New Phytologist

Review

Research review

Tackling redundancy: genetic mechanisms underlying paralog compensation in plants

Author for correspondence:
David Jackson
Email: jacksond@cshl.edu

Received: *17 May 2023*
Accepted: *29 August 2023*

Sessen Daniel Iohannes[1,2] **and David Jackson**[2]

[1]School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724 NY, USA; [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, 11724 NY, USA

New Phytologist (2023)
doi: 10.1111/nph.19267

Key words: dosage, evolution, gene duplication, paralog compensation, redundancy, responsive backup circuits.

Summary

Gene duplication is a powerful source of biological innovation giving rise to paralogous genes that undergo diverse fates. Redundancy between paralogous genes is an intriguing outcome of duplicate gene evolution, and its maintenance over evolutionary time has long been considered a paradox. Redundancy can also be dubbed 'a geneticist's nightmare': It hinders the predictability of genome editing outcomes and limits our ability to link genotypes to phenotypes. Genetic studies in yeast and plants have suggested that the ability of ancient redundant duplicates to compensate for dosage perturbations resulting from a loss of function depends on the reprogramming of gene expression, a phenomenon known as active compensation. Starting from considerations on the stoichiometric constraints that drive the evolutionary stability of redundancy, this review aims to provide insights into the mechanisms of active compensation between duplicates that could be targeted for breaking paralog dependencies – the next frontier in plant functional studies.

Introduction

How biological innovations arise at the molecular, cellular, and organismal levels is a question that has long fascinated evolutionary and molecular biologists. The conventional understanding is that new functional proteins are often obtained by co-opting existing genes to perform new tasks, a concept known as molecular tinkering (Jacob, 1977). Probably, the best illustration of this tinkering process is gene duplication (Jacob, 1977). Gene duplication is a powerful source of genetic novelty and provides new substrates for evolution. Gene duplications have expanded the regulatory gene repertoire (Maere *et al.*, 2005), facilitated new ecological interactions among species and between hosts and pathogens (Panchy *et al.*, 2016), and increased the morphospace that allowed organisms to colonize new habitats (van de Peer *et al.*, 2009). Several mechanisms can give rise to duplicated genes, including transposition, tandem gene duplication, and whole-genome duplication (WGD), with the latter being the primary source of gene duplicates in plants (Wendel, 2000). For instance, WGDs have been suggested to explain the incredible radiation experienced by flowering plants in the early Cretaceous (Wendel, 2000; Jiao *et al.*, 2011).

The versatility of gene duplicates (hereafter, paralogs) arises from their diverse fates acquired throughout evolution. The central paradigm in the theory of duplicate evolution posits that the ultimate fate of a paralog is either to accumulate deleterious mutations so that it becomes nonfunctional or to acquire beneficial mutations that enable it to develop novel functions (Ohno, 1970). Despite the various modes of divergent evolution among paralogs, several duplicated genes persist in the genome with some degrees of functional overlap (Force *et al.*, 1999). Paralogs with overlapping functions are referred to as 'redundant'. This functional overlap enables a gene to serve as a backup copy of its paralog, and to compensate for its function in case of disruption.

Paralog redundancy also greatly limits our ability to link genotypes to phenotypes. The different evolutionary histories of paralogs might cause single-gene mutant loss-of-function phenotypes in one genetic background or species, but not in related backgrounds or species (Vaddepalli *et al.*, 2019; Kwon *et al.*, 2022). In polyploid species with more than two sets of chromosomes, redundant genes on homeologous chromosomes further complicate functional studies, and efficient genome editing usually requires the simultaneous targeting of multiple loci (Botella, 2019). Paralog dependencies also complicate inferences of cell-type

conservation across species using single-cell analysis. These inferences are often conducted by assessing expression patterns of one-to-one orthologous marker genes while disregarding paralog relationships that may be drivers of cell type divergence (Kajala *et al.*, 2021; Guillotin *et al.*, 2023). On top of this, redundancy often depends on specific biological contexts, such as developmental stage and/or cell type, which further complicates the prediction of paralog relationships (Ewen-Campen *et al.*, 2017).

Starting from considerations on the evolutionary stability of redundancy, this review aims to address the redundancy paradoxes and provide insights into mechanisms of paralog compensation that could be targeted for breaking paralog dependencies.

Redundancy in the light of evolution

A compelling paradox in evolutionary genetics is the stability of genetic redundancy. In general, redundancy is prone to rapid degradation due to the accumulation of mutations in duplicated genes and their regulatory regions (Force *et al.*, 1999). Indeed, paralogs tend to functionally diverge (neofunctionalization) or revert to a singleton state (nonfunctionalization) over evolutionary time. However, comparative analyses of duplicate gene retention in Angiosperms revealed that gene families that are involved in gene regulation, signal transduction, and metabolic processes are enriched for duplicates from ancient whole-genome duplication events that occurred at least 75 million yr ago (Ma) (Li *et al.*, 2015; Jia *et al.*, 2023). These genes tend to be dosage-sensitive, meaning their function depends on stoichiometric relationships between products (Birchler *et al.*, 2005). Defoort *et al.* (2019) found that several retained WGD were involved in protein–protein interactions, and diverged more slowly than paralogs arising from segmental duplications. Protein–protein interactions require optimal stoichiometries among interacting partners, so the enrichment of redundant genes encoding interacting proteins again reflects the role of dosage in maintaining redundancy. Additionally, the abundance of retained duplicates arising from WGD compared with the more nonfunctionalization-prone tandem duplications and transpositions may reflect the role of *cis*-regulatory regions in redundancy. Indeed, WGD events lead to greater retention of *cis*-regulatory elements than small-scale duplications (Arsovski *et al.*, 2015). In grasses, for example, paralogs arising from WGD events with a higher number of conserved noncoding sequences (CNSs) are more likely to be retained as a duplicate pair (Schnable *et al.*, 2011). These observations further confirm two long-standing concepts for addressing the redundancy paradox: (1) dosage balance is a powerful force to maintain redundancy between duplicate genes; (2) *cis*-regulatory control is a cornerstone of dosage balance, and hence, redundancy. In the next sections, we will delve into these two aspects and attempt to explain why the redundancy paradox is less clear than it seems.

Dosage balance and compensatory drift drive the stability of redundancy

To a first, quite misleading, interpretation, it might appear that redundancy has been selected to maintain mutational robustness. This is fundamentally not the case, as the selective advantage to maintaining redundancy in natural populations, where nonfunctionalizing mutations are very rare, is too small (Pires & Conant, 2016). Instead of selection for robustness, the retention of specific duplicates over evolutionary time can be explained by the gene dosage hypothesis (Birchler *et al.*, 2001, 2005). Stoichiometric ratios of gene products allow efficient function, and deviations from this balance lead to aberrant phenotypes. Stoichiometry can be controlled by adjusting mRNA or protein production and degradation rates to maintain steady-state expression levels. How is gene dosage involved in maintaining redundancy? During paralog evolution, duplicates often accumulate mutations that reduce their expression to levels insufficient to perform a function, a process known as hypofunctionalization (Qian *et al.*, 2010; Veitia, 2017; Birchler & Yang, 2022) (Fig. 1b). For duplicates with dosage-sensitive interactions with other genes, a steady state is reached when the duplicates are in dosage balance, that is when they reach the optimal combined expression necessary for function (Fig. 1c). In this scenario, there are negative fitness consequences when one duplicate is deleted and the stoichiometric ratio of the paralogs and their interactors is altered, so selection constraints act to maintain dosage-sensitive redundant duplicates (Birchler & Yang, 2022). Conversely, if the expression of just one duplicate is sufficient for function, mutations in one copy will not be selected against and nonfunctionalization of the copy will occur rapidly.

Dosage balance often implies that the duplicates retain similar expression levels. Yet, stoichiometric constraints can also be met if one paralog drifts to lower expression levels, while the other gains a higher expression that enables the pair to maintain constant total expression (Thompson *et al.*, 2016; Birchler & Yang, 2022). This phenomenon is known as compensatory drift (Fig. 1d).

Dosage balance and compensatory drift are a further refinement of subfunctionalization models that explain the retention of paralog redundancy, such as the duplication–degeneration–complementation (DDC) model (Table 1). According to the DDC model, and several subsequent models, ancestral functions need to be partitioned between paralogs in order to preserve redundancy (Force *et al.*, 1999; Birchler *et al.*, 2005; Qian *et al.*, 2010; Gout & Lynch, 2015). This includes the partitioning of stoichiometric requirements between paralogs, also termed 'quantitative subfunctionalization' (Gout & Lynch, 2015). The stable conservation of genetic redundancy over long evolutionary time could also be explained by the piggyback model (Vavouri *et al.*, 2008). This model posits that redundancy can be maintained if each duplicate retains functional overlap while developing nonoverlapping functions, such that mutations will act on both functions (Vavouri *et al.*, 2008). The piggyback model is an alternative version of the redundancy–pleiotropy model first proposed by Nowak *et al.* (1997), and it is similar to the structural and functional entanglement model recently developed by Kuzmin *et al.* (2020) (Nowak *et al.*, 1997; Table 1). In the structural and functional entanglement model, intermediate levels of structural and functional constraints allow duplicated genes to retain functional overlap while expanding some nonoverlapping functions (Kuzmin *et al.*, 2020, 2022). This trade-off between overlap and

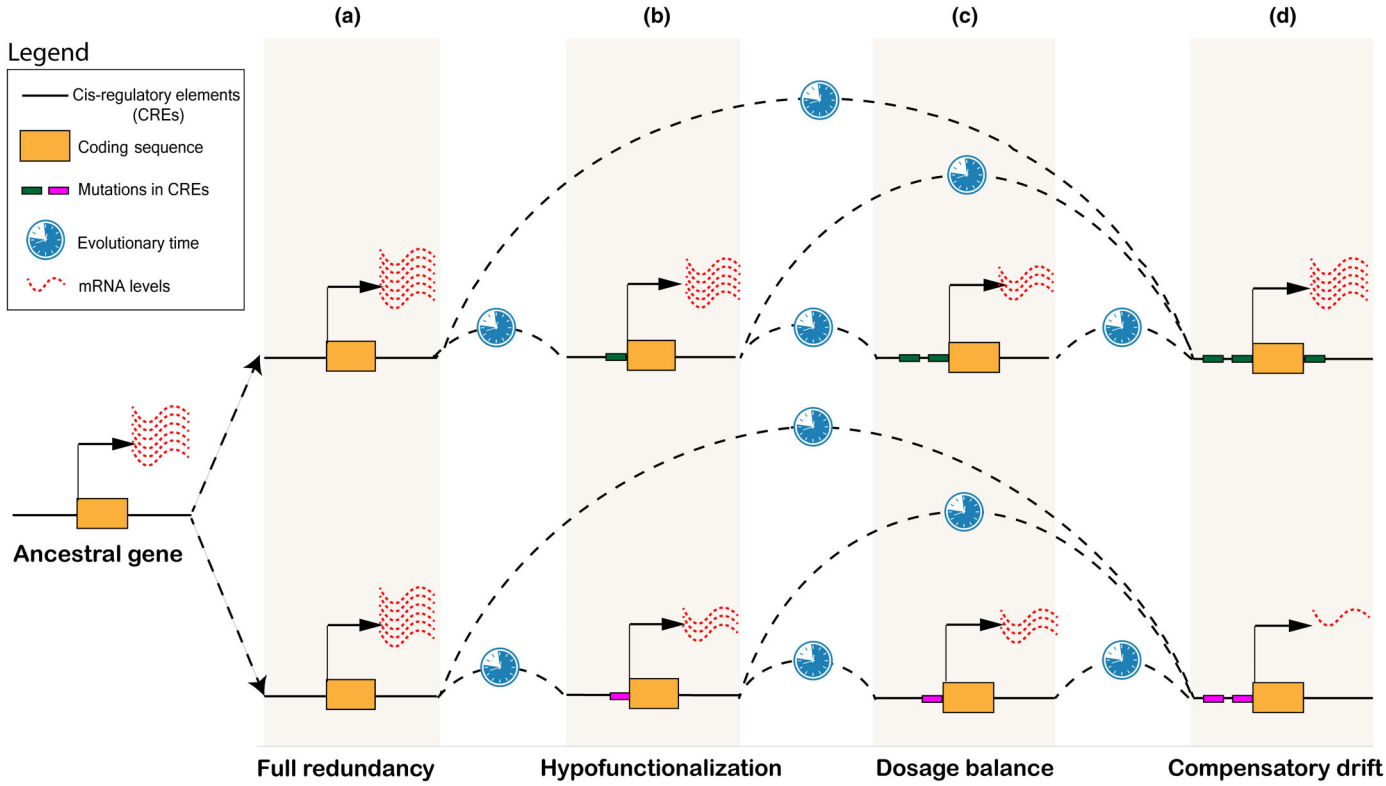

Fig. 1 Model for the evolution of genetic redundancy. (a) Immediately following a duplication event, duplicates are likely to be fully redundant. (b) Accumulation of mutations in *cis*-regulatory elements (green and pink bars) might initially lead to hypofunctionalization, in which duplicates drift to reduced expression levels that are insufficient for function. (c) For paralogs involved in dosage-sensitive interactions with other genes, a steady state is reached when duplicates are subjected to dosage balance constraints: both paralogs are 'stoichiometrically' needed to perform the function. Loss of function of either paralog would lead to negative fitness consequences, so selection constraints act to maintain redundancy. However, dosage balance eventually decays over time. (d) Dosage balance constraints may also act to maintain redundancy in the event of compensatory drift, where one paralog drifts to lower expression levels and the other compensates by evolving higher expression, while the pair maintains constant total expression. Compensatory drift can result from dosage balance decay or be an independent outcome of redundancy evolution. Red dashed lines represent mRNA levels, as in the legend. Black dashed lines just indicate the following step, so they are not significant.

subfunctionalization enables paralogs to reach an evolutionary steady state (Kuzmin *et al.*, 2020, 2022).

All of these models converge into a unified evolutionary framework that establishes a key role of dosage balance and compensatory drift in the evolutionary stability of redundancy. Yet, compensatory drift still represents a paradox. After sustained divergence in expression levels, the lower expressed paralog could lose its function, and thus compensatory drift could be considered a path toward nonfunctionalization (Thompson *et al.*, 2016). There is a workaround to this paradox. In the presence of higher dosage-balance constraints and larger population sizes, paralogs undergoing compensatory drift could be selected for new beneficial functions (Thompson *et al.*, 2016). These include the ability to reprogram their expression to compensate for dosage and limit gene expression noise in response to paralog loss. This reprogramming and its regulation are discussed below.

Cis-regulatory control is the cornerstone of paralog redundancy

A tempting first hypothesis to explain redundancy is that paralogs must undergo similar regulation. This seems to hold for recent paralogs. For example, in recent yeast paralogs, a high overlap in promoter motifs correlates with higher expression similarity and ability to be redundant (Kafri *et al.*, 2005). Similarly, for recent paralogs in maize, the proportion of shared CNSs is positively correlated with their expression similarity (Song *et al.*, 2021). This is in line with the fact that recent duplicates are more likely to be fully redundant. However, almost paradoxically, having only a partial overlap in promoter motifs is a strong predictor of redundancy between ancient paralogs (Kafri *et al.*, 2005). This result demonstrates one of the corollaries of the DDC model, namely that degenerative mutations in regulatory elements can increase rather than reduce the probability of duplicate gene preservation (Force *et al.*, 1999). Ancient paralogs undergo extensive rearrangements in their *cis*-regulatory regions, so it is not surprising that these are only partially conserved. What is more surprising is how ancient paralogs are still co-expressed and maintain some degree of redundancy, despite a seemingly divergent *cis*-regulatory control. There are two possible explanations for this paradox. First, partial conservation of regulatory motifs could enable paralogs to partition dosage requirements, hence leading to dosage balance or compensatory drift (Fig. 1). Alternatively, these partially shared motifs could allow duplicates to reprogram their

Table 1 Theoretical models explaining the maintenance of redundancy.

Concept	Model	Definition	Reference
Subfunctionalization	The redundancy–pleiotropy model	Redundancy occurs only with respect to a given overlapping function, while the genes are maintained by selection because of another nonoverlapping function.	Nowak *et al*. (1997)
	The duplication–degeneration–complementation model	Redundancy is maintained if there is a partitioning of ancestral functions between the paralogs.	Force *et al*. (1999)
	The piggyback model	Redundancy is maintained if duplicates have overlapping and nonoverlapping functions that are coselected.	Vavouri *et al*. (2008)
	The structural–functional entanglement model	Redundancy is maintained by intermediate levels of structural and functional constraints.	Kuzmin *et al*. (2020)
Dosage constraints	The dosage balance model	Redundancy is maintained if there is a partitioning of dosage requirements between the paralogs.	Birchler *et al*. (2005)
	The expression reduction model	Redundancy is maintained by dosage constraints resulting from expression reduction of each duplicate.	Qian *et al*. (2010)
	The absolute-dosage subfunctionalization model	Redundancy is maintained if there is quantitative subfunctionalization between paralogs.	Gout & Lynch (2015)

gene expression to buffer perturbations in dosage. This exciting hypothesis was originally proposed in yeast by Kafri *et al*. (2005, 2006) and recently demonstrated in Drosophila (Loker & Mann, 2022) and Arabidopsis (Ye *et al*., 2016). This reprogramming is what we most commonly refer to as 'active compensation'.

The genetics of compensation in plants

The various phenotypic outcomes of genetic redundancy can be explained in terms of dosage balance and compensatory drift. We often think of genetic redundancy as full redundancy, in which mutants of individual paralogs have no phenotype. However, ancient redundant paralogs are more likely to exhibit partial or unequal redundancy. Partial redundancy is usually an outcome of dosage balance between two duplicates, where each single mutant has a milder phenotype that is enhanced in the double mutant (Fig. 2). Several plant transcription factor genes show partial redundancy. These include the APETALA2 (AP2) class transcription factors *PLETHORA1* and *PLETHORA2* that control stem cell maintenance in the root (Aida *et al*., 2004), the bHLH proteins *GLABRA3* and *ENHANCER OF GLABRA3* that specify root epidermal cell fates (Bernhardt *et al*., 2003), and the MADS-box genes *AGAMOUS*, *SHATTERPROOF1*, *SHATTERPROOF2*, and *SEEDSTICK* that control carpel and ovule development (Pinyopich *et al*., 2003). In contrast to partial redundancy, in unequal redundancy, a mutation in one paralog has a phenotype, while mutations in the other paralog do not. The mutant phenotype is again enhanced in the double mutant (Fig. 2). Paralogs in compensatory drift are more likely to display unequal redundancy, as highlighted by several classical examples from Arabidopsis. These include unequal redundancy between the MADS-box transcription factors *APETALA1* and *CAULIFLOWER* that control the formation of floral meristems (Bowman *et al*., 1993; Kempin *et al*., 1995), the leucine-rich repeat receptor-like kinases *ERECTA* and *ERECTA LIKE1*/*ERECTA LIKE2* that control organ growth and development (Shpak *et al*., 2004), and the MYB-related transcription factors *LONG HYPOCOTYL5* and *HY5 HOMOLOG* involved in light signaling (Holm *et al*., 2002; Briggs *et al*., 2006). The numerous cases of partial and unequal redundancy suggest that these are evolutionary stable states for ancient paralogs (Briggs *et al*., 2006).

At a molecular level, redundancy often results from compensation mechanisms between paralogs. The most well-characterized of these, called passive compensation, occurs when a gene does not change its expression pattern in response to a loss of its paralog (Diss *et al*., 2014) (Fig. 2a). Passive compensation often underlies full redundancy between recent paralogs, the initial state of paralog evolution (Diss *et al*., 2014) (Fig. 2a). However, genes that exhibit passive compensation could also become partially redundant in the case of dosage balance or subfunctionalization between two paralogs (Fig. 2a,b). In this case, the loss of one or the other copy without additional compensation would lead to mutant phenotypes in either single mutant, which is then enhanced in the double mutant. An example is the E-class *SEPALLATA* genes in the early-diverging dicot *Thalictrum thalictroides*. Within this family, the paralogs have subfunctionalized for determining sepal and stamen identity, do not compensate for one another, and exhibit partial redundancy (Soza *et al*., 2016). Another interesting example of partial redundancy driven by passive compensation is represented by the genes encoding for the peptide ligands FON2-LIKE CLE PROTEIN1 (FCP1) and CLAVATA3/ESR-RELATED7 (CLE7) controlling stem cell proliferation in maize (Rodriguez-Leal *et al*., 2019). In both *fcp1* and *cle7* single mutants, inflorescence meristems are increased in size (fasciated) compared with the wild-type, and the double mutant is additive (Fig. 2b). Both *FCP1* and *CLE7* are upregulated in the *cle7* single mutant compared to the wild-type (Rodriguez-Leal *et al*., 2019). While this would appear to show a reprogramming of both paralogs, the apparent change in expression is likely due to the sampling of overproliferating stem cells in the larger meristems from the *cle7* mutant fasciated ears. The two paralogs thus lack an active compensation mechanism (Rodriguez-Leal *et al*., 2019).

Passive compensation could lead to unequal redundancy in the case of compensatory drift, where one paralog is intrinsically more highly expressed and the loss of its duplicate alone does not lead to any apparent phenotype (Fig. 2a). This can be illustrated by the cell

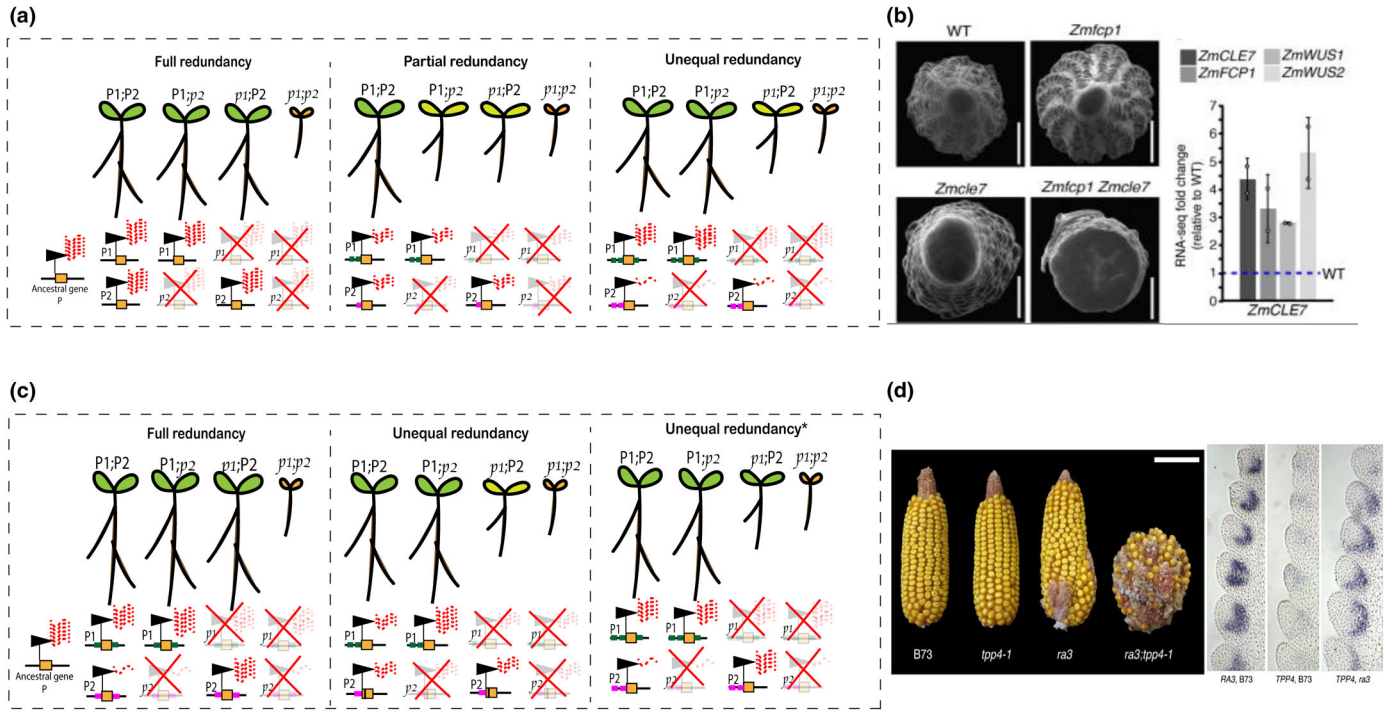

Fig. 2 The genetics of compensation. The different outcomes of redundancy can be explained in terms of passive (a, b) or active (c, d) compensation mechanisms between duplicates. RNA levels are indicated with red dashed lines, and mutations in *cis*-regulatory elements are indicated with green and pink colored bars as in **Fig. 1**. (a) In passive compensation, a gene is not reprogrammed upon genetic perturbation of its duplicate. Under this scenario, recent paralogs are more likely to display full redundancy, paralogs in dosage balance are more likely to display partial redundancy, and paralogs in compensatory drift are more likely to display unequal redundancy. (b) Example of passive compensation leading to partial redundancy between *FCP1* and *CLE7* in maize. (Left) Both *fcp1* and *cle7* single mutants develop larger (fasciated) inflorescence meristems, and this phenotype is additively enhanced in the double mutant. (Right) Both *FCP1* and *CLE7* are upregulated in the *cle7* single mutant, but this is most likely due to sampling of larger meristems in the fasciated ear of the *cle7* mutant. *FCP1* and *CLE7* thus lack active compensation. Bars, 500 μm. Adapted from Rodriguez-Leal *et al*. (2019). (c) Active compensation is a reprogramming mechanism where the expression level or pattern of a gene changes to buffer genetic perturbations of its paralog. This type of compensation can lead to full redundancy for paralogs in compensatory drift that compensate each other, to unequal redundancy for paralogs that are in dosage balance, or to an alternate state of unequal redundancy (*) for paralogs in compensatory drift where only one of the two compensates. Mutations in *cis*-regulatory elements are indicated as colored bars. (d) Example of active compensation leading to unequal redundancy between *RA3* and *TPP4* in maize. (Left) The *ra3* knockout leads to ectopic branching in maize ears. The tpp4 single mutant does not have a phenotype, but the branching phenotype is significantly enhanced in the *ra3;tpp4* double mutant. (Right) The *RA3* transcript is expressed in a region subtending developing spikelet pair meristems. *TPP4* is expressed in the same domain at a much lower level, and its expression is upregulated approximately twofold in the *ra3* background. Bars, 1 mM. Adapted from Claeys *et al*. (2019).

wall-localized glucanase ZERZAUST (ZET) and its homolog ZERZAUST HOMOLOG (ZETH) in the Landsberg *erecta* (L*er*) accession of *Arabidopsis thaliana* (Vaddepalli *et al*., 2019). In this system, only the *zet* mutant has a phenotype, and it is enhanced in the *zet*;*zeth* double mutant. *ZETH* is poorly expressed in flowers and other tissues compared with *ZET*, and its expression levels are not altered in the *zet* mutant (Vaddepalli *et al*., 2019).

Another largely understudied but common redundancy mechanism is active compensation. Active compensation occurs when a gene is upregulated to buffer against the loss of its paralog (Diss *et al*., 2014) (Fig. 2c). This phenomenon can lead to full redundancy between paralogs that have drifted in expression levels and both undergo compensatory upregulation, such as in the case of the aforementioned *ZET* and *ZETH* paralog pair in a different *Arabidopsis thaliana* accession, Columbia (Col) (Fig. 2c). In Col, *ZETH* expression levels are higher than L*er*. However, unlike in L*er*, *ZET* and *ZETH* expression is upregulated in the Col *zeth* and *zet* single mutants, respectively. This leads to a strong phenotype only in the double mutants, and no phenotype in the single mutants in the Col background (Vaddepalli *et al*., 2019). Importantly, this example illustrates how active compensation can be genetic background-dependent, adding yet another confounding factor to the predictability of functional studies.

When the reprogramming of gene expression occurs asymmetrically between paralogs that are in dosage balance or have similar expression levels, it can lead to unequal redundancy (Fig. 2c). This can be illustrated by the AUXIN RESPONSE FACTOR19 and its homolog NONPHOTOTROPIC HYPOCOTYL4 (Okushima *et al*., 2005; Li *et al*., 2006). Active compensation can also lead to an alternate state of unequal redundancy for paralogs that have subfunctionalized or are in compensatory drift, as in the case of the trehalose-6-phosphate phosphatases RAMOSA3 (RA3) and TREHALOSE 6 PHOSPHATE PHOSPHATASE 4 (TPP4) that control branching in the maize inflorescence (Claeys *et al*., 2019) (Fig. 2c,d).

Active compensation is observed across distantly related eukaryotes, including yeast and angiosperms, but it is not yet clear whether it was actively selected to minimize noise in gene

expression or is just a side effect of subfunctionalization (Kafri *et al.*, 2006; Diss *et al.*, 2014; Pires & Conant, 2016). Transcriptional regulatory circuits have been proposed as mechanisms for reducing noise in gene expression, and reprogramming mechanisms among duplicates may have evolved to keep gene expression levels within appropriate limits (Raser & O'Shea, 2005; Pires & Conant, 2016). Whether selected or not, the mechanisms through which paralogs are able to actively compensate for each other are still poorly understood, and they could become an intriguing target for breaking paralog dependencies.

Responsive backup circuits (RBCs) are at the basis of active compensation

How can a gene be reprogrammed upon mutation of its paralog? Studies in different model systems point to responsive backup circuits (RBC) acting at the gene or protein levels as a mechanism underlying active compensation (Fig. 3). An excellent example of a transcriptional RBC is represented by the *Arabidopsis thaliana APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*) genes, two paralog MADS-box genes whose expression levels diverged in the floral primordia and developing sepals and petals (Bowman *et al.*, 1993). In the *ap1* single mutant, flower meristems are partially converted into inflorescence meristems (Bowman *et al.*, 1993; Kempin *et al.*, 1995). This phenotype is significantly enhanced in the *ap1;cal* double mutant (Bowman *et al.*, 1993; Kempin *et al.*, 1995). *AP1* expression is significantly lower in the *ap1;cal* double mutant, but not in the *ap1* single mutant, leading to the hypothesis that *AP1* is positively regulated by *CAL* (Bowman *et al.*, 1993). Indeed, Ye and colleagues demonstrated that CAL binds to a CArG transcription factor binding site found in the promoter of *AP1*, but not in its paralog (Ye *et al.*, 2016).

While this example illustrates a positive RBC between two paralogs, negative circuits underlying paralog compensation have been described in other systems. A recent example is that of the Drosophila *NUBBIN* (*NUB*, also known as *PDM1*) and *PDM2* genes that encode for POU-type homeodomain transcription factors (Loker & Mann, 2022). These genes arose from an ancient tandem duplication event *c.* 200 Ma and are expressed at different levels in the wing progenitor cells. *NUB* represses the expression of *PDM2* through a silencer element in the *PDM2* promoter. In the absence of *NUB*, *PDM2* is upregulated and compensates for its paralog (Loker & Mann, 2022). These examples illustrate how *cis*-regulatory elements can serve as sites for reprogramming between paralogs (Fig. 3a).

Transcriptional circuits depend not only on sequence specificity of transcriptional factors but also on their occupancy of DNA binding sites. The occupancy of transcription factors is also influenced by chromatin architecture, which can affect transcription factor binding by nucleosome remodeling and spatial partitioning of regulatory motifs and factors into topological regions of the genome (Li *et al.*, 2012). Recent work by *Levo et al.*

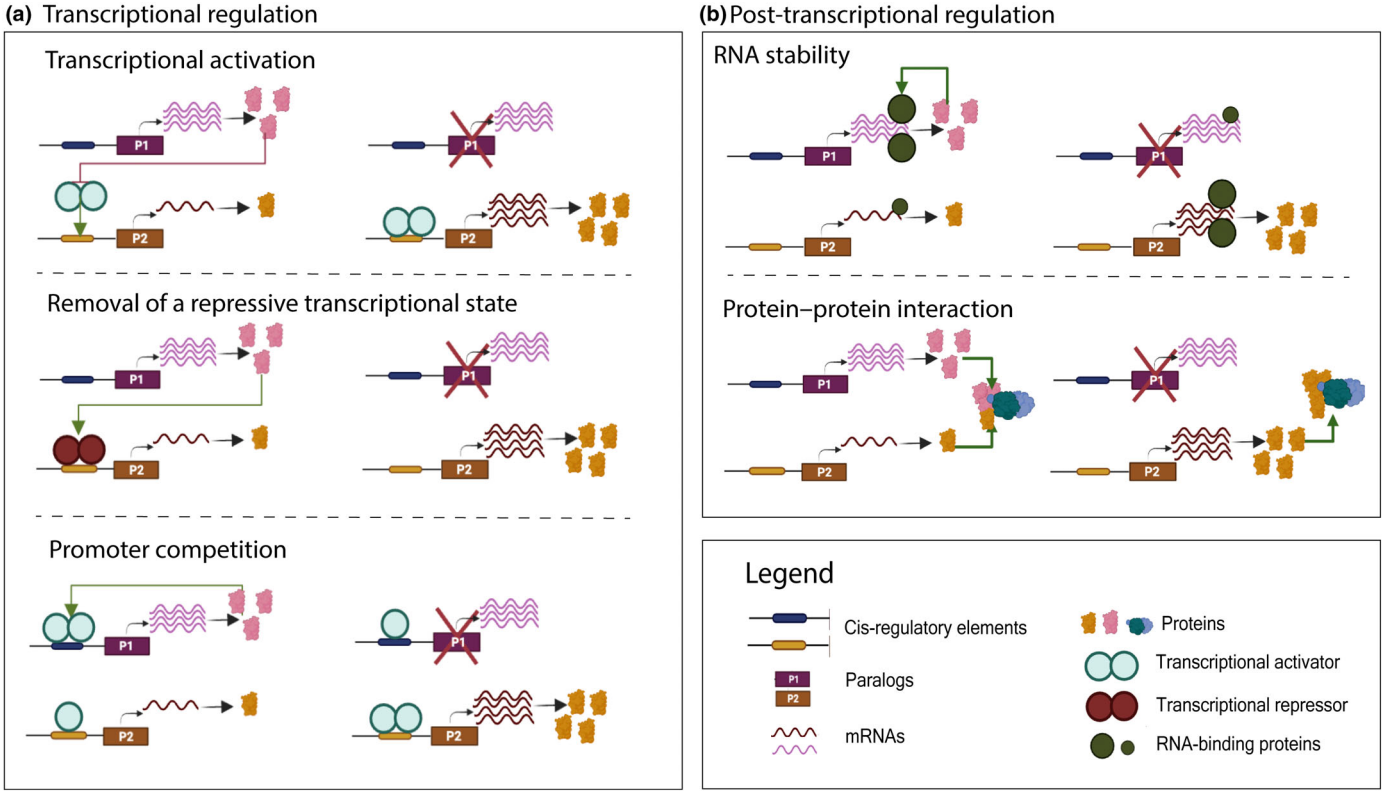

Fig. 3 Hypothetical responsive backup circuits underpinning active compensation. Active compensation could be controlled at the transcriptional (a) or post-transcriptional (b) levels. (a) Transcriptional regulation could involve direct transcriptional activation, removal of a repressive state, or promoter competition for transcriptional resources. (b) Post-transcriptional regulation could involve competition for stabilizing RNA-binding proteins, or reprogramming of protein–protein interactions to compensate for the loss of a binding partner.

showed that >50% of paralogous genes expressed in the Drosophila embryo form focal contacts through shared enhancers and tethering elements (Levo *et al.*, 2022). This spatial association enables the formation of a cotranscriptional hub that allows the coordinated expression of paralogous genes (Levo *et al.*, 2022). Interestingly, the study suggested that active compensation between paralogs could be explained by promoter competition for shared but limiting transcriptional resources within this common hub (Levo *et al.*, 2022). Promoter competition involves the exclusion of transcription factors from certain promoters due to a higher affinity for competing promoters. This pattern of transcriptional interference has also been described for paralogous B-globin genes and the mammalian Hox genes (Conte *et al.*, 2002). Under this model, perturbations in transcriptional cofactors or in the *cis*-regulatory regions of a paralog could make transcriptional resources more available to its duplicate and lead to compensation (Fig. 3a).

Transcription might not be the only regulatory mechanism underlying compensation. Studies in yeast and zebrafish suggest a role for mRNA stability in active compensation (El-Brolosy & Stainier, 2017). As functionally related mRNAs tend to be coregulated by the same RNA Binding Proteins (RBPs), compensation could lie in the stabilization of a transcript by RBPs following the knockout of a related gene (Keene, 2007; El-Brolosy & Stainier, 2017). Under this model, if a mutant mRNA undergoes nonsense-mediated decay, RBPs normally acting on the wild-type version of the mRNA become more available to stabilize the compensating paralog's mRNA (Fig. 3b). Another post-transcriptional mechanism through which compensation could be controlled is protein–protein interaction, which may be common for paralogs encoding proteins that form complexes (Diss *et al.*, 2014) (Fig. 3b). The yeast nuclear pore complex is a great example of paralogous compensation at the protein level. A protein interactome study conducted by Diss *et al.* (2013) showed that the robustness of the nuclear pore complex to the deletion of some of its subunits was due to redundant paralogs that were able to form new protein–protein interactions to compensate for the absence of their duplicate. The transcriptional and post-transcriptional mechanisms of active compensation could become exciting targets for overcoming the redundancy barrier in functional studies.

Active compensation as a target to break the redundancy barrier

Active compensation is an intriguing outcome of the evolution of redundancy. It is a mechanism that enables paralogs that have subfunctionalized to back up each other, thereby limiting the effects of gene expression noise. Breaking active compensation among redundant duplicates has the potential to overcome the redundancy barrier and improve the predictability of functional studies in plants. This strategy must start with the identification of redundancy and compensation patterns. Genome editing technologies now allow for the implementation of multiplex loss-of-function studies (Ewen-Campen *et al.*, 2017; Hu *et al.*, 2023), which can then inform us about the redundant relationships among paralogs. A genome-wide application of this approach has recently been demonstrated in *Arabidopsis thaliana* (Hu *et al.*, 2023). However, its implementation in other plant species must first address the challenges associated with large-scale transformation efforts but could become a new frontier in plant functional studies. Furthermore, cell-type-specific omics in multiple loss-of-function backgrounds might allow the identification of the underlying active compensation mechanisms between duplicates (Xu & Jackson, 2023). Indeed, as redundancy is often cell type-specific (Ewen-Campen *et al.*, 2017), stoichiometric constraints are also more likely to vary between cell types. Paralogs could be in dosage balance or compensatory drift and display active compensation in one cell type, but not in others. While this issue has been difficult to address, single-cell analyses may add some insight into the context dependence of this phenomenon.

Once the paralog dependency is identified, active compensation mechanisms can be targeted. For instance, CRISPR targeting *cis*-regulatory elements could enable us to break the *cis*-regulatory control of compensation, potentially achieving higher expressivity of knock-out mutations. As opposed to multiparalog knockout systems, which often lead to lethality or pleiotropic phenotypes, this approach could also enable us to generate *cis*-regulatory alleles that enable fine-tuning of dosage and provide novel quantitative variation. Although CRISPR editing of promoters is a key method to introduce quantitative trait variation in crops, its application is hindered by the lack of predictability of phenotypic outcomes of *cis*-regulatory alleles (Rodríguez-Leal *et al.*, 2017; Liu *et al.*, 2021; Song *et al.*, 2022). On the contrary, targeting *cis*-regulatory elements that might control active compensation in loss-of-function backgrounds, such as those identified through chromatin accessibility, transcription factor binding, and sequence conservation, could enable us to elucidate the nonlinear relationships between expression changes and resulting phenotypes.

Conclusion

Here, we reviewed theoretical models and genetic studies aimed at addressing one of the greatest puzzles of evolutionary genetics: the maintenance of redundancy between ancient duplicate genes despite the accumulation of mutations. The consensus in the field posits that the maintenance of redundancy derives from selection constraints on dosage. A more fascinating paradox is how redundancy can impart robustness to biological systems despite the stoichiometric dependencies conferred by such dosage constraints. One potential solution is active compensation. Active compensation enables paralogs that have subfunctionalized to buffer dosage perturbations and limit noise in gene expression, therefore potentially solving the robustness/fragility conundrum. Active compensation could also underlie partial and unequal redundancy and the nonlinear control of quantitative trait variation. Although the nature of active compensation among paralogs is still poorly understood, studies in yeast and plants have suggested transcriptional and post-transcriptional responsive backup circuits as mechanisms underlying the phenomenon. These mechanisms could become exciting targets for fine-tuning dosage to engineer quantitative variation and overcome the redundancy barrier in functional studies. In turn, this could enable

us to improve the predictability of genome editing outcomes and inferences about cell-type homologies across genetic backgrounds and species.

Acknowledgements

This work was supported by the National Science Foundation, USA (NSF IOS-2129189). We thank members of the Jackson Laboratory for their helpful feedback. We are grateful to the reviewers and the editor for providing valuable comments that improved the manuscript.

Competing interests

None declared.

ORCID

Sessen Daniel Iohannes https://orcid.org/0000-0002-7831-8997

References

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B. 2004. The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell* **119**: 109–120.

Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL. 2015. Evolution of *cis*-regulatory elements and regulatory networks in duplicated genes of arabidopsis. *Plant Physiology* **169**: 2982–2991.

Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, Schiefelbein J. 2003. The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. *Development* **130**: 6431–6439.

Birchler JA, Bhadra U, Bhadra MP, Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Developmental Biology* **234**: 275–288.

Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: biological implications. *Trends in Genetics* **21**: 219–226.

Birchler JA, Yang H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell* **34**: 2466–2474.

Botella JR. 2019. Now for the hard ones: is there a limit on CRISPR genome editing in crops? *Journal of Experimental Botany* **70**: 734–737.

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR. 1993. Control of flower development in *Arabidopsis thaliana* by APETALA1 and interacting genes. *Development* **119**: 721–743.

Briggs GC, Osmont KS, Shindo C, Sibout R, Hardtke CS. 2006. Unequal genetic redundancies in Arabidopsis - a neglected phenomenon? *Trends in Plant Science* **11**: 492–498.

Claeys H, Vi SL, Xu X, Satoh-Nagasawa N, Eveland AL, Goldshmidt A, Feil R, Beggs GA, Sakai H, Brennan RG *et al.* 2019. Control of meristem determinacy by trehalose-6-phosphate phosphatases is uncoupled from enzymatic activity. *Nature Plants* **5**: 352–357.

Conte C, Dastugue B, Vaury C. 2002. Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons. *EMBO Journal* **21**: 3908–3916.

Defoort J, van de Peer Y, Carretero-Paulet L, Golding B. 2019. The evolution of gene duplicates in angiosperms and the impact of protein–protein interactions and the mechanism of duplication. *Genome Biology and Evolution* **11**: 2292–2305.

Diss G, Ascencio D, Deluna A, Landry CR. 2014. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **322**: 488–499.

Diss G, Dubé AK, Boutin J, Gagnon-Arsenault I, Landry CR. 2013. A systematic approach for the genetic dissection of protein complexes in living cells. *Cell Reports* **3**: 2155–2167.

El-Brolosy MA, Stainier DYR. 2017. Genetic compensation: a phenomenon in search of mechanisms. *PLoS Genetics* **13**: e1006780.

Ewen-Campen B, Mohr SE, Hu Y, Perrimon N. 2017. Accessing the phenotype gap: enabling systematic investigation of paralog functional complexity with CRISPR. *Developmental Cell* **43**: 6–9.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Gout JF, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Molecular Biology and Evolution* **32**: 2141–2148.

Guillotin B, Rahni R, Passalacqua M, Mohammed MA, Xu X, Raju SK, Ramírez CO, Jackson D, Groen SC, Gillis J *et al.* 2023. A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* **617**: 785–791.

Holm M, Ma LG, Qu LJ, Deng XW. 2002. Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in Arabidopsis. *Genes and Development* **16**: 1247–1259.

Hu Y, Patra P, Pisanty O, Shafir A, Belew ZM, Binenbaum J, Ben Yaakov S, Shi B, Charrier L, Hyams G *et al.* 2023. Multi-Knock—a multi-targeted genome-scale CRISPR toolbox to overcome functional redundancy in plants. *Nature Plants* **2023**: 1–16.

Jacob F. 1977. Evolution and tinkering. *Science* **196**: 1161–1166.

Jia Y, Xu M, Hu H, Chapman B, Watt C, Buerte B, Han N, Zhu M, Bian H, Li C *et al.* 2023. Comparative gene retention analysis in barley, wild emmer, and bread wheat pangenome lines reveals factors affecting gene retention following gene duplication. *BMC Biology* **21**: 1–22.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Kafri R, Bar-Even A, Pilpel Y. 2005. Transcription control reprogramming in genetic backup circuits. *Nature Genetics* **37**: 295–299.

Kafri R, Levy M, Pilpel Y. 2006. The regulatory utilization of genetics redundancy through responsive backup circuits. *Proceedings of the National Academy of Sciences, USA* **103**: 11658.

Kajala K, Gouran M, Shaar-Moshe L, Mason GA, Rodriguez-Medina J, Kawa D, Pauluzzi G, Reynoso M, Canto-Pastor A, Manzano C *et al.* 2021. Innovation, conservation, and repurposing of gene function in root cell type development. *Cell* **184**: 3333–3348.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics* **8**: 533–543.

Kempin SA, Savidge B, Yanofsky MF. 1995. Molecular basis of the cauliflower phenotype in Arabidopsis. *Science* **267**: 522–525.

Kuzmin E, Taylor JS, Boone C. 2022. Retention of duplicated genes in evolution. *Trends in Genetics* **38**: 59–72.

Kuzmin E, Vandersluis B, Ba ANN, Wang W, Koch EN, Usaj M, Khmelinskii A, Usaj MM, Van Leeuwen J, Kraus O *et al.* 2020. Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science* **368**, 1446.

Kwon CT, Tang L, Wang X, Gentile I, Hendelman A, Robitaille G, Van Eck J, Xu C, Lippman ZB. 2022. Dynamic evolution of small signalling peptide compensation in plant stem cell control. *Nature Plants* **8**: 346–355.

Levo M, Raimundo J, Bing XY, Sisco Z, Batut PJ, Ryabichko S, Gregor T, Levine MS. 2022. Transcriptional coupling of distant regulatory genes in living embryos. *Nature* **605**: 754–760.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J *et al.* 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**: 84–98.

Li J, Dai X, Zhao Y. 2006. A role for auxin response factor 19 in auxin and ethylene signaling in Arabidopsis. *Plant Physiology* **140**: 899–908.

Li Z, Defoort J, Tasdighian S, Maere S, Van De Peer Y, De Smet R. 2015. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* **28**: 326–344.

Liu L, Gallagher J, Arevalo ED, Chen R, Skopelitis T, Wu Q, Bartlett M, Jackson D. 2021. Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize CLE genes. *Nature Plants* **7**: 287–294.

Loker R, Mann RS. 2022. Divergent expression of paralogous genes by modification of shared enhancer activity through a promoter-proximal silencer. *Current Biology* **32**: 3545–3555.

Maere S, de Bodt S, Raes J, Casneuf T, van Montagu M, Kuiper M, van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences, USA* **102**: 5454–5459.

Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. *Nature* **388**: 167–171.

Ohno S. 1970. *Evolution by gene duplication*. New York, NY, USA: Springer-Verlag New York.

Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D *et al.* 2005. Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* **17**: 444–463.

Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant Physiology* **171**: 2294–2316.

van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews. Genetics* **10**: 725–732.

Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**: 85–88.

Pires JC, Conant GC. 2016. Robust yet fragile: expression noise, protein misfolding, and gene dosage in the evolution of genomes. *Annual Review of Genetics* **50**: 113–131.

Qian W, Liao BY, Chang AYF, Zhang J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* **26**: 425–430.

Raser JM, O'Shea EK. 2005. Molecular biology – noise in gene expression: origins, consequences, and control. *Science* **309**: 2010–2013.

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**: 470–480.

Rodriguez-Leal D, Xu C, Kwon CT, Soyars C, Demesa-Arevalo E, Man J, Liu L, Lemmon ZH, Jones DS, Van Eck J *et al.* 2019. Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nature Genetics* **51**: 786–792.

Schnable JC, Pedersen BS, Subramaniam S, Freeling M. 2011. Dose-sensitivity, conserved non-coding sequences, and duplicate gene retention through multiple tetraploidies in the grasses. *Frontiers in Plant Science* **2**: 2.

Shpak ED, Berthiaume CT, Hill EJ, Torii KU. 2004. Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. *Development* **131**: 1491–1501.

Song B, Buckler ES, Wang H, Wu Y, Rees E, Kellogg EA, Gates DJ, Khaipho-Burch M, Bradbury PJ, Ross-Ibarra J *et al.* 2021. Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Research* **31**: 1245–1257.

Song X, Meng X, Guo H, Cheng Q, Jing Y, Chen M, Liu G, Wang B, Wang Y, Li J *et al.* 2022. Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nature Biotechnology* **40**: 1403–1411.

Soza VL, Snelson CD, Hewett Hazelton KD, Di Stilio VS. 2016. Partial redundancy and functional specialization of E-class SEPALLATA genes in an early-diverging eudicot. *Developmental Biology* **419**: 143–155.

Thompson A, Zakon HH, Kirkpatrick M. 2016. Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics* **202**: 765–774.

Vaddepalli P, Fulton L, Schneitz K. 2019. Asymmetric redundancy of ZERZAUST and ZERZAUST HOMOLOG in different accessions of *Arabidopsis thaliana*. *G3: Genes, Genomes, Genetics* **9**: 2245–2252.

Vavouri T, Semple JI, Lehner B. 2008. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends in Genetics* **24**: 485–488.

Veitia RA. 2017. Gene duplicates: agents of robustness or fragility? *Trends in Genetics* **33**: 377–379.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**: 225–249.

Xu X, Jackson D. 2023. Single-cell analysis opens a goldmine for plant functional studies. *Current Opinion in Biotechnology* **79**: 102858.

Ye L, Wang B, Zhang W, Shan H, Kong H. 2016. Gains and losses of *cis*-regulatory elements led to divergence of thearabidopsis APETALA1 and CAULIFLOWER duplicate genes in the time, space, and level of expression and regulation of one paralog by the other. *Plant Physiology* **171**: 1055–1069.