stringency ($1e^{-75}$) and refined by analyses using LTC software (Frenkel *et al.* 2010). Sequencing was conducted at 5–8 \times coverage and the Solexa-Illumina sequencing was conducted at >300 \times coverage using 75 base pair reads. Additional 454 sequencing using 8-kb mate-pair libraries resolved ambiguities in the sequence assembly. The 480-kb portion of ctg506 (1.3 Mb) analysed in this paper has been submitted to GenBank (GenBank submission pending).

Identification of IVR1 isoforms and sequence annotation

Access to the genome survey sequences for the chromosome arms of wheat before publication (<http://urgi.versailles.inra.fr/Species/Wheat/Sequence-Repository>, accessed 10 May 2012) was provided by the International Wheat Genome Sequencing Consortium (IWGSC). To screen for the presence of previously unidentified IVR1 isoforms a BLASTN query was conducted across the entire wheat survey sequence using two invertase-specific motifs: (1) 5'CAGATCCAAATGGTACGTA'3, residues which span a β -fructosidase motif (NDPN) that is common to all invertase genes (i.e. CW-INV, V-INV, C-INV) and (2) 5'TGGGAATGCCCCGATTTC'3, residues which span a catalytic domain (WECPDF) that is present only in CW-INV isoforms (Tymowska-Lalanne and Kreis 1998; Koonjul *et al.* 2005). Determination of sequence-based IVR1 categorisation required the presence of both the NDPN and WECPDF motifs. The BLASTN analysis was conducted across all available chromosome arm datasets and yielded subject sequence contigs identified by accession numbers, which were then used to obtain the respective contigs from the specific chromosome arm on which they were located (Table 1). The contig sequences were imported into Geneious Pro (ver. 5.5.6, <http://www.genious.com/>) for alignment and annotation using the IVR1 full-length cDNA sequence, accession AF030420.1, as a reference.

Plant material and DNA extraction

Wheat (*Triticum aestivum* L. cv. Westonia and cv. Kauz) seed stocks were provided by the Department of Agriculture and Food Western Australia (DAFWA). Seed stocks were pre-germinated at 4°C for 14 days and then planted into free draining pots with a uniformly prepared soil medium consisting of two parts loam, one part river sand and 3 g of basal fertiliser (mg kg⁻¹: 1220 N, 368 P, 819 K) and grown in controlled glasshouse conditions with ambient temperature during the plant growth period of 9°C (night) and 18°C (day). Pots were watered three times daily using an irrigation mat system to maintain average soil volumetric water content of 60%. Leaf tissue samples were taken for DNA extraction at the early tillering phase (Zadok's growth scale stage Z20), frozen in liquid nitrogen and DNA extraction was performed using a generic phenol-chloroform extraction method. The extracted DNA was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher, Scoresby, Vic., Australia).

IVR1 sequencing and analysis

The genomic sequence ctg506 (GenBank submission pending), as defined in this manuscript, was located on chromosome 3B of Chinese Spring and contained IVR1.1–3B. This isoform was used as the reference sequence to design primers (each

Table 1. IVR1 isoform accessions identified in the IWGSC wheat survey sequence using the β -fructosidase (NDPN) and catalytic domain (WECPDF) motifs and isoform homology to the full length IVR1 cDNA (AF030420.1)

Note: tabulated percentage values for isoform homology to the reference sequence AF030420.1 were deduced after annotation of isoform intron-exon boundaries and based only on homology of coding region sequence

IWGSC survey sequence contig ID	Chromosome location	Contig length (base pairs)	Proposed IVR1 nomenclature	Query motif protein ID	Isoform homology to AF030420.1 (%)
10503511	3B	4630	IVR1–3B	WECPDF, NDPN	82.6
10595081	3B	11 558	IVR1–3B	WECPDF, NDPN	84.4
529524	3A	2756	IVR1–3A	NDPN	82.7
7174266	4A	14 238	IVR1–4A	WECPDF, NDPN	83.9
10924301	5B	25 605	IVR1–5B	WECPDF, NDPN	92.9
4492773	5D	13 924	IVR1–5D	WECPDF, NDPN	95.9

Table 2. TD-PCR primer used for amplification of overlapping fragments spanning IVR1.1–3B

Primer pair (name)	Primer nucleotide sequence	Fragment length (base pairs)	TD-PCR annealing temperatures (°C)
IVR1–1 Forward	GTCTCTATCGTTCGATCGAGCTA	572	58–53
IVR1–4 Reverse	TGTAGTACATTGGCCCTGCATC		
IVR1–4 Forward	GATGCAGGGCCAATGTACTACA	577	58–53
IVR1–6 Reverse	CTCCAGTGTCTGCACATATATC		
IVR1–6 Forward	GATATATGTGCAGACACTGGAG	1050	60–55
IVR1–7 Reverse	GTGTAAGCTAGTGACCTGGATC		
IVR1–7 Forward	GATCCAGGTCAGTACTAGCTTACAC	513	60–55
IVR1–9 Reverse	ATATGTTACCTTGTGGGTCA		
IVR1–9 Forward	TGACCCAACAAGTAACATAT	570	58–53
IVR1–13 Reverse	CTACTGCTCCAGCAGCTTGTTCATC		

spanning intron–exon junctions) to amplify a total of five overlapping fragments spanning the entire length of IVR1.1–3B (Table 2). Primers were designed with Primer3 ver. 0.4.0 (<http://primer3.sourceforge.net/>, accessed 5 March 2011). Each PCR reaction was 20 µl, containing: 200 ng of template DNA, 10 pmol each of a forward and reverse primer pair (specific to an individual genomic fragment), 200 µM dNTP, 2.5 mM MgCl₂, 1× reaction buffer, 6 µl of 1% cresol red and 2.5 units of Taq polymerase (Fisher Biotech, Perth, WA, Australia). Touch down PCR (TD-PCR) conditions were: incubation for 3 min at 94°C, followed by 35 cycles of 1 min denaturation (94°C), 30 s annealing (decreasing by 1°C per cycle until the primer-specific annealing temperature was reached, as in Table 2) and 30 s extension (72°C). An additional 10 minutes extension at 72°C was also performed.

The TD-PCR products were separated on a 2% agarose gel and DNA purified using a Wizard SV Gel and PCR clean up system (Promega Corp., Annandale, NSW, Australia). The purified genomic DNA fragments were cloned by ligation to the pGEM-T Easy cloning vector (Promega Corp.). Both strands of the cloned IVR1 isoform fragments were sequenced with the Sp6 and T7 promoter using BigDye3.1 (BDV3.1 plus dgt) termination chemistry and were run on an ABI Prism XL 3730 sequencer (Applied Biosystems, Foster City, CA, USA).

Results

The genome level structure of wheat IVR1 isoforms on chromosome 3B

Our sequence assembly of a BAC contig from chromosome 3B (ctg506) provided the genome sequence of two IVR1 isoforms, IVR1.1–3B and IVR1.2–3B, in addition to the two previously characterised genes on chromosome 1A (Francki *et al.* 2006).

The ctg506 is between the *Sr2* rust resistance and fusarium head blight resistance (*Fhb1*) loci and includes the molecular marker STS104 (genetic map in Fig. 1). The assembly initially comprised 12 BAC clones, with estimated insert sizes of 70–180 kb, compiled into a minimum tiling path. The linear sequence was established using a combination of Sanger and Solexa-Illumina sequencing technologies. The initial assembly had 21 contigs. Subsequent analyses of the FPC based assemblies and additional 454 sequencing of 8-kb mate pair libraries refined the assembly.

Figure 2 shows the details for 480 kb of ctg506 to indicate the overall structure of this genome sequence and the location of the IVR1 isoforms. The highly repetitive region upstream from IVR1.1–3B comprises members of the Sc119.2 repetitive DNA family that are widely used as chromosome markers (left-hand panel Fig. 1; McIntyre *et al.* 1990) and are distributed in the ctg506 region of chromosome 3B at a relatively low representation compared with other sites in the genome (note the more intense band slightly distal to the region of interest). A SNP (BF293133) was found further downstream on the ctg506 genome segment shown in Fig. 2 and has been assigned a genetic location on chromosome 3B in the position expected for ctg506 (Zhang *et al.* 2008a), providing additional confirmation of the chromosomal location of the sequences analysed.

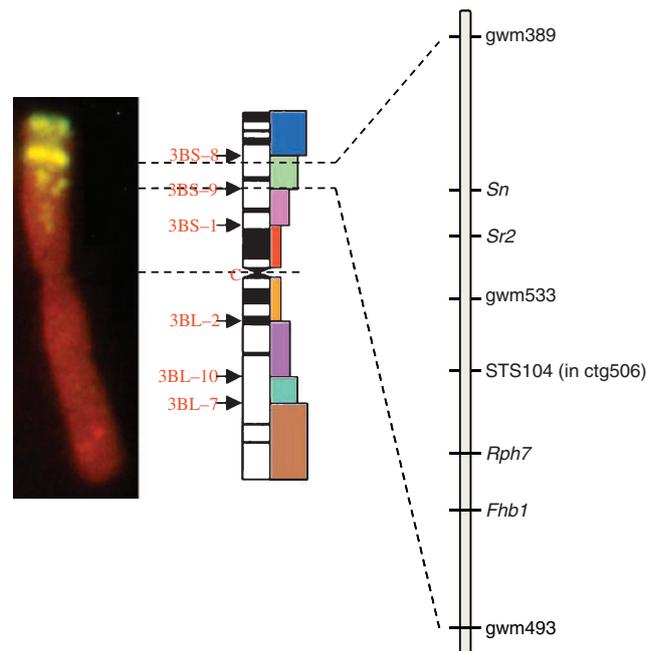


Fig. 1. A summary molecular-genetic map of the region of chromosome 3B in which the genome segment characterised as ctg506 is located. The left panel indicates the *in situ* distribution of a repetitive sequence family found in ctg506, whereas the middle panel indicates the deletion bin location (modified from Sourdille *et al.* 2004) of the genetic map shown in the right-most panel (modified from the 3B consensus map, <http://ccg.murdoch.edu.au/index.php/CMap>, accessed 10 October 2011). The molecular marker STS104 defines the location of ctg506.

Annotation of the 480-kb genome sequence was conducted using GenScan (<http://genes.mit.edu/GENSCAN.html>, 9 May 2012) as well as Augustus and EuGene (in TriAnnot <http://urgi.versailles.inra.fr/Species/Wheat/Triannot-Pipeline/>, accessed 9 May 2012, Leroy 2012). The results consistently showed two IVR1 isoforms at coordinates 300 kb and 454 kb in the genome sequence. In Fig. 3, the primary structure for the IVR1 isoforms is presented based on manual refinement of the annotations in the Geneious program, using AF030420.1 as a reference sequence to identify intron–exon boundaries. The IVR1.1–3B isoform has seven exons and is clearly closely related to the reference cell wall invertase AF030420.1, including a characteristic small exon of three amino acids (DPN, part of the NDPN motif) and a conserved catalytic domain in exon 4 (WECPDF). The IVR1.2–3B isoform also has the DPN and WECPDF motifs but has a nine exon structure and greater sequence divergence in other parts of the coding regions (as compared with the reference AF030420.1), particularly in exon 3.

The IVR1 gene family in wheat

Sequence analysis of the cloned genomic DNA TD-PCR products amplified from exon 3 and exon 7, in both Westonia and Kauz, provided the initial evidence for the existence of additional IVR1 isoforms in the wheat genome that had not yet been identified. The primers used to assay the fragments in question spanned exon-

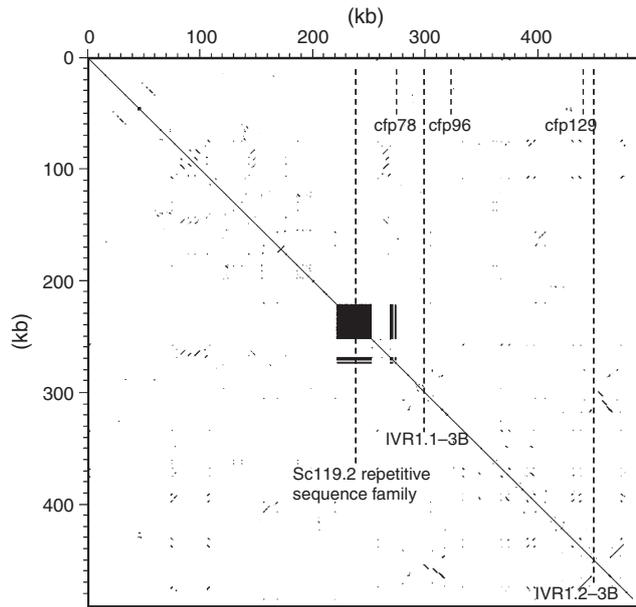


Fig. 2. A dot plot of the 480-kb sequence defined within ctg506 against itself indicating that it is relatively free of major repeat sequences except for the cluster of sequences at 220–260 kb. The repetitive sequence located at 220–260 kb comprises variants of the 120 bp, pSc119.2, sequence which is well known as a molecular marker (McIntyre *et al.* 1990) and is particularly prominent in the terminal region of chromosome 3B (left panel in Fig. 1).

intron boundaries and thus were expected to have high specificity to assay IVR1.1–3B, given they were designed using the Chinese Spring chromosome 3B sequence. Three homologous groupings were apparent in both the Westonia and Kaus sequences of exons 3 and 7, one group of which was homologous to IVR1.1–3B (ctg506) but the remaining two did not show homology to any of the available annotated IVR1 sequences for wheat, including: IVR1.1–3B (ctg506), IVR1.1–1A (GenBank submission pending), IVR1.2–1A (GenBank submission pending) and the full length IVR1 cDNA AF030420.1. This evidence suggested at least two additional IVR1 isoforms existed and that there was likely to be an IVR1 subfamily in wheat.

The IWGSC survey sequences for the chromosome arms of wheat (<http://urgi.versailles.inra.fr/Species/Wheat/>) enabled four

previously unidentified isoforms of the IVR1 subfamily to be identified using the invertase-specific motif NDPN and the CW-INV-specific motif WECPDF, as probes. Using these probes, to screen the wheat D genome donor sequence from *Triticum tauschii* (J. Jia, unpubl. data) and the ctg506 chromosome 3B assembly, two additional uncharacterised IVR1 isoforms were also identified. Figure 4 shows the isoform gene structure determined using the full-length IVR1 cDNA AF030420.1 as a reference. The characterised and annotated IVR1 isoforms show the conserved motifs DPN (exon 2) and WECPDF (exon 4). Conservation around the WECPDF motif spans the entire length of exon 4 (see protein sequence of exon 4 in the lower panel of Fig. 4). Divergence in intron-exon structure is apparent, with five members for the IVR1 family having eight exons (IVR1.1–1A, IVR1.2–1A, IVR1–4A, IVR1–5B, IVR1–5D), two members having seven exons (IVR1.1–3B, IVR1–3A) and IVR1.2–3B comprising nine exons.

The evidence presented in Fig. 4 was based on the screening of individual chromosome arm libraries, as distinct from counting sequence variants. These data suggest that at least eight IVR1 isoforms exist in wheat, including: IVR1.1–1A, IVR1.2–1A, IVR1–3A, IVR1.1–3B, IVR1.2–3B, IVR1–4A, IVR1–5B and IVR1–5D. From these results we propose that using their chromosomal locations is a useful way to identify the isoforms. In Fig. 5 the nucleotide sequence of exon 4 (containing the WECPDF motif) is shown and illustrates that IVR1 isoforms are identifiable by their distributions of SNPs, enabling clear differentiation between the isoforms. We note that defined segments of IVR1.1–1A and IVR1.2–1A were assigned to 1A using genetic stocks (Francki *et al.* 2006), but our survey sequence analysis did not support this chromosomal location. Thus, it is possible that fewer IVR1 genes exist and that IVR1.1–1A is actually IVR1–4A and IVR1.2–1A is IVR1–5B. This needs further confirmation.

Screening of the IWGSC survey sequence, the ctg506 chromosome 3B assembly and the D genome sequence with PCR primers previously used to study the expression of IVR1 in wheat (Koonjul *et al.* 2005; Ji *et al.* 2010), indicated that several IVR1 isoforms on different chromosomes were assayed by these primers. As shown in Fig. 6a, the primers by Ji *et al.* (2010) are predicted to assay IVR1–1A, IVR1–4A and IVR1–5B. The primers used by Koonjul *et al.* (2005) would similarly assay IVR1.1–3B and IVR1–5B (Fig. 6b).

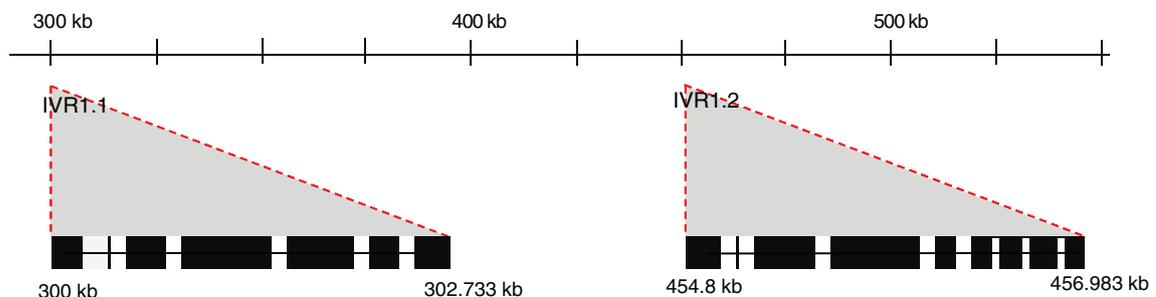
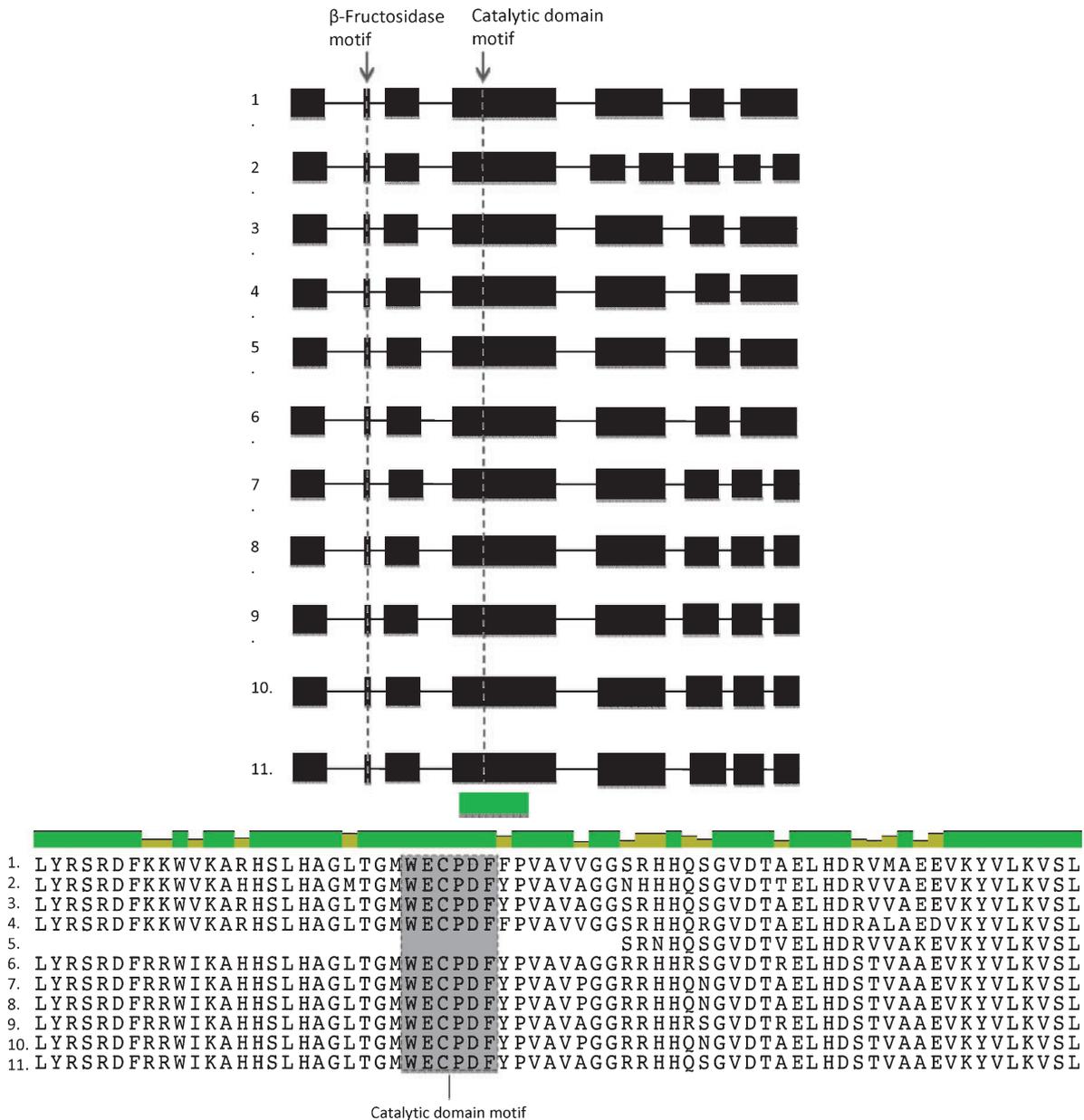
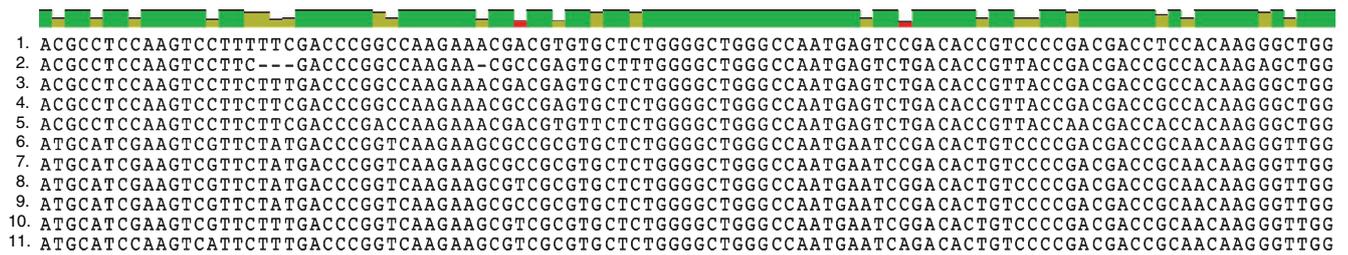


Fig. 3. Detailed genome level structural arrangement of the IVR1 isoforms located on ctg506 from chromosome 3B. A seven exon and nine exon structure is observed in IVR1.1–3B (left) and IVR1.2–3B (right), respectively, and divergence between the isoforms exists at an amino acid level (data not shown) particularly in exon 3. The black boxes indicate exons and ascending order from left to right corresponds to exon numbers 1 to 9.



Legend

- 1.IVR1.1-3B ctg 506 (GenBank accession number pending)
- 2.IVR1.2-3B ctg 506 (GenBank accession number pending)
- 3.IVR1-3B Wheat survey sequence accession: 10595081
- 4.IVR1-3B Wheat survey sequence accession: 10503511
- 5.IVR1-3A Wheat survey sequence accession: 529524
- 6.IVR1-1A GenBank accession: AF030420.1
- 7.IVR1.1-1A (GenBank accession number pending)
- 8.IVR1.2-1A (GenBank accession number pending)
- 9.IVR1-5B Wheat survey sequence accession: 10924301
- 10.IVR1-4A Wheat survey sequence accession: 7174266
- 11.IVR1-5D *Ae tauschii* scaffold 60450



Legend

1. IVR1.1-3B ctg 506
2. IVR1.2-3B ctg 506
3. IVR1-3B Wheat survey sequence accession: 10595081
4. IVR1-3B Wheat survey sequence accession: 10503511
5. IVR1-3A Wheat survey sequence accession: 529524
6. IVR1-1A NCBI accession: AF030420.1
7. IVR1.1-1A GenBank submission ID:1527570
8. IVR1.2-1A GenBank submission ID pending
9. IVR1-5B Wheat survey sequence accession: 10924301
10. IVR1-4A Wheat survey sequence accession: 7174266
11. IVR1-5D *Ae tauschii* scaffold 60450

Fig. 5. Comparison of wheat IVR1 isoform nucleotide sequence in exon 4. In the multiple sequence alignment identical, conserved and analogous regions between the various IVR1 isoforms are demarcated by the colour of the line seen above the alignment; where the dark green line represents homology, light olive green represents a high degree of conservation and red represents low conservation.

Comparative genomics of cell wall invertases in wheat and rice

Within the rice genome sequence eight cell wall invertase isoforms has been identified (Cho *et al.* 2005; Sturm and Tang 1999). The extensive phenotypic analyses associated with genes and regions of rice provide a reference for studies in wheat. Phylogenetic analysis indicated that among the cell wall invertase isoforms in rice, the LOC_Os04g33720 sequence on chromosome 4 was most closely related to the AF030420.1 sequence. The LOC_Os04g33720 gene is preferentially transcribed in anther tissues although there is expression at lower levels (i.e. 30% or less) in other tissues (Sharma *et al.* 2012). It was of interest to align the 5' upstream regions from the LOC_Os04g33720 gene of rice (−507 to 0) with available upstream regions from the wheat IVR1 isoforms, IVR1.1–3B (−1069 to −562) and IVR1-5D (−1777 to −1356), to identify conserved motifs.

The analysis shown in Fig. 7 provides evidence that motifs conserved between rice and wheat are located upstream of IVR1.1–3B, indicating that these may be significant for anther-specific expression of IVR1 isoforms. Similar results

were obtained for IVR1.2–3B and IVR1-5D upstream regions aligned with the respective rice region.

Discussion

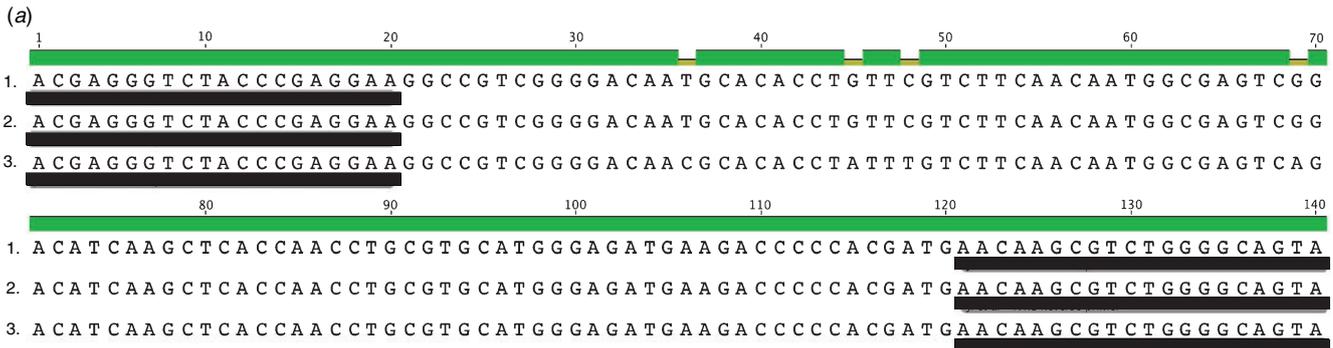
The data in the present paper have further defined the IVR1 subfamily in wheat at the genome level. Complexity at this level was suggested by Koonjul *et al.* (2005) and the genome-level analysis of IVR1 has indicated that the primers used to assay transcripts in drought experiments (Koonjul *et al.* 2005; Ji *et al.* 2010) actually assay several isoforms from different chromosomal locations. The significance of these findings is that transcription studies need to define the chromosomal location of the isoform being assayed so that suitable molecular markers can be transferred to wheat breeding.

As in cotton (Udall *et al.* 2007; Rapp *et al.* 2010), wheat shows homeologous gene silencing (Mochida *et al.* 2004; Bottley and Koebner 2008; Zhang *et al.* 2008b). In an analysis of a panel of 16 wheat varieties, Bottley and Koebner (2008) identified homoeolog non-expression for 15 single-copy genes. The detailed expression profiles of eight genes indicated that only two varieties of the 16 shared the same pattern of silencing and

Fig. 4. Detailed intron-exon structural arrangement of IVR1 isoforms and a portion of the deduced amino acid sequence from exon 4. In the upper panel of the figure exons are depicted by black boxes, with exon number ascending from left to right (corresponding to exons 1–9). Isoform intron-exon structure is highly conserved in the region extending from exons 1–5, as are the DPN and WECPDF motifs. Divergence is apparent at the 3' end of the gene where the isoforms have either a total number of seven exons (sequence #1,3,4,5), eight exons (sequence #7,8,9,10,11), or nine exons (sequence #2). The bar underneath exon 4 indicates the location of the sequences shown in the lower panel. The lower panel of the figure shows the high degree of conservation that exists between the IVR1 isoforms at an amino acid level, for the region spanning exon 4. In the protein alignment identical, conserved and analogous regions between the various IVR1 isoforms are demarcated by the colour of the line seen above the alignment; where the dark green line represents homology and light olive green line represents a high degree of conservation. Note: the IWGSC wheat survey sequence contig (529 524), from which IVR1–3A was annotated, contained the amino acid sequence spanning the DPN motif in exon two (not shown) but was missing the portion of the sequence spanning the first half of exon 4 (where the WECPDF motif is usually located). The protein translation of the available portion of contig 529 524 showed a high degree of homology with the annotation IVR1 reference sequence AF030420.1, grounds which in parallel with the conservation of the DPN motif in contig 529524, indicate it is likely to be an IVR1 isoform.

thus, the homoeologous silencing observed in wheat is likely to be significant in generating phenotypic variation. In this context it is evident that in relating variation in expression of IVR1 isoforms in wheat to drought tolerance it is important to define the chromosomal origin of the IVR isoform being studied in transcriptional analyses.

There is growing evidence to suggest that sucrose conversion by cell wall invertases in the apoplastic space is one of the key rate limiting steps in carbohydrate metabolism in reproductive structures and supplying photosynthetic assimilates for storage (Eschrich 1980). The basis of this dependency is that developing pollen microspores are symplastically isolated and thus, must



Legend

- 1. IVR1-4A
- 2. IVR1-1A
- 3. IVR1-5B



Legend

- 1. IVR1-5B
- 2. IVR1-3B

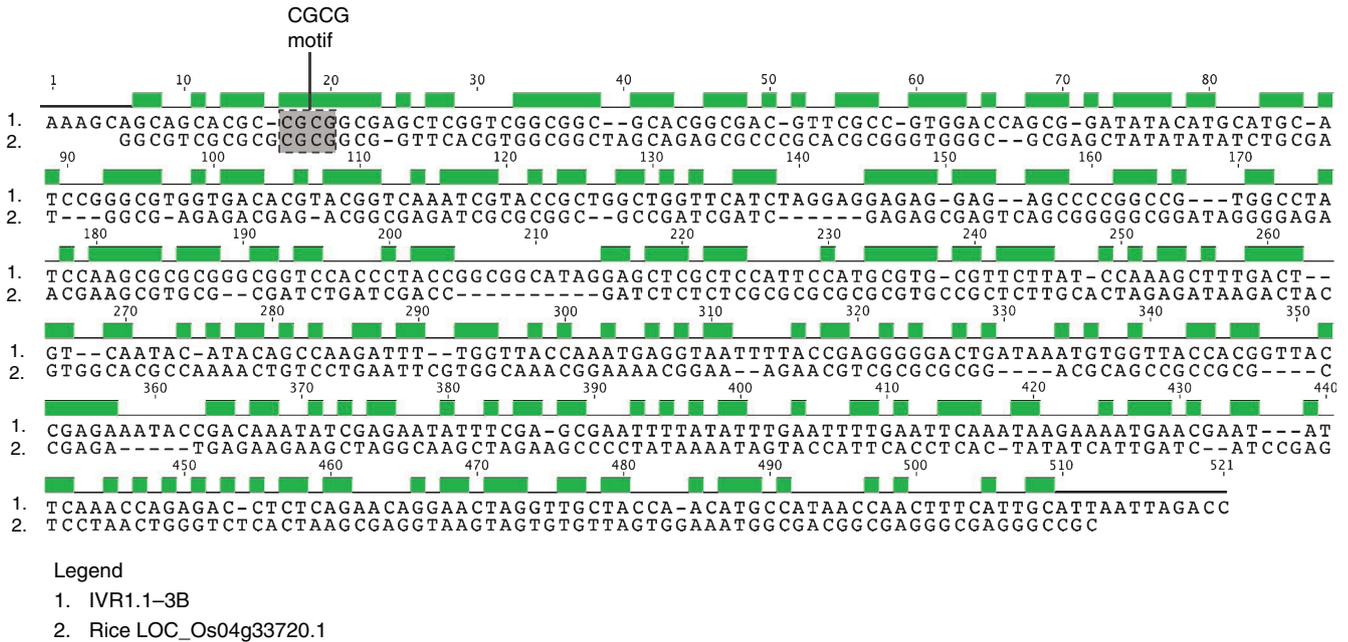


Fig. 7. Comparison of amino acid sequences in the 5' region upstream of IVR1.1-3B in wheat and the most closely related CW-INV in rice LOC_Os04g33720.1 (OsCIN3, Cho *et al.* 2005), located on chromosome 4. Scrutiny of identical regions within the alignment (demarcated by a green line above the homologous regions) show a conserved motif (i.e. the CGCG box highlighted in grey) located in the alignment from residues 17-20. An important pollen-specific promoter motif (AGAAA) is also located upstream from the wheat IVR1.1-3B isoform from residues 355-359 (ProFITS site: <http://bioinfo.cau.edu.cn/ProFITS/>, accessed 27 April 2012).

receive carbohydrates via apoplastic transport in which IVR1 isoforms play a critical role (Roitsch 1999; Sturm and Tang 1999). Impeded apoplastic transport can be attributed to the downregulation of IVR1 in anthers, which results in pollen being starved of the energy substrate, as there is no alternate pathway for the delivery of sucrose. The breakthrough in establishing the central role of CW-INV's in phloem uploading of sugars was in the study of the Nin88 CW-INV gene in tobacco (Goetz *et al.* 2001), which was specifically expressed in the tapetum and pollen. In their study, Goetz *et al.* (2001) demonstrated that tissue-specific suppression of the Nin-88 gene, under the control of its own promoter sequence, resulted in defects in the early stages of pollen development and thus sterile pollen.

In studies examining the selective downregulation of IVR1 isoforms under water deficit conditions in wheat, Koonjul *et al.* (2005) reported that gene expression was highly specific to individual isoforms both spatially and temporally. Similar

conclusions were drawn by Godt and Roitsch (1997) and Sturm and Tang (1999). In the Koonjul *et al.* (2005) study, water deficit applied specifically during the period just before and following pollen meiosis showed immediate and significant downregulation of one IVR1 isoform (IVR1S) but no decline in another (IVR1L). Conversely, in the control plants (where watering was maintained) the IVR1S isoform also declined over the duration of pollen development whereas IVR1L increased. According to the results shown in Fig. 6b, which highlights the PCR products assayed by the Koonjul *et al.* (2005) IVR1 primers, we propose that the two IVR1 isoforms assayed in this gene expression analysis were located on chromosome 3 (IVR1.1-3B) and chromosome 5 (IVR1-5B). However, determination of which isoform was located on chromosomes 3 and 5 was not possible without access to isoform-specific PCR product sequences.

The Koonjul *et al.* (2005) invertase gene expression analysis also highlighted that two genes (IVR3 and IVR5), closely related

Fig. 6. (a) Comparison of nucleotide sequence of IVR1 PCR products in exon 7, assayed by primers utilised by Ji *et al.* (2010) for observing IVR1 gene expression under water deficit conditions. The Ji *et al.* (2010) primer sequences are demarcated in the alignment by a solid black line under the nucleic acid sequence. The forward primer is positioned in the alignment at coordinates 1 to 20 (5'-3') and the reverse primer is positioned at the co-ordinates 121-140 (5'-3'). The degree of homology amongst the IVR1 isoforms is demarcated by the colour of the line seen above the alignment, where the dark green line represents homology and light olive green represents a high degree of conservation. Note: the Ji *et al.* (2010) primer sequence is homologous in the IVR1 isoforms on chromosome 4A, 1A and 5B and therefore lacks specificity to assay a chromosome-specific isoform of the gene. (b) Comparison of nucleotide sequences of IVR1 PCR products in exon 4, assayed by primers utilised by Koonjul *et al.* (2005) for observing IVR1 gene expression under water deficit conditions. The Koonjul *et al.* (2005) primer sequences are demarcated in the alignment by a solid black line above the nucleic acid sequence. The forward primer is positioned in the alignment at coordinates 1-24 (5'-3') and the reverse primer is positioned at the coordinates 703-726 (5'-3'). The degree of homology amongst the IVR1 isoforms is demarcated by the colour of the line seen above the alignment, where the dark green line represents homology and light olive green represents a high degree of conservation. Note: the Koonjul *et al.* (2005) primer sequence is homologous in the IVR1 isoforms on chromosomes 5B and 3B and therefore, lacks specificity to assay a chromosome-specific isoform of the gene.

to the IVR1 family, did not show decline in expression under drought stress. The IVR3 and IVR5 genes were reported as being members of the CW-INV and V-INV families, respectively. In the work related to the current study, detailed examination of the IVR3 (accession AF030421) and IVR5 (accession AF069309) gene sequences confirmed that IVR5 belongs to the V-INV family but that IVR3 is not a CW-INV or a V-INV isoform. Our results show that the IVR3 sequence is: (i) highly divergent from the IVR1 family both at a primary sequence level and exon structure; (ii) located on chromosome 6A in wheat (IWGSC Survey Sequence contig-4429358); and (iii) a member of the fructan 1-exohydrolase gene family, with its closest relative being the 1-FEHw1 (accession AJ516025) gene in wheat.

The studies in the present paper provide greater resolution in defining IVR1 isoforms for expression studies. The comparison in Fig. 7 showed that the 5'-upstream region for IVR1.1-3B (ctg506: 298 676 kb to 300 000 kb) had significant similarity to an equivalent region for LOC_Os04.g33720 (ch4: 20 411 920 kb to 20 412 390 kb). The shared motifs included the CGCG box and CGACG motif. The CGCG box has been characterised in *Arabidopsis* (Yang and Poovaiah 2002) and often occurs as multiple elements as is the case for region shown in Fig. 7. The CGACG motif has been studied in the *cis*-control of amylase (Amy3D) expression in rice (Hwang *et al.* 1998). Within the IVR1.1-3B region the ACTCAT (related to osmoregulation), AGAAA (within a pollen-specific promoter) and CATGTG (related to drought response) motifs can be identified (ProFITS site: <http://bioinfo.cau.edu.cn/ProFITS/>, accessed 27 April 2012) and their presence is consistent with IVR1 expression being significant in maintaining a supply of hexose sugars to the developing pollen. We note that IVR1.1-3B is located in the region of chromosome 3BS where a QTL for variation in early head development, a stage that is particularly sensitive to the effects of water stress (Maccaferri *et al.* 2008).

Gene duplication is widely recognised as a platform from which gene families can evolve through mechanisms of chromosomal remodelling, locus deletion or point mutation (Wang *et al.* 2010). The data in this manuscript further support this hypothesis, with regards to the evolution of complex gene families in wheat (derived from a common ancestral gene), by providing evidence of the existence of a multi member IVR1 subfamily, which has a shared functionality in hydrolysing hexoses in reproductive structures. We propose the sequence divergence that exists amongst the eight IVR1 isoforms, located on four separate chromosome arms of the wheat genome, is likely to have causal links to subtle differences in isoform functionality and regulatory mechanisms, particularly in response to changing water availability conditions. It is plausible to suggest that sequence variation in the IVR1 subfamily provides wheat the flexibility to maintain carbohydrate supply to its developing reproductive structures, thus ensuring viable pollen development even in drought conditions. Gene expression studies remain the foundation methodology by which the relationship between the sequence variation of the IVR1 isoforms and their corresponding functional differences can be better understood. The identification of the chromosomal origins of the IVR1 isoforms undertaken in this study will now enable more accurate transcriptional analysis and

improved capacity to relate variation in IVR1 isoform expression to drought tolerance in wheat.

Acknowledgements

The authors wish to thank Dr Philippe Leroy for his help in annotation of the ctg506 wheat genome sequence.

References

- Akhunov ED, Goodyear AW, Geng S, Qi L, Echalié B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng J, Lapitan NL, Gonzalez-Hernandez JL, Anderson JA, Choi DW, Close TJ, Dillbirligi M, Gill KS, Walker-Simmons K, Steber C, McGuire PE, Qualset CO, Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Research* **13**, 753–763. doi:10.1101/gr.808603
- Bottley A, Koebner RM (2008) Variation for homoeologous gene silencing in hexaploid wheat. *The Plant Journal* **56**, 297–302. doi:10.1111/j.1365-313X.2008.03586.x
- Bush DR (1999) Sugar transporters in plant biology. *Current Opinion in Plant Biology* **2**, 187–191. doi:10.1016/S1369-5266(99)80034-X
- Cho JI, Lee SK, Ko S, Kim HK, Jun SH, Lee HY, Bhoo SH, Lee KW, An G, Hahn TR, Jeon JS (2005) Molecular cloning and expression analysis of the cell-wall invertase gene family in rice (*Oryza sativa* L.). *Plant Cell Reports* **24**, 225–236. doi:10.1007/s00299-004-0910-z
- Crismani W, Kapoor S, Able JA (2011) Comparative transcriptomics reveals 129 transcripts that are temporally regulated during anther development and meiotic progression in both bread wheat (*Triticum aestivum*) and rice (*Oryza sativa*). *International Journal of Plant Genomics* **2011**, 931 898. doi:10.1155/2011/931898
- Eschrich W (1980) Free space invertase, its possible role in phloem unloading. *Berichte Der Deutschen Botanischen Gesellschaft* **93**, 363–378.
- Francki MG, Walker E, Forster JW, Spangenberg G, Appels R (2006) Fructosyltransferase and invertase genes evolved by gene duplication and rearrangements: rice, perennial ryegrass, and wheat gene families. *Genome* **49**, 1081–1091. doi:10.1139/g06-066
- Frenkel Z, Paux E, Mester D, Feuillet C, Korol A (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics* **11**, 584. doi:10.1186/1471-2105-11-584
- Godt DE, Roitsch T (1997) Regulation and tissue-specific distribution of mRNAs for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. *Plant Physiology* **115**, 273–282. doi:10.1104/pp.115.1.273
- Goetz M, Godt DE, Guivarc'h A, Kahmann U, Chriqui D, Roitsch T (2001) Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6522–6527. doi:10.1073/pnas.091097998
- Hwang YS, Karrer EE, Thomas BR, Chen L, Rodriguez RL (1998) Three *cis*-elements required for rice alpha-amylase Amy3D expression during sugar starvation. *Plant Molecular Biology* **36**, 331–341. doi:10.1023/A:1005956104636
- Ji X, Shiran B, Wan J, Lewis DC, Jenkins CLD, Condon AG, Richards RA, Dolferus R (2010) Importance of pre-anthesis anther sink strength for maintenance of grain number during reproductive stage water stress in wheat. *Plant, Cell & Environment* **33**, 926–942. doi:10.1111/j.1365-3040.2010.02130.x
- Kaufman PB, Ghosheh NS, Lacroix JD, Soni SL, Ikuma H (1973) Regulation of invertase levels in *Avena* stem segments by gibberellic acid, sucrose, glucose, and fructose. *Plant Physiology* **52**, 221–228. doi:10.1104/pp.52.3.221

