Fate of duplicated genes in polyploid wheat

Mutti JS and Gill KS
Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420

INTRODUCTION

Polyploidy or whole genome duplication has been proposed as the primary force shaping the evolution of flowering plants. In addition to the 50 to 70% of the flowering plants that are believed to be polyploids, (Masterson, 1994) DNA sequence analyses have suggested that even the diploid organisms such as *Arabidopsis*, yeast and humans may in fact have undergone cycles of polyploidization.

The fate of duplicated genes following polyploidization may include silencing of one of the duplicated copies (*nonfunctionalization)*, divergence of duplicated genes to result in new functions (*neofunctionalization),* reduction of the combined expression of the two copies to the level of a single copy gene, or acquisition of different tissue-specificities for gene expression (*subfunctionalization)* (Lynch and Force 2000). Major consequences of polyploidization reported in different polyploids range anywhere from sequence elimination, chromosomal translocations and inactivation of genes in some plants, to appearance of novel sequences and sub-functionalization of gene expression in others (Chen, 2007). These unexpected major differences in the fate of duplicated genes among different plants may be a result of different approaches taken to study the effects or are based on the study of very few genes. Thus, a comprehensive analysis presenting the fate and subsequent balance upon polyploidization in naturally occurring polyploids is needed.

The effects of polyploidization on gene expression has been studied for synthetic polyploids, but the concerted fate of homoeologues in natural polyploids, the interdependence of homoeologous gene expression and the corresponding mechanisms controlling these processes, remain largely unexplored. In this study, we chose polyploid wheat to address these questions. Wheat, *Triticum aestivum* L. ($2n = 6x = 42$), is a natural allohexaploid with three relatively collinear genomes, designated A, B, and D. Two independent polyploidization events occurring at different times: ~0.5 and 0.01 million years ago (MYA) led to the evolution of present day hexaploid wheat (Huang et al. 2002). Furthermore, hexaploid wheat also has a wealth of aneuploid stocks available that are ideal to understand the functional organization of the wheat genome (for example to study homoeologous gene expression balance) and to analyze the effect of gene dosage on gene expression.

MATERIALS AND METHODS

Wheat Nullisomic-tetrasomic (NT) lines and Ditelosomic (DT) lines produced in cultivar 'Chinese spring' (CS) (Sears, 1954), along with cultivar CS, grown under greenhouse conditions, were used for various experiments. Gene expression corresponding to each homoeologue was identified by sequence comparison of cultivar CS ESTs and the results were confirmed by single stranded conformation polymorphism (SSCP) analysis of RNA using nulli-tetra lines. The methodology of sequence comparison is explained below in the results section. For SSCP analysis, plants were grown under highly controlled conditions in a growth chamber. The poly(A)$^+$ RNA from eleven different plant development stages, namely five day old seedling, root from seedling and adult plants, 28-day old plant, flag leaf, early flowering (Feekes scale: 6), meiosis, pre-anthesis, post-anthesis, Seed at 5 days post-germination (DPA) and 30 DPA stage and adult plant (Feekes scale: 10.5), was extracted using standard methods. SSCP analysis was optimized to resolve fragment bands corresponding to each of the three homoeologues. The PCR reactions were performed on the first stranded cDNA with Advantage® PCR Kits & Polymerase Mixes (Clontech, Catalog #639101), in the presence of S^{35}dATP in a total volume of 20µl. The PCR product was mixed with an equal volume of loading buffer. About 5µl of this mixture was loaded onto 0.4mm thick denaturing 8% polyacrylamide/8M urea gels (Sambrook et al. 1989). Each sample was size separated both on gels run under standard conditions as well as on gels run for SSCP. For standard runs, the gels were pre-run at a 33mA constant current for 30 min and then at 70W constant power for 4 hours. For SSCP runs, the gels were pre-chilled at 4°C for at least 5-6 hr before running it at 10 W for 12-13 hrs at 4°C. An X-ray film was placed on the dried gels and was exposed for three to seven days.

RESULTS

Gene expression corresponding to each homoeologue was studied by sequence comparison of cultivar 'Chinese spring' (CS) ESTs. Of the 9,400 ESTs and genes that have been physically mapped to wheat chromosomes (http://wheat.pw.usda.gov/NSF/), gel-blot analysis images of 6000 that detected less than six bands, were evaluated to select 854 ESTs that showed a clear hybridization pattern and where every restriction fragment band was physically mapped. The selected ESTs were compared with the full-length rice cDNAs and the rice homologues were identified by pairwise comparison using the Blast algorithm at a cut off e-value of e^{-70} (KOME, http://cdna01.dna.affrc.go.jp/cDNA/). The selected wheat ESTs and the rice full-length cDNAs were compared with a 'Chinese spring' (CS) wheat specific EST database using the megablast algorithm (http://www.ncbi.nlm.nih.gov/BLAST/) and were aligned with the 'ContigExpress' Module of the *Vector NTI Advance*™ 8.0 software. A cutoff value of ≥80% identity within minimum overlap of 30 bp was used for

wheat-rice comparisons. These two comparisons yielded CS wheat EST contigs for each of the selected genes. In order to account for sequencing errors, a sequence pattern was considered to be unique only if it was present in two or more ESTs and the differences/similarities among the ESTs within a pattern is consistent along the length of the assembled contig.

An example of the analysis and the approach to identify ESTs corresponding to each homoeologue is given in Figure 1a. For example, the EST *BE497160* (marked red) was physically mapped between FL 0.71 and 1.00 on the long arm of wheat homoeologous group *4* (http://wheat.pw.usda.gov/NSF/). The corresponding gene has three structural copies in the wheat genome as the deletion mapping detected three fragment bands, one

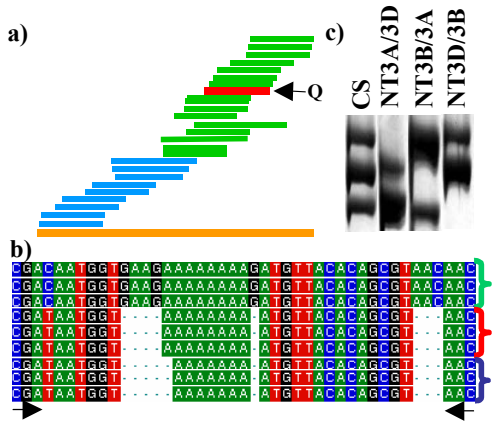

Figure 1. (a) Contig assembly of wheat homolog EST's. Green bars denote the EST's identified using wheat EST as a query (Q), whereas blue bars represents the additional wheat EST's identified for the gene using rice full-length (orange bar) cDNA sequence.. **(b)** Sequence alignment of the EST's within the contig. ESTs corresponding to each of the three wheat homoeologues were distinguished based on different patterns as bracketed on right side by green, red and blue brackets. Brackets of different color represent EST's of different patterns, whereas bracket with same color represent EST's in one pattern. Black arrows on the bottom represents the conserved region among the three patterns with no sequence differences, used to design primers for SSCP analysis. **(c)** *In silico* results confirmed by cDNA-SSCP for gene *BE586090.* CS denotes chinese spring; NT3A\3D, NT3B\3A, NT3D\3B indicates the nulli-tetra lines for group *3* chromosome. Gel electrophoresis pattern indicates three structural copies of the gene, three expresses.

on each of the group *4* homoeologues. Upon its megablast comparison with the CS EST database, 24 ESTs with more than 80% homology (green bars) were identified. A similar comparison using the full-length rice cDNA homologoue (Figure 1a, orange bar) identified an additional 11 wheat ESTs (Figure 1a, blue bars) to extend the contig across the entire gene. Since this gene was represented by three distinct cDNA sequence patterns and all ESTs were from the same cultivar 'Chinese Spring', we conclude that all three copies of this gene are expressed (Figure 1b). This approach was used to study the expression pattern of the 854 putative wheat genes.

Out of the 854 genes, contigs for 309 (36%) had less than five ESTs, thus were not included in the analysis. Three structural copies were identified for 632 of the 854 genes. Of these all three homoeologues were expressed for 55% , with a further 36% expressed from two, and 16% from only one.

In order to confirm the *in silico* analysis results, cDNA-SSCP and standard acrylamide/urea gel analysis was performed for 31 randomly selected genes representing a range of structural copy number and expression patterns. This analysis was performed using RNA of various aneuploid stocks mixed in equal proportion from eleven major developmental stages (see material and methods). Primers for the SSCP analysis were designed from the conserved regions flanking the sequences differentiating the homoeologues. An example of the SSCP analysis is given in Figure 1c. The gene *BE586090* has three structural copies one each on the three homoeologous group *3* chromosomes (http://wheat.pw.usda.gov/NSF/). The *in silico* expression analysis showed three distinct sequence patterns suggesting that all three copies are expressing (Figure 1b). The SSCP analysis of CS and group *3* nulli-tetra (NT) lines showed three bands in CS corresponding to each of the three group *3* homoeologues (Figure 1c).

Both standard acrylamide/urea gels as well as SSCP gels were run for the 31 randomly selected genes. For 13 genes the SSCP gels worked better as it resolved all of the bands where as for 18 genes the standard gels resolved the bands better. Between these two types of gels, all bands for 18 genes were resolved by the NT lines and the results of number of homoeologues expressed matched with that from the *in silico* analysis.

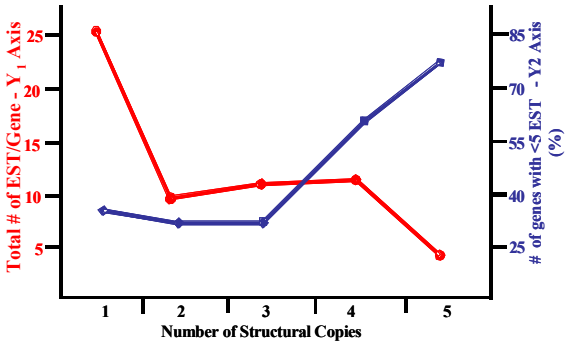

Figure 2. Total number of ESTs identified per gene (Y_1-axis) and number of genes with <5 ESTs identified (Y_2-axis). X-axis gives the structural copy number of the gene physically mapped to three homoeologous chromosomes of wheat. Red curve gives the distribution of total number of ESTs identified per gene per structural copy number, where as blue curve represents number of genes with <5 ESTs identified per structural copy.

A summary of *in silico* expression analysis results for 854 wheat genes is given in Table 1. It was interesting to notice that genes with a higher copy number showed a higher proportion of genes with less than five ESTs. This proportion was highest (43/56, 76%) for genes with five structural copies followed by genes with four copies (46/76, 60%) (Figure 2). The least proportion showing less than five ESTs were the genes with three structural copies (194/632, 30%). The number of ESTs identified per structural copy also showed a range, with the least number of EST/gene (4) for genes with five structural

copies and a maximum number in genes with one structural copy (30). The fraction of expressed copies was generally found to be inversely related to the overall number of structural copies in the genome as none of the genes with five structural copies expressed from all five (Table 1). Fifty-four percent of these expressed from four copies, 23% from three, 15% from two and only 8% from one of the five copies. In the case of genes with four structural copies, a larger proportion (23%) of the genes was expressed from one locus where as in the case of genes with three copies only 8.6% expressed from one locus. In general, the proportion of expressed copies decreased with the increase in homoeologue copy number.

Table 1. Expression pattern of homoeologues in hexaploid wheat**.**

# of Structural copies	# of Genes	Number of copies expressed					
		5 ESTs	4	3	2	1	<5
5	56	-	7	3	2	1	43
4	76		11	3	9	7	46
3	632			254	146	38	194
2	50				26	8	16
1	40					30	10
Total	**854**		**18**	**260**	**183**	**84**	**309**

DISCUSSION

Wheat is one of the most important cereal crops used for human consumption and thus any genetic insights into its genome can eventually be translated into crop improvement. Our *in silico* approach to identify ESTs corresponding to each of the three wheat homoeologues seems to be reliable as the standard acrylamide/urea gel or the SSCP analysis confirmed the *in silico* results. The gel analysis for 18 of the 31 genes resolved all of the fragments and the results matched with that observed with the *in silico* analysis. For the remaining 13 genes not all fragment bands were resolved by the nulli-tetrasomic lines with either of the two gel running approaches but the number of the resolved fragment bands were equal or less than the expected number. Only for three of the 31 genes was the observed number of bands more than expected. This could be due either to alternate splicing, or the primer sequences for these genes may be present in some other unrelated genes that were not detected by the gel blot analysis during physical mapping.

We are reporting that 55% of the wheat genes are expressing from all three homeologues, 36% from two, and 16% from only one. Based on the analysis of 116,232 ESTs, Mochida et al. 2003, reported that 26% of the wheat genes expressed from all three loci, 46% expressed from two and 21% expressed from one of the three homoeologues. In our analysis, a more comprehensive set of ESTs (268,582) from 41 cDNA libraries representing 14 different wheat tissues or developmental stages were included- which in turn is very important for analysing gene expression pattern as a lower number of EST datasets can give misleading results. The percentage of genes expressing from only one homoeologue can be over represented if comprehensive EST datasets from various tissues and developmental stages is not utilized. Another disadvantage of analysing homoeologous gene expression based merely on ESTs is- for genes exceeding 1 kbps in size, two non-overlapping EST contigs corresponding to the same gene may be counted more than once. In our analysis the full-length rice cDNA sequences were used to help assemble the full-length wheat contigs, such that two smaller ESTs contigs were assembled into one for more accurate analysis of homoeologue expression. Along with this we also included structural copy number for an accurate estimation of the number of homoeologues/gene expressed.

Expression of homoeologues in wheat is balanced with most of the homoeologues expressing at all times. More than one copy of a gene is expressed for the most part with silencing of two or more homoeologues kept at a minimum. Based on a sample of 652 with three structural copies, 91% of the wheat genes expressed from two or three copies with only 9% showing silencing of two copies. A balance of homoeologue expression was also observed, as any deviation in copy number of the gene resulted in altered expression levels. For example, on an average gene with five structural copies identified, four ESTs/gene compared to 30 ESTs/gene for genes with one structural copy. A balance of expression was seen in the case of genes with three structural copies, expressing 12 ESTs/gene. Any deviation in genomic copy number leads either to increase in expression of the gene (as in case of genes with one structural copy) or decrease in expression with increase in copy number (as in case of genes with five structural copies.

REFERENCES

1 Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annual Review of Plant Biology* **58:** 377–406.

2 Huang, S., A. Sirikhachornkit, X. Su, J. Faris, 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. *Proceedings of the National Academy of Sciences* **99:** 8133-8138.

3 Lynch, M. and A. Force. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151-1155.

4 Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264:** 421-424.

5 Mochida, K., Y. Yamazaki, and Y. Ogihara. 2003. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs

grouped from a large number of expressed sequence tags. *Molecular Genetic Genomics* **270:** 371-377.

6 Sambrook, J., E.F. Fristch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.

8 Sears, E.R. 1954. The aneuploids of common wheat. *Research Bulletin, University of Missouri Agricultural Experiment Station***:** 1-58.