

NEWS AND COMMENTARY

Plant polyploidy: gene expression and genetic redundancy

LD Gottlieb

Heredity (2003) 91, 91–92. doi:10.1038/sj.hdy.6800317

Allotetraploid plant species originate when the genomes of diploid species are brought together in hybrids and then duplicated, a process apparently initiated by fertilization involving at least one unreduced gamete containing a diploid rather than a haploid complement of chromosomes. In an allotetraploid, the genomes of the diploid parents become homoeologous subgenomes. Many genes in the two subgenomes are expected to be similar in sequence and regulation, but others might be divergent. It is now believed that polyploidy, of one sort or another, characterizes about 70% of the angiosperms including a large proportion of our most important crops (eg, bread wheat, oats, cotton, maize, potato, soybeans, sugarcane), and that even species with small genomes such as *Arabidopsis thaliana* have polyploid ancestry (The Arabidopsis Genome Initiative, 2000). How the genomes of two species that have evolved independently and may be adapted to different environments become integrated in a common tetraploid nucleus has become a topic of great interest (Comai, 2000; Wendel, 2000; Pikaard, 2001; Kashkush *et al*, 2002; Osborn *et al*, 2003).

Many homoeologous genes in a newly formed polyploid might be redundant because they have similar sequences; if so, one or the other of them might be silenced, a possibility frequently noted (review in Wendel, 2000). However, similarity is rarely perfect, and would not be so when genes differ even slightly in sequence or mode of regulation, and certainly not when the differences provide functions that affect the quantity, time or place of appearance of some metabolite or binding factor. In contrast, the duplication of a gene in a diploid species results in a new copy that is essentially identical to the original copy and unlikely to show functional difference, at least initially.

Whether genetic similarity equates to genetic redundancy and how the subgenomes of an allotetraploid interact are forcefully addressed in a recent analysis of homoeologous gene expression in allotetraploid cotton (*Gossypium hirsutum*) by Adams *et al* (2003). The study

assessed the level of mRNA transcripts from 40 pairs of genes in a wide variety of plant parts and provides, for the first time, information about the relative contributions of homoeologs and shows that they can be regulated differently in adjacent plant organs.

The cotton genus *Gossypium* provides an excellent model for studying polyploidy. The genus includes five allotetraploid species that derive from a single polyploid event believed to have occurred about 1–2 million years ago (Seelanan *et al*, 1997). Two of the allotetraploid species, *G. hirsutum* (the source of 'Upland' cotton) and *G. barbadense* (the source of 'Pima' or 'Egyptian' cotton) were independently domesticated within the last 5000 years for their seed fiber and cultivars derived from them now dominate world cotton commerce (Wendel, 1995). The five allotetraploid cottons all carry the A and D genomes (AADD; $2n = 4 \times = 52$) and originated following hybridization between an African or an Asian diploid species (genome AA; $2n = 26$) and an American diploid (genome DD; $2n = 26$). Numerous homoeologous genes have been mapped and sequenced in the two subgenomes of *G. hirsutum* and their molecular evolution characterized (Brubaker *et al*, 1999; Cronn *et al*, 1999; Liu *et al*, 2001).

Transcript levels from both homoeologs of all 40 genes were assessed in whole ovules and the attached fibers. Levels were the same for 30 pairs, but differed for 10 others. More than a five-fold difference was found in four gene pairs and a 1.5–4-fold difference was found in five others; for one gene pair, the transcripts of one homoeolog were not detected. In five cases, genes from the AA subgenome were more highly expressed than those from the DD subgenome and the reciprocal result was observed in four other cases, suggesting that biases in expression were not genome-dependent.

Transcript levels of 16 of the gene pairs were also assessed in 8–10 other plant organs, and 11 of them showed biased expression or absence of expression in at least one organ. Thus, for *AdhA*, the homoeolog from the D

genome was more highly expressed than the one from the A genome in leaves and bracts, whereas in cotyledons and roots the gene from the A genome was more highly expressed. In petals and stamens, only the D genome member was detected, whereas in the carpels only the A genome member was found, a kind of reciprocal expression.

The presence of transcripts of one or the other *AdhA* in different organs could mean that organ-specific expression was selected during the lengthy time period since the origin of *G. hirsutum* or was a direct and immediate consequence of its polyploid origin. The question was examined by assessing the expression of both *AdhAs* in a recently bred synthetic allotetraploid with similar genomic composition as *G. hirsutum*. In the synthetic allotetraploid, the two *AdhAs* showed the same patterns of bias across organs (present in some, absent in others) that were found in cultivated *G. hirsutum*, strongly suggesting that the differences were an immediate consequence of polyploidy.

An alternate hypothesis that the difference in expression patterns was a legacy from the diploid progenitors was rejected by the finding that transcripts were present in all of the tested organs of plants representing both diploid progenitors. However, the tested plants are more than a million years away from the progenitors of cotton and it may be that some differences in organ-specific transcript levels reflect a legacy from the true progenitors. An analysis similar to the one carried out by Adams *et al* should now be carried out on very recent allotetraploids with extant and identified diploid parents (Ownbey, 1950; Roose and Gottlieb, 1976, 1980; Ford and Gottlieb, 2002).

The absence of transcripts of one or another homoeologous gene in newly synthesized polyploids has been correlated with the so-called epigenetic influences involving cytosine methylation, chromatin modifications, and dosage-related effects since it could often be reversed by experimental manipulations (Lee and Chen, 2001; Shaked *et al*, 2001; Kashkush *et al*, 2002; Madlung *et al*, 2002; Osborn *et al*, 2003). Understanding how such 'global' factors operate when a polyploid arises is obviously important, but it will require attention to learn how they affect particular genes and chromosomal segments. Since it is likely that genes encoding different types of proteins, for example, enzymes of metabolism, transcription factors, signal transduction factors, structural proteins, cell-surface receptors, will respond differently to placement in a tetraploid

nucleus with two subgenomes, explanations may eventually prove less global.

From a technical point of view, the tasks carried out by Adams *et al* were not simple because the homoeologous transcripts they studied were nearly identical in length and sequence. They used a technique that involves a novel application of single-strand conformational analysis (SSCP) that is described in Cronn and Adams (2003). In this protocol, one of the strands of a short, double-stranded DNA fragment, produced by PCR from cDNA templates synthesized from the pool of RNA in each of the tested plant organs, is radioactively labeled, denatured, and run out on a special denaturing gel. Under such conditions, single strands differing in even one nucleotide show a difference in electrophoretic mobility. The electrophoretic difference made it possible to distinguish single strands produced from homoeologous cDNA fragments amplified from the same plant; differences in their relative concentrations were determined with a phosphoimager.

Whether the procedure will find wide use remains to be seen. Cronn and Adams claim that the ratio of SSCP products is the same as that of the homoeologs in the PCR pool. This is based on the results of an experiment in which they mixed homoeologous PCR fragments amplified from cDNAs of the A- and D-genome diploids in various proportions, resolved them by the SSCP analysis, and found close agreement with the ratios predicted by the mixtures. The basic question yet to be resolved, which they acknowledge, is

whether the ratio of homoeologs in the pool of PCR products is in turn the same as the ratio of homoeologous transcripts in the sampled plant parts. It should be noted that small amplification biases often occur during early cycles of PCR that may lead to large biases in fragment representation after the typical protocol of 30 or more cycles. Analysis based on a linear rather than exponential multiplication would seem preferable and, eventually, may be accommodated by a microarray analysis (Aharoni and Vorst, 2001; Donson *et al*, 2002). However, with the best present technology, mRNA probes still hybridize to cDNA targets printed on microarrays even when they differ by as much as 10% in sequence (Fernandes *et al*, 2002). Homoeologous cDNAs from *G. hirsutum* or other polyploids that have nucleotide divergence of 1–2% could not, therefore, be resolved by these means.

The study of Adams *et al* is certain to stimulate additional analyses of gene function in polyploids. Their discovery, among many, that a single homoeolog of *AdhA* carries out all the functions required in a particular organ, but the alternate homoeolog does so in a different organ, provides important information for evaluating genetic redundancy. It should now be obvious that differences in expression are complex and that sequence similarity need not predict genetic redundancy. Gene expression has a fine structure that has to be examined organ by organ. Considerations of genetic redundancy must also deal with other issues. One example has to do with the fact that many proteins

are multimeric. Heteromers formed by association of different protein monomers encoded by homoeologous genes in a polyploid plant may have novel functions that are less likely to result from the association of allelic monomers in diploids. The study of Adams *et al* brings these and other issues to the forefront.

LD Gottlieb is at Section of Evolution & Ecology, Division of Biology, University of California, Davis, CA 95616, USA

e-mail: ldgottlieb@ucdavis.edu

- Adams KL *et al* (2003). *Proc Natl Acad Sci USA* **100**: 4649–4654.
- Aharoni A, Vorst O (2001). *Plant Mol Biol* **48**: 99–118.
- Brubaker CL *et al* (1999). *Genome* **42**: 184–203.
- Comai L (2000). *Plant Mol Biol* **43**: 387–399.
- Cronn RC, Adams KL (2003). *BioTechniques* **34**: 726–733.
- Cronn RC *et al* (1999). *USA Proc Natl Acad Sci* **96**: 14406–14411.
- Donson J *et al* (2002). *Plant Mol Biol* **48**: 75–97.
- Fernandes J *et al* (2002). *Plant Physiol* **128**: 896–910.
- Ford VS, Gottlieb LD (2002). *Evolution* **56**: 699–707.
- Kashkush K *et al* (2002). *Genetics* **160**: 1651–1659.
- Lee H-S, Chen ZJ (2001). *USA Proc Natl Acad Sci* **98**: 6753–6758.
- Liu B *et al* (2001). *Genome* **44**: 321–330.
- Madlung A *et al* (2002). *Plant Physiol* **129**: 733–746.
- Osborn TC *et al* (2003). *Trends Genet* **19**: 141–147.
- Owensby M (1950). *Amer J Bot* **37**: 487–499.
- Pikaard CS (2001). *Trends Genet* **17**: 675–677.
- Roose ML, Gottlieb LD (1976). *Evolution* **30**: 818–830.
- Roose ML, Gottlieb LD (1980). *Biochem Genet* **18**: 1065–1085.
- Seelanan T *et al* (1997). *Syst Bot* **22**: 259–290.
- Shaked H *et al* (2001). *The Plant Cell* **13**: 1749–1759.
- The *Arabidopsis* Genome Initiative (2000). *Nature* **408**: 796–814.
- Wendel JF (1995). In: Simmonds N, Smartt J (eds.), *Evolution of Crop Plants*, Longman: London, pp 358–366.
- Wendel JF (2000). *Plant Mol Biol* **42**: 225–249.