# Evolutionary Genetics of Genome Merger and Doubling in Plants

Jeff J. Doyle,[1] Lex E. Flagel,[2] Andrew H. Paterson,[3] Ryan A. Rapp,[2] Douglas E. Soltis,[4] Pamela S. Soltis,[5] and Jonathan F. Wendel[2]

[1] L. H. Bailey Hortorium, Department of Plant Biology, Cornell University, Ithaca, New York 14850; email: jjd5@cornell.edu

[2] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011; email: flagel@iastate.edu; rrapp@iastate.edu; jfw@mail.adp.iastate.edu

[3] Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30602; email: paterson@dogwood.plantbio.uga.edu

[4] Department of Botany, University of Florida, Gainesville, Florida 32611; email: dsoltis@botany.ufl.edu

[5] Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611; email: psoltis@flmnh.ufl.edu

## Key Words

## Abstract

Polyploidy is a common mode of evolution in flowering plants. The profound effects of polyploidy on gene expression appear to be caused more by hybridity than by genome doubling. Epigenetic mechanisms underlying genome-wide changes in expression are as yet poorly understood; only methylation has received much study, and its importance varies among polyploids. Genetic diploidization begins with the earliest responses to genome merger and doubling; less is known about chromosomal diploidization. Polyploidy duplicates every gene in the genome, providing the raw material for divergence or partitioning of function in homoeologous copies. Preferential retention or loss of genes occurs in a wide range of taxa, suggesting that there is an underlying set of principles governing the fates of duplicated genes. Further studies are required for general patterns to be elucidated, involving different plant families, kinds of polyploidy, and polyploids of different ages.

## INTRODUCTION

It is now known that flowering plant genomes are fundamentally polyploid. Great strides have been made in understanding polyploidy (often called genome doubling), and the pace of discovery has accelerated as the importance of the phenomenon has become increasingly clear. In particular, molecular systematic studies have provided insights into the patterns of polyploid origins; genome sequences and genome-scale data have shown that polyploidy is far more prevalent than expected and has been a key force in shaping plant genomes; and studies of synthetic and recently formed natural polyploids have shown that polyploidy can have immediate and profound genetic and epigenetic consequences.

An example of how this work has transformed our understanding is the realization that allopolyploidy often leads to unexpected and unexplained departures from predicted genomic additivity. This includes gene loss (16, 79, 103), widespread modification of methylation patterns (65, 91, 113), and nonreciprocal chromosomal exchanges (82, 109). From an evolutionary or ecological standpoint these phenomena may be viewed as novel generators of genomic variation, as demonstrated in *Brassica* for several environmentally important phenotypic characters, including flowering time (82), leaf morphology, and seed set (31). Much of this has been reviewed elsewhere (1, 4, 19, 20, 23, 53, 98, 115) but is reiterated here as a reminder of the myriad and diverse genomic responses to polyploidy.

Notwithstanding these rapidly accumulating insights, important questions remain about every stage of polyploid evolution. How do individual polyploid plants form from one or more diploid progenitors? How do recently formed polyploids become established in natural populations? How do polyploids achieve discrete evolutionary trajectories separate from those of their progenitor(s)? To what degree has polyploidy provided a stimulus for diversification in the larger tree of life? Despite the frequency of polyploidy in eukaryotes, and particularly in plants, we do not know why polyploidy is so prevalent, or, conversely, why polyploidy is not universal if it confers some general adaptive advantage and promotes lineage diversification. Nor do we understand the dynamics that cause polyploids, once formed, to return to a functionally diploid state, as in cryptic polyploids such as *Arabidopsis* and rice.

What are the essential attributes of polyploidy that make it such an important evolutionary mechanism? Is it the ability of polyploidy to promote heterozygosity, either through fixed hybridity or by polysomic inheritance? Is it the presence of duplicate copies, not only of every gene in the genome, but of every genetic network? Is it the accelerated mutational activity in early generations, through "genomic shock"? All have been proposed as key reasons for the success of polyploids, and all are probably important, but perhaps to varying degrees in different lineages. Finally, are polyploids in general more "successful" than their diploid progenitors? This is an exceedingly difficult question to answer, in large part because "success" is an ill-defined term that can refer to anything from short-term proliferation of individuals to long-term effects on lineage diversification.

Answering these and other questions will require comparisons of diploids and polyploids by researchers in such diverse disciplines as ecology, population biology, and physiology. Most of these areas have received far less attention than have the genetics and genomics of polyploidy (see 107). Nevertheless, even in these better-studied areas much remains to be learned. In particular, it is only by moving beyond the idiosyncrasies of a handful of model systems (most of which are crops) that emergent properties of polyploidy can be detected. Here we focus on the genetic end of the organizational spectrum, dealing with gene expression and epigenetics, the process of diploidization, and the fates of duplicated genes (**Figure 1**). Our aim is to provide an updated entry to the literature as well as to highlight what we view as major unanswered questions in these areas.

# GENE EXPRESSION

## Gene Expression Variation Arising from Polyploidy

The alteration of gene expression patterns is a prominent cause of variation within and between species and may be the primary source of developmental novelty (14, 21, 22, 28, 50, 88). Recent large-scale microarray studies in a range of polyploid plant species have confirmed that gene expression is radically altered by polyploidy (39, 102, 112; R.A. Rapp & J.F. Wendel, unpublished). Furthermore, by evaluating synthetic polyploids, several studies have shown that massive expression changes accompany polyploid formation (although hybridization, rather than genome doubling, per se, may be a major cause of such changes; see below). The magnitude of effects varies greatly between species, but enough data now exist to reveal some general trends.

By comparing global gene expression profiles in synthetic allotetraploids with their parental diploid genome donors, work in *Gossypium* (cotton; R.A. Rapp & J.F. Wendel, unpublished) and *Arabidopsis* (112) has specifically addressed the transcriptional effects of combining differentiated genomes, with their divergent regulatory machinery, into a common nucleus. Wang et al. (112) showed that a synthetic *Arabidopsis* allotetraploid, formed by combining *A. arenosa* with *A. thaliana*, exhibits strong expression dominance of the *A. arenosa* parent, coupled with suppression of the *A. thaliana* genome. The extent of this suppression is impressive; approximately 94% of the genes up-regulated in the *A. thaliana* parent relative to the *A. arenosa* parent are subsequently down-regulated (suppressed) to the level of the *A. arenosa* parent after allotetraploidy. In cotton, comparison of two synthetic allopolyploids with their parents (*G. arboreum* x *G. thurberi* and *G. arboreum* × *G. bickii*; R.A. Rapp & J.F. Wendel, unpublished) revealed substantial dominance of the maternal, *G. arboreum* expression phenotype. The amount of dominance varied significantly between the two crosses. In the *G. arboreum*

X *G. thurberi* allotetraploid, the level of dominance is about 1:12, meaning that for every gene expressed at the parental *G. arboreum* level in the tetraploid there are 12 expressed at the *G. thurberi* level. This ratio is nearer to 1:2 in the *G. arboreum* × *G. bickii* allotetraploid, indicating a weaker, but nonetheless significant effect. Both studies highlight an important and emergent property of allopolyploidy; with regard to expression, many genes do not behave as simple additive combinations of the parental genomes; that is, in cotton and *Arabidopsis*, the best-studied systems to date, genomic dominance appears to be quite common.

This lack of additivity in gene expression levels raises several fundamental questions about the mechanistic underpinnings and evolutionary dynamics of gene expression following genomic merger. From a mechanistic standpoint, what is responsible for nonadditivity in gene expression, and why does this vary so much among genes and between different genomic combinations? Why do the two cotton crosses demonstrate such a large disparity in degree of suppression of the gene expression phenotype contributed by the *G. arboreum* parent? Is this a property of genomic divergence, and if so is it divergence in structural, regulatory, or epigenetic features? From an evolutionary point of view, how do genomic merger and genome doubling individually impact expression variation during the formation of allopolyploids, and what are the potential phenotypic effects of each of these sources of variation?

The relative importance of these two processes can be addressed by comparing $F_1$ diploid hybrids and allopolyploid individuals, as has been done in cotton. In a study of approximately 1400 duplicated gene pairs, one quarter of the cases of genes with biased homoeolog expression ratios were shown to have arisen immediately as a consequence of genomic merger (27), with the expression bias maintained following genome doubling. These biases are also observed in natural cotton allopolyploids that are ~1–2 million years old, demonstrating remarkable evolutionary stability for a phenomenon that was saltational in its origin.

**Synthetic polyploids:** polyploids created in the laboratory, usually using the mitotic spindle inhibitor, colchicine, but sometimes (e.g., potato) using naturally unreduced gametes

**Homoeolog:** referring to genomes, chromosomes, or genes originating from different genome donors in an allopolyploid

**Autopolyploidy:**
autopolyploids are formed within a species and/or exhibit tetrasomic segregation

**Triploid bridge model of polyploid origin:** two-step model of tetraploid formation involving unreduced gametes; contrasted with polyploid formation by direct genome doubling

Similar findings were reported for the proteome of synthesized *Brassica* allotetraploids and diploid hybrids, where Albertin et al. (5) found that ~89% of protein expression differences could be attributed to hybridization rather than to polyploidization.

Such results suggest that autopolyploidy might produce less dramatic expression changes than allopolyploidy, and this has been observed in *Arabidopsis* (112). In potato, around 10% of the 9000 genes assayed by Stupar et al. (102) showed expression differences between diploids and synthetic autotetraploids, and phenotypic differences were also observed. However, the expression differences were characterized as "subtle." All of this suggests that hybridity may have a more profound effect than genome doubling, per se, in the case of allopolyploids.

This is not to say that ploidal level does not play an important role in gene expression, particularly in the early stages of polyploid formation. In maize (6) and *Senecio* (39), dosage balance (an even-numbered ploidal level) was found to play a crucial role in establishing stability in gene expression. When compared to diploids, triploid individuals from both species were shown to exhibit radically different and novel expression profiles. In *Senecio*, Hegarty et al. (39) showed that tetraploidy returned the transcription profile back to a state most similar to that of diploid individuals.

Together, these findings regarding hybridity and ploidy point to an interplay between the "genomic shock" caused by hybridization and dosage imbalance during allopolyploid formation. Because triploidy is often thought to be a necessary "bridge" during allopolyploid formation (85), and because hybridization is involved in the incipient stages of any allopolyploidization event, both dosage balance and the particularities of genomic combination during merger likely contribute to novel expression phenotypes. This raises several interesting and unanswered questions. What level of hybridity is required to trigger such "genomic shock?" Could hybridization within species cause genomic shock? Genotypes within species cross routinely, but given the variability that exists within some plant species [e.g., maize (29)] some genotypic combinations could well produce dramatic effects when doubled. If the degree of differentiation between genomes involved in polyploid formation is a determinant of the amount of change produced by polyploidization, does the genetic distance between diploid progenitors play a role in promoting or inhibiting polyploid formation (13, 18)?

## A Genic Perspective

One of the more spectacular recent revelations with respect to gene expression in polyploids is that homoeologous genes commonly make unequal contributions to the transcriptome, as shown most thoroughly to date in cotton and in wheat (2, 3, 11, 27, 40, 72). Adams et al. (2)

#### Figure 1

Schematic of evolutionarily relevant genetic and genomic dimensions to genome merger and genome doubling. (*a*) Chance cross-pollination between two divergent species or populations leads to hybridization and, often, polyploidization. (*b*) This genomic merger and doubling evokes myriad genetic and epigenetic responses (*center panels*; described in the text), giving rise to various evolutionary outcomes including sub- and neofunctionalization, novel regulatory interactions, and gene loss (not shown). (*c*) Novel expression variation in the nascent lineage takes several forms, including transgressive up-regulation or down-regulation, silencing, unequal parental contributions, and altered expression times and locations. Illustrated are possible expression patterns for three developmental stages (1, 2, and 3) for a hypothetical gene in paternal (*blue*), maternal (*pink*), and polyploid (*green*) cells, tissues, or organs. In turn, this novel expression diversity, in conjunction with genetic alterations, generates evolutionarily novel phenotypes and outcomes, including ecological range expansion, invasiveness, novel pathogen resistance and secondary chemistry, altered phenologies and sexual systems, and novel morphologies. (*d*) Over longer evolutionary periods (millions of years), the polyploid genome undergoes fractionation and diploidization, as discussed in the text. This process may be evolutionarily episodic, recurring several times in various lineages, as has been discovered in many different angiosperms.

demonstrated that 10 out of 40 homoeologs from the A- and D-genomes of allotetraploid cotton exhibit biased expression, including several cases of reciprocal silencing among adjacent floral whorls. This demonstration of unequal contributions of gene duplicates to the transcriptome has been verified and expanded in several subsequent studies, including one in which homoeolog ratios for approximately 1400 gene pairs were studied during development of cotton "fibers," which are single-celled epidermal trichomes on the ovular surface. Biases in homoeologous expression even extend to this fine scale, where temporal variation across developmental stages was shown (40).
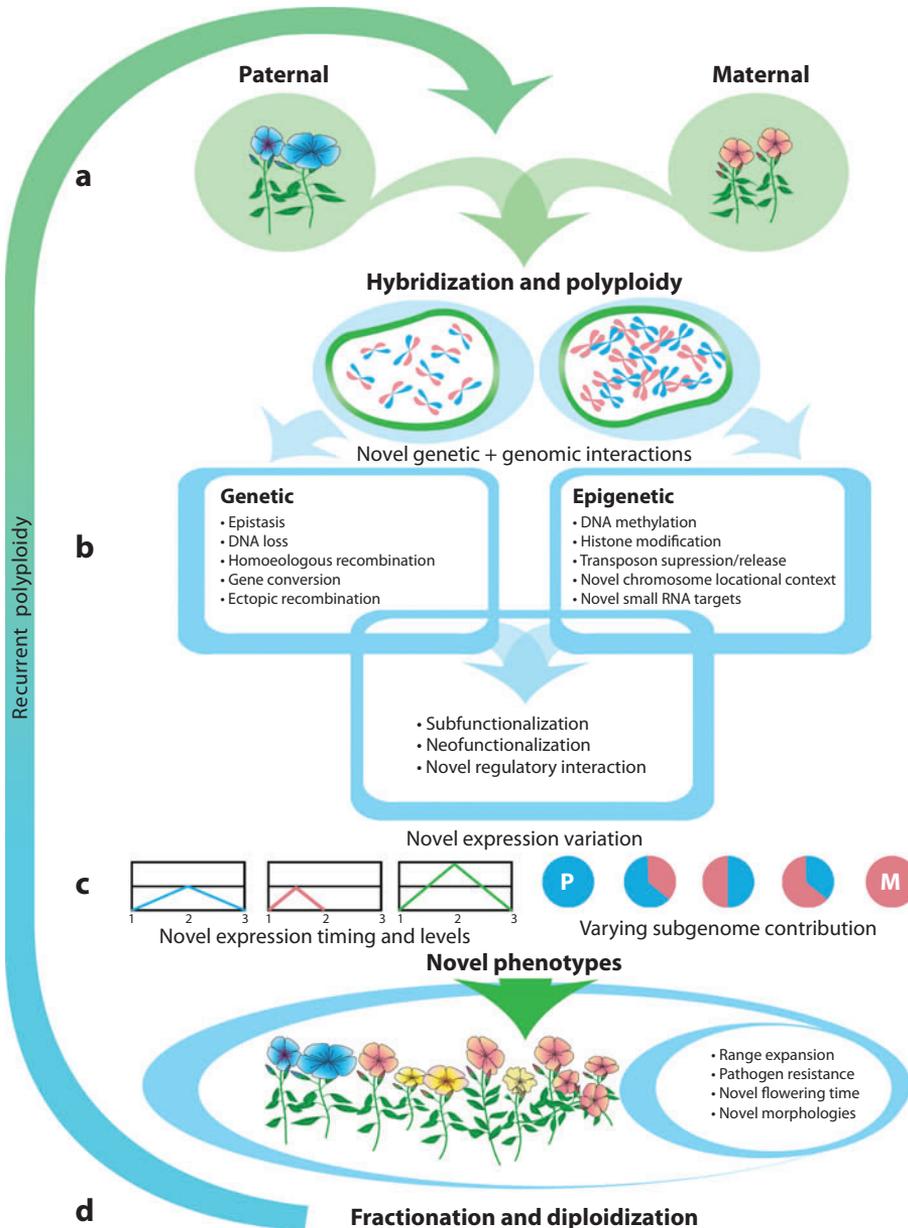
**Subfunctionalization:**
partitioning ancestral
functions/expression
patterns between
duplicated genes,
leading to retention of
both genes because
loss of either would be
lethal

These and other biases in homoeologous expression have been described as expression subfunctionalization, which is the partitioning of ancestral expression domains among duplicate genes (homoeologs in this case). Because there is little sequence divergence between the A- and D-genomes in cotton (~1% in exons) and because ancestral and derived expression states in a range of tissues are readily compared among diploids and polyploids, this observation suggests that homoeolog expression subfunctionalization is controlled by an epigenetic mechanism, a suggestion further bolstered by demonstrations of expression alteration in synthetic allopolyploids, whose genomes have not undergone any subsequent evolution. This mode of subfunctionalization is of considerable evolutionary interest, as it may enhance the probability of duplicate gene retention (62) (see below) following polyploid formation (2, 86). Also, expression subfunctionalization may act far more quickly than the classic conceptualization of subfunctionalization arising from stochastic mutations in coding or noncoding regions, presumably a much slower process. From an evolutionary perspective, it has been hypothesized that expression subfunctionalization may initially preserve a large number of homoeologous pairs from mutational decay, thus retaining additional raw material for subsequent evolutionary tinkering (2, 86).

Expression subfunctionalization in polyploids can perhaps be considered a special case of allelic variation in expression. Allele-specific expression patterns have been observed in a wide range of organisms (60, 116), and the diversity of expression among alleles is thought to be a mechanism underlying heterosis (38). Polyploidy brings together and preserves new combinations of alleles, each with characteristic patterns of expression that may be quite different, particularly in the case of alleles in allopolyploids formed from different species. Gene expression is context dependent (90), and alleles from the diploid progenitor(s) will experience a novel context in the polyploid. Here again, comparison among diploids, synthetic diploid hybrids, and synthetic polyploids can distinguish expression subfunctionalization that is due to hybridity from that due to genome doubling.

Expression subfunctionalization is undoubtedly an important phenomenon at the transcriptional level, and we have argued that it is important for duplicate gene retention. Extrapolating from existing studies to a hypothetical case where complete knowledge of duplicate gene expression patterns was available for all genes in nascent polyploids, it is highly probable that every gene pair experiences expression subfunctionalization. Thus, to the extent that this buffers genomes against gene loss due to mutational decay, an a priori case may be made that this is an evolutionarily highly significant mechanism. At present, however, empirical evidence is largely lacking that expression subfunctionalization has led to the later genesis of physiological or phenotypic novelty and hence diversity in plants. This is not to diminish the possibility, but rather to emphasize a fruitful avenue for future work.

Hints regarding this last point are beginning to emerge. In cotton, comparison of homoeolog expression biases for ~1400 genes in a modern synthetic $F_1$ and a 1- to 2-million-year-old allotetraploid showed that most of the 235 genes with shared expression biases between the $F_1$ and the allotetraploid displayed more extreme biases in the allotetraploid (27). This indicates that, in cotton, the long-term effect of expression subfunctionalization is an apparent enhancement of the bias initially established by genomic merger. However, it remains unclear whether this process facilitated duplicate gene retention or evolutionarily relevant functional diversification. Evidence from *Arabidopsis* indicates that gene retention following polyploidy (the *Arabidopsis* lineage experienced its most recent polyploidy event 20–60 mya) can, to some extent, be explained by expression subfunctionalization (10, 15, 32). However, Casneuf et al. (15) have shown that paleo-homoeologs in *Arabidopsis* show more highly correlated expression patterns (i.e., less subfunctionalization) than do other types of duplicates, indicating that expression subfunctionalization may have played only

a weak role in preserving paleo-homoeologs. Other mechanisms, such as the retention of dosage-sensitive genes (106) and the buffering of critical functions through redundancy (17), have also been suggested. Even without knowing the specific mechanism(s) involved, there is clearly some disconnect between the observation of high levels of expression subfunctionalization in recent homoeologs of cotton and wheat (11, 72) and the apparent absence of such subfunctionalization among *Arabidopsis* paleo-homoeologs. It is difficult to resolve these two contradictory observations; however, they may simply reflect differences in the time since polyploid formation, lineage-specific differences between these species, or may highlight our poor understanding of the processes that return polyploid genomes to a diploid state.

# EPIGENETICS

What controls the altered gene expression observed in polyploids? Any or all of the molecular mechanisms associated with epigenetics (44) could be involved. Here we use a broad circumscription of the term, applicable to any molecular or morphological phenotypic change occurring in the absence of nucleotide change in the coding sequence or upstream promoter of the affected gene. Such a definition encompasses an array of molecular mechanisms, including DNA methylation, histone modification, small RNA-mediated gene silencing, and nuclear/chromosomal context with respect to location. Often, epigenetic phenomena are operationally defined solely by the absence of nucleotide change, as is commonly the case in studies involving synthetic polyploids created from extant diploid progenitors.

The best-studied epigenetic phenomenon in polyploids is methylation. Methylation-sensitive isoschizomers have been used in conjunction with Amplified Fragment Length Polymorphism (AFLP) methodologies to assay polyploidy-induced changes in DNA methylation in several species. Results have been very variable. At one extreme, in synthetic *Gossypium* allohexaploids and allotetraploids, plants

in the first through third polyploid generations displayed no nonadditive banding patterns across 22,000 AFLP bands (56). In contrast, in both wild and resynthesized allohexaploid wheat, nonadditivity was found for 20% of bands (25). In *Spartina anglica*, an invasive allopolyploid European grass formed from a cross between the native *S. maritima* and the introduced North American *S. alterniflora*, 30% of the parental methylation patterns were nonadditive both in the hybrids and allopolyploids, suggesting that methylation changes were due to hybridity (91). Polymorphic methylation changes accompanied polyploidization in 47 synthetically resynthesized lines of *Brassica napus*, with an average of 9% nonadditivity across those lines; reexamination four generations later showed additional changes at a much lower ~3% frequency (31, 61). Similarly, synthetically reconstituted *Arabidopsis suecica*, from a doubled *A. thaliana* by *A. arenosa* cross, yielded an average of 8% nonadditivity across three $F_3$ generations (65).

Although the number of polyploid systems studied in this manner remains low, a possibly notable feature of the results to date is that the species that are relatively static with respect to polyploidy-induced methylation (*Gossypium*, *Brassica*, *Arabidopsis*) are eudicots, whereas the more "active" taxa (*Triticum*, *Spartina*) are grasses. GC content is elevated in grasses, and genomic regions rich in GC have high levels of methylation, so it is possible that grasses, with their high GC, rely on methylation more heavily than do lower-GC eudicots (30, 35, 49) to regulate some aspects of gene expression.

Methylation is the only epigenetic phenomenon that has been studied in several polyploid species, but other mechanisms also regulate gene expression in nascent allopolyploids. Wang et al. showed that synthetic *Arabidopsis* polyploids continued to suppress some loci even after their DNA methylation machinery was knocked out (113). Reactivation of two genes was observed, but others remained silenced. Perhaps these results are not surprising, given the demonstration from using knockouts in *A. thaliana* that various epigenetic marks are

linked, such as methylation and histone modification, each of which has the ability to silence genes (68, 83, 87). Because of this linkage, a potentially valuable approach would be to examine histone and chromatin states in various polyploid/diploid systems using methods such as chromosomal immunoprecipitation on microarray chips (ChIP-on-Chip) (58, 89, 95, 110, 118).

Our understanding of gene regulation has been revolutionized by the discovery of the roles of small RNAs (miRNA and siRNA). Yet the only study of miRNA expression in a polyploid is from the allotetraploid *Gossypium hirsutum*, and it enumerates the miRNAs expressed in a heterogeneous mixture of tissues without reference to evolutionary questions regarding the significance of miRNA controls in the evolution of polyploids (119). Clearly, much remains to be learned about such issues as how homoeologous copies from closely related diploid progenitors, showing near-identity at the sequence level, could be regulated by small RNAs.

## DIPLOIDIZATION

What becomes of the newly formed polyploid genome or the variation that is created during the earliest stages of polyploid evolution? The term "diploidization" is often used to describe the process by which raw polyploids ["neopolyploids" in the sense of Ramsey and Schemske (84)] become stabilized. This term could therefore include all responses to polyploidy, including the genetic and epigenetic effects discussed above, and has been used that way, for example, in a recent review of allopolyploidy (64). As noted above, it appears in general that doubling the genome is a less radical change than combining two genomes (20). If so, then polyploids derived from two very similar genomes may require less evolutionary resolution than those derived by wider hybridization. For example, allopolyploids formed by wide hybridization are expected to undergo many of the regulatory changes described above, in at least some cases also experiencing sequence loss and structural rearrangements (e.g., 31).

From the standpoint of chromosome behavior, genetic allopolyploids are by definition diploids with extra sets of chromosomes, showing diploid chromosomal behavior including bivalent formation and disomic segregation. Chromosomal diploidy is not always achieved immediately, even in interspecific allopolyploids (84), and karyotypic changes can occur through recombination between homoeologs (e.g., 31). Disomic segregation can preserve the contributions of progenitors indefinitely, unless one homoeologous locus is lost or undergoes concerted evolution (e.g., gene conversion) to the other homoeolog. In contrast, hybrid polymorphism can be lost by segregation in polyploids with tetrasomic (or polysomic, for higher polyploids) inheritance. Genome-wide neutral allele coalescence profiles are therefore expected to differ for the two kinds of polyploids. Genetic allopolyploids should have uniform coalescence times across neutrally evolving loci. Polyploids of hybrid origin but with tetrasomic inheritance should have a bimodal coalescence pattern (34), with some loci retaining contributions from both parents (deep coalescence) and others retaining alleles from only one parent (shallow coalescence).

It is generally assumed that because polysomic associations can lead to meiotic aberrations, the polysomic condition is a transient stage leading to chromosomal diploidization and disomy. This may well be true in allopolyploids, which though disomic may exhibit multivalent formation in early stages after formation (77, 84). But is this also true of autopolyploids? Can it be assumed that older polyploids with disomic behavior did not experience polysomic segregation through much of their history? To test that hypothesis, genome-wide studies of gene coalescence, or data on chromosomal behavior and segregation ratios of natural polyploids of known ages, are required. Instances of old polyploids with disomic behavior do not provide evidence one way or another—their ancestry is often unknown (e.g., *Glycine max*, *Arabidopsis thaliana*), and even if progenitors were identified, the chromosome behavior of the polyploid at its formation would remain

---

**Neopolyploids, paleopolyploids:** refers to recent vs older polyploidy, respectively. The time-scales are not absolute, and the terms neopolyploidy and paleopolyploidy are used variously by different authors

unknown. Thus it remains unclear whether genetic autopolyploids must necessarily undergo chromosomal diploidization. In fact, many autopolyploids appear to undergo meiosis with no chromosomal irregularities, and thus selection for chromosomal diploidization in at least these species is expected to be weak.

What is known is that exchange among chromosome sets occurs in many polyploids, including allopolyploids. Moreover, although homoeologous exchanges may occur more frequently during early stages of allopolyploid evolution (84)—well documented, for example, in synthetic *Brassica napus* (31, 109), the process continues over millions of years. In *Nicotiana*, homoeologous chromosome complements are readily observed by genomic in situ hybridization (GISH) in young allopolyploids (e.g., *N. tabacum*, 200,000 years old), but not in older allopolyploids (e.g., *N. quadrivalvis*, 1 million years old; *N. nesophila*, 4.5 million years old), presumably due to homogenization of repeated sequences across homoeologues (55). The chromosomes of older polyploids are mosaics of homoeologous segments, and although it is not known exactly how this occurred, it must have involved mixing of homoeologous chromosome complements. Translocations between homoeologous chromosomes have been detected in *N. tabacum* [(55) and references therein]. In *Oryza*, exchange leading to a mosaic pattern of homoeologous segments is very well documented, with homogenization occurring across "paleologs" (paleo-homoeologs) involving genes as well as repeated sequences (114). In *Glycine*, after around 10 million years the doubled number of chromosomes (relative to generic allies) still betrays its polyploid past, but homoeology is segmental, not chromosomal (96); in *Arabidopsis thaliana* scrambled homoeologous segments occur in a "diploid" number of chromosomes in a small genome.

The reduction of the *Arabidopsis* genome has involved the loss of around 80% of homoeologous genes from the most recent polyploid event, and recent work has shown that the loss has been nonrandom, with many duplicated segments showing preferential gene loss from one homoeolog (106). Although it is not known whether all of the losses involve the same homoeologous genome, this might be expected to be the case because, as noted above, in synthetic allopolyploids the contribution of one parent may be preferentially repressed (112).

## PATTERNS OF GENE RETENTION AND LOSS

### Fates of Duplicated Genes: Are There Patterns?

Polyploidy increases the number of genes by whole-genome multiples. Because the diploid progenitors of the polyploid were functional with a single set of genes, at formation the nascent polyploid is logically viewed as having an "extra" set. Is this redundancy advantageous, deleterious, or neutral? What is the spectrum of consequences of having a suddenly doubled suite of genes and genomes provided by polyploidy? Are the answers to these questions consistent among plant lineages or between polyploidy events within a single species?

Traditional views maintained that allopolyploidy promoted the fixation of homoeologous loci, referred to by some as fixed heterozygosity (52). It was suggested that this increased heterozygosity was itself advantageous (36, 52, 100). Classical views also suggest that genome duplication is potentially advantageous as a primary source of genes with new functions (76, 101). However, such functions take time to evolve. Analysis of whole genome sequences shows that the majority of genes are restored to singleton status following genome duplication, reducing the time available for such adaptive divergence to occur.

Analyses of large EST sets and genome sequences show that retention/loss of duplicated gene copies is not random. Some genes duplicate and reduplicate, whereas others are iteratively returned to singleton status (17, 93), and specific gene functional categories preferentially retain or lose copies (10). Classification of genes into functional groups based

on shared protein functional (Pfam) domains permits genome comparisons across broad taxonomic distances for which orthology is not readily established. This approach has been used to quantify the tendencies of individual Pfam domains to occur among duplicates or singleton genes resulting from independent genome duplication events that occurred 20–60 and ~70 mya in the lineages of *Arabidopsis* (12) and *Oryza* (80), respectively, to one another and to independent genome duplications in yeast (117) and *Tetraodon* (45).

Although the members of most gene functional groups are randomly distributed between singleton and duplicated genes, in angiosperms nonrandom patterns of both gene retention and gene loss are evident (81). For example, in *Arabidopsis*, 16 domains showed highly significant enrichment in singletons, and 4 in duplicates. For *Oryza*, 12 domains showed highly significant enrichment in singletons, and 2 in duplicates, both of which are more frequent than would be expected to occur by chance. These findings generally support tendencies suggested by broader Gene Ontology (GO) categories used in prior studies, which showed that genes involved in signal transduction and transcription are preferentially retained in duplicate, and that duplicate genes involved in DNA repair are preferentially lost (10, 66). However, analyses of Pfam domain-based groupings (81) sometimes reveal heterogeneity within the broader GO categories used in other studies (10, 66), for example, showing one abundant protein-protein interaction domain (LRR) to be usually preserved in duplicate whereas two less-abundant domains (SET, TPR) are usually restored to singleton state.

Classical views about the possible advantages of genome duplication (76, 101) focus largely on one extreme in a spectrum of possible fates for duplicated genes, specifically the subset of genes for which duplicate copies are preserved for a long time. However, as noted above, the fully sequenced angiosperm genomes available to date show nonrandom patterns of both gene retention and gene loss following genome

duplication. Nonrandom gene loss, reflected as genes or gene functional groups that are consistently restored to singleton status [duplication-resistant (81)], represents an underexplored dimension in the spectrum of duplicated gene fates. It may be no coincidence that this dimension was found only recently, with the complete sequencing of multiple angiosperm genomes— the antiquity of genome duplication in most animals and microorganisms precludes detection of nonrandom gene loss, because most genes have been returned to singleton status (81). Early estimates suggest that duplication-resistant genes number in the hundreds and are widely distributed across the genome, although their loss may sometimes be as members of larger blocks rather than as single genes (106).

Retention/loss of gene functional groups has been convergent following independent genome duplications. A total of five and two domains showed highly significant enrichment in singleton and duplicated genes, respectively, in both *Oryza* and *Arabidopsis*, a correspondence unlikely to occur by chance (81). By calculating for each gene functional group (Pfam domain) the fraction(s) of cases in which only one member of a duplicate pair (singletons) is retained following a particular genome duplication, one can compare the fates of most genes across a pair of genomes. These fractions are closely correlated in *Arabidopsis* and *Oryza*, with lesser but still highly significant correlations of the angiosperms to yeast and to *Tetraodon*. The finding that many gene functional groups show convergent patterns of retention or loss following independent duplications that are separated by hundreds of millions of years of evolution (81) is supported by postduplication convergence of gene copy number in divergent yeasts (92), and by the finding that genes from the same metabolic pathway show similar retention/loss trends in *Paramecium* (7). Collectively, these observations suggest that an underlying set of principles of molecular evolution, largely obscure at present, contribute to the fates of genome duplications across

divergent taxa. Elucidating the molecular mechanism and evolutionary forces that shape this process remains an important avenue for future research.

Pooled data from multiple genomes may reveal duplication-resistant gene families that previously were undetectable. The most extreme possible case of duplication resistance, a gene functional group adaptive when only one group member is present in a genome, would provide too little information to be inferred as duplication-resistant within a single genome by statistical methods requiring evaluation of several members of a functional group (81). However, if a gene is repeatedly restored to singleton status following several independent duplications in different genomes, then duplication resistance might be inferred. For example, 17 Pfam domains occurring in only 1–3 *Arabidopsis* or *Oryza* genes (too few to detect significant patterns in any one genome) were inferred to be duplication resistant by pooling data for independent duplications in the two genomes (81). Such pooling across additional duplications in *Populus*, *Medicago*, *Glycine*, *Lycopersicon*, *Vitis*, and *Carica* will permit further tests of the *Arabidopsis-Oryza* inferences and identify additional candidate duplication-resistant genes, even single genes that are convergently restored to singleton status.

At present, relatively little is understood about the temporal dynamics of long-term fractionation of polyploid genomes that leads to genic diploidization, notwithstanding both its theoretical underpinnings and the extensive data that have been generated on present gene content in several model genomes, as discussed above. Some neopolyploids show rapid gene loss (26, 48, 79), whereas others show scant evidence (56) but a surprisingly high incidence of instantaneous (with the onset of polyploidy) expression subfunctionalization (2, 3). One of the challenges in understanding the evolution of polyploid genomes is to reconcile this latter observation with the empirical reality of long-term genome fractionation: Are the early responses of genomes to polyploidization part of the process of adaptation to the dupli-cated state, or merely symptoms of imminent extinction?

## The Evolution of Genes in Modern Populations is Influenced by the Presence of 20- to 60-Million-Year-Old Duplicated Copies at Unlinked Loci

Under classical models, gene duplication is proposed to be a primary source of genetic material available for evolution of genes with new functions (76, 101, 104); one member of a duplicated gene pair may mutate and acquire unique functionality with the fitness of the organism insulated by the homoeolog (neofunctionalization), or the pair may subdivide the ancestral function (63, 108). Such models agree with theory (43) in predicting that in natural populations, higher levels of polymorphism would occur in duplicated genes than in singletons, and that the ability of duplicates to provide functional compensation for one another would erode as their functions diverged (111).

However, several findings raise perplexing questions about this classical functional divergence model. Analysis of 17 nonallelic duplicates in *Xenopus laevis* shows evidence of purifying selection on each duplicate gene (41). For three recently duplicated (ca. 0.25–1.2 million years) *Arabidopsis* genes, both progenitor and derived copies show significantly reduced species-wide polymorphism (73). Paleo-duplicated yeast genes provide a discernible degree of functional compensation for a very long time (37) and even appear to undergo gene conversion (33).

Contrary to classical predictions that duplicated genes may be relatively free to acquire unique functionality, among both *Arabidopsis* ecotypes and *Oryza* subspecies single-nucleotide polymorphisms encode less radical amino acid changes in genes for which there exists a duplicated copy at a paleologous locus, than in singleton genes (10). Genes encoding long and complex proteins are preferentially preserved in duplicate, and evolve conservatively. Genes for which an ancient duplicated

**Neofunctionalization:** the evolution of a new function in a duplicated gene

copy has been preserved are 25% and 112% longer, on average, than singletons in *Arabidopsis* and *Oryza*, respectively, and a much higher fraction of the coding regions comprises identifiable domains. The greater retention in duplicate of longer and more complex proteins, together with their lower tolerance of nonsynonymous mutations, suggest that the immediate benefits of buffering crucial functionality (i.e., ensuring that the functions of essential genes and domains are met even after mutation of one copy) are accrued by the presence of a second copy, regardless of the potential advantage conferred by the possibility of future neofunctionalization. The tendency for duplicated gene copies to accumulate only less severe amino acid substitutions than do singletons in natural populations is in contrast to the widely held belief that a primary advantage of polyploidy is the freedom for duplicated genes to acquire new functions (76, 101, 104). However, it is consistent with recent results about the evolution of recently evolved duplicates (41, 73) and with long-term functional compensation (24, 37, 105). Genetic buffering may be especially important for surviving the early genomic shock that can immediately follow polyploid formation (26, 48, 78, 79, 94, 99). However, such buffering still markedly affects diversity among modern ecotypes of *A. thaliana* and subspecies of *Oryza sativa* (10), species whose genomes were shaped by duplications that occurred 60 million years or more prior to speciation (12, 80).

What mechanisms might preserve the sequence of thousands of pairs of genes at paleologous sites across a genome for 60 million years or more? Occasional nonhomologous associations between chromosomes are observed during mitosis in many taxa, including recently formed polyploids (84) and even in diploids such as rice (51), and might periodically permit homogenization processes to act between paleologs. Mechanisms such as gene repair (57) and gene conversion (75) have been suggested to occur between ancient duplicated genes (33), but can be difficult to quantify (120).

The availability of genome sequences for closely related paleopolyploid taxa offers the opportunity to take a phylogenetic approach to making inferences about concerted evolution of genes, identifying cases in which ancient paralogs that inhabit a common nucleus are more similar than recently diverged orthologs in different species. By applying this approach to genome sequences for two *Oryza* subspecies, Wang et al. (114) reveal appreciable gene conversion in the ∼0.4 million years since their divergence, with a gradual progression toward independent evolution of older paralogs. Sequence similarity analysis in proximal gene clusters suggests more conversion between younger than older paralogs. Domain-encoding sequences are more frequently converted than nondomain sequences, suggesting a sort of circularity: Sequences conserved by selection may be further conserved by relatively frequent conversion. The prevalence of concerted evolution among homoeologous and tandem paralogs is surprising but consistent with previous findings for genes involved in olfaction, immune response, HLA, MHC, sex or reproductive isolation, mating type, multiallelic systems, and tissue- or time-specific expression (9, 93) and involving a wide range of species including yeast, flies, plants, and mammals. Is the preservation of duplicated sequences merely an accident, an occasional aberrant result of the mechanics of DNA recombination and repair? Or, could it have become an integral part of the adaptation of polyploid genomes to the duplicated state?

## Might Genome Duplication be Truly Cyclical in the Angiosperms?

In taxa that lack major obstacles to genome duplication such as sex chromosomes, might the purported benefits of genome duplication, followed by their gradual deterioration resulting from genetic divergence, impart cyclicality to this process? Most angiosperm lineages sequenced to date have experienced more than one whole genome duplication event; the only

exceptions are grape (46) and papaya (70), the latter with incipient sex chromosomes (59). The sequences of most duplicated genes may be sheltered as a consequence of concerted evolutionary mechanisms (for example, gene conversion) during the period of instability immediately following genome duplication. Sheltering may be especially important for avoiding the deleterious effects of Muller's ratchet (74) under asexual reproductive systems, perhaps partly explaining why so many apomicts (8) and other clonally propagated angiosperms are polyploids. For domain-rich genes, which tend to be preferentially preserved in duplicate (17), the exons might be subject to concerted evolution for much longer time periods.

However, since ectopic recombination in plants is dramatically reduced by even small variations in DNA sequence (54), one can envision an exponential decline in conversion frequencies with eventual independence of these once-conserved sequences. Is each angiosperm genome on a sort of timeline? Does the occasional genome duplication reinvigorate an otherwise continuously declining supply of raw material remaining available for the evolution of novelty, with a continuously declining ability to shelter essential genes and functions with duplicated copies? Conversely, is there a minimum period between polyploid events in a single lineage, measured by time or by degree of diploidization? Answers to questions such as these may require in-depth structural and functional analysis of a broad sampling of natural polyploid lineages resulting from polyploidizations that range from recent to ancient.

## CONCLUSIONS

What are the key attributes that make polyploidy such a prevalent phenomenon? Is polyploidy even a single phenomenon, given the likelihood that different features of polyploidy may predominate among various lineages? All polyploids by definition have in common a larger than diploid number of genomes, hence the widespread use of the term "genome doubling" as a synonym for polyploidy. But is this synonymy warranted? Are the most important evolutionary properties of polyploidy due to genome doubling, per se, or is hybridization—genome merger–just as important? Furthermore, are all of these important properties even shared by all polyploids?

There are numerous genetic and genomic differences among various types of polyploids, and it is likely that these differences are more significant than the commonality of possessing multiple genomes. One principal distinction has to do with the genetic and genomic degree to which a given polyploid is hybrid. This quantitative distinction may be critical, in that the interactions established by the initial conditions propagate from the time of initial origin through periods of stabilization and longer-term evolutionary outcome.

Based on the still-limited number of available examples, when genomic shock occurs in nascent polyploids, hybridity is a more significant source of genetic and epigenetic disruption than is genome doubling. However, $F_1$ hybrids between *Tragopogon dubius* and *T. porrifolius* and between *T. dubius* and *T. pratensis* show genetic profiles that are additive of their diploid parents, demonstrating that hybridity alone is not responsible for the genomic changes observed in the allotetraploid derivatives of these crosses, *T. mirus* and *T. miscellus*, respectively (103). Polyploids that do not merge differentiated genomes presumably do not experience the full benefit of genomic shock—if indeed it is a benefit. Most mutations are detrimental, so most hybrid-induced mutation is likely not adaptive. Perhaps intraspecific hybridization (one definition of autopolyploidy) is a safer path to a doubled genome than interspecific crossing—our increasing appreciation of the prevalence of autopolyploidy certainly supports the notion that this is a very successful evolutionary pathway (97). However, interspecific crossing (one definition of allopolyploidy) may more frequently produce a novel evolutionary

trajectory. In other words, "nothing ventured, nothing gained."

Two potential benefits of hybridization are heterosis and the evolution of adaptive transgressive traits, be they physiological or morphological. Polyploidy promotes both, but different types of polyploids are expected to differ in their ability to do so. Classic disomic allopolyploids are fixed for heterozygosity at homoeologous loci and therefore may be more likely to experience such phenomena as allele-specific expression subfunctionalization and its attendant evolutionary possibilities. In contrast, a polysomic autopolyploid possesses a greater than diploid number of alleles at its origin, but these remain vulnerable to segregational loss; also, the ability to experiment with subfunctionalization and genetic novelty likely is more limited in autopolyploid systems. Transgressive effects should also be preserved by fixed hybridity in disomic systems more effectively than in polysomic autopolyploids. Indeed, experimental analysis of autopolyploid sugarcane populations suggests that pyramiding of individually favorable alleles may yield diminishing or even negative returns (69, 71). Thus the breadth of the original cross and the genetic system of the polyploid should together have a major effect on the creation and retention of evolutionary novelty.

The discussion of preservation of duplicate genes over longer evolutionary timescales in plants usually has been framed in terms of extensively diploidized taxa, such as *A. thaliana* and *O. sativa*. The genome donors of these taxa are unknown, as is the degree to which hybridity shaped their evolution. The fact that overlapping patterns of gene loss and retention are observed among plants, animals, and yeast suggests that there are widely applicable principles that guide adaptation of genomes to duplication. As noted above, little is understood about the temporal dynamics by which this long-term fractionation occurred, and hence the study of more recently generated polyploids, both natural and synthetic, holds promise for connecting genome fractionation to the early responses of genomes to polyploidization. Thus, it would be useful to know more about how the process of gene retention and loss plays out in a wider range of polyploid examples, and over wider timescales (e.g., young polyploids), including plants with polysomic segregation.

Indeed, a general conclusion of the state of genetic and genomic research in polyploid evolution is that more models are needed to complement "classic" systems such as wheat, cotton, and *Brassica* and newer models such as *Arabidopsis*. We need to extend from a few models (mostly crops) to encompass natural systems that represent a diversity of plant families, as well as ages, stages, and types of polyploids. Progress is certainly being made, and studies of recently formed polyploids in *Senecio* (39) and *Spartina* (91) hold great promise for revealing the dynamics of polyploid evolution in an ecological context, as do young polyploids such as *Tragopogon*, where studies of the transcriptome have been initiated (67, 103). Somewhat older allopolyploids such as perennial *Glycine* species are also being exploited in studies of gene expression (47) and offer future promise. Some species that are classics in their own right, but as population genetic and evolutionary models rather than genomic ones, such as autopolyploid *Chamerion* (42), are excellent subjects for studies of gene expression and genomic change, and undoubtedly will make future contributions in these areas.

Broadening the sampling of polyploids, and performing comparable surveys and experiments across taxa, will ultimately reveal whether polyploidy is characterized by emergent, widely applicable principles, or, conversely, whether different classes of polyploids, or even different genotypes of the same polyploid, behave idiosyncratically. In addition, this broadly comparative approach will help provide the necessary links between initial conditions at the onset of genome merger and genome doubling and the long-term evolutionary outcomes evident in modern, sequenced genomes. Only with this information can the genetic and genomic aspects of the question "Why are polyploid plants so successful and prevalent" be addressed in a meaningful way.

# DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

# ACKNOWLEDGMENTS

# LITERATURE CITED

1. Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8:135–41
2. **Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* 100:4649–54**
3. Adams KL, Percifield R, Wendel JF. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168:2217–26
4. Adams KL, Wendel JF. 2004. Exploring the genomic mysteries of polyploidy in cotton. *Biol. J. Linn. Soc.* 82:573–81
5. Albertin W, Balliau T, Brabant P, Chevre A, Eber F, et al. 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics* 173:1101–13
6. Auger DL, Gray AD, Ream TS, Kato A, Coe EH Jr, Birchler JA. 2005. Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics* 169:389–97
7. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–78
8. Bayer RJ, Stebbins GL. 1987. Chromosome-numbers, patterns of distribution, and apomixis in *Antennaria* (Asteraceae, Inuleae). *Syst. Bot.* 12:305–19
9. Bettencourt BR, Feder ME. 2002. Rapid concerted evolution via gene conversion at the *Drosophila* hsp70 genes. *J. Mol. Evol.* 54:569–86
10. Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–91
11. Bottley A, Xia GM, Koebner RMD. 2006. Homoeologous gene silencing in hexaploid wheat. *Plant J.* 47:897–906
12. **Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–38**
13. Buggs RJA, Soltis PS, Mavrodiev EV, Symonds VV, Soltis DE. 2008. Does phylogenetic distance between parental genomes govern the success of polyploids? *Castanea*. 73:74–79
14. Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245
15. Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biol.* 7:R13
16. Chantret N, Salse J, Sabot F, Rahman S, Bellec A, et al. 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 17:1033–45
17. Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. *Proc. Natl. Acad. Sci. USA* 103:2730–35
18. Chapman MA, Burke JM. 2007. Genetic divergence and hybrid speciation. *Evolution* 61:1773–80
19. Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58:377–406

Early demonstration of expression subfunctionalization in a polyploid species, suggesting wide scope of this phenomenon.

Integration of genomic and phylogenetic approaches improved understanding of angiosperm evolutionary history.

20. Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28:240–52

21. Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA* 101:700–7

22. Clark RM, Wagler TN, Quijada P, Doebley J. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38:594–97

23. Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6:836–46

24. Conant GC, Wagner A. 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. R. Soc. London Ser. B* 271:89–96

25. Dong YZ, Liu ZL, Shan XH, Qiu T, He MY, Liu B. 2005. Allopolyploidy in wheat induces rapid and heritable alterations in DNA methylation patterns of cellular genes and mobile elements. *Russ. J. Genet.* 41:890–96

26. Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* 147:1381–87

27. **Flagel L, Udall JA, Nettleton D, Wendel JF. 2008. Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.* 6:11**

28. Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap E, et al. 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289:85–88

29. Fu H, Dooner HK. 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* 99:9573–78

30. Fujimori S, Washio T, Tomita M. 2005. GC-compositional strand bias around transcription start sites in plants and fungi. *BMC Genomics.* 6:26

31. **Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19:3403–17**

32. Ganko EW, Meyers BC, Vision TJ. 2007. Divergence in expression between duplicated genes in *Arabidopsis*. *Mol. Biol. Evol.* 24:2298–309

33. Gao LZ, lnnan H. 2004. Very low gene duplication rate in the yeast genome. *Science* 306:1367–70

34. Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* 94:6809–14

35. Goff SA, Ricke D, Lan TH, Presting G, Wang R, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100

36. Grant V. 1981. *Plant Speciation*. New York: Columbia Univ. Press. 2nd ed.

37. Gu ZL, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421:63–66

38. Guo M, Yang S, Rupe M, Hu B, Bickel D, et al. 2008. Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS) reveals *cis*- and *trans*-effects on gene expression in maize hybrid meristem tissue. *Plant Mol. Biol.* 66:551–63

39. Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr. Biol.* 16:1652–59

40. Hovav R, Udall JA, Chaudhary B, Flagel L, Rapp R, Wendel JF. 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl. Acad. Sci. USA* 105:6191–95

41. Hughes MK, Hughes AL. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol. Biol. Evol.* 10:1360–69

42. Husband B, Ozimec B, Martin S, Pollock L. 2008. Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids. *Int. J. Plant Sci.* 169:195–206

43. Innan H. 2003. The coalescent and infinite-site model of a small multigene family. *Genetics* 163:803–10

44. Jablonka E, Lamb MJ. 2002. The changing concept of epigenetics. *Ann. NY Acad. Sci.* 981:82–96

45. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–57

46. **Jaillon O, Aury JM, Noel B, Policriti A, Clepet C. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–67**

A custom microarray was used to partition homoeolog expression bias into those arising from genomic merger and allopolyploidy.

Homoeologous recombination creates novel phenotypic diversity among 47 synthetic allopolyploid lines.

Unusually high retention of duplicated (triplicated) genes in grape clarifies the nature of an early polyploidization that shaped much of eudicot evolution.

47. Joly S, Rauscher JT, Sherman Broyles SL, Brown AHD, Doyle JJ. 2004. Evolutionary dynamics and preferential expression of homeologous 18S-5.8S-26S nuclear ribosomal genes in natural and artificial *Glycine* allopolyploids. *Mol. Biol. Evol.* 21:1409–21

48. Kashkush K, Feldman M, Levy AA. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics.* 160:1651–59

49. Kawabe A, Miyashita NT. 2003. Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.* 78:342–52

50. King M, Wilson A. 1975. Evolution on two levels in humans and chimpanzees. *Science* 188:107–16

51. Lawrence WJC. 1931. The secondary association of chromosomes. *Cytologia* 2:352–84

52. Levin DA. 1983. Polyploidy and novelty in flowering plants. *Am. Nat.* 122:1–25

53. Levy AA, Feldman M. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biol. J. Linn. Soc.* 82:607–13

54. Li LL, Jean M, Belzile F. 2006. The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. *Plant J.* 45:908–16

55. Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, et al. 2007. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytol.* 175:756–63

56. Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* 44:321–30

57. Liu L, Parekh-Olmedo H, Kmiec EB. 2003. The development and regulation of gene repair. *Nat. Rev. Genet.* 4:679–89

58. Liu X. 2007. Getting started in tiling microarray analysis. *PLoS Comput. Biol.* 3:e183

59. Liu ZY, Moore PH, Ma H, Ackerman CM, Ragiba M, et al. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427:348–52

60. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, et al. 2003. Allelic variation in gene expression is common in the human genome. *Genome Res.* 13:1855–62

61. Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T. 2005. Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. *Plant Physiol.* 140:336–48

62. Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–73

63. Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–804

64. Ma X-F, Gustafson JP. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet. Genome Res.* 109:236–49

65. **Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiol.* 129:733–46**

66. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, et al. 2005. Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* 102:5454–59

67. Matyasek R, Tate JA, Lim YK, Srubarova H, Koh J, et al. 2007. Concerted evolution of rDNA in recently formed *Tragopogon* allotetraploids is typically associated with an inverse correlation between gene copy number and expression. *Genetics* 176:2509–19

68. Mette MF, Aufsatz W, Van Der Winden J, Matzke MA, Matzke AJ. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* 19:5194–201

69. Ming R, Wang Y, Draye X, Moore P, Irvine J, Paterson AH. 2002. Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. *Theor. Appl. Genet.* 105:332–45

70. Ming R, Hou S, Feng Y, Yu QY, Dionne-Laporte A, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya*, Linnaeus). *Nature* 452:991–96

71. Ming R, Liu S, Moore PH, Irvine JE, Paterson AH. 2001. QTL analysis in a complex autopolyploid: Genetic control of sugar content in sugarcane. *Genome Res.* 11:2075–84

Treating reconstituted allopolyploids with a methylation disruptor caused novel phenotypes in the allopolyploids.

72. Mochida K, Yamazaki Y, Ogihara Y. 2003. 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* 270:371–77

73. Moore RC, Purugganan MD. 2003. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. USA* 100:15682–87

74. Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 1:2–9

75. Newman T, Trask BJ. 2003. Complex evolution of 7E olfactory receptor genes in segmental duplications. *Genome Res.* 13:781–93

76. Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag

77. Ownbey M. 1950. Natural hybridization and amphiploidy in the genus *Tragopogon*. *Am. J. Bot.* 37:487–99

78. Ozkan H, Levy AA, Feldman M. 2002. Rapid differentiation of homeologous chromosomes in newly-formed allopolyploid wheat. *Isr. J. Plant Sci.* 50:S65–76

79. Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13:1735–47

80. Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. USA* 101:9903–8

**81. Paterson AH, Chapman BA, Kissinger J, Bowers JE, Feltus FA, et al. 2006. Convergent retention or loss of gene/domain families following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces*, and *Tetraodon*. *Trends Genet.* 22:597–602**

82. Pires JC, Zhao J, Schranz ME, Leon EJ, Quijada PA, et al. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biol. J. Linn. Soc.* 82:675–88

83. Pontes O, Li C, Nunes PC, Haag J, Ream T, et al. 2006. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126:79–92

84. Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* 33:589–639

85. Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29:467–501

86. Rapp R, Wendel JF. 2005. Epigenetics and plant evolution. *New Phytol.* 168:81–91

87. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. 2002. MicroRNAs in plants. *Genes Dev.* 16:1616–26

88. Rifkin SA, Kim J, White KP. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* 33:138–44

89. Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, et al. 2002. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* 109:437–46

90. Rockman MV, Kruglyak L. 2006. Genetics of global gene expression. *Nat. Rev. Genet.* 7:862–72

**91. Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol. Ecol.* 14:1163–75**

92. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–45

93. Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* 20:461–64

94. Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13:1749–59

95. Shannon MF, Rao S. 2002. Of chips and ChIPs. *Science* 296:666–69

96. Shoemaker RC, Schlueter J, Doyle JJ. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Curr. Opin. Plant Biol.* 9:104–9

97. Soltis DE, Soltis PS, Schemske DW, Hancock JS, Thompson JN, et al. 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon* 56:13–30

98. Soltis DE, Soltis PS, Tate JA. 2004. Advances in the study of polyploidy since *Plant Speciation*. *New Phytol.* 161:173–91

99. Song KM, Lu P, Tang KL, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. USA* 92:7719–23

**Early exploration of the concept that differential gene loss may be associated with adaptation to the polyploid state.**

**Invasive allopolyploid weed lacking genetic diversity has extensive methylation variation compared to its parental donors.**

100. Stebbins GL. 1971. *Chromosomal Evolution in Higher Plants*. London: Addison-Wesley

101. Stephens S. 1951. Possible significance of duplications in evolution. *Adv. Genet.* 4:247–65

102. Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, et al. 2007. Phenotypic and transcriptomic changes associated with potato autopolyploidization. *Genetics* 176:2055–67

103. Tate JA, Ni Z, Scheen AC, Koh J, Gilbert CA, et al. 2006. Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173:1599–611

104. Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genetics* 38:615–43

105. Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. *Nat. Genetics* 36:492–96

**106. Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* 16:934–46**

107. Thompson JD, Nuismer SL, Merg K. 2004. Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biol. J. Linnean Soc.* 82:511–19

108. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. 2005. Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proc. Natl. Acad. Sci. USA* 102:8933–38

109. Udall JA, Quijada PA, Osborn TC. 2005. Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* 169:967–79

110. van Steensel B. 2005. Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat. Genet.* 37:s18–24

111. Wagner A. 2000. Robustness against mutations in genetic networks of yeast. *Nat. Genet.* 24:355–61

**112. Wang J, Tian L, Lee H, Wei NE, Jiang H, et al. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172:507–17**

113. Wang J, Tian L, Madlung A, Lee H, Chen M, et al. 2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genetics* 167:1961–73

114. Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH. 2007. Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics* 77:1753–63

115. Wendel JF. 2000. Genome evolution in polyploids. *Plant. Mol. Biol.* 42:225–49

116. Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430:85–88

117. Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–13

118. Wu J, Smith LT, Plass C, Huang THM. 2006. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res.* 66:6899–902

119. Zhang B, Wang Q, Wang K, Pan X, Liu F, et al. 2007. Identification of cotton microRNAs and their targets. *Gene* 397:26–37

120. Zhang LQ, Vision TJ, Gaut BS. 2002. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19:1464–73

**Gene loss among *Arabidopsis* paleoduplicate genome segments is shown to proceed in a biased manner.**

**Synthetic *Arabidopsis* polyploids experience genomic dominance: maternal fraction of differentially expressed genes is down-regulated.**