Journal of Evolutionary Biology, 2026, **39**, 1192–1199
https://doi.org/10.1093/jeb/voag074
Advance access publication 31 August 2026
Research Article

JOURNAL OF Evolutionary Biology · eseb · OXFORD

Editor's choice

Adaptive deletion of functional duplicate genes in *Drosophila*

Andre Luiz Campelo dos Santos [1,2,*], **Raquel Assis** [1,3,4]

[1]Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, United States
[2]Present address: School of Biological Sciences, University of Oklahoma, Norman, OK 73019, United States
[3]Department of Biomedical Engineering, Florida Atlantic University, Boca Raton, FL 33431, United States
[4]Institute for Human Health and Disease Intervention, Florida Atlantic University, Boca Raton, FL 33431, United States
Handling editor: Rebekah Rogers, Associate editor: Erica Larson
*Corresponding author: Andre Luiz Campelo dos Santos, Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA. Email: andre.santos@ou.edu

Abstract

Gene deletion is traditionally viewed as a nonadaptive mechanism that eliminates functional redundancy, yet emerging evidence indicates that it disproportionately affects tissue-specific duplicates with unique functions. Here, we test whether gene deletion preferentially removes weakly constrained, degenerating duplicates or instead eliminates functionally active duplicates through an adaptive process. To identify the evolutionary and functional factors that determine which duplicates are lost, we systematically analyzed 100 gene deletion events in *Drosophila* by integrating sequence, expression, interaction, and structural data. We uncovered a strong bias toward the loss of younger child copies among functionally unique duplicates, whereas no such bias was observed for redundant duplicates. Contrary to expectations under relaxed constraint, deleted functionally unique genes evolve more slowly, show higher expression, engage in more protein–protein interactions, and do not exhibit elevated structural divergence or intrinsic disorder relative to redundant duplicates. When compared with single-copy genes, deleted functionally unique genes display similar evolutionary rates, slightly lower expression, greater network connectivity, comparable structural divergence, and lower intrinsic disorder. These patterns suggest that deletion frequently affects functionally active rather than degenerate genes. Collectively, our results support the hypothesis that gene deletion in *Drosophila* can represent an adaptive process acting on transiently functional duplicates, potentially driven by either genome streamlining or context-dependent deleterious effects.

Keywords gene deletion, gene duplication, gene loss, evolution, genome streamlining

Introduction

Gene deletion has long been viewed as a largely nonadaptive process that removes redundant duplicate genes (Albalat & Cañestro, 2016; Assis, 2019). However, mounting evidence shows that deletion often affects duplicates with distinct expression profiles and biological functions (Albalat & Cañestro, 2016; Assis, 2019; Campelo dos Santos et al., 2024; Hottes et al., 2013; Kvitek & Sherlock, 2013), suggesting a potential role in adaptation. The "less-is-more" hypothesis posits that deletion can be beneficial when the function of a gene becomes obsolete, deleterious, or incompatible with newly advantageous traits (Olson, 1999). In such cases, deletion may represent an adaptive mechanism for shedding unnecessary or harmful genes—conserving energy, preventing detrimental interactions, or facilitating specialization and niche adaptation (Olson, 1999).

A recent study in *Drosophila* revealed that deleted genes are often tissue-specific, with primary expression in testis and accessory gland male reproductive tissues (Campelo dos Santos et al., 2024). Such genes are also expected to evolve rapidly at the sequence level and engage in relatively few protein-protein interactions, reducing the risks of deleterious pleiotropic effects upon their loss (Birchler & Yang, 2022; DeLuna et al., 2010; Kuzmin et al., 2022). Intriguingly, these same features are characteristic of recently duplicated genes (Assis & Bachtrog, 2013; Assis 2014; Kaessmann, 2010; Kondo et al., 2017; Luis Villanueva-Cañas et al., 2017; Ma et al., 2024), raising the possibility that deletion preferentially removes the younger "child" copy rather than the ancestral "parent" of a duplicate pair (Assis, 2019; Assis & Bachtrog, 2013; Han & Hahn, 2009). Together, these observations motivate two alternative models: one in which weakly constrained, degenerating duplicates are preferentially lost, and another in which function-

Received: 4 February 2026. **Revised:** 2 July 2026. **Accepted:** 7 July 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of the European Society of Evolutionary Biology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com

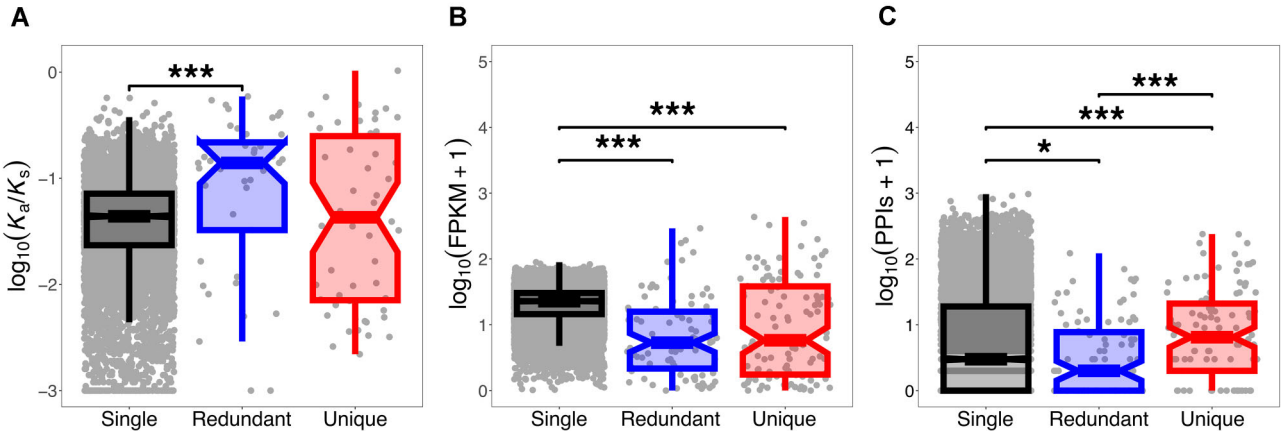


Figure 1. Sequence, expression, and protein-level metrics of deleted genes in *Drosophila*. Boxplots showing the distributions of (A) sequence evolutionary rates measured as $\log_{10}(K_a/K_s)$ between surviving (S) and lost (L) gene copies, (B) gene expression levels measured as the median $\log_{10}(\text{FPKM} + 1)$ across six tissues for each gene copy, and (C) numbers of protein–protein interactions (PPIs) measured as $\log_{10}(\text{PPIs} + 1)$ for single-copy, deleted functionally redundant, and deleted functionally unique genes. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (see *Methods*). Each gene pair contributes one value to panel A, whereas each gene copy contributes one value to panels B and C. Sample sizes: single-copy genes (n = 9,177 in panel A; n = 18,354 in panels B and C), deleted functionally redundant genes (n = 45 in panel A; n = 90 in panels B and C), and deleted functionally unique genes (n = 55 in panel A; n = 110 in panels B and C).

ally active duplicates are removed as an adaptive mechanism that promotes genome efficiency, regulatory stability, or the elimination of deleterious functions. Discriminating between these models requires systematic comparison of the evolutionary and functional properties of deleted genes.

Another important framework relevant to these questions is the Duplication-Degeneration-Complementation (DDC) model (Force et al., 1999), which proposes that degenerative mutations in regulatory elements can partition ancestral gene functions between duplicate genes through subfunctionalization. Under this model, duplicate genes may become jointly required to maintain the full ancestral function, thereby promoting their retention over evolutionary time. Consistent with this framework, studies in *Drosophila* have shown that young duplicate genes can rapidly diverge in expression and occasionally acquire specialized functions (Grill et al., 2023; Kondo et al., 2017; Lee et al., 2024). At the same time, the eventual deletion of apparently functional duplicates suggests that divergence in expression profiles alone may not necessarily ensure long-term evolutionary retention.

Here, we address these questions by examining 100 recent gene deletion events in *Drosophila* (Assis, 2019; Campelo dos Santos et al., 2024). Each event involves a gene duplication in the common ancestor of *D. melanogaster* and *D. pseudoobscura,* followed by deletion of one copy in a descendant lineage. In these cases, one species represents the derived state after deletion and carries a single "derived" copy (D), whereas the other species represents the ancestral state and harbors two copies: the orthologous "survived" copy (S) and the nonorthologous "lost" copy (L; see Figure 1 of Campelo dos Santos et al., 2024). In a prior study, we classified these deletion events as affecting either functionally redundant or unique genes from paired tissue expression data (Campelo dos Santos et al., 2024; see *Methods* and Table S1), providing a valuable framework for dissecting the evolutionary dynamics of gene loss. In the present study, we build on that work by asking two central questions: (1) Does deletion preferentially affect parent or child duplicate gene copies? and (2) What are the sequence, expression, and protein-level features of deleted genes? By answering these questions in the context of redundant versus unique duplicates, we aim to distinguish between nonadaptive gene loss and adaptive purging of functionally active duplicates, thereby illuminating the evolutionary processes shaping gene deletion in *Drosophila.*

Results

To determine whether deletion preferentially affects parent or child copies of duplicate gene pairs, we classified parent–child relationships for the 100 gene deletion events in *Drosophila* (Assis, 2019). Assignments were based on literature review, ortholog retention across *Drosophila* species, and evidence of retrotransposition (see *Methods*). Using these combined criteria, we identified parent and child copies for 90 of the 100 deletion events. Stratifying by functional classifications from Campelo dos Santos et al. (2024), parent–child relationships could be resolved for 37 of 45 redundant deletions and 53 of 55 unique deletions. Functionally unique genes were significantly more likely than functionally redundant genes to have resolvable parent–child relationships ($p = 1.97 \times 10^{-4}$, binomial test; see *Methods*).

Among the 37 functionally redundant deletions with resolved relationships, 14 involved deletion of the parent and 23 involved deletion of the child, a difference that did not statistically deviate from the null expectation of equal loss ($p = 0.19$, binomial test; see *Methods*). In contrast, among the 53 functionally unique deletions with resolved relationships, only nine involved deletion of the parent, revealing a strong bias toward child loss ($p = 1.22 \times 10^{-6}$, binomial test; see *Methods*). Moreover, the distributions of parent and child deletions differed significantly between functionally redundant and functionally unique genes ($p = 0.03$, Fisher's exact test; see *Methods*). Specifically, deletion shows no detectable parent–child bias among functionally redundant duplicates, but strongly favors the loss of the child copy among functionally unique duplicates. These results suggest a mechanism for purging recently evolved functions that are no longer bene-

ficial, as child genes in *Drosophila* often acquire novel roles after duplication (Assis & Bachtrog, 2013; Assis 2014; DeGiorgio & Assis, 2021; Kondo et al., 2017).

To assess the robustness of these patterns, we repeated the parent–child analyses after excluding potential sources of bias. Because retrotransposed duplicates are expected to exhibit divergent expression profiles relative to their parental copies, we first excluded all 22 retrotransposed gene pairs and reevaluated the parent–child relationships. All major patterns remained qualitatively and statistically consistent, including the significant bias toward child loss among unique deletions (Table S2), indicating that the observed trends are not driven solely by retrotransposition events. Similarly, because additional duplication events within gene families could potentially confound parent–child analyses, we conducted the same analyses after excluding all families with copy numbers greater than two in at least one of the 12 *Drosophila* species ($n = 58$ gene families excluded). The resulting patterns remained consistent with those obtained using the full dataset, including the significant bias toward child loss among unique deletions (Table S3). Therefore, all subsequent analyses were performed using the full dataset of 100 deletion events.

To better understand the evolutionary dynamics of gene deletion, we next examined sequence divergence between the S and L gene copies using K_a/K_s ratios, which compare the rates of nonsynonymous (K_a) and synonymous (K_s) substitutions (Li 1993; see *Methods*). We used single-copy genes as a baseline (Table S4), comparing distributions of their K_a/K_s ratios to those of deleted functionally redundant and deleted functionally unique genes. Intriguingly, deleted functionally redundant genes showed significantly elevated sequence evolutionary rates relative to single-copy genes, whereas deleted functionally unique genes exhibited rates comparable to those of single-copy genes (Figure 1A). Thus, deleted functionally unique genes generally did not exhibit signatures consistent with relaxed constraint or positive selection, instead showing levels of constraint similar to single-copy genes.

To gain further insight into this finding, we examined gene expression levels and numbers of protein-protein interactions (PPIs) of deleted genes, again using single-copy genes as a baseline. Both deleted functionally redundant and deleted functionally unique genes had significantly lower expression than single-copy genes, with no significant difference in expression levels between the two deleted classes (Figure 1B). Analyses of PPIs revealed contrasting patterns: deleted functionally redundant genes participate in fewer interactions than single-copy genes, whereas deleted functionally unique genes engage in more interactions than both single-copy and deleted functionally redundant genes (Figure 1C). Taken together, these findings indicate that deleted functionally unique children exhibit characteristics consistent with retained functionality.

Last, we leveraged recent advances in protein structure prediction (Jänes & Beltrao, 2024) to compare the structural properties of single-copy, deleted functionally redundant, and deleted functionally unique genes. Predicted protein structures encoded by the S and L genes were obtained from AlphaFold (Jumper et al., 2021; Varadi et al., 2022, 2024) and aligned to compute TM-scores (see *Methods*), which range from 0 when structures show no similarity to 1 when they are identical (Zhang & Skolnick, 2004). We did not detect statistically significant differences in TM-scores among gene categories (Figure 2A), potentially reflecting limited statistical power. We therefore also examined structural predicted local distance difference test (pLDDT) scores, which range from 0 to 100 and quantify confidence in predicted protein structures, with lower values typically associated with greater intrinsic disorder (Jumper et al., 2021). In this analysis, both deleted functionally redundant and deleted functionally unique genes generally exhibited higher pLDDT scores than single-copy genes (Figure 2B), indicating that deleted genes tend to be more structurally ordered than single-copy genes. One possible interpretation is that proteins with greater structural order may tolerate duplicate loss more readily following functional divergence, whereas intrinsically disordered proteins may be preferentially retained because of their roles in regulatory interactions and context-dependent functions. Structural superimpositions further showed that, in both classes, deleted genes with the lowest structural similarity to their surviving paralogs tended to be shorter and structurally simpler than those with the highest structural similarity (Figure 2C–F). These patterns suggest that structurally divergent deleted genes may have undergone more extensive functional divergence following duplication, whereas structurally conserved genes may have preserved more ancestral protein features and functions.

To further evaluate the robustness of our findings, we repeated all main analyses after excluding the 37 parent–child assignments resolved using orthologous gene counts (see *Methods*). All primary results remained qualitatively consistent and statistically significant following this exclusion (Figures S1 and S2), except the comparison of pLDDT values between single-copy and redundant genes, which was no longer statistically significant (Figure S2B), likely reflecting reduced statistical power. Overall, these analyses indicate that the major sequence, expression, interaction, and structural patterns identified here are robust to alternative parent–child assignment strategies.

Discussion

Gene deletion plays a key role in shaping genome architecture and evolution, yet the factors driving this process remain poorly understood. In this study, we focused specifically on gene deletion following duplication. Accordingly, the patterns identified here may differ from those underlying the deletion of single-copy genes. We systematically characterized 100 gene deletion events in *Drosophila* to uncover the evolutionary and functional patterns underlying gene loss. By integrating sequence, expression, interaction, and structural data, we identified distinct molecular signatures that distinguish functionally redundant and unique deletions. Together, these results provide insight into how gene deletion contributes to the refinement of genome function and the evolution of novel traits.

Our analyses revealed that child copies are preferentially deleted when duplicate genes have unique functions, whereas no such bias occurs for redundant duplicates. This finding suggests that the loss of unique child genes is not random but may reflect their transient evolutionary status. Two main hypotheses could explain this pattern. First, unique child genes may possess nonadaptive or weakly adaptive functions that are not maintained by selection, leading to relaxed constraint and eventual loss through drift. Second, functionally active duplicates may be selectively purged because maintaining them imposes genomic, regulatory, or fitness costs. Such costs could arise through selection favoring genome streamlining, stabilization of regulatory networks, or

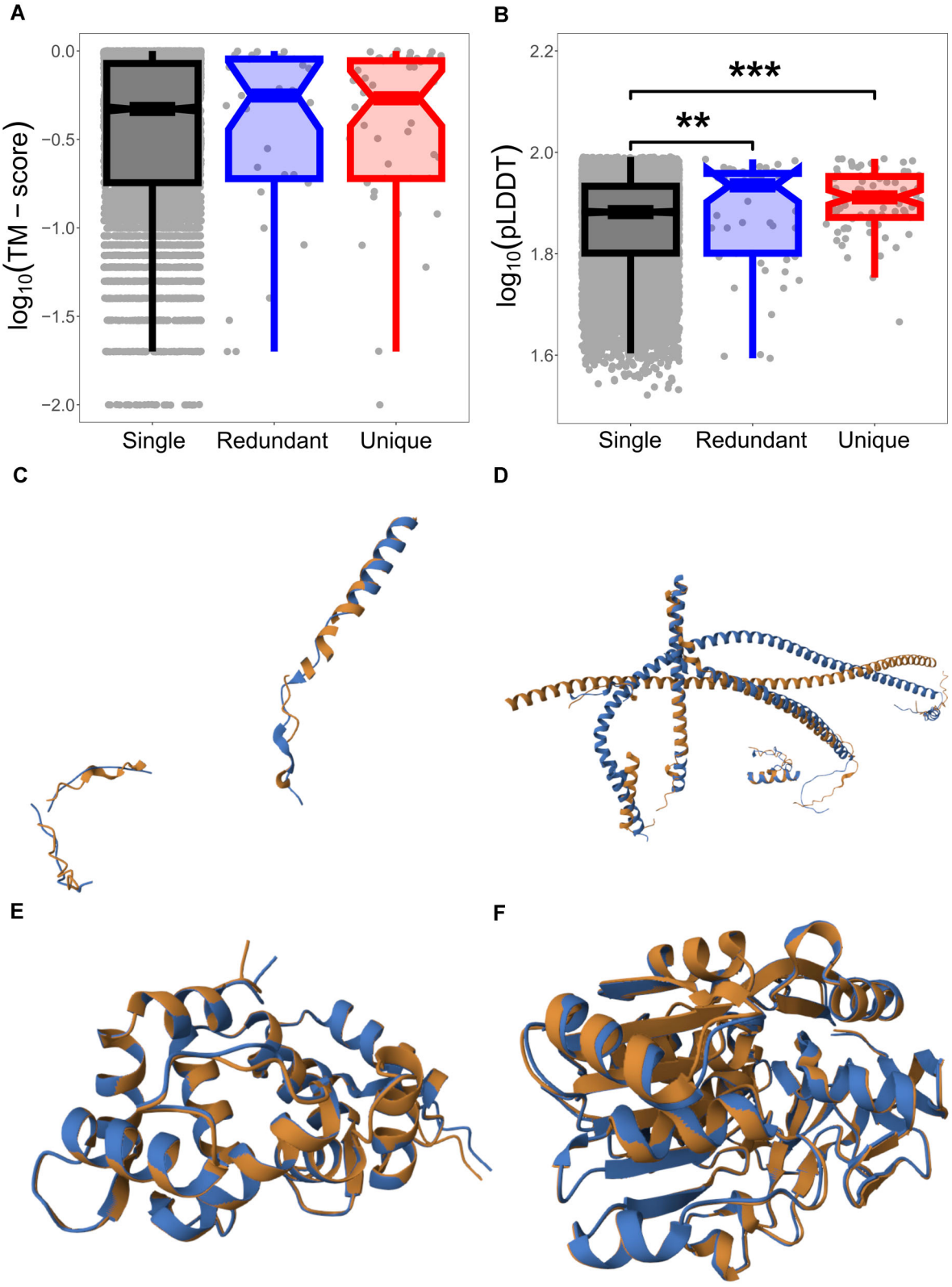


Figure 2. Protein structural properties of deleted genes in *Drosophila*. Boxplots showing the distributions of (A) protein structural similarity measured as $\log_{10}$(template modeling [TM]scores) between surviving (S) and lost (L) proteins, and (B) protein intrinsic disorder approximated by $\log_{10}$(predicted local distance difference test [pLDDT] scores) for single-copy, deleted functionally redundant, and deleted functionally unique genes. (C-F) Structural superimpositions of deleted genes illustrating the lowest and highest structural similarity. The S copy is shown in blue and the L copy in orange. Panels C and D show the least similar functionally redundant (FBgn0030539 and FBgn0052603; TM-score = 0.2) and functionally unique (FBgn0051161 and FBgn0267592; TM-score = 0.1) gene pairs, respectively, whereas panels E and F show the most similar functionally redundant (FBgn0248875 and FBgn0248873; TM-score = 1) and functionally unique (FBgn0022359 and FBgn0024289; TM-score = 1) gene pairs, respectively. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (see *Methods*).

elimination of functions that have become deleterious in changing environments or ecological contexts. Our data do not permit us to formally distinguish between these possibilities, and identifying their relative contributions represents an important direction for future work.

The first hypothesis predicts that deleted functionally unique genes should exhibit signatures of relaxed constraint, such as elevated K_a/K_s ratios, low expression, limited protein-protein connectivity, and increased structural divergence (Assis & Bachtrog, 2013; Birchler & Yang, 2022; Kaessmann, 2010; Kuzmin et al., 2022), consistent with the characteristics of young duplicates in *Drosophila* (Assis & Bachtrog, 2013). However, our results do not support this model. Unique deleted genes show K_a/K_s ratios comparable to those of single-copy genes and lower than redundant duplicates, indicating stronger selective constraint rather than relaxation. Likewise, their higher expression, greater network connectivity, and comparable structural divergence suggest that they were not degenerating prior to deletion. Taken together, these properties are more consistent with the removal of genes that remained functionally active prior to deletion than with the passive loss of weakly constrained duplicates. Additionally, these properties place deleted functionally unique genes closer to single-copy genes than to redundant duplicates.

Because deletion disproportionately affects unique genes in *Drosophila* (Assis, 2019; Campelo dos Santos et al., 2024), these findings challenge the expectation that weakly expressed, low-connectivity genes are the most dispensable (Birchler & Yang, 2022; DeLuna et al., 2010). Instead, the observed properties may reflect a transient phase of integration into complex regulatory networks before removal, similar to patterns observed in other systems where young genes temporarily acquire extensive interactions before being lost (Jin et al., 2013; Li et al., 2005; Zhang et al., 2005). This dynamic suggests that gene deletion may contribute to the fine-tuning of molecular networks by adjusting connectivity and expression balance rather than merely eliminating redundancy. Collectively, these observations are more consistent with the second model, in which adaptive purging removes duplicates whose functional contributions are no longer advantageous.

Although some of the expression differences observed between surviving and deleted duplicate copies may be compatible with DDC-like dynamics, such patterns are not uniquely indicative of subfunctionalization. Indeed, several large-scale studies in *Drosophila* suggest that neofunctionalization may be more prevalent than subfunctionalization in duplicate gene evolution (Assis & Bachtrog, 2013; Cardoso-Moreira et al., 2016; DeGiorgio & Assis, 2021; Jiang & Assis, 2017; Yampolsky and Bouzinier 2014). Moreover, because the classical DDC model primarily focuses on the retention of duplicate genes following functional partitioning, it does not directly address the eventual deletion of apparently functional duplicates observed in our dataset. One possible interpretation is that duplicate genes transiently acquire distinct expression or functional profiles that are not maintained over long evolutionary timescales, ultimately permitting the loss of one copy without substantial fitness consequences.

Overall, our findings suggest that gene deletion in *Drosophila* is not a purely degenerative process but a dynamic force in genome evolution. The preferential loss of unique child genes, despite their functional activity, highlights the selective balance between innovation and constraint: while gene duplication promotes evolutionary novelty, deletion represents a counterforce that removes duplicates whose innovative functions have become dispensable or maladaptive over time. This interplay likely maintains genome efficiency and regulatory stability while permitting bursts of evolutionary experimentation. Future comparative and functional studies will be critical to determine whether similar patterns of adaptive gene loss occur across taxa and to elucidate the mechanisms by which deletion shapes the long-term evolution of gene networks.

Methods

Data acquisition and processing

The 100 deletion events analyzed in this study were obtained from the Dryad dataset associated with a previous study (Assis, 2019; available at https://doi.org/10.5061/dryad.742564m). Each event represents the deletion of one copy from a previously duplicated gene pair. Deletions were inferred to have occurred in either *D. melanogaster* or *D. pseudoobscura* after their divergence, based on phylogenetic comparisons across 12 fully sequenced and annotated *Drosophila* species (Assis, 2019; Drosophila 12 Genomes Consortium et al., 2007). The dataset also included 9,177 single-copy genes used as a comparative baseline, along with quantile-normalized expression levels in six tissues (carcass, male head, female head, testis, accessory gland, and ovary) for both single-copy and deleted genes.

Deletion events were previously classified by Campelo dos Santos et al. (2024), using CLOUDe, a machine learning framework trained on simulated gene expression data generated under an Ornstein-Uhlenbeck model of duplicate gene evolution. CLOUDe evaluates expression patterns among the D, S, and L gene copies. The relationship among the D, S, and L copies is illustrated in Figure 1 of Campelo dos Santos et al. (2024). Deletion events are then classified according to whether D, S, and L copies exhibit broadly similar or distinct expression patterns and inferred expression optima across tissues. In the original CLOUDe analysis, 45 deletion events were classified as redundant and 55 as unique using the best-performing neural network classifier (Campelo dos Santos et al., 2024). Importantly, these classifications are based on differences in expression patterns and do not directly correspond to specific mechanisms of duplicate gene retention. Consequently, unique classifications may be compatible with multiple evolutionary scenarios, including neofunctionalization, subfunctionalization, or combinations of these processes.

For all pairwise comparisons of sequence, expression, PPI, and structural features involving deleted genes, we analyzed the S and L copies, which represent paralogous genes within the species that retained both copies. Accordingly, K_a/K_s ratios and TM-scores were calculated from pairwise comparisons between the S and L genes (or their encoded proteins), whereas expression, PPI, and pLDDT values were analyzed separately for each S and L gene copy. Summaries of all findings for deleted genes are presented in Table S1. We note that chromosome-level genome assemblies for the *Drosophila* species analyzed here are now available via NCBI, derived from long-read sequencing technologies. Although these improved assemblies were not used in the present study, they may further refine orthology calls and parent–child assignments in future work.

We supplemented the expression data from Assis (2019) with sequence, PPI, and protein structure information from additional sources. Coding sequences for all genes used in this study were downloaded from FlyBase release FB2017_05 (dmel r6.18 and dpse_3.04; Gramates et al., 2017), and the longest transcript of each gene was retained. Sequences were aligned using MACSE v2 (Ranwez et al., 2018), and K_a/K_s values were estimated with the codeml program in PAML (Yang, 2007). PPI data were obtained from the STRING database (Szklarczyk et al., 2023). Analyses were restricted to physical links between proteins, and STRING alias files were used to identify the corresponding protein product for each gene. Predicted protein structures and associated pLDDT scores were obtained from AlphaFold v3 (Jumper et al., 2021; Varadi et al., 2022, 2024), and structural similarity was quantified using the template modeling (TM) score (Zhang & Skolnick, 2004) via the Pairwise Structure Alignment tool (Bittrich et al., 2024) at the Protein Data Bank (Berman et al., 2000) with the jFATCAT (flexible) algorithm (Ye & Godzik, 2003).

Assignment of parent–child relationships

We began by reviewing the literature for each S and L gene to identify experimental evidence of their duplication history. Definitive parent–child relationships were found for only 10 of the 100 deletion events, all involving deletions in the *D. pseudoobscura* lineage such that both copies were present in *D. melanogaster*. This bias reflects the greater number of studies available for *D. melanogaster* relative to *D. pseudoobscura.*

Because the literature provided evidence for only a small subset of cases, we next examined additional features distinguishing S and L genes. First, we assessed ortholog retention across 12 *Drosophila* species (Drosophila 12 Genomes Consortium et al., 2007), under the expectation that parent copies are more likely to have detectable orthologs across species than children. This expectation reflects the tendency of child copies in *Drosophila* to evolve under relaxed purifying selection or positive selection following duplication due to their novel genomic and regulatory environment (Assis & Bachtrog, 2013; Jiang & Assis, 2017; Kondo et al., 2017), making them less likely to be retained across all 12 species compared to the parent copy. Using whole-genome alignments from the UCSC Genome Browser (Nassar et al., 2023) and the flyDIVaS database (Stanley & Kulathinal, 2016), we designated the copy retained across all 12 species as the parent, resolving 16 additional deletion events. We then examined gene structures for evidence of retrotransposition, in which an mRNA transcript is reverse-transcribed into cDNA and reinserted into the genome (Kaessmann et al., 2009). Retrotransposed duplicates are typically intronless, in contrast to their multi-exon ancestral genes (Langille & Clark, 2007). Among the remaining 74 deletion events, we identified 22 in which one gene lacked introns while its paralog contained at least one, allowing us to designate the intronless copy as the child.

To further evaluate ortholog retention, we queried OrthoDB v11 (Kuznetsov et al., 2023) for orthologs of the remaining S and L genes across 37 *Drosophila* species, designating the copy with more orthologs as the parent, consistent with greater evolutionary age. This approach resolved 17 additional cases, including 14 with orthologs present in all 37 species. We also reanalyzed the OrthoFinder (Emms & Kelly, 2019) results from Assis (2019), which identified 21,944 orthogroups across 12 *Drosophila* species. For the unresolved events, we designated the copy belonging to the larger orthogroup as the parent, resolving 20 of the remaining 35 deletion events. To assess whether additional duplication events within gene families could confound our analyses, we used OrthoFinder results from Assis (2019) to examine copy numbers across the 12 *Drosophila* species included in the dataset. Gene families in which any species contained more than two copies were considered to show evidence of additional duplications. Parent–child analyses were then repeated after excluding these families.

Finally, we applied the same ortholog retention criteria with OrthoDB to assess D genes in the remaining 15 unresolved events. This analysis yielded definitive parent–child assignments for five additional cases, while the remaining ten lacked sufficient evidence and were left unclassified.

Statistical analyses

All statistical analyses were performed in R (R Core Team, 2021). Exact two-sided binomial tests were conducted using the binom.test function in the *stats* package (R Core Team, 2021) to assess whether the observed proportions of different subclasses of deleted genes deviated significantly from null expectations. For the comparison of resolved parent–child relationships between redundant and unique deletions, we set the number of successes $x = 37$ to represent the number of resolved redundant deletions, the number of trials $n = 45$ to represent the total number of redundant deletions, and the probability of success $p = 53/55$ to represent the proportion resolved among unique deletions. For the comparison of proportions of deleted parents and children, we set $x = 14$ to represent the number of deleted parents, $n = 37$ to represent the total number of redundant deletions with resolved parent–child relationships, and $p = 0.5$ to represent the null expectation of no bias. For the comparison of proportions of deleted parents and children among unique deletions, we set $x = 9$ to represent the number of deleted parents, $n = 53$ to represent the total number of unique deletions with resolved parent–child relationships, and $p = 0.5$ to represent the null expectation of no bias. To directly compare the distributions of deleted parents and deleted children between redundant and unique gene sets, we conducted a two-sided Fisher's exact test using the fisher.test function in the stats package (R Core Team, 2021). Differences between distributions shown in Figures 1 and 2 were assessed with two-sided Mann-Whitney *U* tests implemented with the wilcox.test function in the *stats* package (R Core Team, 2021). For expression analyses, expression values were first summarized as the median across the six tissues for each gene copy, yielding a single expression value for statistical analyses (Table S1).

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

The data underlying this article are available in Zenodo, at https://doi.org/10.5281/zenodo.21517327.

Author contributions

Andre Luiz Campelo dos Santos (Data curation [equal], Formal Analysis [lead], Investigation [equal], Methodology [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Raquel Assis (Conceptualization [lead], Data curation [equal], Funding acquisition [lead], Investigation [equal], Methodology [equal], Resources [lead], Supervision [lead], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Funding

This work was supported by National Institutes of Health grant R35GM142438 and National Science Foundation grant DBI-2130666 to R.A.

Conflicts of interest

The authors declare no conflict of interest.

References

Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics*, *17*, 379–391. https://doi.org/10.1038/nrg.2016.39

Assis, R. (2014). Drosophila duplicate genes evolve new functions on the fly. *Fly*, *8*, 91–94. https://doi.org/10.4161/fly.29131.

Assis, R. (2019). Out of the testis, into the ovary: biased outcomes of gene duplication and deletion in Drosophila. *Evolution; Internation Journal of Organic Evolution*, *73*, 1850–1862. https://doi.org/10.1111/evo.13820.

Assis, R., & Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in Drosophila. *Proceedings of the National Academy of Sciences*, *110*, 17409–17414. https://doi.org/10.1073/pnas.1313759110

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*, 235–242. https://doi.org/10.1093/nar/28.1.235

Birchler, J. A., & Yang, H. (2022). The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell*, *34*, 2466–2474. https://doi.org/10.1093/plcell/koac076

Bittrich, S., Segura, J., Duarte, J. M., & Rose, Y. (2024). RCSB protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. *Bioinformatics*, *40*, btae370. https://doi.org/10.1093/bioinformatics/btae370.

Campelo dos Santos, A. L., DeGiorgio, M., & Assis, R. (2024). Predicting evolutionary targets and parameters of gene deletion from expression data. *Bioinformatics Advances*, *4*, vbae002. https://doi.org/10.1093/bioadv/vbae002

Cardoso-Moreira, M. (2016). Evidence for the fixation of gene duplications by positive selection in Drosophila. *Genome Research*, *26*, 787–798. https://doi.org/10.1101/gr.199323.115

DeGiorgio, M., & Assis, R. (2021). Learning retention mechanisms and evolutionary parameters of duplicate genes from their expression data. *Molecular Biology and Evolution*, *38*, 1209–1224. https://doi.org/10.1093/molbev/msaa267

DeLuna, A., Springer, M., Kirschner, M. W., & Kishony, R. (2010). Need-based up-regulation of protein levels in response to deletion of their duplicate genes. *PLoS Biology*, *8*, e1000347. https://doi.org/10.1371/journal.pbio.1000347

Drosophila 12 Genomes Consortium. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*, 203–218. https://doi.org/10.1038/nature06341

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 238. https://doi.org/10.1186/s13059-019-1832-y

Force, A. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, *151*, 1531–1545. https://doi.org/10.1093/genetics/151.4.1531

Gramates, L. S. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Research*, *45*, D663–D671. https://doi.org/10.1093/nar/gkw1016

Grill, S., Riley, A., Selvaraj, M., & Lehmann, R. (2023). HP6/Umbrea is dispensable for viability and fertility, suggesting essentiality of newly evolved genes is rare. *Proceedings of the National Academy of Sciences*, *120*, e2309478120. https://doi.org/10.1073/pnas.2309478120

Han, M. V., & Hahn, MW. (2009). Identifying parent-daughter relationships among duplicated genes. *Pacific Symposium on Biocomputing*, 114–125.

Hottes, A. K. (2013). Bacterial adaptation through loss of function. *PLoS Genetics*, *9*, e1003617. https://doi.org/10.1371/journal.pgen.1003617

Jänes, J., & Beltrao, P. (2024). Deep learning for protein structure prediction and design-progress and applications. *Molecular Systems Biology*, *20*, 162–169. https://doi.org/10.1038/s44320-024-00016-x

Jiang, X., & Assis, R. (2017). Natural selection drives rapid functional evolution of young Drosophila duplicate genes. *Molecular Biology and Evolution*, *34*, 3089–3098. https://doi.org/10.1093/molbev/msx230

Jin, Y., Turaev, D., Weinmaier, T., & Makse, HA. (2013). The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient networks. *PLoS ONE*, *8*, e58134. https://doi.org/10.1371/journal.pone.0058134

Jumper, J. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Research*, *20*, 1313–1326. https://doi.org/10.1101/gr.101386.109

Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics*, *10*, 19–31. https://doi.org/10.1038/nrg2487

Kondo, S. (2017). New genes often acquire male-specific functions but rarely become essential in Drosophila. *Genes & Development*, *31*, 1841–1846. https://doi.org/10.1101/gad.303131.117

Kuzmin, E., Taylor, J. S., & Boone, C. (2022). Retention of duplicated genes in evolution. *Trends in Genetics*, *38*, 59–72. https://doi.org/10.1016/j.tig.2021.06.016

Kuznetsov, D. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, *51*, D445–D451. https://doi.org/10.1093/nar/gkac998

Kvitek, D. J., & Sherlock, G. (2013). Whole genome, whole population sequencing reveals that loss of signaling networks is the

major adaptive strategy in a constant environment. *PLoS Genetics*, *9*, e1003972. https://doi.org/10.1371/journal.pgen.1003972

Langille, M. G. I., & Clark, DV. (2007). Parent genes of retrotransposition-generated gene duplicates in Drosophila melanogaster have distinct expression profiles. *Genomics*, *90*, 334–343. https://doi.org/10.1016/j.ygeno.2007.06.001

Lee, U. (2024). The three-dimensional genome drives the evolution of asymmetric gene duplicates via enhancer capture-divergence. *Science Advances*, *10*, eadn6625. https://doi.org/10.1126/sciadv.adn6625

Li, W.-H. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution*, *36*, 96–99. https://doi.org/10.1007/BF02407308

Li, W.-H., Yang, J., & Gu, X. (2005). Expression divergence between duplicate genes. *Trends in Genetics*, *21*, 602–607. https://doi.org/10.1016/j.tig.2005.08.006

Luis Villanueva-Cañas, J. (2017). New genes and functional innovation in mammals. *Genome Biology and Evolution*, *9*, 1886–1900.

Ma, F., Lau, C. Y., & Zheng, C. (2024). Young duplicate genes show developmental stage- and cell type-specific expression and function in Caenorhabditis elegans. *Cell Genomics*, *4*, 100467. https://doi.org/10.1016/j.xgen.2023.100467

Nassar, L. R. (2023). The UCSC Genome Browser database: 2023 update. *Nucleic Acids Research*, *51*, D1188–D1195. https://doi.org/10.1093/nar/gkac1072

Olson, MV. (1999). When less is more: gene loss as an engine of evolutionary change. *The American Journal of Human Genetics*, *64*, 18–23. https://doi.org/10.1086/302219

Ranwez, V., Douzery, E. J. P., Cambon, C., & Delsuc, F. (2018). MACSE v2: Toolkit for the alignment of Coding Sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution*, *35*, 2582–2584. https://doi.org/10.1093/molbev/msy159

R Core Team. (2021). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing.

Stanley, C. E., Jr, & Kulathinal, R. J. (2016). FlyDIVaS: A comparative genomics resource for Drosophila divergence and selection. *G3 Genes|Genomes|Genetics*, *6*, 2355–2363. https://doi.org/10.1534/g3.116.031138

Szklarczyk, D. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*, D638–D646. https://doi.org/10.1093/nar/gkac1000

Varadi, M. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061

Varadi, M. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, *52*, D368–D375. https://doi.org/10.1093/nar/gkad1011

Yampolsky, L. Y., & Bouzinier, M. A. (2014). Faster evolving Drosophila paralogs lose expression rate and ubiquity and accumulate more non-synonymous SNPs. *Biol Direct*, *9*, 2. https://doi.org/10.1186/1745-6150-9-2

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Ye, Y, & Godzik, A, eds. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists" by Yuzhen Ye and Adam Godzik. *Bioinformatics*, *19*, 246–255.

Zhang, Z., Luo, Z. W., Kishino, H., & Kearsey, MJ. (2005). Divergence pattern of duplicate genes in protein-protein interactions follows the power law. *Molecular Biology and Evolution*, *22*, 501–505. https://doi.org/10.1093/molbev/msi034

Zhang, Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, *57*, 702–710. https://doi.org/10.1002/prot.20264

Received: 4 February 2026. **Revised:** 2 July 2026. **Accepted:** 7 July 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the European Society of Evolutionary Biology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com