

Evolutionary constraints imposed by gene dosage balance

Han Liang¹, Ariel Fernandez²

¹Department of Ecology and Evolution, University of Chicago, IL, 60637, USA, ²Department of Bioengineering, Rice University, Houston, TX, 77005, USA, Department of Computer Science, University of Chicago, Chicago, IL, 60637, USA

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Gene dosage imbalance as a basis of genetic dominance
4. The impact of gene dosage balance on gene duplicability
5. Perspective
6. Acknowledgements
7. References

1. ABSTRACT

The gene dosage balance hypothesis states that a concentration imbalance among components of macromolecules is often deleterious. Thus this notion potentially provides a mechanistic explanation for understanding genetic dominance and gene duplicability. Accumulating evidence emerged from recent genomic data has strongly supported this hypothesis. Further efforts are needed to understand dosage sensitivity in the context of organismal complexity and co-regulation of genes under dosage-balance constraints.

2. INTRODUCTION

The biological function of a protein not only requires the correct folding of its three-dimensional structure, but also may crucially depend on its expression level. The concept of gene dosage traces back to the early days of genetics (1). More recently, Veitia proposed a very important hypothesis, called “gene dosage balance hypothesis”, which states that the stoichiometric imbalance among components of macromolecular complexes is harmful (2). Later this “balance” notion has been suggested to operate on gene regulatory processes as well, such as signal transduction or genetic pathway (3-5). In a broad sense, the dosage balance hypothesis can be further extended to all the genes that require precise binding partnership to perform their normal function.

According to this hypothesis, both under- or over-expression of a dosage-sensitive gene can lower fitness, and therefore the causing mutations would be removed by purifying selection. The potential mutations can be complete gene deletion, gene duplication, regulatory mutations affecting promoter activity and intragenic changes with abolished product activity (null allele). In particular, the gene dosage balance hypothesis has a direct link with two fundamental biological phenomena: genetic dominance and gene duplication. In this article, we will review how recently available large-scale genomic data contributes to our understanding about these two topics in the light of the gene dosage balance hypothesis.

3. GENE DOSAGE IMBALANCE AS A BASIS OF GENETIC DOMINANCE

In a diploid organism, when a single functional gene copy (with the other copy inactivated by mutation) at a given locus is dominant, it is called haplo-sufficient (wide-type dominant) (2). In contrast, sometimes the lost-of-function mutation can be dominant and lead to an abnormal or disease phenotype. This phenomenon is called haplo-insufficient (wide-type recessive), since half of the normal amount of active gene product is not sufficient to maintain the associated normal phenotype (2). The gene dosage balance hypothesis was first proposed as a

The gene dosage balance hypothesis

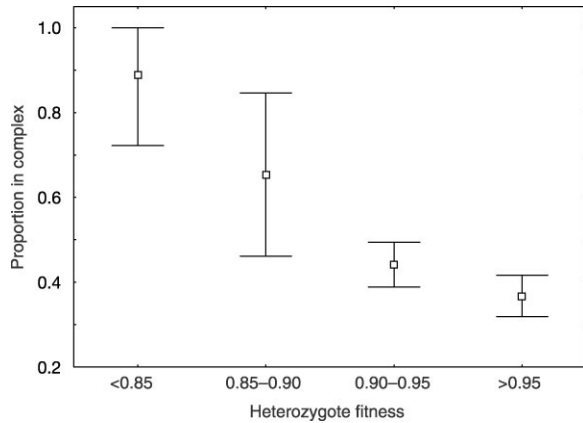


Figure 1. Proportion of genes in protein complex as a function of heterozygote fitness. Only essential genes are considered. Confidence intervals (95%) were obtained by the bootstrap method. Reproduced with permission from reference 6.

mechanistic explanation for the dominance in haplo-insufficient genes: substantially reduced amount of gene products alters stoichiometry required for normal gene activity, resulting in a fitness defect.

The first evidence for such balance comes from examining the fitness effect of heterozygous yeast strains in the context of protein complexes. In a seminal study, Papp and colleagues showed that genes with low heterozygote fitness tend to encode proteins in complexes (i.e., dosage-sensitive genes are at least twice as likely to participate in protein complexes as genes with low dosage sensitivity, Figure 1) (6). Furthermore, they found that interacting subunits with relatively high fitness deficiency are more likely to be co-expressed, suggesting that precise transcriptional co-regulation of interacting proteins has been maintained due to the constraints of dosage balance (6). A second line of evidence is the functional bias of haplo-sufficient and haplo-insufficient genes in humans and yeast. According to the dosage balance hypothesis, a key prediction is that genes whose protein products are supposed to be functionally “balanced” would be haplo-insufficient. Kondrashov and Koonin (7) verified this prediction by showing that enzymatic functions are significantly over-represented among haplo-sufficient genes (enzymes are known to be particularly dosage-insensitive [8]); whereas haplo-insufficient genes preferentially encode regulators, signal transduction elements and structural proteins, which generally work in a dosage-dependent manner (2, 9). Thus the dosage balance notion is compatible with the dominance of observed haplo-insufficient or haplo-sufficient genes.

4. THE IMPACT OF GENE DOSAGE BALANCE ON GENE DuplicABILITY

Gene duplication is a major mechanism for increasing genome complexity and generating new genes (10, 11). As a direct consequence, a duplication event increases gene copy number in the genome and may alter

the original dosage balance. Thus, analyses on gene duplicability (i.e. the number of paralogs) frame a vantage point to investigate the impact of dosage-balance constraints on gene and genome evolution.

Several lines of evidence appear to support the notion of dosage balance. First, Papp and colleagues showed that 33% of single-copy (singleton) genes participate in protein complexes, whereas this frequency drops to 21% for genes with three or more paralogs, suggesting that single-gene duplications involved in protein complexes are strongly disfavored by selection (6). Second, Yang and colleagues further demonstrated that the proportion of unduplicated genes increases with the number of subunits in a protein complex, suggesting that the chance of imbalance after duplication is increased with protein complexity (12). Third, Lin and colleagues found that as for the duplicability of genes in protein complexes, the major distinction exists between hetero-complexes (consisting of at least two different types of subunits) and homo-complexes (consisting of only one polypeptide subunit) (13). Again, this is consistent with the dosage balance hypothesis, because duplication of subunits in a hetero-complex is more likely to cause dosage imbalance. Fourth, through the analyses on yeast protein-protein interaction data, two recent studies showed that hub proteins (i.e. proteins with many interacting partners) tend to have a low gene duplicability, probably because proteins with many interactions are under strong constraints of dosage balance (14, 15). Fifth, Teichmann and Veitia found that a statistically significant fraction of yeast genes coding for subunits of stable complexes are located in the neighboring regions of each other, which may ensure better co-regulation and maintain the stoichiometry of complexes during duplication of chromosomal segments (16).

In the above studies, protein complex or protein interaction data was mainly used to infer the neediness of dosage balance. While straightforward and informative, these approaches have some inherited limitations (e.g., representing the potential dosage imbalance effect in a very abstract way). Recent advances in structural genomics and biophysics enable us to examine the dosage balance hypothesis in a more detailed way. We have focused on a specific attribute of protein three-dimensional structure, the so-called under-wrapping (17-19). Based on the amount of dehydrons (water-accessible hydrogen bonds, representing structural vulnerabilities), this attribute quantifies the extent to which the protein structure is reliant on the interactive context to maintain its integrity. As protein structures become more under-wrapped, their functional competence becomes more reliant on binding partnership (20, 21). From this point of view, the gene dosage balance hypothesis predicts that duplicates of a highly under-wrapped protein should be more sensitive to dosage imbalance and be less likely to be retained in evolution. Indeed, we found a universal negative correlation between protein under-wrapping and gene duplicability in all the six model organisms examined (although this dependence becomes weak in complex organisms, Figure 2) (22). Importantly, this study provides crucial insights into the molecular/structural basis of dosage sensitivity.

The gene dosage balance hypothesis

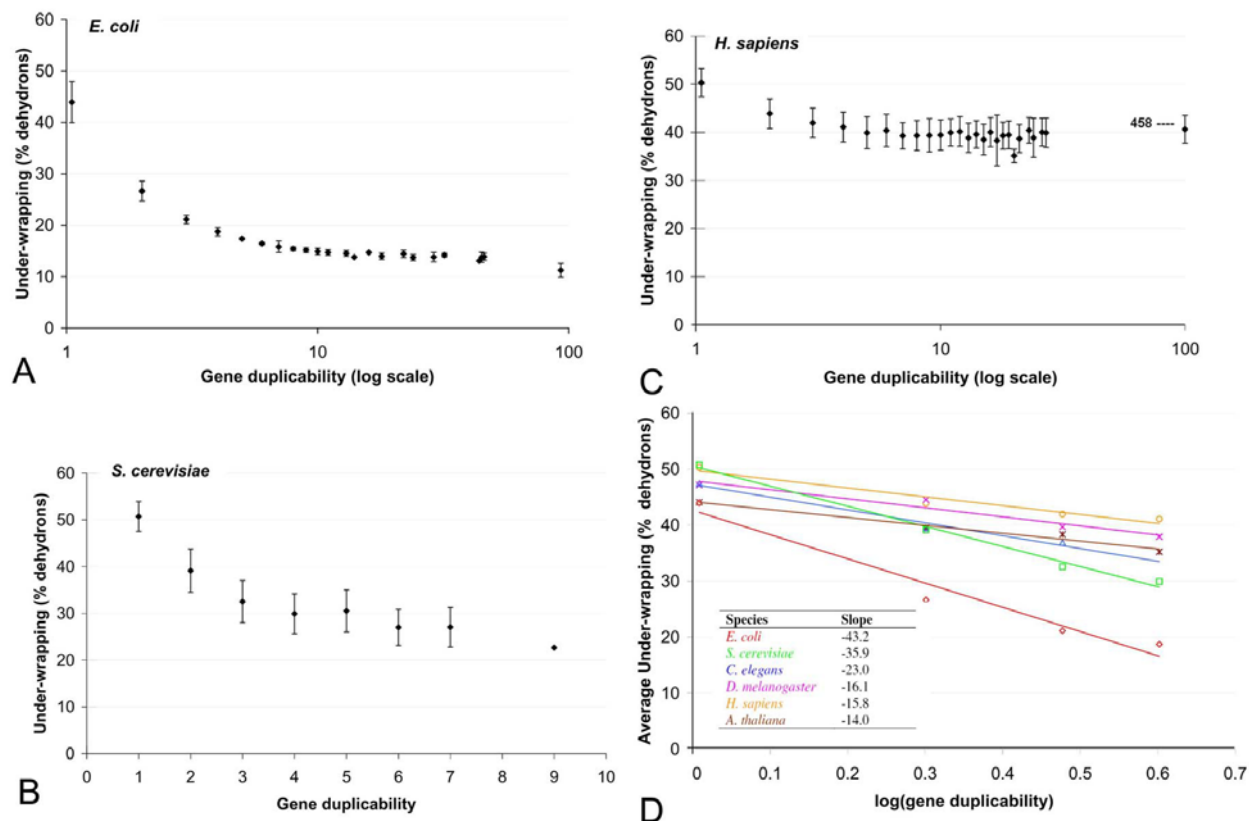


Figure 2. Negative correlations between protein under-wrapping extent and gene duplicability: (A) in *E. coli*, (B) in yeast, (C) in human and (D) slopes in six organisms. Here gene duplicability is defined as the gene family size (m). Because of the huge spread in duplicability for *E. coli* and *H. sapiens*, a log scale was adopted on the abscissas for (A) and (C). The mean extent of wrapping is determined by averaging over all genes binned by gene duplicability value. The error bars indicate \pm a standard deviation from the mean values. The slopes in (D) are determined by the least squares linear regression from $m = 1$ to 4. Reproduced with permission from reference 22.

In a single-gene or segmental duplication, the binding balance of duplicated genes is often broken. However, a whole-genome duplication maintains all the binding partnership and causes little dosage imbalance. From this unique perspective, several observations have supported the dosage balance hypothesis. First, as typical dosage-sensitive genes, ribosomal genes have a much higher frequency among yeast whole-genome duplicates than non-whole-genome duplicates (6). Next, duplicate proteins survived from a whole-genome duplication are structurally more under-wrapped (dosage-sensitive) than those otherwise, consistent with the expectation that dosage-balance constraints were relaxed during the whole genome duplication (22). Finally, substantial support comes from the functional bias of retained duplicated genes in *Arabidopsis* and rice, which are known to have experienced several rounds of whole-genome duplication (ployploidization). It has been found that “connected” genes such as signal transduction components and transcription factors are overrepresented in whole-genome duplicates (23-27). As the dosage balance hypothesis predicts, these genes are resistant to removal because they are in balance with each other and selection would prevent their rapid loss.

5. PERSPECTIVE

While the gene dosage balance hypothesis has received considerable support from recent analyses on genomic data, obviously it is not the only factor that determines genetic dominance and gene duplicability. This seems particularly true in more complex organisms, such as humans. Unexpected from the dosage-balance notion, haplo-insufficient genes (dosage-sensitive) have more paralogs than halpo-sufficient genes in the human genome, and this pattern appears to be independent from the similarity threshold used to classify gene families (7). This suggests that increased dosage for haplo-insufficient genes may have been favored by selection. Moreover, Liang and Li recently showed a positive correlation between protein connectivity and gene duplicability in humans, that is - hub proteins tend to have more paralogs than non-hub proteins (28). One possible explanation is that positive selection for functional diversification after gene duplication may have played an important role in determining gene loss and gene retention in humans.

Several key questions still remain unclear. First, why do different species display different dosage

The gene dosage balance hypothesis

sensitivity? In general, complex organisms (e.g. higher eukaryotes) are less dosage sensitive than simple organisms. This can be considered from the following aspects. (i) The effective population size is usually decreased with organismal complexity. Thus, given the same fitness defect of dosage imbalance, selection is more effective in simple organisms. (ii) Alternative splicing is a wide-spread phenomenon in higher eukaryotes, by which different protein products can be generated from the same gene locus. What's the role of alternative splicing in terms of affecting dosage sensitivity? Do altered spliced products represent "escape routes" from dosage-balance constraints after gene duplication? (iii) How does allostery (i.e., dimerization or oligomerization) influence dosage sensitivity, since self-binding interactions can partly alleviate dosage imbalance? (iv) Higher eukaryotes have much more complex regulatory systems (e.g., chaperons and small non-coding RNA) than simple organisms. Does increased complexity at the regulatory level confer a greater capability to deal with dosage variation?

For genes whose functions are under strong balance constraints, how is the tight stoichiometry among binding partners achieved? It seems pretty challenging when the balance has to be maintained across many tissue types. To address this question, we may not only need to consider both transcriptional and post-transcriptional regulation, but also epigenetic regulation (e.g. chromatin structures).

6. ACKNOWLEDGEMENTS

This work was supported by NIH grants.

7. REFERENCES

1. Sewall Wright.: Physiological and evolutionary theories of dominance. *American Naturalist*, 68, 24-53 (1934)
2. Reiner Veitia: Exploring the etiology of haploinsufficiency. *Bioessays*, 24, 175-184 (2002)
3. James Birchler, Nicole Riddle, Donald Auger, Reiner Veitia: Dosage balance in gene regulation: biological implications. *Trends Genet*, 21, 219-226 (2005)
4. Reiner Veitia: Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics*, 168, 569-574 (2004)
5. Reiner Veitia: Gene dosage balance: deletions, duplications and dominance. *Trends Genet*, 21, 33-35 (2005)
6. Balazs Papp, Csaba Pal, Laurence Hurst: Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424, 194-197 (2003)
7. Fyodor Kondrashov, Eugene Koonin: A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet*, 20, 287-290 (2004)
8. Allen Orr: A test of Fisher's theory of dominance. *Proc Natl Acad Sci U S A*, 88, 11413-11415 (1991)
9. Elizabeth Fisher, Peter Scambler: Human haploinsufficiency--one for sorrow, two for joy. *Nat Genet*, 7, 5-7 (1994)
10. Michael Lynch, John Conery: The origins of genome complexity. *Science*, 302, 1401-1404 (2003)
11. Susumu Ohno: Evolution by Gene Duplication. *Springer-Verlag*, New York (1970)
12. Jing Yang, Richard Lusk, Wen-Hsiung Li: Organismal complexity, protein complexity, and gene duplicability. *Proc Natl Acad Sci U S A*, 100, 15661-15665 (2003)
13. Yeong-shin Lin, Jenn-Kang Hwang, Wen-Hsiung Li: Protein complexity, gene duplicability and gene dispensability in the yeast genome. *Gene*, 387, 109-117 (2007)
14. Diana Ekman, Sara Light, Asa Bjorklund, Arne Elofsson: What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol*, 7, R45 (2006)
15. Anuphap Prachumwa, Wen-Hsiung Li: Protein function, connectivity, and duplicability in yeast. *Mol Biol Evol*, 23, 30-39 (2006)
16. Sarah Teichmann, Reiner Veitia: Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*, 167, 2121-2125 (2004)
17. Ariel Fernandez: Keeping dry and crossing membranes. *Nat Biotechnol*, 22, 1081-1084 (2004)
18. Ariel Fernandez, Stephen Berry: Proteins with H-bond packing defects are highly interactive with lipid bilayers: Implications for amyloidogenesis. *Proc Natl Acad Sci U S A*, 100, 2391-2396 (2003)
19. Ariel Fernandez, Jozsef Kardos, Ridgway Scott, Yuji Goto, Stephen Berry: Structural defects and the diagnosis of amyloidogenic propensity. *Proc Natl Acad Sci U S A*, 100, 6446-6451 (2003)
20. Ariel Fernandez, Stephen Berry: Molecular dimension explored in evolution to promote proteomic complexity. *Proc Natl Acad Sci U S A*, 101, 13460-13465 (2004)
21. Ariel Fernandez, Harold Scheraga: Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci U S A*, 100, 113-118 (2003)
22. Han Liang, Kristina Plazonic, Jianping Chen, Wen-Hsiung Li, Ariel Fernández: Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. *PLoS Genetics*, 4, e11 (2008)

The gene dosage balance hypothesis

23. Guillaume Blanc, Kenneth Wolfe: Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell*, 16, 1679-1691 (2004)

24. Brad Chapman, John Bowers, Frank Feltus, Andrew Paterson: Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. *Proc Natl Acad Sci U S A*, 103, 2730-2735 (2006)

25. Michael Freeling, Brain Thomas: Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res*, 16, 805-814 (2006)

26. Steven Maere, Stefanie De Bodt, Jeroen Raes, Tineke Casneuf, Marc Van Montagu, Martin Kuiper, Yves Van de Peer: Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*, 102, 5454-5459 (2005)

27. Brain Thomas, Brent Pedersen, Michael Freeling: Following tetraploidy in an Arabidopsis ancestor, genes were removed referentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res*, 16, 934-946 (2006)

28. Han Liang, Wen-Hsiung Li: Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends Genet*, 23, 375-378 (2007)

Key Words Dosage balance, Gene duplicability, Haplo-insufficient, Dominance, Whole-genome duplication, Review

Send correspondence to: Han Liang, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, IL, 60637, USA, Tel: 773-834-3964, Fax: 773-702-9740, E-mail: hliang@uchicago.edu

<http://www.bioscience.org/current/vol13.htm>