Consequences of genome duplication

Marie Sémon and Kenneth H Wolfe

Polyploidy has been widely appreciated as an important force in the evolution of plant genomes, but now it is recognized as a common phenomenon throughout eukaryotic evolution. Insight into this process has been gained by analyzing the plant, animal, fungal, and recently protozoan genomes that show evidence of whole genome duplication (a transient doubling of the entire gene repertoire of an organism). Moreover, comparative analyses are revealing the evolutionary processes that occur as multiple related genomes diverge from a shared polyploid ancestor, and in individual genomes that underwent several successive rounds of duplication. Recent research including laboratory studies on synthetic polyploids indicates that genome content and gene expression can change quickly after whole genome duplication and that cross-genome regulatory interactions are important. We have a growing understanding of the relationship between whole genome duplication and speciation. Further, recent studies are providing insights into why some gene pairs survive in duplicate, whereas others do not.

Addresses
Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland

Corresponding authors: Sémon, Marie (msemon@ens-lyon.fr) and Wolfe, Kenneth H (khwolfe@tcd.ie)

Current Opinion in Genetics & Development 2007, **17**:505–512

This review comes from a themed issue on
Genomes and evolution
Edited by Susan Wessler, Kelly Dawe and Jim Leebens-Mack

Available online 19th November 2007

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2007.09.007

Introduction

A change in ploidy is typically expected to be deleterious and an evolutionary dead-end [1]. Despite the problems that might arise in early polyploid generations, the hallmarks of whole genome duplication (WGD) are evident in many sequenced genomes. The prevalence of polyploidy varies across eukaryotic lineages, but evidence of WGD is particularly rampant in plants. It has been demonstrated recently that most eudicot plants descended from an ancient hexaploid ancestor [2••], followed by lineage-specific tetraploidizations in some taxa: one in *Populus* [3•], two in *Arabidopsis* [4–6], one in legumes [7], but none in *Vitis* [2••]. Consequently, a gene that was single-copy in an ancestral angiosperm about 200 million years (Myr) ago could, in principle, have turned into a 12-member family in *Arabidopsis* by means of polyploidizations alone. In practice, of course, each round of polyploidization was followed by many gene deletions, and gene duplications have also happened by mechanisms other than polyploidization.

Detecting natural polyploidization events can be challenging, especially if the events are ancient. Recent duplications can be detected by the identification of species whose karyotypes contain twice as many chromosomes as those of closely related species. Time erases this signal, however: WGD is typically followed by a period of diploidization, at the end of which the genome looks like a diploid. This period involves extensive gene loss, genomic rearrangements, and distinctive modes of evolution of the genes retained in duplicate. The diploidization process has been extensively studied using whole genome data in different paleopolyploid plants [2••,3•,8•,9–11], teleost fishes [13•,14], yeasts [15••], *Paramecium* [12••], and basal vertebrates [16]. Here we review some of these studies, from a wide range of eukaryotic taxa, with emphasis on the consequences of WGD for speciation and the diversification of gene families.

Genomic changes after WGD

Genomic modifications that occur in the first few generations after WGD can be monitored in synthetic polyploid plants (reviewed in reference [17]). For example, *Brassica napus* genomes in the first polyploid generation [18] display few rearrangements but numerous and recurrent CpG methylation changes. To study the longer term evolutionary effects of WGD, however, comparative genomic analyses are required.

Interchromosomal rearrangements are a frequent feature of post-WGD evolution. In a recent elegant study, Kasahara *et al.* [19] reconstructed the chromosomal content of the teleost fish ancestor from the genome sequences of a pufferfish, zebrafish, and medaka, allowing the subsequent history of rearrangement in each lineage to be inferred. They found that eight major interchromosomal rearrangements occurred soon after the fish-specific WGD, but then no further major rearrangements occurred in the medaka lineage. This result suggests that the process of WGD caused a transient increase in the rate of chromosomal rearrangement in the common ancestor of teleosts [19]. Similarly, the frequency of disruption of linkage between adjacent genes increased after the fish WGD [20]. Future comparative genome analyses in plants could address a more challenging problem, the evolution of rearrangements after multiple successive WGDs. A good start in

this direction was recently made by Jaillon *et al.* [2••] who were able to infer the approximate chromosome set of the common ancestor of eudicots, thanks to the remarkably low level of subsequent rearrangement that happened in the *Vitis* genome. *Vitis*, medaka, and *Paramecium* (see below) all illustrate how the fortuitous discovery of genomes that have undergone little rearrangement can assist greatly in reconstructing ancient karyotypes and detecting very old polyploidization events.

We are also beginning to gain an understanding of the rate at which genes are lost from polyploid genomes. The process of gene loss is illustrated neatly by the ciliate *Paramecium tetraurelia* whose genome underwent at least three successive rounds of WGD, obscured by very few other interchromosomal rearrangements [12••]. In *Paramecium*, the frequency of duplicate gene survival is highest for the most recent WGD (51%), intermediate for the second one (24%), and much smaller for the most ancient WGD (8%). This result led Aury *et al.* [12••] to suggest that the rate of gene loss in *Paramecium* had been slow and relatively constant. By contrast, comparative genomics studies on yeast species have suggested that the rate of paralogous gene loss is very rapid shortly after WGD, subsequently slowing down [21••].

Duplicate gene fate(s) and speciation

Most gene pairs formed by a WGD have only a brief lifespan before one copy becomes deleted, leaving the other to survive as a single-copy locus. We might expect that the probability of retention is initially equal for both duplicates following WGD, but recent results have suggested that the one duplicate may be more susceptible to loss than the other. It was shown that in *Arabidopsis thaliana*, one paralogon (duplicated genomic region) tends to contain significantly more genes than the other [22•]. This bias was interpreted as the consequence of an initial inequality between the two paralogons, possibly due to epigenetic marks. If this hypothesis is true, epigenetic marks before WGD, or epigenetic changes occurring immediately after WGD, can influence the long-term fates of duplicated genes. In yeast species, the choice of survivor appears to have been arbitrary for two species that diverged soon after the WGD, but non-random for gene losses that happened at later time points [15••].

These observations are consistent with a model where, soon after an autopolyploidization (WGD due to doubling of a single genome), the two copies of any particular gene are equally likely to be lost because few sequence differences exist between the duplicates. After more time elapses, however, differences between copies emerge as their sequences diverge. The selective consequences of deleting one copy may become different from those of deleting the other copy, with the result that we observe non-random survivor choice: in situations where the same gene pair has sustained a gene deletion in two independent lineages, we tend to see the same copy surviving in both lineages [15••,21••]. Such a distinction between early and late gene losses is expected after autopolyploidization, but not necessarily after allopolyploidization (WGD caused by merging two divergent parental genomes), where the copies are already different by the time the new polyploid is formed.

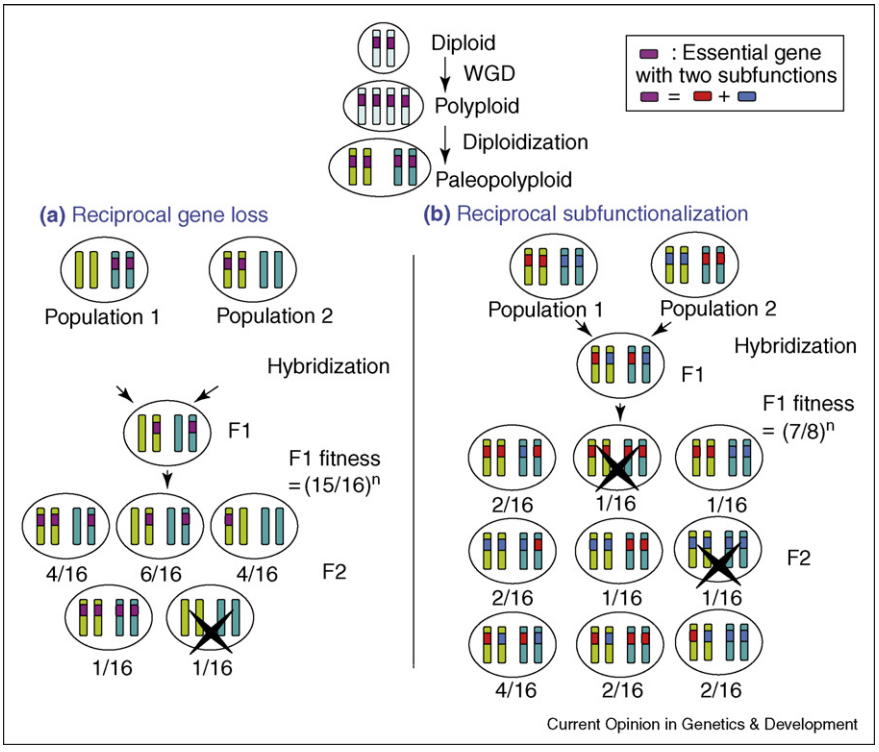
An increase in species diversity (in teleost fishes [23], see also reference [24]) or even a radiation (in *Paramecium* [12••]) has often been noticed in clades that have a WGD at their base. WGD has been proposed to be a lineage-splitting force because of the subsequent occurrence of gene losses independently in different populations [23,25,26]. In particular, reciprocal gene loss (RGL) occurs when two paralogs created by WGD are retained until speciation, after which each species loses a different copy (Figure 1a). Hybrid fitness diminishes exponentially with the number of RGL loci, and the neutral loss of alternate copies of duplicated genes therefore contributes to speciation. We expect that the rate at which RGL occurs, and hence the rate of lineage splitting, will be highest immediately after WGD for two reasons: first, the overall rate of gene loss is maximal at this point [21••] and second, gene copy choice is more likely to be arbitrary for early gene losses, as discussed above. However, RGL can continue to occur for tens to hundreds of millions of years after a WGD [21••,27] and can therefore promote speciation over a long timeframe. The potential influence of ancient polyploidy events on the origin and diversification of angiosperms – thanks to the duplicate retention of genes involved in development, transcriptional regulation and signaling – has already been discussed [28], but RGL has not yet been documented on a whole-genome scale in plants. In view of the numerous WGDs that have occurred during angiosperm evolution, the inevitable RGLs that can be inferred to have happened afterwards must be considered as prime suspects for agents of plant speciation. Hard evidence implicating RGLs as a cause of speciation is, however, still lacking in both plants and fishes, in contrast to *Drosophila* [29].

Rapid functional divergence as an explanation for duplicate gene retention

After duplication, one of the two redundant copies of a gene should theoretically be free to degenerate and become lost from the genome without consequence. As we have seen, contrary to this prediction some genes survive in duplicate long after WGD. Several models, some implying a functional divergence between the two copies, have been proposed to account for these observations. We summarize these models in Figure 2 and discuss them briefly below.

In plants, it is possible to quantify the immediate impact of WGD by studying gene expression changes in recent natural or artificial polyploids (reviewed in reference

Figure 1



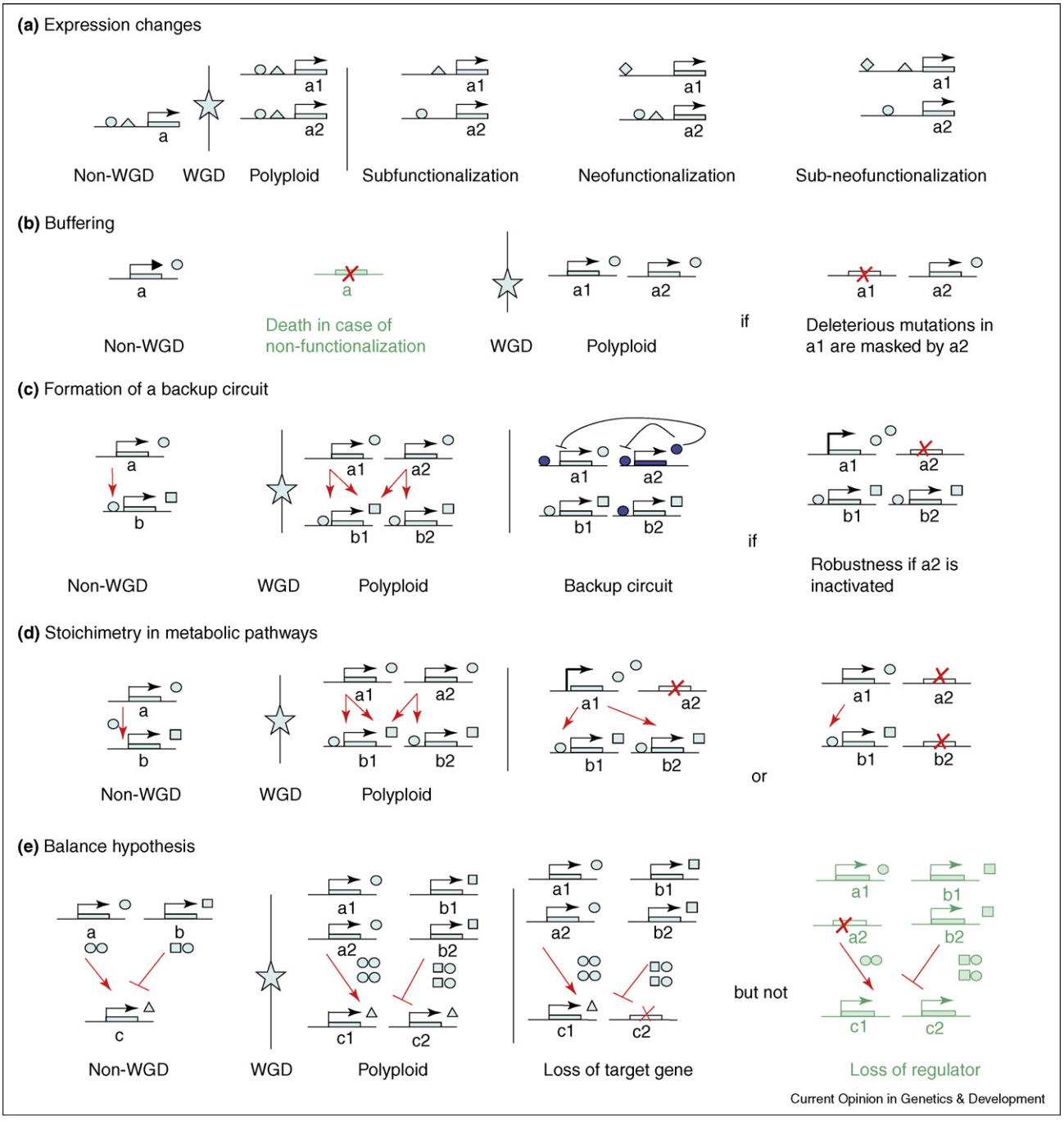
WGD contributes to speciation by the neutral loss of genes or by neutral changes in expression pattern. **(a)** After WGD, different populations may retain a different copy of the same essential gene. 1/16 of the F_2 individuals obtained after crossing these two populations do not carry any copy of the essential gene. F_1 fitness decreases exponentially with the number of RGLs (*n*) and therefore multiple RGLs favor speciation. An even greater effect on fitness is expected if the lost gene is essential for gamete survival (1/4 of the F_1 gametes are inviable) or in the case of reciprocal subfunctionalization as presented in **(b)**. Subfunctionalization is the complementary loss of different subfunctions of an ancestral gene, in each of the duplicate copies. The subfunctions are represented by the colors blue and red, and the full ancestral function is purple. A phenomenon similar to RGL is expected after reciprocal subfunctionalization, which occurs when an essential gene undergoes subfunctionalization in each of the populations, but the same copy retains a different subfunction in each population.

[30]). One analysis performed just after artificial allopolyploidization in cotton found that one paralog is silenced or downregulated in 5% of the gene pairs and that silencing is often organ-specific [31]. Another method of analysis involves testing whether the total expression of a gene pair in a polyploid deviates from the average expression level of the corresponding genes in the two parents. In *Arabidopsis*, deviations like this occur at many loci after allopolyploidization but are much rarer after autopolyploidization [32••]. They are most frequently observed in genes whose expression differs between the parental lines [32••]. The underlying mechanism for these transcriptional changes may involve regulatory proteins encoded by one genome acting *in trans* on the other genome, as illustrated by the way that a cross-genome interaction between the *FRI* and *FLC* loci affects flowering time in a natural *Arabidopsis* allotetraploid ([33••]; Figure 3). Polyploidy also doubles the quantity of DNA and therefore increases volume of the cell: even if doubling the number of genes simply doubled the amount of proteins, the consequences on protein concentration would not be straightforward [34].

These rapid changes in expression between gene duplicates in polyploid plants are compatible with one group of models to explain the retention of two copies after WGD (presented in Figure 2a). This group includes the familiar neofunctionalization and subfunctionalization models of Lynch and Force [35,36], as well as the recent subneofunctionalization model (that is, neofunctionalization of a gene pair after its initial preservation in duplicate by subfunctionalization) of He and Zhang [37]. Several clear-cut examples of subfunctionalization have been described in plants, including the case of the duplicated genes *PLENA* and *FARINELLI* in *Antirrhinum* (*Arabidopsis* orthologs *SHATTERPROOF1&2* and *AGAMOUS*, respectively) [38]. This case is even an example of differential subfunctionalization, because one subfunction (the C-function) has been retained in paralogous copies in *Arabidopsis* (*AGAMOUS*) and in *Antirrhinum* (*PLENA*). Differential subfunctionalization is expected to promote speciation, just like RGL (Figure 1b).

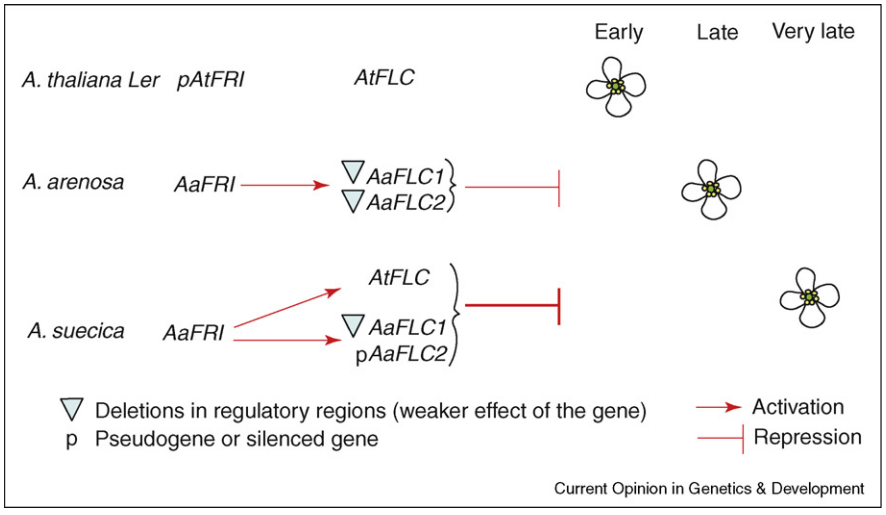
Subfunctionalization and neofunctionalization models make different predictions about the rates of sequence

Figure 2



Simplified examples of several models to explain the retention of both copies of a gene after WGD. The situation before duplication is shown on the left of each panel (before the star), the polyploid state just after WGD is shown in the middle, and possible outcomes after diploidization are shown on the right. Cases in green represent non-viable outcomes. **(a)** Rapid functional evolution is responsible for duplicate retention, either because of acquisition of new function in one duplicate (neofunctionalization), or because of the complementary loss of the subfunctions of the ancestral gene in each of the duplicate copies (subfunctionalization), or a mix of both (subneofunctionalization). Diamond, circle and triangle represent regulatory regions driving expression in distinct tissues, considered as subfunctions. **(b)** Essential genes are maintained in duplicate while their sequences are sufficiently similar (young duplicates). Without duplication, a deleterious mutation (red cross) is lethal (in green on the left). After duplication, a deleterious mutation in one of the duplicates is buffered by the presence of the other paralog. **(c)** Example of a backup circuit involving negative feedback between duplicates that explains the redundancy between a1 and a2. a2 represses a1 in normal conditions, and both a2 and a1 regulate the same genes b1 and b2. When a2 is deleted a1 is more expressed, so b1 and b2 are still regulated and the deletion is not lethal. **(d)** The numbers of copies of all the genes in a complex or a metabolic pathway are correlated, at least for recent duplications; subsequently either all the genes in a pathway return to a single-copy state, or regulatory modifications can occur that

Figure 3



Flowering time in the natural allopolyploid *Arabidopsis suecica* is more extreme than in either of the two parents, *A. thaliana* and *A. arenosa* (from reference [33••]). *A. suecica* is a natural allopolyploid that was formed 0.02–1.5 Myr ago between *A. thaliana* and *A. arenosa*. The two parental lines diverged 6 Myr ago. *FRI* and *FLC* are two of the genes responsible for the flowering time in *Arabidopsis*. In *A. arenosa*, *FRI* activates *FLC* (red arrow), which delays flowering (red bar). In *A. thaliana* (*Ler* ecotype), *FRI* is not functional, so *FLC* is not activated and flowering is early. In *A. suecica*, the combination of the two parental genomes is responsible for *FLC* upregulation and very late flowering.

evolution of the duplicates (summarized in reference [39]). In the tetraploid frog *Xenopus laevis*, for example, 6% of duplicated gene pairs have asymmetrical rates of protein sequence evolution, which may indicate neofunctionalization in the accelerated copy [39]. Similarly, evolutionary rate asymmetry has been reported in 11–16% of duplicated pairs in *Paramecium* [12••], 25–36% in fish [13•,14], and >20% in *Arabidopsis* [10]. Subfunctionalization has been detected by complementary amino acid substitutions in protein pairs [39] and by differential loss of regulatory regions [40•]. Divergence of gene expression between duplicates has been reported in many studies [3•,9,10,41], but because the pattern of expression before duplication is unknown, these comparisons cannot differentiate between subfunctionalization and neofunctionalization. The problem was solved recently in yeast species by comparing the patterns of expression in an outgroup (*Candida albicans*, whose genome is not duplicated and therefore can be used to approximate the ancestral expression state) to those of *Saccharomyces cerevisiae* gene pairs that were formed by WGD [42]. In any case, the models of subfunctionalization and neofunctionalization are not incompatible (Figure 2a; [37]).

Results from recent polyploid plants and ancient WGDs suggest that subfunctionalization and neofunctionalization were responsible for the retention of many loci in two copies after WGD. Rapid functional divergence may not be the only explanation, however, and we discuss some alternatives below. Moreover, the relative importance of different mechanisms may differ among taxa, depending on factors such as population size or the mode of reproduction.

Other models for duplicate gene retention

Buffering of the genome against the consequences of mutations that would otherwise be deleterious has been suggested as a mechanism for duplicate gene retention. The idea is that if one gene is knocked out, the other provides a backup (Figure 2b). Chapman *et al.* [8•] found that genes with a recent WGD paralog in rice and *Arabidopsis* tend to show lower levels of within-species non-synonymous nucleotide polymorphism than do singleton genes. Similar observations were made in poplar [3•]. This result is inconsistent with the prediction of rapid sequence evolution during neofunctionalization and, together with the observation that retained duplicate genes tend to be large and to encode multidomain proteins, led Chapman *et al.* [8•] to propose instead that selection for a buffering effect was a mechanism for duplicate gene retention after plant WGDs. There are, however, theoretical arguments why buffering alone should only rarely lead to the preservation of a pair of genes that are completely redundant in function (see

(Figure 2 Legend Continued) maintain the dosage when imbalanced numbers of the genes are lost. **(e)** One possible circuit for the balance hypothesis. Retention of regulatory genes in two copies is necessary, because deletion of one copy is lethal (in green on the extreme right), which is not true for target genes.

discussion of Model 1 in reference [43]). In our view the low non-synonymous polymorphism seen in retained duplicates is another manifestation of the phenomenon first noted in yeast and in nematode by Davis and Petrov [44] (see also, in fish [13•]) that genes with inherently slow rates of protein sequence evolution (and hence expected to have low levels of non-synonymous polymorphism) tend to survive more often after duplication. A variety of hypotheses have been posed for the mechanisms responsible for the association between duplicate gene retention and low rates of non-synonymous substitution. The association may be due to the fact that a slow rate of evolution correlates with other factors predisposing genes to be preserved in duplicate, like the number of regulatory regions, the number of protein–protein interactions, or a high level of expression [44,45].

In addition to the buffering hypothesis, there are several other models explaining why gene pairs having redundant functions can be retained. One reason why some gene pairs have been retained despite having apparently redundant functions is that their redundancy could be an illusion caused by our incomplete knowledge of their full repertoire of functions. In yeast, numerous ancient paralogous gene pairs can back each other up: one copy compensates for the loss of the other if knocked out. However, this sort of backup may only occur in particular growth conditions [46], or alternatively it may be the result of transcriptional changes caused by the knockout itself [47]. Many WGD paralogs in yeast have different patterns of expression under normal conditions and most of their regulatory motifs differ, but in circumstances where one copy is knocked out, 'responsive backup circuits' [47] can permit the transcriptional reprogramming of one paralog to compensate for the loss of the other (Figure 2c). Some examples of these circuits, involving feedback loops and *trans*-regulation between the two copies, are listed in reference [48]. Similarly, half of the WGD gene pairs in yeast tend to belong to the same protein complexes, and these functionally overlapping pairs show more divergence in expression than do duplicates that do not belong to the same complex [49].

Another model is related to the fact that the relative stoichiometry of proteins belonging to the same complex or to the same metabolic pathway must be maintained after the duplication (Figure 2d). The model proposes that interacting duplicates are retained for stoichiometry just after WGD, but that gene regulation later evolves and eventually allows one copy to be lost without affecting protein levels. This effect is illustrated neatly in *Paramecium* [12••]. After the most recent WGD, there is a strong pattern of retention of both copies of genes belonging to known protein complexes (or metabolic pathways), but this is not true of the older duplications.

A final model attempts to explain the finding that transcription factors and genes involved in signal transduction are over-retained in duplicate after WGD, but not after smaller scale DNA duplication events [11,16]. To explain this, the balance hypothesis proposes that regulatory genes are retained in duplicate because a change in their relative number modifies the expression of the target genes ([50], Figure 2e). Freeling and Thomas [51] extrapolated these observations to make a chain of arguments that we summarize as: (i) WGD tends to cause complete sets of regulatory genes and of genes whose products participate in protein–protein interactions to be retained in duplicate, because of the need for balance; (ii) the retention of these particular types of genes after each successive WGD can be viewed as a 'drive', similar to meiotic drive in the sense that it is inevitable and not adaptive; (iii) these retained genes may have been used for new purposes; and hence (iv) the increases in morphological complexity that occurred during the evolution of both plants and animals may therefore have been passive consequences of this 'drive', caused by WGD. Freeling and Thomas are not arguing that morphological evolution is non-adaptive, but that the limits to the amount of complexity that an organism can attain were pushed upwards by this process. Unfortunately, this model will be very difficult if not impossible to test.

Conclusion

The plant kingdom is the uncontested big kahuna of polyploidization, but simpler non-plant systems still offer many lessons that can help us understand the waves of successive WGDs that have washed over angiosperm evolution. Recurrent trends can be observed in very different taxa, such as the tendency to retain regulatory genes in duplicate in many paleopolyploid genomes. It would be a mistake, however, to think that the outcomes of all WGDs are the same. (i) Different models of duplicate gene retention are favored depending on the biology of the species in question; for example, subfunctionalization should be rare in species with large population sizes [36]. (ii) Different processes are expected to be at work depending on the age of the WGD as well. For example genetically based functional divergence between gene copies is expected to increase with time since WGD, whereas epigenetic effects such as transcriptional silencing are immediate. (iii) Differences are expected between autopolyploidization and allopolyploidization. Most discussions of the process of duplicate gene retention have assumed that the two copies are identical at birth, but this is untrue in the case of allopolyploidization.

Acknowledgements

This work was supported by Irish Research Council for Science, Engineering and Technology, and Science Foundation Ireland. We apologize to authors whose work we could not discuss because of space limits. We thank Kevin Byrne, Gavin Conant, Brian Cusack, Carolin Frank, Jonathan Gordon, Nora Khaldi, Jeffrey Mower, Devin Scannell, Matthew Webster, and Meg Woolfit for helpful discussions.

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

2. •• Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.*: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**. *Nature* 2007, **449**:463-467.

The genome sequence of grapevine (*Vitis vinifera*) reveals that this species – unlike *Populus*, *Arabidopsis* or rice – has not undergone a recent WGD. This property, together with an apparently low rate of chromosome rearrangement in *Vitis*, provides Jaillon *et al*. with a very clear view of early events during eudicot genome evolution. They find beautiful evidence of a triplicated structure, indicating that the common ancestor of the sequenced eudicot genomes was hexaploid. An important question arising from this paper is the phylogenetic relationship among the three sequenced eudicot genomes. Jaillon *et al*. show a tree where *Arabidopsis* is an outgroup to *Vitis* and *Populus*. This accords with their finding that *Vitis* proteins have higher sequence identity to those of *Populus* than to those of *Arabidopsis*, but it contradicts two thorough molecular phylogenetic studies [52,53] that reported *Vitis* to be an outgroup to both the Eurosid I and Eurosid II clades (which include *Populus* and *Arabidopsis*, respectively). It has already been reported [3•] that the rate of sequence evolution in *Arabidopsis* is six times faster than in *Populus*, which could account for the differences in divergence if *Vitis* is the outgroup.

3. • Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**:1596-1604.

This paper reports the analysis of the first tree genome sequence, *Populus trichocarpa*. The *Populus* genome has undergone two successive WGDs, the older of which might be shared with *Arabidopsis*.

4. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**:433-438.

5. 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.

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. Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T *et al.*: **Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes**. *Proc Natl Acad Sci USA* 2006, **103**:14959-14964.

8. • Chapman BA, Bowers JE, Feltus FA, Paterson AH: **Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication**. *Proc Natl Acad Sci USA* 2006, **103**:2730-2735.

Proposal of a model of functional buffering to account for the reduced amino acid polymorphism of genes with retained paralogs compared to single-copy genes in *Arabidopsis* and in *Oryza*. See main text for discussion.

9. Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y: **Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana***. *Genome Biol* 2006, **7**:R13.

10. Blanc G, Wolfe KH: **Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution**. *Plant Cell* 2004, **16**:1679-1691.

11. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: **Modeling gene and genome duplications in eukaryotes**. *Proc Natl Acad Sci USA* 2005, **102**:5454-5459.

12. •• Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Segurens B, Daubin V, Anthouard V, Aiach N *et al.*: **Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia***. *Nature* 2006, **444**:171-178.

The *Paramecium* genome paper presents an elegant and detailed analysis of the consequences of WGD. Because the genome has undergone at least three successive WGDs, it provides insights into the short-term consequences of duplications, particularly relating to the influence of gene dosage constraints on duplicate gene retention.

13. • Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: **Gene loss and evolutionary rates following whole-genome duplication in teleost fishes**. *Mol Biol Evol* 2006, **23**:1808-1816.

Careful analysis of rates of sequence evolution and functional bias in duplicates created by the fish-specific WGD, showing notably that the genes retained in duplicate evolved slowly before duplication.

14. Steinke D, Salzburger W, Braasch I, Meyer A: **Many genes in fish have species-specific asymmetric rates of molecular evolution**. *BMC Genomics* 2006, **7**:20.

15. •• Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH: **Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication**. *Proc Natl Acad Sci USA* 2007, **104**:8397-8402.

See note from reference [21••].

16. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y: **The gain and loss of genes during 600 million years of vertebrate evolution**. *Genome Biol* 2006, **7**:R43.

17. Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V *et al.*: **Understanding mechanisms of novel gene expression in polyploids**. *Trends Genet* 2003, **19**:141-147.

18. Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T: **Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids**. *Plant Physiol* 2006, **140**:336-348.

19. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y *et al.*: **The medaka draft genome and insights into vertebrate genome evolution**. *Nature* 2007, **447**:714-719.

20. Semon M, Wolfe KH: **Rearrangement rate following the whole-genome duplication in teleosts**. *Mol Biol Evol* 2007, **24**:860-867.

21. •• Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH: **Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts**. *Nature* 2006, **440**:341-345.

See also reference [15••]: These papers describe the patterns of gene loss observed in several yeast species that diverged after a WGD in their common ancestor. In many cases, different copies of the same gene have been retained in different species, causing reciprocal gene losses that link WGD to speciation. Comparison of different species that diverged earlier or later after WGD reveals that the patterns of gene loss from duplicated loci changes with time.

22. • Thomas BC, Pedersen B, Freeling M: **Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homologue leaving clusters enriched in dose-sensitive genes**. *Genome Res* 2006, **16**:934-946.

This paper reports that the paralogons created by the last WGD in *Arabidopsis* are asymmetric (one of them retains more genes than the other), which might be explained by epigenetic marks that favor deletions from one particular paralogon. Regulatory genes are retained in duplicate, in accordance with the balance hypothesis, and tend to be clustered in the genome.

23. Taylor JS, Van de Peer Y, Meyer A: **Genome duplication, divergent resolution and speciation**. *Trends Genet* 2001, **17**:299-301.

24. Donoghue PJC, Purnell MA: **Genome duplication, extinction and vertebrate evolution**. *Trends Ecol Evol* 2005, **20**:313-319.

25. Lynch M, Force A: **The origin of interspecies genomic incompatibility via gene duplication**. *Am Nat* 2000, **156**:590-605.

26. Werth CR, Windham MD: **A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression**. *Am Nat* 1991, **137**:515-526.

27. Semon M, Wolfe KH: **Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor**. *Trends Genet* 2007, **23**:108-112.

28. De Bodt S, Maere S, Van de Peer Y: **Genome duplication and the origin of angiosperms**. *Trends Ecol Evol* 2005, **20**:591-597.

29. Masly JP, Jones CD, Noor MA, Locke J, Orr HA: **Gene transposition as a cause of hybrid sterility in *Drosophila***. *Science* 2006, **313**:1448-1450.

30. Adams KL: **Evolution of duplicate gene expression in polyploid and hybrid plants**. *J Hered* 2007, **98**:136-141.

31. Adams KL, Percifield R, Wendel JF: **Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid**. *Genetics* 2004, **168**:2217-2226.

32. •• Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L *et al.*: **Genomewide nonadditive gene regulation in *Arabidopsis allotetraploids***. *Genetics* 2006, **172**:507-517.

Microarray analysis of expression variation between *Arabidopsis* neopolyploids and their parents reveals considerable differences, more so after allopolyploidization than after autopolyploidization. These changes are mainly downregulations and are biased with regard to the parental origin of the copy.

33. •• Wang J, Tian L, Lee HS, Chen ZJ: **Nonadditive regulation of FRI and FLC loci mediates flowering-time variation in *Arabidopsis allopolyploids***. *Genetics* 2006, **173**:965-974.

The flowering time in an *Arabidopsis allopolyploid* is later than in both of its parents. This engaging paper shows that the phenotype of the allopolyploid is a consequence of *trans*-regulation between genes from the two parental genomes.

34. Veitia RA: **Gene dosage balance: deletions, duplications and dominance**. *Trends Genet* 2005, **21**:33-35.

35. 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.

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

37. He X, Zhang J: **Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution**. *Genetics* 2005, **169**:1157-1164.

38. Causier B, Castillo R, Zhou J, Ingram R, Xue Y, Schwarz-Sommer Z, Davies B: **Evolution in action: following function in duplicated floral homeotic genes**. *Curr Biol* 2005, **15**:1508-1512.

39. Chain FJ, Evans BJ: **Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis***. *PLoS Genet* 2006, **2**:e56.

40. • Woolfe A, Elgar G: **Comparative genomics using Fugu reveals insights into regulatory subfunctionalization**. *Genome Biol* 2007, **8**:R53.

This is the first study that detects subfunctionalization after WGD by looking at regulatory region partitioning between duplicates. The method uses conserved non-coding sequences between fish and mammals and could certainly be applied to larger datasets and other organisms.

41. Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW: **Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis**. *Mol Biol Evol* 2006, **23**:469-478.

42. Tirosh I, Barkai N: **Comparative analysis indicates regulatory neofunctionalization of yeast duplicates**. *Genome Biol* 2007, **8**:R50.

43. Nowak MA, Boerlijst MC, Cooke J, Smith JM: **Evolution of genetic redundancy**. *Nature* 1997, **388**:167-171.

44. Davis JC, Petrov DA: **Preferential duplication of conserved proteins in eukaryotic genomes**. *PLoS Biol* 2004, **2**:E55.

45. Davis JC, Petrov DA: **Do disparate mechanisms of duplication add similar genes to the genome?** *Trends Genet* 2005, **21**:548-551.

46. Harrison R, Papp B, Pal C, Oliver SG, Delneri D: **Plasticity of genetic interactions in metabolic networks of yeast**. *Proc Natl Acad Sci USA* 2007, **104**:2307-2312.

47. Kafri R, Bar-Even A, Pilpel Y: **Transcription control reprogramming in genetic backup circuits**. *Nat Genet* 2005, **37**:295-299.

48. Kafri R, Levy M, Pilpel Y: **The regulatory utilization of genetic redundancy through responsive backup circuits**. *Proc Natl Acad Sci USA* 2006, **103**:11653-11658.

49. Musso G, Zhang Z, Emili A: **Retention of protein complex membership by ancient duplicated gene products in budding yeast**. *Trends Genet* 2007, **23**:266-269.

50. Birchler JA, Riddle NC, Auger DL, Veitia RA: **Dosage balance in gene regulation: biological implications**. *Trends Genet* 2005, **21**:219-226.

51. Freeling M, Thomas BC: **Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity**. *Genome Res* 2006, **16**:805-814.

52. Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H: **Phylogenetic analyses of Vitis (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids**. *BMC Evol Biol* 2006, **6**:32.

53. Soltis PS, Soltis DE, Chase MW: **Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology**. *Nature* 1999, **402**:402-404.