
Article:
Harper, Andrea Louise orcid.org/0000-0003-3859-1152, Trick, Martin, He, Zhesi orcid.org/0000-0001-8335-9876 et al. (4 more authors) (2015) Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. Plant biotechnology journal. ISSN 1467-7644

https://doi.org/10.1111/pbi.12486

Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat

Journal:	*Plant Biotechnology Journal*
Manuscript ID:	PBI-00279-2015.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Harper, Andrea; University of York, Biology Trick, Martin; John Innes Centre, He, Zhesi; University of York, Biology Clissold, Leah; The Genome Analysis Centre, Fellgett, Alison; University of York, Biology Griffiths, Simon; John Innes Centre, Crop Genetics Bancroft, Ian; University of York, Biology
Keywords:	differential expression, bread wheat, genome dominance

Title: Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat

A. L. Harper[1], M. Trick[2], Z. He[1], L. Clissold[2*], A. Fellgett[1], S. Griffiths[2], I. Bancroft[1†]

Affiliations:

[1] Department of Biology, University of York, Heslington, York, YO10 5DD, UK

[2] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK

[†]Correspondence to: Prof. Ian Bancroft, Department of Biology, University of York, Heslington, York YO10 5DD +44 (0)1904 328778

Running title: Differential homoeologue expression in bread wheat

Keywords: differential homoeologue gene expression bread wheat

Sequences: Illumina sequence reads are available from SRA (accession number ERA283619) and transcript assemblies from: http://opendata.tgac.ac.uk/associative_transcriptomics/wheat/v1/Trinity_ABD_cured.fasta.gz.

Word count: 6672

Andrea L. Harper (andrea.harper@york.ac.uk)
Martin Trick (martin.trick@jic.ac.uk
Zhesi He (zhesi.he@york.ac.uk
Leah Clissold (leah.clissold@tgac.ac.uk)
Alison Fellgett (alison.fellgett@york.ac.uk)
Simon Griffiths (simon.griffiths@jic.ac.uk)
Ian Bancroft (ian.bancroft@york.ac.uk)

[*] Present address: The Genome Analysis Centre, Norwich Research Park, Norwich, NR4 7UH, UK.

Summary

Using a combination of *de novo* transcriptome assembly, a newly-developed 9495-marker transcriptome SNP genetic linkage map and comparative genomics approaches, we developed an ordered set of non-redundant transcripts for each of the sub-genomes of hexaploid wheat: A (47,160 unigenes), B (59,663 unigenes) and D (40,588 unigenes). We used these as reference sequences against which to map Illumina mRNA-seq reads derived from young leaf tissue. Transcript abundance was quantified for each unigene. Using a 3-way reciprocal BLAST approach, 15,527 triplet sets of homoeologues (one from each genome) were identified. Differential expression ($P<0.05$) was identified for 5,248 unigenes, with 2906 represented at greater abundance than their two homoeologues and 2342 represented at lower abundance than their two homoeologues. Analysis of gene ontology terms revealed no biases between homoeologues. There was no evidence of genome-wide dominance effects, rather the more highly transcribed individual genes were distributed throughout all three genomes. Transcriptome Display Tile Plot (TDTP), a visualization approach based on CMYK colourspace, was developed and used to assess the genome for regions of skewed homoeologue transcript abundance. Extensive striation was revealed, indicative of many small regions of genome dominance (transcripts of homoeologues from one genome more abundant than the others) and many larger regions of genome repression (transcripts of homoeologues from one genome less abundant than the others).

Introduction

The high incidence of polyploidy found in the history of many of our modern crops supports the theory that duplication of genetic material is a powerful facilitator of speciation via ecological diversification and adaptation. Polyploids also exhibit novel phenotypes and biosynthetic pathways (Kliebenstein, 2008; M. Schranz, 2011). Following genome doubling, many gene copies are rapidly lost (Scannell et al., 2007), although this process appears not to be completely random, with similar types of genes such as kinases and transcription factors (where stoichiometry of interactions might be important) more likely to be retained in duplicate (Blanc and Wolfe, 2004; Schnable et al., 2009; Seoighe and Gehring, 2004; Tian et al., 2005).

Several mechanisms have been proposed for enabling the retention of duplicate gene copies. Neofunctionalization (Hughes, 1994) occurs when one gene copy is diverted towards a new beneficial function. Alternatively, subfunctionalization (Force et al., 1999) may occur when normally multifunctional gene copies divide their functional workload. Another possibility is that one of the gene copies may become silenced and in this case also, not necessarily at random. The recent allopolyploid *Senecio cambrensis* exhibits different patterns of gene expression relative to the progenitor species, even when re-synthesised (Hegarty et al., 2006). Alternatively, the expression patterns of nascent wheat exhibited Genome Expression-Level Dominance (ELD), (Li et al., 2014) where the offspring resembled the expression patterns of one of the parents more than the other. These studies suggest that expression changes may occur rapidly and non-randomly after hybridization, and that they may have important functional significance.

Bread wheat (*Triticum aestivum*) is an interesting model for studying the evolution of gene expression across polyploid genome compartments. It is an allohexaploid comprising three genomes: A, B and D, derived from multiple hybridization events between its diploid and tetraploid ancestors (Chantret et al., 2005). The first hybridization is thought to have occurred around 5.5 million years ago, between the ancestors of the A and B genome lineages leading to the formation of the D genome lineage by homoploid hybrid speciation (*i.e.* hybridization without change in chromosome number). Less than 800,000 years ago, a hybridization between the A genome progenitor, *Triticum urartu,* and the B genome progenitor, thought to be a close relative of *Aegilops speltoides,* formed the AABB allotetraploid emmer wheat *T. turgidum*. Finally, less than 400,000 years ago, hybridization between emmer wheat and the D genome progenitor, *Aegilops tauschii* (Marcussen et al., 2014) formed the AABBDD allohexaploid bread wheat, *T. aestivum.*

We hypothesize that some regions of the hexaploid wheat genome may differ from the expected equal contributions for homoeologues to the transcriptome, and aimed to test this. In order to analyse the spacial patterns of gene expression for each gene across the three genomes of bread wheat, we developed two resources using methodology adapted from previous work in *Brassica napus* (Bancroft et al., 2011; Harper et al., 2012; Higgins et al., 2012; Trick et al., 2009b). Although a chromosome-based draft sequence for wheat is now available (International Wheat Genome Sequencing, 2014), the numerous small contigs produced by whole-genome and chromosome-shotgun

sequencing of large repetitive genomes, can lead to underrepresentation and mis-assembly of repetitive sequences, as well as collapsed assemblies of related genes both within and (especially) between genomes. On the other hand, transcriptome assemblies have the benefit of reducing the genome complexity by focusing on only the expressed genes and the independent assembly of transcripts from each of the three progenitor species ensures the correct allocation of gene assembly to genome. In addition, *de novo* assembled transcripts are not prone to modelling errors of gene prediction programs and matching of tissue and developmental stage of mRNA used for construction of the reference sequence with that used for the analysis ensures its suitability for the experimental design. For these reasons, we decided to develop a new transcriptome reference capable of discriminating between homoeologous transcripts, created by *de novo* assembly of transcripts from the diploid progenitor species, which were then improved using the tetraploid progenitor to “cure” the B genome reference. We then developed a set of pseudomolecules to infer the order of the reference genes within the genome by developing a high density gene-based genetic linkage map and exploiting the conserved synteny between wheat genomes and that of *Brachypodium distachyon*, which has been sequenced to a very high standard and shared a common ancestor with wheat 32-39 million years ago (International Brachypodium, 2010). The high degree of conserved synteny between the genomes of grasses (Moore et al., 1995) and the availability of a high quality genome sequence provides the opportunity to use comparative genomics to infer gene order based on that of orthologues in *B. distachyon*. Indeed, such cross-species inference has already been used in estimation of genome organization in barley (International Barley Genome Sequencing et al., 2012;

Mayer et al., 2011). By focusing on the analysis of transcribed sequences (mRNA-seq) we aimed to bypass the difficulties associated with genome sequence assembly in bread wheat, such as the large proportion of repetitive sequences between genes (Choulet et al., 2010) and the hexaploid nature of the species.

Results

Reference assembly

The underpinning resource for robust SNP-calling and transcript quantification by RNA-Seq is a reference sequence comprising non-redundant assemblies of transcribed sequences, *i.e.* unigenes. The first step was to assemble RNA-Seq reads into unigenes representing each of the three progenitor genomes of bread wheat. We generated Illumina 100-base paired-end reads from mRNA isolated from young leaf tissue of *Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii* (representing the A, B and D genomes, respectively) and assembled sets of unigenes using the Trinity package (Grabherr et al., 2011). As the B genome in hexaploid wheat is much more closely related to that in tetraploid wheat, *Triticum turgidum* ssp. *dicoccoides*, the B genome unigenes were "cured" (Higgins et al., 2012) to more closely represent those of bread wheat, using Illumina RNA-Seq data from that tetraploid. The result was a reference transcriptome sequence for hexaploid wheat comprising 105,069, 132,363 and 85,296 unigenes representing its A, B and D genomes, respectively. For each unigene, the gene model with the greatest sequence similarity was identified in each of

Brachypodium distachyon (Brachypodium), rice, Sorghum and Arabidopsis, and annotation extracted for the Brachypodium, rice and Arabidopsis models, as listed in Supplementary Dataset 1.

Linkage map construction

To support the *in silico* rearrangement of the Brachypodium genome to represent that of hexaploid wheat, we first constructed a high density unigene-based SNP linkage map. This involved generating Illumina reads from mRNA isolated from young leaf tissue of 47 lines of a single seed descent linkage mapping population (http://www.wgin.org.uk) and the parents of the population (cultivars Paragon and Chinese Spring), then using them to simultaneously identify and score polymorphisms using the unigenes as reference sequences (Trick et al., 2009a; Trick et al., 2009b). The scoring strings were used, in conjunction with the hypothetical fine-scale order of the unigenes in which the polymorphisms were identified (based on sequence similarity to Brachypodium gene models), to construct a linkage map of hexaploid wheat based on recombination bins. The linkage map, for which details are provided in Supplementary Dataset 2, contained 9,495 markers, with fewer in the less polymorphic D genome (981) compared with the more polymorphic A and B genomes (4,021 and 4,493 respectively). The map established 624 consensus recombination bins, as shown in Supplementary Dataset 3, and defining in detail collinearity with the Brachypodium genome, as shown in Supplementary Dataset 4.

Brachypodium-based pseudomolecule assembly

Based on the ranges of Brachypodium gene models identified in each of 56 collinearity blocks, the Brachypodium genome sequence was rearranged *in silico* to establish a new order of Brachypodium genome sequences, to which we refer hereafter as pseudomolecules, representative of the organization of orthologous sequences in the wheat genomes, according to the design specified in Supplementary Dataset 5. Finally, the positions of the assembled A, B and D unigenes on these pseudomolecules were established by sequence similarity. The resulting resource comprised 175,103 hypothetically ordered unigenes (58,320, 67,829 and 48,954 for the A, B and D genomes, respectively), as shown in Supplementary Dataset 6. As some of these represented alternative splice forms, redundancy was reduced further by selecting the longest unigene where multiple unigenes mapped to the same location in the genome, resulting in a final transcriptome reference sequence for mapping sequence reads of 147,411 unigenes (47,160 for the A genome, 59,663 for the B genome and 40,588 for the D genome). We compared, as an example, the order of 17,315 unigenes along the B genome with the order in the Brachypodium genome of gene models with the greatest sequence similarity. The result provides a detailed gene-based analysis of collinearity of the genome representations (Figure 1). We also compared, as an example, the order of 8,187 unigenes along the B genome with the order inferred in the barley (*Hordeum vulgare*) genome (Mayer et al., 2012) for orthologues of the corresponding Brachypodium gene models. The result showed good overall collinearity between the wheat and barley genomes, but in detail the genomes are differentiated by numerous rearrangements (Figure 2). For completeness, we also compared the order of 21,483

unigenes against the wheat chromosome 3B annotated pseudomolecule (Figure S1), and the V5 wheat genome zipper (Figure S2), and the (both downloadable from http://wheat-urgi.versailles.inra.fr), which also showed extensive, but imperfect, collinearity.

Visualization of regional unbalanced expression

In order to assess genome expression level dominance patterns, 15,527 triplets of homoeologues were identified via a 3-way reciprocal BLASTn analysis (threshold e-value 1E-30). Although the reciprocal blast parameters are particularly stringent to exclude as many spurious homoeologous triplets as possible from the analysis, this is the largest panel of candidate homoeologous genes identified in hexaploid wheat to date.

Using the 147,411-unigene transcriptome reference sequence, mRNA-Seq reads from juvenile leaves of 54 bread wheat accessions were mapped. Transcript abundance was quantified and normalized as reads per kb per million aligned reads (RPKM) and the values extracted for the 15,527 homoeologue triplets.

We developed a visualisation method based on assigning quantitative transcript abundance (RPKM) for each member of the 15,527 homoeologue triplets a value in CMYK color space where the contributions from the A, B and D genome copies were coded to cyan, magenta and yellow channels, respectively, and displaying the results using tile plots. We termed the method Transcriptome Display Tile Plot (TDTP). As

controls, and to provide a visual key, mRNAseq reads (down-sampled to 33 Million reads) from juvenile leaves of each of the diploid wheat species were mapped onto the unigene reference sequence (incorporating all three genomes) and the relative expression across the homoeologous triplets visualised, as shown in Figures 3 and S3. Although there is slight color distortion from cross-mapping of reads to alternate homoeologues of some triplets (the expected consequence of stretches of identical sequences of ~100 bases shared by homoeologous genes), the signals are predominantly as expected: cyan for *T. urartu*, magenta for *A. speltoides* and yellow for *A. tauschii*. Using the mRNAseq reads from combinations of these diploids to simulate the visualization of polyploids, these *in silico* combinations generated, with some color distortion arising from cross-mapping, the expected predominant signals: blue for A plus B genomes, green for A plus D genomes, red for B plus D genomes and grey for A plus B plus D genomes. As a final control, mRNA-Seq reads for *T. turgidum* (the AABB allotetraploid) were mapped, producing the expected predominantly blue signal.

Comparison to genome sequence data

The 15,527 triplet sequences were compared to the complete panel of high confidence gene models (not including splice variants), and 8,605 homoeologous triplets identified from genome sequence data (International Wheat Genome Sequencing, 2014; Pfeifer et al., 2014). By comparing our triplet panel to the IWGSC A, B and D gene models (totalling 32,081, 34226 and 33,079 models respectively), our models showed high similarity with 54.5%, 50.4% and 51.2% of IWGSC A, B and D gene models respectively. When we compared our triplets to the IWGSC triplets, we matched 76.8%,

76.2%, 76.0% of genes from the A, B and D genome respectively. Conversely, when the IWGSC triplets were compared to our set, they only matched 48.8%, 46.0% and 45.6% of our genes respectively.

These results suggest that as well as identifying the majority of triplets that had been found in previous studies, this approach was able to detect roughly double the total number of candidate homoeologous triplets. However, as the estimated number of genes on each of the sub-genomes is likely to be between 28,000 and 36,000 (Brenchley et al., 2012; Choulet et al., 2010; Hernandez et al., 2012; Massa et al., 2011), it is likely that there are many triplets that could not be identified using such a stringent method. Consistent with this, we found that around half of the total IWGSC gene models were represented in our triplet panel.

Assessment of genome expression level dominance

A Tukey test was used to identify homoeologues with transcripts from one copy varying in abundance from the other two. 5,248 unigenes showed significant differential expression from one of the genomes ($P<0.05$), and were thus assigned as up or down-regulated (Figure S4). Of these, 1,074, 978 and 854 A, B and D homoeologues respectively were found to be up-regulated, and 574, 1132 and 636 A, B and D homoeologues were found to be down-regulated (Supplementary Dataset 7). These unigenes exhibited some localised clustering, with those clusters distributed throughout the genome.

Validation of differential expression of homoeologues

The method typically used for validating gene expression differences inferred from mRNA-seq data is Quantitative Reverse Transcription PCR (qRT-PCR). This method is, however, slow and expensive, and in polyploid genomes, assays frequently cannot be designed. We therefore developed a method to quantify the *relative* contributions to the mRNA pool of each homoeologue based on the relative proportions in derived cDNA of each base at inter-homoeologue polymorphisms (IHPs). As an example, the A, B and D transcript assemblies A_comp34236_c0_seq1, B_comp4776_c0_seq1 and D_comp10830_c0_seq1, the products of which are so far uncharacterized, were one of the homoeologue triplets exhibiting significant genome expression imbalance identified using Tukey tests (Figure S4), in this case with the A genome copy being 3.5-fold higher on average than the B genome homoeologue, and 4-fold higher than the D genome homoeologue (Figure S5). These three reference sequences were aligned (Figure S6), and putative IHPs identified and genome assigned depending on the genome showing the alternative base. A set of primers was then designed to amplify and Sanger sequence all three transcripts simultaneously from cDNA. All predicted IHP positions could be confirmed as they appeared as double peaks in the subsequent sequence traces (Figure S7). As the PCR was performed on cDNA, these cDNA-based polymorphisms exhibit peak height variation proportional to the relative abundance in the RNA samples of the transcripts of the underlying genes. At any given IHP, peak height ratios were calculated as polymorphic base to ancestral base using Softgenetics Mutation Surveyor software. The null hypothesis (no expression dominance) would be for the A, B and D-linked IHP variant bases to have roughly equal peak height ratios.

However, consistent with the mRNA-seq data, we found that the A genome ratios were approximately 4.7-fold higher than the B and 5-fold higher than the D peak height ratios, confirming the mRNA-seq-based calling of differential expression (Supplementary dataset 8).

Discussion

As an allohexaploid with several hybridization events spread across the last 5 million years of its history, bread wheat is an interesting model for studying the evolution of gene expression following the sudden increases in transcript levels that must have occurred after each event. We hypothesized that the genomes would not contribute equally to the pool of transcripts, and consistent with this, we found extensive variation in the relative expression of triplets of homoeologues in the tissue analyzed (leaves). On the whole, no one genome dominated the mRNA pool. In fact, the imbalances appeared to be largely random from gene to gene, with no functional bias identified by analysis of gene ontology (data not shown). Some small regions did show distinct banding patterns however, suggesting regional genome dominance (yellow, cyan and magenta colours, as illustrated in Figure 3 and Figure S3). More prominent are the larger regions for which one genome appears to have been repressed (red, green and blue colours, as illustrated in Figure 3 and Figure S3). A rationale for such regions would be the supression of one or more genes with detrimental effects. Both genome dominance and genome repression effects are remarkably consistent across the panel of 54 varieties of bread wheat analysed.

These results are consistent with indications from earlier studies which found signs of genome asymmetry amongst neighbouring genes (Pfeifer et al., 2014) and a lack of global genome dominance (International Wheat Genome Sequencing, 2014; Pfeifer et al., 2014). In addition, there were just a few regions, all large, for which imbalances can be observed in individual (or just a few) varieties, which indicate specific long-range silencing or perhaps genome deletion/homoeologous exchange events as have been observed in other polyploidy crops, such as oilseed rape (Chalhoub et al., 2014).

We were able to study the relative contributions of each genome to the total pool of transcripts in bread wheat through the development of a set of resources based on mRNA-seq. As the cost of sequencing declines, GBS approaches are becoming ever more popular for genetic analyses. The focus for reduced representation methods (to reduce complexity, and hence costs) has been firmly on genome-targeted approaches such as RAD-seq, which generates sequences adjacent to restriction endonuclease cleavage sites and has been applied successfully in plants (Baird et al., 2008; Elshire et al., 2011; Miller et al., 2007) including for linkage map construction (Chutimanitsakun et al., 2011; Poland et al., 2012). There are important limitations to the approach, particularly for species with large, complex genomes such as that of bread wheat. The ordering of markers, for example by alignment to a genome sequence resource, is necessary to realize the full power of approaches such as Genome-Wide Association Scans (GWAS), as a cluster of markers in LD with a locus controlling variation for a trait provides more compelling evidence than single marker associations. This ordering can be difficult for RADseq markers; for example, as the majority of polymorphisms

identified will be in non-coding sequences, which necessitates the use of the genome sequence of the species being studied. In contrast, focusing on transcribed sequences, by mRNA-Seq, enabled us to fully exploit comparative genomics, by utilizing collinearity of coding regions between wheat and Brachypodium. In more generalized applications of the approach, even relatively distantly related orthologous genes can readily be identified, although conservation of synteny can be expected to decrease with increasing genetic distance. Furthermore, mRNA-Seq data can be analyzed for SNP variation as well as transcript abundance, both of which can then be used in association genetic approaches such as Associative Transcriptomics [15].

Experimental procedures

Growth of Materials

Triticum urartu, *Aegilops speltoides*, *Aegilops tauschii* and *Triticum turgidum dicoccoides*, 47 lines of a single seed descent linkage mapping population derived from a Chinese Spring x Paragon cross and 54 lines from a diversity panel of *Triticum aestivum* plants were grown for transcriptome sequencing and validation. The seeds were placed on moist filter paper and placed in a refrigerator at 6°C for 2 days before being transferred to a germinator at 20°C overnight. Seeds were then transferred to pots with a peak/sand mix and arranged in four block, one-way randomized design with one plant of each of the accessions per block and randomized within each block. Plants were grown in long-day glasshouse conditions (16 hour photoperiod) at 15 °C (400W HQI metal halide lamps).

RNA extraction and cDNA synthesis

Second true leaves from each of four plant replicates per accession were harvested approximately 14 days after pricking out (21 d after sowing) as close to the midpoint of the light period as possible, pooled and immediately frozen in liquid nitrogen. Samples were extracted using the Omega Biotek EZNA Plant RNA Kit according to manufacturer's instructions. cDNA was synthesised using standard protocols from 2ul of total RNA.

Leaf transcriptome sequencing

Illumina sequencing, quality control and data processing were conducted as described previously (Bancroft et al., 2011). The HiSeq2500 platform was used to generate 100 base reads, paired-end for the progenitor lines and single-end for the mapping lines and diversity panel.

Development of Brachypodium-based pseudomolecules

First, using Trinity r2012-03-17 (Grabherr et al., 2011), transcriptome assemblies were constructed separately for each of the diploid species, *T. urartu* (101 million reads), *Ae. speltoides* (124 million reads) and *Ae. tauschii* (98 million reads), yielding 123,236 A assemblies, 169,009 B assemblies and 98,063 D assemblies respectively. Redundancy within each set was then removed with CD-HIT v4.5.4 (Li and Godzik, 2006) using an identity threshold of 0.95 and a word length of 5. This step reduced the assemblies to 105,069, 132,363 and 85,296 A, B and D clusters respectively. The sequence identifiers generated by Trinity were prepended with their genome of origin of in order to

distinguish them and the three fasta files were then simply combined to produce a transcriptome of 322,728 unigenes. When this was used as a reference sequence for alignment of hexaploid Chinese Spring RNA-Seq reads it was found that a relatively low number of reads mapped to the B genome assemblies, suggesting that *Aegilops speltoides* was not an ideal B genome proxy. The B genome assemblies within this set were therefore "cured" using reads obtained from the tetraploid *Triticum turgidum dicoccoides*, which contains a B genome more closely related to that in hexaploid wheat, using a method previously described (Higgins et al., 2012). Curing refers to the correction of reference sequences by repeated rounds of alignment and adjustment. Reads were aligned to the un-cured reference and mismatched sites corrected in the B genome reference. After six cycles of curing, a total of 198,607 bases over the B genome assemblies had been modified. The edited file was then used as the reference for alignment with reads for transcript quantification. The unigenes were aligned, using BLASTn (Altschul et al., 1990) with an E-value threshold of 1E-30, against annotated gene sets for *Brachypodium distachyon* (MIPS v1.2), rice (MSU v5) and Sorghum (MIPS v1.4) with functional annotation being extracted and recorded for each hit.

A linkage map of bread wheat, based on the transcriptome SNP markers, was constructed, essentially as described previously for the oilseed rape transcriptome SNP-based linkage map (Bancroft et al., 2011). After scoring polymorphisms across a population of 47 single seed descent lines from a Chinese Spring x Paragon cross plus the parent lines, the markers were filtered in order to retain only those with BLAST hits to Brachypodium gene models. As the number of markers was too large to be

processed by conventional linkage mapping software, the linkage map was constructed using string-matching in MS Excel spreadsheets to first identify SNP markers with similar scoring strings (and therefore grouped close to each other in the genome) from a pool of unmapped markers. The clusters of markers identified were then ordered based on the order of the orthologous *Brachypodium* gene models and runs of collinear markers incorporated into recombination bins (624 in all) comprising sets of markers with identical allele calls. Non-collinear markers returned to the unmapped pool and the process repeated iteratively until the whole genome was covered. Unlinked blocks of markers were positioned relative to each other based on published collinearity analyses between the genomes of Brachypodium and wheat (International Brachypodium, 2010). A graphical genotype was generated and inspected manually to ensure that there were no inconsistencies that might indicate false assembly of the map.

Based on the BLAST hits of the unigenes containing the distal SNP markers in each block of collinearity between the wheat SNP map and the Brachypodium genome sequence, the Brachypodium genome sequence was split into segments, with the division mid-way between the BLAST HSP coordinates defining the ends of rearranged blocks. These blocks were then re-ordered and re-oriented, based on their span of recombination bins in the linkage map, to establish a set of pseudomolecules representing the hypothetical organization of the orthologous sequences in the wheat genome, essentially as described previously for the *B. napus* pseudomolecules (Harper et al., 2012). Finally, the Trinity unigenes were aligned against the constructed

pseudomolecules using BLASTN with an E-value threshold of 1E-30 and the chromosome and coordinates recorded for each best hit.

Assessment of quantitative genome contributions to the transcriptome

In order to assess regions of genome dominance, redundant transcript assemblies were filtered in cases where multiple unigenes with the same Brachypodium BLAST hit were found with the same location on the genome. In this case, only the longest unigene was selected, reducing the number of unigenes to 147,411, comprising 47,160, 59,663 and 40,588 in the A, B and D genomes, respectively. Homoeologues were identified as unigenes in all three genomes with top reciprocal BLAST hits to each other (threshold e-value 1E-30). 15,527 triplets of homoeologues were identified in this way. mRNAseq reads from the wheat progenitor species and 54 varieties from the hexaploid wheat diversity panel were then mapped to the non-redundant 147,411-unigene reference sequence and, using methods and scripts described in Bancroft *et al.* (Bancroft et al., 2011) and Higgins *et al.* (Higgins et al., 2012), expression of each unigenes was estimated for each accession. Transcript abundance was quantified and normalised as reads per kb per million aligned reads (RPKM) separately for each unigene. Filtered data for the 15,527 triplets of homoeologues were retrieved for further analysis. Mapping statistics and average transcript abundance for each accession are provided in Supplementary dataset 9.

To identify homoeologues that may be differentially expressed across the genomes, a Tukey test was applied to each pair within a homoeologue triplet (*i.e.* AB, AD, BD).

Where two of these pairwise tests (*e.g*. AB and AD) were significant (P<0.05), the genome contributing to both (*i.e*. A) was defined as being differentially expressed. The mean RPKM values across the diversity panel were then compared to define relative up- or down-regulation of that genome orthologue. Cases where a single test, or all three pairwise tests were significant, were ignored. All homoeologues were then placed in genome order and significant results colour-coded as seen in Figure S6.

To identify expression structure in the genomes, a normalised tile plot was created. As the reference is based on diploid species representing the hexaploid genome progenitors, all with different evolutionary distances from bread wheat, it is reasonable to assume that there will be unequal efficiency of read mapping to the three genomes. To counteract this, a normalisation based on the total number of mapped reads to each genome was used. RPKM values for each of the putative homoeologues was adjusted to a range between 1 and 0, where 1 is the individual with the lowest and 0 the individual with the highest expression value across the diversity panel. These values were then converted to RGB hexcodes and arranged in genome order to create a tile plot where the colour of each tile is converted corresponding to the given intensities of the red, green and blue primaries. According the standard CMYK colorspace, if cyan, magenta and yellow intensities are equal, the resulting tiles will be a shade of grey, where boundaries are black CMY(1,1,1) and white CMY(0,0,0). A non-greyscale colour will show an expression difference between the homoeologous unigenes on the A, B and D genomes.

Validation by cDNA sequencing

As an example, a triplet of homoeologues from a region identified as having increased A genome expression on linkage group 2 was selected for validation of relative expression, based on sequencing of cDNA. PCR primers were designed to amplify the set of homoeologues (A_comp34236_c0_seq1, B_comp4776_c0_seq1 and D_comp10830_c0_seq1) from cDNA (349bp). Primer sequences were as follows: forward, GATGTATCAAGTTCTGCTCTTC; reverse, CTTATAGTGTCACCACCAATAAC. Capillary sequencing was then performed using the reverse primer to assess relative expression based on peak heights of inter-homoeologue polymorphisms, as measured using the Softgenetics Mutation Surveyor software. This approach was used for all further instances of validation of transcript abundance differences identified on the basis of mRNA-seq data.

Comparison of triplets to wheat genome data

A list of 8,605 homoeologous triplets (http://wheat-urgi.versailles.inra.fr/; IWGSC, 2014) based on wheat genome data were compared to our triplets following concatenation of splice variants to construct a single full length transcript for each gene. Reciprocal BLASTn (Evalue <1E-30) was used to assess significant similarity between the sets of triplets. Using the same BLAST parameters, our homoeologous triplets were also compared to the full set of transcript sequences for all high-confidence (HCS) gene models with a home on the IWGSC sequence assembly (excluding splice variants) (https://urgi.versailles.inra.fr), as these are likely to represent the approximate gene complement of the bread wheat sub-genomes.

Acknowledgements

We thank V. Unwin for technical support and The Genome Analysis Centre and Source Bioscience for generating Illumina sequence data.

This work was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC BB/H004351/1 (IBTI Club), BB/L002124/1, BB/L027844/1).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one* **3**, e3376.

Bancroft, I., Morgan, C., Fraser, F., Higgins, J., Wells, R., Clissold, L., Baker, D., Long, Y., Meng, J., Wang, X., Liu, S. and Trick, M. (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat Biotechnol* **29**, 762-766.

Blanc, G. and Wolfe, K.H. (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *The Plant cell* **16**, 1679-1691.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F., Edwards, K.J., Bevan, M.W. and Hall, N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705-710.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H.,

Guan, C. and Wincker, P. (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**, 950-953.

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., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P. and Chalhoub, B. (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and Aegilops). *The Plant cell* **17**, 1033-1045.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M.C., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., Anderson, J.A., Gill, B.S., Appels, R., Keller, B. and Feuillet, C. (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *The Plant cell* **22**, 1686-1701.

Chutimanitsakun, Y., Nipper, R.W., Cuesta-Marcos, A., Cistue, L., Corey, A., Filichkina, T., Johnson, E.A. and Hayes, P.M. (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* **12**, 4.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* **6**, e19379.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P. and Bancroft, I. (2012) Associative transcriptomics of traits in the polyploid crop species Brassica napus. *Nat Biotechnol* **30**, 798-802.

Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J. and Hiscock, S.J. (2006) Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Current biology : CB* **16**, 1652-1659.

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., Jouve, N., Šimková, H., Valárik, M., Doležel, J. and Mayer, K.F.X. (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *The Plant Journal* **69**, 377-386.

Higgins, J., Magusin, A., Trick, M., Fraser, F. and Bancroft, I. (2012) Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species Brassica napus. *BMC Genomics* **13**, 247.

Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proceedings Biological sciences / The Royal Society* **256**, 119-124.

International Barley Genome Sequencing, C., Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711-716.

International Brachypodium, I. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **463**, 763-768.

International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**, 1251788.

Kliebenstein, D.J. (2008) A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PloS one* **3**, e1838.

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., Wu, J., Yin, L., Zhang, R., Wu, L., Zheng, Y. and Mao, L. (2014) mRNA and Small RNA Transcriptomes Reveal Insights into Dynamic Homoeolog Regulation of Allopolyploid Heterosis in Nascent Hexaploid Wheat. *The Plant cell* **26**, 1878-1900.

Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659.

M. Schranz, P.E., J. Pires, N. van Dam, C. Wheat (2011) Comparative genomics in the Brassicales: Ancient genome duplications, glucosinolate diversification and Pierinae herbivore radiation. In: *Genetics, Genomics and Breeding of Brassica Oilseeds* (Dave Edwards, J.B., Isobel Parkin, Chittaranjan Kole ed) pp. 206-218. Enfield NH: Science Publishers.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing, C., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F. and Olsen, O.A. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092.

Massa, A.N., Wanjugi, H., Deal, K.R., O'Brien, K., You, F.M., Maiti, R., Chan, A.P., Gu, Y.Q., Luo, M.C., Anderson, O.D., Rabinowicz, P.D., Dvorak, J. and Devos, K.M. (2011) Gene Space Dynamics During the Evolution of Aegilops tauschii, Brachypodium distachyon, Oryza sativa, and Sorghum bicolor Genomes. *Molecular Biology and Evolution* **28**, 2537-2547.

Mayer, K.F., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubalakova, M., Suchankova, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Dolezel, J., Waugh, R. and Stein, N. (2011) Unlocking the barley genome by chromosomal and comparative genomics. *The Plant cell* **23**, 1249-1263.

Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711-716.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A. and Johnson, E.A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* **17**, 240-248.

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. (1995) Cereal genome evolution. Grasses, line up and form a circle. *Current biology : CB* **5**, 737-739.

Pfeifer, M., Kugler, K.G., Sandve, S.R., Zhan, B., Rudi, H., Hvidsten, T.R., International Wheat Genome Sequencing, C., Mayer, K.F. and Olsen, O.A. (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091.

Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.L. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one* **7**, e32253.

Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M. and Wolfe, K.H. (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8397-8402.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. and Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112-1115.

Seoighe, C. and Gehring, C. (2004) Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. *Trends in genetics : TIG* **20**, 461-464.

Tian, C.G., Xiong, Y.Q., Liu, T.Y., Sun, S.H., Chen, L.B. and Chen, M.S. (2005) Evidence for an ancient whole-genome duplication event in rice and other cereals. *Yi chuan xue bao = Acta genetica Sinica* **32**, 519-527.

Trick, M., Cheung, F., Drou, N., Fraser, F., Lobenhofer, E.K., Hurban, P., Magusin, A., Town, C.D. and Bancroft, I. (2009a) A newly-developed community microarray resource for transcriptome profiling in Brassica species enables the confirmation of Brassica-specific expressed sequences. *BMC Plant Biol* **9**, 50.

Trick, M., Long, Y., Meng, J. and Bancroft, I. (2009b) Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. *Plant Biotechnol J* **7**, 334-346.

Figure legends

Figure 1. Collinearity between inferred gene order in wheat and Brachypodium. The plot shows the order in each genome of 17,315 wheat unigenes and their Brachypodium orthologues with points color-coded by sequence similarity to the chromosome assignment of Brachypodium gene models: blue for chromosome 1, orange for chromosome 2, purple for chromosome 3, brown for chromosome 4 and green for chromosome 5.

Figure 2. Collinearity between inferred gene order in wheat and barley. The plot shows the order in each genome of 8,187 wheat unigenes and their barley orthologues with points color-coded by sequence similarity to the chromosome assignment of Brachypodium gene models: blue for chromosome 1, orange for chromosome 2, purple for chromosome 3, brown for chromosome 4 and green for chromosome 5.

Figure 3. Tile plots illustrate relative transcript contributions for the A, B and D copies of 2,571 triplets of homoeologous genes on linkage group 2. Represented are 54 bread wheat accessions, diploid ancestors Triticum urartu (AA), *Aegilops speltoides* (BB) and *Aegilops tauschii* (DD), tetraploid ancestor *Triticum dicoccoides* (AABB), and *in silico* tetra- and hexaploid combinations. The A genome is represented by cyan, B genome magenta and D genome yellow. The homoeologous genes are arranged in pseudomolecule order (which is largely identical for all three genomes). Regions of interest are marked, including the region used for validation (*).

Supporting information

Figure S1. Collinearity between inferred gene order and the V5 wheat genome zipper for the A, B and D genomes.

Figure S2. Collinearity of wheat pseudomolecule and 3B genome assembly.

Figure S3. Tile Plots

Figure S4. Tukey Plots

Figure S5. Dot histogram

Figure S6. Homoeologue Alignment

Figure S7. Inter-homoeologue polymorphisms visualized by capillary sequencing

Figure S8. Workflow diagram for visualising homoeologue expression patterns

Supplementary Data file 1. Annotation of unigenes

Supplementary Data file 2. Transcriptome SNP linkage map for hexaploid wheat

Supplementary Data file 3. Consensus recombination bins representing the wheat SNP linkage map

Supplementary Data file 4. Collinearity of wheat SNP linkage map and Brachypodium genome

Supplementary Data file 5. Pseudomolecule specification

Supplementary Data file 6. Mapping of unigenes to pseudomolecules

Supplementary Data file 7. Significant Tukey Tests

Supplementary Data file 8. Genome dominance validation

Supplementary Data file 9. Mapping Statistics

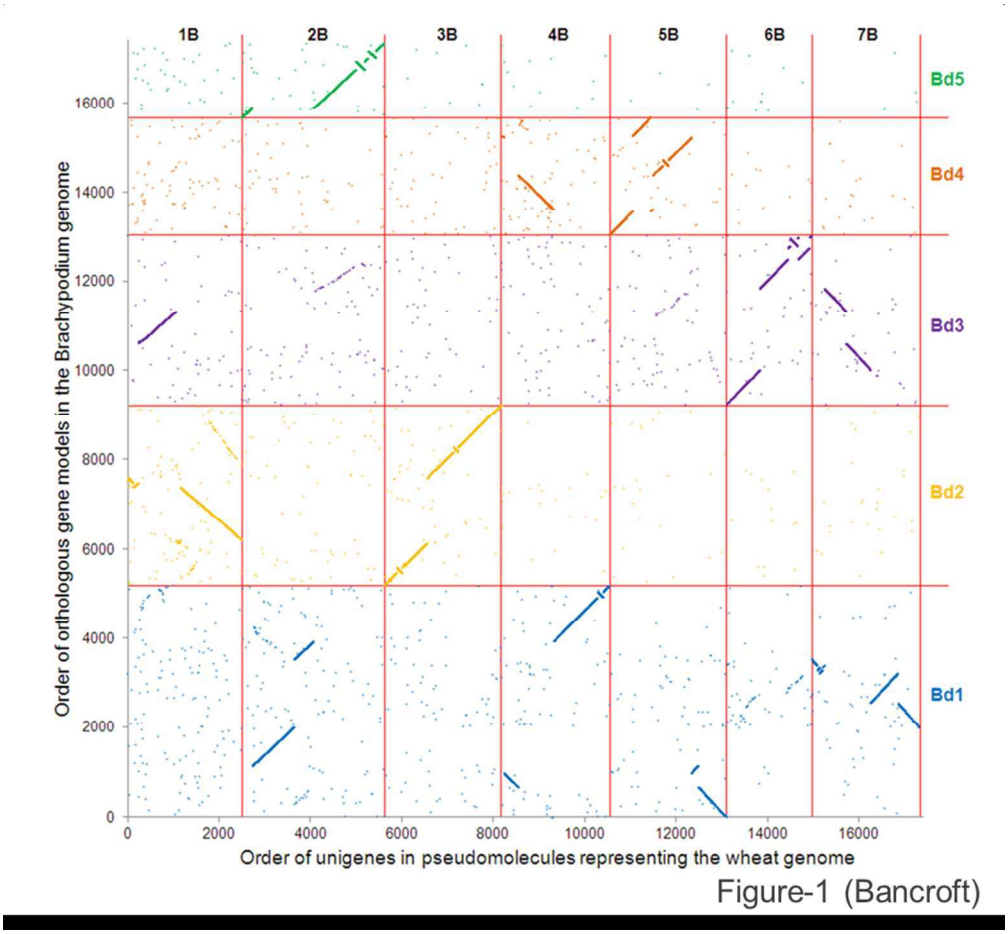

Figure 1. Collinearity between inferred gene order in wheat and Brachypodium. The plot shows the order in each genome of 17,315 wheat unigenes and their Brachypodium orthologues with points color-coded by sequence similarity to the chromosome assignment of Brachypodium gene models: blue for chromosome 1, orange for chromosome 2, purple for chromosome 3, brown for chromosome 4 and green for chromosome 5.
177x163mm (150 x 150 DPI)

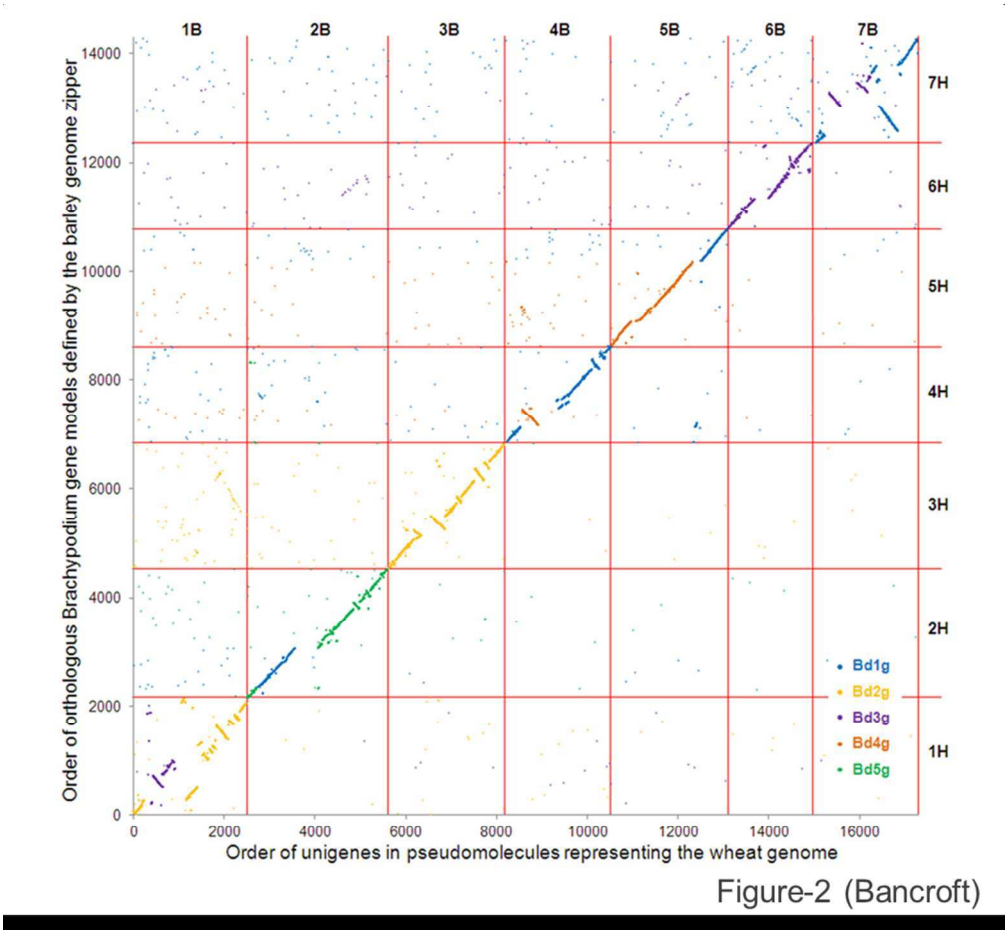

Figure 2. Collinearity between inferred gene order in wheat and barley. The plot shows the order in each genome of 8,187 wheat unigenes and their barley orthologues with points color-coded by sequence similarity to the chromosome assignment of Brachypodium gene models: blue for chromosome 1, orange for chromosome 2, purple for chromosome 3, brown for chromosome 4 and green for chromosome 5.
177x163mm (150 x 150 DPI)

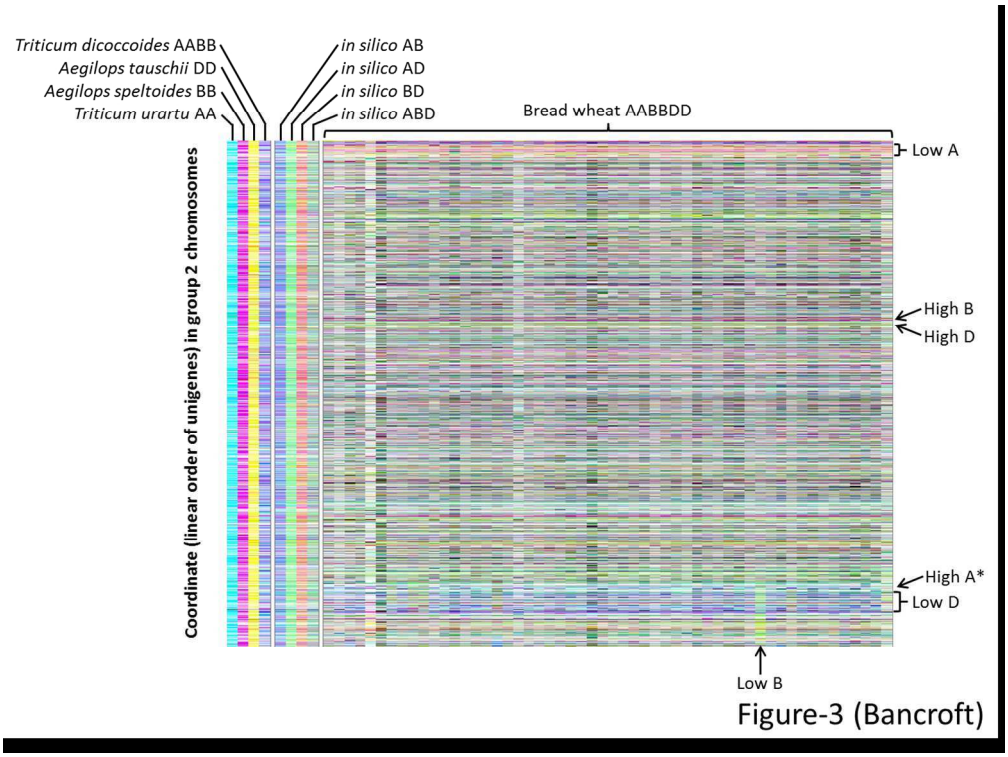

Figure 3. Tile plots illustrate relative transcript contributions for the A, B and D copies of 2,571 triplets of homoeologous genes on linkage group 2. Represented are 54 bread wheat accessions, diploid ancestors Triticum urartu (AA), Aegilops speltoides (BB) and Aegilops tauschii (DD), tetraploid ancestor Triticum dicoccoides (AABB), and in silico tetra- and hexaploid combinations. The A genome is represented by cyan, B genome magenta and D genome yellow. The homoeologous genes are arranged in pseudomolecule order (which is largely identical for all three genomes). Regions of interest are marked, including the region used for validation (*).
260x194mm (150 x 150 DPI)