Protein–protein and protein–DNA dosage balance and differential paralog transcription factor retention in polyploids

*James A. Birchler[1] and Reiner A. Veitia[2,3] ***

[1] Division of Biological Sciences, University of Missouri, Columbia, MO, USA
[2] CNRS UMR7592, Institut Jacques Monod, Paris, France
[3] Université Paris Diderot, Paris, France
*Correspondence to: veitia.reiner@ijm.univ-paris-diderot.fr

A commentary on

Dose-sensitivity, conserved noncoding sequences and duplicate gene retention through multiple tetraploidies in the grasses
by Schnable, J. C., Pedersen, B. S., Subramaniam, S., and Freeling, M. (2011). Front. Plant Sci. 2:2. doi: 10.3389/fpls.2011.00002

Most eukaryotes have an evolutionary history of repeated polyploidization followed by fractionation (or diploidization; Makino and McLysaght, 2010; Jiao et al., 2011). The progression to the near diploid level is not random with regard to the classes of genes that are retained (Freeling et al., 2008; Freeling, 2009; Makino and McLysaght, 2010). Typically, the genes that are preferentially retained are involved with macromolecular machines or heavily connected in the interactome. This differential progression of genic retention is unlikely to be only due to changes in protein function of the different members of a duplicate pair in processes referred to as subfunctionalization (subdivision of function) and neofunctionalization (gain of a novel function; Freeling, 2009). There are two arguments why this should be the case. First, the same classes of genes that are preferentially retained following whole genome duplication are preferentially underrepresented in segmental duplications (Freeling, et al., 2008; Makino and McLysaght, 2010). Both processes will produce duplicate genes that are available for divergence but the reciprocal distribution suggests that other factors are operative. Secondly, the duplicates that are retained for longer periods of evolutionary time very often eventually decay to the diploid state indicating that there has been no bona fide subdivision of function that would maintain both copies. It should be noted, however, that subdivision or gain of

function has certainly been documented for duplicate genes in evolution and the retention of regulatory genes for longer periods of evolutionary time provides greater opportunity for these changes in function to accumulate.

The types of genes that are preferentially retained following whole genome duplications and depleted in segmental copy number changes are quite similar to those shown to exhibit dosage effects in aneuploids (Birchler, 1979; Birchler and Newton, 1981; Guo and Birchler, 1994; Birchler et al., 2001). An analogy can be made to the generalized lack of effects on gene expression by whole genome changes but a regular and consistent set of modulations that occur in aneuploids (Birchler and Newton, 1981; Guo and Birchler, 1994; Guo et al., 1996). This set of observations led to the concept that the stoichiometry of members of regulatory macromolecular complexes involved in the control of transcription was important in affecting the expression of the target genes (Birchler and Newton, 1981; Birchler et al., 2001). These types of dosage effects can often be reduced to the action of single genes (Birchler et al., 2001) and indeed heterozygous mutations of transcription factors were recognized to produce human clinical conditions (Veitia, 2002, 2003, 2004). The stoichiometry of members of macromolecular complexes was postulated to explain this (semi-) dominance (Veitia, 2002). An issue pertinent to this discussion is the relationship of gene copy number to protein expression level. For instance, in a study in diploid yeast, knockouts of every gene were examined for protein concentration (Springer et al., 2010). Only 5% showed no correlation and 80% of genes showed a strong correlation, i.e., 50% expression of normal. The connection between gene dosage and the phenotype can be traced back to classical genetics in which it was known that

changes in whole ploidy would produce some level of morphological change but alterations in the copy number of portions of the genome could be quite detrimental or indeed lethal (Birchler and Veitia, 2007). Thus, the change in stoichiometry of dosage balanced gene products would have negative fitness consequences manifested in the phenotype and be selected against (Papp et al., 2003; Birchler et al., 2005; Veitia et al., 2008).

Biophysical evidence suggests that the more interaction partners a particular protein has, the less likely it is to be involved with a duplication event, indicating further that macromolecular complexes require a balance of subunits to maintain good fitness (Liang et al., 2008). Examinations of protein databases also indicate that proteins with many interactions display lower expressional noise and are underrepresented in copy number variants (Schuster-Bockler et al., 2010). Thus, from the biochemical level to the phenotype, there is evidence for a balance of gene products involved in such complexes, which provides implications in biophysics, evolution, gene expression, and quantitative trait analysis. This synthesis is referred to as the Gene Balance Hypothesis (Birchler and Veitia, 2007, 2010). To reiterate, the underlying theme of the above synthesis is that the amounts of different subunits and mode of assembly of multisubunit complexes will affect the final yield and that this fact will impact the phenotype. One of the tenets of this concept is that during the assembly of multi-subunited complexes, a relative excess of one subunit might lead to the production of potentially inactive subcomplexes. Such a circumstance will produce a different quantity of the whole complex under consideration and affect the functional output.

Schnable et al. (2011) highlight another aspect for the study of retained genes following ancient tetraploidy. These authors

examined conserved non-coding sequences (CNS) associated with genes encoding transcription factors and found that they too can exhibit an extended retention in duplicate over evolutionary time beyond the standard deletion frequency. This observation suggests that there may be negative fitness consequences of deletion of one member of a duplicate pair and, as such, a requirement for the proper balance of these sequences relative to other factors (namely, DNA-binding proteins) in the genome. This concept is based on the idea that transcription factor genes encode proteins that very often function in multi-protein complexes in interaction with DNA. The typical example of this situation is the complex enhanceosome, which is a higher order nucleoprotein "aggregate" that works as transcriptional pre-initiation/stimulatory complexes (Carey, 1998; Levine, 2010). Enhanceosomes are thought to ensure the formation of a specific activation surface that is "complementary" to other co-activators and the transcription machinery. These considerations led the authors to hypothesize that protein–DNA interactions should be sensitive to the "concentration" of the transcription factors and the binding sites of the *cis*-regulatory regions of the genes encoding transcription factors. The concept of dosage sensitive protein–DNA interactions, would be an important confirmation and extension of the Gene Balance Hypothesis.

To address whether the retention of CNS associated with transcription factor genes was simply coincidental, Schnable and colleagues asked whether there was a preferential retention of CNS-rich genes compared to CNS-poor genes, which indeed was the case. Consistently, their analysis showed that the less CNS-rich genes were significantly less likely to have both duplicate copies retained in a second round of whole genome duplication in the maize lineage. Indeed, this finding of preferential retention of some cases of CNS from whole genome duplications suggests that protein–DNA interaction is an important aspect of stoichiometric balance. In terms of complex assembly, the kinetics and stoichiometry of binding to DNA of transcription factors could certainly influence the final amount of functional complexes and hence their biological activity. The change in copy number of either genes encoding transcription factors or their cognate binding sites might influence the dynamics and outcome of complex formation. Indeed, it would not be surprising that the concentration of the DNA-binding sites and the concentration of the relevant factors that recognize them would have evolved preferred stoichiometries. In such a case, fractionation (deletion) of a copy of the gene encoding a transcription factor would be counter-selected because this would change the relative concentration of binding sites and binding factors. From the perspective of the deletion of the DNA-binding sites, deletion of only one gene is not likely to alter much the protein/DNA stoichiometry. However, one must note that a transcription factor can be controlling hundreds or thousands of target genes that can be undergoing fractionation.

In this discussion we cannot overlook the fact that DNA-binding proteins may also establish non-specific interactions with DNA. Given the size of plant genomes, there may be a substantial amount of non-specific interactions. A transcription factor normally recognizes many fewer specific binding sites with high affinity than non-specific ones. Mathematical simulations show, for instance, that increasing the concentration of a transcription factor for a smaller concentration of non-specific binding sites (due to DNA deletion), can lead to a non-linear increase in the concentration of specific transcription factor-DNA complexes. As previously suggested, a strategy to maintain non-specific interactions at optimal levels involves pseudogenization without deletion or replacement of deleted DNA by repetitive DNA (Veitia and Bottani, 2009).

One correlate of the proposition of the authors would be that CNS-rich genes would be less represented in segmental copy number changes than CNS-poor genes, an issue that has yet to be examined. Also, the rich collection of data about the classes of genes that are preferentially retained in whole genome duplications and depleted in segmental changes has yet to inspire molecular biological experiments that will clarify aspects of the dynamics of protein–protein and protein–DNA interactions in producing these ultimate balance consequences. If the findings of Schnable and colleagues are confirmed with further genomic and biochemical evidence, the gene dosage balance concept should be broadened to include DNA–protein interactions.

REFERENCES

Birchler, J. A. (1979). A study of enzyme activities in a dosage series of the long arm of chromosome one in maize. *Genetics* 92, 1211–1229.

Birchler, J. A., Bhadra, U., Pal Bhadra, M., 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 Newton, K. J. (1981). Modulation of protein levels in chromosomal dosage series of maize: the biochemical basis of aneuploid syndromes. *Genetics* 99, 247–266.

Birchler, J. A., Riddle, N. C., Auger, D. L., and Veitia, R. A. (2005). Dosage balance in gene regulation: biological implications. *Trends Genet.* 21, 219–226.

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.

Carey, M. (1998). The enhanceosome and transcriptional synergy. *Cell* 92, 5–8.

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., Lyons, E., Pedersen, B., Alam, M., Ming, R., and Lisch, D. (2008). Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Res.* 18, 1924–1937.

Guo, M., and Birchler, J. A. (1994). Trans-acting dosage effects on the expression of model gene systems in maize aneuploids. *Science* 266, 1999–2002.

Guo, M., Davis, D., and Birchler, J. A. (1996). Dosage effects on gene expression in a maize ploidy series. *Genetics* 142, 1349–1355.

Jiao, Y., Wickett, N. J., Ayyampalayam, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., and dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763.

Liang, H., Rogale-Plazonic, K., Chen, J., Li, W. H., and Fernandez, A. (2008). Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. *PLoS Genet.* 4, e11. doi: 10.1371/journal.pgen.0040011

Makino, T., and McLysaght, A. (2010). Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9270–9274.

Papp, B., Pal, C., and Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197.

Schnable, J. C., Pedersen, B. S., Subramaniam, S., and Freeling, M. (2011). Dose-sensitivity, conserved non-coding sequences, and duplicate gene retention

through multiple tetraploidies in the grasses. *Front. Plant Sci.* 2:2. doi: 10.3389/fpls.2011.00002

Schuster-Bockler, B., Conrad, D., and Bateman, A. (2010). Dosage sensitivity shapes the evolution of copy-number varied regions. *PLoS ONE* 5, e9474. doi: 10.1371/journal.pone.0009474

Springer, M., Weissman, J. S., and Kirshner, M. W. (2010). A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* 6, 368.

Veitia, R. A. (2002). Exploring the etiology of haploinsufficiency. *Bioessays* 24, 175–184.

Veitia, R. A. (2003). Nonlinear effects in macromolecular assembly and dosage sensitivity. *J. Theor. Biol.* 220, 19–25.

Veitia, R. A. (2004). Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics* 168, 569–574.

Veitia, R. A., and Bottani, S. (2009). Whole genome duplications and a "function" for junk DNA? Facts and hypotheses. *PLoS ONE* 4, e8201. doi: 10.1371/journal.pone.0008201

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.