The Plant Journal (2006) doi: 10.1111/j.1365-313X.2006.02841.x

Homoeologous gene silencing in hexaploid wheat

A. Bottley[1,*], G. M. Xia[2] and R. M. D. Koebner[1]
[1]*Department of Crop Genetics, John Innes Centre, Norwich, NR4 7UH, UK, and*
[2]*School of Life Science, Shandong University, Jinan, Shandong, 250100, People's Republic of China*

Received 7 March 2006; revised 26 May 2006; accepted 30 May 2006.
*For correspondence (fax +44 (0) 1603 450045; e-mail andrew.bottley@bbsrc.ac.uk).

Summary

The vast majority of angiosperms are (or were once) polyploid, and as hexaploid bread wheat has undergone two ploidy events separated by approximately 0.5 million years, it represents an elegant model to study gene silencing over time in polyploids. Using an SSCP platform, we have analysed patterns of transcriptional silencing (frequency, genome identity and organ specificity) within 236 single-copy genes, each mapping to one locus on one of the three homoeologous chromosomes within groups 1, 2, 3 and 7 of wheat. In about 27% of unigenes expressed in leaf, and about 26% of those in root, one (rarely two) members of a gene set (homoeoalleles) were not present in the cDNA template. Organ-specific regulation is commonplace, with many homoeoalleles transcribed in leaf but not root (and vice versa). There was little indication of extensive bias towards selective silencing of a particular genome copy. Expression of some of the silenced homoeoalleles was restored in certain aneuploid lines and varieties, and these displayed a significant degree of genetic variation for the silencing of a given homoeoallele. We propose that a substantial proportion of this phenomenon is effected by an epigenetic mechanism, and suggest that this form of genetic variation may be a significant player in the determination of phenotypic diversity in breeding populations.

Keywords: polyploidy, gene silencing, wheat, SSCP, epigenesis.

Introduction

Most of our important crop plants have evolved as a result of one or more wide hybridization events, each followed by a chromosome doubling step to restore fertility – the end product of these processes being a stable allopolyploid. It is thought that many plant species undergo speciation via this process, and that up to 95% of all known angiosperms have experienced a ploidy event(s) at some stage during their evolutionary history (Grant, 1981). A few species have undergone chromosome doubling in the absence of wide hybridization, giving rise to autopolyploids. To ensure both sexual compatibility for successful fertilization and genetic compatibility for the generation of viable zygotes, the progenitor species of an allopolyploid need to be closely related to one another. As a direct consequence of the combining of two (or more) very similar genomes into a single nucleus, all genes, except for those few that are genome-specific, become functionally duplicated. This redundancy may be resolved through loss of function, either irreversibly via wholesale deletion or mutation to pseudogenes, or more subtly, by epigenetic silencing of excess copies; alternatively, functional duplicated sequences may be fully tolerated, where the alternate genes may or may not specify identical function.

Extensive genome changes – deletions and mutations – have been observed to occur in the early generations following the creation of synthetic *Brassica* spp. tetraploids (Song *et al.*, 1995), a process that could certainly contribute to the proposed differentiation of the homoeologous chromosome segments of *B. oleracea*, such that each is characterized by a unique pattern of gene presence interspersed with absence, in comparison to the assumed ancestral sequence of *Arabidopsis thaliana* (O'Neill and Bancroft, 2000). In contrast, in tetraploid cotton, gene maintenance, not gene silencing, is the rule rather than the exception for genes duplicated by polyploidization (Cronn *et al.*, 1999). However, a significant proportion of homoeologous copies of cotton genes appear to be differentially transcribed/silenced, and the pattern of this bias has been shown to be organ-specific (Adams *et al.*, 2003). In monocots, both deletion and epigenetic silencing of genes have been documented in *de novo* hybrids between *Aegilops sharonensis* and *Triticum monococcum* (Kashkush *et al.*, 2002),

but this is not evident in equivalent hybrids between diploid *Gossypium* spp. (Liu *et al.*, 2001). In general, it appears that wide hybridization can significantly affect gene expression immediately upon allopolyploid formation, but that any genomic perturbation that accompanies the polyploidization event is followed by relative stasis.

Tetraploid wheat, the major cultivated form of which is *T. turgidum* var. *durum*, evolved from a hybridization event between the B genome diploid progenitor species (yet to be identified, but thought to belong to the Sitopsis section of the genus *Aegilops*) and *T. urartu*, a close relative of the ancestor of cultivated diploid einkorn wheat (*T. monococcum*). The timing of this polyploidization event has been placed by molecular clock analysis at less than 0.5 Mya (Huang *et al.*, 2002). Hexaploid bread wheat (*T. aestivum*), in which the D genome of goat grass (*Ae. squarrosa*) was incorporated into the BA tetraploid, first appears in the archaeological record approximately 10 kya. Under current models of polyploidy-induced silencing, there have therefore been two opportunities for silencing effects to become fixed in the bread wheat genome. Using a whole-transcript profiling approach, Kashkush *et al.* (2002) estimated that 1–5% of genes in newly synthesized wheat hexaploids are silenced. This is comparable with the estimates of He *et al.* (2003), which proposed that 7–8% of genes are silenced in established wheat hexaploids. The same authors also suggested that the genomic origin of silenced genes is non-random, and specifically that D genome copies are silenced at a higher frequency than their A or B counterparts.

Because the three progenitor genomes A, B and D are so similar to one another, most hexaploid genes are present in three homoeologous forms (commonly referred to as 'homoeoalleles'). That this is the case can be demonstrated both by the high frequency with which a three-banded RFLP profile (one fragment per genome) is generated when restricted hexaploid genomic DNA is hybridized with a cloned, single-copy cDNA sequence, and by the observation that many isoenzyme loci fall into a homoeologous series. The presence of three homoeoallelic coding sequences does not necessarily imply that three independent mRNAs are transcribed, let alone successfully translated. But until now, resolving individual homoeoallelic transcripts from one another has been technically difficult, as these molecules differ only slightly from one another in sequence. In contrast, at the translational level, it has long been possible to distinguish between the homoeoalleles of many isoenzymes. Thus, of 54 bread wheat isoenzymes whose genetic control has been elucidated using histological staining to visualize enzyme activity, 42 show co-expression of all three homoeoalleles, but for 12, only one homoeoallele can be identified using aneuploid analysis (data extracted from Mcintosh *et al.*, 1998). The inability to define all three homoeoalleles in these cases may be interpreted as reflecting the electrophoretic co-migration of peptides or the transcription of silent copies. Furthermore, the existence of a significant level of functionally redundant homoeoalleles can be inferred from the long history of induced mutation carried out by both breeders and geneticists. As an example, the relative frequencies of chlorophyll mutations induced in tetraploid and diploid wheats suggest that, for around 20% of duplicated genes, one copy is inactivated (Comai, 2000).

Homoeologous sequences are, by definition, highly similar but non-identical. Recently, we have shown that the single-strand conformation polymorphism (SSCP) technique is capable of distinguishing between homoeologous wheat gDNA sequences (Forsström *et al.*, 2003). By extension, the same technique can be used to distinguish between homoeologous cDNA molecules, an approach also used for the identification of homoeologous sequences of cotton (Cronn *et al.*, 1999). In the present paper, we exploit this capability to identify gene silencing in hexaploid wheat at the transcriptional level, using as evidence for silencing the situation where a homoeoallele is represented in the SSCP profile of gDNA but not in the profile of the equivalent cDNA. We suggest that this approach is well suited for exploration of the extent of homoeologous silencing in wheat, the existence of genomic patterning in silencing, and the degree to which the expression of homoeoalleles has become organ-specific. This approach has already been applied in cotton (Adams *et al.*, 2003), and lends itself naturally, by extension, to any other polyploid genome.

Results

In pilot experiments designed to test the fidelity of the phi-amplification process, we used ten primer combinations, and detected no discernable differences between SSCP profiles derived from phi-amplified and non-amplified cDNA template. Furthermore, identical SSCP profiles were generated from two independently amplified cDNAs that had each been reverse-transcribed from separate RNA preparations derived from a single tissue sample. For the main experiment, all cDNA amplifications, based on 236 ESTs, were conducted using phi-amplified template.

All primer pairs amplified adequately from genomic DNA of 'CS' wheat, and nullisomic–tetrasomic lines were therefore employed to assign genomic origin to the individual amplicon components, by correlating band loss with the chromosome known to be absent in each particular nullisomic–tetrasomic template (Figure 1). Generally this approach allowed the identification of all three expected genomic copies for each set of homoeoalleles. However, in some cases, it was only possible to recognize two (and rarely one) of the three copies. More than one product was frequently assigned to a single genome. This could reflect the presence of multiple loci on a particular chromosome, but more likely is a property of SSCP itself, which can

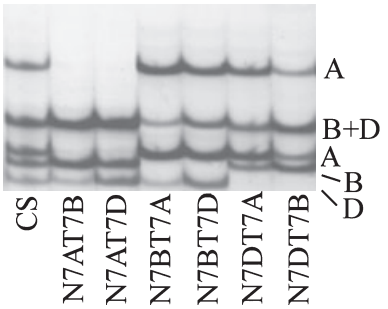

Figure 1. SSCP separates the three homoeologous components of an EST-based amplicon (EST BF201109).
The genomic origin of each PCR product is indicated on the left. N7AT7B refers to the line nullisomic 7A–tetrasomic 7B, etc.

generate multiple conformers from a single sequence. Usually this does not complicate the analysis of profiles, because the most stable comformer is recognized as the band of the highest intensity within a group of products.

Direct comparisons were possible between gDNA profiles and their corresponding cDNA-derived profiles because the amplicons represent intron-free sequence. Only one primer pair out of the 236 (EST BE497590) amplified an intron in wheat, where one is not present in rice. Of the 236 amplicons, 42 were completely absent in leaf cDNA of CS, 142 showed matching gDNA and cDNA profiles, and therefore none of these homoeoalleles was silenced, and in the remaining 52 amplicons, one homoeoallele of the two or three amplified from the gDNA template was absent from the cDNA amplicon, and was therefore taken to be silenced in the leaf sample. An example showing both active and silenced homoeoalleles in leaf and root is EST BF201560 shown in Figure 2, and an overall summary of loci where silencing was detected is given in Table 1. Thus, in CS seedling leaf, one homoeoallele was silenced for 27% of the genes expressed in this tissue. The frequency of silenced homoeoalleles was greatest in the D genome, although not statistically significantly so ($P > 0.05$), as assessed by a χ^2 test. A similar analysis of root cDNA profiles revealed non-expression of 95 genes, full expression of 105, and silencing for 36 (26% of the genes expressed in this tissue). In 35 of these 36 genes, one homoeoallele was silenced, but in one gene (BE438771) neither the A nor the D genome homoeoallele is expressed. The genomic distribution of silenced homoeoalleles was again not different from random.

An unexpected observation was that, in some cases, homoeoalleles silenced in CS (either leaf or root) were expressed in one or more of the nullisomic–tetrasomic lines (Figure 3). This is noteworthy, as these cytogenetic stocks were largely derived from CS. We investigated the possibility that the suppression of transcription in CS, and its restoration in a particular nullisomic–tetrasomic line, is

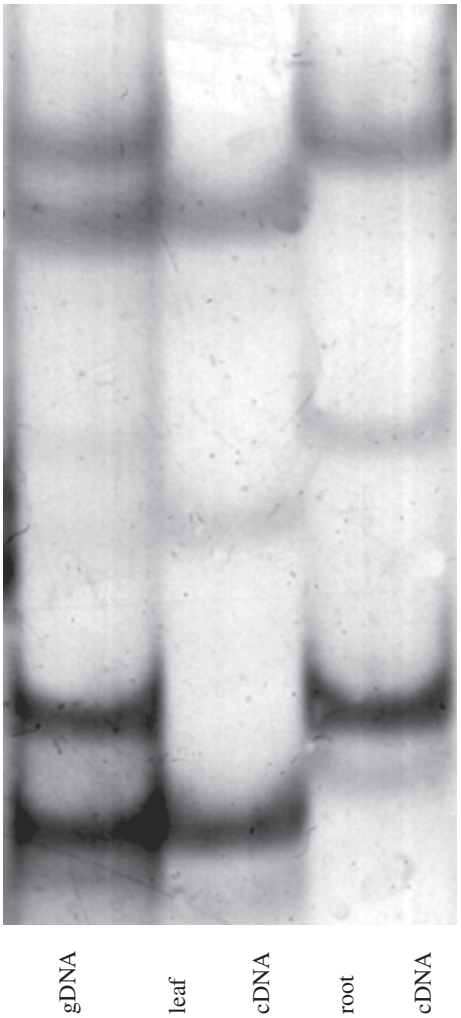

Figure 2. Organ-specific transcript silencing of one homoeoallele in leaf, and a different one in root, for EST BF201560.

due to the presence of a discrete repressor locus on the chromosome deleted in the line for which transcription was restored. For example, in the case of a group 7 gene in which the 7A copy is silenced in CS, and a repressor is present on 7B, then transcription should be restored in both nullisomic 7B–tetrasomic 7A and nullisomic 7B–tetrasomic 7D, but will remain suppressed in both nullisomic 7D–tetrasomic 7A and nullisomic 7D–tetrasomic 7B. Among those loci in homoeologous group 7 showing restored transcription, four exhibit this pattern; however, for three loci, restoration is only observed in one of these four nullisomic–tetrasomic lines, while for the remaining four loci, restoration is observed at least three nullisomic–tetrasomic lines (see, for example, Figure 4, which shows restoration of 1A transcription in the absence of both 1B and 1D). Similar patterns obtain for silenced homoeoalleles mapping to the other homoeologous groups

investigated. In addition, we have occasionally noted the complementary situation, where a locus is transcribed in the euploid individual, but not in the aneuploid one, even though the critical genomic copy is present in both. Therefore, we can exclude the possibility that in many, although not all, cases, conventional repressor genes are responsible for the phenomenon of variable expression of homoeoalleles.

Having established the phenomenon of reversion of expression, we extended the screen to a panel of diverse wheat genotypes, to investigate whether homoeoallele silencing represents a cryptic source of genetic variation. We assayed non-Phi-amplified cDNAs from the leaf and root tissue using primers designed to amplify a number of homoeoallelic sets that display the restoration phenotype among the aneuploid stocks. An example of this is EST BF202265, in which some entries show no silencing (identical gDNA and cDNA profiles), while others show loss of expression of one homoeoallele (Figure 4a). For EST BE404737, one of the three homoeoalleles (putatively the D homoeoallele) is silenced in the variety Avalon, but not in Cadenza (Figure 4b), with the remaining homoeoalleles being expressed in both varieties. In contrast, for EST BE443527, the D genome homoeoallele is silenced in both varieties (Figure 4c).

We attempted to establish whether there is any correlation between function and propensity to be silenced, by performing BLAST searches for all sequences in which silencing was identified. A diverse range of function emerged (Table 1), and no meaningful pattern could be ascertained. Similarly, we have tested the hypothesis that the propensity to be silenced is restricted to particular chromosomal regions, by determining whether silenced EST sequences tended to be concentrated in particular bin locations; however, once again, no obvious pattern emerged.

Discussion

Modern hexaploid wheat represents the end result of two independent polyploidization events, the first occurring 0.5 Mya and the second 0.01 Mya. This contrast between an old and a recent event allows analysis of the effect of time on the rate of the evolution of silencing, and therefore makes wheat a useful model for the study of polyploidy-associated gene silencing. In contrast, cotton, which is the other major subject of investigation to date, is an ancient tetraploid, and thus only allows a view of the effect of a single and ancient polyploidization event. In addition, wheat enjoys the luxury of a substantial and growing resource of EST sequences and associated mapping, along with well-established, characterized aneuploid and deletion stocks, and many mapping populations, all of which facilitate a high-throughput approach to survey the genomic distribution of silencing. Finally, wheat is an important food crop, and thus genetic insights have the potential to be directly exploited in the breeding process.

Heritable gene silencing can be achieved either genetically by deletion or by the formation of pseudogenes, or epigenetically, most typically by C-methylation of promoter and/or coding sequence. The former represents, in evolutionary terms, a more long-term solution than the latter to the problem of gene redundancy in polyploids, as it involves an irreversible change in genomic DNA sequence. However, generating deletions/point mutations is more difficult than achieving an epigenetic change, leading to the prediction that silencing in young polyploids is likely to be dominated by methylation, and in older ones by sequence alteration. Consistent with this prediction is the pattern of gene duplication in yeast, which appears to be an ancient tetraploid, where relic duplicated sequences are interspersed with single-copy sequences, indicative of large-scale deletion events (Wolfe and Shields, 1997). In contrast, in tetraploid cotton, where the polyploidization event is thought to have occurred some 0.5–2.0 Mya, polyploidization does not seem to have been accompanied by any perceptible rate of deletion in homoeoallelic pairs (Cronn *et al.*, 1999).

Our estimates of the frequency of silencing need to be put into context with those emerging from published studies of both wheat and other species. Using a cDNA-AFLP assay, Kashkush *et al.* (2002) estimated that about 5% of genes are silenced in a newly synthesized allohexaploid, a figure comparable with that arrived at by He *et al.* (2003) using a similar approach. Although this level is substantially lower than our estimate, importantly it reflects gene silencing immediately following polyploidization, rather than over the thousands or hundreds of thousands of generations through which established polyploids have passed. Exploiting large collections of EST data, Mochida *et al.* (2004) concluded that silencing affected 11 out of 90 sets of homoeoalleles (12%). A similar difference between *de novo* and established hybrids has been demonstrated in cotton, where Adams *et al.* (2003, 2004 used a cDNA-AFLP assay to show that about 5% of all genes are silenced in a newly synthesized allotetraploid, but that about 25% of genes are affected in established tetraploid cotton. In the newly synthesized *Arabidopsis thalianai* × *A. arenosa* wide hybrids described by Comai *et al.* (2000), 0.4% of genes are silenced as a direct result of polyploidization, a figure substantially lower than that derived in wheat and cotton. The reason(s) for this disparity are unclear, but may be a consequence of a lower level of homoeology between the two contributing genomes, and therefore the induction of a lower level of interference in their independent expression. Silencing certainly appears to be a common phenomenon in established polyploids, and the frequency of silencing seems to increase over time. Cryptic polyploids, such as maize, are

Table 1 Single-copy EST loci at which silencing was observed

Genbank identifier	Function/putative function	Tissue silenced	Genomes identifiable	Genome silenced	E number	Score (bits)
BE398439	NADH dehydrogenase	Leaf/Root	A B D	D	8E-49	202 bits
BE426364	Glyceraldehyde-3-phosphate dehydrogenase	Leaf/Root	A D	A	3E-42	180 bits
BE404371	NADH glutamate dehydrogenase	Root	B D	B	4E-19	103 bits
BE518213	Putative ABC transporter	Root	B D	D	6E-11	75.8 bits
BE498622	Unknown	Leaf	A B	A	–	–
BE499478	FAT domain-containing protein/phosphatidylinositol 3- and 4-kinase family protein	Root	B*	B	5E-16	93.7 bits
BE445693	Unknown	Leaf	A D	D	–	–
BE637228	Unknown	Root	A B	A	–	–
BF145580	Senescence-associated putative protein	Root	A B	B	1E-07	65.9 bits
BE399113	Unknown	Leaf/Root	B D	B D	–	–
BE406808	Reversibly glycosylated polypeptide (*rgp1* gene)	Leaf/Root	B D	B	3E-15	89.7 bits
BE496834	Peptidylprolyl isomerase	Leaf	A B	A	2E-144	519 bits
BF201129	Plastid division protein FtsZ	Leaf/Root	A B D	A/D	0	965 bits
BF473379	Unknown	Leaf	B D	D	–	–
BE443527	Unknown	Leaf/Root	A B D	B/B	–	–
BE443833	Unknown	Leaf/Root	A B D	D/A	–	–
BE444894	Saline-responsive OSSRIII protein	Leaf/Root	A B D	D/B	3E-64	254 bits
BE591763	Unknown	Leaf	A B D	A	–	–
BF201235	Rubisco subunit binding protein α subunit	Leaf	A B D	D	0	3124 bits
BF202681	Unknown	Root	A B	B	–	–
BF484100	Unknown	Leaf	A B D	D	–	–
BF485266	Unknown	Leaf	A D	A	–	–
BF483382	Unknown	Leaf	A D	A	–	–
BF145484	Unknown	Leaf/Root	A B D	D/D	–	–
BF478825	Unknown	Leaf	A B D	D	–	–
BE406148	Cyclophilin	Leaf/Root	A B D	B/B	4E-13	83.8 bits
BE404724	Polyadenylate-binding protein	Root	A B D	D	1E-10	75.8 bits
BE404737	Putative *S*-adenosylmethionine: 2-demethylmenaquinone methyltransferase	Leaf	D*	D	2E-07	65.9 bits
BM138650	Monodehydroascorbate reductase	Leaf	A B D	A	0	737 bits
BM138439	Unknown	Root	A B D	D	–	–
BF201019	Unknown	Leaf	A B D	B	–	–
BE606458	Sorbitol dehydrogenase	Leaf	A B D	A	4E-16	93.7 bits
BE606217	Putative DegP protease	Leaf/Root	A B D	A/A	7E-23	117 bits
BE489493	Unknown	Leaf	A D	D	–	–
BE443396	Glucose-6-phosphate/phosphate translocator	Leaf	A D	D	0	1130 bits
BE443540	Unknown	Leaf	A B	B	–	–
BE443554	Unknown	Leaf/Root	A*	A	–	–
BE445506	Unknown	Leaf/Root	A B D	D	–	–
BE446012	Unknown	Leaf/Root	A D	D/D	–	–
BE495400	Unknown	Root	A B D	B	–	–
BF428786	Unknown	Leaf	A D	D	–	–
BE604737	Unknown	Leaf/Root	A B D	B/B	–	–
BF201560	DRP1 protein	Leaf/Root	A D	A/D	0	1370 bits
BF483361	Unknown	Leaf	A D	A	–	–
BF482273	Unknown	Leaf/Root	B D	D/B	–	–
BG263922	Expressed protein	Leaf	B D	B	1E-42	182 bits
BG274853	FtsH-like protein	Leaf	A B D	D	9E-116	424 bits
BF484735	Unknown	Root	A D	D	–	–
BG607581	Phosphatidylinositol 3- and 4-kinase-like protein	Root	A B D	A	2E-18	103 bits
BI479219	Unknown	Root	A D	D	–	–
BE404461	Unknown	Root	A B D	A	–	–
BE497566	Unknown	Root	A B D	B	–	–
BE586090	OsNAC8 protein	Root	A B D	B	3E-132	482 bits
BE423249	Unknown	Leaf	A B D	D	–	–
BE517931	Unknown	Leaf	A B D	D	–	–
BF200563	Unknown	Leaf	A*	A	–	–
BF473348	Unknown	Leaf	A B D	B	–	–

Table 1 Continued

Genbank identifier	Function/putative function	Tissue silenced	Genomes identifiable	Genome silenced	E number	Score (bits)
BG604766	1,3-beta-glucan synthase	Root	A B	B	0	918 bits
BE499982	Holocarboxylase synthetase	Leaf	A*	A	0	1037 bits
BF202265	Unknown	Leaf	A*	A	–	–
BE494911	Unknown	Leaf	A B D	A	–	–
BE438771	Glycosyltransferase	Root	A B D	A D	8E-38	165 bits
BE438866	Unknown	Root	A B D	A	–	–
BE425898	Unknown	Leaf	A B D	B	–	–
BE438469	ADP-glucose pyrophosophorylase	Leaf	A B D	B	E-168	597 bits
BE442851	Unknown	Leaf	B D	B	–	–
BE443905	Unknown	Leaf	A B D	D	–	–
BE471203	Unknown	Leaf	A D	D	–	–
BM136937	Unknown	Leaf	A D	D	–	–
BM138152	Putative peptide methionine	Leaf	B D	D	4E-24	119 bits

*Additional copy(ies) mapping outside of the expected homoeologous group present.
– indicates no significant BLAST results.

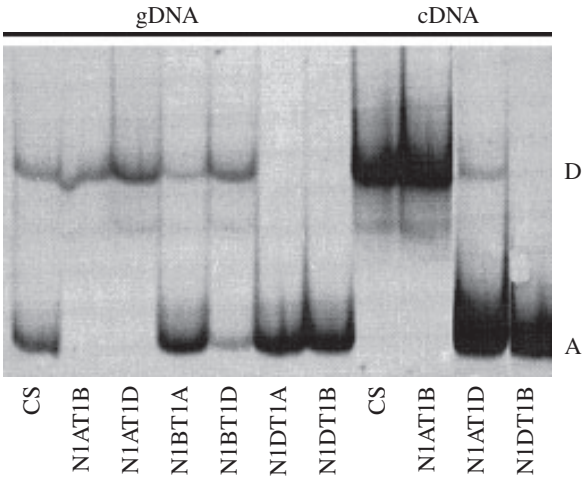

Figure 3. The 1A homoeoallele silenced in CS seedling leaf is transcribed in the aneuploid derivatives N1BT1D and N1DT1B for EST BE494911. The genomic origin of each PCR product is indicated on the right.

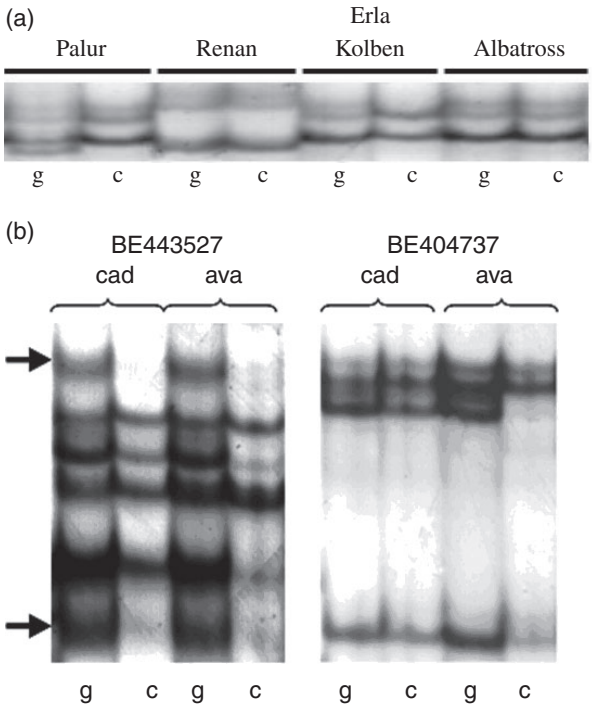

Figure 4. (a) Genotypic variation for silencing in 21-day-old seedling leaf for EST BF202265. In two accessions (Renan, Albatross), all genomic loci are expressed, while in the other two (Palur, Erla Kolben), a different genomic copy is silenced. g, genomic DNA; c, cDNA.
(b) RT-PCR profiles of two single-copy EST sequences for varieties Cadenza (cad) and Avalon (ava). For BE443527, the D genome copy (arrows) is absent from the cDNA (c), but not from the gDNA (g) profiles of both varieties; for BE404737, absence of the D genome copy is only observed in the Ava c profile (arrow).

likely to have evolved from ancient polyploids by a process of pseudogene formation followed by sequence loss. In a study of the fate of duplicated genes in the maize genome, Lai *et al.* (2004) have suggested that, in as little as five million years, about 50% of duplicated genes have been lost through deletion. Nonetheless, gene duplication in maize *per se* via (auto)polyploidization may be associated with detectable increases in expression level, as demonstrated by Guo *et al.* (1996).

The availability of characterized aneuploid lines gives the unique opportunity to assign the genomic origin of the components of the RT-PCR amplicon. It is thus possible to estimate the relative contribution each genome makes to the full transcription profile, and whether the gene product of one genome is more frequently unrepresented than expected. It has been documented that homoeoalleles are not necessarily each expressed at an equivalent level (e.g. Nomura *et al.*, 2005). Genome-wide, as opposed to for an individual

set of homoeoalleles, our data suggest that, for leaf transcripts, there is a modest bias towards silencing of the D genome copies, but this pattern does not extend to root transcripts. The existence of such a bias is controversial. He *et al.* (2003) found D genome homoeoalleles to be silenced twice as frequently as those from the A or B genomes in a newly synthesized wheat hexaploid line, but there was no pattern of preferential silencing in the newly synthesized wheat diploid × diploid *Aegilops* hybrid studied by Kashkush *et al.* (2002). In an analysis of the natural hexaploid by Mochida *et al.* (2004), there was little evidence for any genomic bias among the 11 loci exhibiting homoeoallele silencing. Interestingly, in cotton, Adams *et al.* (2004) suggested a two-to-one bias in silencing against the G genome in a newly synthesized AG tetraploid, but no such bias in the natural AD tetraploid. These data suggest that genomic bias may decay over time, as the component genomes become integrated with one another, which would predict a lack of bias with respect to the A and B genome copies in natural tetraploid wheat, but some residual negative selection against D genome copy transcription in the hexaploid. However, our data, like those of Kashkush *et al.* (2002) and Mochida *et al.* (2004), are not supportive of such a model.

The preferential expression of homoeoalleles in specific plant organs has been explored in cotton by Adams *et al.* (2003), and *in silico* for wheat by Mochida *et al.* (2004). The latter study showed that, of 54 genes represented in pistil EST libraries, over half showed selective expression from a particular genome, but a much lower proportion behaved this way in the emerging spike. Overall, only about 20% of the genes sampled were uniformly expressed from all three genomes across all ten organs sampled. How and why genomes are preferentially expressed in different tissue types is unknown, and an understanding of this may give some insight into both the mechanism and function of polyploidy-related silencing. We speculate that it reflects the preliminary stages of sub-functionalization through mutagenesis, which manifests itself as changes at one homoeoallele, involving the retention of coding sequence but the mutation of transcription binding sites, thereby changing the process of initiation of expression to one directed to the requirements of a particular tissue type.

We have presented here a body of data that establishes the reality of silencing of homoeoalleles in hexaploid wheat, by demonstrating the absence of an mRNA output from loci that are apparently intact (inasmuch as the target sequence is amplifiable) in the genomic DNA. The extent of the phenomenon is not clear, but some indication of its impact on inheritance patterns can be gauged from consideration of the genetic basis of traits that have been built up over many years of genetic experimentation. At the outset of the study of Mendelian inheritance in wheat, as in other organisms, monogenic characteristics were sought. Certainly the number of such characteristics is less favourable in wheat than in diploid species (which led naturally to the establishment of barley as an early 'model' species for the cereals), but many of the most critical genes involved in domestication and improvement do fall into this class. Predominant examples are the three genes that together define the major hexaploid wheat types: *Q* (speltoid, located on chromosome arm 5AL; Mackey, 1954), *C* (*compactum*, located on chromosome arm 2DL; Unrau, 1950) and *s1* (*sphaerococcum*, located on chromosome arm 3DL; Schmidt *et al.*, 1963), and the three genes responsible for awn formation (*Hd*, *B1* and *B2*, on 4AS, 5AL and 6BL, respectively; Sears, 1954). Although, of the above, only *Q* has as yet been characterized at the DNA level (Faris *et al.*, 2003), there is no evidence, based on either natural or induced genetic variation, for the presence of functional homoeoalleles for any of them. The chromosome pairing control locus *Ph1* (5B) has recently been shown to lack equivalents at its homoeologous sites, as it has arisen from a unique insertion event (Griffiths *et al.*, 2006). On the other hand, a number of 'morphological' genes are present as three, or at least two, members of a probable homoeoallelic series. Thus, for example, the red coleoptile genes *Rc* have functional copies on all three group 7 homoeologues (Gale and Flavell, 1971; Rowland and Kerber, 1974), as do the *R* genes responsible for red testa colour, present on each of the group 3 homoeologues (Metzger and Silbaugh, 1970; Sears, 1954). The Green Revolution semi-dwarfing genes *Rht1* and *Rht2* represent homoeoalleles on chromosomes 4B and 4D (Gale and Marshall, 1975; Gale *et al.*, 1975), and cloning of the gene has since established that a copy is also present on 4A (Peng *et al.*, 1999), although functional dwarf mutants have yet to be identified at *Rht-A1*. Similarly, *v* (virescent), *w* (leaf glaucousness) and *D* (grass clump) genes have been mapped to two out of their respective three homoeologues (Sears and Sears, 1968; Tsunewaki, 1966; Worland and Law, 1980).

Perhaps one of the more surprising results generated by this study has arisen from determination of the expression profiles of the aneuploid lines, for which transcripts absent in the euploid profile, despite the known presence of exon sequence, were restored in one or more nullisomic–tetrasomic lines. A possible basis for this phenomenon is the presence of repressor genes that are lost in a specific nullisomic state, but this cannot explain two-thirds of the instances observed among the homoeologous group 7 loci, where restoration is not confined to genotypes lacking a specific chromosome. Departure from disomy of sub-chromosomal segments in maize generates variation in the transcript level of genes mapping both within and outside the critical genomic region (Guo and Birchler, 1994), and could therefore underlie transcription reactivation in aneuploid genotypes of wheat, as the nullisomic–tetrasomic lines differ from euploid with respect to dosage of both the nullisomic chromosome (zero doses) and the tetrasomic chromosome (four doses). However, we have been unable to

establish any reliable association between the identity of the high-dose chromosome and the restoration of expression, and, in any case, our observations relate to extreme levels of repression, rather than the level of quantitative changes that led to the idea that a substantial proportion of phenotypic variation is generated by polymorphism in such *trans*-acting effects.

Our preferred view is that epigenetic modification is the likely mechanism for the majority of this silencing, as the aneuploid lines were bred from the euploid parent within the last 50 years. Kashkush *et al.* (2002) have suggested that C-methylation is responsible for silencing of about half of the transcripts not lost as a result of genome rearrangements, but this estimate was made on the strength of differential Southern profiles obtained from *Hpa*II and *Msp*I restriction, a method that can only identify methylation polymorphism associated with bases in the restriction recognition site, and thus tends to underestimate overall methylation levels. So, although C-methylation (and hence epigenetic silencing) is generally faithfully inherited across generations, it has been noted that it can vary around the wheat *Glu-1* High Molecular Weight (HMW glutenin) locus between individuals of a specific wheat variety (Flavell and O'Dell, 1990). Varieties of wheat naturally vary among themselves at the sequence level, and this variation is assumed to be the basis of their variation in phenotype. Our demonstration that varieties also display variation in patterns of homoeoallele silencing strongly suggests that some phenotypic variation may be generated by epigenetic differences. We propose, therefore, that because a measurable proportion of the gene content is not expressed, or is expressed in an organ-specific manner, that a significant component of the genetic variation that underpins breeding is related to non-sequence-based differences between genotypes. The frequency of sequence variation in genic DNA of wheat is known to be low (Bryan *et al.*, 1999), and so the possibility arises that a significant part of the phenotypic variability, so successfully exploited worldwide by breeders over many years, is governed by epigenetic, rather than by genetic polymorphism.

Experimental procedures

Plant materials, RNA extraction, cDNA synthesis and Phi amplification

The following wheat stocks and varieties were studied: (1) Chinese Spring (CS) wheat, and the set of nullisomic–tetrasomics in homoeologous groups 1, 2, 3 and 7 developed by Sears (1966) and maintained at the John Innes Centre (Norwich, UK) by S. Reader, and (2) a panel of varieties selected on the basis of marker classification to be representative of the diversity of European winter wheat. These were Avalon, Palur, Renan, Erla Kolben, Albatross, Soissons, Hereward, Cadenza and the durum (tetraploid) variety Regina. Genomic DNA from all lines was obtained from curated stocks. RNA was extracted from leaf and root material harvested from four or five 21-day-old seedlings per genotype, grown in vermiculite under controlled temperature and lighting conditions.

Plant tissue was snap-frozen in liquid N_2, and total RNA was extracted using the Trizol™ reagent (Sigma, Gillingham, Dorset, UK), following the manufacturer's instructions. Crude total RNA preparations were treated with DNase (Amersham Bioscience, Chalfont, Bucks, UK) and phenol/chloroform-extracted. The presence/absence of contaminating genomic DNA was tested by trial PCR with a range of appropriate PCR primers, and the quantity and quality of RNA present were assessed on agarose gels, by comparison with a quantitative RNA standard. cDNA was synthesized using Superscript II™ (Invitrogen, Paisly, Renfrewshire, UK), using oligo(dT) as the polyA primer and following the manufacturer's protocol. Newly synthesized cDNA was again tested with a range of PCR primers spanning introns to detect any carry-through of gDNA. To supply sufficient cDNA template for the number of assays required for the CS-based materials, we pre-amplified this cDNA using the GenomiPhi™ kit (Amersham Bioscience) as per the manufacturer's instructions, and established the fidelity of this pre-amplification by comparing amplicon profiles (generated from ten independent gene targets – see Table S1) derived from Phi-amplified cDNA with those derived from a non-amplified stock of the same cDNA.

EST selection, primer design, PCR amplification and SSCP analysis

Unigene ESTs mapping exclusively to a set of loci on one of homoeologous chromosome groups 1, 2, 3 or 7 were identified on the basis of Southern blot patterns, as reported at http://wheat.pw.usda.gov/NSF/. Their distribution among the chromosomal bins was in similar proportion to the EST collection as a whole. In order to establish flanking sequence, each EST was allocated, by BLAST algorithm, to a specific wheat EST contig (http://wheat.pw.usda.gov/ITMI/WheatSNP/blastWheatSNP.html). The matching rice sequence, both genomic and cDNA, was then identified by BLAST, to allow a prediction of both intron position and its approximate size in wheat, assuming a reasonable level of conservation across this species comparison. On the basis of these predictions, primers were designed to amplify intron-free wheat sequence of size 200–250 bp, using Primer3 software (Rozen and Skaletsky, 2000). As the intention was to co-amplify all homoeoallelic copies, we made no attempt to derive genome-specific primers. A list of all EST loci assayed, along with their associated primer sequences, is given in Table S1.

Phi-amplified cDNA was diluted 1:20, and 1 μl of this dilution was used as template for a 10 μl reaction, together with 5 μl of Hotstar Master Mix™ (Qiagen, Crawley, West Sussex, UK), 3.5 μl of water and 0.25 μl of each primer (10 mM concentration). PCR consisted of 15 min at 95°C (to heat-activate the Taql enzyme and denature the template), followed by 35 cycles of 95°C for 30 sec, 59°C for 45 sec and 72°C for 60 sec, and completed with a single extension step of 10 min at 72°C. Amplicons were electrophoretically separated by the SSCP method, for which MDE (Cambrex, Wokingham, Berkshire, UK) gels containing 12% glycerol were prepared according to the manufacturer's instructions. Prior to loading, samples were denatured by 1:1 dilution in formamide-containing loading buffer. Gels were run for 18 h at 3–5 W at 4°C. The power setting was optimized for each amplicon depending on its electrophoretic mobility. Glycerol-tolerant TTE (National Diagnostics, Hessle, Yorkshire, UK) electrophoresis running buffer was used both as part of the gel mixture and as running buffer. Band profiles were visualized using silver staining as described by Bassam *et al.* (1991).

Pattern analysis

To study the patterns of gene expression within each set of single-copy homoeoalleles, we compared SSCP profiles generated from genomic DNA with the equivalent from cDNA templates. If all copies identified in the gDNA template were also present in the cDNA profile, this was taken as an indication of full expression of all homoeoalleles. If, despite the presence of a gDNA locus, no cDNA was detected, we inferred that the relevant locus was 'silenced'. We recognize that the assay assesses the presence/absence of steady-state transcript(s), and that specific products may be absent either because of transcription failure and/or because of rapid degradation or cleavage of specific transcripts. However, we have chosen to use the term 'silenced' to retain consistency with cognate published material.

No attempt was made to quantify band intensity as we believe that this can be misleading, because amplification often distorts the ratio of the relevant molecules present in the original template. For every set of homoeoalleles where silencing was indicated, triplicate assays were carried out to ensure the reliability and reproducibility of the result.

Acknowledgements

AB acknowledges the financial support of a BBSRC studentship. We particularly thank the US National Science Foundation (NSF) wheat EST project (http://wheat.pw.usda.gov/NSF/) for their generous assistance with access to the wheat EST database.

Supplementary Material

The following supplementary material is available for this article online:

Table S1 A detailed list of ESTs analysed

This material is available as part of the online article from http://www.blackwell-synergy.com

References

Adams, K.L., Cronn, R., Percifield, R. and Wendel, J.F. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl Acad. Sci. USA* **100**, 4649–4654.

Adams, K.L., Percifield, R. and Wendel, J.F. (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics*, **168**, 2217–2226.

Bassam, B.J., Caetano-Anolles, G. and Gresshoff, P.M. (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196**, 80–83.

Bryan, G.J., Stephenson, P., Collins, A., Kirby, J., Smith, J.B. and Gale, M.D. (1999) Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. *Theor. Appl. Genet.* **99**, 192–198.

Comai, L. (2000) Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* **43**, 387–399.

Comai, L., Tyagi, A.P., Winter, K. *et al.* (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell*, **12**, 1551–1567.

Cronn, R.C., Small, R.L. and Wendel, J.F. (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl Acad. Sci. USA*, **96**, 14406–14411.

Faris, J.D., Fellers, J.P., Brooks, S.A. and Gill, B.S. (2003) A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. *Genetics*, **164**, 311–321.

Flavell, R.B. and O'Dell, M. (1990) Variation and inheritance of cytosine methylation patterns in wheat at the high molecular-weight-glutenin and ribosomal-RNA gene loci. *Development* (Suppl. S), 15–20.

Forsström, P.O., Koebner, R. and Merker, A. (2003) The conversion of wheat RFLP probes into STS markers via the single-stranded conformation polymorphism technique. *Genome*, **46**, 19–27.

Gale, M.D. and Flavell, R.B. (1971) Genetic control of anthocyanin biosynthesis by homoeologous chromosomes in wheat. *Genet. Res.* **18**, 237–244.

Gale, M.D. and Marshall, G.A. (1975) Nature and genetic control of gibberellin insensitivity in dwarf wheat grain. *Heredity*, **35**, 55–65.

Gale, M.D., Law, C.N., Marshall, G.A. and Worland, A.J. (1975) Genetic control of gibberellic-acid insensitivity and coleoptile length in a dwarf wheat. *Heredity*, **34**, 393–399.

Grant, V. (1981) *Plant Speciation*. New York: Columbia Univ. Press.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid locus in polyploid wheat. *Nature*, **439**, 749–752.

Guo, M. and Birchler, J.A. (1994) Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science*, **266**, 1999–2002.

Guo, M., Davis, D. and Birchler, J.A. (1996) Dosage effects on gene expression in a maize ploidy series. *Genetics*, **142**, 1349–1355.

He, P., Friebe, B.R., Gill, B.S. and Zhou, J.M. (2003) Allopolyploidy alters gene expression in the highly stable hexaploid wheat. *Plant Mol. Biol.* **52**, 401–414.

Huang, S., Sirikhachornkit, A., Su, X.J. *et al.* (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl Acad. Sci. USA*, **99**, 8133–8138.

Kashkush, K., Feldman, M. and Levy, A. (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics*, **160**, 1651–1659.

Lai, J.S., Ma, J.X., Swigonova, Z. *et al.* (2004) Gene loss and movement in the maize genome. *Genome Res.* **14**, 1924–1931.

Liu, B., Brubaker, C.L., Mergeai, G., Cronn, R.C. and Wendel, J.F. (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome*, **44**, 321–330.

Mackey, J. (1954) Neutron and X-ray experiments in wheat and a revision of the speltoid problem. *Hereditas*, **40**, 65–180.

Mcintosh, R.A., Hart, G.E., Devos, K.M., Gale, M.D. and Rogers, W.J. (1998) *Catalogue of Gene Symbols for Wheat*.

Metzger, R.J. and Silbaugh, B.A. (1970) Location of genes for seed color in hexaploid wheat. *Crop Sci.* **10**, 495–496.

Mochida, K., Yamazaki, Y. and Ogihara, Y. (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics*, **270**, 371–377.

Nomura, T., Ishihara, A, Yanagita, R.C., Endo, T.R. and Iwamura, H. (2005) Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc. Natl Acad. Sci. USA*, **102**, 16490–16495.

O'Neill, C.M. and Bancroft, I. (2000) Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J.* **23**, 233–243.

Peng, J.R., Richards, D.E., Hartley, N.M. *et al.* (1999) 'Green revolution' genes encode mutant gibberellin response modulators. *Nature*, **400**, 256–261.

Rowland, G.G. and Kerber, E.R. (1974) Telocentric mapping in hexaploid wheat of genes for leaf rust resistance and other characters derived from *Aegilops squarrosa*. *Can. J. Genet. Cytol.* **16**, 137–144.

Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. and Misener, S., eds). Totowa, NJ: Humana Press, pp. 365–386.

Schmidt, J.W., Weibel, D.E. and Johnson, V.A. (1963) Inheritance of an incompletely dominant character in common wheat resembling *Triticum sphaerococcum*. *Crop Sci.* **3**, 261–264.

Sears, E.R. (1954) The aneuploids of common wheat. *Missouri Agric. Exp. Sta. Res. Bull.* **572**, 59.

Sears, E.R. (1966) Nullisomic–tetrasomic combinations in hexaploid wheat. *Heredity*, **20**, S29–S45.

Sears, L.M.S. and Sears, E.R. (1968) The mutants chlorina-1 and Hermsen's virescent. In *Proceeding of the 3rd International Wheat Genetics Symposium* (Findlay, K.W. and Shepherd, K.W., eds). Canberra: Australian Academy of Science, pp. 299–304.

Song, K.M., Lu, P., Tang, K.L. and Osborn, T.C. (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proc. Natl Acad. Sci. USA*, **92**, 7719–7723.

Tsunewaki, K. (1966) Comparative gene analysis of common wheat and its ancestral species. *Jpn. J. Bot.* **19**, 175–254.

Unrau, J. (1950) The use of monosomes and nullisomes in cytogenetic studies in common wheat. *Sci. Agric.* **30**, 66–89.

Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**, 708–713.

Worland, A.J. and Law, C.N. (1980) The genetics of hybrid dwarfing in wheat. *Z. Pflanzenz.* **85**, 28–39.

