Roles of young serine-endopeptidase genes in survival and reproduction revealed rapid evolution of phenotypic effects at adult stages

Sidi Chen,^{1,†} Haiwang Yang,^{2,†} Benjamin H. Krinsky,² Anthony Zhang² and Manyuan Long^{2,*}

¹Department of Biology; and Koch Institute for Integrative Cancer Research; Massachusetts Institute of Technology; Cambridge, MA USA;

²Department of Ecology and Evolution; The University of Chicago; Chicago, IL USA

[†]These authors contributed equally to this work.

Our recent study found that 30% of young genes were essential for viability that determines development through stages from embryo to pupae in *Drosophila melanogaster*, revealing rapidly evolving genetic components involved in the evolution of development. Meanwhile, many young genes did not produce complete lethal phenotype upon constitutive knockdown, suggesting that they may not be essential for viability. These genes, nevertheless, were fixed by natural selection, and might play an important functional role in their adult stage. Here we present a detailed demonstration that a newly duplicated serine-type endopeptidase gene that originated in the common ancestor in the *D. melanogaster* subgroup 6–11 million years ago, named *Slfc*, revealing a strong effect in post-eclosion. Although animals survived constitutive knockdown of *Slfc* to adult stage, however, their life span reduced significantly by two-thirds compared to wild-type. Furthermore, the *Slfc-RNAi* males dropped their fertility to less than 10% of the wild-type level, with over 80% of these males being sterile. The *Slfc-RNAi* females, on the other hand, showed a slight reduction in fertility. This case study demonstrates that a young gene can contribute to fitness on the three important traits of life history in adults, including the life expectancy, male fertility and female fertility, suggesting that new genes can quickly evolve and impact multiple phenotypes.

New genes arose frequently in various organisms.^{1,2} Many young genes were recently identified in various *Drosophila* species.^{3–5} These young genes are often observed to be evolving under positive selection^{6–8} and many of them have been found to be associated with reproductive and other biological functions.^{2,9} These observations implicated significant phenotypic effects of new genes. We therefore attempted to investigate the phenotypic effects of these young genes in attempt to understand how they contribute to the phenotypic evolution of *Drosophila*. Unraveling their phenotypes becomes critical to detecting the underlying processes and mechanisms that drove the evolution of *Drosophila*.

In our recent genome-wide phenotypic study of the lineage-specific genes for their viability effects of lineage-specific genes, we knocked down the 195 genes that originated in recent 35 mys using constitutive RNAi. We found that 30% of these new genes showed complete lethality when silenced, which were thus essential genes.¹⁰ This suggests that a new gene can evolve essential functions rapidly and frequently. The remaining 70% of the new genes appear to be non-essential for viability; when constitutively knockdown, they eclose to viable adults.¹⁰ We sought to understand how these non-lethal young genes contribute to the phenotypes of *Drosophila* at the adult stage. Here, we present a genetic analysis of a young serine endopeptidase gene, *Short-lived and few-children (Slfc, CG4259)* for its important phenotypic effects and evolution. Our

Keywords: gene evolution, viability and reproduction, phenotype evolution, natural selection

Abbreviations My, million years

Submitted: 06/16/11

Revised: 08/15/11

Accepted: 08/18/11

<http://dx.doi.org/10.4161/fly.5.4.17808>

*Correspondence to: Manyuan Long;
Email: mlong@uchicago.edu

Extra view to: Chen S, Zhang YE, Long M. New genes in *Drosophila* quickly become essential. *Science* 2010; 330:1682–5; PMID:21164016; <http://dx.doi.org/10.1126/science.1196380>

Figure 1. Evolution of *Slfc*, a young duplicated serine endopeptidase. The simplified phylogenetic tree shows the origin of *Slfc* gene in the internal branch leading to the common ancestor of the *Drosophila melanogaster* species subgroup. The syntenic block from UCSC genome browser supports the presence of *Slfc* orthologs (orange star) in *Drosophila melanogaster* subgroup species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*), and the absence *Slfc* orthologs in outgroup species (*D. ananassae*, *D. pseudoobscura*, *D. persimilis*). Orange star in internal branch denotes the origin of the initial *Slfc* gene. Tree not drawn to scale. Levels 1–6 indicate the UCSC genome browser syntenic alignment level between *D. melanogaster* and another species. Level 1 alignments with *D. simulans*, *D. sechellia*, *D. erecta* and *D. yakuba* show continuous syntenicity in these species; non-level 1 alignment with *D. ananassae*, *D. pseudoobscura*, *D. persimilis* and *A. gambiae* (*Anopheles gambiae*) show syntenic breaks. Black and blue blocks are annotated gene regions. To view a color version of this figure, go to <http://www.landesbioscience.com/journals/21/article/17808>.

data show that this gene, after a short period of evolution, has been involved in multiple important biological processes and displayed several distinct phenotypic effects.

Results

Evolution of *Slfc*, a young duplicated serine endopeptidase. It has been reported that the serine endopeptidase class of protein coding genes has expanded into a large gene family by gene duplication,¹¹ implying their important roles in the adaptive evolution of *Drosophila*. A young serine-type endopeptidase gene, *Slfc*, arose by DNA-based interspersed duplication, after the divergence between *D. melanogaster* and *D. annanassae*, and before the divergence between *D. melanogaster* and *D. yakuba*, aged between 6–11 mys (Fig. 1). It is a young gene specific to the *Drosophila melanogaster* subgroup and has evolved rapidly, showing the distribution pattern of presence of the copy within the melanogaster subgroup and the absence of the copy in all tested outgroup species (Fig. 1) (Long M lab, unpublished data). The SLFC protein likely maintained the serine endopeptidase activity, because its active sites are present in the homology-based structure modeling (Fig. 1).

Using a similar method described previously in reference 10 and 12, we employ the RNAi procedure with the ubiquitous Act5C-Gal4 driver to knockdown constitutively the gene transcript of *Slfc* in order to examine the consequential phenotypic effect. We measured the level of *Slfc* mRNA transcript in *Slfc-RNAi* and control animals by semi-quantitative RT-PCR, and confirmed that RNAi knockdown reduced the mRNA level of the gene to an undetectable level, while the wild-type level showed an obviously high level of expression (Fig. 2). This indicates that the RNAi protocol successfully knocked down the expression of *Slfc*. We then measured several different aspects of the fitness phenotypes of this gene by constitutive knockdown.

The phenotypic effects on life expectancy. We measured how long the *Slfc-RNAi* animals lived under normal laboratory conditions. We grew the

Figure 2. Semi-quantitative RT-PCR confirms the reduction of *Slfc* in RNAi animals. Wild-type, control and *Slfc* RNAi animals were labeled on the right. The transcript level of *Slfc* is abundant in wild-type and control but reduced to merely detectable level in *Slfc* RNAi animals. Multiple PCR cycles were shown above the lanes.

Slfc-RNAi males and females in individual vials respectively and recorded their life span. We observed that the *Slfc-RNAi* animals lived for a much shorter time than the wild-type animals and all other controls (Fig. 3). Both males and females exhibited a lifespan reduction under *Slfc-RNAi*, suggesting that this gene is required for longevity in adult stages (Kormogorov-Smirnov (KS) test, $p < 0.01$ for all comparisons) (Fig. 3).

While both sexes showed phenotypic effects of life span reduction from the *Slfc-RNAi* experiments, the males and females were impacted to different degrees; males had significantly shorter lifespans than females. This observation suggests that the effect of gene knockdown is stronger in males than females (KS test, $p < 0.01$ for all comparisons) (Fig. 3).

We then tested whether temperature had an effect on the longevity of these flies. We reared animals at 17°C and 21°C. Under both conditions, the *Slfc-RNAi* animals lived for a shorter time than the wild-type flies and controls. At 17°C, within 25 days, the number of the *Slfc-RNAi* males and females subject to RNAi reduced by 25% and 15%, respectively. In the same period, only 5% of individuals died in the wild-types and controls. In 21°C, the effect of RNAi to the reduction of life expectancy in both males and females appears to be more pronounced, where 45% males and 24% females of the

RNAi animal died, compared to 5% in wild-types. At both conditions the RNAi death rates are significantly higher than wild-types and controls (KS test, $p < 0.01$ for all comparisons) (Fig. 3). The data indicated that the new gene *Slfc* is required for the well-being of wild-type flies and removal of the gene product led to severely shortened life expectancy.

Fertility effects of *Slfc*. We assay the fertility of *Slfc-RNAi* animals. We found that *Slfc* has a strong effect in male reproduction (Student's t test, $p = 3 \times 10^{-52}$). The number of offspring an average *Slfc-RNAi* male produce is less than 10% of the wild-type level (Fig. 4), and 86% of the *Slfc-RNAi* males are completely sterile, while in wild-type flies the sterile proportion is no larger than 5% (Fig. 4). The UAS and Gal4 controls do not shown significant reduction in male fertility compared to wild-type. However, because these observations were based on the knockdown experiments, potential residual transcripts, albeit in low concentration within the cells, might keep a low level of functions. Thus, the observed low fertility and high but incomplete sterility may just give a conserved experimental conclusion; the complete sterility cannot be ruled for a thorough knockout experiment. These data suggested that *Slfc* plays an important role in male reproduction.

Conversely, we also detected a small but significant effect on female reproduction:

Figure 3. Survival phenotype of *Slfc*. Standard survival curve of wild-type, control and *Slfc* RNAi animals at 17°C (A) and 21°C (B) aged in single vials in normal laboratory condition. Genotypes were labeled for each curve. Stars denote statistical significant comparisons ($p < 0.01$) by KS test.

the number of offspring an average RNAi female is reduced by one quarter (t test, $p = 0.0009$) (Fig. 5). There are 23% of *Slfc*-RNAi females that are completely sterile, compared to 2%–7% in wild-type females (Fig. 5). The UAS control also showed reduced female fertility due to slightly increased sterility (16%). The P-element insertion of this construct (UAS-*Slfc*-IR), when homozygous, seems to adversely

affect female fertility. Due to the fact that UAS-*Slfc*-IR and Act5C-Gal4 are both in heterozygous state, the effect did not come from homozygous insertion, but is still likely come from the interaction between UAS and Gal4, i.e., RNAi against *Slfc*, which is shown by RT-PCR (Figs. 2 and 5). These data imply that the gene *Slfc* might also play a significant role, although smaller than male, in female reproduction.

Discussion

The genetic basis underlying phenotypic evolution is critical for understanding the causes of adaptive evolution. For example, the explicit genetic changes in three interacted genes in beach mouse have been identified and found to be the causes for the adaptation to the sand color on beach by regulating the animals' coat color.^{13,14} In this paper, we reported a new genetic change responsible for the phenotype evolution: the presence of new genes, which evolved phenotypic effects on several distinct traits in life history. It is known that young genes have frequently originated in recent evolution. The number of adults that survive to reproductive age and the number of offspring a male or a female can reproduce are direct measurements of fitness, the phenotypic components critical to the understanding of the roles of new genes in evolution of *Drosophila*. These data revealed that the young gene *Slfc* have strong survival and reproductive phenotypes, implying that it contributes significantly to the overall fitness of the organism in multiple aspects. We found that the impact of *Slfc* varied due to different temperatures and sexes. In general, males are more affected when knocking down this young gene. These observations are in accordance with previously detected pattern of new genes that the excess of new genes evolved male functions, e.g., in expression in testis of fruit flies^{5,9} and mammals.^{2,15} The observed greater effects on the male fertility than on the female fertility is consistent with what was observed in the difference between the male and female expression.¹⁶

Structural modeling supports the functionality of the SLFC protein as a serine endopeptidase/protease (Fig. 6). The serine proteases are important components of extracellular signaling cascades in various biological processes, especially in embryonic development of invertebrates.¹⁷ The SLFC structure model displays a canonical SPL domain, which contains key catalytic residues (Histidine 72 and Glycine 210) in putative active sites (Fig. 6). Moreover, this protein has evolved rapidly among the multiple members. The rapid evolution of gene sequences as well as divergent expression patterns might be the reason why

Figure 4. Male reproduction phenotype of *Slfc*. (A) Overall fertility of males in fertility test, with genotypes labeled underneath. The *Slfc*-RNAi animal showed strong reduction in number of offspring produced when mated to wild-type virgin female, while related controls did not. (B) Percent of completely sterile males in fertility test, with genotypes labeled underneath. The *Slfc*-RNAi animal showed high percentage in complete sterile males, while related controls did not. *Statistical significance (Wilcoxon rank test, $p < 0.001$).

Figure 5. Female reproduction phenotype of *Slfc*. (A) Overall fertility of females in fertility test, with genotypes labeled underneath. The *Slfc*-RNAi animal showed a slight reduction in number of offspring produced when mated to wild-type virgin males. (B) Percent of completely sterile females in fertility test, with genotypes labeled underneath. The *Slfc*-RNAi animal showed an elevated percentage in complete sterile females compared to wild-types. The effects from UAS insertion are discussed in the text.

other paralogs are unable to provide full redundancy, thus depletion of *Slfc* leads to multiple severe phenotypic effects.

Duplicate genes which diverge from their paralogs generally face the

evolutionary fates of subfunctionalization, neofunctionalization and nonfunctionalization.¹⁸⁻²⁰ The new genes we are discussing are functional with important phenotype, thus rejecting the possibility

of nonfunctionalization. New genes usually evolve novel functions by neofunctionalization, which were advantageous and fixed by positive selection.^{1,18,21} Our data show that *Slfc* is involved in

Figure 6. Homology-based structural model of SLFC protein. Swiss-model of SLFC protein based on template structure 2b9IA. The overview structure shows a canonical SPL domain (A), which contains key catalytic residues (Histidine 72 and Glycine 210) in putative active sites (B).

multiple phenotypes, and evolved rapidly in protein sequences (Long M lab, unpublished data), which implies a consequence of neofunctionalization. More investigations on the evolution and function of the parental copy are needed to provide more conclusive evidence to test the alternative models. Along with evidence from molecular evolution and gene expression evolution, we proposed that new genes can quickly evolve novel functions and important phenotypes with a combination of paths.

Materials and Methods

Fly stocks and husbandry. The driver stock *yw;Act5C-Gal4,mini-w⁺/CyO,y⁺* was obtained from the Bloomington stock center. Wild-type stocks of *D. melanogaster* i.e., Oregon-R, Canton-S were generous gifts from Maolien Wu. RNAi line targeting *Slfc* (GD45271) was obtained from Vienna Drosophila RNAi Center (VDRC). All flies were raised in room temperature or 25°C.

Life span assay of *Slfc*. The adult flies of *Slfc-RNAi* (genotyped *yw; Act5C-Gal4, mini-w⁺/Slfc-IR*) and related controls (see Tables and Figs.) were collected and aged in single vials, depending on the assay. The number of surviving and dead individuals was scored manually each day.

Adult male and female reproduction (Fertility tests). Male and female fertility tests were carried out similarly to previously described methods (Chen et al., in revision). Briefly, for male fertility test, a single male of certain genotype being tested was crossed to three virgin females, allowed mating for 8-9 days, and the number of offspring was counted 10 days after. Female fertility test was performed by crossing a virgin female being tested to three virgin males. In all experiments, independent wildtype and control flies were set up for comparison.

Molecular biology. All molecular biology assays such as DNA/RNA extraction, PCR, semi-quantitative RT-PCR were carried out using standard protocols and reagents (Invitrogen). Primers for *Slfc* are

SLFCF1: 5'-ACT TTG CGT TTC
TCA CGA CA

SLFCR1: 5'-AGT TCA CAA TGC
CCA CCT GT

SLFCF2: 5'-CTT TGC GTT TCT
CAC GAC AC

SLFCR2: 5'-AGT TCA CAA TGC
CCA CCT GT

Evolutionary Analyses were carried out as described previously in references 10 and 16. Briefly, serine-endopeptidase gene sequences were retrieved from Flybase and ensemble. Gene origination events were carried out as previously described in references 10 and 16. Homology-based modeling was performed using Swiss-model.²²

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank DRSC, VDRC, DGRC, Bloomington Stock Center and the fly community for providing fly stocks, the Long lab members for discussions and assistance. This work is supported by NIH grant (NIH R01GM078070-01A1) and NSF grant (MCB1051826) to Manyuan Long and NSF Doctoral Dissertation Improvement Grant (DEB-1110607) to Sidi Chen.

References

1. Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 2003; 4:865-75.
2. Kaessmann H. Origins, evolution and phenotypic impact of new genes. *Genome Res* 2010; 20:1313-26.
3. Yang S, Arguello JR, Li X, Ding Y, Zhou Q, Chen Y, et al. Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genet* 2008; 4:3.
4. Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, et al. On the origin of new genes in *Drosophila*. *Genome Res* 2008; 18:1446-55.
5. Vibranovski MD, Zhang Y, Long M. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res* 2009; 19:897-903.
6. Long M, Langley CH. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 1993; 260:91-5.
7. Jones CD, Begun DJ. Parallel evolution of chimeric fusion genes. *Proc Natl Acad Sci USA* 2005; 102:11373-8.
8. Llopart A, Comeron JM, Brunet FG, Lachaise D, Long M. Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proc Natl Acad Sci USA* 2002; 99:8121-6.
9. Betran E, Thornton K, Long M. Retroposed new genes out of the X in *Drosophila*. *Genome Res* 2002; 12:1854-9.

10. Chen S, Zhang YE, Long M. New genes in *Drosophila* quickly become essential. *Science* 2010; 330:1682-5.
11. Ross J, Jiang H, Kanost MR, Wang Y. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene* 2003; 304:117-31.
12. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 2007; 448:151-6.
13. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. On the origin and spread of an adaptive allele in deer mice. *Science* 2009; 325:1095-8.
14. Manceau M, Domingues VS, Mallarino R, Hoekstra HE. The developmental role of Agouti in color pattern evolution. *Science* 2011; 331:1062-5.
15. Emerson JJ, Kaessmann H, Betran E, Long M. Extensive gene traffic on the mammalian X chromosome. *Science* 2004; 303:537-40.
16. Zhang YE, Vibranovski MD, Krinsky BH, Long M. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Res* 2010; 20:1526-33.
17. Piao S, Song YL, Kim JH, Park SY, Park JW, Lee BL, et al. Crystal structure of a clip-domain serine protease and functional roles of the clip domains. *EMBO J* 2005; 24:4404-14.
18. Walsh JB. How often do duplicated genes evolve new functions? *Genetics* 1995; 139:421-8.
19. Lynch M, O'Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. *Genetics* 2001; 159:1789-804.
20. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999; 151:1531-45.
21. Ohno S. *Evolution by Gene Duplication*. New York: Springer-Verlag 1970.
22. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006; 22:195-201.