Research Focus

In polyploids, one plus one does not equal two

Sarah P. Otto

Department of Zoology, University of British Columbia, Vancouver, Canada, BC V6T 1Z4

Polyploidization has occurred often in the evolutionary history of eukaryotes, yet we are only now learning about the genetic and epigenetic events that occur in newly formed polyploids. A long-standing assumption has been that each member of a gene pair would contribute equally to expression levels after polyploidization. This assumption has now been shown to be false in cottons. Adams *et al.* studied expression levels in different plant tissues from a newly formed allotetraploid cotton (bearing two copies of the so-called 'A' genome and two copies of the 'D' genome). The authors found that the proportion of mRNA levels transcribed from the A and D genomes varied dramatically among tissues. This result implies that duplicated genes can undergo an immediate divergence in function as a pleiotropic effect of hybridization or polyploidization, providing a novel explanation for why selection often preserves duplicated genes.

Every once in a while, a paper is published that shatters our simple notions of a biological phenomenon and replaces it with a richer and more interesting understanding. Such a paper was recently published by Keith Adams and colleagues [1], addressing the phenomenon of gene expression in polyploids.

When two genomes come together in a polyploid, one would expect that each member of a duplicated gene pair would contribute equally to the phenotype of the organism. For this reason, evolutionary biologists have long argued that pairs of recently duplicated genes should witness relaxed selection as each gene can 'cover' for the function of the duplicate gene partner (the 'homoeologue'). This form of relaxed selection is akin to a natural mutation accumulation experiment; mutations that would otherwise be deleterious (if selection were not relaxed) tend to accumulate in such experiments [2]. From this perspective, we would expect rapid degradation and ultimately gene loss of one member of each gene pair. Thus, it is puzzling that, in many cases, >50% of gene pairs are retained over tens of millions of years following genomic doubling [3]. This surprising observation has led to a flurry of hypotheses to explain how natural selection might act to preserve duplicated gene pairs [4–7]. Several studies (reviewed in [8]) demonstrate, however, that our initial assumption is incorrect: duplicated gene pairs do not contribute equally to the phenotype of a polyploid. The new work of Adams *et al.* [1] confirms this finding and takes it one step further, showing that the contribution of each duplicated gene varies among different tissues within both ancient and newly formed polyploids; this finding opens up new possibilities for the long-term maintenance of gene duplicates.

Adams *et al.* [1] studied naturally occurring and newly synthesized allotetraploid cottons that resulted from the merger of two diploid genomes (the 'A' and the 'D' genome), which have diverged from each other over a period of ~7.5 million years (MY). Using reverse-transcription PCR (RT-PCR) and a single-strand conformation polymorphism analysis of the cDNA (cDNA-SSCP), they measured the contribution of the A and the D genomes to the RNA pools of a tetraploid. Importantly, they examined multiple genes and multiple tissues in this fashion. Their results for naturally occurring tetraploid strains (~1.5 MY post-polyploidization) were fascinating. For 11 out of 18 genes, the A and D homoeologues were expressed at different levels in different tissues. These expression patterns were repeatable among individuals and were often extreme. The most extreme example was *adhA*, which was transcribed entirely from the D gene in petals and stamens but entirely from the A gene in carpels.

Differences in the timing or tissue-specificity of duplicate gene expression have been observed previously [5,9,10] and have been ascribed either to beneficial mutations that restrict expression of the homoeologues to different tissues or points in time, thereby enabling specialization [10], or to deleterious mutations in different regulatory regions of a gene pair, effectively dividing up the initial expression pattern among the copies of the gene ('sub-functionalization') [5,9,10]. This is where the work of Adams *et al.* [1] on newly synthesized tetraploid cotton comes in. If beneficial or deleterious mutations were responsible for the tissue-specific nature of gene expression, we would expect the A and D genes to be expressed in the same manner in all tissues in newly formed polyploids, which have not accumulated any mutations. Astoundingly, this expectation is false.

In two tetraploid individuals descended recently (2–5 generations) from a polyploid produced by colchicine treatment of a sterile diploid A–D hybrid, the proportion of RNA transcripts from each genome varied among tissues for all four genes examined (Fig. 1). Although there was no statistical assessment of this variation, the similarity in two different individuals (see *adhA* and *G8* bars in Fig. 1) indicates that tissues do differ in the proportion of mRNAs transcribed from each homeologue, and this pattern is repeatable. Thus, differences in expression pattern among duplicate genes result, at least in part, from immediate and automatic responses of expression levels to hybridization and/or tetraploidization. Although expression patterns undoubtedly evolve, we must be

---

Corresponding author: Sarah P. Otto (otto@zoology.ubc.ca).

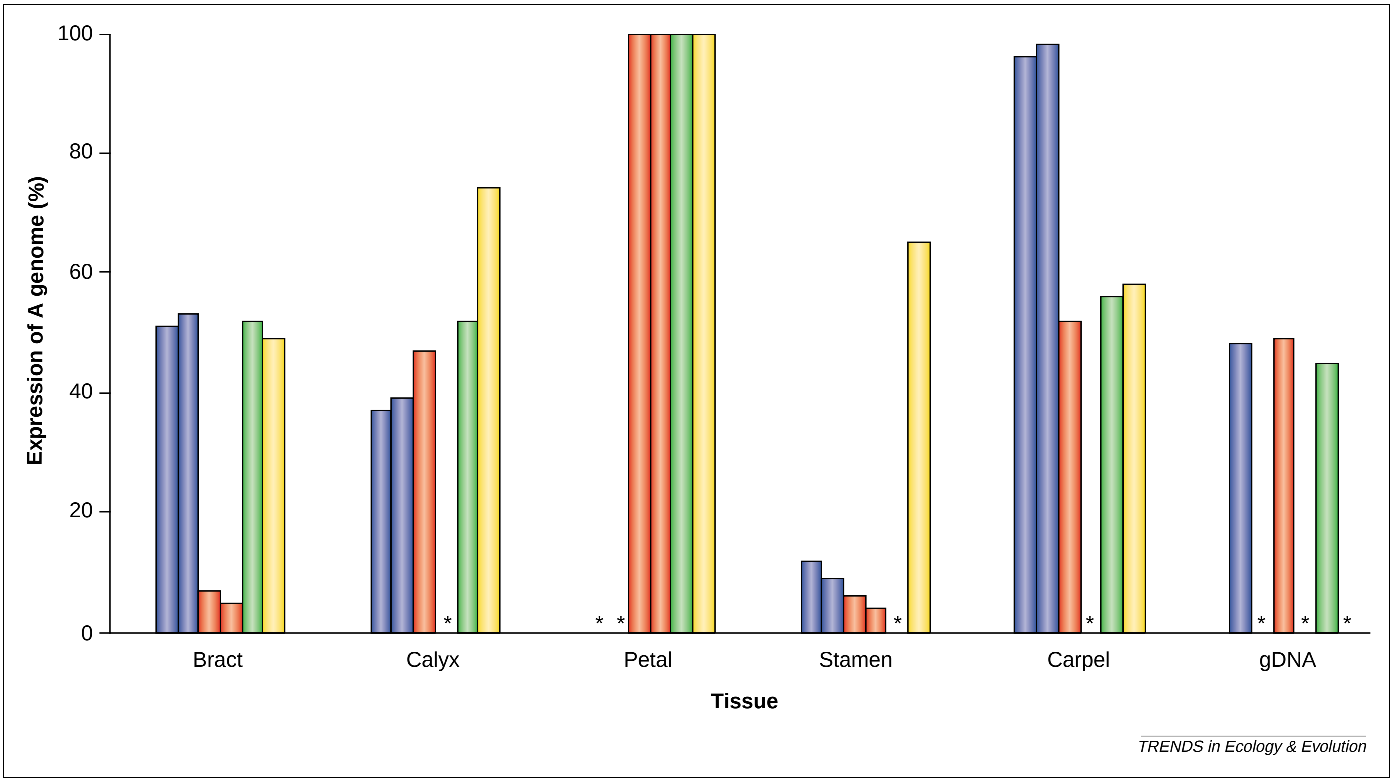

Fig. 1. Proportion of mRNA transcribed from the A genome in different tissues of newly synthesized allotetraploid cottons that carry two A and two D genomes. Expression levels of four genes [*adhA* (blue bars) *G8* (red bars) *myb1* (green bars) and *G7* (yellow)] were measured by Adams *et al.* [1] by extracting mRNA from different tissues, reverse transcribing the mRNA, amplifying the resulting cDNA using PCR, and running gels that distinguish and quantify the amount of cDNA from each genome. The entire procedure was performed on two different plants for *adhA* and *G8*, providing a measure of the repeatability of the results. * indicate expression ratios that were not determined.

careful in ascribing any observed differences in natural polyploids to either beneficial or deleterious mutations until we know whether these differences might have arisen simply as a pleiotropic side effect of polyploidization. Perhaps most importantly, these results cast doubt over the assumption that selection is entirely relaxed for newly duplicated genes. If, as the results of Adams *et al.* [1] suggest, there are developmental stages or tissues where one copy dominates expression and other times and places where the second copy dominates, then deleterious mutations in either copy will have their full impact on a component of fitness (i.e. the component that depends on functioning during those stages or tissues where the mutant copy dominates expression). When duplicated gene copies cannot fully cover for one another, then selection will act to maintain both copies. This result might help to explain why duplicate genes are retained for so long following polyploidization. It also predicts that the homoeologues that are most variable in relative expression immediately following polyploidization would be the most likely to be preserved in duplicate.

There are a few caveats worth mentioning. The methods, whilst ingenious, require that the homoeologues are sufficiently diverged that they appear as separate bands on cDNA-SSCP gels (the A and D genomes of cotton differ by ~1% at the nucleotide level, and the authors chose gene regions that were particularly polymorphic). A related caveat is that the study does not determine whether differences in expression of the A and D gene copies result from hybridization or polyploidization. Furthermore, the study only examined a handful of genes in newly synthesized polyploids, and these genes were a subset chosen because they exhibited tissue-specific differences in expression in natural polyploids. Finally, the authors examined only one time point during development and examined fairly large tissue samples, making it difficult to assess the true extent of expression variation.

Future work in this area promises to address these caveats and improve our understanding of the genomic processes that occur during polyploidization. Polyploid hybrids of more closely related genomes should be examined to determine if they also exhibit automatic differentiation of expression patterns. A comparison of diploid hybrids with polyploid hybrids as well as with non-hybrids would enable us to assess the relative importance of hybridization and polyploidization to the patterns observed by Adams *et al.* [1]. Similarly, it would be interesting to determine whether the expression patterns depend on which genome was maternally inherited by the diploid hybrid, which would indicate whether imprinting and/or cytoplasmic factors are involved. Finally, analyzing a time course of samples drawn from different cell types within each tissue, choosing more randomly chosen genes, and examining multiple individuals would provide a more precise picture of the nature, extent, and replicability of expression variation.

At this point, the mechanisms underlying automatic sub-functionalization are completely unknown, but scientists are beginning to have a greater appreciation for the myriad of ways that cells recognize duplicate genes and alter their expression and/or translation [11]. As Adams *et al.* [1] have shown us, we have much to learn about the genetic processes that occur upon polyploid formation. And what we will learn promises to reshape our understanding of genomic evolution.

References

1 Adams, K.L. *et al.* (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4649–4654

2 Keightley, P.D. and Lynch, M. (2003) Toward a realistic model of mutations affecting fitness. *Evolution* 57, 683–685

3 Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437

4 Takahata, N. and Maruyama, T. (1979) Polymorphism and loss of duplicate gene expression: a theoretical study with application to tetraploid fish. *Proc. Natl. Acad. Sci. U. S. A.* 76, 4521–4525

5 Force, A. *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545

6 Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473

7 Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. Ser. B* 256, 119–124

8 Osborn, T.C. *et al.* (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 19, 141–147

9 Ferris, S.D. and Whitt, G.S. (1979) Evolution of the differential regulation of duplicate genes after polyploidization. *J. Mol. Evol.* 12, 267–317

10 Hughes, M.K. and Hughes, A.L. (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol. Biol. Evol.* 10, 1360–1369

11 Dunlap, J.C. and Wu, C-T., eds (2002) *Homology Effects*, Academic Press

0169-5347/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0169-5347(03)00213-1

The end of the beginning for neutral theory

Sean Nee and Graham Stone

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh, UK EH9 3JT

The point of ecology is to understand the distribution and abundance of species. Neutral models of species' abundances, which suppose that these fluctuate entirely at random, are currently enjoying considerable attention. To date, much of the credibility of such models has been based on their ability to mimic observed abundance distributions. A new paper by Brian McGill shows that, at least for some data, neutral theory performs worse than does an older, simpler null model. However, this is not the end for neutral theory: rather, it is the 'end of the beginning'.

In the 1960s, gel electrophoresis revealed an unexpected amount of molecular variation in natural populations, which, it was thought, was too high to be explained by the action of natural selection. The neutral allele theory of molecular evolution [1] was a proposed explanation, in which selectively neutral variants (alleles) arise by mutation and then fluctuate at random in abundance. Tropical forests have posed a similar problem for ecologists: how can we understand their enormous tree diversity in the traditional terms of the 'one species/one niche' requirement for coexistence? After all, surely all trees want basically the same thing, a patch of ground, some water and some light?

The unified neutral theory of biodiversity and biogeography [2] proposes the same explanation: species are, ecologically, all the same, originate at random and their abundances simply fluctuate randomly over time. Originally proposed a quarter of a century ago [3,4], this idea is currently enjoying considerable interest for numerous reasons, not least of which is that it might be a very pure null model with utility far beyond understanding tropical forest diversity [2,5]. And it appears to pass the minimum requirements for serious consideration: with suitable parameter tuning, the stochastic theory can generate patterns that appear similar to observed patterns of relative abundances of species, such as the tree species studied in the famous 50-ha plot on Barro Colorado Island in Panama [2].

It is not the only candidate for a simple null model, of course ('null' in the sense of not containing any explicit consideration of the usual biological ingredients that ecologists consider to be important [6,7]). Another simple model is this: suppose that numerous factors influence the size of any particular population and they do so in a multiplicative fashion. So, for example, an epidemic decimates the population, a good season doubles the per capita birth rate, a fire halves the amount of suitable habitat for the species and so on. This model predicts a lognormal distribution of species abundances; that is, the logarithms of their abundances should be normally distributed [8].

Best-fit models

So, an obvious question arises: which model fits data better? This is the question addressed by Brian McGill in a recent *Nature* paper [9]. There are actually two interpretations of this question. First, which model is more readily fitted to data? There is no contest: the lognormal is trivially easy to fit from the mean and variance of the logged species' abundances. Fitting the neutral distribution is hard and requires extensive simulation: the procedure requires five pages and a flow chart for its description [2] and, even then, it is

---

Corresponding author: Sean Nee (sean.nee@ed.ac.uk).