

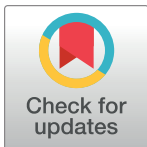
PRIMER

# Regulation plays a multifaceted role in the retention of gene duplicates

Johan Hallin<sup>1,2,3,4,5</sup> , Christian R. Landry<sup>1,2,3,4,5\*</sup> 

**1** Département de biochimie, microbiologie et bio-informatique, Faculté des sciences et de génie, Université Laval, Québec, Canada, **2** Département de biologie, Faculté des sciences et de génie, Université Laval, Québec, Canada, **3** Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada, **4** PROTEO, Le réseau québécois de recherche sur la fonction, la structure et l'ingénierie des protéines, Université Laval, Québec, Canada, **5** Centre de Recherche en Données Massives (CRDM), Université Laval, Québec, Canada

\* [christian.landry@bio.ulaval.ca](mailto:christian.landry@bio.ulaval.ca)



## Abstract

A gene duplication can lead to all sorts of problems in a cell. However, it can also lead to all sorts of benefits. Beneficial or not, the gene duplicates might be kept in the genome because of several different reasons. For instance, if natural selection works towards optimizing one function of a gene at the expense of another, then gene duplication could resolve this conflict by separating the functions in two genes. Here, we outline evolutionary incentives to keep a duplicated gene in the genome, focusing on divergence in expression and trade-off resolution as featured in a new and exciting paper published in this edition of *PLOS Biology*.

## OPEN ACCESS

**Citation:** Hallin J, Landry CR (2019) Regulation plays a multifaceted role in the retention of gene duplicates. *PLoS Biol* 17(11): e3000519. <https://doi.org/10.1371/journal.pbio.3000519>

**Published:** November 22, 2019

**Copyright:** © 2019 Hallin, Landry. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by a Human Frontiers Science Program fellowship (LT000182/2019-L) to JH, a Canadian Institute for Health Research (CIHR) foundation grant (387697) to CRL. CRL holds the Canada Research Chair in Evolutionary Cell and Systems Biology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** RAR, retinoic acid receptor; SSD, small-scale duplication; WGD, whole-genome duplication.

Genes have life cycles of their own. They are, most of the time, born from the duplication of other genes ([Box 1](#)) and may eventually die and become pseudogenes. During the period between birth and death, the sequence and regulatory elements of a new gene change through mutations. This dynamic gain and loss of genes and the associated changes to regulation and function contribute to phenotypic differences between species and among populations of the same species [1]. Numerous studies over the past 50 years have investigated the role of different evolutionary forces such as natural selection and drift in shaping these life cycles, for example, by investigating the contribution of nucleotide and amino acid substitutions to the divergence of new genes. One category of molecular changes that appears to play a key role in the evolution of genes that originate from gene duplication (duplicates or paralogs) are regulatory changes, i.e., changes in the gene itself or elsewhere in the genome that determine when, where, and at what level a gene is transcribed and translated.

## Regulatory evolution and the maintenance of gene duplicates

The immediate effect of gene duplication is typically an increase in gene dosage [4] ([Fig 1](#)). Higher dosage, however, does not always translate into increased fitness [5]. This means that at this stage, natural selection could favor gene retention or loss, or if the expression change is effectively neutral, the duplicate could evolve neutrally for extended periods of time ([Box 1](#)). If an increase in total expression is favored by internal or external conditions, a gene duplication

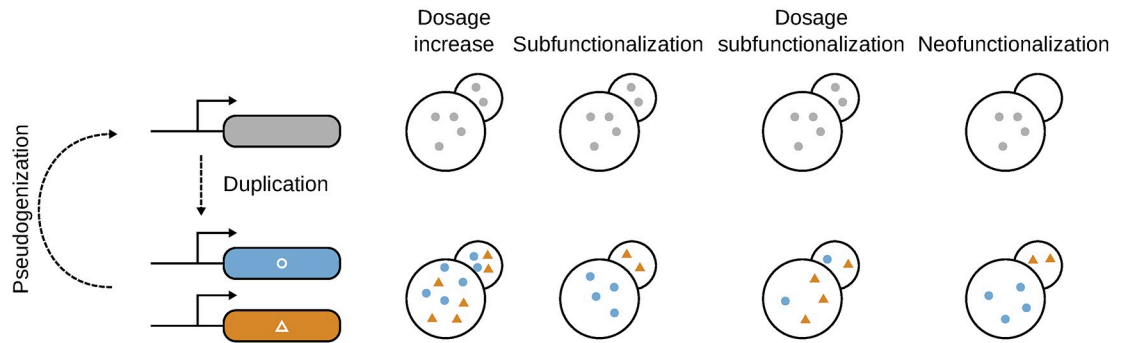
### Box 1

Gene duplicates originate mainly by two mechanisms: small-scale duplication (SSD) and whole-genome duplication (WGD) [30]. In SSDs, only one or a few genes are duplicated, whereas in WGD, all genes are duplicated simultaneously. These two mechanisms have specific features that influence the retention of duplicates, which in turn influences the properties of genes that originated from either mechanism. One of the key differences is that SSD genes first originate in a single individual and must increase in frequency by drift or selection to be maintained. WGD would also occur in one individual, but it could potentially incite or co-occur with a speciation event [31], which would coincide with a population bottleneck and thus the fixation of all duplicates without the need for natural selection. However, WGDs have been associated with performance traits in plants, for instance [32], which means natural selection can also favor their fixation.

There is a major difference between SSDs and WGDs if we consider interactions among gene products—for instance, for proteins forming protein complexes. WGD will likely maintain the stoichiometric balance of the complexes, whereas the duplication of a single subunit through SSD would perturb the balance [33]. In a WGD, this principle predicts the preferential maintenance of proteins that are dosage sensitive and whose loss would lead to a fitness defect because it would perturb the balance. WGD genes being more dosage-balance sensitive [34] is supported by observations that they have fewer copy-number variations in human populations [35] and are overrepresented among genes with copy-number variations that are pathogenic [36]. The properties of genes may therefore influence the probability that their duplicates are maintained after SSDs or WGDs, thereby determining what is the extent of novelties that can evolve from these mechanisms. In the case of Chapal and colleagues [25], the *Msn* duplicates show a fitness trade-off when expression is increased and are thus dosage sensitive, which suggests that their duplication may initially have been maintained specifically because it originated via a WGD.

could provide an immediate benefit. For example, genes coding for digestive enzymes, such as amylases that hydrolyze starch, vary in copy number among human populations. Copy-number correlates with the diet such that high-starch diets are associated with more copies, whereas low-starch diets are associated with fewer copies [6]. Diet is such a strong selective force that multiple copies have been maintained in many mammals [7]. Selection for higher dosage can sometimes lead to the maintenance of a large number of gene copies. An extreme example of this is the hundreds of duplicated copies of ribosomal RNA genes in certain microbial genomes: the adaptiveness of this most likely derives from some life-history traits that demand an increase in protein synthesis machinery [8]. When paralogs are maintained because of dosage effects, gene copies are maintained without the need for the individual copies to gain new functions.

Regulatory evolution could also favor the retention of a new gene by changing the tissue or timing of expression, a process called neofunctionalization (Fig 1). The duplicate's newly gained expression pattern would favor its retention by contributing a new function to, for instance, a tissue. The gain of new expression specificity can also be accompanied by and facilitate the gain of new molecular functions at the protein level. Indeed, the change of cellular



**Fig 1. Regulatory evolution in the retention and divergence of gene duplicates.** The cells represent a mother and a daughter cell to illustrate different phases of development. The top row shows these cells between the duplication event and the divergence of the duplications. The same concept applies to multicellular organisms with different cell types. Colored dots represent gene products from the ancestral gene (gray, in the cells at the top) and from the duplicated genes (blue circles and orange triangles, in the cells at the bottom). Retention by dosage effects refers to a gain in fitness caused by a larger amount of gene product. The dosage change does not need to correspond to an exact doubling (as illustrated) but could be higher [2] or lower than that if, for instance, some mechanisms of attenuation are present [3]. Retention by subfunctionalization refers to the maintenance of the two copies by the splitting of the ancestral function, here illustrated by the different localizations. Dosage subfunctionalization refers to a special case in which the total expression is maintained, but the abundance of each duplicate can change. Neofunctionalization refers to the evolution of a new function by a paralog, here shown by the new localization. In the absence of sufficient selection pressure to maintain two copies, the system can revert to a single gene system through pseudogenization (or simple loss by other mechanisms) of one copy.

<https://doi.org/10.1371/journal.pbio.3000519.g001>

context for the protein can represent new opportunities for selective forces to act on the protein itself. The retinoic acid receptors (RARs) have evolved following this path in vertebrates. RARs are nuclear receptors that are bound by specific ligands and that activate the transcription of genes during key developmental steps. Three paralogs originated from the whole-genome duplications (WGDs) at the origin of vertebrates, two of which have evolved new ligand-binding specificities and expression patterns during development [9].

Although it is easy to conceive that natural selection may favor the maintenance of gene duplicates because of dosage effects or new regulatory programs, it may be less intuitive to imagine that gene duplicates could be maintained by degenerative mutations that lead to the “specialization” of each duplicate (subfunctionalization, Fig 1). The theory behind the role of this mechanism was formally derived by [10,11]. Briefly, the model showed that if a gene has multiple functions or tissues of expression, its duplication could be followed by the loss of different functions in each copy while still preserving all the ancestral functions. However, because now the two genes are required to perform the functions previously performed by the single progenitor, natural selection will act to maintain both copies. The power of this model is that it does not require the evolution of new and adaptive functions, which may be inaccessible for many genes and thus could not explain why gene duplicates would be maintained. Regulatory subfunctionalization was recently hypothesized to occur at the level of alternative splicing and subcellular localization of the plastid ascorbate peroxidase in plants, a key detoxifying enzyme. Some plants have a single gene that produces two distinct proteins by alternative splicing that localize in different cell compartments; others have two independent genes, each producing a single protein that localizes to one compartment or the other [12].

Recent work showed that subfunctionalization could also take place at the level of gene dosage (dosage subfunctionalization, Fig 1) and not necessarily implying the loss of other molecular functions such as tissue or timing expression specificity. Gout and Lynch [13] showed that natural selection to maintain the expression level of a gene could act on total expression of a gene pair rather than on each of them individually. This allows the two genes to accumulate

mutations that change the expression levels without being filtered by natural selection, if total expression is maintained. For instance, if one copy accumulates mutations that reduce expression, then the other copy could accumulate mutations that increase expression, all the while maintaining the total expression level. Eventually, one copy could be expressed at such a low level that its loss would be effectively neutral.

## Other dimensions of gene expression regulation

The examples mentioned thus far consider only a few dimensions of gene regulation (level, timing, localization). Yet gene expression systems are highly dynamic, and other features may contribute to the evolution and retention of gene duplicates. Two important expression features are (1) the responsiveness, which refers to the magnitude and propensity to change gene expression levels in response to intra- and extracellular signals, and (2) expression noise. Responsiveness has mostly been studied in single-celled organisms such as yeast, in which expression level has been studied in entire populations across hundreds of growth and stress conditions as well as at the single-cell level using various reporters. There are important differences in the sensitivity of genes to environmental changes and mutations: some genes rarely change expression levels, whereas others do so easily [14]. Interestingly, responsiveness also appears to correlate with divergence of gene expression levels among species [15]. More responsive genes show more differences in expression regulation between species. This observation suggests that responsiveness could be a gene property that favors divergence between species. However, this is not a universal trend, as responsiveness could also be selected against for many genes that would rather require stable dosage [16].

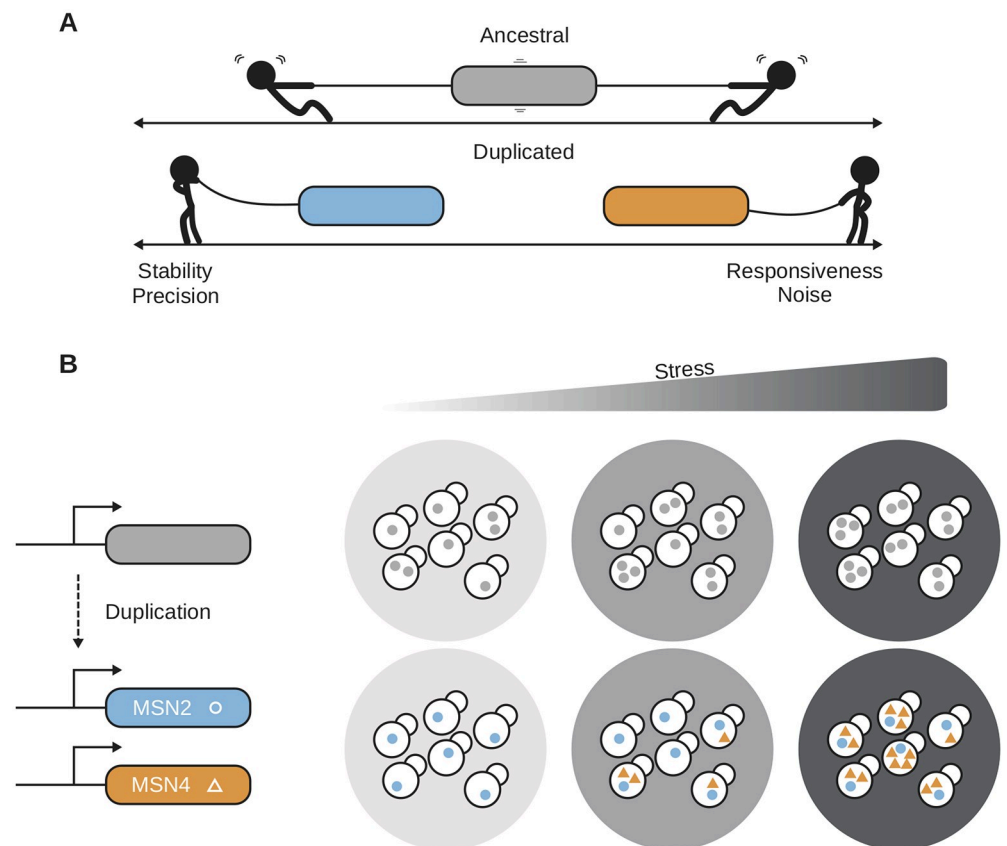
Expression noise appears to be strongly associated with responsiveness. Noise is linked to the architecture of the genes themselves and is manifested by expression differences among cells that are genetically identical. Noisiness is not always easily assessed because it requires cells to be examined individually. Although some studies have suggested that noise in gene expression could be advantageous [17], it is also likely to be deleterious because it prevents a large fraction of a population from attaining the optimal expression level at a given time [18]. Attesting to the importance of low noise, essential genes and (most importantly) those that reduce fitness when their dosage is reduced (haploinsufficient) tend to be less noisy than genes that do not show measurable effects on fitness upon deletion [19,20]. Furthermore, the study of the fitness consequences of noise and changes in average expression has revealed that noise could, in some cases, be as detrimental as changes in mean expression [21,22].

Nevertheless, gene expression noise is prevalent. This prevalence could be explained by the fact that for some gene classes and promoter architectures (those with a TATA box), responsiveness and noise seem to be intrinsically coupled [23]. Although natural selection may favor responsiveness, the inability of cells to reach a precise expression level comes as an unavoidable cost. This correlation between responsiveness and noise was detailed by Lehner [24], who also suggested that this trade-off could be alleviated by gene duplication because it would allow the system to maintain responsiveness while reducing noise. If the deviation from optimality of expression level of the two genes is not correlated, their average expression will be closer to optimal level than the expression of an individual gene with the same average expression and noise level. Gene duplication in this case would allow for two responsive genes but with reduced absolute noise. The consideration of gene regulation at the single-cell level therefore allowed geneticists to uncover potential mechanisms for the maintenance of gene duplication. However, a detailed example of what role these features of gene expression play in the maintenance of duplicates was yet to come.

## Single-cell biology offers a new perspective on the role of regulatory evolution in the retention of gene duplicates

This edition of *PLOS Biology* [25] brings forth an elegant example of a gene duplication that did not result in neofunctionalization or subfunctionalization as typically defined. The transcription factor Msn was duplicated during the WGD in the budding yeast *Saccharomyces cerevisiae* and has since diverged into Msn2 and Msn4. Previous studies of these two genes have differed in their conclusion as to the divergence of function between them. Despite previous suggestions that these two transcription factors may have diverged in terms of function [27], Chapal and colleagues provide convincing evidence that they regulate the same target genes. That raises the question: How and why would yeast have kept these paralogs for the last 100 million years?

Chapal and colleagues bring forward compelling evidence that these two transcription factors are cooperating in the cell to minimize growth speed defects while maximizing stress responsiveness. The authors show that higher levels of Msn2 are detrimental to the growth of the cells but beneficial when cells are in stressful conditions. They propose this simple trade-off between growth speed and environmental responsiveness as the incentive for the retention of the two copies of Msn, even though they have the same target genes. In accordance with



**Fig 2. Gene duplication as a force for trade-off resolution.** (A) Nonduplicated genes are forced to sacrifice expression responsiveness for precision because of the noise that accompanies responsiveness. Gene duplication can resolve this conflict by providing two genes with the same function at the protein level, one that is responsive and one that has a precise expression level. (B) Chapal and colleagues [25] provide a description of how the paralogs Msn2 and Msn4 cooperate to supply the cell with a precise but stable (Msn2) and responsive but noisy (Msn4) expression. Msn4 is hardly expressed during nonstress conditions but increases in expression as stress levels go up, whereas Msn2 remains constantly expressed at a low level. *Panel A inspired from [26].*

<https://doi.org/10.1371/journal.pbio.3000519.g002>

this, Chapal and colleagues [25] found that Msn2 has a low but steady expression with little noise, whereas Msn4 is environmentally responsive with a high level of noise. This allows for a regulatory dynamic that solves the conflict between a dynamic response, which comes with the trade-off of noisy expression, and a steady number of proteins in the cell during nonstress conditions (Fig 2). The expression of Msn2 does not change during the growth of a population, whereas Msn4 increases gradually along the growth curve and with it, the resistance to stress.

Interestingly, they compare the expression of the two paralogs with an ortholog from *Kluyveromyces lactis*, which diverged from *S. cerevisiae* before the WGD event (Box 1), and found that it had an expression profile that was intermediate to that of Msn2 and 4. The ortholog was induced throughout the growth curve, although at a lower level than Msn4, and its noisiness was intermediate between the two (Fig 2B). The authors suggest the following scenario: After the WGD, Msn2 gained a more stable expression by its transcription start site moving farther away from the open reading frame to the boundary of a nucleosome free region. Msn4, on the other hand, increased its dynamic range and noise by gaining new transcription factor binding sites.

The model proposed by this new study [25] is not, strictly speaking, about a case of subfunctionalization, because the initial model by Lynch and colleagues [10,11] does not require that the division of labor occurs with a gain in fitness. The case documented here rather suggests that division of labor allows for a gain in fitness by resolving a trade-off, as has been proposed for other pairs of paralogs that may have conflicting protein functions [28,29]. It is unclear how frequent this form of adaptive subfunctionalization is, given that many more types of conflicts may exist between the different functions of a given gene and may not necessarily be resolvable by simple mutational events. All cases of putative subfunctionalization may need to be dissected in detail as Chapal and colleagues did, to make sure that what appears to be a simple division of labor may not be accompanied with an exquisite functional specialization.

## Acknowledgments

We thank Angel Cisneros, Simon Aubé, Damien Biot-Pelletier, Anna Fijarczyk, David Bradley, and Laurence Hurst for comments.

## References

1. Chen S, Krinsky BH, Long M. New genes as drivers of phenotypic evolution. *Nat Rev Genet.* 2013; 14: 645–660. <https://doi.org/10.1038/nrg3521> PMID: 23949544
2. Loehlin DW, Carroll SB. Expression of tandem gene duplicates is often greater than twofold. *Proc Natl Acad Sci U S A.* 2016; 113: 5988–5992. <https://doi.org/10.1073/pnas.1605886113> PMID: 27162370
3. Dephore N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, et al. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *Elife.* 2014; 3: e03023. <https://doi.org/10.7554/eLife.03023> PMID: 25073701
4. Ohno S. Other Mechanisms for Achieving Gene Duplication. In: Ohno S, editor. *Evolution by Gene Duplication.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1970. pp. 107–110.
5. Moriya H, Makanae K, Watanabe K, Chino A, Shimizu-Yoshida Y. Robustness analysis of cellular systems using the genetic tug-of-war method. *Mol Biosyst.* 2012; 8: 2513–2522. <https://doi.org/10.1039/c2mb25100k> PMID: 22722869
6. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 2007; 39: 1256–1260. <https://doi.org/10.1038/ng2123> PMID: 17828263
7. Pajic P, Pavlidis P, Dean K, Neznanova L, Romano R-A, Garneau D, et al. Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife.* 2019; 8. <https://doi.org/10.7554/eLife.44628> PMID: 31084707



8. Nelson JO, Watase GJ, Warsinger-Pepe N, Yamashita YM. Mechanisms of rDNA Copy Number Maintenance. *Trends Genet.* 2019; <https://doi.org/10.1016/j.tig.2019.07.006> PMID: 31395390
9. Escrava H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, et al. Neofunctionalization in vertebrates: the example of retinoic acid receptors. *PLoS Genet.* 2006; 2: e102. <https://doi.org/10.1371/journal.pgen.0020102> PMID: 16839186
10. Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. *Genetics.* 2000; 154: 459–473. PMID: 10629003
11. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 1999; 151: 1531–1545. PMID: 10101175
12. Qiu Y, Van Tay Y, Ruan Y, Adams KL. Divergence of duplicated genes by repeated partitioning of splice forms and subcellular localization. *New Phytol.* 2019; <https://doi.org/10.1111/nph.16148> PMID: 31469915
13. Gout J-F, Lynch M. Maintenance and Loss of Duplicated Genes by Dosage Subfunctionalization. *Mol Biol Evol.* 2015; 32: 2141–2148. <https://doi.org/10.1093/molbev/msv095> PMID: 25908670
14. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. *Science.* 2007; 317: 118–121. <https://doi.org/10.1126/science.1140247> PMID: 17525304
15. Tirosch I, Weinberger A, Carmi M, Barkai N. A genetic signature of interspecies variations in gene expression. *Nat Genet.* 2006; 38: 830–834. <https://doi.org/10.1038/ng1819> PMID: 16783381
16. Duveau F, Yuan DC, Metzger BPH, Hodgins-Davis A, Wittkopp PJ. Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 2017; 114: E11218–E11227. <https://doi.org/10.1073/pnas.1713960115> PMID: 29259117
17. Zhang Z, Qian W, Zhang J. Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol.* 2009; 5: 299. <https://doi.org/10.1038/msb.2009.58> PMID: 19690568
18. Wang Z, Zhang J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci U S A.* 2011; 108: E67–76. <https://doi.org/10.1073/pnas.1100059108> PMID: 21464323
19. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression. *PLoS Biol.* 2004; 2: e137. <https://doi.org/10.1371/journal.pbio.0020137> PMID: 15124029
20. Batada NN, Hurst LD. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet.* 2007; 39: 945–949. <https://doi.org/10.1038/ng2071> PMID: 17660811
21. Schmiedel JM, Carey LB, Lehner B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nat Commun.* 2019; 10: 3180. <https://doi.org/10.1038/s41467-019-11116-w> PMID: 31320634
22. Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. Selection on noise constrains variation in a eukaryotic promoter. *Nature.* 2015; 521: 344–347. <https://doi.org/10.1038/nature14244> PMID: 25778704
23. Blake WJ, KAERN M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature.* 2003; 422: 633–637. <https://doi.org/10.1038/nature01546> PMID: 12687005
24. Lehner B. Conflict between noise and plasticity in yeast. *PLoS Genet.* 2010; 6: e1001185. <https://doi.org/10.1371/journal.pgen.1001185> PMID: 21079670
25. Chapal M, Mintzer S, Brodsky S, Carmi M, Barkai N. Resolving noise-control conflict by gene duplication. *PLoS Biol.* 2019; 17(11): e3000289. <https://doi.org/10.1371/journal.pbio.3000289>
26. Moriya H, Shimizu-Yoshida Y, Kitano H. In vivo robustness analysis of cell division cycle genes in *Saccharomyces cerevisiae*. *PLoS Genet.* 2006; 2: e111. <https://doi.org/10.1371/journal.pgen.0020111> PMID: 16839182
27. Kuang Z, Pinglay S, Ji H, Boeke JD. Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth. *Elife.* 2017; 6. <https://doi.org/10.7554/eLife.29938> PMID: 28949295
28. Storz JF. Genome evolution: gene duplication and the resolution of adaptive conflict. *Heredity.* 2009; 102: 99–100. <https://doi.org/10.1038/hdy.2008.114> PMID: 18971957
29. Piatigorsky J. The recruitment of crystallins: new functions precede gene duplication. *Science.* 1991; 252: 1078–1079. <https://doi.org/10.1126/science.252.5009.1078> PMID: 2031181
30. Ohno S. *Evolution by Gene Duplication.* Berlin: Springer Berlin Heidelberg; 2014.
31. Vanneste K, Maere S, Van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos Trans R Soc Lond B Biol Sci.* 2014; 369. <https://doi.org/10.1098/rstb.2013.0353> PMID: 24958926
32. del Pozo JC, Ramirez-Parra E. Whole genome duplications in plants: an overview from Arabidopsis. *J Exp Bot.* 2015; 66: 6991–7003. <https://doi.org/10.1093/jxb/erv432> PMID: 26417017

33. Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature*. 2003; 424: 194–197. <https://doi.org/10.1038/nature01771> PMID: 12853957
34. Qian W, Zhang J. Gene dosage and gene duplicability. *Genetics*. 2008; 179: 2319–2324. <https://doi.org/10.1534/genetics.108.090936> PMID: 18689880
35. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A*. 2010; 107: 9270–9274. <https://doi.org/10.1073/pnas.0914697107> PMID: 20439718
36. McLysaght A, Makino T, Grayton HM, Tropeano M, Mitchell KJ, Vassos E, et al. Ohnologs are overrepresented in pathogenic copy number mutations. *Proc Natl Acad Sci U S A*. 2014; 111: 361–366. <https://doi.org/10.1073/pnas.1309324111> PMID: 24368850