Escape from Preferential Retention following Repeated Whole Genome Duplications in Plants

James C. Schnable
*University of Nebraska-Lincoln*, schnable@unl.edu

Xiaowu Wang
*Chinese Academy of Agricultural Sciences*

J. Chris Pires
*University of Missouri*

Michael Freeling
*University of California - Berkeley*, freeling@berkeley.edu

Follow this and additional works at: [https://digitalcommons.unl.edu/agronomyfacpub](https://digitalcommons.unl.edu/agronomyfacpub)

Part of the [Agricultural Science Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Agriculture Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Agronomy and Crop Sciences Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Botany Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Genomics Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Horticulture Commons](https://digitalcommons.unl.edu/agronomyfacpub), [Other Plant Sciences Commons](https://digitalcommons.unl.edu/agronomyfacpub), and the [Plant Biology Commons](https://digitalcommons.unl.edu/agronomyfacpub)

Schnable, James C.; Wang, Xiaowu; Pires, J. Chris; and Freeling, Michael, "Escape from Preferential Retention following Repeated Whole Genome Duplications in Plants" (2012). *Agronomy & Horticulture -- Faculty Publications*. 1007.
[https://digitalcommons.unl.edu/agronomyfacpub/1007](https://digitalcommons.unl.edu/agronomyfacpub/1007)

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Escape from preferential retention following repeated whole genome duplications in plants

James C. Schnable¹, Xiaowu Wang², J. Chris Pires³ and Michael Freeling¹*

1 Freeling Lab, Plant and Microbial Biology, University of California – Berkeley, Berkeley, CA, USA
2 Molecular Genetics Lab, Biotechnology Department, Institute of vegetables and flowers, Chinese Academy of Agricultural Sciences, Beijing, China
3 Biological Sciences, Bond Life Sciences Center, University of Missouri, Colombia, MO, USA

EDITED BY:
Elena R. Alvarez-Buylla, Universidad Nacional Autonoma de Mexico, Mexico

REVIEWED BY:
Paula Casati, Centro de Estudios Fotosinteticos-CONICET, Argentina
Amy Louise Lawton-Rauh, Clemson University, USA

*Correspondence:
Michael Freeling, Freeling Lab, Plant and Microbial Biology, University of California – Berkeley, 111 Koshland Hall, PMB, Berkeley, CA 94720, USA.
e-mail: freeling@berkeley.edu

INTRODUCTION

Plants have been colorfully labeled the “big kahuna of polyploidy” (Sémon and Wolfe, 2007). The lineages leading to the two preeminent models for plant genetics – Arabidopsis (a eudicot) and maize (a monocot) – each show evidence of multiple independent whole genome duplications (Figure 1) since monocots and eudicots diverged approximately 120 million years ago (Soltis et al., 2009). Recent evidence suggests at least two additional, shared, whole genome duplications prior to the monocot/eudicot split (Jiao et al., 2011). The cumulative ploidy numbers relative to a pre-seed plant ancestor are listed in parentheses in Figure 1. Whole genome duplication creates duplicate, potentially redundant, copies of all the genes within a genome. The loss of these duplicate copies from the genomes of ancient polyploid species is known as fractionation (Langham et al., 2004) and – over evolutionary time scales – the majority of genes duplicated by polyploidy will be reduced back to a single copy. If fractionation did not occur, an ancestral genome of 10,000 genes would grow to an unrealistically large 640,000 genes in maize, and 1.44 million genes in Brassica rapa.

Some classes of genes, particularly those encoding organelle, preferentially revert to single copy status following whole genome duplications (Duarte et al., 2010). However, other classes of genes – such as subunits of large multiprotein complexes, transcription factors, and signal transduction machinery tend to resist fractionation following whole genome duplication (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Maere et al., 2005). This observation has been explained by the Gene Dosage Hypothesis (Birchler and Veitia, 2007) which predicts that fractionation of genes encoding proteins involved in dose–sensitive interactions will be selected against, as the loss of either gene copy is expected to throw the dosage of that gene pair’s product out of balance with its interaction partners, partners that also tend to remain duplicated. The topic of the influence of gene dosage-constraints on post-tetraploidy genome evolution has been well-reviewed (Sémon and Wolfe, 2007; Edger and Pires, 2009; Freeling, 2009; Birchler and Veitia, 2010). A previous study of multiple sequential tetraploidies in the Arabidopsis lineage found a general tendency for genes retained following one tetraploidy to also be retained following a second one (Seoighe and Gehring, 2004).

Since the divergence of the Arabidopsis and grape lineages, Arabidopsis has experienced two additional rounds of whole genome duplication. The rate of duplicate gene retention for transcription factors after single polyploidies have been observed to be approximately 25% (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004). If no mitigation of gene dosage occurred, our expectation after two rounds of whole genome duplication is that Arabidopsis should contain approximately 156% as many transcription factors encoding genes as grape. However, a detailed annotation of transcription factors using conserved protein domains found the number of transcription factors in the Arabidopsis genome is only 25.4% greater than the number found in grape
The fitness cost of changes in relative gene dosage must, to some extent, be mitigated over multiple whole genome duplications or the genomes of plants would long ago have become over-burdened with genes encoding life’s most complicated machines.

This paper provides evidence that duplicate genes do not equally maintain their progenitor’s preference for duplicate gene retention. Duplicate genes produced by whole genome duplication are not equivalent. Parental genomes originating from different species within a polyploid almost immediately differentiate into dominant and non-dominant subgenomes (Chang et al., 2010), and these expression differences are preserved for millions of years (Flagel and Wendel, 2010; Schnable et al., 2011a). Bias in gene loss between duplicate regions (fractation bias) has been observed in Arabidopsis (Thomas et al., 2006) and maize (Woodhouse et al., 2010) and seems to be a general rule for whole genome duplications ranging from paramecium to fish (Sankoff et al., 2010). Bias in fractionation and genome dominance are linked because it is expected that genes on the underexpressed, non-dominant subgenome simply matter less to purifying selection and dosage-constraints (Schnable et al., 2011a). In maize, genes with known mutant phenotypes are indeed preferentially found on the dominant subgenome (Schnable and Freeling, 2011). As bias in expression predicts which subgenome will experience more fractionation following polyploidy, either subgenome identity or the expression patterns of individual gene pairs may also predict which copy of a duplicate gene pair will be more prone to duplicate gene retention in future polyploidies.

We addressed the issue of mitigation of gene dosage-constraints with two experimental systems, the grasses, and the crucifers. Both clades have roughly parallel histories of polyploidy among species with sequenced genomes (Figure 1; Table 1). Both grasses and crucifers contain a more ancient whole genome duplication which is shared by all sequenced species in the clade (Bowers et al., 2003; Paterson et al., 2004) and in both clades one well studied species with a sequenced genome has experienced a second subsequent whole genome duplication – maize in the grasses (Gaut and Doebley, 1997) and B. rapa in the crucifers (Lysak et al., 2005). In both cases any duplicate genes retained from the older clade-wide
polyploidy did not retain additional duplicate copies in the subsequent lineage-specific polyploidy. Therefore we were able to carry out parallel experiments to identify characteristics associated with preferential retention. It was possible to control, to some extent for the effect of protein function, by focusing on pairs of duplicate genes retained in the clade-wide polyploidy which had different fates in the subsequent lineage-specific polyploidy. A model is proposed to explain how the duplicate copies of dose-sensitive genes escape preferential retention in later polyploidies.

**MATERIALS AND METHODS**

**DATA SOURCES**

The genome assemblies and annotation used in this study were TAIR 10 (*Arabidopsis thaliana*), *Arabidopsis lyrata* v1.0 (Hu et al., 2011), the initial release of the *B. rapa* genome (The *Brassica rapa* Genome Sequencing Project Consortium, 2011), MSU 6 (*Oryza sativa*; Goff et al., 2002), *Sorghum bicolor* 1.4 (Paterson et al., 2009), and B73_refgen1 (*Zea mays*; Schnable et al., 2009).

**GENE PAIR IDENTIFICATION**

Orthologous genes between *A. thaliana* and *A. lyrata* were identified using SynMap (Lyons et al., 2008) with QuotaAlign settings of 1:1 (Tang et al., 2011). *Arabidopsis–Brassica* orthologous relationships were taken from Tang et al. (2012). All orthologous and homoeologous relationships between grass species are those published in Schnable et al. (2012).

**EXPRESSION CALCULATIONS**

Gene expression levels were calculated using previously published RNA-seq data from wild type seedlings of *A. thaliana* (SRX019140: 44.7 million reads; Deng et al., 2010) and rice (SRX020118: 8.9 million reads; Zemach et al., 2010). These datasets were selected because, at the time these analysis were originally conducted they represented the RNA-seq experiments with the most sequencing depth for these two species deposited in the sequence read archive. Reads were aligned to reference genomes using Bowtie (Langmead et al., 2009) and gene expression levels were quantified using Cufflinks (Trapnell et al., 2010). Bowtie does not perform spliced alignments, which means some reads from regions of mRNA molecules which span exon junctions were not recovered in our analysis. However, given that homoeologous genes will in almost all cases possess the same intron–exon structure, any bias introduced by this approach will be equivalent between gene copies.

**MEASURING PURIFYING SELECTION**

Synonymous and non-synonymous substitution rates were calculated using the synonymous_calculation package included with bio-pipeline1 using the Nei–Gojobori method (Nei and Gojobori, 1986). All other settings remained as default.

**IDENTIFICATION OF RICE CNSs**

Rice CNSs were identified using version 3 of the CNS Discovery pipeline2 (Schnable et al., 2011b).

**STATISTICS**

*p*-Values for the difference in retention frequencies between singleton genes and homoeologously paired genes were calculated using Fisher’s Exact Test. In the crucifers, *Arabidopsis* genes with two or three retained co-orthologs in *B. rapa* were grouped together as ”retained.”

---

1 https://github.com/tanghaibao/bio-pipeline/
2 https://github.com/gturco/find_cns

---

**Table 1 | Whole genome duplications.**

| Footnote ID from Figure 1 | One name (often of many) | One citation (often of many) |
|---------------------------|-------------------------|-----------------------------|
| 1                         | Pre-seed plant          | Jiao et al. (2011)          |
| 2                         | Pre-flowering plant     | Jiao et al. (2011)          |
| 3                         | Sigma1                  | Tang et al. (2010)          |
| 4                         | Sigma2                  | Tang et al. (2010)          |
| 5                         | Pre-grass/Rho           | Paterson et al. (2004)      |
| 6                         | Maize Lineage WGD      | Gaut and Doebley (1997)     |
| 7                         | Gamma/pre-eudicot       | Jaillon et al. (2007)       |
|                           | hexaploidy              |                             |
| 8                         | Beta                    | Bowers et al. (2003)        |
| 9                         | Alpha                   | Bowers et al. (2003)        |
| 10                        | Brassica hexaploidy     | Lysak et al. (2005)         |

**FIGURE 2 | Impact of retention in a previous whole genome duplication on retention in future whole genome duplications. (A) Proportion of *Arabidopsis* genes with one, two, or three co-orthologs in *Brassica rapa*. (B) Proportion of genes syntenically retained in sorghum and rice with one, or two co-orthologs in maize.**
RESULTS

Genes syntenically conserved through the crucifers or grasses were categorized as (1) those without a homeologous duplicate from the older polyploidy in each lineage (2) those with a retained homeolog from the older polyploidy in each lineage. In the crucifer lineage, the older tetraploidy is Arabidopsis lineage alpha (23–40 MYA); in the Poales, the earlier tetraploidy was “pre-grass” (about 70 MYA; Figure 1). In crucifers, these genes are classified by the number of co-orthologs conserved in B. rapa after the hexaploidy shared by all Brassica species (Figure 2A). In grasses, genes were classified by whether maize retained only one or both co-orthologs following the more recent tetraploidy of the Zea/Tripsacum lineage (Figure 2B). Retention in older polyploidies does predict retention in future polyploidies ($p < 2.2 \times 10^{-16}$ for both crucifers and grasses), as previously showing in Arabidopsis (Seoighe and Gehring, 2004). However in both experiments approximately half of genes previously retained as a duplicate pair in the older whole genome duplication – and therefore presumed to be sensitive to changes in gene dosage – fractionated to a single copy in the more recent whole genome duplication.

The crucifer dataset consisted of 817 Arabidopsis gene pairs where one copy was orthologous to only a single gene in B. rapa and the other possessed either two or three co-orthologs (Data Sheet S1 in Material). The grass dataset consisted of 407 gene pairs conserved in both rice and sorghum where one copy was orthologous to only a single gene in maize, its duplicate having been fractionated and the other represented by two co-orthologs in maize (Data Sheet S2 in Supplementary Material). Gene pairs result from more ancient whole genome duplications were identified and removed, as these tend to introduce confounding factors. Members of gene pairs were assigned to under and over fractionated subgenomes using differences in the number of genes syntenically retained in multiple species between homoeologous regions of the rice and Arabidopsis genomes (Schnable et al., 2011a, 2012). In both datasets, the analysis of the relative levels of RNA encoded by duplicate genes pairs – measured by RNA-seq – was carried out in an outgroup lineage which shared only the older clade-wide polyploidy. In the grasses we used the expression of syntenic orthologs in rice and in the crucifers syntenic orthologs in A. thaliana (see Materials and Methods). The relative levels of purifying selection acting on each members of a gene pair were also compared using the ratio of non-synonymous substitutions to synonymous substitutions between orthologous genes in A. thaliana and A. lyrata (for the crucifiers) and between rice and sorghum (for the grasses; see Materials and Methods). Promoter complexity, as measured by number of conserved non-coding

![Graph A](image1.png) ![Graph B](image2.png)

**FIGURE 3** | Correlation between subsequent duplicate gene retention and a number of predicting factors including gene expression, ratio of non-synonymous to synonymous substitutions, and subgenome identity for (A) crucifer and (B) grass gene pairs. $P$-values relative to a 50/50 binomial distribution.
sequences, has previously shown to influence the odds a gene will be retained as a duplicate pair following polyploidy in the grasses (Schnable et al., 2011b) – so gene pairs were also sorted based on number of conserved non-coding sequences, in the grasses, and total quantity of upstream non-transposon sequence in Arabidopsis, this length being a crude proxy for promoter complexity having previously been shown to correlate with complexity of gene expression patterns (Sun et al., 2010).

All four potential markers examined showed significant power to predict which copy of a homeologous gene pair would be more resistant to fractionation in subsequent whole genome duplications (Figure 3). In general the gene copy retained in duplicate tended to also be the higher expressed copy, show evidence of greater purifying selection and to be associated with greater amounts of non-coding regulatory sequence. These genes also tended to be located on the dominant subgenome.

**DISCUSSION**

Following polyploidy, a genome possesses two or more homeologous genes, each with the same coding sequence and regulatory elements. Yet these gene copies can immediately show very different patterns of expression (Flagel et al., 2008; Buggs et al., 2011). It has been proposed that the deletion of less expressed copy of a gene following polyploidy is more likely to be selectively neutral (Schnable and Freeling, 2011; Schnable et al., 2011a). When combined with the observation that expression levels are unequal between parental subgenomes in allotetraploids (Chang et al., 2010; Flagel and Wendel, 2010; Schnable et al., 2011a), this model may explain the bias fractionation bias which has been found in ancient polyploids species (Schnable et al., 2011a).

Here we have shown that that the dominant gene copy – more expressed, under higher purifying selection, associated with more regulatory sequence – of a homeologous gene pair is more likely to retain the ancestral characteristic of preferential retention of duplicate copies in subsequent polyploids. A number of explanations could be proposed for the link between expression and future resistance to fractionation. We propose a model based on the same link between expression and which predicts fractionation bias between parental subgenomes. If all the co-orthologs of a single ancestral gene contribute to a single pool of gene-product, the loss of less expressed gene copies would result in the smallest change in total gene-product dosage. If the total expression of a group of homeologous genes is constrained in either relative or absolute terms (Bekaert et al., 2011) smaller changes in total gene-product dosage – created by the loss of a less expressed gene copy – are predicted to be more often selectively neutral, and therefore more common (Figure 4). This model also predicts that, for gene pairs in Arabidopsis where only one copy possesses any orthologous genes in Brassica, it should more often be the more expressed copy; as is indeed the case (Table A1 in Appendix).

When combined with previous results linking genome dominance with biased fractionation (Chang et al., 2010; Schnable et al., 2011a), our results suggest the Gene Dosage Hypothesis could perhaps be better thought of as the Gene-Product Dosage Hypothesis in that it can generally be considered to act on the concentration of the proteins encoded by duplicate genes, not gene copy number itself. Even when both copies of a gene are retained following whole genome duplication, the less expressed copy will often be lost in subsequent whole genome duplications. Furthermore, the greater the number of duplicate copies of a gene are found within a genome the less each individual copy contributes to total expression and the more likely it becomes that the loss of individual copies can be tolerated. In other words, the protection against fractionation provided by selection for gene dosage – either absolute or relative – becomes less powerful the less a given gene copy contributes to total expression, and the more total gene copies are present within the genome. This explains, at least in part, why despite being the “big kahuna” of whole genome duplications, plant genomes are not over-burdened with subunits of large dose-sensitive protein complexes.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Plant_Genetics_and_Genomics/10.3389/fpls.2012.00094/abstract

**REFERENCES**

Bekaert, M., Edger, P. P., Pires, J. C., and Conant, G. C. (2011). Two-phase resolution of polyploidy in the Arabidopsis metabolic network gives rise to relative and absolute dosage constraints. Plant Cell 23, 1719–1728.

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., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16, 1679–1691.

Schnable et al. (2011a). Mitigation of gene dosage balance.
Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422, 433–438.

Buggs, R. J. A., Zhang, L., Miles, N., Tate, J. A., Gao, L., Wei, W., Schnable, P. S., Brad Barbazuk, W., Solits, P. S., and Solits, D. E. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allotetraploid plant. Curr. Biol. 21, 551–556.

Chang, P. L., Dilkes, B. P., McMahon, M., Comai, L., and Nuzhdin, S. V. (2010). Homeodomain-specific retention and use in allotetraploid Arabidopsis suecica depends on parent of origin and network partners. Proc. Natl. Acad. Sci. U.S.A. 107, 19114–19119.

Duarte, I. M., Wall, P. K., Edger, P. P., Landherr, L. M., Ma, H., Pires, J. C., Leebens-Mack, J., and dePamphilis, C. W. (2010). Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10, 61. doi:10.1186/1471-2148-10-61

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.

Flagel, L. E., Juttellet, D., and Wendel, J. (2008). Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. BMC Biol. 6, 16. doi:10.1186/1741-7007-6-16

Flagel, L. E., and Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression during allotetraploid cotton speciation. New Phytol. 186, 184–193.

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, 435–453.

Gaut, B. S., and Doebely, J. F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. U.S.A. 94, 6809–6814.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G. W., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T. X, Yu, Budiworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Collard, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T., C., Mao, L., Quail, P., Wing, R., Dean, B., Yu, Y., Zharkikh, A., Shen, R., Saharabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavignian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tuineем, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Hartman, B., Bhatnagar, S., Dunn, M., Ko, C., Goff, S. A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genome Biol. 5, R25.

Jiao, Y., Wickett, N. J., Ayvampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. F., Hu, Y., Liang, H., Solits, P. S., Solits, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., and dePamphilis, C. W. (2011). Ancient allopolyploidy in seed plants and angiosperms. Nature 473, 97–100.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D. M., Correa, L. G. G., Reski, R., Mueller-Roeber, B., and Rensing, S. A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a time line of loss, gain, expansion, and correlation with complexity. Genome Biol. Evol. 2, 488–503.

Langham, R. J., Walsh, J., Dunn, M., Ko, C., Goff, S. A., and Freeling, M. (2013). The sequence of the rice genome (Oryza sativa ssp. japonica). Science 296, 92–100.

Liang, M., Bhatnagar, S., Adey, N., Rubano, T., Tuineem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Hartman, B., Bhatnagar, S., Dunn, M., Ko, C., Goff, S. A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genome Biol. 5, R25.

Lynch, M., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. Genome Res. 15, 516–525.

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. U.S.A. 102, 4545–4549.

Mei, N., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimson, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., Grigoriev, I. V., Lyons, E., Maher, C. A., Martis, M., Narechania, A., Otil, L. P., Penning, B. W., Salamov, A. A., Wang, Y., Zhang, L., Carpita, N. C., Freeling, M., Gingle, A. R., Hash, C. T., Keller, B., Klein, P., Kresovich, S., McCann, M. C., Ming, R., Peter- son, D. G., Rahman, M., Ware, D., Westhoff, P., Mayer, K. F., Messing, J., and Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556.

Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. U.S.A. 101, 9903–9908.

Sankoff, D., Zheng, C., and Zhu, Q. (2010). The collapse of gene complement following whole genome duplication. BMC Genomics 11, 313. doi:10.1186/1471-2164-11-313

Schnable, J. C., and Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6, e17855.

Schnable, J. C., Freeling, M., and Lyons, E. (2012). Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol. Evol. 4, 265–277.

Schnable, J. C., Springer, N. M., and Freeling, M. (2011a). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. U.S.A. 108, 4069–4074.

Schnable, J. C., Pedersen Brent, S., Sabarinhath, S., and Michael, F. (2011b). Dose-sensitivity, conserved non-coding sequences and duplicate gene retention through multiple tetraploidizations in the grasses. Front. Plant Sci. 2:2. doi:10.3389/fpls.2011.00002

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D., Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehanty, K., Fronick, C., Courney, B., Rock, S. M., Belter, E., Du, F., Kim, K., Abbott, R. M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S. M., Gilliam, B., Chen, W., Yan, L., Hingbootham, J., Cardenas, M., Maligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Than, T., Scimone, A., Than, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Bhatnagar, S., Haiden, J., Aschtheron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, K., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braitdott, M., Ashley, E., Golier, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento,
Mitigation of gene dosage balance

Schnable et al.

Schnable, J. C., Wang, X., Pires, J. C., and Freeling, M. (2012). Escape from preferential retention following repeated whole genome duplications in plants. Front. Plant Sci. 3:94. doi: 10.3389/fpls.2012.00094

This article was submitted to Frontiers in Plant Genetics and Genomics, a specialty of Frontiers in Plant Science. Copyright © 2012 Schnable, Wang, Pires and Freeling. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M. J., McMahan, L., Van Buren, P., Vaughn, M. W., Ying, K., Yeh, C. T., Emrich, S. J., Jia, Y., Kalyanaraman, A., Hsiea, A. P., Barbazuk, W. B., Baecoom, R. S., Bruttnell, T. P., Carpita, N. C., Chaparro, C., Chia, J. M., Deracon, J. M., Estill, J. C., Fu, Y., Jeddeloh, J. A., Han, Y., Lee, H., Li, P., Lisch, D. R., Liu, S., Liu, Z., Nagel, D. H., McCann, M. C., San Miguel, P., Myers, A. M., Nettleton, D., Nguyen, J., Penning, B. W., Ponnala, L., Schneider, K. L., Schwartz, D. C., Sharma, A., Soderlund, C., Springer, N. M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T. K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J. L., Dawe, R. K., Jiang, I., Jiang, N., Presting, G. G., Wessler, S. R., Aluru, S., Martienssen, R. A., Clifton, S. W., McCombie, W. R., Wing, R. A., and Wilson, R. K. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1112–1115.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 January 2012; paper pending publication: 12 March 2012; accepted: 24 April 2012; published online: 15 May 2012.

Citation: Schnable JC, Wang X, Pires JC and Freeling M (2012) Escape from preferential retention following repeated whole genome duplications in plants. Front. Plant Sci. 3:94. doi: 10.3389/fpls.2012.00094

www.frontiersin.org

May 2012 | Volume 3 | Article 94 | 7
### APPENDIX

Table A1 | Expression in *Arabidopsis* and complete gene loss in *Brassica rapa.*

| Category                                                                 | Less expressed copy lost in *Brassica rapa* | More expressed copy lost in *Brassica rapa* | p-Value   |
|--------------------------------------------------------------------------|--------------------------------------------|--------------------------------------------|-----------|
| All alpha pairs where one copy has been completely lost in *Brassica rapa* | 428 gene pairs                            | 217 gene pairs                            | $p = 3.60 \times 10^{-17}$ |
| Alpha pairs where there are multiple co-orthologs in *Brassica rapa* of the retained copy | 271 gene pairs                            | 98 gene pairs                             | $p = 3.48 \times 10^{-20}$ |
| Both copies expressed above five FPKM in *Arabidopsis thaliana*          | 191 gene pairs                            | 128 gene pairs                            | $p = 2.49 \times 10^{-4}$ |