Supporting Online Material for

New Genes in *Drosophila* Quickly Become Essential

Sidi Chen, Yong E. Zhang, Manyuan Long*

*To whom correspondence should be addressed. E-mail: mlong@uchicago.edu

Published 17 December 2010, *Science* **330**, 1682 (2010)
DOI: 10.1126/science.1196380

This PDF file includes:

Materials and Methods

Identification of new gene origination events and gene age dating in *Drosophila*

We used the original gene age assignment in *(S1)* to generate a refined young gene dataset. In brief, for each *D.melanogaster* gene, we first inferred the presence or absence of its ortholog in other *Drosophila* species by syntenic genomic alignment (Fig. S1a, b) generated by the University of California – Santa Cruz (UCSC) bioinformatic group *(S2).* Then, we assigned the origination timing by following the parsimony rule (Fig. S1a). Finally, we identified 947 young genes, which originated after the split of *Sophophora* and *Drosophila* subgroup. Extensive estimation showed that our method tends to generate a conservative dataset of new genes, namely to assign genes with relatively older ages compared to previous related work *(S1).*

We further improved the young gene dataset for experimental characterization. Specifically, first, we considered a gene to be a "new gene" (or a "young gene") if it is less than 35 Myr old (arising after the divergence between *D. melanogaster* and *D. willistoni*). In other words, only genes with branch 2~6 of *(S1)* were considered. Second, we manually verified all these cases based on both UCSC genome-alignment based synteny and FlyBase protein-based synteny map to exclude genes with multiple losses. For example, a gene (*CG7627*) that was lost in *D.grimshawi, D. willistonni* and the *Obscura* group was assigned as branch 4 in *(S1),* and was excluded from the young gene dataset in this study. Moreover, genes with any slight homology in *Drosophila* subgroup or *D.willistoni* revealed by Tblastn track of UCSC were also excluded. Third, we considered a gene to be an "old gene" if it is older than 40 Myr (shared by all 12 sequenced *Drosophila* species). Fourth, genes with orthologs that are present in *D. willistoni* but not in all sequenced *Drosophila* subgenus species were excluded in either group due to the uncertainty in their age inferences. With these improvements, we identified 566 young genes that had originated within the past 35 million years (Myr) and 11,909 old genes that are shared by the 12 *Drosophila* genomes (>40 Myr). The refinement of the young gene dataset has excluded 180 (or 22%) previous cases *(S1),* and expanded with 19 new entries that were not assignment according to the presence/absence information of UCSC, but instead assigned through manual verification of FlyBase protein-based synteny maps. The current new gene dataset is more conservative and has higher accuracy in gene age dating.

As performed previously *(S1, S3, S4),* we then classified the new genes into DNA-based duplicated genes, RNA-based retroposed genes, and *de novo* genes according their DNA sequence signatures (Fig. S1 b-c) including the existence of parental genes, number of exons and introns. Briefly, a DNA-based duplicated gene has more than one exon, and has at least one existing paralogs; an RNA-based retroposed gene is intronless, and has a multi- exon parental gene. A *de novo* gene does not have an identifiable paralog. After classifying a *de novo* gene, we collected three pieces of evidence for confirmation: first, no similarity hits could be found outside the clade where the gene was located in a sequence search against the NCBI NR protein database; second, it is unique in the *D. melanogaster* genome and has no paralogs; finally, it does not encode any known protein domains in the FlyBase annotation.

The identification of gene structure renovation (Fig. S1 b-c) and the calculation of protein sequence divergence were carried out as described previously (*S4, S5*). Peptide sequence data for the new genes were retrieved from mass spectra databases (PeptideAtlas (*S6*)).

RNAi lines targeting young genes

All UAS-IR RNAi lines were ordered from the RNAi library of Vienna Drosophila RNAi Center (VDRC) (*S7*) in Austria. This RNAi library is known to be potent and specific in gene silencing, showing efficient reductions in the mRNA levels of genes of interest (70–90% on average) and low false-positive rates in phenotype detection (2%) (*S7*). Briefly, we first searched for available RNAi lines targeting 566 young genes. To avoid off-target effects, we excluded the RNAi lines with any level of off-target 19-mers (*S7*), retaining only the lines with off-target = 0 (no off-target 19-mer matches anywhere in the genome) and s19 = 1 (100% on-target 19-mer matches). Therefore genes that are highly similar to their paralogs were not included in this study due to technical unavailability. Moreover, we removed lines with lethal or sterile phenotypes caused by insertion of the P-element carrying the UAS-IR construct before the induction of Gal4. By applying these filters, we obtained a set of specific RNAi lines for 195 young genes. For a few genes there are multiple RNAi lines with independent constructs available, which we used for confirmation of phenotypes.

Other fly stocks, crosses and lethality phenotyping

The following Gal4 lines were ordered from the Bloomington stock center:

yw; Act5C-Gal4,w+/CyO,y+

yw; ; TubP-Gal4,w+/Tm3,Sb

w,Bx(ms1096)-Gal4,w+; UAS-dcr2,w+

w,UAS-dcr2,w+; bbg(C96)-Gal4,w+

w,UAS-dcr2,w+; pannier-Gal4,w+/TM3,Ser

All flies were raised on standard cornmeal media (*S8*) under standard laboratory conditions (21–25℃, 40–60% humidity, 12 h: 12 h daylight cycle).

As shown briefly in a schematic representation (Fig. S2), the constitutive RNAi knockdown of new genes was performed by crossing the UAS-RNAi lines to a constitutive Act5C-Gal4 driver line balanced by CyO (*yw; Act5CGal4, w+/CyO,y+)*. We allowed F1 embryos to develop for two to three weeks into adults. Two morphological markers were used to distinguish RNAi and control F1s: red eye, normal-winged offspring were recognized as RNAi F1s, and yellow (or orange) eye, curly winged offspring were recognized as non-RNAi control F1s. We scored approximately 100 flies for each vial of each line crossed and followed the canonical definition of lethality as described previously (*S7, S9*). The ratio of RNAi:control F1s was expected to be 1:1 if the RNAi F1s were fully viable; it was expected to be 0:1 if the RNAi was completely lethal. For a gene targeted by Act5C-Gal4 RNAi, we classified it as lethal if no RNAi F1 hatched into adults or only a few (<5%) RNAi F1 hatched into adults but died very early in adulthood; we classified a gene as semi-lethal if less than 20% of the RNAi F1s hatched into adults. With respect to the sex of lethality, “male lethal” meant that lethality was specific to male flies; “female lethal” meant that lethality was specific to female flies, while “both” meant that lethality occurred in both male and female flies. For the

lines that showed lethal phenotypes, more than four replicate crosses were performed to confirm lethality. As a pilot experiment, we tested genes known to be essential (*Caf40, RpS15Ab) (S10)* or non-essential *(S7)* for viability. As expected, silencing of known essential genes was lethal, whereas silencing of non-essential genes was viable. We then carried out crosses for all young genes and controls in parallel.

Second Gal4 driver verification of RNAi silencing

We used a second constitutive Gal4 driver line, *yw;; TubP-Gal4,w+/Tm3,Sb*, which was crossed with the Act5C-Gal4 lethal RNAi lines. Lethal phenotyping was carried out similar to that performed for the *yw; Act5C-Gal4,w+/CyO,y+* driver line, with the exception of balancer marker (*Tm3,Sb*) recognition.

Determining developmental stages of lethality

We examined the stages in which lethality occurred with two methods: (I) direct observation / dissection and (II) gene inactivation/silencing with fluorescence tracking. These two methods showed consistent results, while each has its own advantage.

In the first method, we allowed all larvae and pupae to develop for two additional weeks to exclude the effect of slow development. If all control F1 pupae hatched and (a) all RNAi F1 pupae (approximately 50% of total pupae) died (necrosis, color became darker), thus failing to form adults, we classified the gene as "pupal lethal;" (b) all RNAi F1s died before starting pupation, we classified the gene as "before pupal lethal"; (c) some RNAi F1s died before pupation, while some died in the pupal case, we classified the gene as "mixed stage lethal." For each case, we verified the developmental stage by examining adults and pupae in multiple vials under a stereoscope (Olympus, Center Valley, PA). For those pupal lethal cases, we dissected out the multiple individual pupal case cuticles to examine their morphological features, especially the formation of rudimentary head structures, early wing structures and leg structures, to determine the developmental stage of the aborted pupae (Fig. S3). We imaged these individuals using an Olympus dissecting microscope with a digital camera DP70 (Olympus, Center Valley, PA).

In the second method, we first recombined the Act5C-Gal4 driver with a green fluorescence reporter UAS-mCD8-GFP to generate a line Act5C-Gal4,UAS-mCD8-GFP with CyO balancer. We used this line to cross to UAS-IR lines for constitutive RNAi, where the GFP reporter allows tracking of the development of the RNAi animal by fluorescence stereoscope, as shown with genotypes:

Act5C-Gal4,UAS-mCD8GFP/CyO X UAS-RNAi
RNAi F1 animal: Act5C-Gal4,UAS-mCD8GFP>>UAS-RNAi
Control F1 animal: CyO/(;)UAS-RNAi

These two methods are consistent in lethal stage classification. Furthermore, we observed that the "before pupal lethal" cases with the first method are all larval lethals, with no embryonic lethals.

Tissue-specific gene silencing and electron microscope (EM) imaging

We crossed representative RNAi lines with constitutive RNAi lethals to the following tissue-specific Gal4 driver lines:

w,Bx-(ms)1096-Gal4,w+; UAS-dcr2,w+ ,

w,UAS-dcr2,w+; bbg(C96)-Gal4,w+
w,UAS-dcr2,w+; pannier-Gal4,w+/TM3,Ser

We then examined multiple individual F1s under a dissecting microscope for the presence of morphological/developmental defects. To image detailed morphological defects at high resolution, we mounted cuticles of RNAi flies on sample holders (Ted Pella, Redding, CA) coated with a Pt/Pd alloy with a sputter coater (Cressington, UK) and imaged them using a scanning electron microscope (FEI, Hillsboro, OR).

Expression analysis of essential young genes

We retrieved updated expression data over the course of the *D. melanogaster* life cycle from http://flybase.org/reports/FBrf0205914.html (*S11, S12*). For a gene of interest, the two channels of fluorescence intensity values corresponded to the mRNA abundance of the sample at a particular time point and the sample of all pooled time points. The log ratio of these two values (M value) was used to measure the relative gene expression levels across all of the 66 time points. We mapped probes to genes based on chromosomal coordinates and discarded those that mapped to different genes. We generated a heatmap of 59 young essential genes (Fig. S4) using the gplots package of the Bioconductor platform (*S13*).

Sequence evolution and population genetics

Orthologous sequences were parsed out from the multiple-species alignment from UCSC. Protein alignments were built and then converted to codon-based alignments (*S14, 15*). *Ka/Ks* ratios were estimated with a free ratio model using the CODEML program (*S14*).

Polymorphism and divergence data were retrieved from DPGP (*S16, S17*). We used the version 1.0 of DPGP containing polymorphism data of 37 lines of *D.melanogaster* and performed population genetic analysis using a similar pipeline to previous work (*S1*). In brief, for each gene, the chromosomal coordinates for the longest coding open reading frame were retrieved from the flyBaseGene table of UCSC genome browser (*S2*). Coding sequences together with flanking 1,000 base pairs were extracted with nucleotides showing low quality score (<20 or >1% sequencing error rate) masked as “N”. Given existence of “N” and deletions or insertions, genewise (*S18*) was used to conceptually translate each coding sequence with the annotated reference protein as the template. Orthologous *D. simulans* coding regions were extracted from UCSC syntenic genome alignment, which was similarly fed into genewise. Conceptual translations for 37 lines together with the orthologous protein of *D. simulans* were aligned with the linsi function of MAFFT package (*S19*). Strains with indels or “N” contributing to more than half of the alignment were filtered out. Then, codon based alignments were reverse translated based on this protein alignment using the PAL2NAL package (*S20*).

Based on the alignments, we then used the PGEToolbox package (*S21*) to count number of synonymous polymorphism (Ps), number of non-synonymous polymorphism (Pn), number of synonymous divergence (Ds) and number of non-synonymous Divergence (Dn), respectively. A maximum-likelihood based package (DOEF(*S22*)) was implemented to estimate group level α. Likelihood ratio tests were performed in the DOEF package.

Calculation of the proportion of essential genes in the old gene group

Data on Act5C-Gal4 RNAi lethality were retrieved from Supplementary Table 1 and Table 4 from (*S7*). We chose the lines belonging to the "random set" and Cross-Ref with CG numbers according to the latest FlyBase annotation. We found 443 genes with CG ID, construct ID and transformation ID that were consistent between all lines and tables. We then applied the same RNAi linefiltering criteria (off-target, position effect) as were used for the young genes. After filtering, we narrowed our set to 300 genes, of which 275 were covered in our branch assignment and age dating and 245 were old genes (branch = 0, age >40 Myr). Among these genes, 86 were Act5C-RNAi lethal. The numbers of genes with lethality at different stages were counted according to the information of "lethal stage" in (*S7*).

Modeling interaction network of young essential genes

Gene-gene interaction data, including protein-protein interactions, genetic interaction, microRNA-target gene interactions, were retrieved from IntAct *(S23)*, BioGrid *(S24)* and miRBase *(S25)*. Direct interactions of young essential genes identified in this study were used to draw interaction network model.

Supplemental Figures

Figure S 1. Schematic representation of young gene identification, age dating and chimeric structure analysis **(A)** schematic representation of young gene identification and age dating; *Drosophila* phylogeny is shown on the left; genome synteny alignment is shown on the right; genomes from each species were shown as purple solid lines; grey blocks represent aligned regions; orange boxes represent orthologous genes existing in the melanogaster subgroup but not the outgroups; orange star denotes the origination event on the phylogeny; evolutionary time estimates are shown under the phylogeny (not drawn to scale); **(B)** An example of chimeric young gene *CG11639 (TfIIA-S-2)* in a UCSC-genomebrowser- based figure; Between-species genomic alignment (Net) tracks show that *CG11639* is present in the melanogaster subgroup revealed by a syntenic chain, but not in the other *Drosophila* species; **(C)** similar to (**B**) the black alignment track shows where the parental gene aligns with *CG11639*. The parental gene contributes to the coding region, but the new copy recruited local sequence to generate its new UTR region.

Figure S 2. Schematic representation of the screening for lethal phenotypes in young genes under constitutive RNAi by Act5C-Gal4 driver; GOI, gene-of interest. Genotypes of the starting parents (F0) and offsprings (F1) were shown in black; phenotypic markers to distinguish F1 were shown in green.

Figure S 3. Representative lethal phenotypes of young essential genes at different developmental stages examined by dissection **(A)** Normal pharate adults dissected from developing pupae of control animals; **(B-E)** dead individuals following constitutive silencing of young essential genes by Act5C-RNAi, with developmental halt at different stages; **(B)** mixed-stage lethals; **(C)** prepupal or early stage pupal lethals; **(D)** pharate or eclosion stage lethals; and **(E)** late pupal or pharate stage lethals. The annotation symbol (CG number) of each young essential gene is labeled above each image; arrowheads point to developing head structures. The images with higher resolution are available upon request.

Figure S 4. Expression profiles of young essential genes during the life cycle of *D. melanogaster* Each row represents one gene; Log-ratio of expression of a particular time point to that of the whole-life-cycle-pool sample is shown in each cell; Color key and histogram were shown on the bottom-right corner; (Raw array data from (*12*)); the developmental stage of lethality of each gene was shown with color coding illustrated at bottom left; abbreviations of stages: L2, second instar larva; L3, third instar larva; P, pupae, PP, prepupae, EP, early pupa; PH, pharate (late pupa); A, adult.

Figure S 5. Chromosomal distribution of young and old genes essential for Viability; The barplot shows proportion of genes being essential for viability across five major chromosomal arms in young or old age groups; autosomal data were shown in grey or black bars, X-chromosome data were shown in orange bars.

Figure S 6. A simplified interaction network map of young essential genes **(A)** core components of young essential gene (YEG) interaction network complex that are linked

by primary connections; **(B)** peripheral components that are not linked to the core with existing data.

Figure S 7. Molecular evolutionary rates of young essential genes and their parental genes; (**A**-**B)** Comparison of *Ka/Ks* ratios between young essential genes and their parental genes with pairwise x-y plot (**A**) and (**B**) Boxplot; Wilcoxon rank test *p* value for comparison between these two gene groups is shown in the panel; **(C**-**D)** Phylogenetic representation of sequence evolution of representative young essential genes; **(C)** a *de novo* young essential gene (*CG31882*) and **(D)** a duplication-generated young essential gene (*RpS28a*); Number of non-synonymous substitutions, number of synonymous substitutions and *Ka/Ks* (if available) are shown above the related evolutionary branches; On branches where no number is shown, no substitutions occurred. Young essential gene lineages are shown in red, parental gene (if applicable) lineages are shown in blue.

Figure S 8. Natural selection on essential genes; α values and their 95% confidence intervals were plotted for five gene groups: branch 4-5 young essential genes (6~11 Myr), parental genes of branch 4-5 young essential genes, branch 2-3 young essential genes (11~35 Myr), parental genes of branch 2-3 young essential genes and old essential genes (>40 Myr).

Supplemental Tables

Table S1. Phenotype of young essential genes by constitutive silencing;
A Transf_ID is the transformation ID of an RNAi line; Act5C-Gal4 RNAi lethality is scored according to materials and methods and schemed in Fig. S2; lethality sex is the particular sex that is lethal under Act5C-Gal4 RNAi; in these cases (female or male lethal), one sex is completely lethal, while the other sex is semilethal or viable; Pupal lethal phenotype class definition: Class I, pharate lethal; Class II, early pupal stage lethal; Class III, mixed pupal stage lethal.

Table S2. Origin and evolution of young essential genes
Footnotes:
(a) NA, not available
(b) Most genes listed are parental gene of the young gene, for a few cases where the parental-child relationship is ambiguous, closest paralogs are listed. In brief, we built a self-chained genome alignment by following UCSC's strategy. We aligned *D. melanogaster* against itself and built a series of homology blocks to look for the most similar paralogous genomic regions for one gene of interest.
(c) Origination mechanism, D = DNA-based duplication, R = RNA-based retroposition, A = de novo origination.
(d) “C” and “R” indicate chimerism and structural re-organization, respectively. In the former case, the young gene recruits the adjacent regions as a new functional element such as 3’UTR. In the later case, the young gene has similar sequence organization to parental gene, but its gene structure is different involving at least one introns gain/loss event. For chimerism, we required at least of 10% of region in the young gene that are not alignable with the parental gene.
(e) N, no structural renovation
(f) Divergence of protein sequences of young gene - parental gene pair
(g) In the case of *CG13559*, there is only one intron loss between parental gene and the young gene. It is difficult to infer whether this is a DNA-level duplication or retroposition with high divergence.

Table S3. Phenotype of constitutive silencing of all young genes tested;

Table S4. Representative F1 adult counting in lethal RNAi crosses
Brief cross scheme: Act5C-Gal4/CyO X UAS-RNAi; detail cross scheme in Fig. S2;
: number greater than 100; * : these RNAi F1s are very sick, died in a few days after hatching.

Table S5. Lethality stage examination with fluorescence tracking;
Brief cross scheme: Act5C-Gal4,UAS-mCD8GFP/CyO X UAS-RNAi; All GFP animals are Act5C-Gal4,UAS-mCD8GFP>>UAS-RNAi; while all non-GFP animals are controls CyO/(;)UAS-RNAi; * Developmental stage abbreviations: E = embryo, L1 = 1st instar larva, L2 = 2nd instar larva, L3 = 3rd instar larva, PP = prepupa, EP = early pupa, PH = pharate, A= adult, na, not available; lethal stage call: L2/L3: lethal at the transition between L2 and L3, similar for L3/PP, PP/EP, EP/PH, PH/A; MIX: mix stage lethal,

called according to dissection; P: pupal stage lethal, where lethality occurs at mixed pupal sub-stages or sub-stage unclear.

Table S6. Viability phenotype consistency of representative young genes; Part (I) Viability phenotype consistency between independent Gal4 driver constructs, where the phenotypes of two constitutive drivers, i.e. TubP-Gal4 and Act5C-Gal4 were compared using the same UAS-IR lines; Part (II) Viability phenotype consistency between independent UAS-IR lines, where the phenotypes of different UAS-IR lines were compared using the same Gal4 driver (in this case Act5C-Gal4).

Table S7. Tissue-specific RNAi phenotype of representative young essential genes; UAS-IR RNAi lines for representative young essential genes were crossed to wing-specific Gal4 driver Bx(ms)1096-Gal4 and notum midline specific Gal4 driver Pnr-Gal4, UAS-dicer2 was used for all these crosses and morphological abnormalities were observed under dissecting scope (details in materials and methods); * : nd: not determined; # : Representative phenotypes were imaged using EM and shown in Fig. 3.

Table S8. Peptide evidence for young essential genes
Data retrieved from PeptideAtlas (materials and methods), source of sample according to Sample_ID legends:

46	Heads from adult flies, soluble proteins
47	Heads from adult flies, soluble proteins
49	S2 cells, nuclear fraction
50	Kc cells membrane fraction
51	S2 cells; membrane fraction
52	Heads from adult flies, membrane fraction
54	Fatbodies from 20%sucrose-treated and untreated larvae
55	Fatbodies from rapamycin-treated and untreated larvae
56	Heamolymph from adult flies
57	Embryos up to the blastoderm stage (2hours after egg laying)
60	Cytoplasmic fractions (soluble proteins) from Adult Drosophila heads
61	Kc cells, Chromatin fraction
62	Membrane fraction from total larval extract
64	Peptides for S2 cells cytoplasmic fraction separated by FFE
67	Larval Haemolymph
68	Cytoplasmic fractions (soluble proteins) from Adult Drosophila heads
69	Golgi fraction from Kc cells
71	S2 cells nuclear fraction
74	S2 cells membrane fraction
76	Dm_S2_nuc_FFE
77	Adult flies Membrane
78	Embryos 0-24h AEL
80	Embryos 0-24h AEL
82	Total Larvae, Protein FFE
83	Total Larvae, Peptide FFE
86	Gelfiltration of exponentially growing Kc-cell total lysate

88 S2cells cytoplasmic fraction
89 Total Larvae, Protein FFE
91 Heads from adult flies, nuclear fraction

Table S9. Proportion of essential genes arisen by different mechanisms;
A statistical summary table for Table S3.

Table S10. Phenotype of parental genes of young essential genes;
Cross scheme for phenotypic characterization of parental genes was identical to that of young genes; the phenotype of related young gene for each parental gene was shown at the last two columns in the same row of the parental gene for comparison.

Table S11. Essential-Nonessential relationship summary table for young gene - parental gene pairs; this is a summary statistic table with data generated from this study (Table S10, where 16 pairs are both essential, 9 pairs are parental gene nonessential young gene essential), pooled with data from (*7*), where the available data include 1 pair both essential, 5 pairs parental gene essential young gene nonessential, and 6 pairs both nonessential. Statistical independence between rows and columns were tested by two-tailed fisher's exact test.

Supplemental References:

S1. Y. Zhang, M. D. Vibranovski, B. H. Krinsky, M. Long, *Genome Res. gr.107334.110; Published in Advance August 26, doi: 10.1101/gr.107334.110* (2010).
S2. R. M. Kuhn *et al.*, *Nucleic Acids Res.* **35**, D668 (Jan, 2007).
S3. Y. Bai, C. Casola, C. Feschotte, E. Betran, *Genome Biol* **8**, R11 (2007).
S4. Q. Zhou *et al.*, *Genome Res.* **18**, 1446 (Sep, 2008).
S5. W.-H. Li, *Molecular evolution.* (Sinauer Associates, Sunderland, Mass., 1997), pp. xv, 487 p.
S6. F. Desiere *et al.*, *Nucleic Acids Res.* **34**, D655 (Jan 1, 2006).
S7. G. Dietzl *et al.*, *Nature* **448**, 151 (Jul 12, 2007).
S8. M. Ashburner, K. G. Golic, R. S. Hawley, *Drosophila : a laboratory handbook*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., ed. 2nd, 2005), pp. xxviii, 1409 p.
S9. N. Perrimon, A. Lanjuin, C. Arnold, E. Noll, *Genetics* **144**, 1681 (Dec, 1996).
S10. G. Goshima *et al.*, *Science* **316**, 417 (Apr 20, 2007).
S11. M. N. Arbeitman *et al.*, *Science* **297**, 2270 (Sep 27, 2002).
S12. Z. Gauhar *et al.*, Flybase, Ed. (2008).
S13. R. C. Gentleman *et al.*, *Genome Biol* **5**, R80 (2004).
S14. Z. Yang, *Mol. Biol. Evol.* **24**, 1586 (Aug, 2007).
S15. M. Suyama, D. Torrents, P. Bork, *Nucleic Acids Res.* **34**, W609 (Jul 1, 2006).
S16. D. J. Begun *et al.*, *PLoS Biol.* **5**, e310 (Nov 6, 2007).
S17. www.dpgp.org
S18. E. Birney, M. Clamp, R. Durbin, *Genome Res.* **14**, 988 (2004).
S19. K. Katoh, H. Toh, *Brief. Bioinform.* **9**, 286 (Jul, 2008).
S20. M. Suyama, D. Torrents, P. Bork, *Nucleic Acids Res.* **34**, W609 (Jul 1, 2006).
S21. J. J. Cai, *J. Hered.* **99**, 438 (Jul-Aug, 2008).
S22. N. Bierne, A. Eyre-Walker, *Mol. Biol. Evol.* **21**, 1350 (2004).
S23. H. Hermjakob *et al.*, *Nucleic Acids Res.* **32**, D452 (Jan 1, 2004).
S24. C. Stark *et al.*, *Nucleic Acids Res.* **34**, D535 (Jan 1, 2006).
S25. S. Griffiths-Jones, *Methods Mol Biol* **342**, 129 (2006).

Table S1. Phenotype of young essential genes by constitutive silencing

Gene Age(Myr)	Gene_ID	Gene_name	RNAi_line Transf_ID	Act5C- Gal4 RNAi lethality	lethality sex	lethality stage	Pupal lethal category
0~3	CG33105	p24-related-2	GD5843	lethal	both	pupal	Class III
3~6	CG11466	Cyp9f2	KK106779	lethal	both	before pupal	
3~6	CG12224		KK102603	lethal	both	pupal	Class I
3~6	CG12842		KK103233	lethal	both	pupal	Class I
6~11	CG10474	*Carna*	GD41455	lethal	both	pupal	Class I
6~11	CG10700		KK107394	lethal	both	pupal	Class I
6~11	CG11639	TfIIA-S-2	KK104539	lethal	both	pupal	Class I
6~11	CG13559		GD6643	lethal	both	before pupal	
6~11	CG15503	Chemosensory protein B 93a	KK102742	lethal	both	before pupal	
6~11	CG15527	Ribosomal protein S28a	KK102725	lethal	both	pupal	Class II
6~11	CG15636	Heterochromatin protein 6	GD13072	lethal	both	mixed	
6~11	CG16992	ethuselah-like 6	KK108048	lethal	both	pupal	Class I
6~11	CG17011	lectin-30A	KK107218	lethal	both	pupal	Class I
6~11	CG17031	ref2	KK105585	lethal	both	pupal	Class I
6~11	CG17268	Proteasome 28kD subunit 1A	KK101203	lethal	both	pupal	Class I
6~11	CG17802		KK102311	lethal	both	pupal	Class I
6~11	CG30083		KK104289	lethal	both	pupal	Class I
6~11	CG31406		KK105072	lethal	both	pupal	Class I
6~11	CG31882		KK100160	lethal	both	pupal	Class I
6~11	CG32282	drosomycin-4	KK103505	lethal	both	mixed	
6~11	CG32301		GD12239	lethal	both	pupal	Class II
6~11	CG33462		KK106216	lethal	both	pupal	Class I
6~11	CG4580		KK105535	lethal	female	pupal	Class I
6~11	CG5348		GD1698	lethal	both	before pupal	
6~11	CG6289		KK105515	lethal	both	before pupal	
6~11	CG6687		GD28425	lethal	male	pupal	Class I
6~11	CG7476	mthl7	KK102811	lethal	both	pupal	Class I
6~11	CG7594	Eig71Eh	KK107495	lethal	male	pupal	Class I
6~11	CG8137	Serine protease inhibitor 2	KK100958	lethal	both	pupal	Class I
11~25	CG8358		GD15161	lethal	both	pupal	Class I
11~25	CG1149	MstProx	KK108034	lethal	both	pupal	Class I
11~25	CG12766		KK104031	lethal	both	pupal	Class I
11~25	CG17176	ACXA	KK104283	lethal	both	pupal	Class I
11~25	CG17240	Serine protease 12	KK100037	lethal	both	pupal	Class I
11~25	CG31061	Gustatory receptor 98d	GD4398	lethal	both	pupal	Class I
11~25	CG31413		KK106151	lethal	both	pupal	Class I

11~25	CG31438	Chemosensory protein B 93b	KK104611	lethal	both	pupal	Class I
11~25	CG31962	Scavenger receptor class C, type III	KK102716	lethal	both	pupal	Class I
11~25	CG32376		KK101917	lethal	female	pupal	Class I
11~25	CG6690		KK101104	lethal	both	pupal	Class III
11~25	CG9120	Lysozyme X	KK104164	lethal	both	pupal	Class I
11~25	CG9873	Ribosomal protein L37b	GD29760	lethal	both	pupal	Class III
25~35	CG13463		KK108170	lethal	both	pupal	Class II
25~35	CG14957		KK102784	lethal	both	pupal	Class I
25~35	CG16960	Odorant receptor 33a	KK101137	lethal	both	pupal	Class I
25~35	CG17673	Accessory gland peptide 70A	KK109175	lethal	both	pupal	Class I
25~35	CG18324		KK105401	lethal	both	pupal	Class unknown
25~35	CG1840		KK108492	lethal	both	mixed	
25~35	CG30395		KK105198	lethal	both	mixed	
25~35	CG31524		KK100877	lethal	both	pupal	Class I
25~35	CG33350	Chemosensory protein B 42c	KK105947	lethal	both	pupal	Class I
25~35	CG33459		KK105903	lethal	both	before pupal	
25~35	CG3347		KK100453	lethal	both	pupal	Class III
25~35	CG3640		KK100399	lethal	both	pupal	Class III
25~35	CG6052		KK106738	lethal	both	mixed	
25~35	CG7931	janus B	KK103516	lethal	both	pupal	Class I
25~35	CG8626	Acp53C14a	KK103960	lethal	both	mixed	
25~35	CG9284		KK102235	lethal	both	pupal	Class II
25~35	CG9722		KK101679	lethal	both	pupal	Class I

Table S2. Origin and evolution of young essential genes

Origination event dating (Myr)	Gene_ID	Chr arm	IPR predicted protein domain	parental gene (b)	origination mechanism (c)	structural renovation (d)	protein sequence divergence (f)
0~3	CG33105	3R	GOLD	CG33104	D	C: 3' CDS	2.4%
3~6	CG11466	3R	Cytochrome P450	CG4486	D	N	56.0%
3~6	CG12224	3R	Aldo/keto reductase	CG3397	D	R: 1st intron and 3' UTR	9.9%
3~6	CG12842	2R	NA (a)	CG13617	D	C: 3' UTR	35.9%
6~11	CG10474	2R	Peptidase T2, asparaginase 2	CG1827	D	N	39.4%
6~11	CG10700	2L	Pyridine nucleotide-disulphide oxidoreductase	CG4199	D	C: 5' UTR, 3' UTR	40.9%
6~11	CG11639	X	Transcription factor IIA	CG5163	D	C: 5' UTR, 3' UTR	44.6%
6~11	CG13559	2R	LPS-induced tumor necrosis factor alpha factor	CG32280	R/D (g)	N	54.2%
6~11	CG15503	3R	Protein of unknown function DM11	CG31438	D	N	41.6%
6~11	CG15527	3R	Ribosomal protein S28e	CG2998	D	N	17.2%
6~11	CG15636	2L	Chromo domain	CG7041	D	C: 3' CDS, 3' UTR	41.3%
6~11	CG16992	3L	GPCR, family 2, secretin-like	CG7476	D	R: exon & introns structure	55.2%
6~11	CG17011	2L	C-type lectin-like	CG17799	D	N	61.3%
6~11	CG17031	2L	Nucleotide-binding, alpha-beta plait	CG1101	R	C: 5' UTR, 1st intron, 3'UTR	57.8%
6~11	CG17268	3R	Proteasome	CG3422	R	C: 3' UTR, 5' UTR	24.9%
6~11	CG17802	3R	Zinc finger, C2H2-type/integrase, DNA-binding	CG17806	D	C: 3' UTR	67.9%
6~11	CG30083	2R	Peptidase S1/S6	CG5302	D	C: 5' UTR	61.1%
6~11	CG31406	3R	NA	CG31374	D	C: 3' UTR	73.7%
6~11	CG31882	2L	NA	NA	A	NA	NA
6~11	CG32282	3L	Knottin	CG10812	D	N	36.2%
6~11	CG32301	3L	Adenylyl cyclase class-3/4/guanylyl cyclase	CG32305	D	N	47.9%
6~11	CG33462	2R	Serine/cysteine peptidase, trypsin-like	CG33461	D	C:3' CDS, 3UTR	50.8%
6~11	CG4580	2L	Peptidase M13, neprilysin	CG9508	D	N	55.9%
6~11	CG5348	2R	Sodium/calcium exchanger membrane region	CG12376	D	R: exon & introns structure	41.5%
6~11	CG6289	3L	Protease inhibitor I4, serpin	CG6663	D	C: 3' UTR	8.0%
6~11	CG6687	3R	ATPase, F0 complex, subunit C	CG18525	D	C: 5UTR	50.1%
6~11	CG7476	3L	methuselah-like 7	CG32853	D	R: exon & introns structure	54.4%
6~11	CG7594	3L	NA	CG7599	D	C: 5' UTR, 3' UTR	46.8%
6~11	CG8137	2L	Protease inhibitor I4, serpin	CG9334	D	C: 5' UTR	27.3%

11~25	CG8358	3R	Peptidase M13, neprilysin	CG5527	D	C: 5’ UTR, 3’UTR	47.3%
11~25	CG1149	3R	Toll-Interleukin receptor, leucine-rich repeat	CG18241	D	C: 3’UTR	57.9%
11~25	CG12766	3L	Aldo/keto reductase	CG10863	D	N	22.2%
11~25	CG17176	2L	Adenylyl cyclase class-3/4/guanylyl cyclase	CG17174	D	N	31.2%
11~25	CG17240	2L	Peptidase S1A, chymotrypsin	CG7754	D	C: 3' UTR	55.6%
11~25	CG31061	3R	7TM chemoreceptor	CG31060	D	C: 3’ UTR	38.7%
11~25	CG31413	3R	Erv1/Alr	CG17843	D	C: 3’ UTR	60.0%
11~25	CG31438	3R	NA	CG15503	D	N	41.6%
11~25	CG31962	2L	Sushi/SCR/CCP	CG8856	D	C: 3’ UTR	55.6%
11~25	CG32376	3L	Serine/cysteine peptidase, trypsin-like	CG32374	D	N	36.3%
11~25	CG6690	3R	Erv1/Alr	CG17843	D	C: 3’ UTR	59.5%
11~25	CG9120	3L	Glycoside hydrolase, family 22	CG1180	D	R: 3’ UTR	25.5%
11~25	CG9873	2R	Ribosomal protein, zinc-binding	CG9091	R	C: 5’ UTR, 3’ UTR	23.0%
25~35	CG13463	3L	NA	CG5784	D	C: 5' UTR, 5' CDS	68.8%
25~35	CG14957	3L	Chitin binding protein, peritrophin-A	CG32284	D	C: 3' UTR	49.5%
25~35	CG16960	2L	Olfactory receptor, Drosophila	CG16961	D	R: exon & introns structure	42.1%
25~35	CG17673	3L	Sex peptide	CG33495	D	R: exon & introns structure	50.9%
25~35	CG18324	2R	Mitochondrial substrate carrier;Adenine nucleotide translocator 1	CG18327	D	R: exon & introns structure	41.3%
25~35	CG1840	X	NA	CG1844	D	C: 3’ UTR	34.7%
25~35	CG30395	2R	NA	NA	A	NA	NA
25~35	CG31524	3R	Prolyl 4-hydroxylase alpha-subunit, N-terminal;Tetratricopeptide TPR2;Tetratricopeptide region;2OG-Fe(II) oxygenase;Prolyl 4-hydroxylase, alpha subunit	CG9720	D	N	41.2%
25~35	CG33350	2R	Chemosensory protein B	CG33351	D	R: exon & introns structure	43.2%
25~35	CG33459	2R	Peptidase S1A, chymotrypsin;Peptidase S1 and S6, chymotrypsin/Hap;Peptidase, trypsin-like serine and cysteine	CG30090	D	N	57.3%
25~35	CG3347	2L	Zinc finger, FYVE/PHD-type;Zinc finger, PHD-type;Zinc finger, RING/FYVE/PHD-type	CG17440	D	C: 3’CDS, 3’ UTR	77.0%
25~35	CG3640	2R	SCP-like extracellular;Allergen V5/Tpx-1 related	CG17575	D	C: 3'CDS	61.2%
25~35	CG6052	3L	ATPase, AAA+ type, core;ABC transporter-like	CG1718	R	C: 3'CDS	53.0%
25~35	CG7931	3R	Janus/Ocnus	CG7929	D	C: 3'UTR	56.6%
25~35	CG8626	2R	Drosophila ACP53EA	CG8622	D	C:3'UTR	71.9%
25~35	CG9284	2R	NA	NA	A	NA	NA
25~35	CG9722	3R	Uncharacterised protein family UPF0005	CG3798	R	C: 5'CDS, 5'UTR	43.5%

Table S3. Phenotype of constitutive silencing of all young genes tested

Gene Age (Myr)	CG	TID	Act5C-Gal4 RNAi lethality	Parental gene	Origination mechanism
25~35	CG13477	KK109920	viable	NA	A
25~35	CG14835	KK100402	semilethal	NA	A
25~35	CG15263	KK102663	viable	NA	A
25~35	CG16741	KK103236	viable	NA	A
25~35	CG2665	KK103004	viable	NA	A
25~35	CG9024	KK107536	viable	NA	A
25~35	CG9284	KK102235	lethal	NA	A
25~35	CG11598	KK103262	viable	CG18530	D
25~35	CG11833	KK102870	viable	CG3373	D
25~35	CG12780	KK101969	viable	CG13422	D
25~35	CG13091	KK100943	viable	CG10097	D
25~35	CG1349	KK103336	viable	CG6646	D
25~35	CG14957	KK102784	lethal	CG32284	D
25~35	CG16960	KK101137	lethal	CG16961	D
25~35	CG17637	KK102424	viable	CG18281	D
25~35	CG18324	KK105401	lethal	CG18327	D
25~35	CG1840	KK108492	lethal	CG1844	D
25~35	CG18530	KK108503	viable	CG11598	D
25~35	CG1946	KK108495	viable	CG1942	D
25~35	CG31508	KK106379	viable	CG31509	D
25~35	CG31524	KK100877	lethal	CG9720	D
25~35	CG33223	KK106782	viable	CG32713	D
25~35	CG33342	KK100826	semilethal	CG17784	D
25~35	CG33483	KK102589	viable	CG33923	D
25~35	CG33920	KK103447	viable	CG33654	D
25~35	CG4052	KK109160	viable	CG2341	D
25~35	CG33350	KK105947	lethal	CG33351	D
25~35	CG11261	KK105927	viable	CG1877	D
25~35	CG13463	KK108170	lethal	CG5784	D
25~35	CG13738	KK104508	viable	CG6912	D
25~35	CG15461	KK103441	viable	CG34200	D
25~35	CG15616	KK105778	viable	CG8626	D
25~35	CG16710	KK100054	viable/semilethal	CG18754	D
25~35	CG17673	KK109175	lethal	CG33495	D
25~35	CG18557	KK107357	viable	CG6639	D
25~35	CG18748	KK109017	viable	CG18746	D
25~35	CG30395	KK105198	lethal	CG15040	D
25~35	CG31313	KK102741	viable	CG8066	D
25~35	CG31335	KK108029	viable	CG31336	D
25~35	CG31493	KK101611	viable	CG31248	D
25~35	CG32115	KK108451	viable	CG32107	D
25~35	CG32186	KK106262	semilethal/viable	CG1718	D
25~35	CG32228	KK104148	viable	CG3996	D
25~35	CG32243	KK101716	viable	CG8588	D
25~35	CG33109	KK107240	viable	CG16826	D
25~35	CG33458	KK105451	viable	CG30090	D
25~35	CG33459	KK105903	lethal	CG30090	D
25~35	CG3347	KK100453	lethal	CG17440	D
25~35	CG34041	KK103005	semilethal	CG31013	D
25~35	CG34235	KK109077	semilethal	CG30058	D

25~35	CG3640	KK100399	lethal	CG17575	D
25~35	CG4907	KK108893	viable	CG13978	D
25~35	CG7676	KK101813	viable	CG6392	D
25~35	CG7931	KK103516	lethal	CG7929	D
25~35	CG8626	KK103960	lethal	CG8622	D
25~35	CG8664	KK101088	viable	CG5107	D
25~35	CG15358	GD13372	viable	CG15818	D
25~35	CG11597	KK104729	viable/semilethal	CG32505	R
25~35	CG6036	KK105568	viable	CG1906	R
25~35	CG9722	KK101679	lethal	CG3798	R
25~35	CG15296	KK104221	viable	CG7603	R
25~35	CG30187	KK106612	viable	CG30083	R
25~35	CG6052	KK106738	lethal	CG1718	R
11~25	CG1338	GD32187	viable	NA	A
11~25	CG17650	KK105047	viable	NA	A
11~25	CG5016	GD29219	viable	NA	A
11~25	CG10246	KK100143	viable	CG10247	D
11~25	CG12766	KK104031	lethal	CG10863	D
11~25	CG16931	GD23534	viable	CG7350	D
11~25	CG17174	GD9748	viable	CG5983	D
11~25	CG17176	KK104283	lethal	CG17174	D
11~25	CG17946	GD23560	viable	CG17945	D
11~25	CG18125	GD47675	viable	CG17234	D
11~25	CG1878	KK103148	viable	CG1367	D
11~25	CG30037	GD12079	viable	CG33145	D
11~25	CG31061	GD4398	lethal	CG31060	D
11~25	CG31438	KK104611	lethal	CG15503	D
11~25	CG5103	GD46603	viable	CG8036	D
11~25	CG6367	GD18541	viable	CG6361	D
11~25	CG8358	GD15161	lethal	CG5527	D
11~25	CG9120	KK104164	lethal	CG1180	D
11~25	CG9720	KK100523	viable	CG31524	D
11~25	CG1149	KK108034	lethal	CG18241	D
11~25	CG32376	KK101917	lethal	CG32374	D
11~25	CG10799	GD16828	viable	CG4716	D
11~25	CG17012	GD15570	viable	CG30031	D
11~25	CG17240	KK100037	lethal	CG7754	D
11~25	CG18096	GD30873	viable	CG7052	D
11~25	CG18228	KK101647	viable	CG9633	D
11~25	CG30450	KK101927	viable	CG8462	D
11~25	CG30494	KK104153	viable	CG33715	D
11~25	CG31208	KK102611	viable	CG31173	D
11~25	CG31413	KK106151	lethal	CG17843	D
11~25	CG31782	GD45455	viable	CG8474	D
11~25	CG31918	GD13220	viable	CG9508	D
11~25	CG31962	KK102716	lethal	CG8856	D
11~25	CG33490	GD22222	viable	CG33489	D
11~25	CG4259	GD44873	viable	CG3117	D
11~25	CG4477	GD39486	viable	CG18223	D
11~25	CG4691	GD21920	viable	CG30362	D
11~25	CG4712	GD21925	viable	CG4714	D
11~25	CG6690	KK101104	lethal	CG17843	D
11~25	CG7800	GD973	viable	CG18249	D
11~25	CG30036	KK101307	viable	CG33145	D
11~25	CG31875	KK104638	viable	CG33525	D
11~25	CG9873	GD29760	lethal	CG9091	R
6~11	CG10852	KK106565	viable	NA	A
6~11	CG1361	GD9706	viable	NA	A

6~11	CG15882	KK104267	viable	NA	A
6~11	CG31882	KK100160	lethal	NA	A
6~11	CG32368	KK103313	viable	NA	A
6~11	CG10090	GD37347	viable	CG14666	D
6~11	CG10232	KK100033	viable	CG31219	D
6~11	CG10474	GD41455	lethal	CG1827	D
6~11	CG10700	KK107394	lethal	CG4199	D
6~11	CG11639	KK104539	lethal	CG5163	D
6~11	CG13656	KK101054	viable	CG11889	D
6~11	CG13977	KK100415	viable	CG10248	D
6~11	CG15527	KK102725	lethal	CG2998	D
6~11	CG17216	GD51996	viable	CG6715	D
6~11	CG17795	GD26815	viable	CG6936	D
6~11	CG18814	KK106349	viable	CG4842	D
6~11	CG31002	GD48226	viable	CG6653	D
6~11	CG31791	GD19241	viable	CG31801	D
6~11	CG32282	KK103505	lethal	CG10812	D
6~11	CG32283	KK105013	viable	CG32282	D
6~11	CG32301	GD12239	lethal	CG32305	D
6~11	CG32580	KK105591	viable	CG41038	D
6~11	CG5348	GD1698	lethal	CG12376	D
6~11	CG5372	KK100120	viable	CG8095	D
6~11	CG5609	GD51123	viable	CG31508	D
6~11	CG6289	KK105515	lethal	CG6663	D
6~11	CG6663	GD49626	semilethal	CG6289	D
6~11	CG6730	GD8955	viable	CG16761	D
6~11	CG7594	KK107495	lethal	CG7599	D
6~11	CG8137	KK100958	lethal	CG9334	D
6~11	CG8652	GD46514	viable	CG13270	D
6~11	CG8856	KK100928	viable	CG4099	D
6~11	CG15503	KK102742	lethal	CG15503	D
6~11	CG33462	KK106216	lethal	CG33461	D
6~11	CG7476	KK102811	lethal	CG32853	D
6~11	CG13686	KK106450	viable	CG2839	D
6~11	CG14027	GD6958	viable	CG33117	D
6~11	CG14610	KK104298	viable	CG12883	D
6~11	CG15636	GD13072	lethal	CG7041	D
6~11	CG16992	KK108048	lethal	CG7476	D
6~11	CG17011	KK107218	lethal	CG17799	D
6~11	CG17802	KK102311	lethal	CG17806	D
6~11	CG2826	KK102844	viable	CG13686	D
6~11	CG30083	KK104289	lethal	CG5302	D
6~11	CG30448	GD12997	viable	CG11218	D
6~11	CG30473	GD27143	viable	CG8462	D
6~11	CG31406	KK105072	lethal	CG31374	D
6~11	CG31769	KK106293	viable	CG15293	D
6~11	CG31932	GD42744	viable	CG31929	D
6~11	CG31941	GD31141	viable	CG8462	D
6~11	CG3217	KK102373	viable	CG13896	D
6~11	CG32853	KK105546	viable	CG7476	D
6~11	CG33235	GD48441	viable	CG34387	D
6~11	CG4580	KK105535	lethal	CG9508	D
6~11	CG5509	GD22143	viable	CG14719	D
6~11	CG6687	GD28425	lethal	CG18525	D
6~11	CG7157	KK105453	viable	CG32133	D
6~11	CG11012	GD6398	viable	CG5724	D
6~11	CG14405	KK102486	viable	CG33320	D
6~11	CG14584	GD8658	viable	CG17382	D

6~11	CG31370	GD21324	viable	CG13659	D
6~11	CG31509	KK106548	viable	CG31508	D
6~11	CG7106	KK104201	semilethal	CG15818	D
6~11	CG11825	GD49834	viable	CG17734	R
6~11	CG17268	KK101203	lethal	CG3422	R
6~11	CG1736	GD32889	semilethal	CG9327	R
6~11	CG5265	KK104754	viable	CG1041	R
6~11	CG6208	KK105575	viable	CG3988	R
6~11	CG7815	GD22567	viable	CG1404	R
6~11	CG13559	GD6643	lethal	CG32280	R
6~11	CG14088	GD5361	viable	CG33460	R
6~11	CG17031	KK105585	lethal	CG1101	R
6~11	CG31495	KK105777	viable	CG33101	R
6~11	CG31883	GD36488	viable	CG7875	R
6~11	CG32110	KK107634	viable	CG12359	R
6~11	CG6639	KK104307	viable	CG18563	R
3~6	CG32021	KK102286	viable	NA	A
3~6	CG12224	KK102603	lethal	CG3397	D
3~6	CG12493	KK102360	viable	CG10630	D
3~6	CG2885	KK103311	viable	CG9807	D
3~6	CG12842	KK103233	lethal	CG13617	D
3~6	CG11466	KK106779	lethal	CG4486	D
3~6	CG17797	KK100995	viable	CG17799	D
3~6	CG3410	GD37836	viable	CG2958	D
3~6	CG10174	GD31195	viable	CG1740	R
3~6	CG1924	GD52347	viable	CG11958	R
3~6	CG9573	GD49820	viable	CG14213	R
0~3	CG32588	KK102410	semilethal	CG33252	D
0~3	CG33105	GD5843	lethal	CG33104	D

Table S4. Representative F1 adult counting in lethal RNAi crosses

		Act5C-Gal4/CyO cross	
Gene_ID	RNAi_line Transf_ID	NonCyO_F1	CyO_F1
CG10474	GD41455	0	>100#
CG10700	KK107394	0	88
CG11466	KK106779	0	>100
CG1149	KK108034	0	81
CG11639	KK104539	0	>100
CG12224	KK102603	0	>100
CG12766	KK104031	0	>100
CG12842	KK103233	0	107
CG13463	KK108170	0	>100
CG13559	GD6643	0	>100
CG14957	KK102784	0	>100
CG15503	KK102742	0	>100
CG15527	KK102725	0	>100
CG15636	GD13072	0	66
CG16960	KK101137	0	>100
CG16992	KK108048	0	67
CG17011	KK107218	0	57
CG17031	KK105585	0	74
CG17176	KK104283	0	>100
CG17240	KK100037	0	>100
CG17268	KK101203	0	>100
CG17673	KK109175	0	>100
CG17802	KK102311	0	>100
CG18324	KK105401	0	>100
CG1840	KK108492	0	>100
CG30083	KK104289	0	>100
CG30395	KK105198	1*	>100
CG31061	GD4398	0	83
CG31406	KK105072	0	>100
CG31413	KK106151	0	115
CG31438	KK104611	0	>100
CG31524	KK100877	0	>100
CG31882	KK100160	0	>100
CG31962	KK102716	0	86
CG32282	KK103505	0	86
CG32301	GD12239	0	>100
CG32376	KK101917	1*	>100
CG33105	GD5843	5*	114
CG33350	KK105947	0	>100
CG33459	KK105903	0	>100
CG33462	KK106216	0	83
CG3347	KK100453	0	>100
CG3640	KK100399	0	>100
CG4580	KK105535	1*	>100
CG5348	GD1698	0	>100
CG6052	KK106738	0	>100
CG6289	KK105515	0	119
CG6687	GD28425	1*	>100
CG6690	KK101104	0	70
CG7476	KK102811	0	>100
CG7594	KK107495	2*	115
CG7931	KK103516	1*	>100
CG8137	KK100958	0	96
CG8358	GD15161	0	66

CG8626	KK103960	0	>100
CG9120	KK104164	0	>100
CG9284	KK102235	0	>100
CG9722	KK101679	0	>100
CG9873	GD29760	0	>100

Table S5. Lethality stage examination with fluorescence tracking

			RNAi animal survival developmental stages with GFP tracking NOTES							
GENE_ID	TID	Lethal stage_call	E*	L1*	L2*	L3*	PP*	EP*	PH*	A*
CG10474	GD41455	PH/A	+	+	+	+	+	+	+	-
CG10700	KK107394	PH/A	na	na	na	na	na	na	na	na
CG11466	KK106779	L2/L3	+	+	+	-	-	-	-	-
CG1149	KK108034	PH/A	+	+	+	+	+	+	+	-
CG11639	KK104539	EP/PH	+	+	+	+	+	+	-	-
CG12766	KK104031	PH/A	na	na	na	na	na	na	na	na
CG12842	KK103233	PH/A	+	+	+	+	+	+	+	-
CG13463	KK108170	EP/PH	+	+	+	+	+	+	-	-
CG13559	GD6643	PH/A	+	+	+	+	+	+	+	-
CG14957	KK102784	EP/PH	+	+	+	+	+	+	-	-
CG15503	KK102742	EP/PH	+	+	+	+	+	+	-	-
CG15527	KK102725	MIX	na	na	na	na	na	na	na	na
CG15636	GD13072	EP/PH	+	+	+	+	+	+	-	-
CG16960	KK101137	EP/PH	+	+	+	+	+	+	-	-
CG16992	KK108048	PH/A	+	+	+	+	+	+	+	-
CG17011	KK107218	EP/PH	+	+	+	+	+	+	-	-
CG17031	KK105585	EP/PH	+	+	+	+	+	+	-	-
CG17176	KK104283	EP/PH	+	+	+	+	+	+	-	-
CG17240	KK100037	EP/PH	+	+	+	+	+	+	-	-
CG17268	KK101203	EP/PH	+	+	+	+	+	+	-	-
CG17673	KK109175	EP/PH	+	+	+	+	+	+	-	-
CG18324	KK105401	P	na	na	na	na	na	na	na	na
CG1840	KK108492	MIX	+	+	+	+	+	+	-	-
CG30083	KK104289	EP/PH	+	+	+	+	-	+	-	-
CG30395	KK105198	EP/PH	+	+	+	+	+	+	-	-
CG31061	GD4398	EP/PH	+	+	+	+	+	+	-	-
CG31406	KK105072	PH/A	+	+	+	+	+	+	+	-
CG31413	KK106151	PH/A	+	+	+	+	+	+	+	-
CG31438	KK104611	EP/PH	+	+	+	+	+	+	-	-
CG31524	KK100877	PH/A	na	na	na	na	na	na	na	na
CG31882	KK100160	EP/PH	+	+	+	+	+	+	-	-
CG31962	KK102716	PH/A	+	+	+	+	+	+	+	-
CG32282	KK103505	MIX	na	na	na	na	na	na	na	na
CG32301	GD12239	PP/EP	na	na	na	na	na	na	na	na
CG32376	KK101917	PH/A	+	+	+	+	+	+	+	+
CG33105	GD5843	EP/PH	+	+	+	+	+	+	-	-
CG33459	KK105903	L3/PP	+	+	+	+	-	-	-	-
CG33462	KK106216	EP/PH	+	+	+	+	+	+	-	-
CG3347	KK100453	EP/PH	+	+	+	+	+	+	-	-
CG3640	KK100399	EP/PH	+	+	+	+	+	+	-	-
CG4580	KK105535	EP/PH	+	+	+	+	+	+	-	-
CG5348	GD1698	PH/A	+	+	+	+	+	+	+	-
CG6052	KK106738	MIX	+	+	+	+	+	+	-	-
CG6289	KK105515	L2/L3	+	+	+	-	-	-	-	-
CG6690	KK101104	PH/A	+	+	+	+	+	+	+	-
CG7594	KK107495	PH/A	+	+	+	+	+	+	+	-
CG7931	KK103516	EP/PH	+	+	+	+	+	+	-	-
CG8137	KK100958	EP/PH	+	+	+	+	+	+	-	-
CG8358	GD15161	L3/PP	+	+	+	+	-	-	-	-
CG8626	KK103960	MIX	+	+	+	+	+	+	-	-
CG9284	KK102235	EP/PH	+	+	+	+	+	+	-	-
CG9722	KK101679	EP/PH	+	+	+	+	+	+	-	-

Table S6. Viability phenotype consistency of representative young genes

Part (I) Viability phenotype consistency between independent Gal4 driver constructs

Gene_ID	RNAi_line Transf_ID	TubP- Gal4 RNAi lethality	Act5C- Gal4 RNAi lethality
CG10474	GD41455	lethal	lethal
CG10700	KK107394	lethal	lethal
CG11466	KK106779	lethal	lethal
CG1149	KK108034	lethal	lethal
CG11639	KK104539	lethal	lethal
CG12766	KK104031	lethal	lethal
CG12842	KK103233	lethal	lethal
CG13559	GD6643	semilethal	lethal
CG15503	KK102742	lethal	lethal
CG15527	KK102725	lethal	lethal
CG15636	GD13072	lethal	lethal
CG16992	KK108048	lethal	lethal
CG17011	KK107218	lethal	lethal
CG17031	KK105585	lethal	lethal
CG17176	KK104283	lethal	lethal
CG17240	KK100037	viable	lethal
CG17268	KK101203	lethal	lethal
CG30083	KK104289	lethal	lethal
CG31061	GD4398	lethal	lethal
CG31406	KK105072	lethal	lethal
CG31413	KK106151	lethal	lethal
CG31438	KK104611	lethal	lethal
CG31882	KK100160	lethal	lethal
CG31962	KK102716	lethal	lethal
CG32282	KK103505	lethal	lethal
CG32301	GD12239	lethal	lethal
CG32376	KK101917	lethal	lethal
CG33105	GD5843	lethal	lethal
CG33462	KK106216	lethal	lethal
CG4580	KK105535	lethal	lethal
CG5348	GD1698	lethal	lethal
CG6289	KK105515	lethal	lethal
CG6690	KK101104	lethal	lethal
CG7594	KK107495	lethal	lethal
CG8358	GD15161	lethal	lethal
CG9120	KK104164	lethal	lethal
CG9873	GD29760	lethal	lethal

Part (II) Viability phenotype consistency between independent UAS-IR lines

Gene_ID	RNAi line Transf_ID	Act5C-Gal4 lethality
CG10090	GD37347	viable
CG10090	KK101139	viable
CG10852	KK106565	viable
CG10852	GD26898	viable
CG10852	GD26899	viable
CG1338	GD32187	viable
CG1338	KK101107	viable
CG13656	KK101054	viable
CG13656	GD23898	viable
CG13686	KK106450	viable
CG13686	GD32507	viable
CG13686	GD31867	viable
CG14027	GD6958	viable
CG14027	GD6957	viable
CG14027	KK106727	viable
CG14719	GD32316	viable
CG14719	KK105565	viable
CG15734	KK103519	viable
CG15734	GD39247	viable
CG15882	KK104267	viable
CG15882	GD38901	viable
CG15882	GD38902	viable
CG16931	GD23534	viable
CG16931	KK106010	viable
CG17011	GD15597	lethal
CG17011	KK107218	lethal
CG17011	GD15596	lethal
CG17216	GD51996	viable

CG17216	KK105265	viable
CG17216	GD51995	viable
CG17650	KK105047	viable
CG17650	GD35628	viable
CG18125	GD47675	viable
CG18125	GD15762	viable
CG18125	GD15763	viable
CG30037	GD12079	viable
CG30037	KK101660	viable
CG30037	GD13474	viable
CG30448	GD12997	viable
CG30448	GD12998	viable
CG30473	GD27143	viable
CG30473	GD27144	viable
CG30494	KK104153	viable
CG30494	GD31006	viable
CG31002	GD48226	viable
CG31002	GD48225	viable
CG31061	GD4398	lethal
CG31061	GD4399	lethal
CG31406	KK105072	lethal
CG31406	GD39272	lethal
CG31406	GD39273	lethal
CG31501	KK101011	lethal
CG31501	GD25773	lethal
CG31509	KK106548	viable
CG31509	GD14415	viable
CG31782	GD45455	viable
CG31782	KK100145	viable
CG31782	GD45454	viable
CG31782	GD21422	viable
CG31875	KK104638	viable

CG31875	GD45070	viable
CG31875	GD40298	viable
CG31918	GD13220	viable
CG31918	GD13218	viable
CG31918	GD39684	viable
CG31932	GD42744	viable
CG31932	KK102860	viable
CG31932	GD42745	viable
CG31941	GD31141	viable
CG31941	KK106893	viable
CG31941	GD31142	viable
CG32110	KK107634	viable
CG32110	GD34062	viable
CG32853	KK105546	viable
CG32853	GD48402	viable
CG33235	GD48441	viable
CG33235	GD19355	viable
CG33490	GD22222	viable
CG33490	KK101678	viable
CG33533	KK107567	viable
CG33533	GD51100	viable
CG3410	GD37836	viable
CG3410	KK107254	viable
CG4259	GD44873	viable
CG4259	GD45271	viable
CG4259	GD52657	viable
CG4259	GD44874	viable
CG4477	GD39486	viable
CG4477	GD39487	viable
CG4477	KK106868	semilethal
CG4712	GD21925	viable
CG4712	GD21924	viable

CG5103	GD46603	viable
CG5103	GD41647	viable
CG5103	GD46604	viable
CG5213	GD43449	viable
CG5213	GD43450	semilethal
CG5509	GD22143	viable
CG5509	GD22142	viable
CG5609	GD51123	viable
CG5609	GD51124	viable
CG5662	GD35030	viable
CG5662	KK102645	viable
CG6367	GD18541	viable
CG6367	GD18543	viable
CG6541	GD27655	viable
CG6541	KK101870	viable
CG6541	GD27656	viable
CG6687	GD28425	lethal
CG6687	GD28423	lethal
CG7800	GD973	viable
CG7800	GD6673	viable
CG7800	GD6672	viable
CG7800	GD27121	viable
CG8953	GD15171	viable
CG8953	KK106162	viable

Table S7. Tissue-specific RNAi phenotype of representative young essential genes

Gene_ID	RNAi_line Transf_ID	Bx(ms)1096-Gal4 phenotype #	Pnr-Gal4 phenotype #
CG10474	GD41455	wings fold laterally, occassional notching	notum growth defect
CG11466	KK106779	wing vestitial-like, bristle overproliferation	semilethal, scutellum growth defect, bristles differentiation defect
CG1149	KK108034	wing posture defect	normal
CG15527	KK102725	wing malformed, deformed or blisterd, tumor spots, bristle cell differentiation defects	semilethal, notum necrosis, bristle growth defect
CG16992	KK108048	wing malformation	bristle pattern defect
CG17011	KK107218	wing necrosis	notum necrosis,semilethal,scutella bristle defect
CG17176	KK104283	wing bent	nd *
CG17240	KK100037	Normal	tumor spots in post-alars-scutella
CG17268	KK101203	nd	normal
CG30083	KK104289	various wing defects	normal
CG31061	GD4398	wing misfolded or curled	bristle growth defect
CG31406	KK105072	wing size and shape defect hinge growth abnormal, bristles overproliferation	severe bristle overproliferation, scutellum growth defect
CG31438	KK104611	wings bent	nd
CG31882	KK100160	wings bent or notched	nd
CG31962	KK102716	wing growth defect, tumor spots at wing tip	tumor spots in post-alars-scutella region
CG32301	GD12239	wing malformation: twisted or blistered, vein disrupted, occasional notching	semilethal, necrosis
CG32376	KK101917	wing size reduced	nd
CG33105	GD5843	nd	normal
CG33462	KK106216	nd	normal
CG4580	KK105535	nd	abnormal dark spots
CG5348	GD1698	wing bent	normal
CG6289	KK105515	wing strongly deteriated	nd
CG6690	KK101104	nd	occasional tumor spots in post-alar-scutella
CG7594	KK107495	wing occasional shrink	nd
CG8137	KK100958	various wing defects	nd
CG9873	GD29760	wing shape defect, random notching (30%)	normal

Table S8. Peptide evidence for young essential genes

Gene_ID	Peptide_evidence	Sample_ID
CG10474	TESVECSANPEKSPSEDDR	55
CG10700	ILGNSRGEVVAVKLLDNSR;STQVVCLGSSFMAVEATANLVSR	52;56
CG11466	FSDENKESIQPFTYFPFGLGQR;LYEEVQQVDQDLEGK;LYEEVQQVDQDLEGKELTY EAIMGMK;SCIPLELITSGFQLSPK;YFENPMK;YLDQVVNEVLR;AMFDIVCDLYTK;AMF DIVCDLYTKGGSK;DITFDVDGQKVEVK;ELTYEAIMGMK;FFGIFEQR;FTNDVIASTAFG LQVNSFK;KSCIPLELITSGFQLSPK;STLSPAFTGSK;VQENELDMKDYCTR;YLDQVVN EVLRK;DRENTFYQMGK;DFDHFINHR	46;50;51;52;54;55;60;61;62; 69;71;74;77;86;88;91
CG12224	VDILQVHDVDAAPNLDIVLNETIPVLEEYVQAGK;NAGPHASHPGSQEILAVAK	76;86
CG12766	LDAEDHAILDSYHNGER	54
CG13559	ATLGIVPQQDQR	51
CG15636	IVFTDEGDEEDLESDNGYETTPSPR;NGFDLGLEPLR	49
CG16960	RYNRSLIILMQLTLVPVNIK;SLIILMQLTLVPVNIK	60;78
CG17011	INPALFQR;NGLFASSLTGR;DEQEFNEIFSR	77;91
CG17176	LESSIKEKNHDEVAFIMATFALDLMR	83
CG17268	MLDSVVISEIVK	64
CG17673	TLALFLVLVCVLGLVQAWEWPWNRKPTK	80
CG17802	GNFNLHVLR	51
CG18324	AQQHAQAVQSIAVGFQHK	77
CG1840	GNGWGSSSWGGGGGGGGGGGGGGGGGGGGGSGYR;RGNGWGSSSWGGG GGGGGGGGGGGGGGGGGGGSGYR;GNGWGSSSWGGGGGGGGGGGGGGGG GGGGGGGSGYRGGGLRPNR	52;55;56;62;77;83;89
CG30083	LSSFIDWILMVVDNYTVR	61
CG31406	TPEAATNDNANPNMKPSLK	82
CG31413	TEEVLHYLKGIYTLK	86
CG31882	THPNELGNQYPPR	49
CG32282	VSGHCSASLKCWCEQC	80
CG32301	IDLRIGVHSGEILAGIIGLTK;SLGNMPGYEWK;TRTIKIIMANILPTHVAEVFK	50;62;82
CG3347	TPKNSKTPK;VTDDAQKPAVEK	50;78
CG3640	QLIQESSTHMGCGVLR;YPNLCALQEEYDVNAVDR	68;77
CG4580	TLELLLKTMINPIDTLNNR	77
CG5348	MAFASAIGGSFFTIVVTGNTVAYVRSLVGNR	89
CG6289	VEQFYNAGGSPAGK	46;60
CG6687	IKVDEVGSTAAAATILLVSRSSR;QSPMELSSANR;VDEVGSTAAAATILLVSR	51;55;67
CG6690	GVKPIFDPIEPGNK;THDEEILWLWEAHNEVNEKLAGDASGDPR;KILAEPAMGKLVQFL NSYCGNCR	49;51;57
CG8137	DLVENSNVHVK;VATLSLANR;VLVAMGIQDAFEK	77;91
CG8358	ALEDADWLDAESR;LPELGFDHIIVHEENISR;SIGGFPAVDGDAWK;TTLDENIADNSGL R	46;47;56;60

Table S9. Proportion of essential genes arisen by different mechanisms

	Essential	Non-essential	Total	percentage
DNA-based duplicated genes	50	106	156	32.05%
RNA-based retroposed genes	6	17	23	26.09%
de novo genes	3	13	16	18.75%
Subtotal	59	136	195	30.26%

Fisher's exact test, two tailed

p = 0.64	DNA-based vs RNA-based
p = 0.4	DNA-based vs de novo
p = 0.71	RNA-based vs de novo

Table S10. Phenotype of parental genes of young essential genes

		Parental gene phenotype *						
Parental gene_ID	Transf_ID	RNAi lethality	lethality sex	lethality stage	RNAi F1_count	Control F1_count	related oung essential gene ID	Young essential gene phenotype
CG10863	KK103433	lethal	both	pharate	0	53	CG12766	lethal
CG1101	KK104471	lethal	both	nd	0	56	CG17031	lethal
CG13096	KK108860	lethal	both	before pupal	0	98	CG9641	lethal
CG15503	KK102742	lethal	both	pharate	0	32	CG31438	lethal
CG1664	KK103715	lethal	both	before pupal	0	24	CG31501	lethal
CG17799	KK101710	lethal	both	pharate	0	52	CG17011	lethal
CG17806	KK101592	lethal	both	prepupal	0	81	CG17802	lethal
CG17843	KK102838	viable	na	na	62	63	CG31413	lethal
CG17843	KK102838	viable	na	na	62	63	CG6690	lethal
CG18241	KK102642	viable	na	na	47	53	CG1149	lethal
CG31438	KK104611	lethal	both	pupal	0	59	CG15503	lethal
CG32280	KK108754	viable	na	na	66	54	CG13559	lethal
CG32305	KK101103	lethal	both	before pupal	0	59	CG32301	lethal
CG33104	KK101388	viable	na	na	29	51	CG33105	lethal
CG33461	KK105904	viable	na	na	33	29	CG33462	lethal
CG3422	KK105712	viable	na	na	10	34	CG17268	lethal
CG4199	KK106170	lethal	male	nd	26	62	CG10700	lethal
CG4486	KK102496	viable	na	na	59	53	CG11466	lethal
CG5163	KK108960	lethal	male or both	mix	2	73	CG11639	lethal
CG5302	KK107439	semi-lethal	male	before pupal	10	62	CG30083	lethal
CG7476	KK102811	lethal	both	pharate	0	45	CG16992	lethal
CG7599	KK102775	lethal	both	pupal	1	37	CG7594	lethal
CG7971	KK101384	lethal	both	before pupal	0	92	CG17240	lethal
CG8856	KK100928	viable	na	na	71	76	CG31962	lethal
CG9091	KK100772	lethal	both	before pupal	0	63	CG9873	lethal

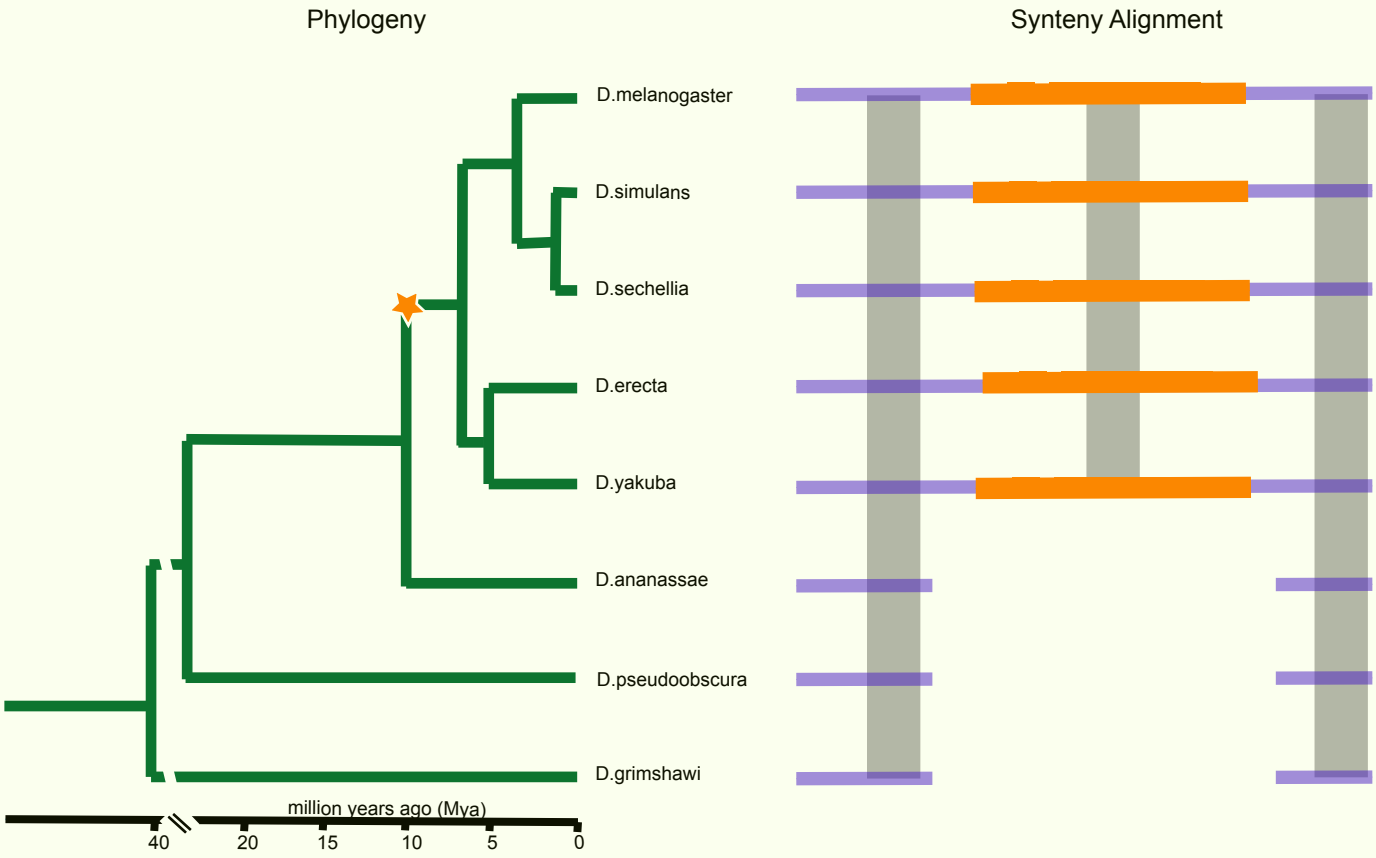
Table S11. Essential-Nonessential relationship for young gene - parental gene pairs

	Young gene essential	Young gene nonessential
parental gene essential	17	5
parental gene nonessential	9	6

Fisher's exact test, two tailed
p = 0.296

Figure S1

A



B

C

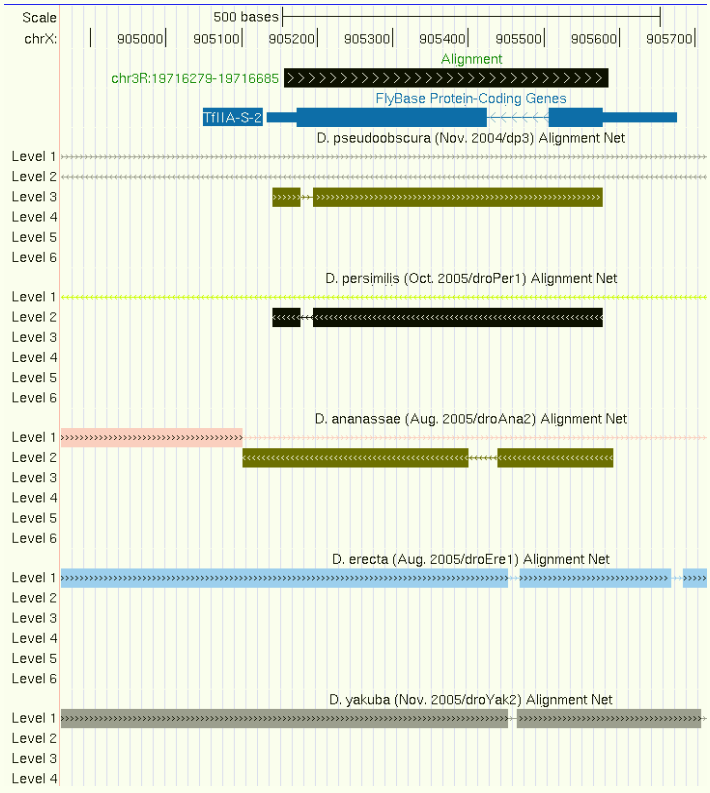


Figure S2

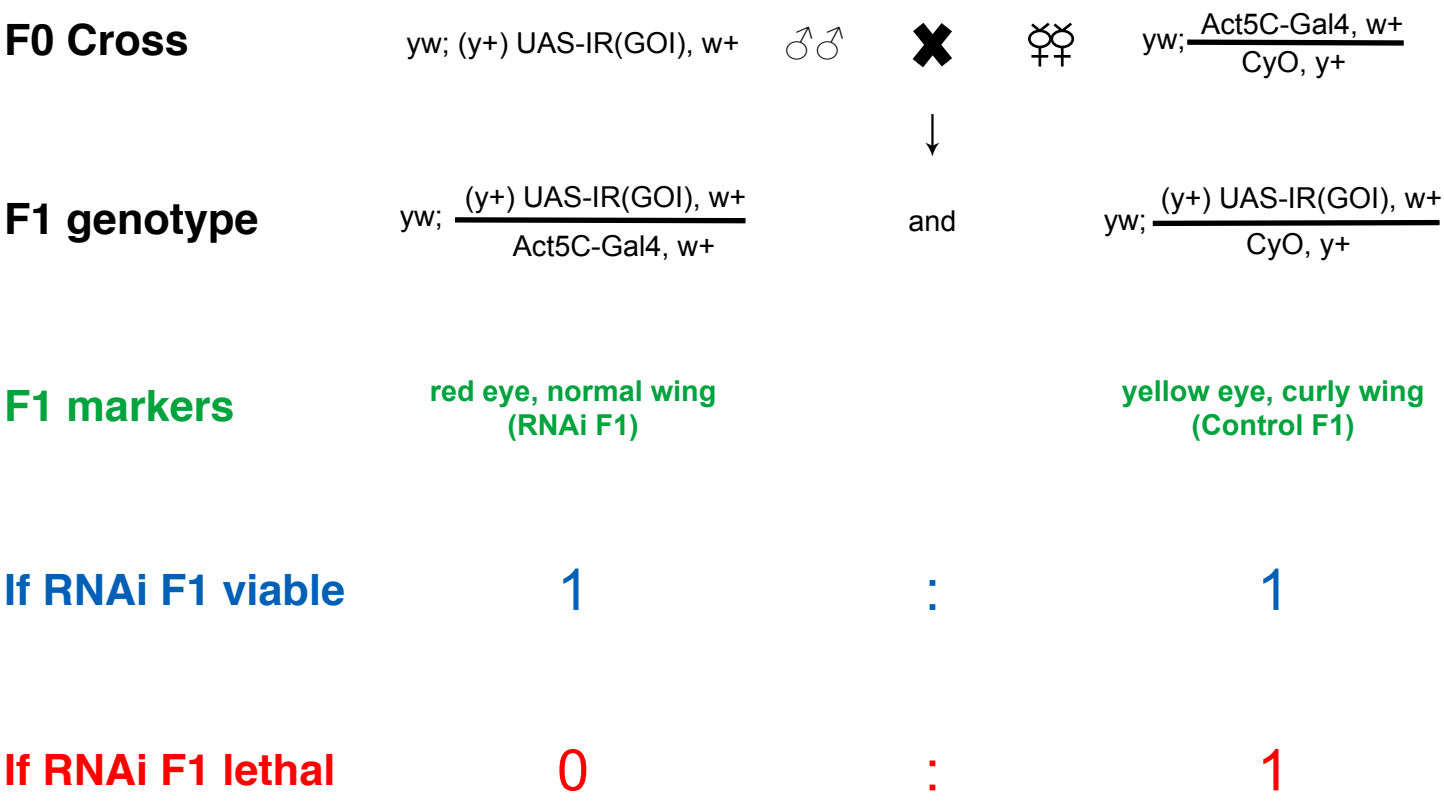
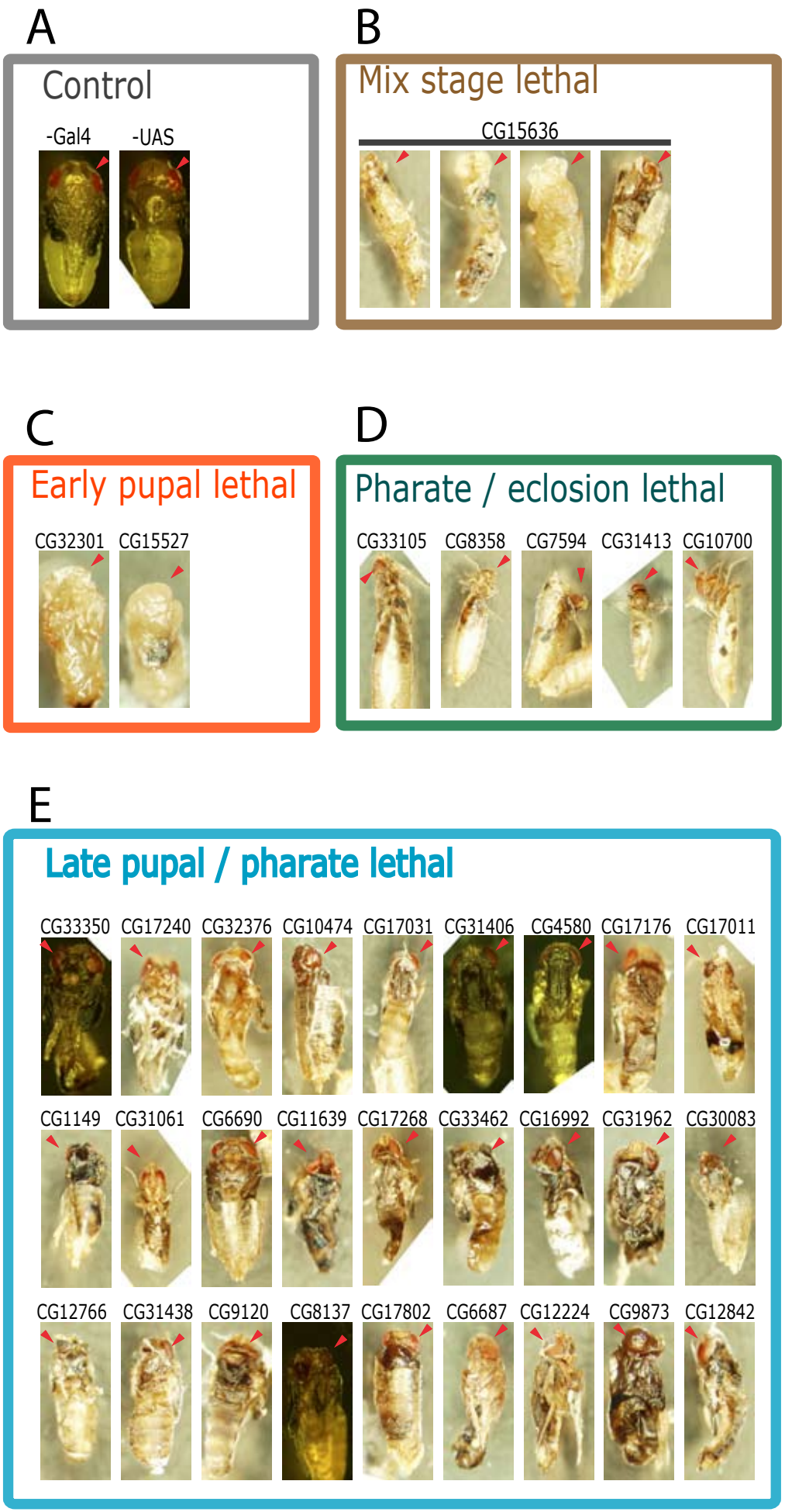
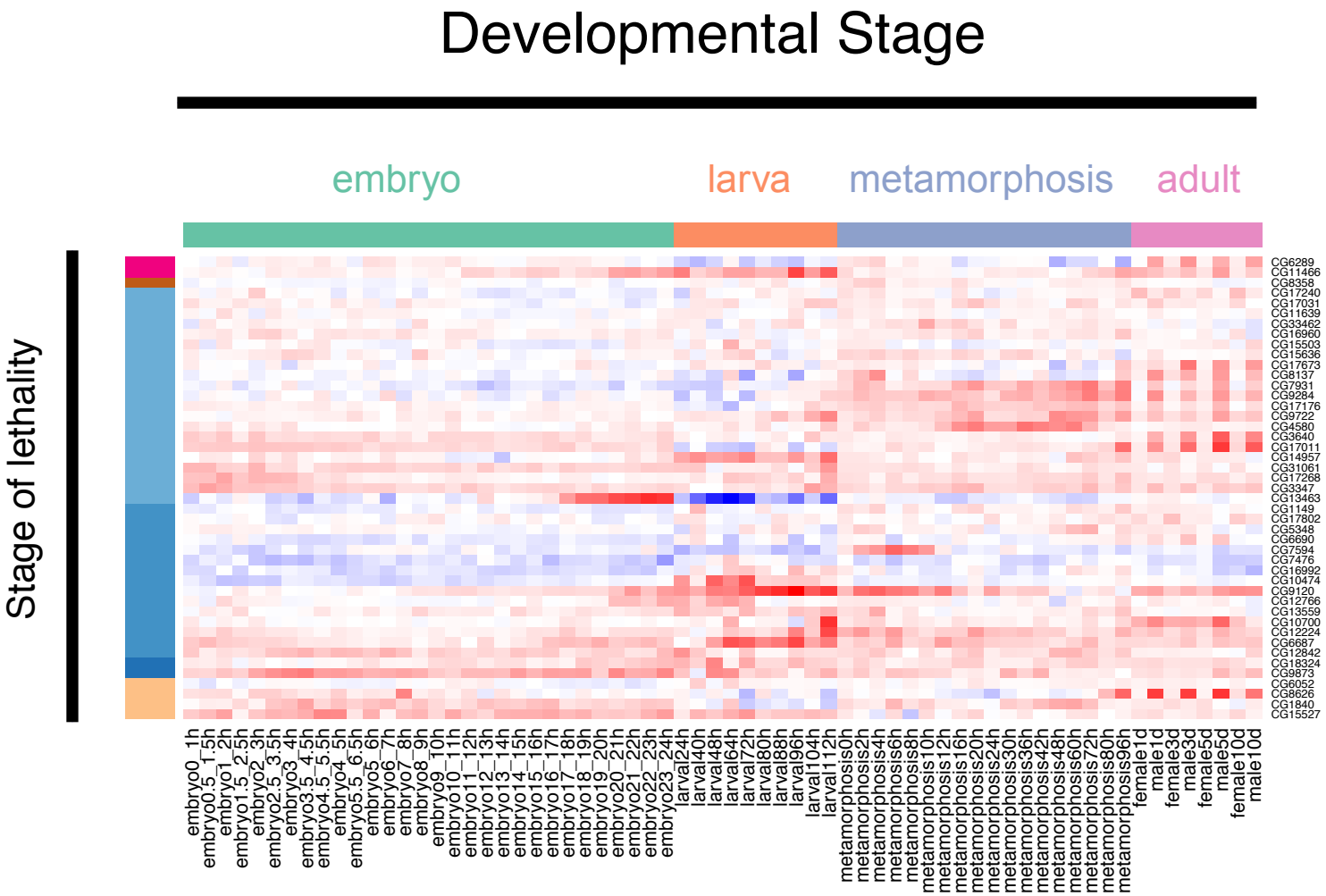

F0 Cross
yw; (y+) UAS-IR(GOI), w+
♂♂
✖
♀♀
yw; Act5C-Gal4, w+ / CyO, y+
F1 genotype
yw; (y+) UAS-IR(GOI), w+ / Act5C-Gal4, w+
and
yw; (y+) UAS-IR(GOI), w+ / CyO, y+
F1 markers
red eye, normal wing (RNAi F1)
yellow eye, curly wing (Control F1)
If RNAi F1 viable
1 : 1
If RNAi F1 lethal
0 : 1

Figure S3



Animal genotypes: Act5C-Gal4>>Gene-X-RNAi

Figure S4



Row color coding scheme
for stage of lethality

- RNAi lethal at the stage of L2/L3
- RNAi lethal at the stage of L3/PP
- RNAi lethal at the stage of PP/EP
- RNAi lethal at the stage of EP/PH
- RNAi lethal at the stage of PH/A
- RNAi lethal at pupal stage
- RNAi lethal at mix stages

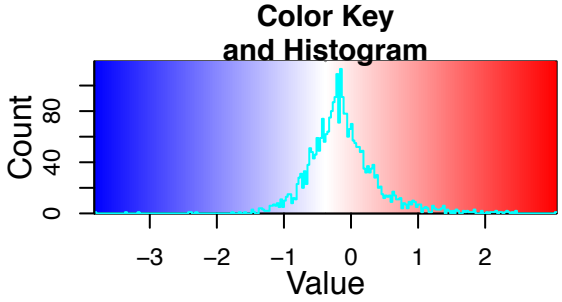


Figure S5

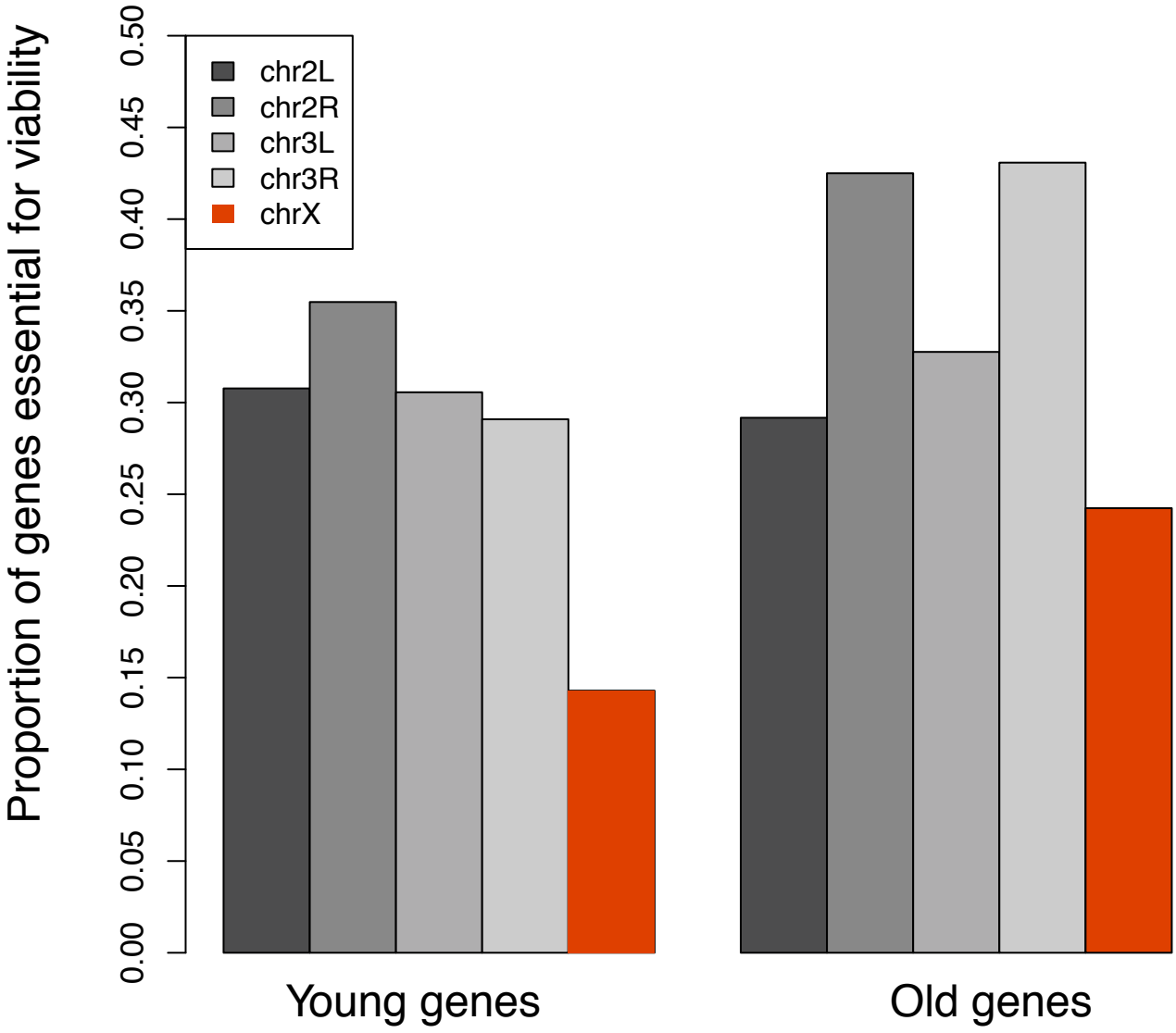
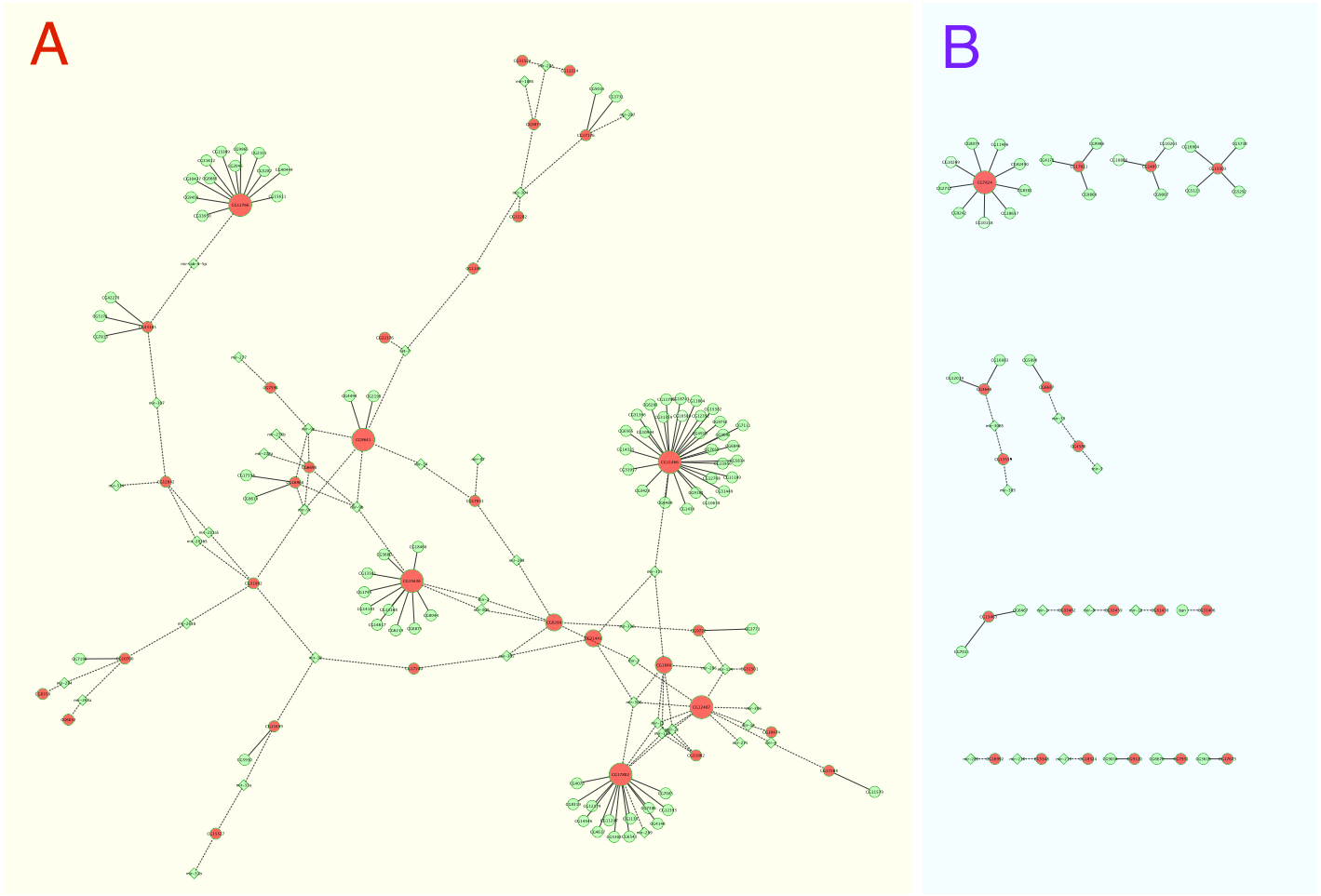

Proportion of genes essential for viability
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
chr2L
chr2R
chr3L
chr3R
chrX
Young genes
Old genes

Figure S6



A: YEG Network Core components

B: Peripheral Components

Network Legends:

Protein-protein interaction

microRNA-Target_Gene regulation

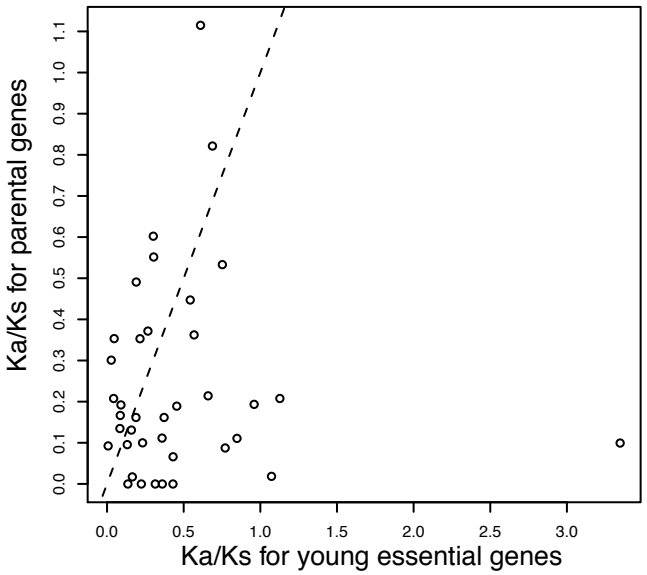
Young Essential Genes (YEGs)

YEG interacting proteins

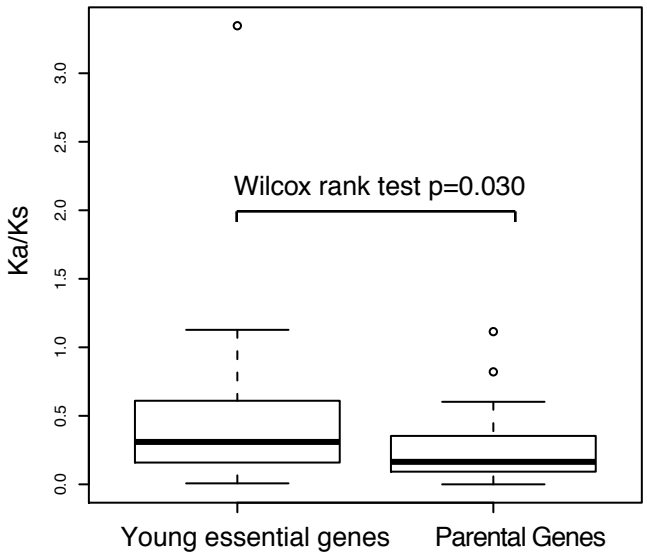
YEG interacting microRNAs

Figure S7

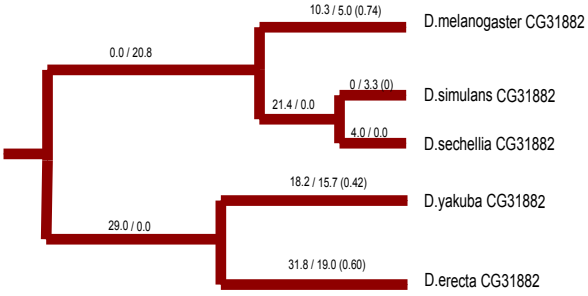
A



B



C



D

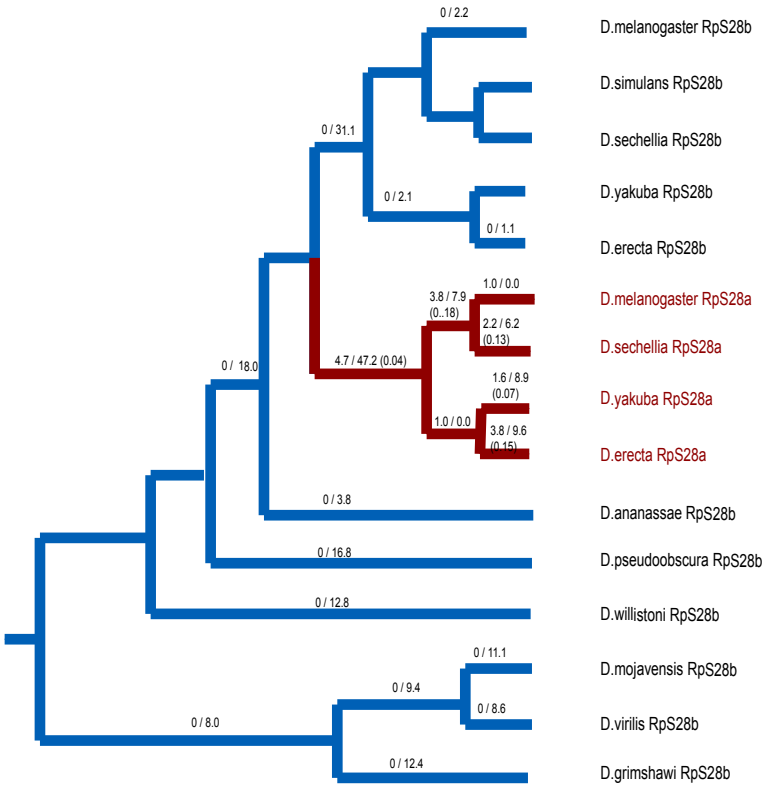
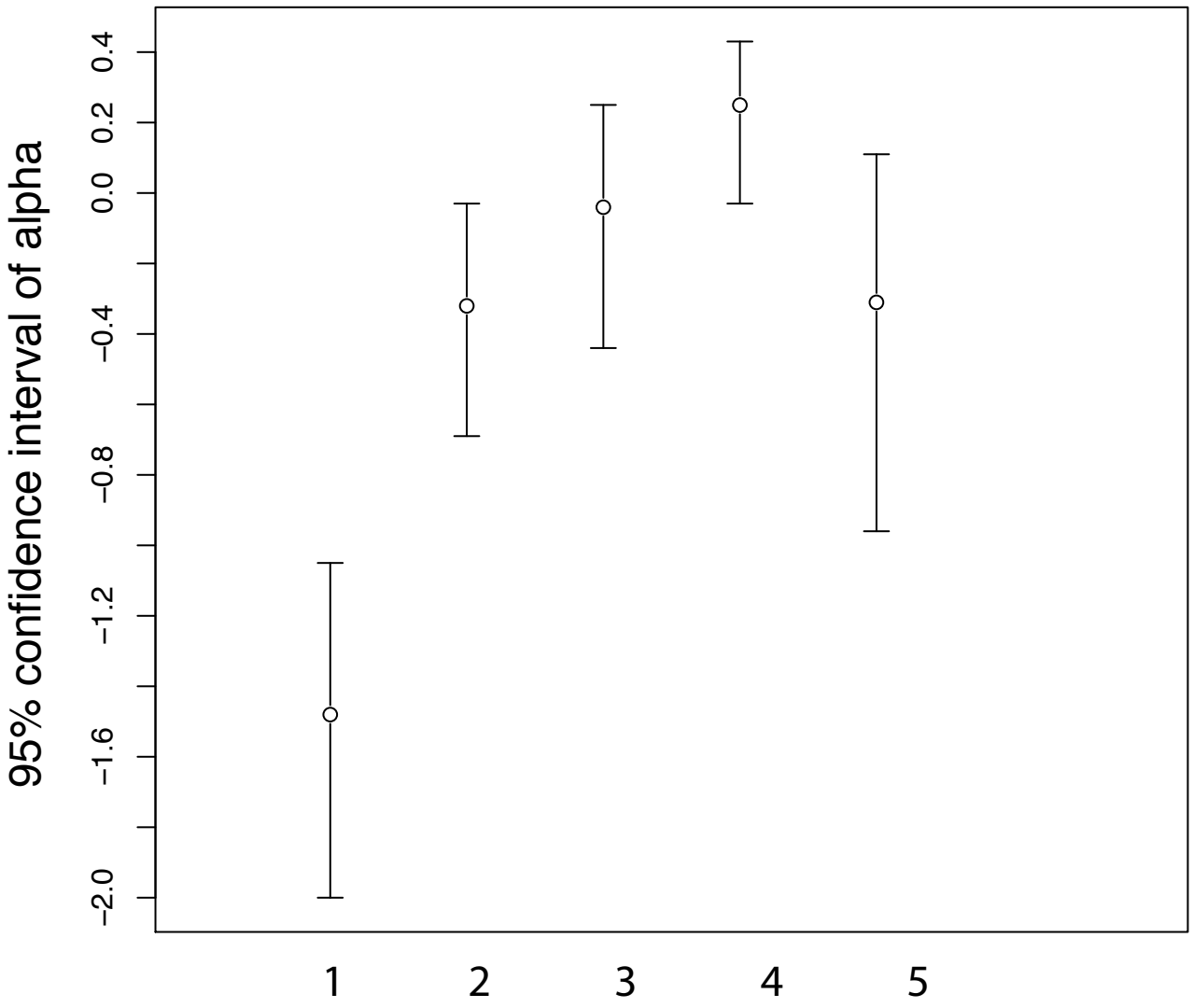


Figure S8



Group legends:

1. Old essential gene (branch 0)

2. Young essential gene (branch 2~3)

3. Parental gene of young essential gene (branch 2~3)

4. Young essentia gene (branch 4~5)

5. Parental gene of young essential gene (branch 4~5)