

# Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes

Patrick P. Edger · J. Chris Pires

© Springer Science + Business Media B.V. 2009

**Abstract** Whole genome duplications (WGDs) followed by diploidization, which includes gene loss, have been an important recurrent process in the evolution of higher eukaryotes. Gene retention is biased to specific functional gene categories during diploidization. Dosage-sensitive genes, which include transcription factors, are significantly over-retained following WGDs. By contrast, these same functional gene categories exhibit lower retention rates following smaller scale duplications (e.g., local and tandem duplicates, segmental duplicates, aneuploidy). In light of these recent observations, we review current theories that address the fate of nuclear genes following duplication events (i.e., Gain of Function Hypothesis, Subfunctionalization Hypothesis, Increased Gene Dosage Hypothesis, Functional Buffering Model, and the Gene Balance Hypothesis). We broadly review different mechanisms of dosage-compensation that have evolved to alleviate harmful dosage-imbalances. In addition, we examine a recently proposed extension of the Gene Balance Hypothesis to explain the shared single copy status for a specific functional class of genes

across the flowering plants. We speculate that the preferential retention of dosage-sensitive genes (e.g., regulatory genes such as transcription factors) and gene loss following WGDs has played a significant role in the development of morphological complexity in eukaryotes and facilitating speciation, respectively. Lastly, we will review recent findings that suggest polyploid lineages had increased rates of survival and speciation following mass extinction events, including the Cretaceous-Tertiary (KT) extinction.

**Keywords** Diploidization · Dosage Compensation · Gene Balance · Paleopolyploidy · Polyploidy · Subfunctionalization

## Abbreviations

CNV Copy Number Variation  
GO Gene Ontology annotation  
WGD Whole Genome Duplication

## The abundance and recurrence of ancient whole genome duplications

Ancient whole genome duplications (WGDs), inferred from analyzed sequenced genomes and comparative genomics, are prevalent and recurring throughout the evolutionary history of higher eukaryotic lineages (Fig. 1a). The complete sequencing of the genome of *Paramecium tetraurelia*, a ciliated protozoan, provides

---

Responsible Editor: Edith Heard.

P. P. Edger · J. C. Pires (✉)  
371 Bond Life Sciences Center,  
Division of Biological Sciences, University of Missouri,  
Columbia, MO 65211, USA  
e-mail: piresjc@missouri.edu

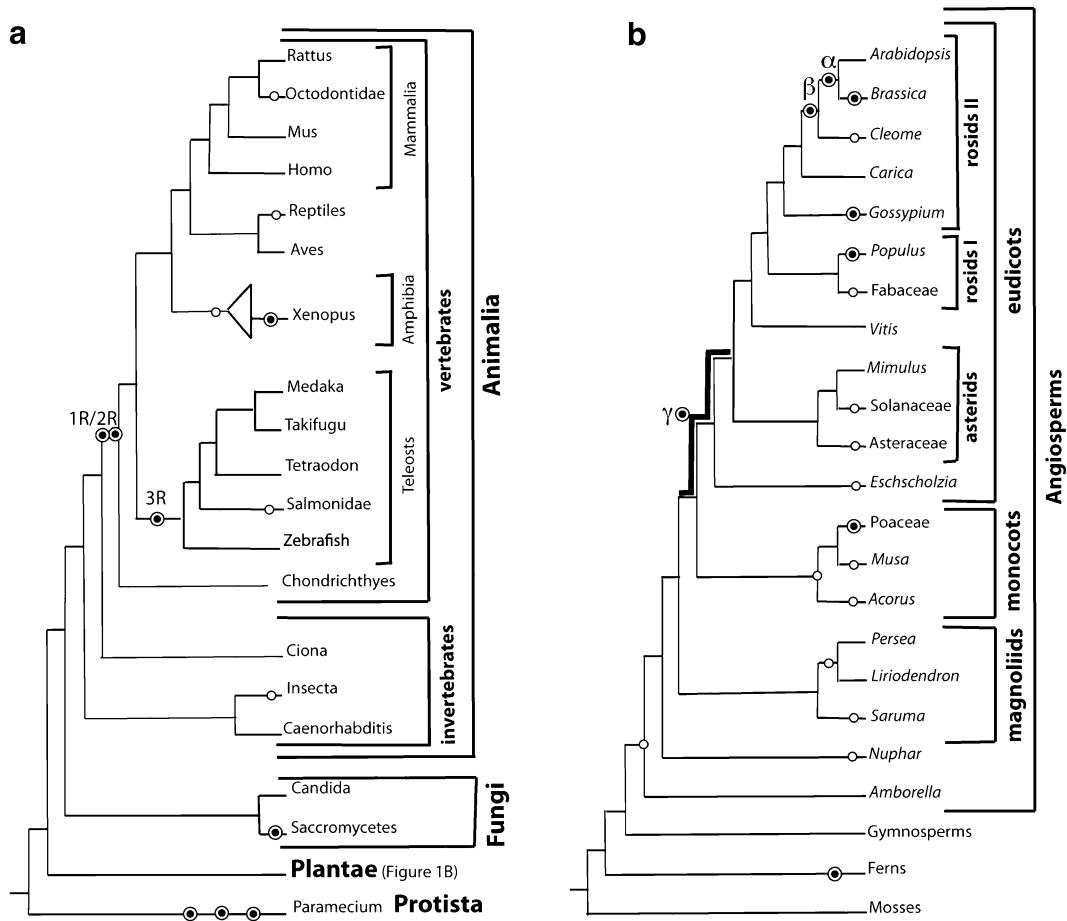
**Fig. 1** Putative positions of whole genome duplications mapped onto a heavily pruned phylogeny of the eukaryotes and angiosperms. **a** The timing of ancient whole genome duplications and distribution of polyploid eukaryotes is modified from published phylogenies (Ciccarelli et al. 2006; Freeling and Thomas 2006; Kasahara 2007). Ancient whole genome duplication events based on substantial evidence, including syntenous blocks in sequenced genomes, are denoted with partially filled in circles. The genome of the unicellular eukaryote *Paramecium tetraurelia* has undergone at least three successive whole-genome duplications (Aury et al. 2006). An ancient whole genome duplication event occurred in the *Saccharomyces cerevisiae* (baker's yeast) lineage following the divergence from a related yeast species *Kluyveromyces waltii* (Wolfe and Shields 1997; Kellis et al. 2004). At least three whole genome duplications have occurred during the evolution of vertebrates. Two rounds of whole genome duplication, 1R and 2R events, preceded the radiation of jawed vertebrates (Ohno 1970; Kasahara 2007). An additional whole genome duplication, termed 3R, occurred in the teleosts fish lineage that includes *Takifugu rubripes* (tiger blowfish) (Christoffels et al. 2004). The African clawed frog *Xenopus laevis* has undergone a recent whole genome duplication (Semon and Wolfe 2008). The distribution of polyploid eukaryotes is indicated with smaller empty circles (Otto 2007). **b** The analysis of sequenced genomes and expressed sequence tag (EST) data sets has provided evidence of multiple whole genome duplications distributed throughout several angiosperm lineages. The position of whole genome duplications shared by flowering plants is modified from a recently published studies (Soltis et al. 2009). Whole genome duplications based on substantial evidence, including analyses of sequenced genomes, are denoted with larger partially filled in circles. There is conclusive evidence based on the

evidence for three WGDs (Aury et al. 2006). An ancestor of the baker's yeast, *Saccharomyces cerevisiae*, underwent a WGD approximately 100 million years ago (Kellis et al. 2004). The analyses of gene content and gene family size data in different animal lineages suggest that two WGDs occurred at the origin of vertebrates (known as the 2R hypothesis) (Lundin 1993; Garcia-Fernandez and Holland 1994; Sidow 1996; Meyer and Schartl 1999). In addition, a more recent WGD event occurred early in the evolution of ray-finned fishes and many orders of fish have polyploid species (Amores et al. 1998; Woods et al. 2000; Amores et al. 2004; Le Comber and Smith 2004; Naruse et al. 2004). In mammals, a WGD was detected in South American desert rodents (Octodontidae) (Gallardo et al. 2004), while WGDs are widespread among insect, amphibian, and reptile lineages (Mable 2004; Otto 2007). These data collectively suggest that many if not all eukaryotes probably have experienced at least one polyploid event in their evolutionary history. The detection of ancient WGDs is often

identification of retained syntenic genomic regions that the *Arabidopsis thaliana* genome has experienced at least three ancient whole genome duplications over the last 300 million years (Vision et al. 2000; Simillion et al. 2002; Blanc et al. 2003; Bowers et al. 2003). The relative timing of these events ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) has been a contentious issue. The recent addition of fully sequenced plant genomes has aided in resolving the history of these paleo-polyploid events. The analyses of the sequenced genomes of *Populus trichocarpa* (poplar) (Tuskan et al. 2006), *Vitis vinifera* (grape) (Jaillon et al. 2007), and *Carica papaya* (papaya) (Ming et al. 2008) suggest that these species share only the most ancient paleo-polyploid event ( $\gamma$  event) with *Arabidopsis thaliana*. The timing of the  $\gamma$  event ranges from the common ancestors of the rosids and asterids to the divergence of the monocot and eudicot lineages (Lyons et al. 2008; Soltis et al. 2009). The  $\beta$  event occurred within the order Brassicales following the divergence from papaya. The  $\alpha$  event, shared with the genus *Brassica*, likely occurred within the Brassicaceae (Schranz and Mitchell-Olds 2006). The genus *Brassica* underwent an additional whole genome duplication event that occurred 7.9–14.6 mya (Lukens et al. 2004; Lysak et al. 2005). The Cleomaceae, family sister to the Brassicaceae, has an independent whole genome duplication (Schranz and Mitchell-Olds 2006). The poplar and *Gossypium* (cotton) genomes each have an additional independent whole genome duplication event (Tuskan et al. 2006; Soltis et al. 2009). The analysis of the rice and sorghum genome revealed an ancient whole genome duplication that occurred in the common ancestor of cereals (i.e. family Poaceae) (Yu et al. 2005; Paterson et al. 2009). Whole genome duplications inferred from paralog pairs filtered out of substantial EST data sets are denoted with smaller empty circles (Blanc and Wolfe 2004b; Cui et al. 2006; Soltis et al. 2009)

intricate (Martin et al. 2007), and may require multiple lines of evidence including identification of syntenic regions, analysis of gene content data, and the analysis of age distributions (i.e. Ks distributions) of duplicated genes. A robust phylogenetic framework including multiple analyzed genomes from both duplicated and preduplicated lineages simplifies the process of inferring and determining the proper placement of ancient WGDs.

Repeated rounds of WGDs, or polyploid events, have been best documented among plants, and in particular the flowering plants (Fig. 1b). Recent data suggest that a WGD occurred early in angiosperm history, possibly near the origin of all flowering plants (Cui et al. 2006; Soltis et al. 2009). In addition, polyploidy has been well documented in the grass (Poaceae; (Paterson et al. 2009)), the sunflower (Asteraceae; (Barker et al. 2008)), and the mustard (Brassicaceae; (Marhold and Lihova 2006)) families. Despite its small genome size, the model organism *Arabidopsis thaliana* (Brassicaceae) underwent at



least three ancient polyploid events ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) over the last 300 million years: the recent sequencing of the grape, papaya, and poplar genomes allow us to place these nested polyploid events into a phylogenomic context (Vision et al. 2000; Simillion et al. 2002; Blanc et al. 2003; Bowers et al. 2003; Freeling and Thomas 2006; Thomas et al. 2006; Tuskan et al. 2006; Jaillon et al. 2007).

Following a WGD, a gene duplicate will experience one of two distinct fates during diploidization (refer to glossary): either retention or loss of one member. These fates may occur through either neutral loss, selection, or via random mechanism (e.g. subfunctionalization / DDC model). Several models have been proposed to explain the retention of gene duplicates from ancient WGDs. The availability of complete genome sequences for multiple species having experienced paleo-polyploid events provides investigators a natural system to investigate changes in

gene content and the fates of nuclear genes following the duplication of entire gene networks. This system allows us to address additional questions. Are the fates of gene duplicates repeated (i.e. due to selection) or are they random? If they are repeated, are the fates for specific functional gene categories consistent across eukaryotic lineages? Which of the proposed model(s) (i.e., Gain-of-Function Hypothesis, Subfunctionalization Hypothesis, Increased Gene Dosage Hypothesis, Functional Buffering Model, and the Gene Balance Hypothesis) fit the data?

After whole genome duplications, proper balance in signaling and regulatory networks is maintained, while other types of duplication events (e.g., local, tandem, segmental, aneuploidy) leave genes out of balance to varying degrees. By investigating the fates of nuclear genes after independent WGDs and smaller scale duplications, researchers have been able to contrast the expansion of specific gene families

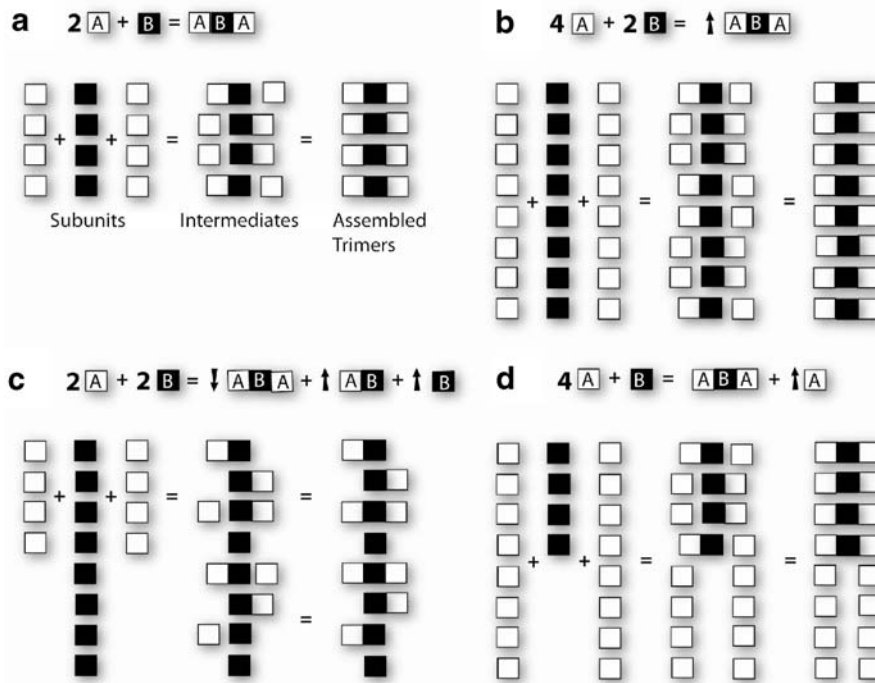
and functional gene categories following different duplication mechanisms. In this review, we summarize the evidence from these recent studies on how the Gene Balance Hypothesis (also called the Dosage Balance Hypothesis) predicts the fate of nuclear genes following both WGDs and smaller scale duplications. We then explain the mechanics of dosage-sensitivity and review different mechanisms of dosage-compensation that have evolved at the transcriptional and protein level. With these predictions, observations, and mechanisms in mind, we evaluate current alternative theories of duplicate gene retention, namely the Gain-of-Function hypothesis (Lewis 1951; Ohno 1970), Subfunctionalization (Hughes 1994; Force et al. 1999), Increased Gene Dosage Hypothesis (Seoighe and Wolfe 1999; Kondrashov and Kondrashov 2006; Conant and Wolfe 2007; Conant and Wolfe 2008), and the Functional Buffering Model (Chapman et al. 2006) in comparison to the Gene Balance Hypothesis (Birchler and Newton 1981; Birchler et al. 2001; Veitia 2002; Birchler et al. 2003; Papp et al. 2003; Birchler et al. 2005; Freeling 2008; Freeling et al. 2008). Given that the Gene Balance Hypothesis has been exclusively used to explain patterns of gene retention, we examine a recently proposed extension of the Gene Balance Hypothesis to explain the shared single copy status for a specific functional class of genes across flowering plants (Duarte et al. 2009). We review the argument that WGDs and other “balanced duplications” have played a significant role in increasing the morphological complexity in animal and plant lineages by preferentially retaining regulatory genes including transcription factors (Freeling and Thomas 2006). Finally, we will review recent findings that suggest polyploids, partly because they tend to be more vigorous and phenotypically plastic, had increased rates of survival following mass extinction events and increased rates of speciation due to evolutionary success and the development of species barriers.

### **Gene balance hypothesis: all gene duplicates are not equally retained post-WGD**

The Gene Balance Hypothesis predicts that an imbalance in the concentration of protein subunits in a macromolecular complex or between proteins with opposing functions in a transcription or signaling

network may either lead to decreased fitness or lethality (Birchler and Newton 1981; Birchler et al. 2001; Veitia 2002; Papp et al. 2003; Veitia 2005; Veitia et al. 2008). Maintaining proper protein and transcriptional balance is vital to sustain normal function. For instance, an imbalance in a highly connected portion of a network likely would result in great negative pleiotropic effects. Likewise, this is true for macromolecular complexes, especially those with regulatory processes. For example, a modification in the relative abundance of subunits in a transcription factor complex may alter the assembled complex and the expression of target genes (Birchler et al. 2001). The Gene Balance Hypothesis, supported by analyzed genomes across eukaryotic lineages, provides the basis for understanding duplicate retention following gene and genome duplication. For instance, dosage-sensitive genes must be retained in duplicate following WGD to maintain proper balance of protein and transcriptional networks. However following a smaller scale duplication (e.g. local and tandem duplicates, segmental duplicates, aneuploidy) (i.e. over-expression), duplicates of dosage-sensitive genes will tend to be eliminated to maintain proper balances.

Whole genome duplications differ from smaller scale duplications in that WGD increase the dosage of all genes simultaneously. Thus, organisms experiencing WGD immediately maintain proper balance in both signaling and transcription networks as well as stoichiometric balance in macromolecular complexes (Fig. 2). During diploidization, the spectrum of remaining duplicates would be expected to be random if gene loss is neutral. However, comparative genomic studies have revealed that gene loss is not random, which begs the question as to whether selection operates to either retain gene duplicates, return genes to single copy, or both. Interestingly, some functional gene categories, including subunits of protein complexes such as transcription factors and ribosomal proteins, and specific signal transduction components, are significantly over-retained in duplicate and have resisted loss during the diploidization process in the *Arabidopsis* (Maere et al. 2005; Freeling 2008), *Paramecium* (Aury et al. 2006), vertebrate (Blomme et al. 2006), and yeast (Papp et al. 2003) genomes. For example within the *Arabidopsis* genome, the three WGDs (alpha, beta, and gamma event) generated approximately 59% of retained duplicates (the remaining 41% are due to smaller-scale duplications); but significantly, the genes



**Fig. 2** A modification in the relative abundance of subunits involved in a complex may result in the lowered assembly of the complex, production of intermediate dimers, and excess unbound free monomers. This figure is modified from Veitia et al. 2008. **a** The assembly of the trimer A-B-A requires a specific relative abundance of A and B subunits to produce diploid levels of the assembled complex. This model assumes that complex assembly is random and that complexes do not dissociate. **b** Following a whole genome duplication, the proper relative abundance of subunits is maintained and yields only an overproduction of the assembled trimer A-B-A. The overpro-

duction of the trimer A-B-A will be consistent to that of other components in the pathway. Pathway balance is also maintained. The scenarios depicted in **C.** and **D.** may result either because of a gene duplication resulting in the over-expression of a subunit or because of gene loss following a whole genome duplication resulting in the under-expression of a subunit. **c** An increase in the abundance of monomer B results in lowered assembly of trimer A-B-A, production of dimers (A-B / B-A), and an increase in unbound free B monomers. **d** An increase in the abundance of monomer A results in the overproduction of unbound free A monomers

duplicated via polyploidy include 90% of all transcription factors, 99% of signal transducers including kinases, and 92% of all developmental genes (Maere et al. 2005). Retained duplicates show evidence for strong purifying selection in *Xenopus laevis*, *Arabidopsis*, rice, and *Paramecium* (Hughes & Hughes 1993; Aury et al. 2006; Chapman et al. 2006), and are preferentially retained in duplicate in subsequent WGDs (Aury et al. 2006; Chapman et al. 2006). Also, genes in most transcription factor families exhibit negative selection against transposition in *Arabidopsis* (Freeling et al. 2008). The co-retention of interacting dosage-sensitive genes, while maintaining balanced expression patterns, following a WGD and during diploidization is necessary to maintain proper balance of dosage-sensitive complexes and networks. Papp et al. (2003) found a significantly large excess of

interacting pairs that retained the same number of paralogs in yeast. In the *Paramecium* genome, genes involved in the common metabolic pathway or same macromolecular complex displayed significant patterns of co-retention (Aury et al. 2006). In yeast, nearly half of artificial over-expression experiments resulting in lethality involved genes encoding subunits of protein complexes (Papp et al. 2003). In addition, interacting proteins tend to have correlated patterns of co-expression with similar expression levels (Ge et al. 2001; Jansen et al. 2002; Papp et al. 2003; Ettwiller and Veitia 2007).

In contrast, these same functional gene categories, which are significantly over-retained after WGDs, are significantly under-retained following smaller scale duplications in the *Arabidopsis* (Maere et al. 2005), yeast (Li et al. 2006) and *Drosophila* genome (Dopman and Hartl 2007). Smaller scale duplicates

are instead enriched for genes whose products function at either flexible steps or at the tips of pathways (i.e. products that participate in fewer protein-protein interactions) (Li et al. 2006; Dopman and Hartl 2007). Li et al. (2006) observed that gene duplicability was negatively correlated with protein connectivity (i.e. number of protein-protein interactions) of a gene product. These results suggest that gene retention after smaller scale duplications preferentially occur to poorly connected genes, while genes retained in duplicate post-WGD tend to be in more connected in pathways. For example, the genes retained in duplicate following smaller scale duplications in the rice and *Arabidopsis* genomes are enriched for various functional gene categories including membrane proteins and proteins that function in abiotic and biotic stress (Rizzon et al. 2006). Consequently, the genes duplicated through smaller scale duplications represent different gene classes than those retained from WGDs (Davis and Petrov 2005; Maere et al. 2005). This reciprocal pattern in retention (i.e. significant over-retention following WGDs and under-retention following smaller-scale duplications) for certain functional gene categories is predicted by the Gene Balance Hypothesis (Freeling 2008; Freeling 2009).

Some highly expressed genes, including ribosomal proteins, are also significantly over-retained following WGDs in the *Paramecium*, yeast, and *Arabidopsis* genome (Seoighe and Wolfe 1999; Blanc and Wolfe 2004a; Maere et al. 2005; Aury et al. 2006; Freeling 2008). In yeast, ribosomal complexes, composed of 150 rRNA genes and 137 ribosomal protein genes and transcriptionally co-regulated to maintain proper stoichiometric balance (Warner 1999), are dosage-sensitive. As would be predicted by the Gene Balance Hypotheses, ribosome proteins are more commonly retained following duplication by WGD events compared to single-gene duplications ( $P < 10^{-21}$ ) (Papp et al. 2003). However, an increase in absolute magnitude of expression (i.e. increased dosage) for highly expressed genes, such as ribosomal proteins, may also be beneficial (Conant and Wolfe 2008). This is the theoretical basis of the Increased Gene Dosage Hypothesis (Seoighe and Wolfe 1999; Kondrashov and Kondrashov 2006; Conant and Wolfe 2007; Conant and Wolfe 2008). Ribosome complexes, which consist of up to 60% of the transcriptome, are required in high titer.

The WGD and diploidization cycle, loss of a subset of gene duplicates and retention of highly expressed dosage-sensitive genes, may yield increased expression for ribosomal proteins and other highly expressed genes. Likewise, an increase in expression of some dosage-insensitive genes may also be beneficial and result in purifying selection to retain both gene duplicates. So in addition to the Gene Balance Hypothesis, it is plausible that increased dosage for some functional genes would be beneficial. For example, an environmental change may favor increased dosage and would facilitate fixation of the duplication (Conant and Wolfe 2008).

### The mechanics of dosage-sensitivity and dosage-compensation

The Gene Balance Hypothesis not only makes predictions supported by comparative genomic data, it also provides a well-supported mechanism to explain the significant over-retention of specific functional gene categories following WGDs. This is in contrast to some alternate hypotheses, such as the Functional Buffering Model, which also attempts to explain the preferential retention of duplicated genes (Chapman et al. 2006). The Functional Buffering Model suggests that certain genes, those that function in essential processes, are retained in duplicate to buffer crucial functions (i.e. to ensure the maintenance of essential gene functions) (Chapman et al. 2006). However, the Functional Buffering Model does not provide a mechanism to explain the retention of gene duplicates, but merely proposes a hypothetical benefit of retaining certain classes of genes in duplicate. Thus, this model has no explanatory power because a mechanism is needed to properly comprehend and more accurately predict changes in gene content following WGDs. For instance, the deletion of genes retained following WGDs, based on the Functional Buffering Model, are not expected to result in lethal (or otherwise detrimental) phenotypes. Various lines of evidence presented in this paper suggest that this is not the case (e.g. over-retention of highly-connected protein coding gene duplicates that are under strong purifying selection). The Gene Balance Hypothesis does explain gene duplicate retention post-WGD in a predictive manner with a biological mechanism that is supported by published

data from across eukaryotic systems. We will compare additional proposed hypotheses (e.g., Gain of Function and Subfunctionalization), to the Gene Balance Hypothesis in a later section. Here, we review what is currently known about the mechanical basis for dosage-sensitivity. In addition, we will broadly discuss mechanisms of dosage-compensation that have evolved to alleviate harmful dosage-imbalances at the transcriptional and protein level.

The Gene Balance Hypothesis posits that, in macromolecular complexes and highly connected portions of networks, maintaining proper gene balance is required for normal function. An under- or over-expression of a dosage-sensitive subunit may induce drastic reductions of the assembled complex, and produces unassembled intermediates and free subunits (Fig. 2) (Veitia et al. 2008). Collectively, these changes may lead to decreased fitness (Birchler et al. 2001; Veitia 2002; Papp et al. 2003; Veitia 2003a; Veitia 2003b; Veitia et al. 2008). Components with greater protein connectivity (e.g. central subunit in a complex) have increased chances of producing unassembled intermediates when over-expressed (Fig. 2c) (Veitia et al. 2008). This prediction is consistent with observations that dosage-sensitivity is influenced by the size (i.e. number of interactors) of the molecular complex (Papp et al. 2003).

In addition to protein connectivity, the function of the protein has a significant influence on the sensitivity to dosage imbalances (Freeling 2008; Veitia et al. 2008). A reduced number of assembled complexes may disrupt the balance of opposing actions in a network (i.e. an inhibitor and activator acting on a common target) (Veitia et al. 2008). For instance, the over-expression of the central subunit B in the trimer A-B-A will lead to decreased assembly of the trimer, production of dimer intermediates, and increased production of free unbound B monomers (Fig. 2c) (Veitia et al. 2008). If the trimer is an activator (e.g. kinase), the decreased production of the complex may result in a network imbalance with the opposing inhibitor (e.g. phosphatase). This is the predicted outcome when complex assembly is random (Veitia et al. 2008).

A variety of dosage-compensation mechanisms have evolved to alleviate harmful dosage-imbalances, including at both the protein and transcriptional level. In the previous paragraph, we discussed how the over-expression of monomer B leads to the decreased

production of trimer A-B-A when complex assembly is random. The effects of over-expressing monomer B (i.e. decreased production of trimer) could be eliminated if complex assembly is non-random. Specifically, the production of intermediate dimers would diminish when the reaction leading to the trimer complex is faster than that leading to the dimers (Veitia et al. 2008). In other words, the dimers (A-B or B-A) would bind another A subunit at a faster rate compared to free unbound B monomers. This non-random assembly, based on kinetics and assembly pathway, would aid in mitigating the effects of over-expressing the B subunit by increasing the production of the trimer (Veitia et al. 2008). This may diminish dosage-sensitivity for many complexes, but is limited to specific scenarios. For example, the under-expression of the A or B subunit will yield a reduction of the trimer A-B-A in either assembly scheme (i.e. via random and non-random assembly). This highlights the importance of maintaining proper gene balance and provides additional support for the Gene Balance Hypothesis.

Various mechanisms have evolved to eliminate the toxic scenarios generated by various free unassembled monomers (Veitia et al. 2008). For example, the protein Rb12p binds to free  $\beta$ -tubulin subunits transiently to eliminate toxicity caused by the accumulation of unbound subunits, which disrupt microtubule assembly and function (Abruzzi et al. 2002). Similarly, the Rad53 protein kinase will form a complex with histone proteins to regulate protein abundance (Gunjan and Verreault 2003). Additionally, Veitia et al. (2008) hypothesized that unassembled monomers are preferentially degraded compared to complexes via exposed degradation signals. They proposed that the formation of complexes might mask monomer degradation signals (i.e. degradation signals are buried inside the complex). The masking of all degradation signals, as a result of complex formation, would lead to the preferential degradation of monomers and intermediates, and protect the assembled complex from protein degradation (Veitia et al. 2008). These mechanisms, both demonstrated and hypothetical, aid in alleviating harmful dosage imbalance effects caused by excess unbound protein monomers.

Dosage-compensation also occurs at the transcriptional level. The loss or gain of genes encoding subunits of a complex may be countered by either the inverse change in gene expression from the alternate copies of the gene or the equivalent change in gene

expression from all of the other genes within the complex to maintain proper balance (Veitia et al. 2008). If one partner in a complex is over-expressed, then the overproduction of its partner(s) is needed to maintain proper stoichiometric balance. This may certainly also involve altering the rates of mRNA degradation (Veitia et al. 2008). Additionally, the involvement of linked regulators that act negatively on the expression of a target gene may result in compensation (Birchler 1981; Birchler et al. 2005; Veitia et al. 2008). Following a segmental duplication (i.e. trisomic state), the linked regulators would down-regulate the expression of the three copies of the target gene. In the monosomic state, the single linked regulator would up-regulate the expression of the target gene. In both examples, the expression of the target gene would be nearly equivalent to that of a diploid (Veitia et al. 2008).

These mechanisms of dosage-compensation may have evolved to aid in both equalizing gene expression and alleviating the toxicity caused by the free unbound monomers. These processes, at both the protein and transcriptional level, are clearly supportive of the prediction that maintaining proper gene balance is required for normal functions. However, dosage-compensation mechanisms certainly are limited to specific genes and scenarios. For example, aneuploids have unlinked effects (i.e. affecting expression of genes whose dosage was not altered) (Guo and Birchler 1994). Additionally, it is unlikely that each gene has a linked regulator to equalize expression or a regulatory protein to bind excess unbound monomers. This is evident in the analysis of the genomic data (i.e. reciprocal pattern in retention for dosage-sensitive genes). If dosage-compensation mechanisms did exist globally for all genes, we would not observe a skewed retention of dosage-sensitive genes post-WGD and under-retention of dosage-sensitive genes following smaller scale duplications.

### **Current alternate hypotheses do not explain retention of duplicate genes**

Given the observation, predictions, and mechanisms that have thus far been reviewed for the retention of duplicate genes after WGD and smaller scale duplications, we can now evaluate additional currently invoked alternative theories of duplicate gene reten-

tion, namely the Gain of Function Hypothesis (Lewis 1951; Ohno 1970) and the Subfunctionalization Hypothesis (i.e. the duplication, degeneration, complementation (DCC) model) (Hughes 1994; Force et al. 1999). These alternate hypotheses are currently widely accepted explanations for the over-retention of gene duplications following WGDs. However, do the gene content data support these alternate hypotheses? Do these hypotheses provide ‘primary’ mechanisms that could be responsible for the retention of the specific functional gene categories observed in all the aforementioned analyzed genomes?

The Subfunctionalization Hypothesis argues that duplicate genes are preserved in pairs through a two-step neutral mutational process that partitions the ancestral functions to different gene copies (Lynch and Force 2000). Following subfunctionalization, both gene duplicates, which now specialize in complementary functions, are under strong purifying selection to retain the entire ancestral function. Subfunctionalization may involve structural domains of genes or gene expression patterns (quanta, spatial, and temporal) (Conant and Wolfe 2008). There is overwhelming evidence from multiple experiments that the subfunctionalization mechanism does occur to retained duplicates (Force et al. 2005; He and Zhang 2005). This raises two important questions. Is subfunctionalization responsible for the retention of specific functional gene categories following WGD? Or, do gene dosage constraints provide a mechanism for initial retention that provides a longer time frame for subfunctionalization? If so, how can retained dosage-sensitive gene duplicates subsequently subfunctionalize without resulting in a detrimental dosage-imbalance?

The Gain of Function Hypothesis asserts that gene duplication followed by innovation, that is the evolution of a novel function (i.e. Neofunctionalization), in one daughter gene is the primary source of new genes (Lewis 1951; Ohno 1970). The other duplicate maintains the ancestral function. Strong positive selection and inactivation of gene conversion drives the fixation of the novel gene (Clark 1994; Lynch and Force 2000; Innan 2003; Beisswanger and Stephan 2008). Neofunctionalization has occurred over evolutionary time and has certainly contributed to lineage-specific differences in gene content. But, is neofunctionalization responsible for the retention of specific functional gene categories following WGD? Or, does neofunctionalization operate only after genes are



retained in duplicate due to gene dosage constraints (similar to subfunctionalization)? If so, how can retained dosage-sensitive gene duplicates subsequently neofunctionalize without resulting in a detrimental dosage-imbalance?

Although neofunctionalization and subfunctionalization do occur and provide specific mechanisms, neither should be accepted as the primary models to explain the over-retention of gene duplicates post-WGD, as reviewed by Freeling (2008). Freeling (2008) argues that the Gene Balance Hypothesis best explains the gene content data from several sequenced genomes across eukaryotic lineages, each with independent WGDs and smaller scale duplications. Neofunctionalization and subfunctionalization do not predict a reciprocal pattern in retention as observed for specific gene categories post-WGD and smaller scale duplications (Fig. 3). Instead, these two alternative hypotheses predict that any gene may be retained following any sort of duplication (i.e. WGD and smaller scale duplications). In other words, these hypotheses make no predictions between functional gene categories (i.e. gene function) and retention frequency post-duplication. We concur with Freeling (2008) that neofunctionalization and subfunctionalization occur largely after genes are retained in duplicate due to gene-dosage constraints. The gene content data, specifically the reciprocal relationship in duplicate retention frequencies for certain GO categories, supports only the Gene Balance Hypothesis. Even though Neofunctionalization and Subfunctionalization may be “downgraded” by the Gene Balance Hypothesis (Freeling 2008), it is important to note that these are not mutually exclusive hypotheses and a pluralistic if not unified framework is likely. For example, the data are suggestive that gene dosage constraints provide a mechanism for initial retention that provides a longer time frame to allow alternate mechanisms (i.e. neofunctionalization and subfunctionalization). Thus, WGDs allow for certain functional gene categories to undergo subfunctionalization and neofunctionalization that would not have occurred following smaller-scale duplications, and vice versa. This concept will be discussed more thoroughly in a subsequent section. Additionally, we want to note that these alternate mechanisms may explain the retention of a subset of dosage-insensitive gene duplicates following WGDs. However, these alternate mechanisms do not occur at a sufficient frequency in the

short periods following WGDs to distort the observed reciprocal pattern in duplicate retention.

### **Extending the gene balance hypothesis to dosage-sensitive single copy genes**

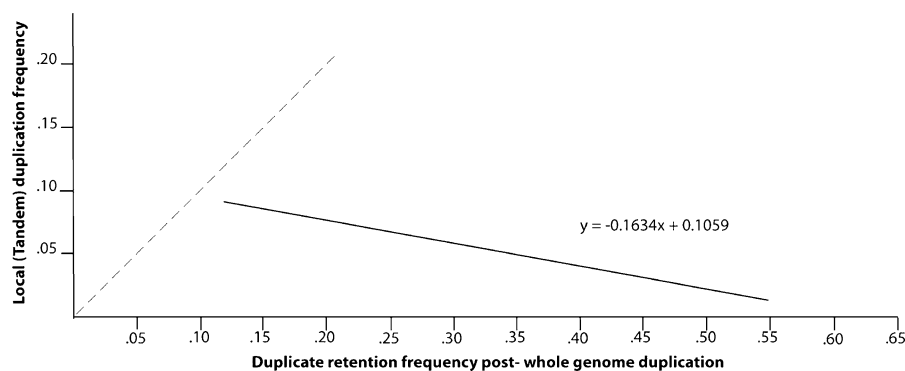
Although most polyploidy comparative genomic studies have focused on the non-random retention of duplicate genes, some studies have also pointed out the non-random single copy status of genes following WGD (Chapman et al. 2006; Paterson et al. 2006). After polyploidy, do only random processes account for the single copy status of genes? Does a unique functional class of genes exist (or perhaps multiple classes) that is under strong selective pressures to repeatedly return to single copy?

Given that the Gene Balance Hypothesis has been exclusively used to explain the over-retention of dosage-sensitive genes post-WGDs, we examine a recently proposed extension of the Gene Balance Hypothesis to explain the shared single copy status for a specific functional class of genes across the plant kingdom (Duarte et al. 2009). Such genes that have repeatedly returned to single copy have been referred to by some as “duplication-resistant” (Paterson et al. 2006). However, this term has been considered misleading or confusing; because the genes are in fact duplicated and then lost (only one duplicate member), we will use the term “selected single copy” genes. Paterson et al. (2006) suggest that the single copy state for these genes was important for the long-term survival of polyploids; however, they do not propose a mechanism by which such genes repeatedly and convergently return to single copy across diverse organisms. But several are easily envisioned, one leading candidate being strong selection against increased dosage of some genes representing certain networks or pathways.

Hypothetically, the single copy status of a gene following a WGD can be explained in two ways. First, the loss of a gene duplicate occurred through random deletion. Second, a biological mechanism exists which repeatedly restores some genes to a single copy state. However, no such biological mechanism had previously been proposed. The Gene Balance Hypothesis would predict that gene duplicates, which are not under selection to be retained in duplicate post-WGD (i.e. dosage-insensitive genes),

**a**

Molecular Function GO Category	GO Term	Retention post- WGD	Local (Tandem) Duplicates
Structural constituent of ribosome	GO:0003735	Over-retained	Under-retained
Structural molecule activity	GO:0005198	Over-retained	Under-retained
Transcription Factor activity	GO:0003700	Over-retained	Under-retained
Transcription Regulator activity	GO:0030528	Over-retained	Under-retained
DNA binding	GO:0003677	Over-retained	Under-retained
Nucleic acid binding	GO:0003676	Over-retained	Under-retained
Cytidine deaminase activity	GO:0004126	Not Significant	Over-retained
Endopeptidase activity	GO:0004175	Not Significant	Over-retained
Hydrolase activity	GO:0016787	Not Significant	Over-retained
Deaminase activity	GO:0019239	Not Significant	Over-retained
Oxygen binding	GO:0019825	Not Significant	Over-retained
Carbohydrate binding	GO:0030246	Not Significant	Over-retained

**b**

**Fig. 3** Duplicate retention frequencies for select GO categories following the alpha whole genome duplication (WGD) and local (tandem) duplications in *Arabidopsis thaliana* (modified from Freeling 2008). These data support the Gene Balance Hypothesis, and discredit alternate hypotheses as the primary mechanism to explain the retention of gene duplicates following whole genome duplications. **a** The retention frequencies, significantly over- / under- retained or not significant, for select GO categories following the alpha event and local duplications were calculated by Gostat (Freeling 2008). The first six GO categories exhibit a reciprocal relationship in retention (i.e. significant over-retention following WGD and under-retention following tandem duplications). This pattern in retention is supportive of the Gene Balance Hypothesis. The last six GO categories are significantly over-retained following tandem duplications. In the *Arabidopsis* and rice genome, tandem duplicated genes are enriched for abiotic and biotic stress, and membrane protein functions (Rizzon et al. 2006). Rizzon et al. (2006) suggested that these products function either at flexible steps in a pathway or at the end of biochemical pathways. The Gene Balance Hypothesis predicts no significant retention frequency for dosage-insensitive genes following whole genome

are lost at random. Dosage-insensitive genes are somewhat likely to repeatedly return to a single copy state, given sufficient time, following repeated rounds of genome duplication. However, dosage-insensitive genes are also more likely to exhibit copy number variation (CNV) (Dopman and Hartl 2007). Based on

duplications. These GO terms are only a subset of GO categories analyzed by M. Freeling. He did not observe a single GO category that was significantly under-retained following whole genome duplications and over-retained following tandem duplications. **b** The duplicate retention distribution following the alpha ( $\alpha$ ) WGD event (X-axis) and tandem duplications frequency (Y-axis), shown as a linear regression, for 16 *Arabidopsis* transcription factor families each with at least 30 genes. The X-axis was calculated (number of alpha pairs / total number of genes in family). The Y-axis was calculated (number of tandem duplicates / total number of genes in family). A skewed distribution (i.e. reciprocal relationship), a negative relationship with 95% confidence, is observed and is drawn with a solid line. This distribution is supportive of the Gene Balance Hypothesis. The null hypothesis distribution, drawn with a dotted line, is clearly rejected and predicts that gene retention is random following both tandem duplications and the alpha whole genome duplication event. Transcription factors are preferentially retained in duplicate following whole genome duplications due to gene dosage constraints, and are significantly under-retained following tandem duplications. This figure was modified from M. Freeling (2008, 2009)

the predictions of the Gene Balance Hypothesis, it is certainly possible that many if not most single copy genes are merely dosage-insensitive genes that have repeatedly and convergently returned at random to single copy following independent WGDs and smaller scale duplications. This scenario of random

loss is a testable hypothesis because the probability of sharing the single copy state would decrease as the number of genomes sampled increases.

The Gene Balance Hypothesis has recently been extended to explain the shared single copy status for a particular set of dosage-sensitive genes, namely the nuclear encoded organellar (plastid and mitochondria) genes (Duarte et al. 2009). This hypothesis, which we call the Selected Single Copy Gene Hypothesis, claims that a subset of nuclear encoded genes might encode dosage-sensitive proteins that function in either organellar signaling networks or macromolecular complexes that must maintain proper stoichiometric balance with interacting partner(s) that are encoded in the organellar genome (Duarte et al. 2009). While the chloroplast proteome is composed of 2,100 to 3,600 proteins, almost all of these proteins are encoded in the nucleus (Abdallah et al. 2000; Leister 2003). Similarly, the plant mitochondrial proteome contains approximately 2,000 to 3,000 gene products (Millar et al. 2005), while the plant mitochondrial genome encodes approximately 30 to 40 proteins (i.e. 1% – 2% of the proteome) (Adams et al. 2002). The Selected Single Copy Gene Hypothesis (Duarte et al. 2009) is an amendment to the Gene Balance Hypothesis that predicts that all gene duplications (i.e. whole genome duplications and smaller scale duplications) of any of these genes would result in a dosage-imbalance and selection against duplicate retention.

The Selected Single Copy Gene Hypothesis still requires substantial further investigation, particularly because it is still unclear how gene balance is coordinated between the nuclear genome (one nucleus per cell) and organellar genomes (many chloroplasts or mitochondria per cell) (Duarte et al. 2009). To what degree is the stoichiometric balance of an organellar protein complex, encoded by both the nuclear and organellar genome, upset following the duplication of a nuclear encoded subunit? For example, the ribosomal protein S13 (rps13) gene for both the plastid and mitochondrial ribosomal complex is encoded in the *Arabidopsis thaliana* nuclear genome by two separate genes (Mollier et al. 2002). As previously discussed, ribosomal complexes are sensitive to dosage-imbalance. As predicted by the Gene Balance Hypothesis, nuclear ribosomal protein genes are significantly over-retained in duplicate following WGDs to maintain proper stoichiometric balance (Papp et al. 2003; Aury et al. 2006). In comparison, the Selected

Single Copy Gene Hypothesis (Duarte et al. 2009) would predict that a gene duplication (i.e. over-expression) of the ribosomal protein rps13 that encodes a subunit of the mitochondrial organellar ribosomal complex would result in a dosage-imbalance and selection against duplicate retention.

If the Selected Single Copy Gene Hypothesis is valid, it may be necessary to investigate mechanisms that might have evolved to mitigate gene dosage imbalances between the nuclear genome and organellar genomes. For instance, are relative “diploid” transcript levels and/or protein levels maintained following a tandem duplication involving a selected single copy gene? Will artificial over-expression of a nuclear encoded mitochondrial gene in yeast lead to a deleterious dosage-imbalance? In addition, it is still entirely unclear how recent duplicates of shared single copy genes are eliminated following duplication. If there is selection for single copy status, there must be a mechanism besides random and neutral processes. This may involve epigenetic gene silencing of either duplicate followed by pseudogenization. Typically gene silencing involves both homologous copies, as observed with introduced transgenes in genetically modified plants (Stam *et al.* 1997). For this mechanism to work, gene silencing would have to target only one gene copy. Alternatively, there could be strong positive selection for the deletion of one member of a pair. In short, future studies need to demonstrate that these genes are not shared in a single copy state due to random chance alone. The skewed distribution toward certain GO functions (i.e. significant overrepresentation of organellar gene functions) for shared single copy genes across four angiosperm genomes (*Arabidopsis*, *Populus*, *Vitis*, and *Oryza*), a moss genome (*Physcomitrella*), and one lycophyte genome (*Selaginella*) suggests that many of these genes did not return at random (Duarte et al. 2009). This compilation suggests that a set of “selected single copy” genes may actually exist, at least in plant genomes. Because only a limited set of plant genomes was analyzed (Duarte et al. 2009), we predict that the percentage of genes shared at random in single copy within the conserved single copy list will decrease as more available sequenced plant genomes are added (e.g. *Carica*, *Sorghum*, and *Mimulus*).

The utility of shared single copy genes as global phylogenetic markers has been proposed and demonstrated at both a high throughput approach using

transcriptome data across the angiosperms and at a family level using a standard reverse transcription—polymerase chain reaction (PCR) protocol (Duarte et al. 2009). In addition to selected single copy genes, which theoretically are strictly orthologous across species since paralogs are selected against, dosage-sensitive genes are excellent phylogenetic markers. Establishing orthology for nuclear genes across divergent species, which is required for constructing accurate phylogenetic estimates (Alvarez and Wendel 2003), is generally not a trivial exercise. Dosage-insensitive genes that are shared at random in single copy or low copy across species have a fifty percent probability between any two species of sharing a paralogous relationship. In contrast, dosage-sensitive genes have characteristics that aid in ascertaining orthologous relationships, including rarely exhibiting CNV (i.e. infrequently duplicated via non-WGDs) and lower transposition frequencies.

### Recurring WGDs facilitate speciation and increases in morphological complexity

Finally, we speculate that both the preferential retention of regulatory genes and loss of genes following WGDs have played a significant role in facilitating speciation, diversification, and increasing morphological complexity. In addition to polyploid reproductive barriers (i.e. reproductive isolation between polyploids and their diploid progenitors) and polyploid heterosis (i.e. polyploids exhibit greater biomass, fertility, and speed of development, thus are more vigorous and tend to out-compete diploid competitors), the neutral loss of alternate members of dosage-insensitive gene duplicates in different species, also known as reciprocal gene loss, has been well established to drive reproductive isolation and correlate with rapid speciation shortly after WGDs (Scannell et al. 2006; Scannell et al. 2007; Semon and Wolfe 2007). Specifically, the reciprocal loss of different copies (i.e. paralogs) between two closely related species can create a Bateson-Dobzhansky-Müller hybrid incompatibility, which reduces the viability and fertility of hybrids (Scannell et al. 2006; Scannell et al. 2007; Semon and Wolfe 2007). These hybrid incompatibilities create species barriers and a robust lineage-splitting force, which appear to have contributed to the rapid speciation of the yeast, teleost and angiosperm lineages following WGDs (Soltis and

Soltis 2004; Scannell et al. 2006; Semon and Wolfe 2007). This is supported by recent observations that lineage specific WGDs had contributed to dramatic increases in species richness in several angiosperm families including Poaceae, Brassicaceae, Solanaceae, and Fabaceae (Soltis et al. 2009). This correlation between WGDs and diversification rates is more apparent when comparing species richness between polyploid lineages to sister lineages lacking the WGD. For example, the teleost clade that shares the 3R event are the largest and most diverse group of vertebrates (~22,000 species), while the sister ‘basal’ ray-finned species lineages have only a few extant species (Van de Peer 2004; Hurley et al. 2007). Similarly, the mustard family (Brassicaceae) are composed of approximately 3,700 species, while the earliest diverging lineage (Tribe Aethionema), which lacks the *Arabidopsis thaliana* alpha ( $\alpha$ ) WGD, has 57 species (Al-Shehbaz et al. 2006; Schranz and Mitchell-Olds 2006).

In addition to promoting speciation, WGD could also increase morphological complexity by providing the ‘building blocks’ (i.e. retained duplicated regulatory networks and transcription factors) that may later evolve novel regulatory functions. The evolution of novel morphologies and morphological variation in living organisms has been of general interest to most biologists. A growing body of evidence suggests that changes in both the coding sequence of regulatory proteins and in the non-coding regulatory sequences of their targets are primarily responsible for developmental novelty (Carroll 2005; Wray 2007; Lynch and Wagner 2008). In short, major organismal differences (e.g. anatomical and behavioral) are largely due to changes in gene expression, rather than protein repertoire. For example, this certainly accounts for a subset of the observed differences between chimpanzees and humans (King and Wilson 1975). It has been argued that morphological complexity has increased over time in both animal and plant lineages (Freeling and Thomas 2006). Is there a pattern for such a trend? Is there a predictable mechanism that contributes to the expansion of regulatory protein families? Freeling and Thomas (2006) proposed that repeated rounds of WGDs have driven the increase in morphological complexity, an observation that has been found by other researchers as well (Blomme et al. 2006). Specifically, Freeling and Thomas argue that the increase in complexity has been driven through the emergence of novel regulatory functions (e.g. tran-

scription factor with a novel function) across new developmental boundaries. As discussed in previous sections, genes encoding transcription factors are typically dosage-sensitive, preferentially duplicated via WGDs, and are under strong purifying selection to be retained in duplicate (Fig. 3). These observations raise important questions. How does functional divergence (i.e. neo- and sub-functionalization) occur between retained transcription factor duplicates without resulting in a detrimental dosage-imbalance? A transcription factor duplicate must first escape from these constraints in order to allow functional divergence to occur. How is a new balance in regulatory proteins that forms a novel regulatory complex achieved? Birchler et al. (2007) suggested that accumulating changes in cis-dominant regulatory regions of critical target loci may allow for a shift in the balance of regulators in the complex. A mutation in a cis-regulatory region could modify gene expression exclusively in a specific spatiotemporal domain (i.e. not globally). These changes would progressively occur to multiple target genes until a new balanced state of the regulatory complex is tolerated or selected to resolve the resulting intra-genomic conflict (Birchler et al. 2007).

Once a new balanced state is tolerated, a spatiotemporally separated gene regulatory relationship may evolve to give rise to novel morphologies. The development of a new gene regulatory network may include the recruitment of additional transcription factors, the evolution of novel elements (e.g. tissue-specific enhancer) that may completely replace an ancestral regulator, and possibly the evolution of novel transcription factor functions (Lynch and Wagner 2008). Recent data indicates that many transcription factors are tissue specific (Yu et al. 2006). In the human genome, approximately 30% of transcription factors were specific to only a single tissue (Yu et al. 2006; Lynch and Wagner 2008). Tissue-specific transcription factors and cis-regulatory elements limit negative pleiotropic effects caused by mutations (Carroll 2005; Lynch and Wagner 2008). For example, a mutation in a single cis-regulatory element will modify gene expression only in the spatiotemporal domain governed by that regulatory element (Carroll 2005). This characteristic will allow adaptive evolution to modify and evolve novel morphologies without having extensive negative pleiotropic effects (Carroll 2005; Wray 2007; Lynch and Wagner 2008).

Does the preferential retention of regulatory proteins following WGDs facilitate the functional divergence of transcription factors? Is there a correlation between WGDs and increased complexity? The divergence in transcription factor functions (i.e. Neofunctionalization) has been well documented in multiple animal and plant lineages (Lynch and Wagner 2008), which tend to co-occur with the origin of novel morphological structures, rapid cladogenesis, and WGDs. For example, the MADS Box transcription factor family in plants regulates many important aspects of plant development including floral organ development and initiation of flowering (Coen and Meyerowitz 1991; Michaels and Amasino 1999). The co-occurrence of many duplicated genes, including MADS-Box subfamilies, near the origin of the angiosperms suggests that these duplicates arose via WGD (Zahn et al. 2005; Soltis et al. 2007a; Soltis et al. 2007b) (Fig. 1b). Soltis et al. (2007b) suggested that functional divergence in MADS-Box duplicates, APETELA3 (AP3) and PISTILLATA (PI), following this event aided in the origin and subsequent diversification of all flowering plants (i.e. the origin of the flower). The AP3 B-class gene subsequently diverged in functions giving rise to TM6 and euAP3 near the base of the core eudicot radiation, which coincides with the gamma ( $\gamma$ ) whole genome duplication in *Arabidopsis thaliana* and the origin of the eudicot flower (Cui et al. 2006; Rijpkema et al. 2006; Hernandez-Hernandez et al. 2007; Soltis et al. 2009) (Fig. 1b). Additionally, recent studies suggest that polyploid plants had increased chances of surviving mass extinction events (Fawcett et al. 2009; Soltis and Burleigh 2009). Fawcett et al. (2009) proposed that polyploid species, demonstrated to be remarkably plastic and highly adaptable (Osborn et al. 2003; Lukens et al. 2004), had increased tolerances to low sunlight and other drastic environmental changes during the Cretaceous-Tertiary (KT) extinction event 65 million years ago (mya). This claim is supported by evidence that shows that many angiosperm lineages, distributed across monocots and eudicots, have independent WGDs that coincide with the timing of the KT event (Fawcett et al. 2009).

A similar pattern in diversification of regulatory functions following WGDs is also observed in animal lineages. For example, the expansion of transcription factor families, including HOX gene functions, following the 1R and 2R event, contributed to the

evolution of complex vertebrates (i.e. origin of the vertebrate skeleton) (Fig. 1a) (Amores et al. 1998; Blomme et al. 2006). Similarly, the 1R/2R events, which occurred approximately 520 to 550 mya (Holland et al. 2008; Putnam et al. 2008), coincide with a mass extinction that occurred at the dawn of the Cambrian period approximately 544 mya (Bowring et al. 1993; Knoll and Carroll 1999). While the teleost 3R event, estimated 226–316 mya, may have coincided with a mass extinction approximately 250 mya at the end of the Permian period (Hurley et al. 2007).

The co-occurrence pattern of WGDs, mass extinction events, and rapid diversifications, observed across angiosperm and animal lineages, suggest that polyploids have increased chances of surviving mass extinction events and rapidly colonizing old and new ecological niches made available by the mass extinction events. Recent polyploids are remarkably plastic and exhibit great variation in novel phenotypes including organ size and changes in developmental timing, which may allow them to differentiate from diploid progenitors, out-compete diploid species, and enter new ecological niches (Pires et al. 2004; Gaeta et al. 2007). In other words, the phenotypic plasticity and vigor of nascent polyploids likely aided them to survive mass extinction events and while providing a competitive edge over formerly existing diploid species to invade novel niches resulting in rapid rates of speciation. Over longer periods of time, the functional divergence of retained duplicated regulatory networks likely was selectively favored following these extinction events gradually giving rise to novel morphologies (e.g. the flower that attracts insects to aid in efficient long-distance pollination). Interestingly, recent data suggest that early fossil angiosperms were insect-pollinated and that the origin of specialized pollinators including bees occurred during the major radiation of the angiosperms (Hu et al. 2008).

Lastly, in some cases novel transcription factors do arise from smaller scale duplications. However, the frequency of retention is much more rare (Fig. 3b), and appears to be biased toward specific transcription factor families and these exceptions may partially be explained by balanced segmental duplications (Freeling and Thomas 2006; Freeling 2008). This alternate duplication mechanism is supported by the fact that genes encoding subunits of the same complex tend to be clustered together on chromosomes

(Lee and Sonnhammer 2003; Teichmann and Veitia 2004). A segmental duplication involving clustered subunits, if all subunits are co-duplicated, would maintain proper stoichiometric balance of the complex. An unbalanced small-scale duplication, based on all the aforementioned data, would have to be followed by a mechanism (e.g. dosage-compensation at the transcriptional level) to maintain proper stoichiometric balance of the transcription factor complex or it would result in decreased fitness of the organism. In conclusion, we argue that the evolution of novel morphologies are highly dependent on WGDs, and likely would not have resulted following a (or even a series) of smaller scale duplications.

## Summary

Maintaining proper balance in pathways (signaling and regulatory) and stoichiometric balance of macromolecular complexes is essential for normal function and development. This provides the basis for predicting which gene duplicates are retained and lost following various duplication mechanisms. Some dosage-sensitive functional gene categories, such as transcription factors, show a reciprocal pattern in retention (i.e. significant over-retention post-WGD and significant under-retention following smaller-scale duplications). This reciprocal pattern in retention is only supportive of the Gene Balance Hypothesis. Alternative hypotheses, namely the Subfunctionalization and Gain-of-Function hypotheses, are not supported by the gene content data from eukaryotic genomes as a ‘primary’ explanation for duplicate gene retention post-WGDs, at least initially. The data suggest instead that these processes occur largely after genes are already retained in duplicate due to gene dosage constraints. In other words, gene dosage constraints retain dosage-sensitive gene duplicates following WGDs, while providing these alternate mechanisms longer periods of time to operate. However, these alternate mechanisms (e.g. Subfunctionalization and Neofunctionalization) may only occur once gene dosage-constraints are alleviated. Additionally, the repeated presence of some genes only as single copy genes may be explained in part by a recent extension of the Gene Balance Hypothesis. This extension, Selected Single Copy Gene Hypothesis, may explain the shared single copy status of some nuclear encoded organellar (plastid and mitochondria) genes across the plant

kingdom, which may be under strong selection to maintain proper balance between the nuclear and organellar genome and repeatedly return to single copy following all modes of duplication (i.e. WGD and smaller scale duplications). Lastly, we support recent arguments that WGDs have contributed to increases in morphological complexity and cladogenesis in eukaryotic lineages. The Gene Balance Hypothesis provides a unifying mechanism to explain the impact of polyploidy simultaneously both in the short term and over longer time periods: the immediate neutral reciprocal loss of dosage-insensitive genes that can lead to rapid speciation post-WGD (e.g., Bateson-Dobzhansky-Müller hybrid incompatibility) as well as the long-term significant over-retention of regulatory genes post-WGD, followed by functional divergence, that have contributed to novel variation and developmental evolution in eukaryotic lineages (e.g. angiosperms and vertebrates) over deep time.

**Acknowledgements** We thank Jim Birchler, Gavin Conant, Eric Lyons, Michael Freeling, Doug Soltis, Lex Flagel, Marta Wayne, Keith Adams, the phylophiles and anonymous reviewers for their thoughtful comments. PPE and JCP are funded by the NSF Plant Genome Program (DBI-0733857 and DBI 063836).

## Glossary

**Whole genome duplications** an event, which arises through either mitotic or meiotic misdivisions (i.e. unreduced gametes) and may involve interspecific hybridizations, that multiplies the number of complete chromosome sets (i.e. genome).

**Polyploids** can either be defined as containing three or more complete genomes derived from a single species termed Autopolyploids or containing three or more complete genomes derived from at least two different species termed Allopolyploids.

**Diploidization** This is an ongoing process, involving a suite of mutational mechanisms, which tends to return a polyploid genome back to a smaller genome and base chromosome size (i.e. ancestral ‘diploid’ state). The

diploidization process, includes genetic (e.g. aneuploidy, chromosomal rearrangements, deletion of repetitive sequences, and changes in gene content), epigenetic (e.g. gene silencing), and transcriptional (i.e. altered gene expression) changes, may begin immediately following a whole genome duplication (Osborn et al. 2003; Adams and Wendel 2005a; Adams and Wendel 2005b; Gaeta et al. 2007; Ha et al. 2007; Ha et al. 2009; Pignatta and Comai 2009; Rapp et al. 2009). This review focuses on the two fates gene duplicates face: either the loss of one gene duplicate member, through a process referred to as fractionation (Freeling 2008; Freeling 2009), or the preferential retention of both duplicate members.

**Statistical evaluation of duplicate retention**

To test for significant over-retention and under-retention of gene duplicates across Gene Ontology (GO) annotation categories, applications such as GOSTat (Beissbarth and Speed 2004) with a rigorous statistical foundation calculates the probability that the observed pattern in duplicate retention did not occur randomly. The GOSTat program was used to calculate the significant over- and under-retention for a subset of GO categories in the *Arabidopsis thaliana* genome (Fig. 3) (Freeling 2008).

## References

- Abdallah F, Salamini F, Leister D (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci* 5:141
- Abruzzi KC, Smith A, Chen W, Solomon F (2002) Protection from free beta-tubulin by the beta-tubulin binding protein Rbl2p. *Mol Cell Biol* 22:138–147
- Adams KL, Wendel JF (2005a) Novel patterns of gene expression in polyploid plants. *Trends Genet* 21:539–543
- Adams KL, Wendel JF (2005b) Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 8:135–141
- Adams KL, Qiu YL, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: high

- and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A* 99:9905–9912
- Al-Shehbaz IA, Beilstein MA, Kellogg EA (2006) Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Syst Evol* 259:89–120
- Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylog Evol* 29:417–434
- Amores A, Force A, Yan YL et al (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282:1711–1714
- Amores A, Suzuki T, Yan YL et al (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-finned fish. *Genome Res* 14:1–10
- Aury JM, Jaillon O, Duret L et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178
- Barker MS, Kane NC, Matvienko M et al (2008) Multiple paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol* 25:2445–2455
- Beissbarth T, Speed TP (2004) Gostat: find statistically overrepresented Gene Ontologies with a group of genes. *Bioinformatics* 20:1464–1465
- Beisswanger S, Stephan W (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. *Proc Natl Acad Sci U S A* 105:5447–5452
- Birchler JA (1981) The genetic basis of dosage compensation of alcohol dehydrogenase-1 in maize. *Genetics* 97:625–637
- Birchler JA, Newton KJ (1981) Modulation of protein levels in chromosomal dosage series in maize: the biochemical basis of aneuploid syndromes. *Genetics* 99:247–266
- Birchler JA, Bhadra U, Bhadra MP, Auger DL (2001) Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol* 234:275–288
- Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. *Plant Cell* 15:2236–2239
- Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: biological implications. *Trends Genet* 21:219–226
- Birchler JA, Yao H, Chudalayandi S (2007) Biological consequences of dosage dependent gene regulatory systems. *Biochimica et Biophysica Acta - Gene Structure and Expression* 1769:422–428
- Blanc G, Wolfe KH (2004a) Functional divergence of duplicated genes formed by polyploidy during arabidopsis evolution. *Plant Cell* 16:1679–1691
- Blanc G, Wolfe KH (2004b) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678
- Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res* 13:137–144
- Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y (2006) The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biology* 7(5), R43
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438
- Bowring SA, Grotzinger JP, Isachsen CE, Knoll AH, Pelechaty SM, Kolosov P (1993) Calibrating rates of early Cambrian evolution. *Science* 261:1293–1298
- Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS Biol* 3:1159–1166
- Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication. *Proc Natl Acad Sci U S A* 103:2730–2735
- Christoffels A, Koh EGL, Chia JM, Brenner S, Aparicio S, Venkatesh B (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* 21:1146–1151
- Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287
- Clark AG (1994) Invasion and maintenance of a gene duplication. *Proc Natl Acad Sci U S A* 91:2950–2954
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353:31–37
- Conant GC, Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Molecular Systems Biology* 3. art. no 129
- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* 9:938–950
- Cui L, Wall PK, Leebens-Mack JH et al (2006) Widespread genome duplications throughout the history of flowering plants. *Genome Res* 16:738–749
- Davis JC, Petrov DA (2005) Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet* 21:548–551
- Dopman EB, Hartl DL (2007) A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 104:19920–19925
- Duarte JM, Wall PK, Edger PP, et al. (2009) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis*, and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* In Review
- Ettwiller L, Veitia RA (2007) Protein coevolution and iso-expression in yeast macromolecular complexes. *Compar Funct Genom* 2007. art. no. 58721
- Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Natl Acad Sci* 106:5737–5742
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Force A, Cresko WA, Pickett FB, Proulx SR, Amemiya C, Lynch M (2005) The origin of subfunctions and modular gene regulation. *Genetics* 170:433–446
- Freeling M (2008) The evolutionary position of subfunctionalization, downgraded. *Genome Dyn* 4:25–40
- Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60:433–453



- Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res* 16:805–814
- Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. *Genome Res* 18:1924–1937
- 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* 19:3403–3417
- Gallardo MH, Kausel G, Jimenez A et al (2004) Whole-genome duplications in South American desert rodents (Octodontidae). *Biol J Linn Soc* 82:443–451
- Garcia-Fernandez J, Holland PWH (1994) Archetypal organization of the amphioxus Hox gene cluster. *Nature* 370:563–566
- Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat Genet* 29:482–486
- Gunjan A, Verreault A (2003) A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. *Cell* 115:537–549
- Guo M, Birchler JA (1994) Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science* 266:1999–2002
- Ha M, Kim ED, Chen ZJ (2009) Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc Natl Acad Sci U S A* 106:2295–2300
- Ha M, Li WH, Chen ZJ (2007) External factors accelerate expression divergence between duplicate genes. *Trends Genet* 23:162–166
- He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164
- Hernandez-Hernandez T, Martinez-Castilla LP, Alvarez-Buylla ER (2007) Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein-protein interaction domains after major gene duplication events. *Mol Biol Evol* 24:465–481
- Holland LZ, Albalat R, Azumi K et al (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 18:1100–1111
- Hu S, Dilcher DL, Jarzen DM, Taylor DW (2008) Early steps of angiosperm-pollinator coevolution. *Proc Natl Acad Sci U S A* 105:240–245
- Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society B: Biological Sciences* 256:119–124
- Hughes MK, Hughes AL (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol Biol Evol* 10:1360–1369
- Hurley IA, Mueller RL, Dunn KA et al (2007) A new time-scale for ray-finned fish evolution. *Proceedings of the Royal Society B: Biological Sciences* 274:489–498
- Innan H (2003) A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. *Proc Natl Acad Sci U S A* 100:8793–8798
- Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467
- Jansen R, Greenbaum D, Gerstein M (2002) Relating whole-genome expression data with protein-protein interactions. *Genome Res* 12:37–46
- Kasahara M (2007) The 2R hypothesis: an update. *Curr Opin Immunol* 19:547–552
- Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624
- King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science* 188:107–116
- Knoll AH, Carroll SB (1999) Early animal evolution: emerging views from comparative biology and geology. *Science* 284:2129–2137
- Kondrashov FA, Kondrashov AS (2006) Role of selection in fixation of gene duplications. *J Theor Biol* 239:141–151
- Le Comber SC, Smith C (2004) Polyploidy in fishes: patterns and processes. *Biol J Linn Soc* 82:431–442
- Lee JM, Sonnhammer ELL (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13:875–882
- Leister D (2003) Chloroplast research in the genomic age. *Trends Genet* 19:47–56
- Lewis EB (1951) Pseudoallelism and gene evolution. *Cold Spring Harbor Symp Quant Biol* 16:159–174
- Li L, Huang Y, Xia X, Sun Z (2006) Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol* 23:2467–2473
- Lukens LN, Quijada PA, Udall J, Pires JC, Schranz ME, Osborn TC (2004) Genome redundancy and plasticity within ancient and recent Brassica crop species. *Biol J Linn Soc* 82:665–674
- Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 16:1–19
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473
- Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62:2131–2154
- Lyons E, Pedersen B, Kane J et al (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol* 148:1772–1781
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. *Genome Res* 15:516–525
- Mable BK (2004) ‘Why polyploidy is rarer in animals than in plants’: myths and mechanisms. *Biol J Linn Soc* 82:453–466
- Maere S, DeBodt S, Raes J et al (2005) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A* 102:5454–5459
- Marhold K, Lihova J (2006) Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. *Plant Syst Evol* 259:143–174
- Martin N, Ruedi EA, LeDuc R, Sun FJ, Caetano-Anoll√s G (2007) Gene-interleaving patterns of synteny in the *Saccharomyces cerevisiae* genome: are they proof of an ancient genome duplication event? *Biology Direct* 2

- Meyer A, Scharl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* 11:699–704
- Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949–956
- Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Møller IM (2005) The plant mitochondrial proteome. *Trends Plant Sci* 10:36–43
- Ming R, Hou S, Feng Y et al (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996
- Mollier P, Hoffmann B, Debast C, Small I (2002) The gene encoding *Arabidopsis thaliana* mitochondrial ribosomal protein S13 is a recent duplication of the gene encoding plastid S13. *Curr Genet* 40:405–409
- Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res* 14:820–828
- Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York
- Osborn TC, Chris Pires J, Birchler JA et al (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends Genet* 19:141–147
- Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* 131:452–462
- Papp B, Paul C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424:194–197
- Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
- Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet* 22:597–602
- Pignatta D, Comai L (2009) Parental squabbles and genome expression: lessons from the polyploids. *J Biol* 8
- Pires JC, Zhao J, Schranz ME et al (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*). *Biol J Linn Soc* 82: 675–688
- Putnam NH, Butts T, Ferrier DEK et al (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071
- Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. *BMC Biology* 7:18. art. no. 18
- Rijkema AS, Royaert S, Zethof J, Van Der Weerden G, Gerats T, Vandenbussche M (2006) Analysis of the *Petunia* TM6 MADS box gene reveals functional divergence within the DEF/AP3 lineage. *Plant Cell* 18:1819–1832
- Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Computational Biology* 2:0989–1000
- Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345
- Scannell DR, Frank AC, Conant GC, Byrne KP, Woolfit M, Wolfe KH (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Natl Acad Sci U S A* 104:8397–8402
- Schranz ME, Mitchell-Olds T (2006) Independent ancient polyploidy events in the sister families *Brassicaceae* and *Cleomaceae*. *Plant Cell* 18:1152–1165
- Semon M, Wolfe KH (2007) Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor. *Trends Genet* 23:108–112
- Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc Natl Acad Sci U S A* 105:8333–8338
- Seoighe C, Wolfe KH (1999) Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* 2:548–554
- Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* 6:715–722
- Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 99:13627–13632
- Soltis PS, Soltis DE (2004) The origin and diversification of angiosperms. *Am J Bot* 91:1614–1626
- Soltis DE, Burleigh JG (2009) Surviving the K-T mass extinction: new perspectives of polyploidization in angiosperms. *Proc Natl Acad Sci* 106:5455–5456
- Soltis DE, Albert VA, Leebens-Mack J et al (2009) Polyploidy and angiosperm diversification. *Am J Bot* 96:333–348
- Soltis DE, Chanderbali AS, Kim S, Buzgo M, Soltis PS (2007a) The ABC model and its applicability to basal angiosperms. *Ann Bot* 100:155–163
- Soltis DE, Ma H, Frohlich MW et al (2007b) The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends Plant Sci* 12:358–367
- Teichmann SA, Veitia RA (2004) Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics* 167:2121–2125
- 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* 16:934–946
- Tuskan GA, DiFazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604
- Van de Peer Y (2004) *Tetraodon* genome confirms Takifugu findings: most fish are ancient polyploids. *Genome Biology* 5(12), art. no. 250
- Veitia RA (2002) Exploring the etiology of haploinsufficiency. *BioEssays* 24:175–184
- Veitia RA (2003a) A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol Rev Camb Philos Soc* 78:149–170
- Veitia RA (2003b) Nonlinear effects in macromolecular assembly and dosage sensitivity. *J Theor Biol* 220:19–25
- Veitia RA (2005) Gene dosage balance: deletions, duplications and dominance. *Trends Genet* 21:33–35
- Veitia RA, Bottani S, Birchler JA (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet* 24:390–397

- Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117
- Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24:437–440
- Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713
- Woods IG, Kelly PD, Chu F et al (2000) A comparative map of the zebrafish genome. *Genome Res* 10:1903–1914
- Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206–216
- Yu J, Wang J, Lin W et al (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* 3:0266–0281
- Yu X, Lin J, Zack DJ, Qian J (2006) Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. *Nucleic Acids Res* 34:4925–4936
- Zahn LM, Kong H, Leebens-Mack JH et al (2005) The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169: 2209–2223