

RESEARCH ARTICLES

Two-Phase Resolution of Polyploidy in the *Arabidopsis* Metabolic Network Gives Rise to Relative and Absolute Dosage Constraints^W

Michaël Bekaert,^{a,1} Patrick P. Edger,^b J. Chris Pires,^{b,c} and Gavin C. Conant^{a,c}

^a Division of Animal Sciences, University of Missouri, Columbia, Missouri 65211

^b Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211

^c Informatics Institute, University of Missouri, Columbia, Missouri 65211

The abundance of detected ancient polyploids in extant genomes raises questions regarding evolution after whole-genome duplication (WGD). For instance, what rules govern the preservation or loss of the duplicated genes created by WGD? We explore this question by contrasting two possible preservation forces: selection on relative and absolute gene dosages. Constraints on the relative dosages of central network genes represent an important force for maintaining duplicates (the dosage balance hypothesis). However, preservation may also result from selection on the absolute abundance of certain gene products. The metabolic network of the model plant *Arabidopsis thaliana* is a powerful system for comparing these hypotheses. We analyzed the surviving WGD-produced duplicate genes in this network, finding evidence that the surviving duplicates from the most recent WGD (WGD- α) are clustered in the network, as predicted by the dosage balance hypothesis. A flux balance analysis suggests an association between the survival of duplicates from a more ancient WGD (WGD- β) and reactions with high metabolic flux. We argue for an interplay of relative and absolute dosage constraints, such that the relative constraints imposed by the recent WGD are still being resolved by evolution, while they have been essentially fully resolved for the ancient event.

INTRODUCTION

Duplication of genetic material is an engine of evolutionary innovation (Taylor and Raes, 2004). In particular, Ohno (1970) argued for an important role for whole-genome duplications (WGD) in the shaping of the genetic and morphological complexity of multicellular eukaryotes. While such associations are of considerable interest, it is quite difficult to associate changes in rates of diversification or complexity with particular WGD events (Sémon and Wolfe, 2007). This difficulty is compounded by the fact that most duplicate genes produced by WGD are quickly lost (Scannell et al., 2007). As a result, understanding the interplay of duplicate gene loss and adaptation after WGD is an important open problem in evolutionary biology.

The model plant *Arabidopsis thaliana* (thale cress) is an attractive system for studying this question because, despite its small size (157 Mb), the *Arabidopsis* genome includes the remnants of at least three ancient WGDs (Vision et al., 2000; Simillion et al., 2002; Blanc et al., 2003; Bowers et al., 2003; Maere et al., 2005; Barker et al., 2009). These WGDs (Bowers et al., 2003), termed α (the most recent event, with 3947 retained duplicate genes in the

genome), β (an intermediate event in time with 2765 retained duplicates), and γ (the most ancient event, 771 retained duplicates), have been validated by both gene synteny and comparative duplicate age (K_c) analyses (Maere et al., 2005). While the γ event is quite widely distributed, having occurred near the origin of all eudicots (Soltis et al., 2009), the two more recent events (α and β) are known to be restricted to the order Brassicales, as papaya (*Carica papaya*), a member of this order, possesses neither (Soltis et al., 2009). Chronologically, α and β are estimated to have occurred roughly 47 to 65 and 65 to 115 million years ago, respectively (Beilstein et al., 2010). Here, we explore the patterns of duplicate preservation after WGD to gain a better appreciation of the evolutionary role of such duplications. That their influence is important is suggested by several observations. For instance, recent studies have demonstrated that the two recent WGD events gave rise to novel pathways for the synthesis of indole and Met-derived glucosinolates (see Kliebenstein, 2008; Schranz et al., 2011, and references therein) and may have increased species diversification rates (Beilstein et al., 2010). Likewise, it has been argued that repeated WGD events have contributed to the great diversity of angiosperm lineages (De Bodt et al., 2005; Soltis et al., 2008; Fawcett et al., 2009).

While the loss of duplicated genes after polyploidy may be rapid (Scannell et al., 2007), it is also distinctly nonrandom. Similar types of genes (kinases, transcription factors, and ribosomal proteins, among others) have been retained in duplicate after independent WGD events in ancestors of modern *Arabidopsis*, rice (*Oryza*

¹ Address correspondence to bekaertm@missouri.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Michaël Bekaert (bekaertm@missouri.edu).

^WOnline version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.110.081281

sativa), paramecium (*Paramecium tetraurelia*), bakers' yeast (*Saccharomyces cerevisiae*), and maize (*Zea mays*) (Seoighe and Wolfe, 1999; Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Tian et al., 2005; Aury et al., 2006; Conant and Wolfe, 2008; Schnable et al., 2009; Wu and Qi, 2010). In addition, other genetic features have been shown to associate with increased frequency of duplicate survival after WGD: they include greater numbers of protein interactions (Guan et al., 2007; Hakes et al., 2007), higher levels of gene expression (Seoighe and Wolfe, 1999), and greater numbers of phosphorylation sites (Amoutzias et al., 2010). Collectively, these trends suggest a pattern of anticorrelation in the functional classes of duplicate genes created by the two major types of gene duplications: those from small-scale duplications (SSDs) and those created by WGD (Maere et al., 2005; Dopman and Hartl, 2007; Wapinski et al., 2007). In contrast with WGD duplicates, SSD-produced duplicate genes tend to have fewer protein interactions and not to have core regulatory functions (Maere et al., 2005; Freeling, 2009). Given this strong differentiation between the two duplication mechanisms, there is a need for theories explaining the rules by which duplicate genes survive, especially following WGD.

In fact, the study of duplicate genes is a flourishing field in molecular evolution, and numerous mechanisms of preservation have been proposed (Innan and Kondrashov, 2010). Two of the most important are neofunctionalization (Hughes, 1994), whereby one of the two duplicate copies acquires a new, selectively beneficial function, and subfunctionalization, whereby the functions of ancestrally multifunctioned genes are subdivided between duplicated copies by a variety of mechanisms (Force et al., 1999; Stoltzfus, 1999; Des Marais and Rausher, 2008; Innan and Kondrashov, 2010). As powerful as these explanations are for exploring SSD events, they are less useful for understanding WGD, both because we lack the detailed historical and functional knowledge to employ them at the genome scale and because they do not seem to account for the kinds of functional biases observed in WGD-produced duplicates (Freeling, 2009).

Instead, at least two (likely complementary) hypotheses have been proposed to explain which genes survive in duplicate after WGD. The first is the dosage balance hypothesis (DBH) also known as the gene balance hypothesis (Papp et al., 2003; Freeling and Thomas, 2006; Birchler and Veitia, 2007). It states that, in eukaryotes, there is selection operating to disfavor duplications of central network genes due to the imbalance in network stoichiometry that results (Maere et al., 2005; Conant and Wolfe, 2008; Edger and Pires, 2009). This situation is reversed in the context of the duplication of the entire genome, since in that case it would be the loss of a second copy of a particular gene that would introduce imbalances relative to the remaining, duplicated, genes. Genes such as ribosomal proteins, kinases, and transcription factors that have pervasive and extensive functional interactions with other biomolecules are precisely the classes of genes that the DBH predicts to be maintained in duplicate post-WGD, and empirical evidence indicates they are dosage sensitive (Birchler et al., 2001). Likewise, the observation that genes coding for proteins of high interaction degree are overretained is in agreement with the DBH. We refer to these retention patterns as relative dosage constraints to distinguish them from cases of absolute dosage constraint, the

second proposed explanation for duplicate gene overretention post-WGD.

An absolute dosage constraint implies that an absolute increase in the concentration of a gene product is beneficial. Such selection has been observed for duplicate genes from a variety of organisms (Kondrashov and Kondrashov, 2006), including plants (van Hoof et al., 2001; Widholm et al., 2001). At first blush, it might appear that this type of selection would be agnostic as to the type of duplication employed. However, as Kacser and Burns (1981) demonstrated in their classic article, flux through a metabolic pathway is unlikely to be limited by the availability of a single enzyme, meaning that no single gene duplication is likely to dramatically increase flux. We have therefore argued that when a dosage increase is favored not for a single reaction but an entire pathway (such as glycolysis in yeast), WGD is a more rapid adaptation than is awaiting the necessary sequence of single gene duplications (Conant and Wolfe, 2007). Our observation is in keeping with a number of analyses that have shown an association between the yeast WGD and the propensity to ferment glucose even in the presence of oxygen (Blank et al., 2005; Piškur et al., 2006; Chen et al., 2008). In an elegant computational evolution experiment, van Hoek and Hogeweg (2009) found that overretention of certain enzymes was a recurrent feature of post-WGD evolution under (simulated) selection for rapid growth in a defined environment. This result suggests an interesting answer to Gould's famous question of the results of replaying the tape of life (Gould, 1989): at least in certain environments, both phenotypic and genotype patterns can recur.

Here, we examine the above two hypotheses of relative dosage and absolute dosage in the context of the patterns of duplicated metabolic gene retention after two of the nested WGD events (α and β) from *Arabidopsis*. In addition, we consider the question of how repeatable post-WGD genome evolution actually is. Maere et al. (2005) have already shown that categories of duplicate genes surviving after the α and β differ in their functional distributions, and our results suggest that such differences are seen even within metabolism. On the other hand, Seoighe and Gehring (2004) found that, over the genome as a whole, genes duplicated in one WGD were also likely to survive in duplicate after a second. One can imagine differing hypotheses in this regard. van Hoek and Hogeweg's results could imply a situation where repeated polyploidy gives rise to increasingly large but similarly constituted gene families. A similar pattern might be observed if the primary selection acting on WGD is internal, as implied by the DBH. In both cases, one argues that the selective environment post-WGD is relatively constant, meaning that one expects the same retention patterns. Differences would be expected if instead each polyploidy is a unique adaptation to a local environment.

Using the primary metabolic network from *Arabidopsis* (de Oliveira Dal'Molin et al., 2010), we explored whether the WGD-produced α and β duplicates tend to cluster together in the metabolic network and whether genes associated with high flux metabolic reactions were more likely to remain duplicated. Our work follows others in finding a general, but not universal, trend for metabolic genes to survive in duplicate post-WGD at higher frequency than genes in the genome at large (Maere et al., 2005; Kliebenstein, 2008). We argue that natural selection immediately

following WGD operates to maintain relative dosage balance in the metabolic network (i.e., clustering of duplicates in the network for the WGD- α duplicates), but, for the older WGD- β duplicates, selection instead operates on absolute gene product abundance. Thus, our results suggest that all polyploids may not be created equal.

RESULTS

Network Construction and Validation

We employed a published *Arabidopsis* primary metabolic network (de Oliveira Dal'Molin et al., 2010) to explore the patterns of post-WGD evolution in duplicated metabolic genes. A metabolic network is an abstract representation of the relationships between the biochemical reactions an organism is capable of (which are generally enzyme catalyzed) and the metabolites used or produced by those reactions (Wagner and Fell, 2001). In this case, we represent each metabolic reaction as a node in the network (a circle in Figure 1). Two nodes (reactions) are connected by an edge if they share a metabolite (lines in Figure 1). For *Arabidopsis*, two groups have produced models describing the metabolites and enzymes of primary metabolism (Poolman et al., 2009; de Oliveira Dal'Molin et al., 2010). The resulting networks include such pathways as photosynthesis, core anabolic and catabolic processes, and bulk biomass synthesis. Generally absent are secondary metabolic processes such as glucosinolate synthesis. The resulting network from de Oliveira Dal'Molin and colleagues consists of 1709 actual metabolites and 1550 metabolic reactions. The full network has 161,309 edges and the following network statistics: diameter, 11; average shortest path, 2.50; and density, 0.067 (Sabidussi, 1966; Coleman and More, 1983). There are 1398 genes annotated in the network, and 1168 reactions are associated with at least one gene. The metabolic network includes 334 reactions annotated with redundant genes (the redundant reactions in Supplemental Data Set 1 online): these are distinct reactions that share the same genes. Many of these genes are close paralogs that have undergone postduplication sub- or neofunctionalization (Table 1; Conant and Wolfe, 2008). For the purposes of this analysis, we thus employed a nonredundant 1217-reaction primary metabolic network created by merging nodes with the same genes and compartmental annotations but different metabolites (see Methods). This filtered network has 102,591 edges and the following network statistics: diameter, 8; average shortest path, 2.47; and density, 0.069.

Duplication Mapping

Arabidopsis duplicate genes were identified and attributed either to the (recent) SSDs (245 duplicates) of Rizzon et al. (2006) or to WGD, as identified by Bowers et al. (2003) and modified by Thomas et al. (2006). Duplications attributed to genome duplication were further subdivided as surviving from the early β duplication (156 duplicates) and/or from the subsequent α event (420 duplicates). We then individually mapped these distinct types of duplicate genes onto the reaction-centered metabolic network (Figure 1; see Supplemental Data Set 2 online).

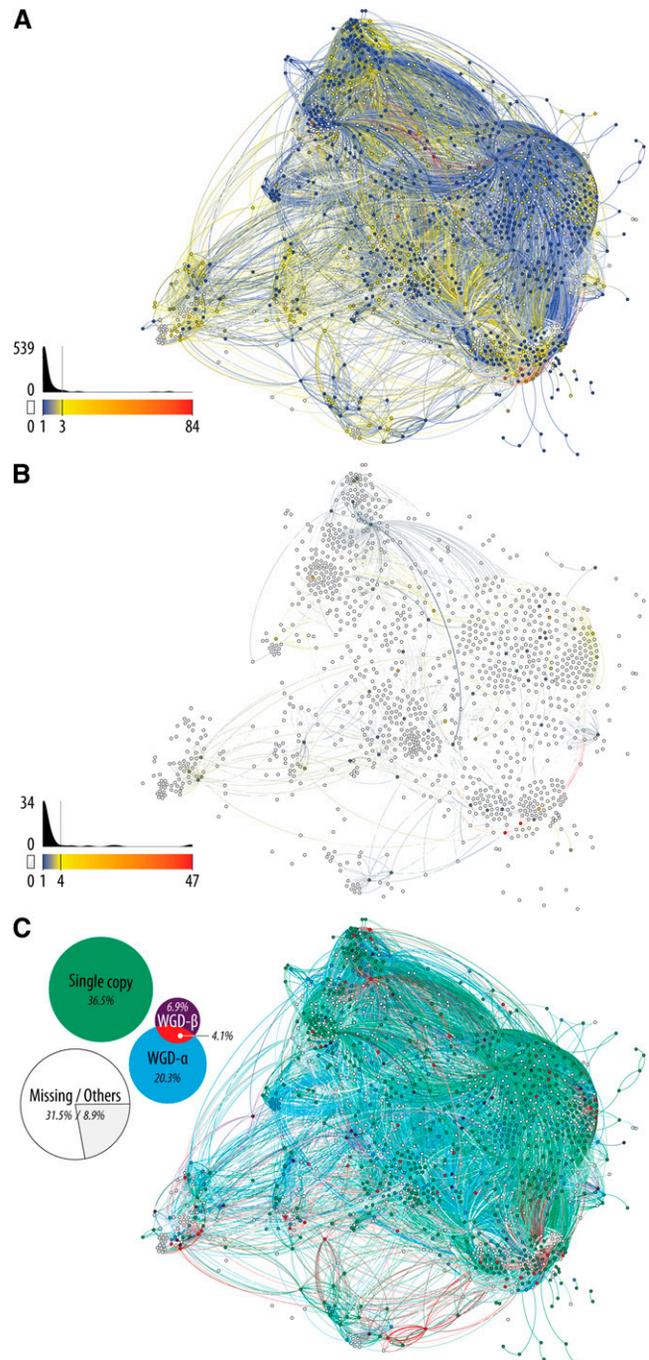


Figure 1. Gene Numbers and Duplication Status in the *Arabidopsis* Primary Metabolic Network.

(A) and (B) For each pair of reactions (nodes), an edge is created if the two reactions share a metabolite. The number of associated enzyme genes (A) or SSD events (B) is indicated for each reaction by a color gradient. The observed mean value was set to the gradient mid color to highlight the actual distribution, shown at the top of the color gradient. (C) The WGD- α and WGD- β duplication events are shown separately. Nodes with at least one such event are reported (the detailed distribution is also illustrated in the associated pie charts). White nodes are nodes without associated genes (31.5% of the reactions).

Table 1. Selected List of Sub- and Neofunctionalized Duplicates

| Genes | Reactions | Event |
|------------------------|------------|---|
| At1g20630 | R_R00009_x | Temporal and spatial expression pattern differences |
| At4g35090 | R_R02670_c | |
| At3g10620 ^a | R_R00184_c | Redundant |
| At5g06340 ^a | R_R01232_c | |
| At2g34890 ^b | R_R00571_c | Temporal and spatial expression pattern differences |
| At3g12670 ^b | R_R00573_c | |
| At4g30910 ^a | R_R00899_c | Redundant |
| At4g30920 ^a | R_R04951_c | |
| At2g01290 | R_R01056_c | Distinct subcellular localization and |
| At3g04790 | R_R01056_p | Temporal and spatial expression pattern differences |
| At3g54420 | R_R01206_c | Temporal and spatial expression pattern differences |
| At5g24090 | R_R02334_c | |
| At5g41670 | R_R01528_c | Temporal and spatial expression pattern differences |
| At3g02360 | R_R01528_p | |
| At3g01850 | R_R01529_c | Temporal and spatial expression pattern differences |
| At5g61410 | R_R01529_p | |
| At5g22300 | R_R01887_c | Distinct molecular functions and |
| At3g44320 | R_R07855_c | Temporal and spatial expression pattern differences |
| At4g22330 ^a | R_R03540_c | Distinct molecular functions |
| At5g56650 ^a | R_R06940_c | |
| At1g71230 | R_R04313_c | Temporal and spatial expression pattern differences |
| At5g27430 | R_R04869_c | |

The analysis of gene annotation (functional, expression, and localization) was made using The Arabidopsis Information Resource (<http://www.Arabidopsis.org/>), the Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>), and the SUB-cellular location database for *Arabidopsis* (<http://suba.plantenergy.uwa.edu.au/>).

^aReactions associated with both duplicates.

^bExample highlighted Supplemental Figure 2B online.

Propensity for Duplicate Preservation after WGD Varies by Functional Role

We first evaluated the frequency of duplicate genes (of both SSD and WGD origins) in the metabolic network compared with the genome at large. There are more surviving duplicate genes from the metabolic network than would be expected: duplicates produced by SSD are observed 29% more often than expected (compared with the duplication level observed in the overall genome), while the corresponding figures for the α and β events are 30 and 12%, respectively (Fisher's exact test, $P < 0.001$). We were further curious whether duplication propensity was associated with the cellular compartment in which an enzyme acts, especially given that these compartments are specialized for differing types of pathways. Furthermore, two of the compartments (the mitochondria and the chloroplast) possess their own genomes whose ploidy level was unaltered by nuclear WGD events (Figure 2). Except for a deficit of duplicates annotated as transporters ($P < 0.003$), the proportion of genes retained following WGD does not differ significantly among the compartments. By contrast, the frequency of SSD was significantly different among the compartments, with the highest frequency of events seen in the nuclear-encoded fraction of the enzymes active in the mitochondria and chloroplast. No SSDs of transporters were detected. Similarly, Table 2 summarizes the frequency of duplicate retention after the WGD- α and WGD- β events for genes of differing functions, inferred using the Gene Ontology framework (Carbon et al., 2009). Retention rates for

genes from the metabolic network are high relative to the genome at large but are lower than that for transcription factors, a class of genes known to be highly retained post-WGD (Maere et al., 2005).

Duplicated Genes Form Clusters in the Metabolic Network

Figure 1 suggests that duplicated metabolic genes are clustered nonrandomly in the network. We used a novel cluster detection algorithm to assess this apparent pattern (see Methods; Bekaert and Conant, 2011). With this algorithm, we analyzed five distinct sets of duplicated genes: (1) SSD duplicates, (2 and 3) genomic loci with duplicates derived from the α and β events (WGD- α and WGD- β , respectively), (4) genomic loci with duplicates surviving from both events (WGD- $\alpha\cap\beta$), and (5) a pooled data set of loci with duplicates from either WGD event (WGD- $\alpha\cup\beta$). To analyze each set, we first removed from the network all nodes not possessing genes from that set. We then calculated the number of connected components (Watts and Strogatz, 1998) among the remaining n_d nodes with appropriate duplicates. We compared this number to the number of components obtained by retaining n_d nodes selected at random from the network (see Methods; Figure 3). For the sets WGD- α (Figure 3B) and WGD- $\alpha\cap\beta$, we found significantly fewer and larger clusters than expected (Figure 3D; see Supplemental Table 1 online). These two duplicate gene sets also have higher in and out degrees than would be expected (i.e., their products are more likely to be reactants in

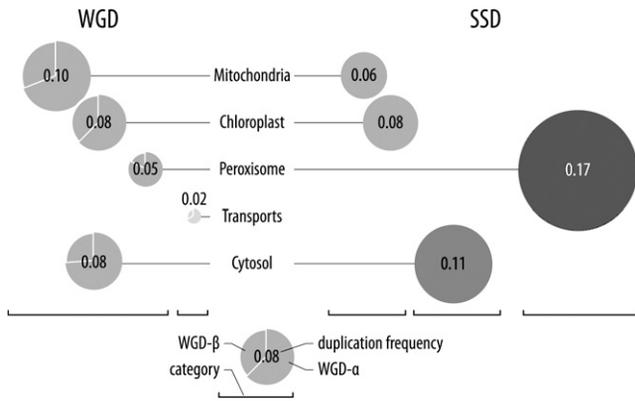


Figure 2. Duplication Frequencies by Cellular Compartment.

For each compartment, the proportion of reactions with duplicate genes surviving from SSD and WGD events is shown. The diameter and shade of gray of each pie is proportional to the frequency of duplicates (inner numbers). The proportion of surviving WGD-produced duplicates is not significantly different between compartments save for the transporters ($P < 0.003$). The cellular compartments fall into three distinct groups/categories with respect to the frequency of SSD events (bottom brackets): [Mitochondria, Chloroplast], [Cytosol], and [Peroxisome] ($P < 0.001$).

other reactions, while their reactants are produced by more other reactions than those of the average reaction). While the WGD- β set (Figure 3C) does not show a significant bias in clustering or node degree compared with the remainder of the network, this may be due to the small number (84 nodes total) of known β duplicates: the pooled set of α and β duplicates (WGD- $\alpha\cup\beta$) show strong biases. We see no tendency for the set of genes produced by SSD to cluster in the network (Figures 1B and 3D). Note that we see less clustering than expected among the reactions catalyzed by only a single gene (Figure 3D). This result is expected because our analysis is essentially zero sum: an excess of clustering among some reactions will be balanced by a deficit among other classes. This fact should be clear when one recognizes that our randomization approach compares clustering in a subset of the network to the average clustering in the whole network (given by the random subsets created for statistical analysis).

Enzymes Duplicated in Ancient Polyploidy Events Are Associated with High Flux

A natural question is how the flux an enzyme carries influences its chances of surviving in duplicate after WGD. To explore this question, we used flux balance analysis (Orth et al., 2010) to estimate the flux through every reaction in the *Arabidopsis* primary metabolic network under both photosynthetic and non-photosynthetic conditions (see Supplemental Figure 1 online). The published network does not allow the synthesis of either DNA/RNA nucleotides (ATP, dATP, etc.) or the amino acids Arg and His. However, we were able to model the biosynthesis of these molecules by adding a transport reaction that imports succinate in the mitochondria in exchange for fumarate (Catoni et al., 2003). This transporter is encoded by the known *Arabi-*

dopsis gene At5g01340 (Catoni et al., 2003). Using these flux values (covering 40% of all reactions; see Supplemental Figure 1 online) and the above five sets of duplicate genes (SSD, WGD- α , WGD- β , WGD- $\alpha\cap\beta$, and WGD- $\alpha\cup\beta$), we built logistic regression models to evaluate whether reaction flux is predictive of membership in a duplication set. Among these sets, high metabolic flux is strongly predictive of membership in WGD- β ($P \approx 0.002$; Figure 4); no other set shows a significant association between flux and duplication status.

DISCUSSION

The degree to which genome duplication is associated with functional or morphological innovation is a long-standing evolutionary question (Kliebenstein, 2009). Ohno (1970) was a strong proponent of such a relationship, but direct links between the two types of event are difficult to establish. Nonetheless, a number of authors have speculated that repeated polyploidy events, especially in plants, might have driven both speciation and increases in morphological complexity (Freeling and Thomas, 2006; Scannell et al., 2006; Edger and Pires, 2009; Wood et al., 2009). One of the few cases where the details of such links are available is in *S. cerevisiae*, where it has been argued that a WGD helped give rise to yeast's propensity to ferment glucose even in the presence of oxygen (Blank et al., 2005; Conant and Wolfe, 2007; Merico et al., 2007; Chen et al., 2008). Importantly, this adaption appears to have arisen through selection for high absolute dosage among metabolic genes, even though relative dosage constraints appear to have been active in other parts of the genome (Seoighe and Wolfe, 1999). Here, we tested the roles of relative dosage constraints and absolute dosage in *Arabidopsis* in influencing the fates of two classes of gene duplications: ancient polyploidy events (WGD- β and - α) and tandem duplications, with the prediction that they would differ (Edger and Pires, 2009; Freeling, 2009).

Is it possible that, as in yeast, absolute dosage selection was active at the time of the WGD- β in *Arabidopsis*? Figure 4 suggests that high flux reactions have been overretained duplicated after this event, and there should be a strong association between high flux and the need for multiple copies of the associated enzymes. However, no such association is detected for the more recent α duplication (Figure 4), suggesting, at a minimum, that selection on absolute dosage is not the dominant force in all polyploidy events. And of course, flux high should not be read as a synonym for gene essentiality or importance: some high flux reactions can be

Table 2. Retained Functions after WGD

| Gene Category | WGD- α Duplicates | WGD- β Duplicates |
|---------------------------------------|--------------------------|--------------------------|
| Metabolic genes (from the network) | 28.0% ($P < 10^{-15}$) | 9.8% ($P = 0.003$) |
| Catalytic activity (GO) | 28.2% ($P < 10^{-15}$) | 11.6% ($P = 10^{-14}$) |
| Transcription factor (GO) | 35.7% ($P < 10^{-15}$) | 12.8% ($P = 0.04$) |
| Unknown function (GO) | 12.7% ($P < 10^{-15}$) | 5.4% ($P < 10^{-15}$) |
| All genes | 20.4% | 8.7% |

Proportion of duplicates and P values (two-tailed Fisher's exact tests). GO, Gene Ontology.

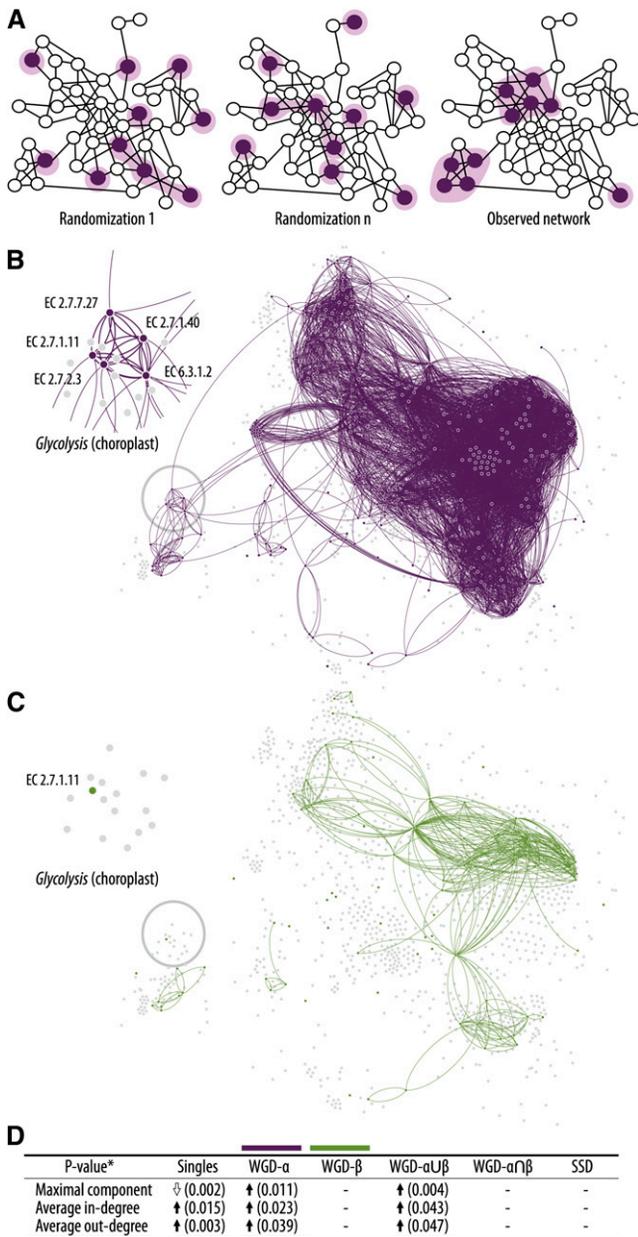


Figure 3. WGD- α Duplicate Genes Cluster in the Metabolic Network.

(A) Overview of the cluster detection method. The number and size of the connected components (shaded clusters in the figure) for the real network are calculated after nonduplicated nodes (white) are removed. These clusters are then compared with those seen in randomized networks with the same number of duplicated nodes (see Methods).

(B) The clusters observed for the 247 reactions with retained WGD- α duplicates (dark nodes). The inset shows a detail of the chloroplast glycolysis. EC 2.7.7.27, glucose-1-phosphate adenylyltransferase; EC 2.7.1.11, ATP:D-fructose-6-phosphate 1-phosphotransferase; EC 2.7.2.3, ATP:3-phospho-D-glycerate 1-phosphotransferase; EC 6.3.1.2, glutamate-ammonia ligase; EC 2.7.1.40, ATP:pyruvate 2-O-phosphotransferase.

(C) The clusters observed for the 84 reactions with retained WGD- β duplicates (dark nodes).

bypassed by alternative metabolic routes (Wagner, 2005), while low flux reactions may be essential. Alternatively, we suggest that both absolute and relative dosage constraints have acted during the post-WGD evolution of *Arabidopsis*. It is intuitive that, under the DBH, clusters of interacting metabolic genes would tend to either all remain duplicated or all return to single copy. Such relative dosage constraints are exactly the pattern we observe after the α event (e.g., the clustering in Figure 3). We suggest that the pattern of flux distribution among WGD- β duplicates can also be partly understood in terms of the DBH, despite the fact that little clustering is seen among these duplicates.

Because polyploidy imparts an immediate and relatively predictable increase in cell volume (Galitski et al., 1999; Leitch and Bennett, 2004; Veitia, 2005), it is reasonable to argue that, on average, mRNA and protein concentrations change relatively little after genome doubling. Thereafter though, the rapid and nonrandom loss of duplicated genes after WGD (Scannell et al., 2007) should alter the dosages of the new genome's genes relative to each other (Conant and Wolfe, 2007). We argue that selection on relative dosage balance is responsible for the immediate preservation of duplicates (consistent with the clustering results discussed above). One caveat to this argument is that it is at least possible that the lack of clustering among the single-enzyme reactions (Figure 3C) is the true evolutionary signal from our analysis and not the clustering among the duplicates (see Results). From this perspective, rather than selection to maintain duplicates with many interactions after WGD, there is actually selection to return certain enzymes to the single copy state (Paterson et al., 2006; Rong et al., 2010). We are skeptical of this argument, however, because it is unclear why such selection would be in force.

While the DBH is an attractive unifying theory, its predictive power is not expected to be absolute (Veitia et al., 2008). One reason is that much of the cellular regulatory apparatus that acts to buffer gene expression noise can also partly buffer ploidy changes (Stelling et al., 2004; Raser and O'Shea, 2005). It seems plausible therefore, that, over the longer term, changes in gene expression and regulation (Wray et al., 2003; Prud'homme et al., 2007; Wagner and Lynch, 2008) will release relative dosage constraints (Carroll, 2000; Lynch and Wagner, 2008; Veitia et al., 2008; Birchler and Veitia, 2010). However, at least for metabolic genes, because relative expression levels remain tuned to the current genome, there may be a set of high flux enzymes originally preserved by selection on relative dosage but for which absolute dosage selection is now acting. Absolute dosage selection will oppose the loss of a duplicate copy in cases where reduced gene expression interferes with the high enzyme copy number needed to maintain flux. Again, note that the history of WGD means that these dosage constraints are acting at the

(D) Results of the clustering tests for the five data sets described in Results. Significant P values (one-sided tests, $\alpha < 0.05$) are shown along with the direction of the test (higher or lower than the expectation). Thus, the WGD- α duplicates show larger clusters than seen in the randomized networks ($P < 0.05$). Comparisons indicated with a dash were not significantly different from randomized networks.

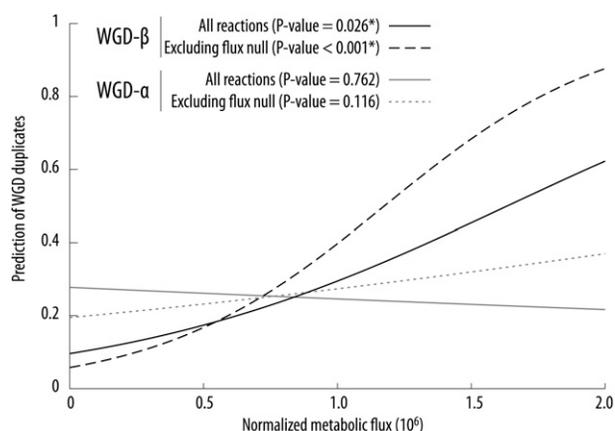


Figure 4. High Flux Reactions Are Enriched for Surviving WGD- β Duplicate Genes.

The association between metabolic flux (x axis) and the likelihood of a surviving WGD-produced duplication (y axis) as predicted by logistic regression (see Methods). Thus, high reaction flux is a predictor of the presence of a WGD- β , but not a WGD- α , duplicate. *Significant P values ($\alpha < 0.05$).

pathway and not the single gene level, meaning that we do not necessarily expect further duplication-based dosage amplifications. Examples of selection on gene copy number to achieve high enzyme activity are known from a number of species (Brown et al., 1998; Guillemaud et al., 1999; Price et al., 2004; Perry et al., 2007). Note that the argument is not that all genes behave in this dosage-dependent manner; merely that it is a general trend. This hypothesis of two types of dosage selection is attractive because it explains the difference in flux bias seen between the α and β duplicates: the α event is not yet fully resolved, meaning that enzymes other than those of high flux still survive due to selection on relative dosage balance, while the β duplicates are being acted on primarily by (late-arriving) selection on absolute dosage.

Once relative dosage constraints are lifted, duplicate genes have a number of alternate fates beyond absolute dosage selection. These fates include gene loss as well as subfunctionalization and neofunctionalization. We previously illustrated an example of post-WGD neofunctionalization (Kliebenstein, 2008; Schranz et al., 2011): the ability to synthesize two major classes of glucosinolates, which gives members of the order Brassicales their unique sharp taste, arose through the α and β WGDs. These compounds are synthesized from novel substrates by retained α or β duplicates, and the novelty is clear from the retained ancestral function in the other duplicate. Here, we identified several other cases of putative neo- and subfunctionalization post-WGD (Table 1). In these cases, functional genomics, gene expression and knockout, and protein localization data indicate that functional divergence has been layered on top of conserved enzyme structures and reaction mechanisms. In this view, the dosage balance hypothesis provides one absolutely key ingredient for functional change: time. By preserving duplicates, it allows a sufficient period for the necessary function-altering mutations to appear, meaning that the relaxation of relative dosage constraints represents the beginning, not the end, of the evolutionary impact of WGD.

An important caveat is that our study did not examine glucosinolates or any other potentially ecologically important secondary metabolic processes. Genome-scale metabolic models allow us to explore the potential for very interesting links between metabolism, genome duplication, and evolution. However, these models do not include most secondary metabolites. One might think this omission would tend to artificially lower certain fluxes in the network and hence mislead us when looking for links between metabolism and evolution. However, even the relatively abundant glucosinolates represent <2% of biomass for adult *Arabidopsis* plants (Brown et al., 2003). This proportion is an order of magnitude less than the fraction of protein by mass (Poolman et al., 2009). Since the highest flux reactions considered here are not even biomass synthesis reactions but those for energy production, it is unlikely that including the synthesis of secondary metabolites would greatly alter the numeric flux values considered here. Indeed, even the artificial doubling of flux through amino acid synthesis pathways has no discernable effect on the results presented in Figure 4 (see Supplemental Table 2 online). The more serious concern is that omitting such evolutionary labile processes gives a false sense of stability as to the role of WGD in metabolic evolution (Kliebenstein, 2008). It will be very interesting in future studies to collect this missing metabolic data to explore the potential for functional changes in secondary metabolism enzymes duplicated at WGD.

Here, we tried to integrate our understanding of the roles of relative dosage and absolute dosage in driving evolution after genome duplication. It is still an open question whether the WGD events in the ancestors of *Arabidopsis* conferred some immediate benefit that led to their fixation. What is clearer is that these events profoundly affected the subsequent evolution of this lineage. One effect that should not be underestimated is the role of WGD in relaxing the epistatic constraints all genomes acquire over their history. Subfunctionalization is one of the best illustrations of this principle: the divergence between the genes in Table 1 occurred post-WGD, and it is intriguing that such changes could take place even in this relatively limited window of divergence. Whether such relaxation could be a route to the increased phylogenetic and phenotypic diversity thought to characterize polyploid lineages is an exciting open question.

METHODS

Data Collection

The complete *Arabidopsis thaliana* primary metabolic network v1.0 was obtained from de Oliveira Dal'Molin et al. (2010). It includes a list of metabolites and their respective cellular compartments as well as the biochemical equation for each reaction. In cases where the enzyme catalyzing a reaction is known, the encoding gene is also noted. Each reaction is defined as a node in our reaction-centric network. Edges between these nodes are defined by shared metabolites between the reactions (see Supplemental Figure 2A online). The network is directed: for irreversible reactions if the product of one reaction is a reactant in the second, we define an edge. Reversible reactions are treated similarly, except that both directions of the reaction are allowed and handled independently. A specific compartment was created for the transporters (which we defined as reactions having metabolites in two compartments). The networks were visualized with Gephi v0.7 alpha (Bastian et al., 2009)

using the Force-based algorithm ForceAtlas. Our goal was to use this network to assign gene duplication events. To do so, we must account for the fact that several of the reactions are associated with the same set of genes (e.g., At1g04710, At2g33150, and At5g48880 all encode enzymes responsible for two different reactions: the conversion of 2-methylacetyl-CoA into acetyl-CoA and the conversion of 3- α 7- α dihydroxy-5- β 24-oxocholestanoyl-CoA into chenodeoxyglycocholoyl-CoA) (see Supplemental Figure 2B online). To account for these ambiguous annotations, we merged 334 such reactions (see Supplemental Figure 2B online) in order to not overestimate patterns of gene duplication and duplicate clustering. The result is the simplified metabolic network (with only 1217 nodes rather than 1550) that we use for mapping the gene duplications (see Supplemental Figure 2C online).

Identifying Enzyme Copy Numbers and Duplicate Gene Origins

The mean, median, and maximum number of genes associated with a reaction were 2.56, 1, and 84, respectively. For visualization purposes (Figure 1A), we use two color gradients in order to distinguish nodes with fewer (blue to yellow shades) or more (yellow to red shades) genes than the mean. To identify duplicate genes produced by SSD, we employed the data from Rizzon et al. (2006) and mapped the presence of SSDs (4043 events under a low stringency criteria with up to 10 spacer genes; Rizzon et al., 2006) onto the set of reactions (mean, median, and maximum number of duplicate genes were 4.30, 1, and 47, respectively). We plotted nodes with more or fewer SSD genes than the mean in Figure 1B. Duplicate genes produced by the α and β events were taken from the list given by Bowers et al. (2003) as modified by Thomas et al. (2006). We then mapped the presence of single copy genes or WGD- α or WGD- β duplicated genes onto the metabolic network (Figure 1C).

Retained Functions

We investigated the functional biases in the set of genes retained in duplicate following WGD. We collected the highest level of the Gene Ontology molecular function ontology (Ashburner et al., 2000). Genes might have more than one distinct function and therefore might be annotated with more than one Gene Ontology term. As a result, the total number of functional classifications is greater than the total number of genes. We calculated the significance of the enrichment with two-tailed Fisher's exact tests under the null hypothesis that there is no bias in the proportion of duplicate genes in the specified category.

Clustering Tests

We were interested in whether reactions with the same history of duplications cluster in the metabolic network. We thus calculated the number and maximal size of the network components containing nodes with each tested event (Figure 2C). To assess whether these components were bigger than would be expected, we used network randomization. We begin by copying the original network and reassigning the gene number or duplication status at random (Figure 3A). We then computed connected components for the random networks. We performed 10,000 permutations and used the distribution of component sizes to determine whether the clusters in the real network were larger than expected. The procedure was implemented in C++ using the Boost Libraries (<http://www.boost.org/>).

Flux Balance Analysis

We used the Systems Biology Research Tool v2.0.0 (Wright and Wagner, 2008) to perform flux balance analysis on the *Arabidopsis* primary metabolic network. Because the original network lacked the ability to synthesize nucleotides, DNA and RNA were not included in the biomass reaction (see Results). Thus, the nucleic acid requirements used here were calculated based on the weight of the each nucleotide and the

composition of the *Arabidopsis* genome. Fluxes computed with this tool were also verified with our own implementation of flux balance analysis. Using the network and the stoichiometry provided (de Oliveira Dal'Molin et al., 2010), we estimated the optimal biomass production rate under photosynthetic conditions (photon import allowed and sugar imports forbidden) and under nonphotosynthetic conditions (photon import forbidden but sugar imports allowed). In each case, we also made every possible reaction knockout whereby the reaction flux is constrained to 0, and the remainder of the network is reoptimized (see Supplemental Figure 1 online). For each such knockout analysis, all fluxes were first normalized by the value of the computed biomass flux for that analysis. For each reaction across all analyses, we then selected the maximum flux (i.e., across all possible conditions). We compared this maximal flux to the duplication status of that reaction node.

Logistic Regression

A logistic regression model (Sokal and Rohlf, 1995) was used to evaluate the relationship between flux and gene duplications. In this framework, reaction flux is used as a predictor of a binary duplication status variable (i.e., to infer whether a given reaction, drawn at random, possesses a duplicate given its flux value). We can use a likelihood ratio test to ask whether adding flux information significantly improves our ability to make such a prediction as opposed to simply using the overall duplication frequency as our predictor. The analysis was implemented in the package mLogit v0.18 in R.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Metabolic Flux Maps.

Supplemental Figure 2. Initial Metabolic Network and Processed Network after Merging Redundant Reactions.

Supplemental Table 1. Clustering Test Results.

Supplemental Table 2. Amino Acids Drain the Secondary Metabolism.

Supplemental Data Set 1. List of All the Case of Reactions Encoded by the Same Set of Genes.

Supplemental Data Set 2. Reactions' Duplication Status and Flux.

ACKNOWLEDGMENTS

We thank T. Arias, J. Birchler, C. Hudson, D. Mayfield, and J. Schultz for helpful discussions. This work was supported by a Research Board grant from the University of Missouri (M.B.), the Reproductive Biology Group of the Food for the 21st Century program at the University of Missouri (G.C.C.), and by the U.S. National Science Foundation (DBI 0501712 and DBI 0638536; P.P.E. and J.C.P.).

Received November 15, 2010; revised April 7, 2011; accepted April 15, 2011; published May 3, 2011.

REFERENCES

- Amoutzias, G.D., He, Y., Gordon, J., Mossialos, D., Oliver, S.G., and Van de Peer, Y. (2010). Posttranslational regulation impacts the fate of duplicated genes. *Proc. Natl. Acad. Sci. USA* **107**: 2967–2971.
- Ashburner, M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**: 25–29.

- Aury, J.M., et al.** (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171–178.
- Barker, M.S., Vogel, H., and Schranz, M.E.** (2009). Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol. Evol.* **1**: 391–399.
- Bastian, M., Heymann, S., and Jacomy, M.** (2009). Gephi: An open source software for exploring and manipulating networks. In Third International AAAI Conference on Weblogs and Social Media. (San Jose, CA: AAAI Publications), pp. 361–362.
- Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S.** (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**: 18724–18728.
- Bekaert, M., and Conant, G.C.** (2011). Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol. Biol. Evol.* **28**: 1111–1121.
- Birchler, J.A., Bhadra, U., Bhadra, M.P., and Auger, D.L.** (2001). Dosage-dependent gene regulation in multicellular eukaryotes: Implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* **234**: 275–288.
- Birchler, J.A., and Veitia, R.A.** (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* **19**: 395–402.
- Birchler, J.A., and Veitia, R.A.** (2010). The gene balance hypothesis: Implications for gene regulation, quantitative traits and evolution. *New Phytol.* **186**: 54–62.
- Blanc, G., Hokamp, K., and Wolfe, K.H.** (2003). A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* **13**: 137–144.
- Blanc, G., and Wolfe, K.H.** (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**: 1679–1691.
- Blank, L.M., Lehmebeck, F., and Sauer, U.** (2005). Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEM. Yeast Res.* **5**: 545–558.
- Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H.** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.
- Brown, C.J., Todd, K.M., and Rosenzweig, R.F.** (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* **15**: 931–942.
- Brown, P.D., Tokuhisa, J.G., Reichelt, M., and Gershenzon, J.** (2003). Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry* **62**: 471–481.
- Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S.; AmiGO Hub; Web Presence Working Group** (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics* **25**: 288–289.
- Carroll, S.B.** (2000). Endless forms: The evolution of gene regulation and morphological diversity. *Cell* **101**: 577–580.
- Catoni, E., Schwab, R., Hilpert, M., Desimone, M., Schwacke, R., Flügge, U.I., Schumacher, K., and Frommer, W.B.** (2003). Identification of an *Arabidopsis* mitochondrial succinate-fumarate translocator. *FEBS Lett.* **534**: 87–92.
- Chen, H., Xu, L., and Gu, Z.** (2008). Regulation dynamics of WGD genes during yeast metabolic oscillation. *Mol. Biol. Evol.* **25**: 2513–2516.
- Coleman, T.F., and More, J.J.** (1983). Estimation of sparse Jacobian matrices and graph coloring blems. *SIAM J. Numer. Anal.* **20**: 187–209.
- Conant, G.C., and Wolfe, K.H.** (2007). Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* **3**: 129.
- Conant, G.C., and Wolfe, K.H.** (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* **9**: 938–950.
- De Bodt, S., Maere, S., and Van de Peer, Y.** (2005). Genome duplication and the origin of angiosperms. *Trends Ecol. Evol. (Amst.)* **20**: 591–597.
- de Oliveira Dal'Molin, C.G., Quek, L.E., Palfreyman, R.W., Brumbley, S.M., and Nielsen, L.K.** (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol.* **152**: 579–589.
- Des Marais, D.L., and Rausher, M.D.** (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**: 762–765.
- Dopman, E.B., and Hartl, D.L.** (2007). A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **104**: 19920–19925.
- Edger, P.P., and Pires, J.C.** (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**: 699–717.
- Fawcett, J.A., Maere, S., and 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. USA* **106**: 5737–5742.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J.** (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Freeling, M.** (2009). Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**: 433–453.
- Freeling, M., and Thomas, B.C.** (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**: 805–814.
- Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S., and Fink, G.R.** (1999). Ploidy regulation of gene expression. *Science* **285**: 251–254.
- Gould, S.J.** (1989). *Wonderful Life: The Burgess Shale and the Nature of History.* (New York: W. W. Norton).
- Guan, Y., Dunham, M.J., and Troyanskaya, O.G.** (2007). Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175**: 933–943.
- Guillemaud, T., Raymond, M., Tsagkarakou, A., Bernard, C., Rochard, P., and Pasteur, N.** (1999). Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* **83**: 87–99.
- Hakes, L., Pinney, J.W., Lovell, S.C., Oliver, S.G., and Robertson, D.L.** (2007). All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol.* **8**: R209.
- Hughes, A.L.** (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **256**: 119–124.
- Innan, H., and Kondrashov, F.** (2010). The evolution of gene duplications: Classifying and distinguishing between models. *Nat. Rev. Genet.* **11**: 97–108.
- Kacser, H., and Burns, J.A.** (1981). The molecular basis of dominance. *Genetics* **97**: 639–666.
- Kliebenstein, D.J.** (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS ONE* **3**: e1838.
- Kliebenstein, D.J.** (2009). Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell* **21**: 1637–1646.
- Kondrashov, F.A., and Kondrashov, A.S.** (2006). Role of selection in fixation of gene duplications. *J. Theor. Biol.* **239**: 141–151.
- Leitch, I.J., and Bennett, M.D.** (2004). Genome downsizing in polyploid plants. *Biol. J. Linn. Soc. Lond.* **82**: 651–663.
- Lynch, V.J., and Wagner, G.P.** (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution* **62**: 2131–2154.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M.,**

- Kuiper, M., and Van de Peer, Y.** (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* **102**: 5454–5459.
- Merico, A., Sulo, P., Piškur, J., and Compagno, C.** (2007). Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS J.* **274**: 976–989.
- Ohno, S.** (1970). *Evolution by Gene Duplication*. (New York: Springer).
- Orth, J.D., Thiele, I., and Palsson, B.Ø.** (2010). What is flux balance analysis? *Nat. Biotechnol.* **28**: 245–248.
- Papp, B., Pál, C., and Hurst, L.D.** (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**: 194–197.
- Paterson, A.H., Chapman, B.A., Kissinger, J.C., Bowers, J.E., Feltus, F.A., and Estill, J.C.** (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.
- Perry, G.H., et al.** (2007). Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**: 1256–1260.
- Piškur, J., Rozpedowska, E., Polakova, S., Merico, A., and Compagno, C.** (2006). How did *Saccharomyces* evolve to become a good brewer? *Trends Genet.* **22**: 183–186.
- Poolman, M.G., Miguet, L., Sweetlove, L.J., and Fell, D.A.** (2009). A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiol.* **151**: 1570–1581.
- Price, R.N., Uhlemann, A.C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., Patel, R., Laing, K., Looareesuwan, S., White, N.J., Nosten, F., and Krishna, S.** (2004). Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet* **364**: 438–447.
- Prud'homme, B., Gompel, N., and Carroll, S.B.** (2007). Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. USA* **104** (suppl. 1): 8605–8612.
- Raser, J.M., and O'Shea, E.K.** (2005). Noise in gene expression: Origins, consequences, and control. *Science* **309**: 2010–2013.
- Rizzon, C., Ponger, L., and Gaut, B.S.** (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput. Biol.* **2**: e115.
- Rong, J., Feltus, F.A., Liu, L., Lin, L., and Paterson, A.H.** (2010). Gene copy number evolution during tetraploid cotton radiation. *Heredity* **105**: 463–472.
- Sabidussi, G.** (1966). The centrality of a graph. *Psychometrika* **31**: 581–603.
- Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S., and Wolfe, K.H.** (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**: 341–345.
- Scannell, D.R., Frank, A.C., Conant, G.C., Byrne, K.P., Woolfit, M., and Wolfe, K.H.** (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl. Acad. Sci. USA* **104**: 8397–8402.
- Schnable, P.S., et al.** (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science* **326**: 1112–1115.
- Schranz, M., Edger, P., Pires, J., van Dam, N., and Wheat, C.** (2011). Comparative genomics in the Brassicales: Ancient genome duplications, glucosinolate diversification and Pierinae herbivore radiation. In *Genetics, Genomics, and Breeding of Brassica Oilseeds*, I.P. David Edwards and J. Batley, eds (Enfield, NH: Science Publishers), pp. 206–218.
- Sémon, M., and Wolfe, K.H.** (2007). Consequences of genome duplication. *Curr. Opin. Genet. Dev.* **17**: 505–512.
- Seoighe, C., and Gehring, C.** (2004). Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* **20**: 461–464.
- Seoighe, C., and Wolfe, K.H.** (1999). Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**: 548–554.
- Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M., and Van de Peer, Y.** (2002). The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **99**: 13627–13632.
- Sokal, R.R., and Rohlf, F.J.** (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*. (New York: W.H. Freeman).
- Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P.K., and Soltis, P.S.** (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**: 336–348.
- Soltis, D.E., Bell, C.D., Kim, S., and Soltis, P.S.** (2008). Origin and early evolution of angiosperms. *Ann. N. Y. Acad. Sci.* **1133**: 3–25.
- Stelling, J., Sauer, U., Szallasi, Z., Doyle III, F.J., and Doyle, J.** (2004). Robustness of cellular functions. *Cell* **118**: 675–685.
- Stoltzfus, A.** (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**: 169–181.
- Taylor, J.S., and Raes, J.** (2004). Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**: 615–643.
- Thomas, B.C., Pedersen, B., and 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.
- Tian, C.G., Xiong, Y.Q., Liu, T.Y., Sun, S.H., Chen, L.B., and Chen, M.S.** (2005). Evidence for an ancient whole-genome duplication event in rice and other cereals. *Yi Chuan Xue Bao* **32**: 519–527.
- van Hoek, M.J., and Hogeweg, P.** (2009). Metabolic adaptation after whole genome duplication. *Mol. Biol. Evol.* **26**: 2441–2453.
- van Hoof, N.A., Hassinen, V.H., Hakvoort, H.W., Ballintijn, K.F., Schat, H., Verkleij, J.A., Ernst, W.H., Karenlampi, S.O., and Tervahauta, A.I.** (2001). Enhanced copper tolerance in *Silene vulgaris* (Moench) Garcke populations from copper mines is associated with increased transcript levels of a 2b-type metallothionein gene. *Plant Physiol.* **126**: 1519–1526.
- Veitia, R.A.** (2005). Paralogs in polyploids: One for all and all for one? *Plant Cell* **17**: 4–11.
- Veitia, R.A., Bottani, S., and Birchler, J.A.** (2008). Cellular reactions to gene dosage imbalance: Genomic, transcriptomic and proteomic effects. *Trends Genet.* **24**: 390–397.
- Vision, T.J., Brown, D.G., and Tanksley, S.D.** (2000). The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117.
- Wagner, A.** (2005). *Robustness and Evolvability in Living Systems*. (Princeton, NJ: Princeton University Press).
- Wagner, A., and Fell, D.A.** (2001). The small world inside large metabolic networks. *Proc. Biol. Sci.* **268**: 1803–1810.
- Wagner, G.P., and Lynch, V.J.** (2008). The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol. (Amst.)* **23**: 377–385.
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A.** (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**: 54–61.
- Watts, D.J., and Strogatz, S.H.** (1998). Collective dynamics of 'small-world' networks. *Nature* **393**: 440–442.
- Widholm, J.M., Chinnala, A.R., Ryu, J.H., Song, H.S., Eggett, T., and Brotherton, J.E.** (2001). Glyphosate selection of gene amplification in suspension cultures of 3 plant species. *Physiol. Plant.* **112**: 540–545.
- Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., and Rieseberg, L.H.** (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* **106**: 13875–13879.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A.** (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.
- Wright, J., and Wagner, A.** (2008). The Systems Biology Research Tool: Evolvable open-source software. *BMC Syst. Biol.* **2**: 55.
- Wu, X., and Qi, X.** (2010). Genes encoding hub and bottleneck enzymes of the *Arabidopsis* metabolic network preferentially retain homeologs through whole genome duplication. *BMC Evol. Biol.* **10**: 145.

Two-Phase Resolution of Polyploidy in the *Arabidopsis* Metabolic Network Gives Rise to Relative and Absolute Dosage Constraints

Michaël Bekaert, Patrick P. Edger, J. Chris Pires and Gavin C. Conant
Plant Cell 2011;23;1719-1728; originally published online May 3, 2011;
DOI 10.1105/tpc.110.081281

This information is current as of August 2, 2011

| | |
|---------------------------------|--|
| Supplemental Data | http://www.plantcell.org/content/suppl/2011/05/03/tpc.110.081281.DC1.html http://www.plantcell.org/content/suppl/2011/05/03/tpc.110.081281.DC2.html |
| References | This article cites 85 articles, 33 of which can be accessed free at: http://www.plantcell.org/content/23/5/1719.full.html#ref-list-1 |
| Permissions | https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X |
| eTOCs | Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain |
| CiteTrack Alerts | Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain |
| Subscription Information | Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: http://www.aspb.org/publications/subscriptions.cfm |