ELSEVIER

The evolutionary dynamics of plant duplicate genes

Richard C Moore and Michael D Purugganan

Given the prevalence of duplicate genes and genomes in plant species, the study of their evolutionary dynamics has been a focus of study in plant evolutionary genetics over the past two decades. The past few years have been a particularly exciting time because recent theoretical and experimental investigations have led to a rethinking of the classic paradigm of duplicate gene evolution. By combining recent advances in genomic analysis with a new conceptual framework, researchers are determining the contributions of single-gene and whole-genome duplications to the diversification of plant species. This research provides insights into the roles that gene and genome duplications play in plant evolution.

Addresses
Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695, USA

Corresponding authors: Moore, Richard C (remoore@unity.ncsu.edu) and Purugganan, Michael D (michaelp@unity.ncsu.edu)

Current Opinion in Plant Biology 2005, **8**:122–128

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Susan Wessler and James C Carrington

Available online 29th January 2005

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2004.12.001

Introduction

The evolutionary diversification of genomes and genetic systems is driven in part by duplication events [1•,2]. Gene duplications contribute to the establishment of new gene functions [3] and underlie the origins of evolutionary novelty [4]. Plants are exceptional among eukaryotic organisms in that duplicate loci compose a large fraction of their genomes, partly because of the frequent occurrence of genomic segmental duplications and polyploidization events in plants. For example, in the *Arabidopsis thaliana* and rice genomes up to 90% and 62% of loci are duplicated, respectively, and it is estimated that 70–80% of angiosperm species have undergone polyploidization at some point in their evolutionary history [5–7,8••,9•].

The high proportion of duplicate genes in plant genomes reflects the rate of retention of duplicate copies among plant species [10,11]. The birth rate of genes in *A. thaliana* (0.002 duplicates per gene per million years) is in the same order of magnitude as that observed for yeast and *Drosophila*, although ten-fold slower than that found in *Caenorhabditis elegans*. The half-life to silencing and loss of a gene duplicate in *A. thaliana*, however, is estimated at 23.4 million years, which is 3–7-fold higher than for animal genomes [11]. It is not clear why duplicate genes persist for longer periods in *A. thaliana* than in animal genomes, nor is it known whether this is a general characteristic of plant genomes. Understanding the mechanisms underlying the retention of duplicate genes in plant genomes continues to remain a topic of intense interest [8••,9•,12,13••,14,15].

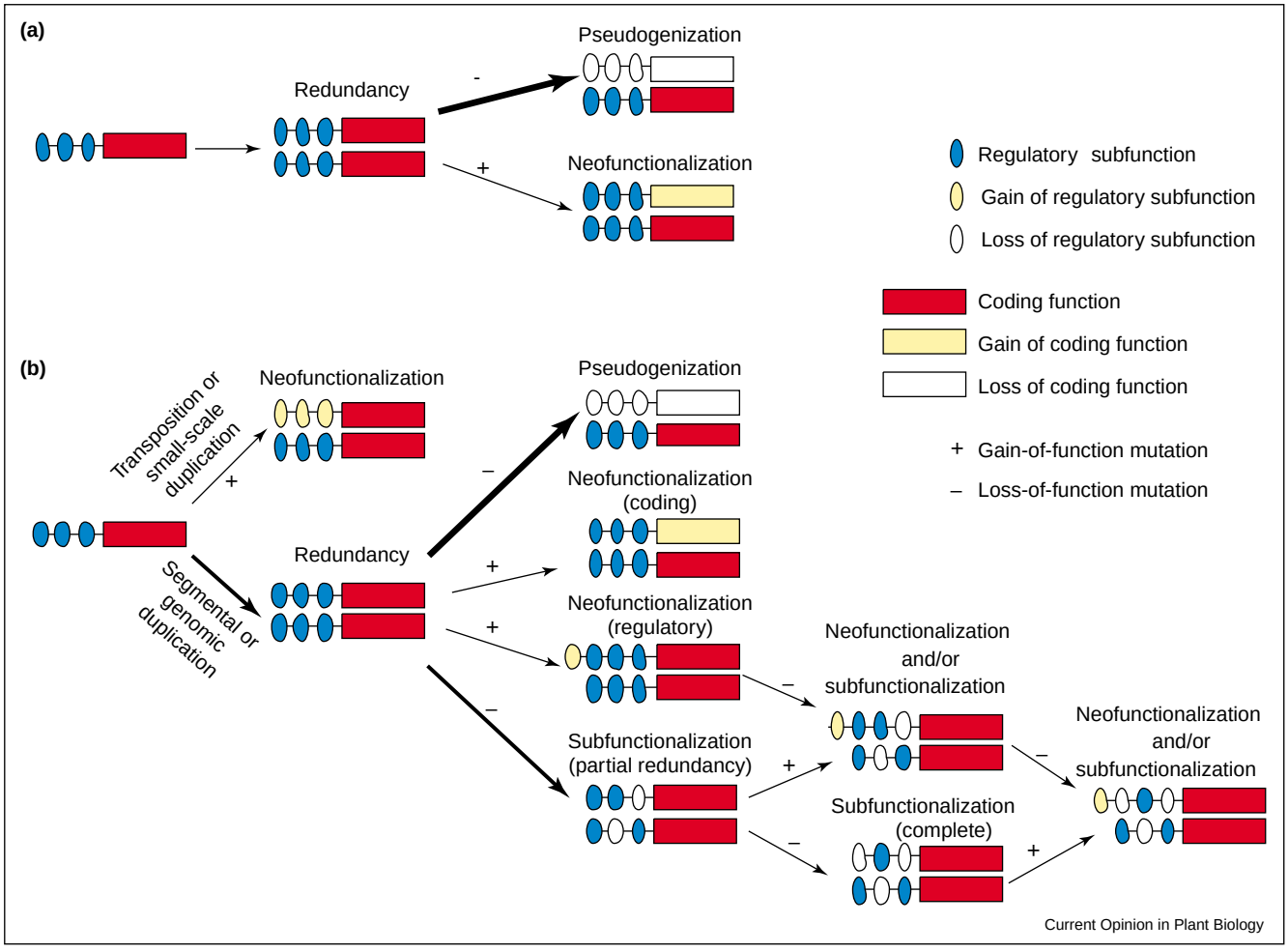
Theoretical and experimental studies in the past few years have advanced our understanding of the evolutionary dynamics of duplicate loci in plant genomes, and have led to a rethinking of the paradigm that provided the conceptual foundations of studies on duplicate gene evolution [10,13••,16,17••]. The confluence of a new conceptual framework for gene duplication with new genomic technologies has allowed investigators to study gene duplications on a genome-wide scale [8••,9•,12, 14,18••]. This, in turn, has reshaped our understanding of the evolution of duplicate genes and genomes.

The fates of duplicate genes: a new paradigm emerges

In the thirty years since the publication of Ohno's landmark book entitled '*Evolution by Gene Duplication*' [2], the reigning paradigm regarding the fate of duplicated genes predicts that one of the duplicates is either lost (pseudogenization) or gains a new function (neofunctionalization) (Figure 1a). Because deleterious mutations are more probable than advantageous mutations, it has been presumed that most duplicate copies are lost and only a few neofunctional loci are maintained by selection.

It has since become apparent that positive selection does play a key role in preserving some gene copies, and indeed can act at a very early stage of the gene duplication process. For example, a population genetic analysis of three unlinked duplicate gene pairs in *A. thaliana* that originated less than 1.2 million years ago (mya) revealed significantly reduced levels of nucleotide polymorphism in the progenitor locus, the duplicate locus or both. This reduced nucleotide variation, which is associated with a recent selective sweep, is evidence that positive selection plays a prominent role in the establishment of duplicate loci [17••]. Although the action of positive selection is consistent with the process of neofunctionalization, the precise targets of selection in these duplicate genes are unclear in the absence of functional data [19]. It will also be of interest to determine whether selection acts to

Figure 1



(a) Classic Ohno model of duplicate gene fates. Mechanisms of duplication and fates of genes are indicated. Thickness of arrows indicate relative frequency of possible fates. **(b)** Recent theoretical work supports a much more complex model for the fates of duplicate genes.

preserve tandem-linked duplicate loci, which are prevalent in plant genomes [14,17••] and for which neofunctionalization is theoretically unlikely [20].

Although Ohno's classic paradigm provides key insights into duplicate gene evolution, it fails to account for the preponderance of retained duplicates in whole genomes [21]. The duplication-degeneration-complementation (DDC) model [16,22], which harkens to the 'gene sharing' concept proposed by Hughes [23], suggests an additional evolutionary fate for duplicate loci. This model, also referred to as subfunctionalization, suggests that duplicate genes acquire debilitating yet complementary mutations that alter one or more subfunctions of the single gene progenitor (Figure 1b). The strength of this model is that it does not rely on the sparse occurrence of beneficial mutations, but on more frequently occurring loss-of-function mutations in regulatory regions [16]. Subsequent empirical studies on expression divergence between duplicate genes suggest that changes in expression regimes occur both frequently and rapidly, consistent with the predictions of this model [13••,24,25].

The plant MADS-box gene family: a case study of the fate of duplicate genes

The diversification of the MADS-box transcription factor family, whose members control key aspects of plant vegetative and reproductive development, is a clear example of the role that subfunctionalization plays in shaping genetic systems. Indeed, one of the earliest-documented examples of subfunctionalization in plants is the lineage-specific duplication of an *AGAMOUS* (*AG*)-like MADS-box gene in maize [26]. In maize, the ancestral function of *AG* (which is expressed in stamens and carpels) is partitioned between the duplicated genes *ZMM2* (expressed primarily in stamens) and *ZAG1* (expressed primarily in carpels) [26]. A more recent example of subfunctionalization involving MADS genes

comes from the dioecious plant *Silene latifolia*, in which the male sex is determined by the presence of a morphologically distinct Y chromosome. In this species, an autosomal homolog of the MADS gene *APETALA3* (*AP3*), which is involved in petal and stamen identity, was duplicatively transferred to the Y chromosome, and subsequently underwent divergence in gene expression [27].

On a larger evolutionary time-scale, the evolutionary forces that govern the fates of duplicate genes are responsible for the incredible diversity of MADS-box genes found in present-day flowering plants. MADS-box genes are found ubiquitously in eukaryotes as type I and II gene subfamilies [28]. The proliferation of type II MADS-box genes, known as the MIKC-class genes, is unique to plants [29]. A comparative genomic survey indicates that although the birth rate of type II MADS-box genes in plants is lower than that of type I genes, the former are preferentially retained in the genome [30•]. This pattern suggests that the loss of type II MADS genes is more deleterious than the loss of type I loci, a prediction supported by the observation that type II genes tend to be subject to stronger purifying selection. The lower death rate of type II MADS genes might be due to subfunctionalization if the divergence in function that is observed across this subfamily necessitates the retention of family members in the genome [31•,32]. The selective retention of type II MADS genes also creates a certain degree of redundancy between closely related paralogs [33], however, suggesting that only partial subfunctionalization occurs in certain clades.

An examination of the *AG* subfamily of MADS-box genes, which controls key aspects of inner whorl floral development in flowering plants [34•], reveals evidence of both subfunctionalization and overlapping redundancy. The *AG* subfamily is subdivided via an ancient duplication at the base of the angiosperms into the C-class and D-class lineages. The D-class genes are almost all expressed in the ovules, where they specify ovule identity, whereas the C-class lineage is primarily involved in carpel and stamen identity [35]. By contrast, an *AG*-like gene in gymnosperms is expressed in both the female reproductive structure (megasporophyll) and the ovule [36–39], suggesting that the genetic function of the ancestral gene in this group was partitioned into two distinct lineages within the angiosperms.

Another issue raised by this study is whether the partitioning of subfunctions is equally shared between lineages. The *Antirhinnum majus PLENA* (*PLE*) gene was presumed to be an ortholog of the *A. thaliana AG* gene because mutant analysis indicated that both genes provide the primary C-class function in their respective species [40,41]. However, the true ortholog of *AG* in *A. majus* is *FARINELLI* (*FAR*), which functions redundantly in specifying stamen identity [34•,42]. By contrast, the orthologs of *PLE* in *A. thaliana* are the *SHATTERPROOF* (*SHP*) genes, which primarily control differentiation of the seed valve margin while sharing overlapping redundancy with *AG* [33,43]. Thus, subfunctions of the progenitor gene were asymmetrically partitioned between *AG/SHP* paralogs in the *A. thaliana* lineage relative to the *PLE/FAR* paralogs in the *A. majus* lineage. This functional asymmetry also underscores the possible difficulties in using functional similarity as evidence for evolutionary orthology because the latter does not necessarily lead to the former.

All is not lost: redundancy remains

One of the predictions of the subfunctionalization model is that duplicate genes will share overlapping redundant functions early in the process of functional divergence, and phylogenetic and functional analyses clearly support this prediction [16,33,34•,43,44]. Within *A. thaliana*, the C-class genes *AG* and *SHP1/SHP2*, along with the D-class gene *SEEDSTICK* (*STK*), share overlapping functions in carpel identity, a remnant of their shared ancestral role [33]. Given that these paralogs diverged at the base of the angiosperm lineage, it is clear that duplicate genes can exist stably in at a partially redundant state over a protracted evolutionary period. Although subfunctionalization can occur rapidly, it is not clear when the transition to complete functional partitioning occurs. Indeed, theoretical considerations of redundancy predict that duplicate genes will reach an evolutionarily stable equilibrium of partial redundancy, potentially delaying the transition to complete subfunctionalization [45].

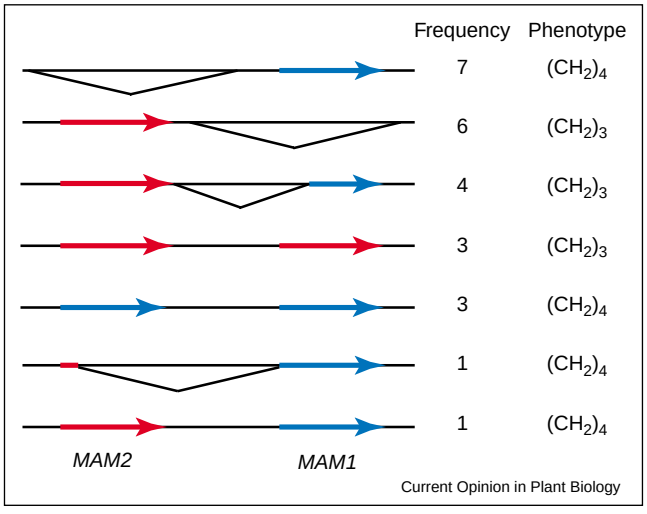
Although complete functional redundancy is difficult to confirm, numerous examples of paralogous genes in plants for which the deletion of one copy yields no observable phenotype do exist [46–52], including the three *SEPALLATA1/2/3* (*SEP1/2/3*) and two *SHP1/SHP2* MADS-box paralogs [43,44,53]. The protein changes in these paralogous redundant genes show no evidence of functional divergence, suggesting that amino-acid replacements are functionally constrained [54]. Although purifying selection apparently constrains divergence between paralogs, population genetic analyses indicate that significant heterogeneity in nucleotide diversity, suggestive of positive selection, occurs within the transcriptional unit of *SHP1* and the promoter of *SHP2*. By contrast, *SEP2* and *SEP1* appear to be evolving neutrally [55]. These differences in evolutionary dynamics between paralogs might be a consequence of the position of these genes in the floral developmental pathway. The *SEP* loci are involved in the highly conserved function of floral patterning, whereas the *SHP* loci control the more variable trait of seed shattering.

Duplicate and defend

The duplication of genes within or between closely related species might lead to phenotypic variation in

specific traits, and manifest itself as functionally relevant, ecologically significant polymorphisms. Gene duplication, for example, contributes to the ability of plants to mount a defense response against disease and herbivory by allowing the functional diversifcation of genes that are involved in pathogen recognition and herbivory defense. This diversification is most evident in the large family of disease resistance (*R*) genes that encode the nucleotide-binding site plus leucine-rich repeat (NBS–LRR) proteins. The *Arabidopsis* genome contains more than 150 NBS–LRR genes found as isolated genes or in tandemly arrayed clusters that are dispersed throughout the genome. Gene duplication of NBS–LRRs followed by positive selection for diverse amino acids in the LRR protein-recognition domain might provide the means by which the plant's recognition of novel pathogens evolves [56]. Interestingly, a genomic analysis of the history and chromosomal locations of NBS–LRR gene duplications revealed that tandem duplications have contributed to the birth of the majority of NBS–LRRs in the *Arabidopsis* genome, whereas the duplicative transfer of NBS–LRRs to dispersed chromosomal locations is largely the result of segmental duplication [55]. Because tandemly arrayed NBS–LRRs are subject to frequent intergenic exchange, it is believed that those genes found in different chromosomal regions have a greater chance of evolving new pathogen recognition functions [56]. Moreover, a recent study of the NBS–LRR *Cf-2* pathogen-resistance gene family in the wild species *Solanum pimpinellifolium* suggests that complex evolutionary dynamics surround duplicate *Cf-2* genes, including the differential selection of gene copies [57].

Figure 2



The genotypic frequencies and structural organization of the *MAM* duplicate loci from a survey of 25 *A. thaliana* ecotypes. The *MAM* loci modify the carbon length of the glucosinolate basic side chain; *MAM2* adds three methylene groups ($[CH_2]_3$), whereas *MAM1* adds four methylene groups ($[CH_2]_4$). Slanted lines indicate the loss of particular duplicate copies. (Based on Figure 1 of [59•].)

Gene duplication also plays a significant role in the evolutionary dynamics of anti-herbivory genes. In *A. thaliana*, resistance to generalist insect herbivores is conferred in part by glucosinolates, a class of plant secondary metabolites. The composition and quality of glucosinlate production varies between *Arabidopsis* accessions. This variation is due, in part, to the tandemly duplicated methylthioalkylmalate synthase (*MAM*) genes, *MAM1* and *MAM2*, that are involved in glucosinolate biosynthesis [58]. *MAM1* and *MAM2* are functionally divergent duplicates: the *MAM2* locus is a stronger deterrent against generalist herbivores than the *MAM1* locus [59•]. A population genetic analysis of the *MAM* genes found evidence of differential gene loss and differential selection between the *MAM* loci within *A. thaliana* accessions (Figure 2; [59•]). These data support the action of balancing selection on the *MAM2* locus but not on *MAM1*. Because glucosinolates can also stimulate feeding in specialist herbivores, selection at the *MAM2* locus might be caused by the ecological trade-off between protection against generalist insect herbivores and increased susceptibility to specialists.

Variable patterns of evolution exist in another class of anti-herbivory genes, the tandemly arrayed family of six *A. thaliana* trypsin inhibitor (*ATTI*) genes. The expression of these six genes varies significantly between loci, among allelic classes within *A. thaliana*, and in response to herbivory [60]. These data suggest that subfunctionalization is acting to maintain divergent functions between duplicate loci. Despite their close proximity to each other (within 10 kb), population genetic analyses suggest that evolutionary history also varies between *ATTI* loci, although the flanking loci, *ATTI1* and *ATTI6*, which are separated from the core loci by recombination, have the greatest opportunity to be acted upon by natural selection [60,61].

Polyploids: ancient and new

Polyploidization is a major mechanism by which duplicate genes are introduced into plant genomes, and this is reviewed elsewhere in this issue (see review by Adams and Wendel). Several aspects of polyploidization do, however, highlight several interesting facets that surround the evolutionary dynamics of duplicate genes in plants. In a recent genomic analysis of 14 model plant species, nine species were documented to have one or more whole-genome duplication events in their evolutionary past [18••]. Up to four whole-genome duplication events appear to have occurred in *A. thaliana* [6,7,8••,62,63]. Genomics technologies might now provide a systematic explanation of the fates of duplicate genes on a whole-genome level.

As predicted by theory [20,64,65], the majority of duplicated genes resulting from polyploidization events are lost in the transition to diploidy, although the percentage of retained duplicates varies between species [10,11]. The type of duplicate genes lost or retained after

polyploidization is correlated with function. Among the retained duplicates from the last whole-genome duplication event in *A. thaliana*, genes involved in signal transduction and transcriptional regulation are over represented, whereas DNA-repair genes have been preferentially lost [13••,15]. The majority of duplicates retained (57%) tend to have divergent expression patterns as predicted by the DDC model, whereas over 20% have asymmetric rates of protein evolution; both are suggestive of functional divergence [13••].

Functional divergence can occur rapidly after polyploidization, perhaps even within a generation, and this phenomenon has been studied in both naturally occurring recent polyploids and in synthetic polyploids. For example, among 40 homeologous gene pairs in natural tetraploid cotton (which was the result of a whole-genome duplication event that occurred 1–2 mya), nine gene pairs exhibited biased expression, with the homeolog from one parental genome contributing more to the transcriptome than the other [66••]. Interestingly, 11 of these 18 homeologs showed tissue-specific silencing or biased expression in at least one of 10 organs tested, indicating that they were at an early phase of subfunctionalization. Similar patterns of biased expression were observed in synthetic *Gossypium* polyploids, demonstrating that functional divergence could occur within one generation of polyploidization [66••]. Similar conclusions were reached in a study of synthetic *Arabidopsis* allotetraploids produced in a cross between *A. thaliana* and *Arabidopsis arenosa* [67,68•].

Conclusions

A new conceptual framework for gene duplication combined with emerging genomic data and technologies has advanced our understanding of the contributions of gene and genome duplication to the evolution of plants. Care must be taken, however, in interpreting the influx of genomic data in a proper functional and phylogenetic context. For instance, although gene expression can diverge rapidly between duplicate genes, expression data alone cannot reveal whether the combined expression domains of the duplicates reflect that of the progenitor gene or whether one or both of the duplicates have gained a novel expression regime. Indeed, in a comparative study of lineage-specific duplicates in humans and mice, changes in expression were more often associated with the gain of novel expression domains, consistent with neofunctionalization [69•]. Similarly, inferring patterns of plant genome evolution requires a 'phlyogenomics' approach; genomic data on the structural arrangements of extant genes must be compared to that from related taxa to determine the extent and timing of genome duplication events [9•]. Only when such comparative approaches are used will we be able to truly understand the contribution of single-gene and whole-genome duplications to the emergence and diversification of plant species.

Acknowledgements

The authors thank members of the Purugganan laboratory for a critical reading of the manuscript. This work was supported in part by grants from the Integrated Research Challenges in Environmental Biology (IRCEB), Frontiers in Integrative Biological Research (FIBR) and Plant Genome Research Programs of the US National Science Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH: **Role of duplicate genes in genetic robustness against null mutations**. *Nature* 2003, **421**:63-66.
By analyzing yeast duplicate gene knockout mutants, the authors found that gene duplication makes a significant contribution to the buffering of deleterious mutations.

2. Ohno S: *Evolution by Gene Duplication*. Springer-Verlag; 1970.

3. Long M, Langley CH: **Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila***. *Science* 1993, **260**:91-95.

4. Gilbert W, de Souza SJ, Long M: **Origin of genes**. *Proc Natl Acad Sci USA* 1997, **94**:7698-7703.

5. Otto SP, Whitton J: **Polyploid incidence and evolution**. *Annu Rev Genet* 2000, **34**:401-437.

6. Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y: **The hidden duplication past of *Arabidopsis thaliana***. *Proc Natl Acad Sci USA* 2002, **99**:13627-13632.

7. Blanc G, Hokamp K, Wolfe KH: **A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome**. *Genome Res* 2003, **13**:137-144.

8. •• Bowers JE, Chapman BA, Rong J, Paterson AH: **Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**:433-438.
The authors combine phylogenetic and genomic approaches to determine the number and timing of genome duplication events in *A. thaliana*'s past.

9. • Paterson AH, Bowers JE, Chapman BA: **Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics**. *Proc Natl Acad Sci USA* 2004, **101**:9903-9908.
The authors undertook a comparative 'phylogenomics' approach in an attempt to ascertain ancient duplication events in cereals. Unlike previous studies, they had access to the completed rice genome, allowing for more sensitive detection of duplicated genomic segments.

10. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

11. Lynch M, Conery JC: **Gene duplication and evolution — a response**. *Science* 2001, **293**:1551.

12. Ilic K, SanMiguel PJ, Bennetzen JL: **A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes**. *Proc Natl Acad Sci USA* 2003, **100**:12265-12270.

13. •• Blanc G, Wolfe KH: **Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution**. *Plant Cell* 2004, **16**:1679-1691.
This is an exemplary functional genomic analysis of the fates of duplicated genes arising from whole-genome duplications. The authors surveyed the types of genes that have survived since the last whole-genome duplication in *Arabidopsis*, and analyzed the divergence in their expression and functions.

14. Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF *et al.*: **Sequence composition and genome organization of maize**. *Proc Natl Acad Sci USA* 2004, **101**:14349-14354.

15. Seoighe C, Gehring C: **Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome**. *Trends Genet* 2004, **20**:461-464.

16. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations**. *Genetics* 1999, **151**:1531-1545.

17. •• Moore RC, Purugganan MD: **The early stages of duplicate gene evolution**. *Proc Natl Acad Sci USA* 2003, **100**:15682-15687.
The authors took a population genetics approach to try to ascertain the evolutionary forces acting during the fixation of duplicate genes. Although much attention has been focused on the theoretical aspects of this crucial time in duplicate gene evolution, this is one of the first empirical studies of this process. The authors found that selection, rather than drift, plays a role in the establishment of duplicate gene loci.

18. •• Blanc G, Wolfe KH: **Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes**. *Plant Cell* 2004, **16**:1667-1678.
Using divergence estimates of duplicated EST unigenes from a total of 14 model plant species, the authors were able to map where and when ancient polyploidy events have occurred across a wide range of angiosperm taxa. By sampling taxonomically diverse sets of genomes, the authors were able to map numerous ancient polyploidy events across a breadth of angiosperm taxa.

19. Zhang J: **The infancy of duplicate genes**. *Heredity* 2004, **92**:479-480.

20. Lynch M, O'Hely M, Walsh B, Force A: **The probability of preservation of a newly arisen gene duplicate**. *Genetics* 2001, **159**:1789-1804.

21. Zhang J: **Evolution by gene duplication: an update**. *Trends Ecol Evol* 2003, **18**:292-298.

22. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization**. *Genetics* 2000, **154**:459-473.

23. Hughes AL: **The evolution of functionally novel proteins after gene duplication**. *Proc R Soc Lond B Biol Sci* 1994, **256**:119-124.

24. Gu Z, Nicolae D, Lu HH, Li WH: **Rapid divergence in expression between duplicate genes inferred from microarray data**. *Trends Genet* 2002, **18**:609-613.

25. Makova KD, Li WH: **Divergence in the spatial pattern of gene expression between human duplicate genes**. *Genome Res* 2003, **13**:1638-1645.

26. Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ: **Diversification of C-function activity in maize flower development**. *Science* 1996, **274**:1537-1540.

27. Matsunaga S, Isono E, Kejnovsky E, Vyskot B, Dolezel J, Kawano S, Charlesworth D: **Duplicative transfer of a MADS box gene to a plant Y chromosome**. *Mol Biol Evol* 2003, **20**:1062-1069.

28. Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martinez-Castilla L, Yanofsky MF: **An ancestral MADS-box gene duplication occurred before the divergence of plants and animals**. *Proc Natl Acad Sci USA* 2000, **97**:5328-5333.

29. Nam J, dePamphilis CW, Ma H, Nei M: **Antiquity and evolution of the MADS-box gene family controlling flower development in plants**. *Mol Biol Evol* 2003, **20**:1435-1447.

30. • Nam J, Kim J, Lee S, An G, Ma H, Nei M: **Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms**. *Proc Natl Acad Sci USA* 2004, **101**:1910-1915.
By analyzing the evolution of both functional and non-functional MADS genes, the authors were able to present the detailed life history of one of the most intensely studied gene family in plants.

31. • Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M: **Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses**. *Mol Biol Evol* 2003, **20**:1963-1977.
The authors performed an expression analysis of 105 MADS-box genes. Interestingly, eight of these genes exhibit gametophytic expression, and might represent ancestral functions of MADS genes retained from the gametophyte-dominant ancestor of land plants.

32. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B *et al.*: **Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world**. *Plant Cell* 2003, **15**:1538-1551.

33. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF: **Assessing the redundancy of MADS-box genes during carpel and ovule development**. *Nature* 2003, **424**:85-88.

34. • Kramer EM, Jaramillo MA, Di Stilio VS: **Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms**. *Genetics* 2004, **166**:1011-1023.
The authors presented a comprehensive phylogenetic survey of the *AGAMOUS* subfamily of MADS-box genes. They demonstrated that an ancient gene duplication followed by functional divergence via subfunctionalization, along with overlapping redundancies, has shaped the evolution of this gene family.

35. Theissen G: **Development of floral organ identity: stories from the MADS house**. *Curr Opin Plant Biol* 2001, **4**:75-85.

36. Tandre K, Albert VA, Sundas A, Engstrom P: **Conifer homologues to genes that control floral development in angiosperms**. *Plant Mol Biol* 1995, **27**:69-78.

37. Rutledge R, Regan S, Nicolas O, Fobert P, Cote C, Bosnich W, Kauffeldt C, Sunohara G, Seguin A, Stewart D: **Characterization of an *AGAMOUS* homologue from the conifer black spruce (*Picea mariana*) that produces floral homeotic conversions when expressed in *Arabidopsis***. *Plant J* 1998, **15**:625-634.

38. Jager M, Hassanin A, Manuel M, Le Guyader H, Deutsch J: **MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family**. *Mol Biol Evol* 2003, **20**:842-854.

39. Zhang P, Tan HT, Pwee KH, Kumar PP: **Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms**. *Plant J* 2004, **37**:566-577.

40. Bowman JL, Smyth DR, Meyerowitz EM: **Genes directing flower development in *Arabidopsis***. *Plant Cell* 1989, **1**:37-52.

41. Carpenter R, Coen ES: **Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus***. *Genes Dev* 1990, **4**:1483-1493.

42. Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z: ***PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development**. *EMBO J* 1999, **18**:4023-4034.

43. Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF: ***SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis***. *Nature* 2000, **404**:766-770.

44. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF: **B and C floral organ identity functions require *SEPALLATA* MADS-box genes**. *Nature* 2000, **405**:200-203.

45. Wagner A: **Redundant gene functions and natural selection**. *J Evol Biol* 1999, **12**:1-16.

46. Kempin SA, Savidge B, Yanofsky MF: **Molecular basis of the cauliflower phenotype in *Arabidopsis***. *Science* 1995, **267**:522-525.

47. Siegfried KR, Eshed Y, Baum SF, Otsuga D, Drews GN, Bowman JL: **Members of the YABBY gene family specify abaxial cell fate in *Arabidopsis***. *Development* 1999, **126**:4117-4128.

48. Ferrandiz C, Gu Q, Martienssen R, Yanofsky MF: **Redundant regulation of meristem identity and plant architecture by *FRUITFULL, APETALA1* and *CAULIFLOWER***. *Development* 2000, **127**:725-734.

49. Eshed Y, Baum SF, Perea JV, Bowman JL: **Establishment of polarity in lateral organs of plants**. *Curr Biol* 2001, **11**:1251-1260.

50. McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK: **Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots**. *Nature* 2001, **411**:709-713.

51. Takada S, Hibara K, Ishida T, Tasaka M: **The *CUP-SHAPED COTYLEDON1* gene of *Arabidopsis* regulates shoot apical meristem formation**. *Development* 2001, **128**:1127-1135.

52. Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M, Chory J: **Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth**. *Genetics* 2002, **162**:1445-1456.

53. Pelaz S, Tapia-Lopez R, Alvarez-Buylla ER, Yanofsky MF: **Conversion of leaves into petals in *Arabidopsis***. *Curr Biol* 2001, **11**:182-184.

54. Moore RC, Grant SR, Purugganan MD: **Molecular population genetics of redundant floral regulatory genes in *Arabidopsis thaliana***. *Mol Biol Evol* 2005, **22**:91-103.

55. Baumgarten A, Cannon S, Spangler R, May G: **Genome-level evolution of resistance genes in *Arabidopsis thaliana***. *Genetics* 2003, **165**:309-319.

56. Mondragon-Palomino M, Meyers BC, Michelmore RW, Gaut BS: **Patterns of positive selection in the complete NBS–LRR gene family of *Arabidopsis thaliana***. *Genome Res* 2002, **12**:1305-1315.

57. Caicedo AL, Schaal BA: **Heterogenous evolutionary processes affect *R* gene diversity in natural populations of *Solanum pimpinellifolium***. *Proc Natl Acad Sci USA* 2004, **101**:17444-17449.

58. Kroymann J, Textor S, Tokuhisa JG, Falk KL, Bartram S, Gershenzon J, Mitchell-Olds T: **A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway**. *Plant Physiol* 2001, **127**:1077-1088.

59. • Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T: **Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus**. *Proc Natl Acad Sci USA* 2003, **100(Suppl 2)**:14587-14592.

The authors use population genetics analyses to reveal complex evolutionary dynamics of two duplicate loci that are involved in the production of glucosinolates. The authors provide evidence that one of the loci, *MAM2*, is subject to balancing selection.

60. Clauss MJ, Mitchell-Olds T: **Functional divergence in tandemly duplicated *Arabidopsis thaliana* trypsin inhibitor genes**. *Genetics* 2004, **166**:1419-1436.

61. Clauss MJ, Mitchell-Olds T: **Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history**. *Mol Ecol* 2003, **12**:1287-1299.

62. Vision TJ, Brown DG, Tanksley SD: **The origins of genomic duplications in *Arabidopsis***. *Science* 2000, **290**:2114-2117.

63. Raes J, Vandepoele K, Simillion C, Saeys Y, Van de Peer Y: **Investigating ancient duplication events in the *Arabidopsis* genome**. *J Struct Funct Genomics* 2003, **3**:117-129.

64. Wagner A: **The fate of duplicated genes: loss or new function?** *Bioessays* 1998, **20**:785-788.

65. Walsh B: **Population-genetic models of the fates of duplicate genes**. *Genetica* 2003, **118**:279-294.

66. •• Adams KL, Cronn R, Percifield R, Wendel JF: **Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific process reciprocal silencing**. *Proc Natl Acad Sci USA* 2003, **100**:4649-4654.

The authors analyzed the expression profiles of 40 gene pairs in both natural and synthetic polyploids of cotton. They found evidence for the reciprocal and differential gene expression of a portion of these gene pairs, consistent with the early phase of subfunctionalization.

67. Lee HS, Chen ZJ: **Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids**. *Proc Natl Acad Sci USA* 2001, **98**:6753-6758.

68. • Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ: **Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids**. *Genetics* 2004, **167**:1961-1973.

The authors followed changes in duplicate gene expression between siblings of multiple selfed generations of synthetic *Arabidopsis* allopolyploids. They found that duplicate gene silencing is rapid and can vary stochastically between siblings. They also performed functional analyses to show that this silencing is potentially controlled by DNA methylation.

69. • Huminiecki L, Wolfe KH: **Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse**. *Genome Res* 2004, **14**:1870-1879.

The authors compared the expression profiles of lineage-specific duplications to those of their single-copy orthologs, an approach previously lacking in similar studies. They found that the overwhelming majority of duplicates in their survey gained novel expression domains, consistent with neofunctionalization. This result suggests that neofunctionalization might be more prevalent than previously considered.