

## RESEARCH ARTICLE

# Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in *Drosophila* development

Shengqian Xia<sup>1‡</sup>, Nicholas W. VanKuren<sup>1‡</sup>, Chunyan Chen<sup>2,3‡</sup>, Li Zhang<sup>1</sup>, Clause Kemkemer<sup>1</sup>, Yi Shao<sup>2,3</sup>, Hangxing Jia<sup>2,3</sup>, UnJin Lee<sup>1</sup>, Alexander S. Advani<sup>1</sup>, Andrea Gschwend<sup>4</sup>, Maria D. Vibranovski<sup>5</sup>, Sidi Chen<sup>6</sup>, Yong E. Zhang<sup>1,2,3,7\*</sup>, Manyuan Long<sup>1\*</sup>

**1** Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois, United States of America, **2** State Key Laboratory of Integrated Management of Pest Insects and Rodents & Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, **3** University of Chinese Academy of Sciences, Beijing, China, **4** Department of Horticulture & Crop Science, The Ohio State University, Columbus, Ohio, United States of America, **5** Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil, **6** Department of Genetics, Yale School of Medicine, West Haven, Connecticut, United States of America, **7** Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan, China

‡ These authors share first authorship on this work.

\* [zhangyong@ioz.ac.cn](mailto:zhangyong@ioz.ac.cn) (YEZ); [mlong@uchicago.edu](mailto:mlong@uchicago.edu) (ML)



## OPEN ACCESS

**Citation:** Xia S, VanKuren NW, Chen C, Zhang L, Kemkemer C, Shao Y, et al. (2021) Genomic analyses of new genes and their phenotypic effects reveal rapid evolution of essential functions in *Drosophila* development. *PLoS Genet* 17(7): e1009654. <https://doi.org/10.1371/journal.pgen.1009654>

**Editor:** Juergen Brosius, University of Münster, GERMANY

**Received:** January 12, 2021

**Accepted:** June 9, 2021

**Published:** July 9, 2021

**Copyright:** © 2021 Xia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

**Funding:** Y.E. Zhang was supported by the National Key R&D Program of China (2019YFA0802600, 2018YFC1406902) and the National Natural Science Foundation of China (91731302, 31771410, 31970565). M. L. supported by NSF1026200 and NIH R01GM116113. The funders had no role in study design, data collection and

## Abstract

It is a conventionally held dogma that the genetic basis underlying development is conserved in a long evolutionary time scale. Ample experiments based on mutational, biochemical, functional, and complementary knockdown/knockout approaches have revealed the unexpectedly important role of recently evolved new genes in the development of *Drosophila*. The recent progress in the genome-wide experimental testing of gene effects and improvements in the computational identification of new genes (< 40 million years ago, Mya) open the door to investigate the evolution of gene essentiality with a phylogenetically high resolution. These advancements also raised interesting issues in techniques and concepts related to phenotypic effect analyses of genes, particularly of those that recently originated. Here we reported our analyses of these issues, including reproducibility and efficiency of knockdown experiment and difference between RNAi libraries in the knockdown efficiency and testing of phenotypic effects. We further analyzed a large data from knockdowns of 11,354 genes (~75% of the *Drosophila melanogaster* total genes), including 702 new genes (~66% of the species total new genes that aged < 40 Mya), revealing a similarly high proportion (~32.2%) of essential genes that originated in various *Sophophora* subgenus lineages and distant ancestors beyond the *Drosophila* genus. The transcriptional compensation effect from CRISPR knockout were detected for highly similar duplicate copies. Knockout of a few young genes detected analogous essentiality in various functions in development. Taken together, our experimental and computational analyses provide valuable data for detection of phenotypic effects of genes in general and further strong evidence

analysis, decision to publish, or preparation of the manuscript.

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

for the concept that new genes in *Drosophila* quickly evolved essential functions in viability during development.

## Author summary

It is conventionally believed that the genes in control of important biological functions, processes and structures are conserved in evolution. Using recently developed RNAi knockdown libraries in combination of CRISPR knockout analyses, we showed that a high proportion in a large number of evolutionary new genes (~32.2% in 702 new genes, aged < 40 million years ago) in *D. melanogaster* are essential for the survival in development. We found that the frequency of essentiality in gene functions is stable in the species' various ancestral stages and unchanged among ancient genes, unveiling a constant gene evolution that has driven evolution of *Drosophila* species in the *Sophophora* subgenus. We detected the transcriptional compensation effect from CRISPR knockout for highly similar duplicate copies. We experimentally examined the reproducibility, knockdown efficiency and performance of knockdown libraries, revealing the validity of RNAi knockdown in detection of essential phenotypic effects of genes. Our experimental and computational analyses provided strong evidence for the concept that evolutionary new genes in *Drosophila* quickly evolved essential functions.

## Introduction

The question of how often evolutionary new genes develop essential functions is a critical problem in understanding the genetic basis of development and general phenotypic evolution. New genes in evolution have widely attracted discussion [1–6], supported by increasing studies with fulsome evidence in various organisms [7–13]. The detected large number of new genes with unexpected rate of new gene evolution [14–16] and the revealed important functions of new genes [2,17–20] challenged a widely held dogma that the genetic basis in control of development is conserved in a long time scale of evolution [21–24].

Our previous genome-wide analysis used the RNAi knockdown in a smaller sample showing that new genes may quickly become essential in *Drosophila* and that potential for a gene to develop an essential function is independent of its age [25]. This work suggests a tremendous and quickly evolving genetic diversity, which had not been previously anticipated. Since then, genomes of better quality from more species have allowed for more reliable new gene annotation [15]. In addition, technical progress in the detection of gene effects has increased with better equipped knockdown libraries and direct CRISPR knockout methods. Related scientific discoveries and technical development in knockdown and knockout techniques can be considered when investigating the evolution of gene essentiality. For example, Green *et al* reported unexpected dsRNA construct landing site (40D3) of a public RNAi library and its phenotypic consequence [26]; Kondo *et al* investigated phenotypic consequences of newly added RNAi triggers into RNAi libraries [27], although its knockout analyses of new essential genes were re-examined by a recent thorough phenotypic analysis, which revealed essential functions in early development in tested young genes [17].

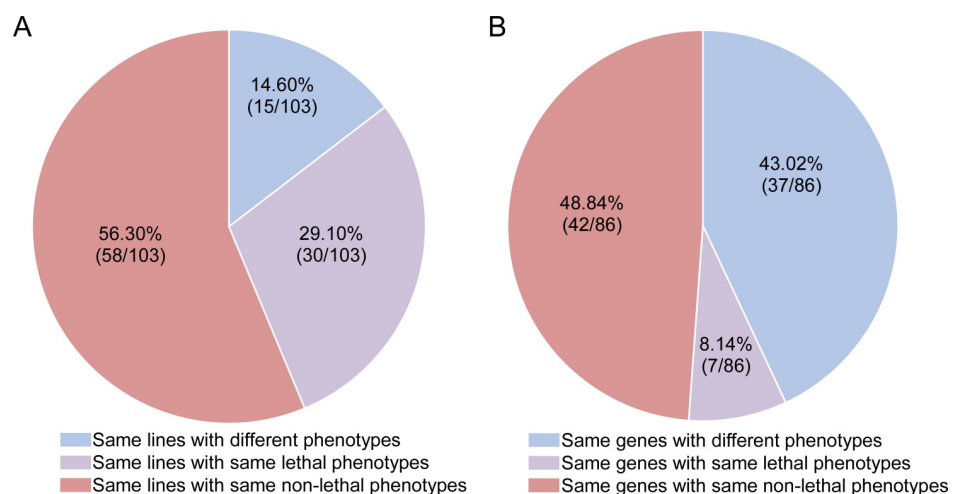
We will present in this report our recent experiments and computational analyses, examining a few important issues raised in recent years that we find to be generally relevant for the detection of the phenotypic effects of genes, particularly of those that recently originated. We

conducted following analyses of newly raised technical problems and scientific issues: 1) the repeatability of knockdown analysis for testing essentiality phenotypes; 2) an evaluation of the knockdown efficiency distribution in RNAi experiments; 3) an understanding of the differences between different RNAi libraries in phenotyping large samples of new genes for viability effects; 4) a detection of compensation effect which may compromise detection of effects of CRISPR mutants in new gene duplicates; 5) a detection of developmental essential effects in a larger sample of new genes that are shown to be also in a higher resolution for their *Drosophila* lineage distribution based on our recent dating gene ages [15] than that previous analysis [25]. Our data, with additional evidence published recently by our group and others, provide ample and strong evidence to further support a concept suggested by the fitness effect analysis of new genes in *Drosophila*: new genes have quickly evolved essential functions in viability during development. Meanwhile, both the technique and the data we created regarding the RNAi knockdown analyses of repeatability, efficiency, and different libraries and detected compensation effect for highly similar gene duplicates will offer wide and valuable reference for detection of phenotypic effects of genes in general.

## Results

### High reproducibility of RNAi knockdown for detecting lethal phenotypes

We investigated the consistency of RNAi experiments with the same lines and the same drivers in different laboratories, conditions, and years. Zeng *et al* independently screened 16,562 transgenic RNAi lines using an Act5C-Gal4 driver to detect the lethality of 12,705 protein-coding genes (~90% of all annotated coding genes) in their study of intestinal stem cell development and maintenance [28]. Their dataset included RNAi lines targeting the same 103 genes that were measured for lethality by Chen *et al* [25]. Chen *et al* and Zeng *et al* obtained the same phenotypes for 88 (85.4%) genes, including 30 (29.1%) of the lethal phenotype and 58 (56.3%) of non-lethal phenotype (Fig 1A, S1 Table). These data suggest that despite differences in independent observers, lab environments, and years to conduct experiments, the vast



**Fig 1. The reproducibility analysis of RNAi experiments by comparing two groups of independent experiments by [25] and [28].** A. Phenotypes of same 103 RNAi lines analyzed by [25] and [28] using same lines; B. Phenotypes of 86 same new genes knocked down by two different drivers or the same drivers with different insertion sites. The old drivers detected 29 genes as lethal while 57 non-lethal; the new drivers detected 20 genes as lethal while 66 non-lethal.

<https://doi.org/10.1371/journal.pgen.1009654.g001>

majority of RNAi knockdown experiments are reproducible for phenotyping lethality and non-lethality.

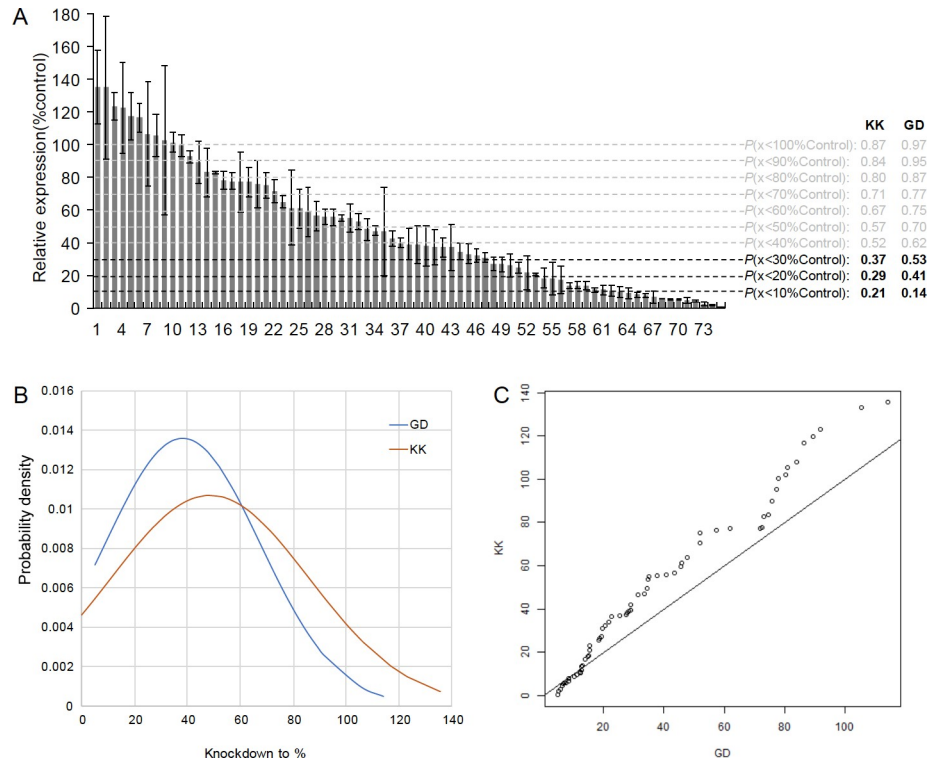
We also tested consistency between RNAi lines with different RNAi drivers (called new drivers) or same drivers in different genome positions. Specifically, the datasets of Chen *et al* and Zeng *et al* shared 86 new genes in knockdown experiments, mostly (81.4%, 70) with different RNAi drivers and fewer (18.6%, 16) same drivers in different genome positions (S2 Table). This dataset showed that: 7 genes were consistently lethal; 42 genes were consistently non-lethal; and 37 genes have different phenotypes (Fig 1B). Thus, the two groups with different drivers or same drivers with different positions reveal that a majority of new genes (57.0%, 49) show the same phenotypes.

### Low knockdown efficiency and better GD library than KK library

We considered an additional factor in RNAi knockdown, sensitivity, in the two widely used RNAi libraries: the Vienna Drosophila Resource Center's (VDRC's) GD and KK libraries [29]. The GD libraries were constructed using P-elements to randomly insert hairpin RNAs (average 321bp) into the genome targeting individual genes, while the KK library inserted constructs carrying hairpin RNAs (average 357bp) into a specific landing site by  $\Phi$ C31-mediated homologous recombination. All KK lines carry an insertion at 30B3, but a proportion (23–25%) also carry an insertion at 40D3 (*tio* locus) that results in pupal lethality when using constitutive drivers like Act5C-GAL4 [26,30]. Unless specified, no lines discussed below contain 40D3 insertions.

Given the intrinsic different designs of GD and KK libraries, we hypothesized that they have different false negative or false positive rates, which cause the inconsistency shown in Fig 1B. Only GD lines were examined previously, and they have a high false negative rate (39.9%) but low false positive rate (<2%) [29]. The high false negative rate is likely caused by insufficient target gene knockdown, while false positives may be due to off target effects [29]. We thus tested the knockdown efficiency of 75 KK lines targeting randomly selected 75 young genes (S3 Table, Fig 2A). We found that the knockdown efficiency of KK lines is generally lower than the efficiency of 64 GD lines as previously reported [29]. Specifically, using the same driver (Act5C), we found that in general, GD lines have significantly higher knockdown efficiency than KK lines, as shown by the knockdown expression as the percentage of the control expression (Fig 2A). That is, the KK lines have an average knockdown efficiency as 48.6% of control expression while the GD lines show an average efficiency as 38.1% (Fig 2B and 2C, *t*-test  $P = 0.031$ ). Notably, the expression reduction to 50–60% level of the wide-type level was observed to have no significant fitness loss due to widespread haplosufficiency [31,32]. Detecting any fitness effect may be expected when the expression drops to a lower level, for example, 20–30% or lower of the control expression. In this range of knockdown efficiency, we observed that only 29% of KK lines but 41% of GD lines reduced target expression levels to  $\leq 20\%$  of control levels; 37% of KK lines but 53% of GD lines were seen to reduce target expression levels to  $\leq 30\%$  of control levels (Fig 2A). Thus, based on knockdown efficiency assay of 75 KK RNAi lines derived by Act5C, these observations derive two expectations: firstly, most of RNAi knockdown have so low efficiency that no phenotypic effects can be expected; secondly, GD lines, having a higher knockdown efficiency than KK lines, should have a higher power in detecting lethal phenotypes as shown in the next section.

To estimate false positive rate of KK lines, we constructed 10 randomly chosen new KK lines targeting one member of a young duplicate gene pair, in addition to one KK line and 3 TRiP lines (Transgenic RNAi Project, BDSC, Materials and Methods). The rationale is that for each gene of interest its paralog is the most likely off target. The same rationale was also followed by [29] when false positive rates of GD lines were estimated. We measured the

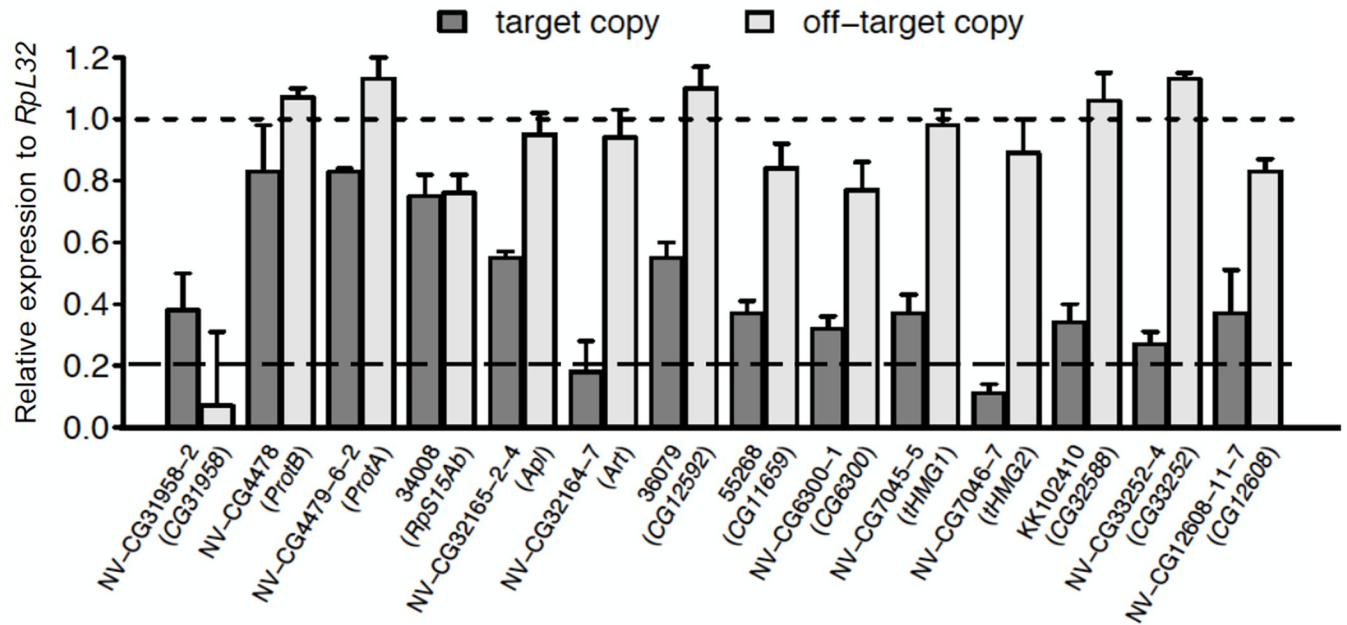


**Fig 2. Knockdown efficiency in the KK and GD libraries revealed GD lines have significantly higher knockdown efficiency than the KK lines.** A. The knockdown efficiency of the 75 KK lines was measured, compared to the expression of the wild-type control and the standard deviation is calculated from the measurement of three repeats; P refers to proportion of genes with the expression lower than a certain threshold while the values of KK lines are generated in this work and that of GD lines are extracted from [29]. B. The Z-score distributions of knockdown efficiency of KK and GD lines. z-score means the number of standard deviations from the mean.  $z = (x - \mu) / \sigma$ , x is the value to be standardized,  $\mu$  is the mean,  $\sigma$  is the standard deviation. 38.1% and 48.6% are average knockdown-to-percent of GD and KK lines, respectively. C. The Q-Q Plot (quantile-quantile plot) between KK and GD lines. Q-Q plot is a probability plot which is used here for comparing probability distributions between knockdown-to-percent of KK and GD lines by plotting their quantiles against each other. Those Q-Q plots don't fit the best fitting straight line, which indicate that the knockdown efficiency distribution between KK and GD lines are different.

<https://doi.org/10.1371/journal.pgen.1009654.g002>

knockdown efficiency and estimated off-target effects using these 14 lines with qPCR experiments in adult whole bodies (Fig 3). We found that two lines likely produce off-target effects (NV-CG31958, 34008 (the TRiP line)), for both of which the expression of paralog is down-regulated to similar or even lower level compared to the corresponding gene of interest. Twelve other lines have significantly higher target effects than off-target effects, among which 10 genes reduced activity to 20–80% expression level of the control (7 genes reduced activity to 20–40%) and only two genes (CG32164, CG7046) reach  $\leq 20\%$  of control levels. Thus, if we take 20% as the cutoff of efficient knockdown, only CG31958 could be counted as the false positive, and CG32164 and CG7046 be counted as the true positives. Collectively speaking, the off-target effects are rare while insufficient knockdowns are pervasive.

These experiments detected an interesting variation of knockdown efficiency among different drivers where newer KK lines have lower efficiency and thus higher false negatives compared to older GD lines. Thus, if a newly constructed RNAi driver is added to the phenotypic analysis, insufficient knockdown is also introduced with a high probability, suggesting the new RNAi driver with the low knockdown efficiency is not revealing for a gene's essentiality.



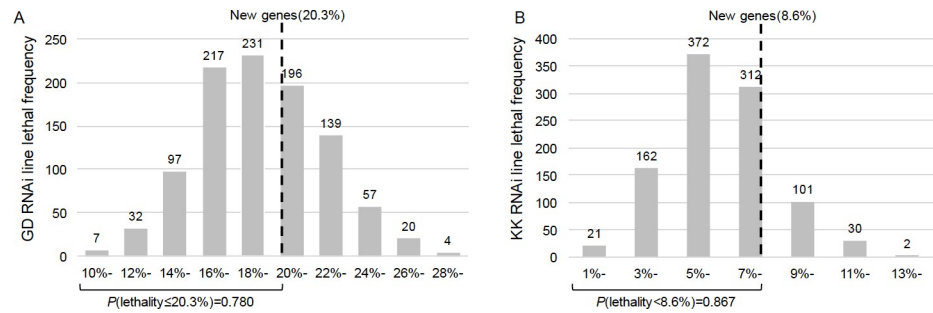
**Fig 3. Experimental comparison of the efficiency and off-target effects explain the conservative nature of RNAi knockdown experiments and limited off-targets propensity.** For each young duplicate gene pair specific for *D. melanogaster* and *melanogaster* species complex, we examined their expression intensity relative to the wide type control (relative expression level to WT normalized by *RpL32*) in whole body flies with qPCR. Only two lines likely produce off-target effects (NV-CG31958, 34008 (the TRiP line)), for both of which the expression of paralog is down-regulated to similar or even lower level compared to the corresponding gene of interest. However, the 34008 line have so low knockdown efficiency that no phenotypic effects can be expected. The standard deviation is calculated based on three replicates.

<https://doi.org/10.1371/journal.pgen.1009654.g003>

### Both GD and KK libraries detected similar proportions of lethality between new and old genes and a higher proportion of lethality in GD

We first investigated differences in measured lethality between the KK and GD libraries used in Chen *et al* [25]. To control for the confounding effect of *tio* insertion in the KK lines, we genotyped these lines using PCR-amplification and found that out of 153 KK lines we collected (S4 Table), 47 (30.7%) had two landing sites and 6 (3.9%) had only 40D3 landing site (the confounding site) [26], which all showed lethal phenotypes. Using the recombination approach [26], we recovered 41 of the 47 lines into the lines that have only the 30B3 site. The RNAi knockdown of 140 KK lines carrying insertions only at 30B3 identified 12 genes (8.6%) with lethal phenotypes. Meanwhile, 12 new genes in GD lines for 59 new genes (20.3%) were detected to have lethal knockdown effects [25], significantly higher than the KK libraries ( $P = 0.0112$ , Fisher's Exact Test). As aforementioned, this difference is likely due to higher false negative rate of KK lines (Fig 2).

By using the essentiality data of 10,652 old genes provided by VDRC ([https://stockcenter.vdrc.at/control/library\\_rnai](https://stockcenter.vdrc.at/control/library_rnai)) that were in branch 0 [15], we characterized the statistical distribution of essential old genes (Fig 4). We independently sampled 1000 times, with each randomly sampling 150 old genes and calculating the proportion of essential ones. We found that in the GD library, the probability to obtain a proportion of essential new genes equal or lower than 20.3% is 0.780. Meanwhile, in the KK library, the probability to observe a proportion of essential new genes equal or lower than 8.6% is 0.867. These analyses of GD and KK libraries reveal similarly that the proportions of new and old genes with lethal phenotypes are not statistically different. We note here that due to the low knockdown efficiency in GD library, despite higher than KK library, the actual proportion of essential genes can be higher than 20.3% as measured by GD library.



**Fig 4.** Comparison of proportions of lethality between new genes and old genes in GD lines (A) and KK lines (B) suggests that in both GD and KK lines, new genes have an equally high probability to be lethal as old genes. By using the essentiality data of 10,652 old genes provided by VDRC ([https://stockcenter.vdrc.at/control/library\\_rnai](https://stockcenter.vdrc.at/control/library_rnai)) that were in branch 0 [15], we characterized the statistical distribution of essential old genes. Since old genes are much more abundant than new genes, we independently sampled 1000 times of old genes with the same number (150) of new genes and then plotted the distribution of proportion of essential genes as histograms. In the GD library, the probability to obtain a proportion of essential new genes equal or lower than 20.3% is 0.780. Meanwhile, in the KK library, the probability to observe a proportion of essential new genes equal or lower than 8.6% is 0.867. These analyses of GD and KK libraries reveal similarly that the proportions of new and old genes with lethal phenotypes are not statistically different.

<https://doi.org/10.1371/journal.pgen.1009654.g004>

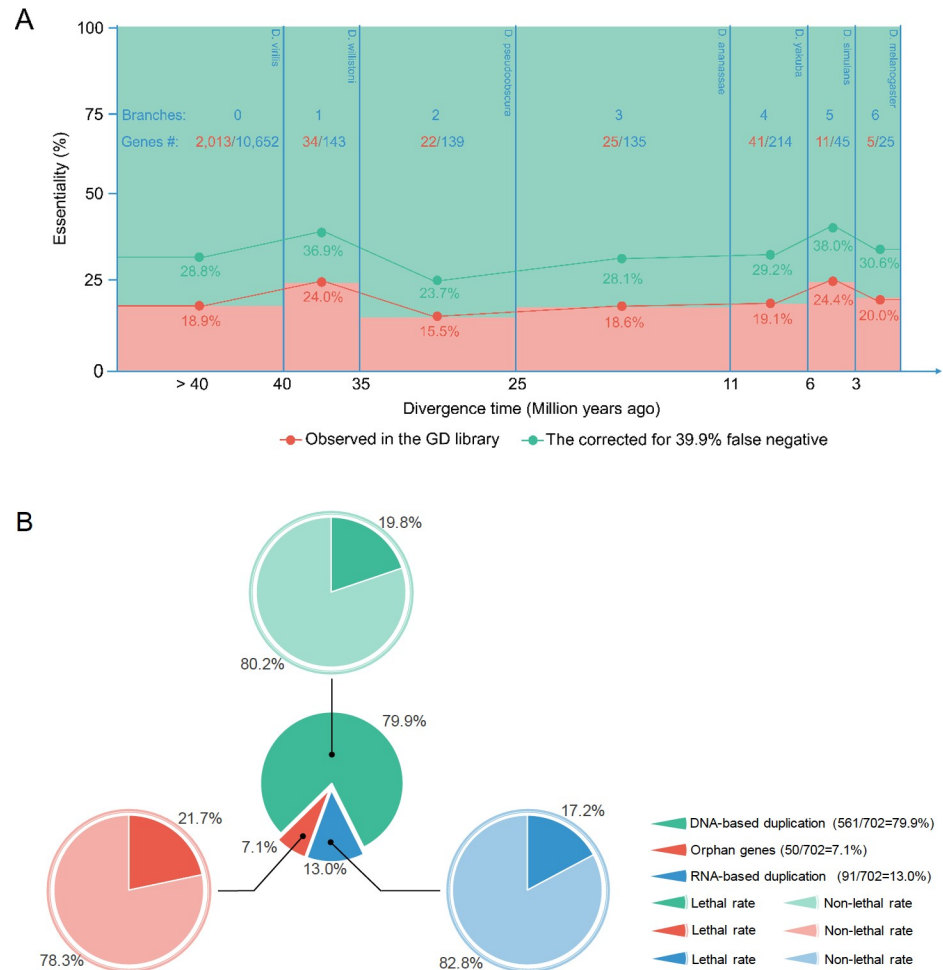
## Genome-wide gene-dating in *Drosophila* phylogeny reveal that new genes evolve essential functions in development quickly

Further analysis of gene essentiality data in a recent version of VDRC libraries detected with increased resolution the proportions of essential genes in six detectable ancestral stages of *D. melanogaster*. We reported the analysis of the GD library, which has a significantly higher knockdown efficiency than the KK library. From several new gene duplicate datasets of *Drosophila* [33–35], we chose two recent datasets to compare [15,27] (S1 Fig, also see the comparison in S1 File) because they are more updated and also used additional gene-synteny information from multiple species genomes besides considering substitution rates among paralogous and orthologous copies. Due to the better quality (S1 Fig) of GenTree and more complete types of new genes including DNA-based and RNA-based duplicates and orphan genes, we used its gene dating results for examining the evolution of essentiality in all *Drosophila* genes. In total, 11,354 genes (72% of 15,682 genes in the species, Ensembl 73) have been identified phenotype for their lethality or non-lethality, including 702 *Drosophila* genus specific genes (66% of 1,070 detected *Drosophila*-specific genes) [15,20,34,36] and 10,652 genes that predated the *Drosophila* divergence 40 Mya.

We parsimoniously mapped the 702 *Drosophila*-specific genes on the six ancestral stages by examining their species distribution in the *Drosophila* phylogeny [15] (Fig 5A). Of the 702 genes, 19.7% (138) are directly observed to be essential, similar to the proportion of essential old genes, 18.9% ( $P = 0.6212$ , Fisher's Exact Test). We considered a low knockdown efficiency as shown by the 47% of GD lines whose knockdowns are expressed at the level of 30% or higher of the control (Fig 2A), suggesting that 47% of RNA lines are invalid for the testing and should be subtracted from the total tested lines.

Thus, the actual proportion of essential genes can be estimated by correcting for the bias of false positives ( $F_p$ ) and false negatives ( $F_n$ ) by following formula:

$$\text{Corrected proportion of essential genes} = [E - (T \cdot F_p)] / [T - (T \cdot F_n)]$$

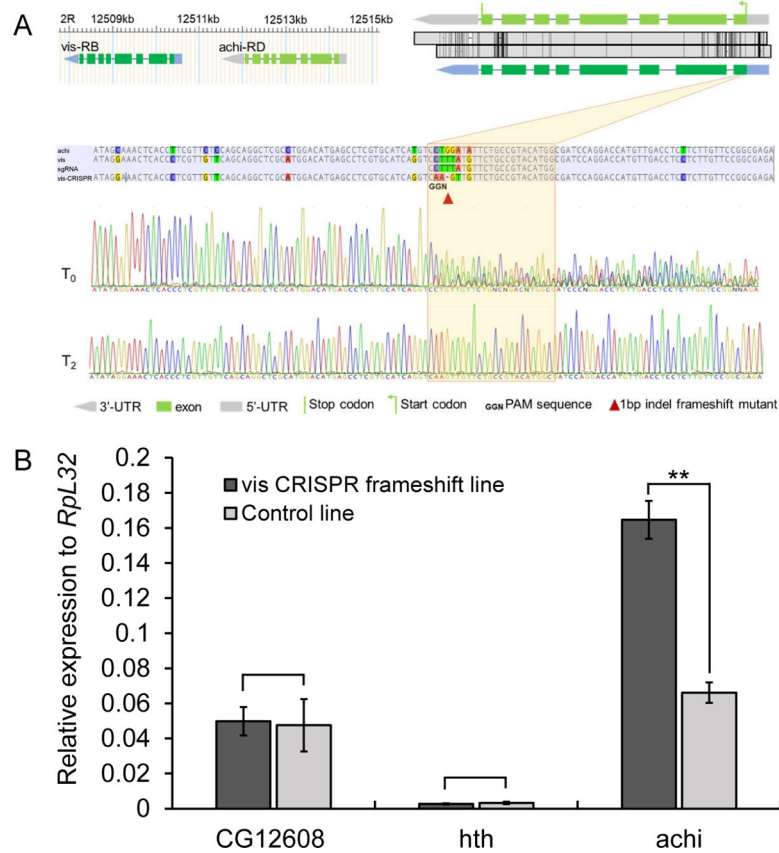


**Fig 5. Lethality proportion of 702 *Drosophila*-specific genes.** A. Lethality proportion of 702 *Drosophila*-specific genes in 6 ancestral stages of extant *D. melanogaster*, compared to the lethality proportion of 10,652 genes older than 40 Mya. No stages show an essentiality proportion significantly different from that of old genes (0.189). B. Lethality proportion of 702 *Drosophila*-specific genes based on three origin mechanism catalogs. No catalog shows a lethality proportion significantly different from that of old genes (0.189).

<https://doi.org/10.1371/journal.pgen.1009654.g005>

Where E and T are observed number of essential genes and total number of genes examined, respectively.  $F_p$  was measured as 1.6% [29] while  $F_n$  as 47% as estimated above or 39.9% as measured previously [29]. Thus, the estimated proportion of essential genes after correcting false positives and false negatives can be as high as 36.5% for the estimated false negative rate of 47% in this study. The corrected proportion can be also as high as 32.2% given the previously measured false negative rate of 39.9%. Furthermore, all six stages show a stable proportion of essential genes; none of the proportions is statistically different from the proportion of old genes (Fig 5A). Meanwhile, lethal rates of new genes which belong to three origin mechanism categories (DNA-based duplication, RNA-based duplication and orphan genes) [15] also show no significant difference (Fig 5B). Interestingly, 21.7% of orphan genes, some of which might be *de novo* genes [20], are essential. These data add new insight into the evolution of essentiality in all ancestral stages: soon after genes originated and fixed in *D. melanogaster*, a stable and high proportion of new genes is essential throughout entire evolutionary process from ancient ancestors to the speciation of *D. melanogaster*.





**Fig 6. CRISPR/Cas9 frameshift mutant could induce compensatory effect in *Drosophila*.** A. Design of CRISPR/Cas9 mutant. We targeted a randomly chosen young gene, *vis*, which emerged via duplication of *achi* in the common ancestor of *melanogaster* species complex. The genomic arrangement of two genes are shown in the upper left panel with the boxes referring to exons and connecting lines as introns. The pair shares a high sequence identity (0.92) in their 9 exons, which is schematically shown in the upper right panel. The middle panel shows the diverged site between *vis* and *achi*, which was chosen to design a short guide RNA (sgRNA) specifically targeting *vis*. The mutation (CTTTA→AAGT) was marked with a red triangle. The raw sanger sequencing data for the initial generation (T<sub>0</sub>) and the second generation of offspring (T<sub>2</sub>) was shown. B. The compensation effect of *achi*. In the frameshift mutant of *vis*, *achi*'s expression is significantly increased (P = 0.0003). By contrast, the unrelated *CG12608* and the remotely related *hth* did not show any significant upregulation. *RpL32* was used as a control as in [32].

<https://doi.org/10.1371/journal.pgen.1009654.g006>

### New gene duplicates show compensation effects in CRISPR frameshift mutants

It is now well documented that vertebrate cells such as mammalian cells or zebrafish cells recognize such aberrant mRNAs and compensate for their loss by increasing expression of genes with high sequence similarity, such as paralogs in zebrafish, worm and other organisms [37–41]. This has the effect of producing false negatives especially for recent duplicates that usually share high sequence similarity with parental duplicate copies. We detected a similar compensation effect in new gene duplicates in *Drosophila*.

We induced a one-nucleotide deletion using CRISPR/Cas9 into the ORF region of *vismay* (*vis*), a *D. melanogaster*-specific gene duplicated from a parental gene, *achintya* (*achi*), 0.8 Mya, with a nucleotide similarity of 92% between the two copies. We found that *achi* in the *vis* mutant was significantly upregulated whereas a randomly selected unrelated gene *CG12608* and the distantly related gene *hth* (nucleotide similarity of 45%) to *vis*, is not impacted by the *vis* mutation (Fig 6). Although the generality of the association between the *vis* mutation and

enhanced expression level of its highly similar duplicate copy has yet to be further tested, its implication to test phenotypic effects of new gene duplicates is clear: the CRISPR knockouts do not offer gold-standard for detecting phenotypic effects.

## Discussion

Our experimental and data analyses yielded fresh insights into the conceptual and technical issues raised with the progresses in understanding of evolution of gene essentiality. We showed that the repeatability of *Drosophila* knockdown experiments from RNA libraries between independent researchers separated by several years is as high as 85.4%, a level inspirable for using the technique to detect phenotypic effects of genes.

On the other hand, we also found that the knockdown efficiency is generally low with publicly available RNAi libraries leading to a high false negative rate. This reveals a high conservative property of the technology, suggesting that actual proportion of essential genes is higher than the measured proportion. A failure to understand this property of RNAi knockdown erroneously led to a confusion of false negative as false positive in detection of essential phenotypes [27] (S1 File). Our genome-wide analysis of essentiality in the *Drosophila* phylogenetic tree reveals that the proportion of essentiality with the proportions of new genes are not significantly variable with evolutionary time periods as short as 3 million years to 14 million years. These genes generated within the *sophophora* subgenus lineages (<35 million years) are as similarly frequent as the older genes dated for more distantly diverged *Drosophila* lineages or the ancient genes in non-*Drosophila* ancestors.

Detailed case analyses for new gene functions provide several lines of evidence in support of the essentiality of several new genes in development. First, Ross *et al* reported a stepwise neofunctionalization evolution in which a centromere-targeting gene in *Drosophila*, *Umbrea*, was generated less than 15 Mya [19]. Both RNAi knockdown, rescue experiments and P-element mediated gene knockout revealed that *Umbrea* evolved a species-specific essentiality to target centromere in chromosome segregation [19,25]. Second, Lee *et al* recently detected stage-specific (embryos/larvae/pupa) lethality associated with RNAi knockdown and CRISPR knockout in *Cocoon*, a gene emerged 4 Mya in the common ancestor of the clade of *D. melanogaster-simulans* [18]. These data show that *Cocoon* is essential for the survival at multiple developmental stages, including the critical embryonic stage. Third, P-element insertion/excision experiments show the essentiality of *K81* as a paternal element in early development. This gene only exists in the *Drosophila melanogaster*-subgroup species that diverged 6 Mya [42]. Fourth, Zeus, a gene that duplicated from the highly conserved transcription factor *CAF40* 4 Mya in the common ancestor of *D. melanogaster* and *D. simulans* rapidly evolved new essential functions in male reproductive functions, as detected in the null mutants and knockdown [43,44]. Fifth, A pair of extremely young duplicates, *Apollo* (*Apl*) and *Artemis* (*Arts*), was found to have been fixed 200,000 years ago in *D. melanogaster* populations [32]. CRISPR-created gene deletions of these genes showed that both evolved distinct essential functions in gametogenesis and *Apl* critical function in development.

Finally, in a comprehensive functional and evolutionary analysis of the ZAD-ZNF gene family in *Drosophila* [17], 86 paralogous copies were identified with phenotypic effects detected by knockdown and knockout in *D. melanogaster*. It was found that the proportion ( $17/58 = 29.3\%$ ) of lethal copies in old duplicates (>40 Mya) and the proportion ( $11/28 = 39.3\%$ ) of lethal copies in *Drosophila*-specific duplicates (<40 Mya) are statistically similar. Development analysis of two recently duplicated copies, *Nicknack* and *Oddjob*, using thorough knockdown and knockout analyses provided compelling evidence that these recent duplicates are essential in development, falsifying the new gene knockout results of Kondo *et al* [27].

A well-supported hallmark of young new genes is their male-biased expression pattern (often testis/accessory gland specific) [45–47]. However, new genes in *Drosophila* quickly evolved essential functions in viability with an approximately equal lethal rate with old genes (Fig 6A). We further compared the lethal rates between the two sex-biased subgroups, especially male-biased genes, based on the same phenotypic data with Fig 6A and found that the lethal rate of those male-bias genes is significantly higher than non-male-bias genes. This is consistent with our previous observation in a pair of extremely young genes, *Apollo* (male-biased expressed) and *Artemis* (female-biased expressed), that were created just 200,000 years ago in *D. melanogaster* [32]. Both *Apollo* and *Artemis* showed lethality in development whereas the *Apollo* showed even stronger effects. These observations reveal a new and interesting phenomenology that may deserve attention: the development is coupling with the sex-specific expression in the importance of gene functionality.

Overall, these data from present and previous studies challenge a conventional belief in the antiquity of important gene functions in general [21,24,48,49] and in development in specific [22–23].

## Materials and methods

### RNAi strain construction

Since species-specific new genes are under-represented in public RNAi lines, we generated new RNAi lines following [29]. Briefly speaking, we designed RNAi reagents using the E-RNAi server (<http://www.dkfz.de/signaling/e-rnai3/>) and kept constructs with all possible 19-mers uniquely matching the intended target gene and excluded designs with >1 CAN repeat (simple tandem repeats of the trinucleotide with N indicates any base) [50]. Constructs were cloned into pKC26 following the Vienna Drosophila Resource Center's (VDRC's) KK library strategy (<http://stockcenter.vdrc.at>, last accessed 2 February 2016). We introgressed the X chromosome from Bloomington Drosophila Stock Center line 34772, which expresses  $\Phi$ C31 integrase in ovary under control of the *nanos* promoter, into the VDRC 60100 strain. Strain 60100 carries attP sites at 2L:22,019,296 and 2L:9,437,482 [26]. We ensured that our RNAi constructs were inserted only at the 2L:9,437,482 site using PCR following Green *et al* [26]. RNAi constructs were injected into the 60100- $\Phi$ C31 at 250 ng/ $\mu$ L. Surviving adult flies were crossed to *sna*<sup>ScO</sup>/CyO balancer flies (BDSC 9325) and individual insertion strains were isolated by backcrossing.

### RNAi screen

We knocked down target gene expression using driver lines constitutively and ubiquitously expressing GAL4 under the control of either the *Actin5C* or  *$\alpha$ Tubulin84B* promoter. We replaced driver line's balancer chromosomes with GFP-marked chromosomes to track non-RNAi progeny. Control crosses used flies from the background strains 60100- $\Phi$ C31, 25709, or 25710 crossed to driver strains. Five males and five virgin driver females were used in each cross. Crosses were grown at 25°C, 40% - 60% humidity, and a 12h:12h light: dark cycle. F1 progeny were counted at day 19 after crossing, after all pupae had emerged. We screened F1 RNAi flies for visible morphological defects in 1) wings: vein patterning and numbers, wing periphery; 2) notum: general bristle organization and number, structure and smoothness; 3) legs: number of segments. We monitored survival of RNAi F1s by counting GFP and non-GFP L1, L3 larvae and pupae. We tested RNAi F1 sterility by crossing individual RNAi F1 flies to 60100- $\Phi$ C31 and monitoring vials for L1 production. Ten replicates for each sex for each line were performed.

## RNAi knockdown specificity and sensitivity

We sought to address two known problems of RNAi technology using RT-qPCR. First, since off-target effects are often discussed in RNAi experiments [29] we need to test whether target gene expression are specifically knocked down, although our constructs are computationally predicted to be specific. Second, since the RNAi knockdown is often incomplete [29], we need to estimate how many genes are adequately knockdown in expression. We targeted a random dataset of 14 *D. melanogaster*-specific genes. We collected qPCR primers from FlyPrimerBank [51]. For those genes not found in FlyPrimerBank, we took Primer-BLAST to design primers by specifically targeting a ~100 bp region of the gene (S5 Table). We confirmed primer specificity with PCR and Sanger sequencing.

We randomly selected 75 KK RNAi lines (no *tio* site insertion) to analyze their knock down efficiency. We cross these 75 KK RNAi lines with same driver which was used in Dietzl *et al* [29] for GD RNAi line knock down efficiency test. We extracted RNA from sets of 8 adult males (2~4 day old) in triplicate from each RNAi cross using TRIzol (Catalog# 15596-026, Invitrogen, USA), treated ~2  $\mu$ g RNA with RNase-free DNase I (Catalog# M0303S NEW ENGLAND Biolabs, USA), then used 1  $\mu$ L treated RNA in cDNA synthesis with SuperScript III Reverse Transcriptase (Invitrogen, USA) using oligo(dT)<sub>20</sub> primers. cDNA was diluted 1:40 in water before using 1  $\mu$ L as template in 10  $\mu$ L qPCRs with Universal SYBR Green Supermix (Catalog# 1725121, Bio-Rad, USA) and 400 nM each primer. Reactions were run on a Bio-Rad C1000 Touch thermal cycler with CFX96 detection system (BioRad, CA). Cycling conditions were 95°C for 30 sec, then 45 cycles of 95°C for 5 sec, 60°C for 30 sec, and 72°C for 15 sec. We normalized gene expression levels using the  $\Delta\Delta C_T$  method and *RpL32* as the control [29,52]. We tested the specificity and efficiency (90% < qPCR Efficiency < 110%) of qPCR primers using an 8-log<sub>2</sub> dilution series for each primer pair [32].

## Testing compensation effects of new gene duplicates

We generated the frameshift mutation line of *vis* using the CRISPR protocol previously developed [32] but with one single sgRNA for one gene as Kondo *et al* did [27]. The sgRNA-*vis* primer below was synthesized (the underlined sequence):

5'-GAAATTAATACGACTCACTATAGGATGTACGGCAGAACATAAGTTTAA  
GAGCTATGCTGGAA-3';

We used the following sequence-specific qRT-PCR primers to test the compensatory expression of *achi*, the duplicate of *vis*. Two control genes including *CG12608* and *hth* were examined too. Since *vis*'s expression is largely testis-specific, we extracted RNAs from testis of mated 4-day males and used qRT-PCR with 3 replicates to assess the expression, as developed previously [32].

248bp:

Achi-RT1F: 5'-AAAGTGACAGGTTTCTCTGTTTG-3';

Achi-RT1R: 5'-CTGATCCTCCTCCACGATGAC-3'.

237bp:

CG12608-RT1F: 5'-CATAGTGGGCACCTACGAG-3';

CG12608-RT1R: 5'-TGCGAGAGTATGATCTGCGAC-3'.

92bp:

hth-RT1F:5'-CCTAGTCATGTATCGCCGGTC-3';

hth-RT1R:5'-AGCGGATGTTTCATAAATCGCA-3'.

Internal control:

113bp:

RpL32-RT1F: 5'-AGCATAACAGGCCCAAGATCG-3';

RpL32-RT1R: 5'-TGTTGTCGATACCCTTGGGC-3'.

## Supporting information

**S1 Fig. Summary of new gene candidates in the K-dataset and G-dataset.** A. phylogenetic distribution of gene origination identified by the K-pipeline and the G-pipeline as shown in the two datasets. B. Evaluation of the two datasets based on individual gene analyses. The two datasets share 471 candidates (red). The G-dataset consist of 101 authentic candidates (deep blue) undetected in the K-dataset, 19 false positives (light purple), 18 dubious cases (green) and 45 cases not applicable for dating (sky blue). By contrast, the K-dataset includes 49 bona fide new gene candidates, 318 false positives, 102 dubious cases and 242 difficult cases. For the 318 false positive genes, 275 actually have orthologs in outgroup species; 32 noncoding or pseudogene models are taken as protein coding genes; 6 redundant entries of same genes are treated as different genes; 5 polycistronic coding genes are misdated. For possible reasons of false positive new gene caused by incorrect ortholog analyses: 1. existing orthologs not identified in the K dataset because of a different cutoff value in identifying an outgroup ortholog; 2. orthologs may not have been annotated in outgroup species; 3. genes without orthologs in adjacent outgroup species but actually having orthologs in more distant outgroup species. Note, the K-dataset mentions 1,182 genes in the main text, however its associated supplemental table includes 1,176 genes with 6 genes listed more than once.

(TIF)

**S2 Fig. Age dating between this work and Kondo et al.** This Figure, following [S1 Fig](#) in this paper, adds specific information on how we classified genes into six major categories or dozens of subcategories. For more details, please refer to [S1 File](#).

(TIF)

**S3 Fig. A representative difficult-to-date locus in the K-dataset.** A. The syntenic view of Ste locus between *D. melanogaster* and *D. simulans* shows fragmented continuity. Due to its multiplicative nature, Ste locus is difficult to assemble. In the UCSC Net track, the most assemblies can only reach level 2 of one-way syntenic mapping, rather than a better reciprocal syntenic mapping as level 1. B. Some orthologous region in *D. simulans* (lifted from *D. melanogaster*) is not anchored to the chromosome (X) and they are arbitrarily assembled as chrU. C. In *D. sechellia*, two scaffolds are assembled with the major scaffold super\_20 spanning 200 kb, in contrast to the assembly of 15 kb for the orthologous region of *D. melanogaster*.

(TIF)

**S1 Table. 103 knockdown experiments repeated by two independent works [25,28].** Note, works from [25] classified phenotypes as lethal, semi-lethal and viable. Since there are only few genes deemed as semi-lethal, we merged them into lethal gene groups to simplify.

(XLSX)

**S2 Table. For 86 new genes with different RNAi drivers, the consistency between different drivers in [25] and [28] is listed.**

(XLSX)

**S3 Table. The knockdown efficiency data of KK library and GD library.**

(XLSX)

**S4 Table. The genotyping results of 153 KK lines, the corrected lines by recombination and knockdown results.**

(XLSX)

**S5 Table. Primers for 75 KK lines knockdown efficiency tests.**

(XLSX)

**S6 Table. The list of 471 genes with the exact ages across the G-dataset and the K-dataset.** (This table consists of three sub-tables to show genes with the exact ages across the G-dataset and the K-dataset, genes with slightly younger ages in the K-dataset and genes with slightly older ages in the K-dataset, respectively).

(XLSX)

**S7 Table. The list of false negatives and false positives in the K-dataset.** Since the Pan-*Drosophilid* age group in the K-dataset corresponds to the age group 0, 1 or 2 in the G-dataset (S1A Fig), we simply replaced the Pan-*Drosophilid* age group as 0/1/2 in the table if applicable.

(XLSX)

**S1 File. Supplementary Materials and Method.**

(DOCX)

## Acknowledgments

We are grateful for valuable discussion with Edwin Ferguson, Urs Schmidt-Ott, Norbert Perri-mon, Emily Mortola, the M. Long lab members and Y.E. Zhang lab members for the technical development in RNAi and scientific issues related to this study. We also appreciate Lisa Meadows from VDRC for supplying haplotyping results for KK lines of the correct insertion site.

## Author Contributions

**Conceptualization:** Shengqian Xia, Nicholas W. VanKuren, Sidi Chen, Yong E. Zhang, Manyuan Long.

**Data curation:** Shengqian Xia, Nicholas W. VanKuren, Chunyan Chen, Li Zhang, Clause Kemkemer, Yi Shao, Hangxing Jia, UnJin Lee, Alexander S. Advani, Andrea Gschwend, Maria D. Vibranovski.

**Formal analysis:** Shengqian Xia, Nicholas W. VanKuren, Chunyan Chen, Li Zhang, Yi Shao, Alexander S. Advani, Manyuan Long.

**Funding acquisition:** Yong E. Zhang, Manyuan Long.

**Investigation:** Shengqian Xia, Nicholas W. VanKuren, Chunyan Chen, Li Zhang.

**Methodology:** Shengqian Xia, Nicholas W. VanKuren, Clause Kemkemer, UnJin Lee, Andrea Gschwend, Maria D. Vibranovski, Sidi Chen, Yong E. Zhang, Manyuan Long.

**Project administration:** Yong E. Zhang, Manyuan Long.

**Resources:** Shengqian Xia, Nicholas W. VanKuren, Chunyan Chen, Li Zhang, Clause Kemkemer, Yi Shao, Hangxing Jia, UnJin Lee, Alexander S. Advani, Andrea Gschwend, Maria D. Vibranovski, Sidi Chen.

**Software:** Shengqian Xia, Nicholas W. VanKuren, Li Zhang, Clause Kemkemer, Yi Shao, Hangxing Jia, Alexander S. Advani, Maria D. Vibranovski, Sidi Chen.

**Supervision:** Yong E. Zhang, Manyuan Long.

**Validation:** Shengqian Xia, Nicholas W. VanKuren, Chunyan Chen, Li Zhang, Hangxing Jia, UnJin Lee.

**Visualization:** Shengqian Xia, Nicholas W. VanKuren, Li Zhang, Clause Kemkemer, Yi Shao, Hangxing Jia, UnJin Lee, Alexander S. Advani, Andrea Gschwend, Maria D. Vibranovski, Sidi Chen, Yong E. Zhang, Manyuan Long.

**Writing – original draft:** Shengqian Xia, Nicholas W. VanKuren, Yong E. Zhang, Manyuan Long.

**Writing – review & editing:** Shengqian Xia, Nicholas W. VanKuren, UnJin Lee, Yong E. Zhang, Manyuan Long.

## References

1. Long M, Langley CH. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science*. 1993; 260(5104):91–5. <https://doi.org/10.1126/science.7682012> PMID: 7682012.
2. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet*. 2003; 4(11):865–75. <https://doi.org/10.1038/nrg1204> PMID: 14634634
3. Chen SD, Krinsky BH, Long MY. New genes as drivers of phenotypic evolution. *Nat Rev Genet*. 2013; 14(9):645–60. <https://doi.org/10.1038/nrg3521> WOS:000323280100012. PMID: 23949544
4. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, et al. Proto-genes and de novo gene birth. *Nature*. 2012; 487(7407):370–4. <https://doi.org/10.1038/nature11184> PMID: 22722833; PubMed Central PMCID: PMC3401362.
5. Ding Y, Zhou Q, Wang W. Origins of new genes and evolution of their novel functions. *Annual Review of Ecology, Evolution, and Systematics*. 2012; 43:345–63. <https://doi.org/10.1146/annurev-ecolsys-110411-160513>
6. McLysaght A, Hurst LD. Open questions in the study of de novo genes: what, how and why. *Nat Rev Genet*. 2016; 17(9):567. <https://doi.org/10.1038/nrg.2016.78> PMID: 27452112
7. Ruiz-Orera J, Verdaguer-Grau P, Villanueva-Cañas JL, Messeguer X, Albà MM. Translation of neutrally evolving peptides provides a basis for de novo gene evolution. *Nat Ecol Evol*. 2018; 2(5):890–6. <https://doi.org/10.1038/s41559-018-0506-6> PMID: 29556078
8. Xie C, Bekpen C, Kunzel S, Keshavarz M, Krebs-Wheaton R, Skrabar N, et al. A de novo evolved gene in the house mouse regulates female pregnancy cycles. *Elife*. 2019; 8. <https://doi.org/10.7554/eLife.44392> PMID: 31436535; PubMed Central PMCID: PMC6760900.
9. Vakirlis N, Acar O, Hsu B, Castilho Coelho N, Van Oss SB, Wacholder A, et al. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nat Commun*. 2020; 11(1):781. <https://doi.org/10.1038/s41467-020-14500-z> PMID: 32034123; PubMed Central PMCID: PMC7005711.
10. Witt E, Benjamin S, Svetec N, Zhao L. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. *Elife*. 2019; 8:e47138. <https://doi.org/10.7554/eLife.47138> PMID: 31418408
11. Jiang X, Assis R. Natural selection drives rapid functional evolution of young *Drosophila* duplicate genes. *Mol Biol Evol*. 2017; 34(12):3089–98. <https://doi.org/10.1093/molbev/msx230> PMID: 28961791
12. Rogers RL, Cridland JM, Shao L, Hu TT, Andolfatto P, Thornton KR. Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. *Mol Biol Evol*. 2014; 31(7):1750–66. <https://doi.org/10.1093/molbev/msu124> PMID: 24710518; PubMed Central PMCID: PMC4069613.
13. Schroeder CM, Tomlin SA, Valenzuela JR, Malik HS. A rapidly evolving actin mediates fertility and developmental tradeoffs in *Drosophila*. *BioRxiv*. 2020. <https://doi.org/10.1101/2020.09.28.317503>
14. Zhang L, Ren Y, Yang T, Li G, Chen J, Gschwend AR, et al. Rapid evolution of protein diversity by de novo origination in *Oryza*. *Nat Ecol Evol*. 2019; 3(4):679–90. <https://doi.org/10.1038/s41559-019-0822-5> PMID: 30858588
15. Shao Y, Chen C, Shen H, He BZ, Yu D, Jiang S, et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res*. 2019; 29(4):682–96. <https://doi.org/10.1101/gr.238733.118> PMID: 30862647; PubMed Central PMCID: PMC6442393.
16. Zhang YE, Vibranovski MD, Landback P, Marais GAB, Long MY. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol*. 2010; 8(10):e1000494. <https://doi.org/10.1371/journal.pbio.1000494> WOS:000283495100002. PMID: 20957185

17. Kasinathan B, Colmenares SU III, McConnell H, Young JM, Karpen GH, Malik HS. Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in *Drosophila*. *Elife*. 2020; 9: e63368. <https://doi.org/10.7554/eLife.63368> PMID: 33169670
18. Lee YCG, Ventura IM, Rice GR, Chen DY, Colmenares SU, Long MY. Rapid Evolution of Gained Essential Developmental Functions of a Young Gene via Interactions with Other Essential Genes. *Mol Biol Evol*. 2019; 36(10):2212–26. <https://doi.org/10.1093/molbev/msz137> WOS:000501734200011. PMID: 31187122
19. Ross BD, Rosin L, Thoma AW, Hiatt MA, Vermaak D, de la Cruz AFA, et al. Stepwise evolution of essential centromere function in a *Drosophila* neogene. *Science*. 2013; 340(6137):1211–4. <https://doi.org/10.1126/science.1234393> WOS:000319972800044. PMID: 23744945
20. Long MY, VanKuren NW, Chen SD, Vibranovski MD. New gene evolution: Little did we know. *Annu Rev Genet*. 2013; 47:307–33. <https://doi.org/10.1146/annurev-genet-111212-133301> PMID: 24050177
21. Ashburner M, Misra S, Roote J, Lewis S, Blazej R, Davis T, et al. An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. *Genetics*. 1999; 153(1):179–219. <https://doi.org/10.1093/genetics/153.1.179> PMID: 10471707
22. Gould SJ. *The structure of evolutionary theory*: Harvard University Press; 2002.
23. Carroll SB. *Endless forms most beautiful: The new science of evo devo and the making of the animal kingdom*: WW Norton & Company; 2005.
24. Krebs JE, Lewin B, Goldstein ES, Kilpatrick ST. *Lewin's essential genes*: Jones & Bartlett Publishers; 2013.
25. Chen SD, Zhang YE, Long MY. New genes in *Drosophila* quickly become essential. *Science*. 2010; 330(6011):1682–5. <https://doi.org/10.1126/science.1196380> PMID: 21164016
26. Green EW, Fedele G, Giorgini F, Kyriacou CP. A *Drosophila* RNAi collection is subject to dominant phenotypic effects. *Nat Methods*. 2014; 11(3):222. <https://doi.org/10.1038/nmeth.2856> PMID: 24577271
27. Kondo S, Vedanayagam J, Mohammed J, Eizadshenass S, Kan LJ, Pang N, et al. New genes often acquire male-specific functions but rarely become essential in *Drosophila*. *Genes Dev*. 2017; 31(18):1841–6. <https://doi.org/10.1101/gad.303131.117> PMID: 29051389
28. Zeng XK, Han LL, Singh SR, Liu HH, Neumuller RA, Yan D, et al. Genome-wide RNAi screen identifies networks involved in intestinal stem cell regulation in *Drosophila*. *Cell Rep*. 2015; 10(7):1226–38. <https://doi.org/10.1016/j.celrep.2015.01.051> WOS:000349918700017. PMID: 25704823
29. Dietzl G, Chen D, Schnorrrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*. 2007; 448(7150):151–6. <https://doi.org/10.1038/nature05954> PMID: 17625558
30. Vissers JHA, Manning SA, Kulkarni A, Harvey KF. A *Drosophila* RNAi library modulates Hippo pathway-dependent tissue growth. *Nat Commun*. 2016; 7:10368. ARTN 10368 <https://doi.org/10.1038/ncomms10368> WOS:000369022100005. PMID: 26758424
31. Huang N, Lee I, Marcotte EM, Hurler ME. Characterising and predicting haploinsufficiency in the human genome. *Plos Genet*. 2010; 6(10):e1001154. <https://doi.org/10.1371/journal.pgen.1001154> PMID: 20976243; PubMed Central PMCID: PMC2954820.
32. VanKuren NW, Long MY. Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nat Ecol Evol*. 2018; 31(4):705–12. <https://doi.org/10.1038/s41559-018-0471-0> PMID: 29459709
33. Zhou Q, Zhang GJ, Zhang Y, Xu SY, Zhao RP, Zhan ZB, et al. On the origin of new genes in *Drosophila*. *Genome Res*. 2008; 18:1446–55. <https://doi.org/10.1101/gr.076588.108> PMID: 18550802
34. Zhang YE, Vibranovski MD, Krinsky BH, Long M. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Res*. 2010; 20(11):1526–33. <https://doi.org/10.1101/gr.107334.110> PMID: 20798392.
35. Rogers RL, Hartl DL. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Mol Biol Evol*. 2012; 29(2):517–29. <https://doi.org/10.1093/molbev/msr184> PMID: 21771717
36. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 2007; 450(7167):203–18. <https://doi.org/10.1038/nature06341> PMID: 17994087.
37. Rossi A, Kontarakis Z, Gerri C, Nolte H, Holper S, Kruger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*. 2015; 524(7564):230–5. <https://doi.org/10.1038/nature14580> WOS:000359386900034. PMID: 26168398
38. El-Brolosy MA, Stainier DYR. Genetic compensation: A phenomenon in search of mechanisms. *Plos Genet*. 2017; 13(7):e1006780. <https://doi.org/10.1371/journal.pgen.1006780> PMID: 28704371; PubMed Central PMCID: PMC5509088.



39. El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, et al. Genetic compensation triggered by mutant mRNA degradation. *Nature*. 2019; 568(7751):193–7. <https://doi.org/10.1038/s41586-019-1064-z> PMID: 30944477
40. Ma Z, Zhu P, Shi H, Guo L, Zhang Q, Chen Y, et al. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature*. 2019; 568(7751):259–63. <https://doi.org/10.1038/s41586-019-1057-y> PMID: 30944473.
41. Serobyán V, Kontarakis Z, El-Brolosy MA, Welker JM, Tolstenkov O, Saadeldein AM, et al. Transcriptional adaptation in *Caenorhabditis elegans*. *Elife*. 2020; 9. <https://doi.org/10.7554/eLife.50014> PMID: 31951195; PubMed Central PMCID: PMC6968918.
42. Loppin B, Lepetit D, Dorus S, Couble P, Karr TL. Origin and neofunctionalization of a *Drosophila* paternal effect gene essential for zygote viability. *Current Biology*. 2005; 15(2):87–93. <https://doi.org/10.1016/j.cub.2004.12.071> WOS:000226858600018. PMID: 15668163
43. Chen SD, Ni XC, Krinsky BH, Zhang YE, Vibranovski MD, White KP, et al. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *EMBO J*. 2012; 31(12):2798–809. <https://doi.org/10.1038/emboj.2012.108> PMID: 22543869.
44. Matteuzzo Ventura I. Functional Evolution of Young Retrogenes with Regulatory Roles in *Drosophila*. 2019.
45. Betrán E, Thornton K, Long M. Retroposed new genes out of the X in *Drosophila*. *Genome Res*. 2002; 12(12):1854–9. <https://doi.org/10.1101/gr.6049> PMID: 12466289
46. Vibranovski MD, Lopes HF, Karr TL, Long MY. Stage-Specific Expression Profiling of *Drosophila* Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. *Plos Genet*. 2009; 5(11). <https://doi.org/10.1371/journal.pgen.1000731> WOS:000272419500026. PMID: 19936020
47. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 2010; 20(10):1313–26. <https://doi.org/10.1101/gr.101386.109> PMID: 20651121
48. Jacob F. Evolution and tinkering. *Science*. 1977; 196(4295):1161–6. <https://doi.org/10.1126/science.860134> PMID: 860134.
49. Mayr EJ. *The Growth of Biological Thought—Diversity, Evolution, and Inheritance*. New York Rev Books. 1982; 29(8):41–2. WOS:A1982NM03000018.
50. Ma Y, Creanga A, Lum L, Beachy PA. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature*. 2006; 443(7109):359–63. <https://doi.org/10.1038/nature05179> PMID: 16964239.
51. Hu Y, Sopko R, Foos M, Kelley C, Flockhart I, Ammeux N, et al. FlyPrimerBank: an online database for *Drosophila melanogaster* gene expression analysis and knockdown evaluation of RNAi reagents. *G3*. 2013; 3(9):1607–16. <https://doi.org/10.1534/g3.113.007021> PMID: 23893746; PubMed Central PMCID: PMC3755921.
52. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*. 2001; 25(4):402–8. <https://doi.org/10.1006/meth.2001.1262> WOS:000173949500003. PMID: 11846609