Gene Duplication, Lineage Specific Expansion and Sub-functionalization of genes encoding MADF-BESS domain proteins in *Drosophila* development

A Thesis

Submitted in partial fulfillment of the requirements
Of the degree of
Doctor of Philosophy

By

Vallari Shukla

20103080

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

2016

CERTIFICATE

Certified that the work incorporated in the thesis entitled, "**Gene Duplication, Lineage Specific Expansion and Subfunctionalization of genes encoding MADF-BESS domain proteins in *Drosophila* development"** submitted by Vallari Shukla was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.

Girish Ratnaparkhi
Supervisor

Date: 24th October, 2016

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have not been properly cited or from whom proper permission has not been taken when needed.

Vallari Shukla

20103080

Date: 24th October, 2016

Acknowledgements

I owe my deepest gratitude to my research supervisor Dr. Girish Ratnaparkhi for being an excellent guide and above all a great person. His enthusiasm towards research and life are very unusual yet inspiring. I am thankful to him for providing me with all the support and encouragements. I thank him tremendously for introducing me to the world of fly genetics which is very exciting

I would like to thank Dr. Richa Rikhy for her critical comments and opinions on my research work and for sharing her lab reagents with me on various occasions. I would like to thank Dr. Anuradha Ratnaparkhi for open for discussing my work and her helpful suggestions. A special thanks to Dr. Girish Despande for helping me shaping up the ovary work in my research with his expert comments and necessary reagents. Their suggestions and critical thinking has proved extremely useful in obtaining major research insights during my PhD

I would also thank all the faculty members of Department of with special thanks to Prof. L S Shashidhara, Prof. Sanjeev Galande, and Dr. Aurnab Ghosh their constant support in my research as a part of my Research Advisory Committee. I would like to thank Dr. Farhat Habib and Dr. Neelesh Dahanukar for their help with the phylogenetics part of my research.

I was blessed to be part of a lab with extremely hardworking and highly motivated lab members. I extend my thanks for all their support during my research work. I owe a lot to my seniors Dr. Mithila Handu and Dr. Senthil Deivasigamani for helping me with the tools of the trade and for their scientific inputs. Also, I would like to acknowledge Miss Varsha Rajshekhar for her data on the eye work and Mr. Apurv Kulkarni for his help in the wing work in my and for actively participating my research work. I would like to thank my other lab Bhagyashree Kaduskar, Kriti Chaplot, Amar Soory, Prajna Nayak, Shewta Tendulkar, Neena Dhiman, Sushmita Hegde, Darshini Ravishankar and Aparna Thulsidharan for their co-operation and maintaining lively atmosphere in the lab. I would like to acknowledge all the previous Ratnaparkhi lab members for their help at various timepoints during my research period especially Srija Bhagvatula, Shalaka Chitale and Lokesh Pimpale. I would like to thank Dr. Richa Rikhy Lab members for their comments and suggestions during our combined lab meetings

I would like to owe my special thanks to my loving husband Dr. Kumar Somyajit for his love and support always. He has played a very important role in motivating me in life always. I would like to extend my special thanks to my good friends Dr. Mithila Handu, Miss Rini Shah, Mr. Ashi Zaheer Rizvi, Mrs Pallavi Singh, Miss Neeharika Gupta and Mrs Aishwarya Singh for their constant care, motivation and concern for me.

I would like to pay my special thanks to IISER Fly Facilty for helping to maintain our fly stocks and timely providing us with fly food. I would like to thank BioManager and her team for making our lives easier by providing us reagents whenever we need them.I extend my special thanks to Mr Vijay Vitthal from IISER microscopy facility for his support in data analysis and and microscopy.

Most importantly, I am grateful to my parents Mr Navanshu Kumar Shukla and Mrs Chitra Shukla for always believing in me and being with me at my worst times.I would like to thank my grandparents especially Mr.B.N Mishra who has always motivated me to do well in life. I would like to thank my brother Mr. Harsh Vardhan Shukla and my aunt Mrs Vithika Dikshit without whose support reaching here would not have been possible. Their invariable encouragement, love and belief have helped me in achieving my dreams. I thank them from the bottom of my heart.

Table of Contents

List of Figures

List OF Tables

List of Publications

Published

SHUKLA V., HABIB F., KULKARNI A., RATNAPARKHI G. S., 2014. Gene duplication, lineage-specific expansion, and subfunctionalization in the MADF-BESS family patterns the *Drosophila* wing hinge. Genetics **196**: 481–96.

Manuscript/ Review in Preparation

SHUKLA V., DESHPANDE G., RATNAPARKHI G. S. *Brickwall* is a gene-duplicate of *stonewall* with functional conservation in *Drosophila* germ-line stem cell development. (Manuscript).

SHUKLA V., DHANUKAR N., RATNAPARKHI G. S. The Drosophila MADF-BESS family – A model to study gene duplication in evolution. (Review).

Synopsis

Gene Duplication, Lineage Specific Expansion and Sub-functionalization of genes encoding MADF-BESS domain proteins in *Drosophila* development

Name of the student: VALLARI SHUKLA

Roll Number : 20103080

Name of the thesis advisor: Girish Ratnaparkhi

Date of Registration: 2nd August 2010

Indian Institute of Science Education and Research (IISER), Pune, India

Introduction

Gene duplication and subsequent diversification is a well-recognized phenomenon. Calvin Bridges was one of the first to put forward the idea of genetic units duplicating, specifically after observing symmetric, adjacent banding patterns in Polytene chromosomes (Bridges 1936). The idea of gene duplication influencing heredity, evolution and speciation was highlighted by Susumu Ohno (Ohno 1970) in his influential book, *Evolution by Gene Duplication*. In recent years, genome sequencing has confirmed the reality of extensive gene duplication in animal and plant genomes and its possible effects on the evolutionary process. The immediate effect of gene duplication at a locus is the duplication of both the coding sequence and the regulatory elements; leading to an increase in levels of transcripts and protein levels. The duplicated gene at that point can be retained in the genome as a functional copy or be lost. If the gene is maintained, it can be a functionally redundant duplicate or the gene sequence may be modified over time. Three possible fates of duplicated genes that are modified have been discussed in the literature (Force *et al.* 1999b; Lynch and Conery 2001; Kondrashov *et al.* 2002; Zhang 2003) as loss of gene (pseudogenization), sharing or dispersion of function (subfunctionalization), and the generation of novel functions (neofunctionalization). Duplication can occur at the level of the whole genome, chromosomal subsets or localized segments. Duplication can be followed by expansion, which leads to not two but multiple copies of the ancestral gene. Many paralogous 'families' in genomes are created by expansion, post duplication. *Drosophila melanogaster* is estimated to

have 674 protein families with biggest gene family as the trypsin family (111 members). One interesting expanded family is a family of ~ 55 proteins containing the DNA binding MADF (Myb-SANT like in Adf) domain. A subset of this family is the sixteen member MADF-BESS gene family, defined as a collection of genes that code for proteins which contain an N-terminal MADF followed by a C-terminal BEAF, Su-Var (3–7), Stonewall like (BESS) domain (Bhaskar and Courey 2002). Very few genes in the MADF-BESS family have been studied; functions of most of the members remain unknown. Adf1, a founding member, has been implicated as a transcriptional activator that contains a TAF-like binding motif (England *et al.* 1992; Cutler *et al.* 1998). Dorsal interacting protein3 (Dlip3) appears to be a co-activator in NF-kB/Twist function (Bhaskar and Courey 2002; Ratnaparkhi *et al.* 2008), while Co-repressor of Pangolin (Coop) is a negative regulator of Wg/Wnt signaling (Song *et al.* 2010) and *stwl* is required for germ-cell development (Clark and McKearin 1996) and germ stem-cell maintenance (Maines *et al.* 2007), acting as a repressor.

The retention of these 16 functional genes in the *D. melanogaster* genome suggests important roles for this family. If these genes have evolved from a common ancestral gene then they may retain ancestral function and thus may be partially or completely redundant. The genes have also diversified have evolved novel functions. In this study we examine redundancy in the MADF-BESS family using tools of phylogeny and genetics and attempt to understand retention or loss of ancestral functions

Results and Discussion

Lineage specific Expansion of MADF and BESS domains in *Drosophila*

To determine if the MADF and BESS domains have expanded in a specific lineage, the numbers of individual MADF and BESS domains were counted in a few representative organisms where genome sequences were available. Analysis of the phylogeny of dipterans and of sequenced *Drosophila* species indicated that the expansion probably occurred in an ancestor common to the *Drosophila* lineage, 40 million years ago. From the InterPro database, we found that the MADF domain alone occurs in 908 curated proteins whereas MADF together with a BESS domain occurs in 353 proteins. The concatenated phylogenetic tree consisting of 576 MADF genes in 12 *Drosophila* species is similar to the *Drosophil*a species tree based on the

whole *Drosophila* genome (Genomes 2007). Similarly the BESS phylogenetic tree consisting of concatenated 240 genes in 12 *Drosophila* species is similar to the Drosophila species tree based on whole *Drosophila* genome. The BESS domain was found to be most strongly associated with the MADF domain, with 16 of the twenty- two BESS-domain-containing genes in *D. melanogaster* being exclusively associated with the MADF domain, with no other intervening domains in the linker region. The 353 proteins containing a MADF-BESS domain were also analyzed on the basis of their distribution in different species.

A striking feature of the MADF-BESS family of proteins is their conserved protein architecture: all members have an N-terminal MADF domain and a C-terminal BESS domain. The proteins also show sequence similarity and/or identity for the two domains, indicating that the proteins have common evolutionary origins and that sequence motifs have been retained over time. Between the MADF and BESS domains is a "linker" region that ranges from 100 to 600 amino acids with little or no homology. An expanded phylogenetic tree of all MADF-BESS-containing genes in dipterans showed that the orthologs of the MADF-BESS genes in *Drosophilids* cluster together as would be expected if the gene duplication had happened before *Drosophila* speciation and the genes had diverged from each other. A phylogenetic tree with all members of the MADF-BESS family in *D.melanogaster* shows that these genes are related suggesting a gene duplication and/or expansion event followed by maintenance of these seemingly redundant genes in the animal. To understand the functions of the members of the MADF-BESS family in *Drosophila*, we decided on a loss-of-function approach, using the UAS-Gal4 system (Brand and Perrimon 1993; Duffy 2002) to reduce the transcript levels of every member of this family individually using double-stranded RNAi (Zamore *et al.* 2000; Kennerdell and Carthew 2000; Dietzl *et al.* 2007). The primary result of this screen was that different genes showed phenotypes in distinct tissues with subsets of MADF-BESS genes showing phenotypes in eyes, wing, ovary and the early embryo. In the subsequent chapters, I report dissection of redundant functions in the wing and the ovary.

Members of the MADF-BESS family pattern the wing hinge

When wing drivers were used to knockdown MADF-BESS domain genes, wing-hinge phenotypes were seen for three genes - CG9437, CG8359, and CG13897. Based on the phenotype, we have named these genes *hinge1 (hng1), hinge2 (hng2),* and *hinge3 (hng3),*

respectively. The phenotype is 100% penetrant and is dose dependent with an increased knockdown leading to a stronger phenotype that also affects the more distal wing blade. The defects can be rescued by co-expression of UAS-*hng1* in the same expression domain suggesting that the RNAi line specifically affects *hng1* transcripts. Over-expression of UAS-*hng1* by itself does not affect normal wing development. The primary phenotype was a bend in the costa region of the wing hinge and a reduction in the size of the hinge, with a dramatic effect on the patterning and size of the alula with respect to wild type. The phenotype was dose dependent, with drastic reduction in alula and wing-blade size with an increase in UAS or Gal4 dosage. Phenotypes similar to *hng1*i were seen with knockdowns of *hng2* and *hng3* using gene-specific RNAi lines expressed in the *MS1096* expression domain. *hng1*, *hng2*, and *hng3* thus appeared to have critical roles in wing-hinge development. Simultaneous knockdown of *hng2* and *hng1* or *hng3* and *hng1* leads to enhanced wing-hinge phenotypes.

***hinge1* interacts with genes of wing-hinge gene regulatory network.**

Since flies lacking *hng1* have a wing-hinge defect, we tested for genetic interactions between *hng1* and genes that play important roles in wing-hinge development. A central pathway involved in wing-hinge development consists of *teashirt (tsh), homothorax (hth),* and *extradenticle (exd)*(Mann and Abu-Shaar 1996; Casares and Mann 2000; Azpiazu and Morata 2000; Wu and Cohen 2002; Zirin and Mann 2004). *tsh* acts like an activator of *hth,* and binding of Hth is necessary for nuclear localization of Exd. The Hth:Exd complex then activates downstream targets that pattern the wing hinge. A double knockdown of *tsh* and *hng1* rescued the hinge defect as well as the size and patterning of alula, indicating that *hng1* is a negative regulator of *tsh* function. A similar rescue was seen upon simultaneous *hth* and *hng1* knockdown.

A major player in wing-hinge development is Wg, which has roles in patterning by restricting the tsh-hth network to the wing hinge. Wg staining in the third instar wing disc marks a near-concentric outer (wg-OR) and inner ring (wg-IR) (Couso *et al.* 1993; Neumann and Cohen 1996b; Russell 2000; Cavodeassi *et al.* 2002) with a gap in between. The two rings are critical regions for wing-hinge development with the members of the wing-hinge gene regulatory network (GRN) interacting with or regulating Wg or being regulated by expression of Wg. wg-IR, regulated by the Wg^{spade}-flag enhancer (Neumann and Cohen 1996b) patterns a major section of the region of the hinge that is affected in the *hng1* knockdown. When *hng1* is knocked

down in the MS1096 expression domain (MS1096/+; UAS-*hng1*i/UAS-*hng1*i), a broadening of the Wg expression domain with an intrusion into the gap region was observed. Knockdown of *hng1* in the patched (*ptc-Gal4)* expression domain (*ptc-Gal4/+; UAS-hng1i/+)* also leads to derepression of Wg in the gap region.

Testing redundancy using double gene knockdowns and heterologous rescue

hng1, *hng2*, and *hng3* appear to be genes with similar or equivalent roles in the wing hinge. Other MADF-BESS knockdowns, with the exception of *stwl*, do not appear to give a hinge phenotype in single knockdown experiments. It is possible that these genes have a partially redundant function in the wing hinge. One method of testing this would be to simultaneously knock down each of these genes, along with *hng1*, and check for an enhancement of the *hng1* phenotype. Double knockdowns on all remaining MADF-BESS genes in the background of *hng1* RNAi indicated that *Dip3, Coop, CG3838, CG11723, and CG4404* had roles in wing-hinge patterning as they enhanced the *hng1* phenotype. This indicates that a substantial fraction of the MADF-BESS family (9 of 16) plays a direct or a supporting role in wing-hinge development. We also attempted to rescue the hng1 phenotype by overexpressing other MADF-BESS genes in the *MS1096* expression domain. On testing the available UAS lines, we found that *CG13204* and *CG11723* expression could partially rescue the wing phenotype. This data again suggested that a subset of MADF-BESS genes retained ancestral function.

***brwl* is required for ovary development in an age dependent manner**

To understand the role of MADF-BESS genes in ovary development we used UAS-Gal4 system (Brand and Perrimon 1993; Duffy 2002) to knockdown these genes one by one in the ovary. Valium 20 and Valium 22 lines available in Bloomington Fly were used for this study. CG3838 upon knockdown with *nanos*-Gal4 gave a fused ovary phenotype and other oocyte defects. Based on the phenotype and the fact that mutants of one another MADF-BESS gene known as *stonewall* shows similar defects (Clark and McKearin 1996; Akiyama 2002) we named CG3838 gene *brickwall* (*brwl*). The *brwl* loss of function phenotype is age dependent; the ovaries of newly emerged flies do not show this phenotype and only flies aged until seven days or beyond show this phenotype. The phenotype also enhances in terms of its penetrance and severity as the flies age from 3rd to 18th day post eclosion. CG3838 mutant ovaries dissected on 10th and 18th day post eclosion show a drastic change in the ovariole architecture; individual egg

chambers are not formed, oocyte fails to specify and apoptotic nuclei are seen in 18-days old fly ovaries. Two P-element insertions, available in BDSC, were obtained. The P-element *brwl*^{KG00824} is inserted at the N-terminal of *CG3838* and is a part of BDGP Gene Disruption Project (Bellen *et al.* 2004) collection. The P-element *brwl*^{MI054561} is inserted at the C-terminal of the gene and is the part of transposon Minos-mediated integration cassette (MiMic) (Venken *et al.* 2011) collection. When tested, both *brwl*^{KG00824} and *brwl*^{MI054561} females showed ovary defects when homozygous. *brwl*^{MI054561} shows a stronger phenotype than *brwl*^{KG00824} at the same stage. A strong phenotype is also seen in ovaries of trans-heterozygous *brwl*^{KG00824} / *brwl*^{MI054561} females. As in the RNAi experiments, *brwl* mutants also show age dependent phenotype however in the case of the insertional mutants, phenotype was seen after 10 days of fly eclosion as opposed to 7 days for that of *nanos>brwlRNAi* . By 18th day the defects are so drastic in case of the mutants that the whole ovarian architecture is lost and the ovarioles seem to be full of undifferentiated cells and similar to tumorous ovarioles. Staining with phalloidin shows loss of ring canals in 10 and 18 day old ovarioles, which in turn possibly affects also affects oocyte development. DAPI stained nuclei in *brwl*^{KG00824} / *brwl*^{MI054561} resemble apoptotic nuclei suggesting probable cell death in the tissue.

The oocyte is initially specified in the germarium and in its mature form is present at the posterior end of the egg chamber, with 15 nurse cells, both being surrounded by follicle cells. *orb* is required in ovary development at multiple levels- for formation of 16-cell cyst, differentiation of egg chamber and polarity establishment (Lantz *et al.* 1994). Orb is expressed in a crescent shaped pattern in the posterior end of the developing oocyte. *brwl*^{KG00824} / *brwl*^{MI054561} ovaries were stained with oocyte specific marker like *orb* and they show loss of oocyte specification, mis-orientated oocyte and sometimes loss of oocytes altogether. 79.6% of the total ovarioles showing the phenotype on 10th day ovaries show mis-orientated oocyte and 20.83% show no oocytes. By the 18th day as more than 63.33% of the ovarioles showing the phenotype show no oocytes while around 36.67% show mis-oriented oocytes. DAPI staining of *brwl*^{KG00824} / *brwl*^{MI054561} mutants showed pycnotic nuclei that suggest cell death. In order to confirm apoptosis, we stained the ovaries with an apoptotic marker. *brwl* mutants ovaries undergo cell death as shown by staining with the apoptotic marker *caspase-3*. 88.88% of ovarioles showing the phenotype are positive for caspase activity on the 10th day after fly eclosion while all the

ovarioles showing the phenotype are positive for caspase activity on the 18th day. Increase in cell death most likely contributes to oogenesis arrest in these mutants.

Fusomes are germline specific organelles that play an important role in cyst formation. In germ stem cells and cystoblast fusomes are round and spherical in shape often referred to as Spectrosomes (Lin *et al.* 1994). Fusomes are made up of membrane skeletal proteins like alpha and beta spectrin and adducin-like protein called as *hu-li tao shao (hts)*. $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries were stained with *alpha-spectrin.* We counted the number of spectrosomes and fusomes. In wild type usually there are 2-3 spectrosomes are present, however in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries only1 spectrosome was present or totally absent in 10 day old flies. This depicts decrease in the number of germ stem cells (GSCs) and cystoblast. Number of fusomes present in the wild type is 9-11, however $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries did not show any significant change in the number on the 10th day (6-9).However by 18th day there was a significant decrease in the number of fusomes (3-5 fusomes). To further verify the decrease in number of GSCs, a GSCmarker *dad-lacZ* was used to stain the GSCs. *daughters against dpp (dad)* is a target gene of dpp and is only expressed in GSCs only. In cystoblast *dad* expression decreases with increasing distance from the dpp niche. $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show 50% decrease in the number of *dad-lacZ* positive cells in 10 day old flies.

brwl mutant phenotype is rescued by over-expression of *stwl* in germ stem cells

stwl is a MADF-BESS gene shown to be essential for germ cell development (Clark and McKearin 1996). *stwl* mutants block oocyte differentiation and cause the presumptive oocyte to develop as nurse cell. It is also required for proper cystoblast differentiation into cysts. *stwl* mutants also show a decrease in the number of GSCs and cysts. Overexpression of stwl with nanos-Gal4 in $brwl^{MI054561}$/ $brwl^{MI054561}$ background leads to the rescue of the ovarian phenotype. Egg chambers are formed with 15 nurse cells and a developing oocyte. 32.8% of total $brwl^{MI054561}$/ $brwl^{MI054561}$*; nanos-Gal4/UAS-stwl-HA* ovarioles show ovary defects which is almost a 50% reduction compared to $brwl^{MI054561}$/ $brwl^{MI054561}$ ovarioles. $brwl^{MI054561}$/ $brwl^{MI054561}$*; nanos-Gal4/UAS-stwl-HA* ovarioles show proper arrangement of ring canals in the egg chambers. Overexpression of *stwl* in $brwl^{MI054561}$/ $brwl^{MI054561}$ background rescues the oocyte

defects depicted by orb staining. Interestingly, overexpression of *stwl* in $brwl^{MI054561}$/ $brwl^{MI054561}$ background did not decrease the apoptotic activity that $brwl^{MI054561}$/ $brwl^{MI054561}$ ovarioles show.

References

AKIYAMA T., 2002 Mutations of stonewall disrupt the maintenance of female germline stem cells in Drosophila melanogaster. Dev. Growth Differ. **44**: 97–102.

AZPIAZU N., MORATA G., 2000 Function and regulation of homothorax in the wing imaginal disc of Drosophila. Development **127**: 2685–93.

BELLEN H. J., LEVIS R. W., LIAO G., HE Y., CARLSON J. W., TSANG G., EVANS-HOLM M., HIESINGER P. R., SCHULZE K. L., RUBIN G. M., HOSKINS R. A., SPRADLING A. C., 2004 The BDGP gene disruption project: Single transposon insertions associated with 40% of Drosophila genes. Genetics **167**: 761–781.

BHASKAR V., COUREY A. J., 2002 The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. Gene **299**: 173–84.

BRAND A. H., PERRIMON N., 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118**: 401–15.

BRIDGES C. B., 1936 THE BAR "GENE" A DUPLICATION. Sci. **83**: 210–211.

CASARES F., MANN R. S., 2000 A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identities in Drosophila. Development **127**: 1499–508.

CLARK K. A., MCKEARIN D. M., 1996 The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development **122**: 937–950.

COUSO J. P., BATE M., MARTINEZ ARIAS A., 1993 A wingless-dependent polar coordinate system in Drosophila imaginal discs. Science (80-.). **259**: 484–489.

CUTLER G., PERRY K. M., TJIAN R., 1998 Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol. Cell. Biol. **18**: 2252–61.

DIETZL G., CHEN D., SCHNORRER F., SU K.-C., BARINOVA Y., FELLNER M., GASSER B., KINSEY K., OPPEL S., SCHEIBLAUER S., COUTO A., MARRA V., KELEMAN K., DICKSON B. J., 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature **448**: 151–6.

DUFFY J. B., 2002 GAL4 System in Drosophila : A Fly Geneticist ' s Swiss Army Knife. Genesis **15**: 1–15.

ENGLAND B. P., ADMON a, TJIAN R., 1992 Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc. Natl. Acad. Sci. U. S. A. **89**: 683–7.

FORCE a, LYNCH M., PICKETT F. B., AMORES a, YAN Y. L., POSTLETHWAIT J., 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**: 1531–45.

GENOMES D. 12, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature **450**: 203–18.

KENNERDELL J. R., CARTHEW R. W., 2000 Heritable gene silencing in Drosophila using double-stranded RNA. Nat. Biotechnol. **18**: 896–8.

KONDRASHOV F. a, ROGOZIN I. B., WOLF Y. I., KOONIN E. V, 2002 Selection in the evolution of gene duplications. Genome Biol. 3 Res. **3**: 1–9.

LANTZ V., CHANG J. S., HORABIN J. I., BOPP D., SCHEDL P., 1994 The Drosophila orb RNA-binding protein is required for the formation of the egg chamber and establishment of polarity. Genes Dev. **8**: 598–613.

LIN H., YUE L., SPRADLING A. C., 1994 The Drosophila fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. Development **120**: 947–56.

LYNCH M., CONERY J. S., 2001 The Evolutionary fvate and Consequences of Duplicate Genes. Science (80-.). **290**: 1151–1155.

MAINES J. Z., PARK J. K., WILLIAMS M., MCKEARIN D. M., 2007 Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development **134**: 1471–1479.

MANN R. S., ABU-SHAAR M., 1996 Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. Nature **383**: 630–3.

NEUMANN C. J., COHEN S. M., 1996 Distinct mitogenic and cell fate specification functions of wingless in different regions of the wing. Development **122**: 1781–9.

OHNO S., 1970 *Evolution by gene duplication.*

RATNAPARKHI G. S., DUONG H. a, COUREY A. J., 2008 Dorsal interacting protein 3 potentiates activation by Drosophila Rel homology domain proteins. Dev. Comp. Immunol. **32**: 1290–300.

RODRÍGUEZ DD D. D. A., TERRIENTE J., GALINDO M. I., COUSO J. P., DÍAZ-BENJUMEA F. J., 2002 Different mechanisms initiate and maintain wingless expression in the Drosophila wing hinge. Development **129**: 3995–4004.

RUSSELL S., 2000 The Drosophila dominant wing mutation Dichaete results from ectopic expression of a Sox-domain gene. Mol. Gen. Genet. **263**: 690–701.

SONG H., GOETZE S., BISCHOF J., SPICHIGER-HAEUSERMANN C., KUSTER M., BRUNNER E., BASLER K., 2010 Coop functions as a corepressor of Pangolin and antagonizes Wingless signaling. Genes Dev. **24**: 881–886.

VENKEN K. J. T., SCHULZE K. L., HAELTERMAN N. a, PAN H., HE Y., EVANS-HOLM M., CARLSON J. W., LEVIS R. W., SPRADLING A. C., HOSKINS R. a, BELLEN H. J., 2011 MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods **8**: 737–743.

WU J., COHEN S. M., 2002 Repression of Teashirt marks the initiation of wing development. Development **129**: 2411–8.

ZAMORE P. D., TUSCHL T., SHARP P. A., BARTEL D. P., 2000 RNAi : Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell **101**: 25–33.

ZHANG J., 2003 Evolution by gene duplication: An update. Trends Ecol. Evol. **18**: 292–298.

CHAPTER 1

Introduction

Genetic redundancies and their evolutionary maintenance

1.1 Summary

In this chapter, I introduce the mechanisms, molecular processes and fate of genes that duplicate and are thereafter modified by natural selection. Representative examples of duplicated genes from literature are described to explain the process of gene duplication and subsequent changes in the sequences of the duplicated genes. At the end of the Chapter, I examine an interesting family of genes that encode proteins with MADF, BESS or MADF-BESS domains. This family codes for a set of transcription factors that may have arisen from the duplication and expansion of an ancestral gene. I describe the reasons for my interest in this family and the need for dissection of redundant function.

1.2 Gene Duplication: Mechanism for the evolution of new genes

Gene duplication and subsequent diversification of one or both of the duplicated genes is a well-recognized phenomenon in evolution. Calvin Bridges was one of the first to put forward the idea of genetic units duplicating, specifically in response to symmetric, adjacent banding patterns in Polytene chromosomes(Bridges 1935). This idea gained further ground in early studies of the Bar locus (Sturtevant 1925) ;reviewed in (Duncan, I., Montgomery 2002; Zhang 2003; Taylor and Raes 2004) and took root as a fundamental concept in genetics as a result of studies of the bithorax complex by E.B. Lewis (Lewis 1978). The Hox genes are today a textbook example of gene duplication and diversification (Liberles *et al.* 2010). The idea of gene duplication influencing heredity, evolution and speciation was enhanced further by Susumu Ohno (Ohno 1970) in his seminal book, *Evolution by Gene Duplication*. Among other ideas, Ohno highlighted the role of the 'extra' copy of the duplicate gene(s) being available for modification by the process of natural selection. In recent years, genome sequencing efforts and their resultant analysis has further cemented the reality of gene duplication (Table 1.1) and its

role in invertebrate and vertebrate evolution. Many mechanisms have been proposed to explain the cause and consequence of gene duplication. Gene duplication and expansion has a direct effect on evolution, speciation and the patterning of new life forms. Studies suggest that duplications at the level of single genes, neighboring loci, chromosomes and whole genomes have occurred regularly as life evolved. Lynch and Connery have calculated that for a number of eukaryote species rate of duplication is 0.01 per gene per million years (Lynch and Conery 2001).

Table 1.1 Occurrence of Gene Duplication in sequenced genomes: Bacteria, Archaebacteria and Eukaryotes (Adapted from (Zhang 2003). Data from (Himmelreich *et al.* 1996) (Klenk *et al.* 1997) (The Arabidopsis Genome Initiative 2000) (Rubin *et al.* 2000) (Li *et al.* 2001))

Organism	**Total number of genes**	**Number of duplicated genes**	**Percentage of genome duplicated**
Mycoplasma pneumoniae	677	298	44
Archaeoglobus fulgidus	2436	719	30
Saccharomyces cerevisiae	6241	1858	30
Caenorhabditis elegans	18424	8971	49
Drosophila melanogaster	13601	5536	41
Arabidopsis thaliana	25498	16574	65
Homo sapiens	40580	15343	38

Apart from gene duplication whole genome can be also duplicated. While gene duplication is more prevalent and gives rise to new genes, whole genomes duplications though rare are responsible for speciation. Ohno stated that whole genome duplications provide raw material for evolution by allowing duplication and divergence of entire pathways. Genomes of yeast, plants and fishes have undergone whole genome duplications. *Saccharomyces cerevisae* is described as degenerate tetraploid resulting from mating of two ancestral genomes and doubling the DNA content (Wolfe and Shields 1997). The genome of *Arabidopsis* has undergone whole genome duplication followed by gene loss and several local duplications (The Arabidopsis Genome Initiative 2000). In vertebrates teleosts fishes provide evidence for whole genome duplication which have occurred between divergence of ray finned fishes from tetrapods (Christoffels *et al.* 2004) .

Tandem duplications or segmental duplications involve duplication of the same exons, which gives rise to paralogous genes which then form gene families. *Drosophila* is estimated to

have 674 protein families with biggest gene family as the trypsin family (111 members). Yeast and *C.elegans* have 540 and 1,219 protein families respectively (Gu *et al.* 2002). The most famous example of tandem duplication are the *Hox* genes which are transcription factors and regulate animal development by specifying body plan all throughout animal kingdom. A *Hox* gene cluster usually comprises of 8-13 genes situated next to each other. They are a result of tandem duplication of a ProtoHox gene. Amphioxus (cephalochordate) has one single Hox cluster and duplication events at the time of vertebrate evolution have produced minimum of four Hox clusters in vertebrates. Another duplication event during teleosts lineage diversification have produced a seven Hox gene clusters in zebrafish (Brunet *et al.* 2006). *Drosophila* has a single cluster consisting 8 genes whereas humans have four clusters with 39 genes. In this chapter we will discuss mechanisms and consequences of gene duplication with particular interest in gene families.

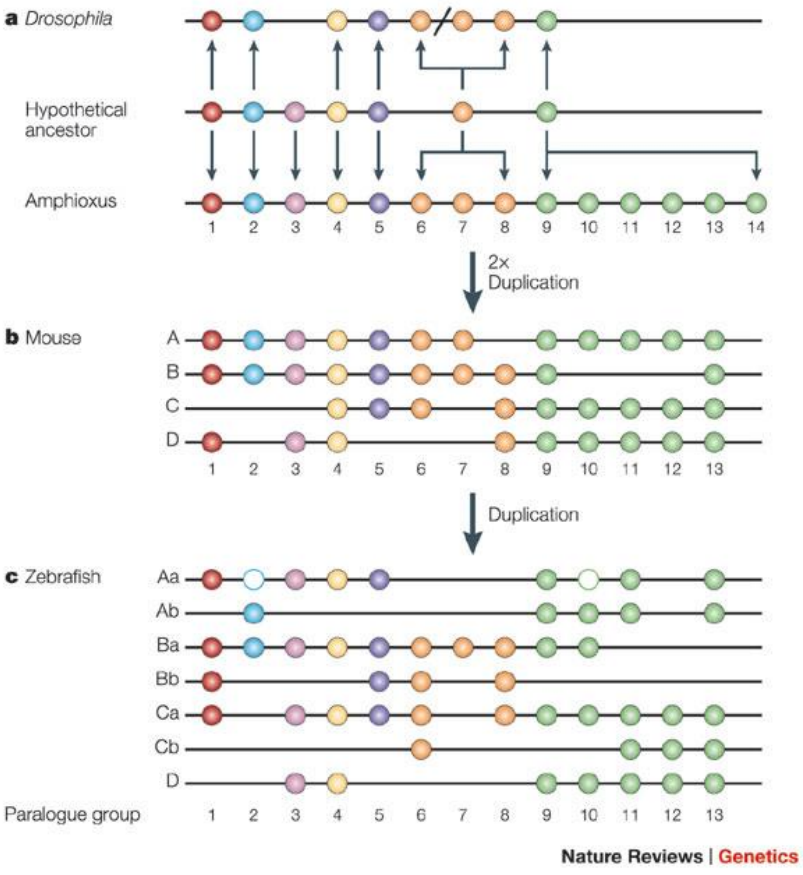


Figure 1.1. Duplication events have caused expansion of the *Hox* genes: a) The *Drosophila* and *Amphioxus* Hox clusters appear to trace their ancestry to an (ancestral) *Hox* cluster, which was itself a result of tandem duplications of a single *Hox* gene. **b)** Tetraplods, including mice and humans have four Hox clusters on four different chromosomes while **c)** Zebrafish has seven clusters as a result of additional duplications. Adapted from (Prince and Pickett, 2002).

1.3 Mechanisms of Gene duplication

There are various mechanisms that contribute to the duplication in all genomes. The mechanism by which a gene is duplicated often depends on the size of the duplication. For example, polyploidy is the result of whole genome duplications while mechanisms like unequal crossing over and replication slippage give rise to small scale or segmental duplications. Unequal crossing over results in tandem duplications as the duplicates situated are side by side. Evidence from *Arabidopsis*, *Saccharomyces cerevisae*, and humans show that there are at least 10%-20 % genes are tandem arrayed duplicates (The Arabidopsis Genome Initiative 2000; Lander *et al.* 2001) while in *C.elegans* the number is as high as 70% (Gu *et al.* 2002). A sudden explosion of Alu elements (in primates) are believed to be a consequence of tandem duplication because of unequal crossing over (Lander *et al.* 2001). Replication slippage is another mechanism that can result in duplication of short segments like microsatellites (Hahn 2009).

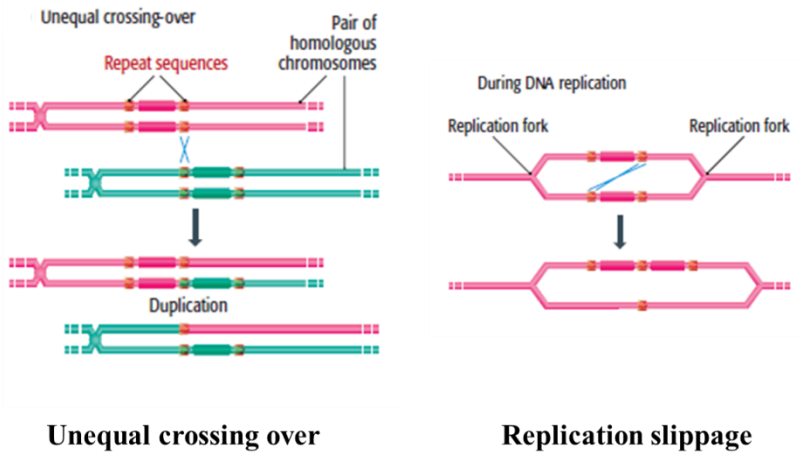


Figure 1.2. Unequal Crossing over and Replication slippage caused tandem duplicates. Unequal crossing over results in deletion of a segment from one strand and duplication in its homologous chromosome. Replication slippage leads to nucleotide expansion during DNA replication (Images Adapted from Sciknowledge Genome evolution webpage (https://sciknowledge.wordpress.com/2013/01/06/genome-evolution/)).

DNA transposition and Retrotransposition employ either DNA or RNA transposon elements to copy and move DNA in the genome. This causes duplication of exons or small genes and their insertion anywhere in the genome. Recent studies show pools of transposable elements at the junctions of interchromosomal duplicated DNA in *Drosophila* and humans (Bailey *et al.* 2001; Fiston-Lavier *et al.* 2007). Retrotransposition of genes occurs from the reverse transcription of mRNA to DNA and its insertion in a new genomic locus. The new paralogs that form will initially lack introns and poly A tail and do not have any flanking DNA; as a result there expression is dependent on whether they land in other coding regions (Hahn 2009).

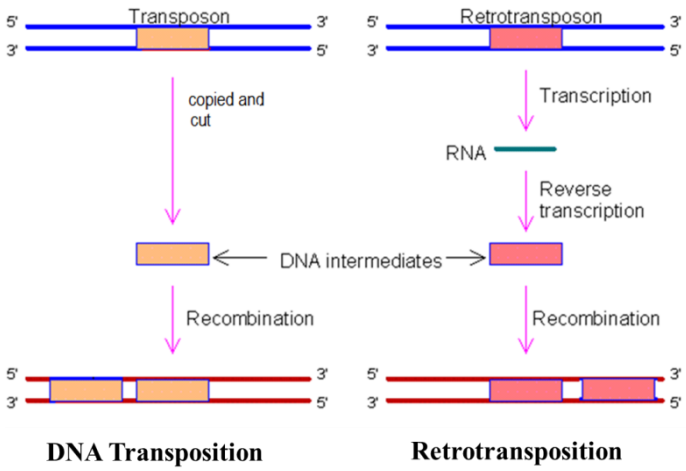


Figure 1.3. DNA Transposition and Retrotransposition cause interchromosomal duplicates. Transposition can be achieved by either DNA Transposable elements or RNA intermediates. Images adapted from Sciknowledge Genome evolution webpage (https://sciknowledge.wordpress.com/2013/01/06/genome-evolution/).

Polyploidization is a powerful mechanism for duplication which leads to whole genome duplication (Figure 1.4). Though whole genome is duplicated, in most cases, with time, only about 10-30% of duplicates are maintained (Wolfe and Shields 1997). The duplicate genes retained after whole genome duplication bear high similarity in many species. Genes retained from polyploidy events in yeast and *Arabidopsis* include ribosomal proteins and kinases. Similarly in *C.elegans* and *Arabidopsis* after whole genome duplication, duplicate genes coding for transcription factors are retained.

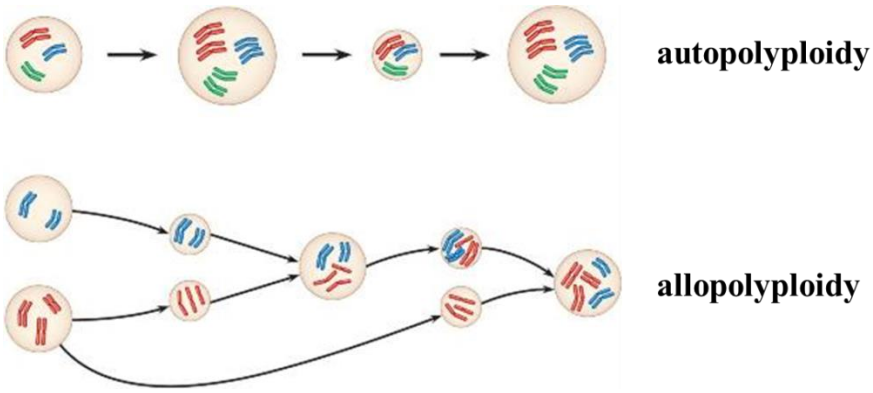


Figure 1.4. Autopolyploidy and allopolyploidy. Autopolyploidy occurs when an individual has two or more pairs of chromosomes derived from the same species. Allopolyploidy are polyploids when chromosome pairs come from the mating of two species. Nearly 80% of plants are allopolyploids (Images Adapted from Sciknow Genome evolution webpage).

1.4 Fate of Duplicate Genes

Gene duplication creates a redundant copy of the gene. The availability of two copies allows one (or both copies) to be modified without detrimental effect on the organism or even lost. If the duplicates are a complete copy; then redundancy is established. This complete redundancy may not be stable over long time scales and usually genes sequences and their subsequent functions are modified. Sometimes certain mutations can lead to significant changes in gene sequences, such as gain of novel gene/protein function. After duplication, an immediate loss of one copy maintains the *status-quo*. However, in certain instances the duplicates also get fixed in the population (Figure 1.5). They can be fixed as the parent gene in which case it provides functional redundancy. A mutation in the coding region or regulatory region or in both which is not deleterious providing either a new function to the gene can preserve the duplicate gene (Figure 5) (Force *et al.* 1999a; Zhang *et al.* 2003; Hahn 2009; Innan and Kondrashov 2010). Ohno outlined three outcomes after gene duplication 1) Neo-functionalization (Force *et al.* 1999b), which is the evolution of a new function in one of the duplicates 2) Sub-functionalization, a phenomenon that involves the division of ancestral functions among duplicates and 3) Gene Conservation or Redundancy, the conservation of all functions in both duplicates (Ohno 1970). In this chapter we will discuss the outcomes of gene duplication in detail and certain models proposed by various studies so far.

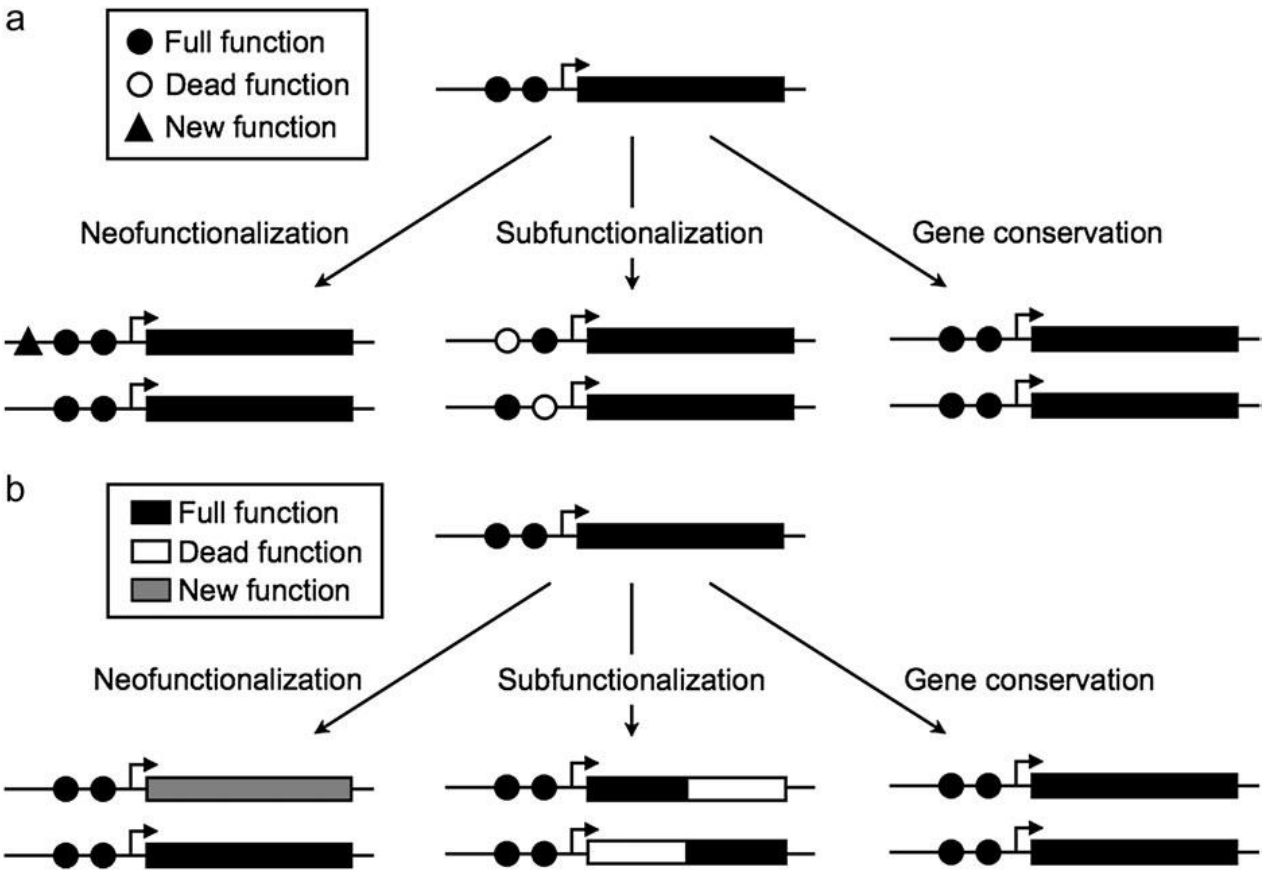


Figure 1.5. Fate of duplicate gene: mutations or deletions in the a) regulatory region or b) in the coding regions can lead to neofunctionalization or subfunctionalization . In very rare cases are the duplicate genes conserved. (Adapted from (Hahn 2009)).

1.4.1 Gene Loss or Psuedogenization

Psuedogenization is a process by which a functional duplicated gene becomes useless. It occurs a few million years after gene duplication if the gene is not selected. Genomes that undergo whole genome duplication have a higher proportion of psuedogenes that are duplications on a smaller scale (Krakauer and Nowak 1999; Zhang *et al.* 2003). It is clear from various genomic studies that a number of genes undergo non functionalization after duplication. For example 80 % genes were lost in 80 My after duplication in yeast (Kellis *et al.* 2004) and 70% genes in Arabidopsis 86 My after duplication (Bowers *et al.* 2003) . In teleosts fishes, gene loss has been estimated to be 85% (Brunet *et al.* 2006) . *C.elegans* has 2158 psuedogenes as

revealed by genomic analysis (Rubin *et al.* 2000; Gu *et al.* 2002). Gene loss is also responsible for the physiological changes. For example, 34 genes from 29 gene families in non legumes species like *Arabidopsis* are lost in legume species. The loss of these genes in the legume species is somehow responsible for the evolution of nitrogen-fixation in legume species (Gu *et al.* 2016). There are also examples where duplicated genes have been selected and maintained in the genome and only recently being psuedogenized like the olfactory receptor families in humans and mice have over over 1000 members but 60% percent of them are psuedogenes in humans as compared to 20% in mice which suggests that after evolution of humans many olfactory receptor genes have been rendered useless (Rouquier *et al.* 2000). Pseudogenes can also be revived as seen in cows, a paralog of pancreatic ribonuclease gene expressed in semen, seminal ribonuclease gene is expressed .These genes are a result of gene duplication that occurred before ruminants diverged 35Mya. In all ruminanats, this gene is not expressed, but it is expressed in cows, for reasons unknown .

1.4.2 Neofunctionalization

One of the most important outcomes of gene duplication is generation of gene novelties. This concept was coined by Ohno in his book *Evolution by gene Duplication* (Ohno 1970). However, the term neofunctionalization was coined by (Force *et al.* 1999b). Ohno stated that as gene duplication creates extra copies of the gene, one copy is sufficient for the function and the other copies are free from the selection pressure and can accumulate nucleotide substitutions (Ohno 1970). These substitutions can also occur in the regulatory region of the gene which can change the expression of the gene and result in a new function . For example during teleost whole genome duplication roughly 300Mya the whole family of sodium channels was duplicated and retained in the genome, totalling the number of sodium channel genes to 8 which expressed and functioned in the same way in the skeletal muscle of the teleost. After 100My a mutation in the regulatory region of one of the sodium channel genes *Scn4aa* affected its expression in the skeletal muscle. This event coincided with the emergence of electric organ in the teleost with expression of *Scn4aa* in it (Brodie 2010). In plants the C_4 photosynthesis is an interesting example for evolutionary novelty after gene duplication. Studies indicate that evolution of C_4 photosynthesis is a result of gene duplication and divergence of various components used by the C_3 plants. Changes in promoter regions in genes like *chloroplastic PPdk* and *NADPmesophyll*

after duplication allow these genes to function in the C_4 photosynthesis pathway (Monson 2003). Another example is duplication and divergence is of *RNase1* gene in the colobine monkeys . These monkeys feed on leaves while other monkeys feed on insects and fruits. This was achieved by several amino acid substituitions in the duplicate copy of *RNase1* gene which allowed these monkeys to eat leaves and adapt to a new nutritional niche (Taylor and Raes 2004). Neofunctionalization of transcription factors can allow evolutionary novelty at mulitple levels as changes in transcription factors affect their downstream targets. KRAB zinc-finger transcription factor genes belong to the class of genes which have undergone duplication during homonid evolution 35Mya. Analysis between parent daughter KRAB-ZNF gene pair shows that despite their recent duplication, there are vast differences between their structure, expression , splicing factors and functions. These genes being transcription factors have evolved into new functions that are responsible for defining regulatory pathways and novel evolutionary diversity in primates (Nowick *et al.* 2010). It has also been seen that evolutionary innovation after gene duplication is not always for the betterment of the species. In humans around 1700 unique genes are associated with over 3000 diseases . Studies show that about 80% of these genes are a result of gene duplication and neofunctionalization (Dickerson and Robertson 2012).

For a gene to evolve into a new function a vast number of nucleotide substitutions are required and if these substitutions are not deleterious (in which case the gene becomes a pseudogene) they are selected and fixed in the genome (Zhang *et al.* 2003; Conant and Wolfe 2008; Innan and Kondrashov 2010). The main question in the field of neofunctionalization is how are these nucleotide substitutions selected. Are all these substitutions positively selected since the duplicate gene is free of any constraints of natural selections to continue the original function. There are two models for neofunctionalizaton which explain as to how these substitutions are retained and selected in the redundant gene locus. The first model, the Dykhuizen–Hartl model was discussed by Kimura in 1983 in his book *The neutral theory for molecular evolution* after they showed that four bacterial alleles for allozymes of 6-phosphogluconate dehydro- genase in Escherichia coli are selectively neutral or nearly neutral (Dykhuizen and Hartl 1980). They proposed that none of the substitutions in the redundant locus are selected for but accumulate due to drift and later when environmental changes happen these new mutations are selected if these are advantageous to the organism. According to Kimura these neutral mutations have a potential for selection under right conditions and at appropriate time.

The rate of non-synonymous substitutions might be higher because of the relaxed constraints but will not be more than that of synonymous substitutions (Kimura 1983).

The second model which is the Adaptation model in which the substitutions are fixed at the redundant locus. After gene duplication a few neutral substitutions create a new but weak function in the redundant locus and then the positive selection accelerates the fixation of these substitutions (Zhang *et al.* 1998). Several cases of positive selection after gene duplication are reported. For example two *RNase A* superfamily genes *eosinophila derived neurotoxin* (*EDN*) and *eosinophil cationic protein* (*ECP*) were generated during gene duplication when hominoid and Old World monkey lineages were generated. After duplication *ECP* acquired a new anti-bacterial activity function. This function is absent in *EDN* in humans and *EDN* of New World monkeys and is not related to its ribonuclease activity. Molecular analysis shows that number of arginine residues increased after duplication which allowed ECP to be toxic to bacteria because it makes their cell membranes porous (Zhang *et al.* 1998).

Following gene duplication, the two duplicate gene may experience different functional restraints. One study involving 250 young human duplicates indicates that after duplication the duplicates evolve at different rates. One copy evolves faster and accumulate substitutions evenly at all sites while the other copy evolves slowly and accumulates substitutions unevenly on all sites (Zhang *et al.* 2003). Another analysis using data from 26 bacterial, six archael and seven eukaryotic genomes indicate that the ratio of nonsynonymous to synonymous substitutions ($Kn/Ks<<1$) is similar in most parologous pairs and they evolve at similar rates, with both the paralogs subject to purifying selection after duplication (Kondrashov *et al.* 2002). Whether the rate of evolution after gene duplication is because of positive selection or because of relaxed constraints on the duplicates is an open question in this research area.

1.4.3 Gene conservation or Redundancy

Ohno proposed gene conservation as " duplication for the sake of producing more". Immediately after gene duplication both the genes are completely redundant however, since complete redundancy does not last long in biological systems for a long time and due to many evolutionary constraints the redundant gene is either pseudogenized or diverged. Recent

evidences suggest that redudancy at times is maintained stably throughout evolution for example in *C.elegans* 14 duplicate gene pairs were found to be conserved for over 80 My (Tischler *et al.* 2006). Quantitative analysis of proportion of redundant gene duplicates is 25% in *S.cerevisae* and 7% in *C.elegans* (Zhang *et al.* 2003) . Also, the fact that many a times knockdown of many genes in various species has led to observation of no phenotype suggests that genetic redundancy does exists and is infact maintained in the species. For example *Myo D* and *Myf-5* are master regulators that specify myogenic fate of somites in skeletal muscle development. Null mutations of either of one genes results in normal skeletal muscle development while the knockout mice for both the genes together lacks skeletal muscles and dies immediately after birth . There are mechanisms by which genes can be retained as redundant and maintained in the system (Ohno 1970; Zhang *et al.* 2003; Hahn 2009).

Redundancy provides robustness in the system by providing a backup (Wagner 1999). Redudancy is maintained in the system by various ways. Strong purifying selection against mutations which can prevent the redundant genes from diverging (Nei 1969). The most common example is retention of genes that increase the gene dosage in the organism. Relationship between gene duplication and gene dosage is simple, as long as the increase in gene dosage is beneficial to the organisms the duplicates post gene duplication will be positively selected for (Innan and Kondrashov 2010). This can be applied to two kinds of genes. First the stress response genes, transport genes and genes with metabolic functions (Ohno 1970) . Second are the genes whose products are required in high doses like ribosomal and histone genes which are needed in translationally intensive stages especially during development (Sugino and Innan 2006). Another example is that of human salivary *AMY1* gene (*amylase1*) which are present in high copy number in individuals who have starch-rich diet. *AMY1* gene duplicates under positive selection are favored in high numbers and fit in with the dosage model (Hahn 2009). Another example is that of anti-microbial peptides in *Drosophila* which evolve rapidly in copy number (Heger and Ponting 2007). Sloppy paired genes *slp1* and *slp2* in *Drosophila* are functionally redundant and biochemically equilvalent. Removal of any one gene does not affect the embryogenesis and removal of both affects the development of embryo. However, there are slight differences by which these genes are expressed spatio-temporally (Cadigan *et al.* 1994)

Genetic Redundancy means more than one gene is performing the same function and if one of the genes is removed it would have no effect on the biological phenotype. But if a gene is truly redundant then it would be not be under selection and would accumulate deletrious mutations. However , genetic redudancy does exist in biological system as described above. In 1997 Nowak described four models for maintenance of genetic redundancy .

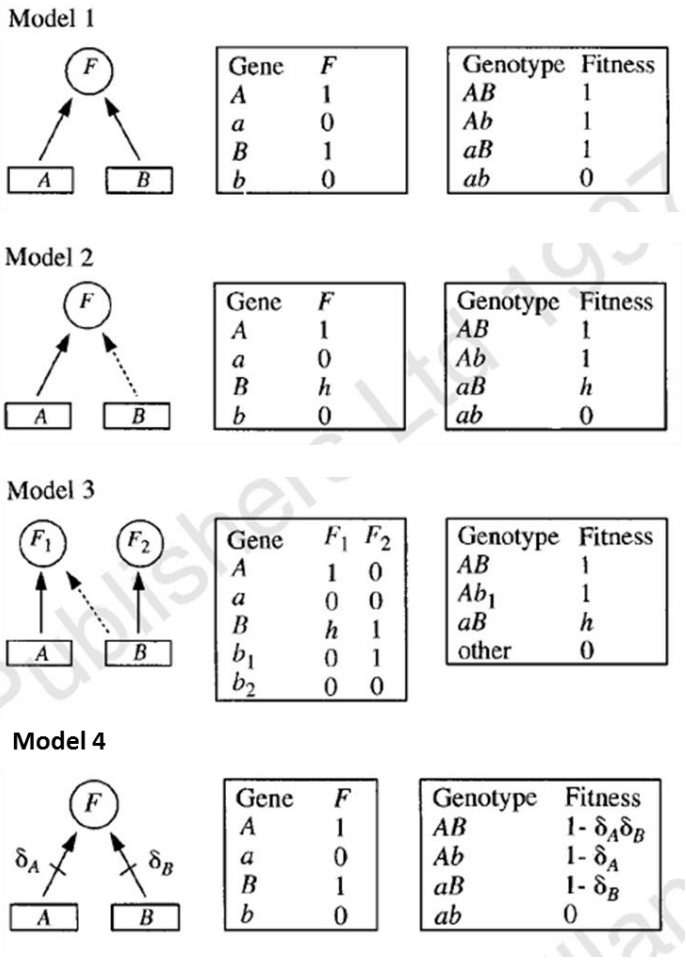


Figure 1. 6. Models of genetic redundancy Model 1) Both the genes A and B perform the function with equal efficacy. Model 2) Both genes A and B perform the same function B does it with a lower efficacy. Model 3) when B performs two functions F1 and F2 but F1 with lower efficacy than A. Model 4) Both genes A and B perform same function but have different mutation rates and different developmental error rates. a and b represent null of A and B. (Adapted from (Nowak *et al.* 1997)).

In model 1, it is assumed that both genes A and B perform the same function with equal efficacy and each gene can function on its own. The fitness of the population with AB, aB and Ab genotypes is 1, while ab is 0. However this is not a evolutionary stable model but if mutation rates between the two genes is smaller, then it will take a long time to eliminate any one of them

(Figure 1.6) (Nowak *et al.* 1997). Model 2 assumes that both the genes A and B perform the same function F1but with unequal efficacy h<1 . The fitness will be 1 in AB and Ab while it will be less for aB and 0 in ab. This leads to equilibrium between the two genes and both the genes are maintained in the system. Model 3 describes pleiotropy and redudancy together. Pleiotropy means one gene performing more than 1 function. In this model, A performs one function F1 and B performs F1 and F2 but it performs F1 with lower efficacy h<1. In this way both the genes are maintained in the population. Thus genetic redudancy is maintained as a result of functional overlap (Figure 1.6) (Nowak *et al.* 1997; Krakauer and Nowak 1999)(Nowak *et al.* 1997) (Krakauer and Nowak 1999).

The first three model assume that redudancy is fixed but Wagner introduced the idea of error buffering function of redudancy. He stated that redundancy is dynamic and can increase when mutation rates are high and lowers when mutation rates are lowered then the genetic duplicates would be maintained in the system (Wagner 1999). Thus, simply put genetic redundancy varies as the environmental pressure and selection differs. As a result Model 4 is based on both genes A and B having different developmental errors and different mutation rates allowing both to be maintained in the population (Figure 1.6)(Nowak *et al.* 1997; Krakauer and Nowak 1999).

1.4.4 Subfunctionalization or Gene Sharing

Subfunctionalization has been defined as division of functions of the parent gene among the duplicated loci. Several genes perform multiple and different functions with different selection pressures on these functions division of these functions would be considered optimal (Ohno 1970; Force *et al.* 1999a). This method of subfunctionalization has been described as duplication, degeneration and complementation or DDC model of subfunctionalization (Force *et al.* 1999b). In this model duplicate genes can be maintained in the genome due to mutations in both duplicates that can remove different subsets of original functions. These mutations are degenerate mutations and reduce the efficacy of both the genes to perform the original function. Once these mutations are fixed in the genome neither of the copy is sufficient to perform the original function by itself and, as a result, both the genes must be maintained by selection. These mutations can occur either in the regulatory regions (Figure 1.7) which will change the

expression profile of these genes or in the coding region which will divide the function between the two copies (Force *et al.* 1999a; Kondrashov *et al.* 2002; Hahn 2009).

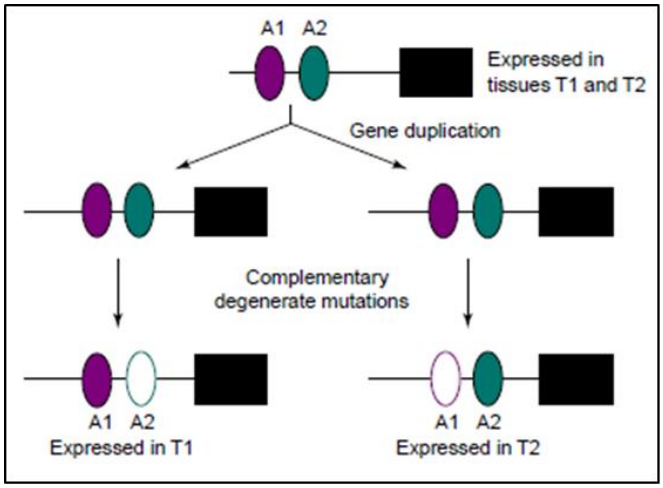


Figure 1.7. DDC Model of subfunctionalization After Gene duplication degenerate mutations accumulate in different regulatory regions of the duplicate genes which ensures that they are expressed in different tissues. Both genes will be selected and maintained in the genome. Figure Adapted from (Force *et al.* 1999b).

DDC model can be described as qualitative or quantitative. In the qualitative model the original gene function can be sub-divided into two or more distinct functions. This division can be protein based or expression based. This will ensure that the two duplicated genes are complemetary for the original function. In the quantiative model or gene dosage requirement expression levels of the duplicate genes plays a role.After gene duplication, both the duplicate genes have lost their ability to individually express enough to carry out the original function. The original function requires expression enough from both the genes, and if one is removed it would partially affect the function of the original gene. When both are removed the effect would be additive and and the phenotype will be similar to that of the removal of the ancestral gene.

DDC model of subfunctionalization can be applied to a number of duplicated genes. Several genes have been observed to evolve in this manner. For example zebrafish transcription factors, *engrailed-1* and *engrailed-2* are a pair of duplicate genes which emerged from chromosomal segmental duplication in ray finned fish lineage. *engrailed-1* is expressed in pectoral appendage bud while *engrailed-2* is expressed in hindbrain/spinal cord. There is only one mouse ortholog for both the genes *engrailed-1* is expressed both in the hindbrain and in the pectoral appendage bud as reviewed in (Zhang 2003). Other examples are the *zag1* and *zmm2*

genes in maize are a result of duplication by allotertaploidization, that occured 11Mya. Both the genes are orthologues to a single copy of floral homeotic gene *AGAMOUS* in *Arabidopsis* which is expressed in carpels and stamens. In maize *zag1* is expressed in high levels during carpel development but very low levels in stamen primordia. *zmn2* on the other hand is expressed in very low levels in carpel but high levels during stamen development; reviewed in (Force *et al.* 1999b). Another example is that of the Hox genes which are a result of tandem duplication of a ProtoHox gene. *Hoxa1* and *Hoxb1*in mice together pattern the anterior derivatives of mesodermal and ectodermal origins in the embryo. *Hoxa1* is important for segment identity in rhombomere5 and development of glossopharyngeal nerve. *Hoxb1* specifies rhombomere4 and development of facial nerve. The ancestral gene from pufferfish and zebrafish from which both the mouse genes have originated had regulatory elements to specify both the r4 and r5 and development of both the nerves; reviewed in (Taylor and Raes 2004).

Hughes proposed another model for subfunctionalization of redundant genes. It is called specialization or escape from adaptive model (EAC) (Conant and Wolfe 2008) . It proposes that after gene duplication assuming the genetic drift is fixed the duplicate copies are positively selected inorder to improve upon the two functions the original gene was performing. The improvement can be in the expression patterns of the original gene or on the secondary function the gene was performing (Oakley *et al.* 2006). Enzymatic proteins called crystallins make up 60% of total protein in the lens of eyes in squids and vertebrates. There are four crystallins *alpha* (belong to heatshock superfamily) *beta, gamma* (belong to calcium binding proteins) and *eta* crystallin, which is a lactate dehydrogenase. Alpha , beta and gamma genes acquired new promoter regions after duplication and underwent divergence and specialization . *eta,* on the other hand is still a functional lactate dehydrogenase and has dual function. This phenomenon is described as gene sharing and specialization (Piatigorsky *et al.* 1988).

Though there are multiple models that explain the fate of duplicate genes after gene duplication, biological systems being complex and dynamic it is not as strightforward to explain the fate of duplicate genes with just a single model.

1.5 MADF-BESS family in *Drosophila melanogaster* as a model to study gene duplication

The MADF-BESS gene family in *Drosophila melanogaster* consists of 16 transcriptional regulators (Figure 1.8), coded by 16 discrete genes. Proteins coded by all 16 members contain an N-terminal Myb-SANT like in Adf (MADF) followed by a C-terminal BEAF, Su-Var (3–7), Stonewall like (BESS) domain (Bhaskar and Courey 2002). Out of these 16 proteins, most range in size from 200 to 500 amino acids. The MADF-BESS family is also in a broader sense a subgroup of the individual, independent MADF and BESS family genes, where both MADF and BESS domains are together, coded by a single continuous polypeptide. The architecture of the domains, sequence identity, and the lack of any additional defined domains in the polypeptide sequences suggest that the MADF-BESS family members may have similar or identical function. Very few genes have actually been studied; functions of most of the members remain unknown (Figure 1.8).

Adf1, a well-studied gene in the family, has been implicated as a transcriptional activator that contains a TFII D binding domain and can interact with certain TAF subunits(England *et al.* 1992; Cutler *et al.* 1998). Dorsal interacting protein3 (Dlip3) appears to be a co-activator in Dorsal/Twist function. It binds to the Rel homology domain of both Dorsal and Relish and by doing so it acts as a co-activator in dorso-ventral patterning and immune response. In early embryos it is shown to localize to the heterochromatic pericentromeric region (Bhaskar and Courey 2002; Ratnaparkhi *et al.* 2008). Dip3 is also expressed in differentiating photoreceptors and is involved in regulating neuronal differentiation (Duong *et al.* 2009). Co-repressor of Pangolin (Coop) is an interactor of Pangolin and is a negative regulator of Wg/Wnt signaling (Song *et al.* 2010). *stwl* is required for germ-cell development (Clark and McKearin 1996) and germ stem-cell maintenance (Maines et al. 2007), by epigenetically repressing the genes that are required for the differentiation of germ stem cells. The MADF domain in Stwl is implicated in chromatin remodeling and histone modification (Boyer et al. 2002). Stonewall is also associated with heterochromatin and colocalizes with HP1 (Yi et al. 2009). The remaining 11 genes have not been named and are predicted to be transcriptional regulators with unknown functions.

The retention of these 16 functional genes in the *D. melanogaster* genome suggests important roles for this family. If the family members do indeed have similar functions, then functionally they may be completely or partially redundant. Since most genetic studies or even large-scale loss-of-function screens knock down only one gene at time, functional roles for

multiple, redundant genes supporting a single function tend to go undiscovered. In this study we examine redundancy in the MADF-BESS family using tools of phylogeny and genetics.

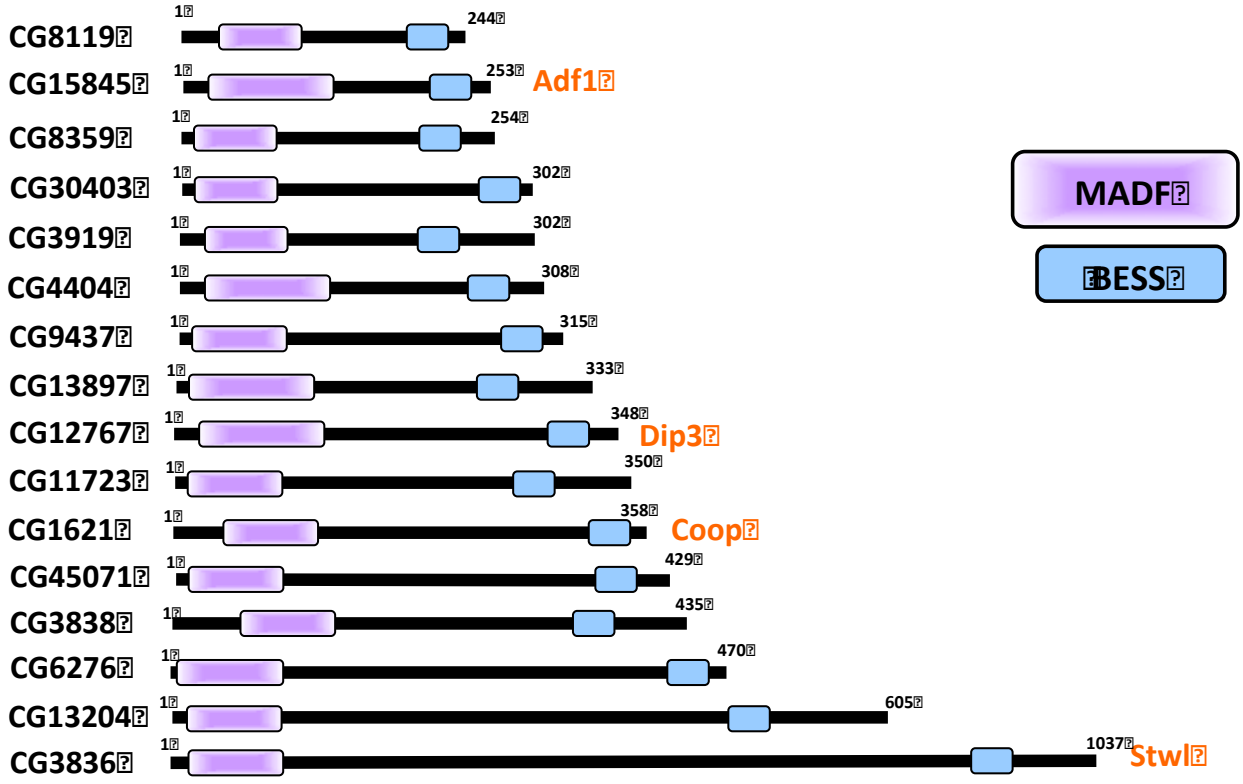


Figure 1.8 The *D. melanogaster* MADF-BESS family. Sixteen genetics loci code for 16 polypeptides that are possibly a consequence of gene duplication. The 16 proteins, represented in the figure, contain an N-terminal DNA-binding Myb/SANT like in Adf (MADF) domain and a C-terminal BEAF-32, Stonewall, Su (var) 3-7 homology (BESS) domain. MADF and BESS domains tend to be found together in a single polypeptide chain. Proteins labeled in orange have been characterized, while others had unknown functions at the start of this study.

1.6 Open Questions regarding the MADF-BESS family

In my thesis, I have attempted to address some of the open questions about the MADF-BESS family of proteins. In the paragraphs below, I have listed these questions, a few of which have been addressed in the work done for my Ph.D.

1. Have the MADF:BESS, MADF-only and BESS-only genes evolved from a common ancestor? Specifically, are these genes duplicates? Which domains of life contain these genes; are they present in all life forms?
2. If phylogenetic and bioinformatics analysis indeed suggests that these genes are duplicates; then can these genes be tested for divergence or functional redundancy by genetic tools available in *D. melanogaster*?
3. The biological roles for the few genes studied, namely *adf1, dlip3, stonewall* and *coop* suggest diverse unrelated functions of these genes. What functions do the 12 unknown proteins perform in the fly? Are any of these functions novel or do they overlap with known functions?
4. Are members of this family transcriptional regulators or co-regulators? Is there DNA binding specificity or do these proteins work by interacting with other transcriptional activators or repressors. Do these family members regulate chromatin packing and dynamics?
5. Can a reverse genetic screen with single or simultaneous, multiple gene knockouts allow us to dissect out partial or complete redundancy in this family?
6. Can one predict the function of the ancestral genes? Will studying all members of the family shed light on common roles for the entire family?
7. The few MADF: BESS proteins studied till date show centromeric localization. Is this a common theme for MADF: BESS or MADF or BESS domain containing proteins? Does this localization have a functional significance for the family?

CHAPTER 2

Gene Duplication and Lineage Specific Expansion of MADF and BESS family in *Drosophila*

2.1 Summary

This chapter examines the possibility of Gene Duplication and Lineage Specific Expansion for the MADF, BESS and MADF-BESS domain containing genes in *Drosophila*. Evidence is gathered by analysis of all DNA and protein sequences of genes that code for or contain these domains in animals. We find that the MADF, BESS and MADF-BESS families in flies appear to be a consequence of a gene duplication and expansion - specifically in the *Drosophila*-lineage. The genes that code for both the MADF and BESS domains (the MADF-BESS family) evolve similarly, when compared to only MADF or only BESS domain coding genes. Nucleotide substitution analysis also shows that evolution of BESS domain is more constrained and that of MADF is more relaxed.

2.2 Introduction

The initial analysis of the sequence of the human genome (Lander *et al.* 2001), amongst other findings reported on the duplications found in the human genome. For example, segmental duplications of 1-200 kb block of the human genome were found to be widespread. These included inter and intra chromosomal duplications and these were associated with other events. Another aspect of gene duplications was the finding that some protein domains were dramatically expanded in humans when compared to other invertebrates or invertebrates. For example, a correlation plot between human and fly genes (Figure 2.1A) taken from the study shows that some protein domains, such as IgG, Zn-finger, protein kinase and GPCR are present in the human in large numbers. Unlike in invertebrates, where IgG domains were utilized as cell surface proteins, in vertebrates their roles are expanded in immune surveillance and recognition of antigens. The human genome contains 30 Fibroblast Growth factors (FGF's) as compared to only two in flies.

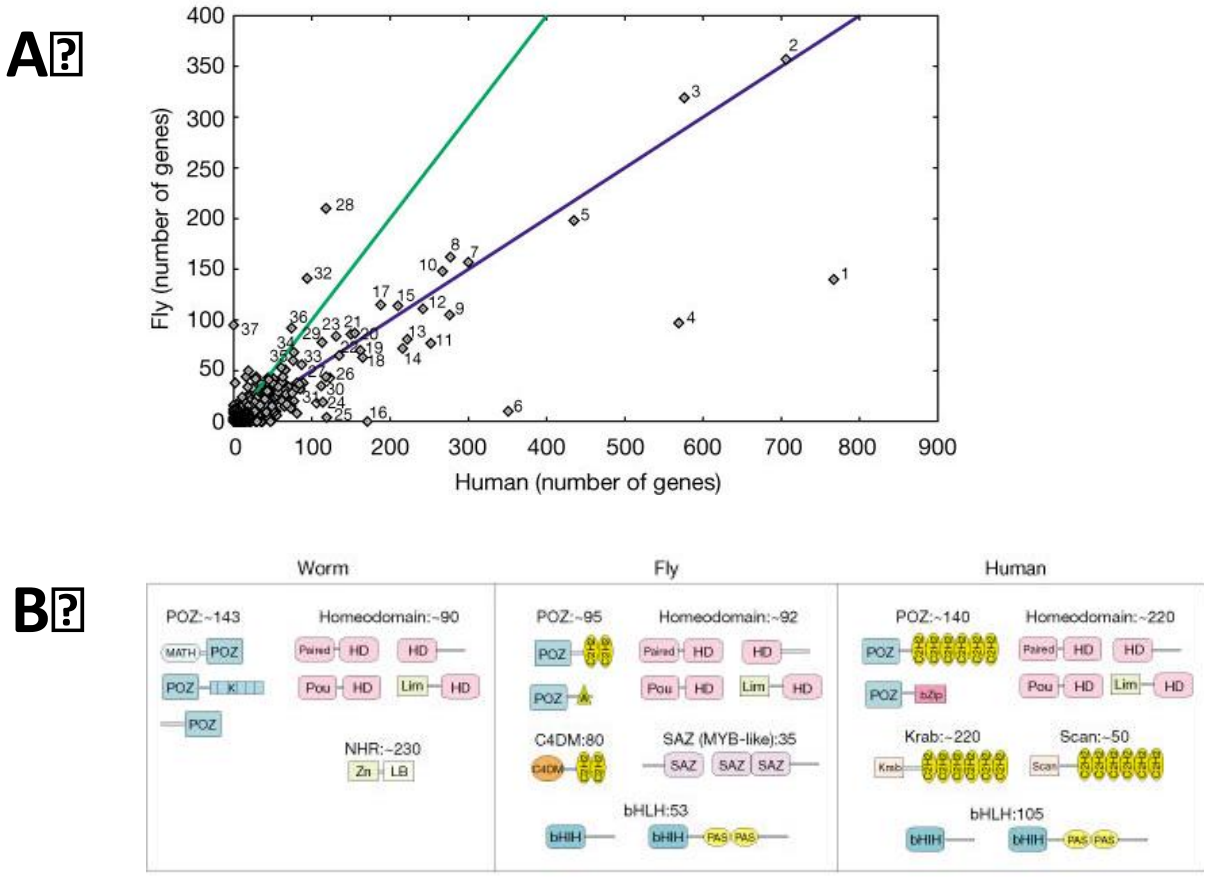


Figure 2.1. Expansion of genes coding for specific protein domains.
(A) Correlation plot between *Drosophila* and human for genes coding for specific domains that are expanded in large numbers in these genomes. The domain(s) with the greatest increase in humans include 1) immunoglobulin domain [IPR003006]; (2) zinc finger, C2H2 type [IPR000822]; (3) eukaryotic protein kinase [IPR000719] and (4) rhodopsin-like GPCR superfamily [IPR000276] whereas the (28) serine proteases, trypsin family [IPR001254] and (32) esterase/lipase/thioesterase [IPR000379] are expanded the most in flies. Blue line, equality between normalized family sizes in the two organisms. Green line, equality between unnormalized family sizes. The data is taken and adapted from (Lander *et al.* 2001).
(B) Expansion of protein domains that are involved in transcriptional regulation indicate that different architectures are expanded in different lineages. The homeodomain is expanded in all three lineages (worm, flies, human) but the SAZ (Myb-Like) domain is only expanded in flies. Approximate numbers of domains identified for the lineages are shown next to the domains. Domains: HD, homeodomain; Zn, zinc-binding domain; LB, ligand-binding domain; C4DM, novel Zn cluster with four cysteines, SAZ, specialized Myb-like helix-turn-helix (HTH); E2F, winged HTH DNA-binding domain; POZ, Pox zinc finger domain; HF, histone fold; ANK, ankyrin repeat; C2H2, classic zinc finger domain; WD, WD40 repeats. Reproduced and adapted from (Lander *et al.* 2001).

Vertebrates are neither unique nor alone in employing domain expansion for function. Many eukaryotic lineages show expansion of domains. For example for transcription factors, the C2H2 family of Zn-finger domains are expanded independently in yeast worm fly and human linages (Figure 2.1B). In humans a C2H2 with POZ/KRAB/SCAN domains are the most

prevalent combinations of expanded domains. Lineage specific expansion of transcription factors is of specific interest because they are critical in development and differentiation. In flies, one example of an expanded family is the SAZ family (Specialized Myb-like helix-turn-helix domain found in Stonewall, Adf1 and Zeste). In 2001, 35 domains had been counted in the recently sequenced *Drosophila melanogaster* genome. This family was later rechristened as the MADF (Myb-like in Adf1) family. The MADF domain is approximately 80 amino-acids long. The MADF domain coding proteins were found to associate with other domains, one of which was the BESS Domain found in BEAF, Stonewall and SuVar(3-7). The BESS domain is approximately 40 amino acids long.

In order to determine phylogenetic relationship between organisms using DNA sequences- a model of evolution is necessary. These are sets of assumptions about process of nucleotide substitutions. These models determine different probabilities of nucleotide substitutions and are important to understand the evolutionary process of DNA sequences(Yang and Goldman 1993; Strimmer and von Haeseler 2009). Last 50 years an array of models have been defined, each with increasing complexity about nucleotide substitutions. Models of nucleotide substitutions take several parameters into considerations: 1) Base frequency which defines number of times a nucleotide base is repeated in the compared sequences. 2) Rate of substitutions which determine the rate at which transitions and transversions occur among the nucleotides. 3) Invariable sites which determine the number of sites that never changes. 4) Rate heterogeneity estimates among-site variation. Different models are defined by different combinations of these parameters(Wakeley 1994; Yang 1996; Posada 2003; Jia *et al.* 2014). In this chapter, we use soft wares to predict model of nucleotide substitutions for all MADF and BESS genes in *Drosophila melanogaster*.

In this chapter, genes that encode MADF and BESS domain proteins are identified from known genome sequences, counted and analysed, as are their polypeptide sequences. We gather evidence to test if the MADF and BESS genes are duplicated and to quantitate the extent of duplication. Further, phylogenetic analysis and nucleotide substitution models are used to understand the genetic distances between the genes after duplication

2.3 Materials and Methods

2.3.1 Phylogenetics Analysis

The number of MADF, BESS, and MADF-BESS domains were counted in different species in the animal kingdom based on information available in databases such as InterPro (Hunter *et al.* 2012), UniProt (Magrane and Consortium 2011), Flybase (Marygold *et al.* 2013), Flymine (Lyne *et al.* 2007), SMART (Schultz *et al.* 1998), 12 Drosophila genomes (Genomes 2007), and ORTHODB (Waterhouse *et al.* 2013). cDNA and protein sequences for all the MADF and BESS genes were obtained from FlyBase and Flymine for Drosophila genes and Vectorbase (Megy *et al.* 2012) for genes from Aedes, Anopheles, and Culex species. Multiple sequence alignment of the genes was performed using MUSCLE (Edgar 2004). Dipteran phylogenetic Tree was constructed using RaxML 7.2.7 (Pfeiffer and Stamatakis 2010) at the CipRES Science Gateway (Miller *et al.* 2011) with default parameters for DNA. A total of 400 bootstrap iterations were performed and the best scoring maximum-likelihood tree was obtained. Similarly, a tree for the 16 genes corresponding to *D. melanogaster* was obtained by performing multiple sequence alignment with MUSCLE (Edgar 2004) and a maximum-likelihood tree with RaxML was obtained with the same parameters as for the dipteran MADF-BESS gene tree. The MADF and the BESS domain coding sequence from 48 MADF genes and 22 BESS genes in 12 *Drosophila* species was predicted using NCBI Conserved Domain (Marchler-Bauer *et al.* 2015). The sequences were aligned using MUSCLE in MEGA6.0(Kumar S, Stecher G 2013). These sequences were then concatenated to form a supergene alignment for both MADF and BESS genes using DAMBE (Xia 2013). Tree analysis and model selection analysis for both the sets was done on IQTree Webserver program (Trifinopoulos *et al.* 2016). Model selection analysis was done by creating partitions for the domain coding sequence and also for first, second and third position for every codon. The nucleotide substitution model predicted above was used to calculate the distance for each gene between all the 12 species using MEGA6.0 which was then used for divergence with time plots using Past3 statistical software. DAMBE was used for codon bias and codon usage analysis. Calculations for positive and negative selection for these genes were done by using MEGA6.0.

2.3.2 Expression Analysis

BDGP *in-situ* data (Tomancak *et al.* 2007)(Tomancak *et al.* 2002) was used for the collection of data for mRNA localization for all the MADF-BESS genes shown in Table 1 and Appendix II. Data from modENCODE consortium (Celniker *et al.* 2009) was used in Table 2, Appendix II.

2.3.3 Chromosomal Locations of all the MADF and BESS genes in *Drosophila melanogaster.*

The base map used in Figure 2.5 is from the Duncan lab web-page (http://willamette.edu/~jduncan/Duncan_Lab_-_Department_of_Biology_-_Willamette_University/Resources.html) and the map locations for all MADF and BESS are based on information in Flybase.

2.4 Results

2.4.1 Lineage-specific expansion of MADF and BESS domains

To determine if the MADF and BESS domains have expanded in a specific lineage, the numbers of individual MADF and BESS domains were counted in a few representative organisms where genome sequences were available. A single species was taken as representative for each genus, and the MADF and BESS domains, as defined in various databases, were counted. The analysis showed an overrepresentation of both MADF and BESS domains in invertebrates with an unusual increase in numbers for *D. melanogaster* (Figure 2.2). Analysis of the phylogeny of dipterans and of sequenced *Drosophila* species indicated that the expansion probably occurred in an ancestor common to the *Drosophila* lineage, 40 million years ago. From the InterPro database, we found that the MADF domain alone occurs in 908 curated proteins whereas MADF together with a BESS domain occurs in 353 proteins. The most frequent other combination was the presence of multiple MADF domains in the same protein.

All other domain combinations such as multiple MADF with a single BESS domain or vice versa were significantly less frequent. The BESS domain was thus found to be most strongly associated with the MADF domain, with 16 of the 22 BESS-domain-containing genes in *D. melanogaster* being exclusively associated with the MADF domain, with no other intervening domains in the linker region. The 353 proteins containing a MADF-BESS domain were also

analyzed on the basis of their distribution in different species. We found that 225 are present in dipterans (National Center for Biotechnology Information taxonomy ID 7147) and, furthermore, of these 225 proteins, 197 proteins are in drosophilids. As multiple proteins may arise from a single gene, we also analyzed the distribution of MADF-BESS domains at the gene level. We found that there are 168 MADF-BESS genes across drosophilids and a further 27 identified MADF-BESS genes from other dipterans. Lineage specific expansion (LSE) is defined as the proliferation of a specific protein family in a genera/species, relative to its sister lineage, with which it is compared (Genomes 2007) . LSE for the MADF-BESS family (Lespinet *et al.* 2002) in the *Drosophila* lineage was confirmed by phylogenetic analysis and by counting the number of MADF and BESS domain family members (InterPro). Of the 1576 MADF domains contained in the databases, 828 were found in Diptera. Furthermore, 755 of these were found in Drosophilidae. At the gene level, we found 576 genes in Drosophilidae containing a MADF domain. By doing a similar analysis for the BESS domain, we found that, of the 644 proteins containing a BESS domain, 388 are in Diptera and 356 of these are in Drosophilidae. At the gene level, we found 240 genes in Drosophilidae with a BESS domain. As mentioned above, among these, 225 proteins in Diptera and 197 proteins in Drosophilidae have an N-terminal MADF and a C-terminal BESS domain with no other domains in the intervening sequence. The number of MADF-BESS genes in the 12 sequenced *Drosophila* species ranges from 13 to 16 except in the case of *D. simulans*, which has only 11 MADF-BESS genes. The smaller number of genes in *D. simulans* may reflect genes that have lost the MADF or BESS domain or incomplete annotation that would have eliminated them from this study (Figure 2.2C). *Aedes aegypti* and *Culex quinquefasciatus* contain 10 and 9 MADF-BESS genes, respectively, and *Anopheles gambiae* and *darlingi* contain only 2 genes each. *Glossina morsitans* (tsetse fly) contains three MADF-BESS genes.

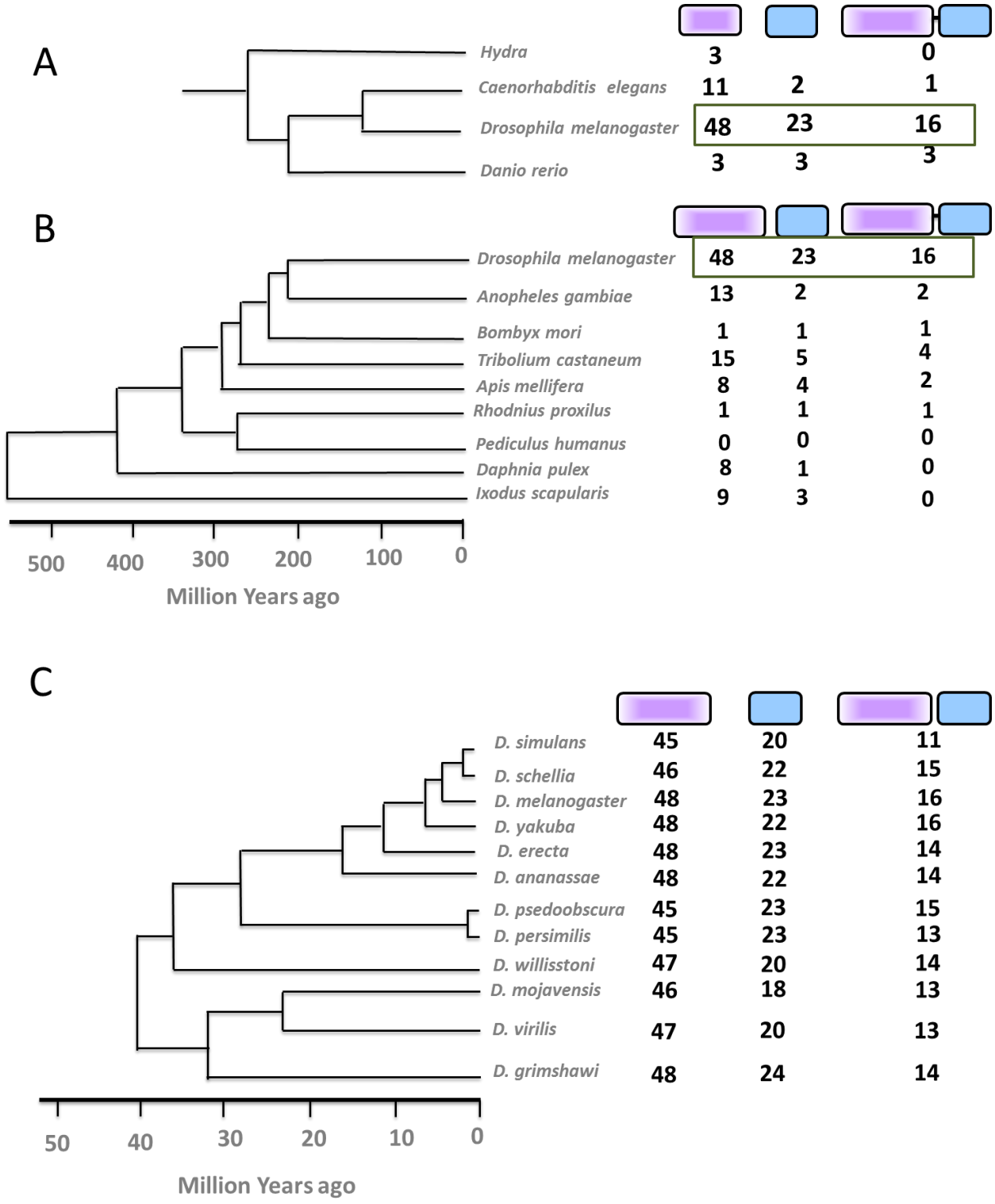


Figure 2.2. MADF and BESS domains are expanded specifically in the *Drosophila* lineage. **(A)** MADF and BESS protein domains, when counted in representative members of the animal kingdom, indicate expansion in *Drosophila* lineage. **(B)** In arthropods, the number of individual MADF and BESS

domains coded by the *Drosophila* genome is higher than in other sequenced species. The dipteran group is marked. **(C)** The expansion is dramatic in all members of the *Drosophila* lineage, confirming that that the expansion occurred in a common drosophilid ancestor.

The concatenated phylogenetic tree consisting of 576 MADF genes in 12 *Drosophila* species (Figure 2.3 A) is similar to the *Drosophila* species tree based on the whole *Drosophila* genome (Genomes 2007). We also see the same with 240 BESS domain containing genes in 12 *Drosophila* species Figure 2.3 B). This indicates that MADF genes and BESS genes duplicated and expanded as *Drosophila* speciation took place 40 million years ago.

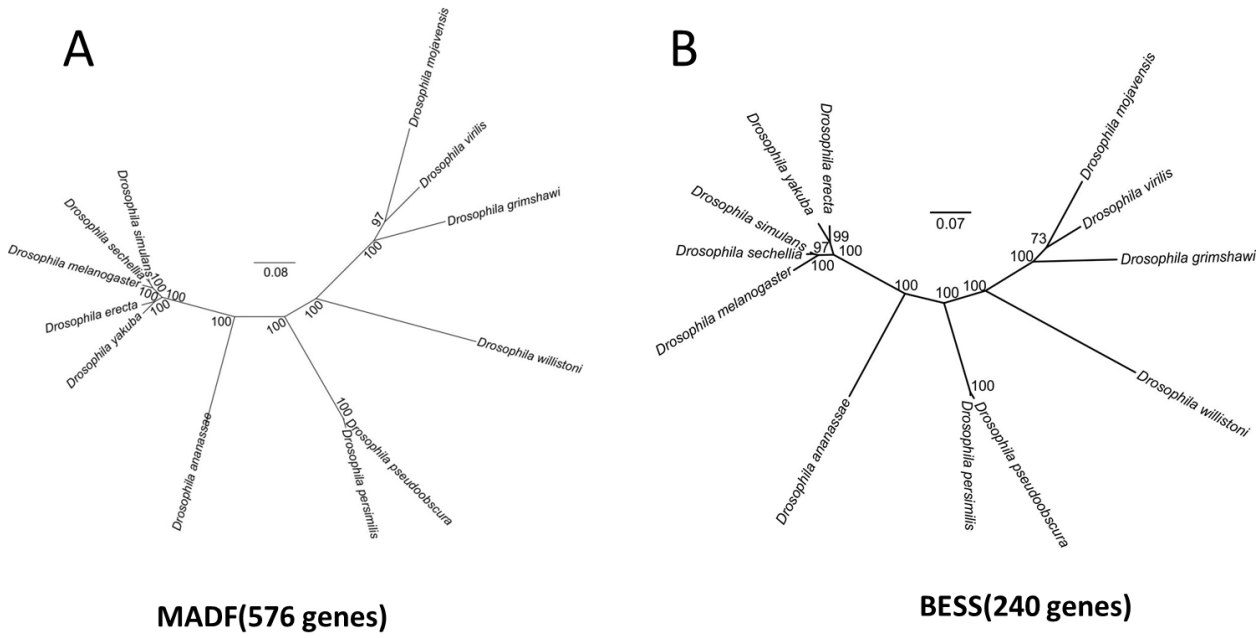


Figure 2.3. **Phylogenetic Trees for MADF and BESS domains in Sequenced fly genomes**. The Maximum-likelihood phylogenetic tree of all **A)** concatenated MADF domain and **B)** concatenated BESS domain containing genes in 12 *Drosophila* species using IQtree program.

2.4.2 The MADF-BESS family

A striking feature of the MADF-BESS family of proteins is their conserved protein architecture: all members have an N-terminal MADF domain and a C-terminal BESS domain (Figure 2.4 A). The proteins also show sequence similarity and/or identity for the two domains, indicating that the proteins have common evolutionary origins. Between the MADF and BESS domains is a "linker" region that ranges from 100 to 600 amino acids. The roles of the linker

polypeptide sequences, if any, are unknown. A phylogenetic tree with all members of the MADF-BESS family in *D.melanogaster* shows that these genes are related (Figure 2.4 B), suggesting a gene duplication and/or expansion event followed by maintenance of these seemingly redundant genes in the animal.

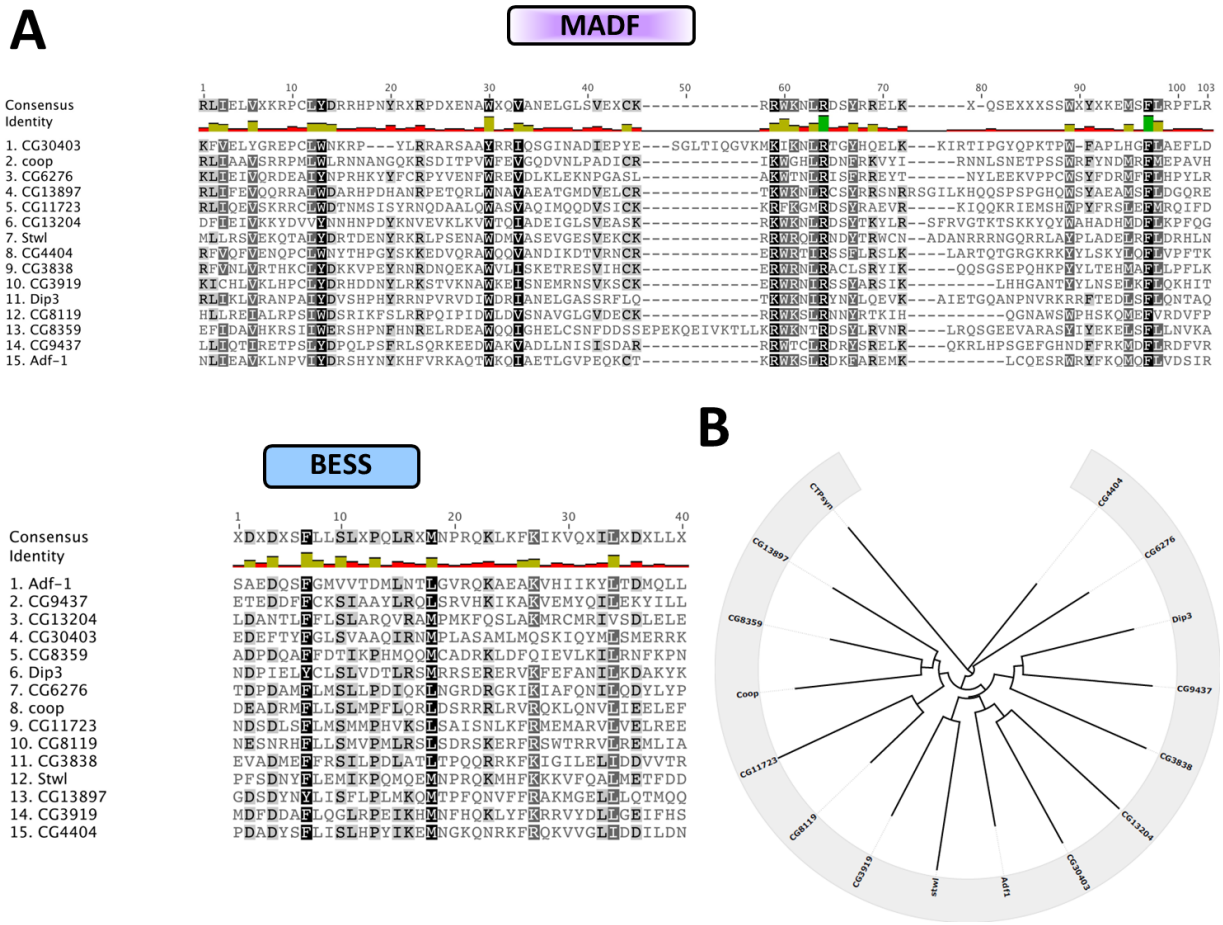


Figure 2.4. Sequence Alignment and phylogenetic tree for the 16 MADF-BESS genes in *Drosophila melanogaster.* A. Sequence alignment of MADF and BESS coding region shows strong conservation. B. Phylogenetic tree of MADF-BESS genes in *Drosophila melanogaster* showing relationship between the sixteen paralogs.

An expanded phylogenetic tree of all MADF-BESS-containing genes in dipterans showed that the orthologs of the MADF-BESS genes in drosophilids cluster together (Figure 2.5) as would be expected if the gene duplication had happened before *Drosophila* speciation and the genes had diverged from each other. In contrast, the MADF-BESS genes in *Culex* and *Anopheles* did not always have equivalent counterparts in drosophilids: their genes clustered separately with each other and were distant from the MADF-BESS genes in drosophilids. From the phylogenetic

tree we can see that despite *Culex* and *Aedes* (Culicinae) having a comparable number of genes to *Drosophila* species, the genes from Culicinae are more closely related between themselves than to genes in Drosophilidae (Figure 2.5). This indicates that Duplication/expansion of genes in Culicinae was separate from that in Drosophilidae lineage. Three major clusters for MADF-BESS genes were observed in *Culex* and *Anopheles*. The first cluster was close to CG8119; the second was near CG4404, and the third was close to CG13204 (Figure 2.5). A small cluster was seen close to CG3838. Thus, the genes in Culex and Anopheles were closer to each other than to genes in other drosophilids. This indicates that gene duplication and expansion in drosophilids was separate from that in other dipterans. Phylogenetic analyses also showed that the MADF-BESS containing genes are highly conserved as evident from the distances on the phylogenetic tree within the "melanogaster group" [the melanogaster group contains *Drosophila simulans (GD), Drosophila yakuba (GE), Drosophila ananassae (GF), Drosophila erecta (GG), Drosophila sechellia (GM),* and *Drosophila melanogaster (CG)]*. For example, in Figure 2.6, the blue asterisk marks strongly conserved sequences for Adf1, hinge1, and Coop where the branches are short and cluster near each other. In comparison, CG8119 and CG4404 DNA sequences do not cluster and are not as strongly conserved (Figure 2.5).

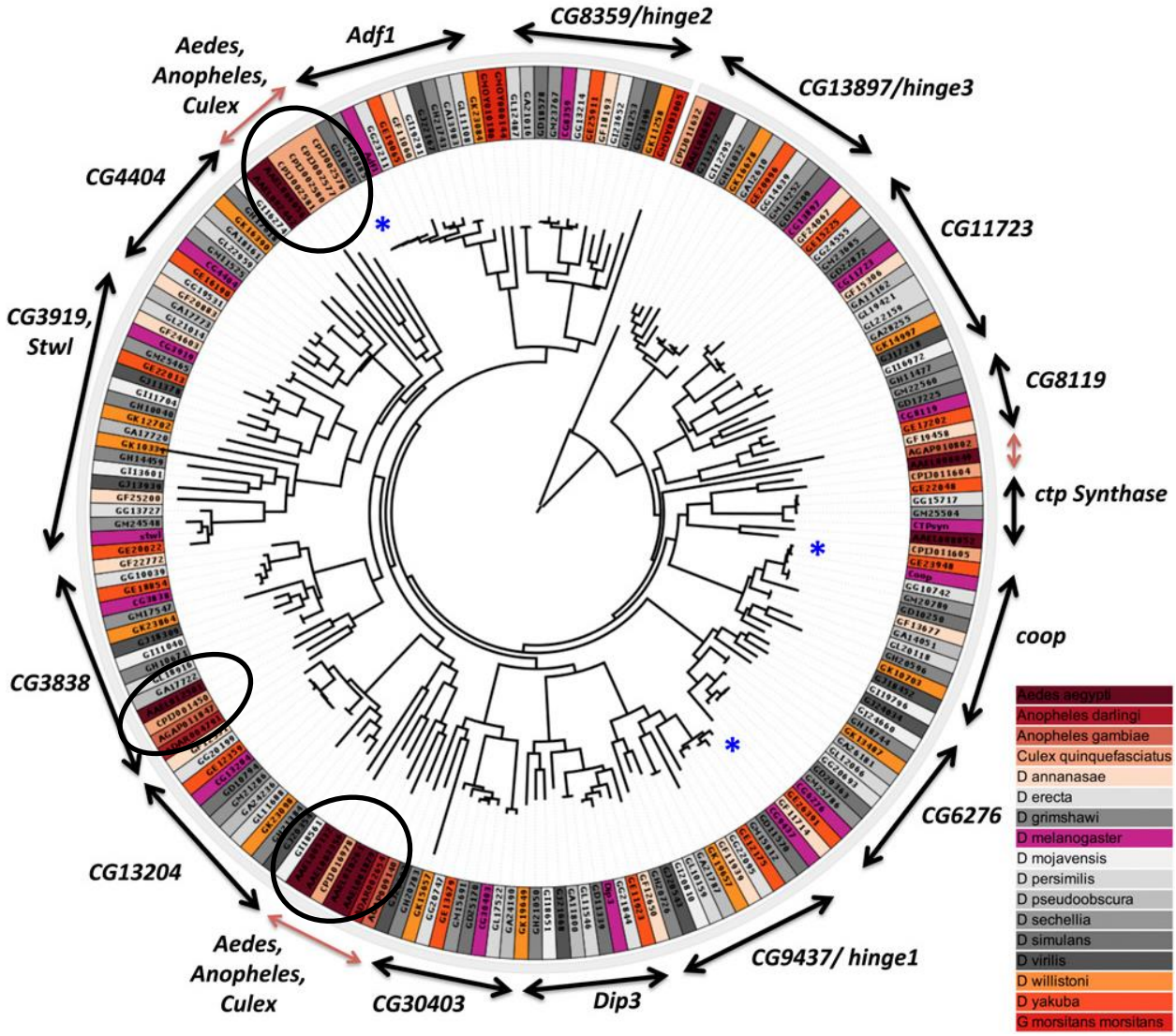


Figure 2.5: Maximum-likelihood phylogenetic tree of all MADF-BESS genes in dipterans. Branch lengths are proportional to mean substitutions per site. Orthologs of the 16 MADF-BESS genes cluster separately with the corresponding *D. melanogaster* gene (labeled on two-sided arrows), indicating that the duplication in the family occurred before the divergence of Drosophilids. MADF-BESS genes from *Culex*, *Anopheles*, and *Glossina morsitans* have fewer orthologs and cluster separately (brown arrow) for the most part, indicating that, in their case, expansion in MADF-BESS was independent from that in Drosophilids. The genes for different sequenced dipterans are color-coded to bring out this feature. The blue asterisk marks the genes that show short branch lengths and thus minimal sequence divergence.

2.4.3 A mechanism for Gene Duplication and expansion of MADF and BESS domains?

In the *D. melanogaster* genome most members of MADF and BESS genes are present in distant locations on multiple chromosomes, pointing to a mechanism for duplication and expansion that is not via a localized duplication as seen for the bithorax complex (Lewis 1978) but instead a possible RNA or transposon-mediated duplication event (Casola *et al.* 2007). In Figure 2.6 we see that there only 3 clusters where genes are right next to each other which might be a result of local duplication. First cluster comprising of *CG1602*, *CG1603* and *az2* on 2R chromosome are all MADF genes containing 2 MADF domains. The second cluster on 3L chromosome comprises of MADF-BESS genes *stwl* and *CG3919*. *Ravus* and *su(var)3-9* make up the third cluster on 3R. This clearly indicates that duplication might have occurred via transposon elements.

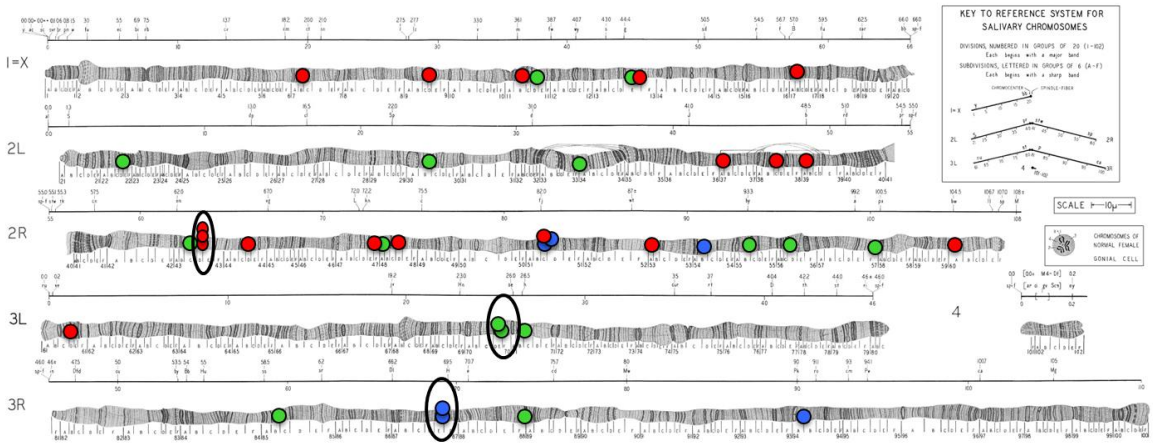


Figure 2.6. Cytological positions of MADF, BESS and MADF-BESS genes. These genes are cytologically positioned all over the genome except in three cases where few of them cluster together. This suggests gene duplication might have occurred via transposon elements (TEs).Red circles represent non-BESS MADF genes. Blue the non-MADF BESS genes and green MADF-BESS genes. Adapted from EM map generated by the Duncan Lab, Willamette University.

2.4.4 Expression Analysis of MADF and BESS genes

The compilation of the expression levels of MADF and BESS genes of *Drosophila melanogaster* (Table 2.1) from the modENCODE tissue expression data and FlyAtlas anatomical

expression data shows that most of these genes are present in moderate levels in the early embryo. Brain, eye, gut and fat body show low levels of most of these genes. On the other hand ovary and testis shows moderate levels of these genes. Interestingly, none of these genes shows very high expression levels in any tissue of *Drosophila melanogaster*.

Table 2.1. Expression Profile of MADF and BESS genes in *Drosophila melanogaster*. Consolidated Expression Profile of MADF transcripts (1-32), MADF-BESS transcripts (33-48) and BESS transcripts (49-54) collated from FlyAtlas Anatomy Microarray data and modENCODE Tissue Expression Data.

S.No	Gene	Embryo (0-2 Hr)	Brain	Eye	Gut	Fat body	Ovary	Testis
1	*ADD1*	Mod High	Low	Low	Low	Low	High	Low
2	*az2*	Mod	Low	Low	Low	Low	Mod	Low
3	*CG1602*	Mod	Low	Low	Low	Low	Low	Low
4	*CG1603*	Mod	Low	Low	Low	Low	Mod	Low
5	*CG3386*	Low	Low	Mod	Low	Low	Low	Low
6	*CG4004*	Low	Low	Low	Low	Low	Low	No
7	*CG5180*	High	Mod	Low	Low	Low	Mod	Mod
8	*CG5953*	Mod	Low	Low	Mod	Low	Mod	Low
9	*CG6163*	Low	Low	NA	Low	Low	Mod	Mod
10	*CG6175*	Mod	Low	Low	Low	Low	Low	No
11	*CG6683*	Mod	Low	Low	Low	NA	Mod	Low
12	*CG7745*	Mod High	Mod	Low	Low	Low	Mod	Low
13	*CG8281*	Low	Low	Low	Low	Low	Low	Low
14	*CG8765*	Mod High	Mod	Mod	Low	Mod	Mod	Low
15	*CG8944*	Mod High	Low	Low	Low	Low	Low	No
16	*CG9948*	Mod	Low	Low	Low	Low	Low	Low
17	*CG10151*	Low	Mod	Low	No	No	No	Mod
18	*CG10904*	Mod	Mod	Mod	Low	Low	Mod	Low
19	*CG10949*	Mod	Low	Low	Low	Low	Mod	Low

20	*CG11504*	High	Mod	Low	Low	NA	Mod	Low
21	*CG12155*	High	Mod	Low	Low	Low	Mod	Low
22	*CG12609*	No	No	No	No	No	No	Mod
23	*CG12768*	No	Mod	NA	No	No	No	No
24	*CG15601*	Low	Low	Low	Low	No	Low	Low
25	*CG31627*	No	Low	No	No	Low	High	NA
26	*CG33017*	No	No	No	No	No	No	Mod
27	*CG42526*	Low	NA	NA	NA	NA	Low	No
28	*hmr*	Mod High	Low	Low	Low	Low	Mod	Low
29	*jigr-1*	High	Mod	Mod	Mod	Low	High	Low
30	*l(3) j2D3*	Mod High	Mod	Low	Low	Low	Mod	Low
31	*mes2*	Mod High	Mod	Low	Low	High	Mod	No
32	*rgr*	Mod	NA	NA	NA	NA	Mod	Low
33	*adf-1*	High	High	Low	Low	Low	High	Low
34	*brwl*	Mod	Low	Low	Low	Low	Mod	Mod
35	*CG3919*	Low	no	no	no	no	Low	Mod
36	*CG4404*	High	Mod	Mod	Mod	Mod	High	Low
37	*CG6276*	High	Mod	Mod	Mod	Mod	Mod	Low
38	*CG8119*	NA	No	No	No	No	No	Low
39	*CG11723*	Mod	Mod	Mod	Mod	Mod	Mod	Mod
40	*CG13204*	V Low	Mod	Low	Low	No	Low	No
41	*CG30403*	Mod	NA	No	No	No	No	NA
42	*CG45071*	Mod High	NA	NA	NA	NA	Very high	Mod
43	*coop*	Low	NA	NA	NA	NA	Low	Low
44	*dlip3*	Mod	Mod	Low	No	No	Low	Low

45	*hng1*	Mod High	Low	Low	Low	Low	Mod	Mod
46	*hng2*	Mod	Low	Low	Low	Low	Mod	Mod
47	*hng3*	Mod High	Low	Low	Low	Low	Mod	Low
48	*stwl*	Mod High	Low	No	Low	No	Mod	Low
49	*BEAF-32*	Very high	Mod	Mod	Low	Low	High	Low
50	*CG10209*	Mod	Mod	Mod	Mod	Low	Mod	Low
51	*CG34149*	Low	NA	NA	NA	NA	Mod	Low
52	*lhr*	Low	Low	Low	NA	No	Mod	Low
53	*ravus*	Mod	Low	Low	Low	NA	Mod	Low
54	*su(var)3-9*	Mod High	NA	NA	NA	NA	Mod	Low

Note: NA, Not Avaliable. Mod High, Moderately High.

2.4.5 MADF and BESS domains have different evolutionary restraints.

To understand the way these genes are evolving IQ Tree program was used to predict the nucleotide substitution model for the MADF domain coding regions for all the 54 genes in *Drosophila melanogaster*. Partitions were made in 2 ways: one to analyse the complete MADF coding region and second to analyse every 1st 2nd and 3rd codon position in the complete MADF coding region. Table 2.2 lists all the MADF domain containing genes along with the corresponding nucleotide substitution model for both the partitions. Interestingly, for complete domain partition 29 out of 48 genes follow Kimura 2 parameter (K2P) (Kimura 1980) nucleotide substitution model. Out of these 29 genes 12 have BESS domain associated with them. 17 out of 48 genes have more invariant sites in their MADF domain while others show increased rate substitution across sites of the MADF domain. Out of the 16 MADF genes which are associated with BESS only 5 show more invariant sites which suggests that MADF domain is less constrained for evolutionary forces to act on. However on analysing the first second and third position for every codon in the MADF domain coding region we see that the sites at the first position follow more complex models like TIM2e, TIM3e, TNe and TVM (25 genes) than the ones at the second position (for 35 genes) which follow simpler models like JC (Jules Canton), K2P and F81. This shows that the first codon position is evolving in a more complex

manner than the second position. The third codon position being a wobble mostly evolves in a complex way.

Table 2.2: Nucleotide Substitution Model for the MADF domain. MADF coding sequence in 48 MADF domain containing genes in *Drosophila melanogaster* were used to predict nucleotide substitution model using IQTree. The genes from 1-32 contain single or multiple MADF domains while the genes from 33-48 contain MADF domain along with the BESS domain.

S.No.	Gene	Complete domain	1st codon	2nd codon	3rd codon
1	*ADD1*	TPM2+I	SYM+I	F81+I	TPM2
2	*az2*	K2P+I	TIM2e+I	JC	K2P+G4
3	*CG1602*	TNe+I	K2P+I	TPM3u+I	TNe
4	*CG1603*	HKY+G4	TIM3e+G4	HKY+G4	K2P
5	*CG3386*	K2P+I	JC+I	TPM2+G4	TNe+G4
6	*CG4004*	K2P+I	TIM2e	JC	TPM2
7	*CG5180*	TN+G4	TIM2e	F81+I	TN+G4
8	*CG5953*	K2P+G4	JC+I	JC	TPM2+G4
9	*CG6163*	K2P+G4	TNe+I	JC	TPM2+G4
10	*CG6175*	TIM3e+G4	TIM2e	JC+I	TIM2+G4
11	*CG6683*	HKY+I	TVMe+G4	TIM3e+G4	K2P
12	*CG7745*	K2P+I	TIM2e+I	F81	TPM2
13	*CG8281*	TIM3e+G4	TIM2e+I	JC	TPM2
14	*CG8765*	HKY+I	JC+I	HKY+G4	TPM2
15	*CG8944*	TIM3e+G4	K3P+I	F81+I	TPM2+G4
16	*CG9948*	TIM2e+G4	K2P	JC	TIM2e+G4
17	*CG10151*	K2P+G4	JC+I	JC+I	TPM2u+I
18	*CG10904*	K2P+I	K2P+I	TPM3+G4	K2P
19	*CG10949*	TIM2e+G4	TIM2e+I	HKY+G4	TPM2u
20	*CG11504*	K2P+I	K2P+I	F81	TPM2u+G4
21	*CG12155*	K2P+G4	TIM2e+G4	F81+I	TIM+G4
22	*CG12609*	K2P+I	K2P	TPM3u+I	TPM2
23	*CG12768*	TIM3e+G4	JC	JC	TPM2u+G4
24	*CG15601*	K2P+G4	TIM2e+G4	JC+I	TPM2+I
25	*CG31627*	K2P+I	TIM3e	TIM3+I	TN
26	*CG33017*	K2P+I	K2P+I	TNe	K2P
27	*CG42526*	JC	JC	F81	JC
28	*Hmr*	K2P+G4	TIM+G4	K3Pu	TPM2
29	*jigr-1*	K2P+G4	JC+G4	K3P+G4	K2P
30	*l(3) j2D3*	TPM3u+G4	TIM3+G4	JC+G4	K2P+G4
31	*mes2*	K2P+G4	TIM2e	JC	HKY+G4
32	*Rgr*	TIM3e+G4	TNe+G4	TVM+G4	HKY+G4
33	*adf-1*	K2P+I	TIM2e+I	JC	K2P+G4
34	*Brwl*	K2P+G4	JC	JC+I	TPM2u
35	*CG3919*	K2P+G4	K2P+G4	F81+G4	TPM2u+I
36	*CG4404*	K2P+I	TIM2e	JC	TPM2
37	*CG6276*	K2P+G4	TIM2e+I	JC	HKY+G4
38	*CG8119*	HKY+G4	TNe+G4	K2P+G4	K3Pu

39	*CG11723*	K2P+G4	JC+G4	K2P+I	TPM2u
40	*CG13204*	K2P+G4	TIM3e	F81	K2P+G4
41	*CG30403*	K2P+G4	JC+G4	K2P+G4	TPM2u+G4
42	*CG45071*	K2P+I	TIMe+G4	F81	TPM2u
43	*Coop*	TIMe+G4	K2P+I	TNe+I	TPM2u+G4
44	*dlip3*	K2P+I	TIMe	JC	K2P+I
45	*hng1*	K2P+G4	JC+I	K2P+I	TNe+G4
46	*hng2*	K2P+I	JC+I	TPM3u+G4	K2P+G4
47	*hng3*	TN+G4	K3P+G4	K2P+G4	TPM2u
48	*Stwl*	K2P+G4	K2P+G4	K2P+G4	TPM2

Similar analysis for genes with the BESS domain revealed that 16 out 22 genes containing the BESS domain follow the K2P (kimura 2 parameter) nucleotide substitution model (Table 2.3). Out of these 16 genes 12 are associated with the MADF domain. Interestingly, 9 out of 16 MADF-BESS genes have more number of invariant sites whereas most of the BESS only genes show more rate substitutions across the sites. This is opposite to what is seen with respect to the MADF domain. This holds true even when first and second position of every codon of BESS domain for each gene are analysed. There are more invariant sites for first and second position for the each codon for 9 and 11 BESS domain genes respectively. The first codon and second position sites for each codon seem to follow simpler nucleotide substitution models like JC, K2P and F81. First position site of the each codon for 12 genes follow K2P, K3P and JC models and second position site of each codon for 16 genes follow K2P JC and F81 models. Third position for each codon being a wobble follows more complex models like TPM3, TIM2, HKY and TIM3e models.

Table 2.3. Nucleotide Substitution Model for the BESS domain. BESS domain coding sequence in 22 BESS domain containing genes in *Drosophila melanogaster* were used to predict nucleotide substitution model using IQTree. The genes from 1-6 contain BESS domain while the genes from 7-22 contain BESS domain along with the MADF domain.

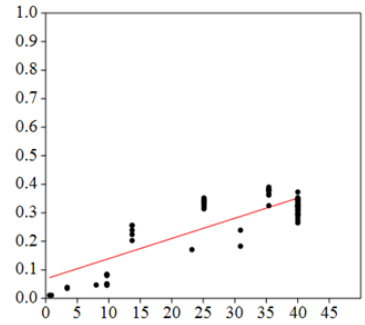
S.No.	Gene	Complete domain	1st position	2nd position	3rd position
1	*BEAF-32*	K2P+G4	K2P+I	JC+I	K2P+G4
2	*CG10209*	K2P+G4	Tne	HKY+I	HKY+G4
3	*CG34149*	K2P+G4	K2P+I	JC	K2P+G4
4	*Lhr*	TIM3e+G4	K2P	JC	TPM3
5	*Ravus*	TPM3u+I	TIM3e+I	JC+I	TPM3
6	*su(var)3-9*	K2P+I	K2P+G4	JC+I	TPM3

7	*adf-1*	K2P+I	TIM2e+I	JC	K2P+G4
8	*CG3838*	K2P+G4	JC+I	JC+I	TIM2
9	*CG3919*	TNe+G4	K3P	F81	K2P
10	*CG4404*	TIM3e+G4	K3P	HKY	TIM3e+G4
11	*CG6276*	K2P+G4	TIM3e	JC	K2P+G4
12	*CG8119*	JC	JC	JC	HKY
13	*CG11723*	K2P+G4	K3P+G4	K2P+G4	K2P+G4
14	*CG13204*	K2P+I	K2P+I	JC	TPM2u
15	*CG30403*	K2P+G4	JC+G4	K2P+G4	TPM2u+G4
16	*CG45071*	K2P+I	TNe+I	F81	K2P
17	*Coop*	K2P+I	TIM2	HKY	HKY+I
18	*dlip3*	K2P+I	TIM2e	JC+I	K2P
19	*hng1*	TIM3e+I	TNe+G4	F81+I	TPM3
20	*hng2*	K2P+I	JC	TPM3u+I	K2P
21	*hng3*	K2P+I	K2P+I	F81+I	HKY
22	*Stwl*	K2P+I	JC+I	F81+I	HKY

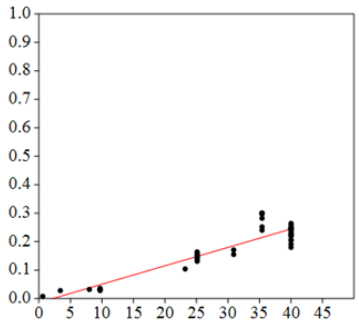
2.4.6 MADF and BESS domains show saturation divergence with time

To understand how the MADF domain for a particular gene has evolved with respect to its orthologs analysis for MADF domain of 48 genes in 12 *Drosophila* species was done using the nucleotide substitution model as predicted by IQTree program as described in the previous section. Divergence with respect to time (in million years) was plotted as in figure 2.7 which shows that that out of 48 genes, MADF domains of 6 genes show a linear relationship with evolutionary time. These are ADD1, CG3838, CG8119, CG12155, CG33017 and jigr-1 out of which CG3838 and CG8119 are associated by BESS domain (Figure 2.7). Others fit the curve of saturation divergence. The linear vs Michaels relationship was based on the Akaike information criterion (AIC value) (Posada 2001) calculated and the one that gave lower value was selected.

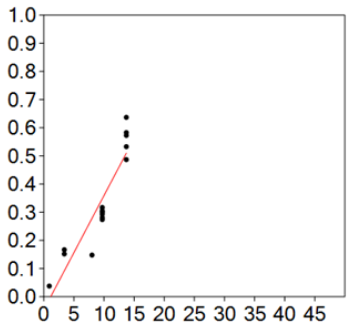
Similar analysis when done for BESS domain reveals three different kinds of plots (Figure 2.8). First, the linear relationship, which is shown by CG8119, Ravus and CG45071 out of which CG8119 and CG45071 are MADF-BESS genes. Second, the scattered plots which are shown by CG11723, CG13204 and CG34149 out of which the first two are the MADF-BESS genes. For the rest 16 genes, the BESS domain shows saturation divergence with time.



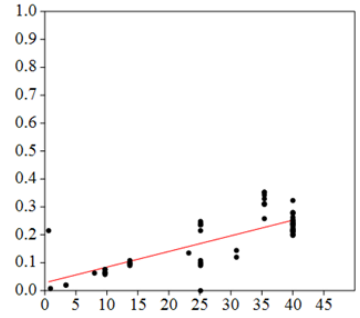
add1



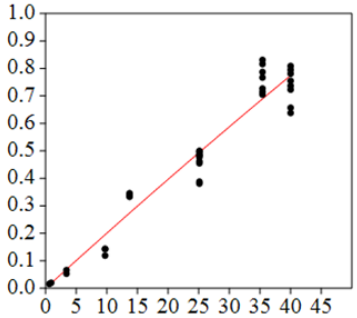
CG3838



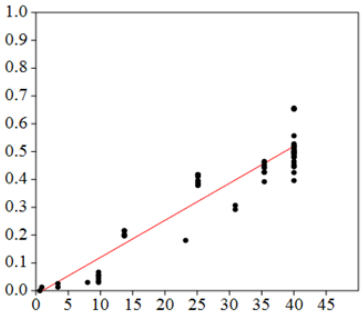
CG8119



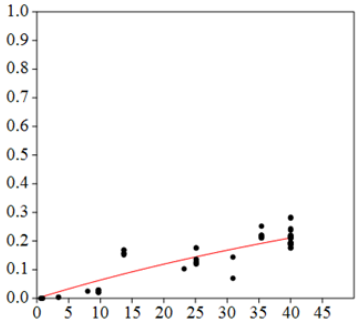
CG12155



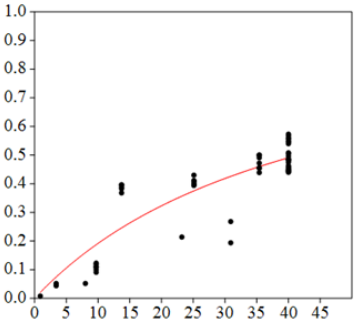
CG33017



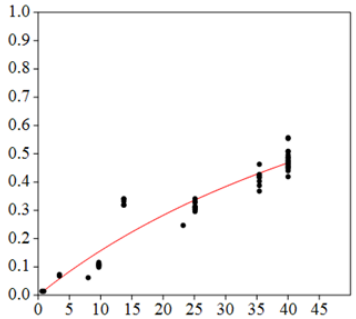
jigr-1



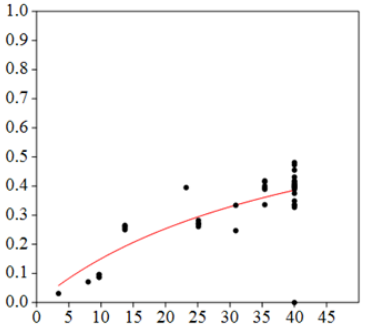
adf-1



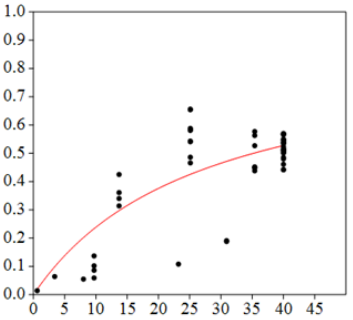
az2



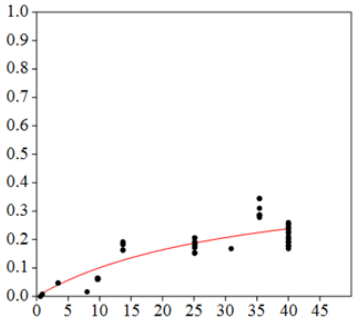
CG1602



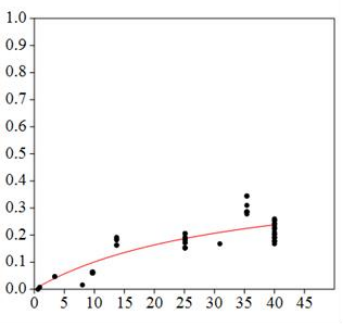
CG3386



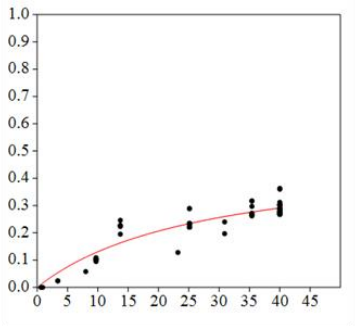
CG3919



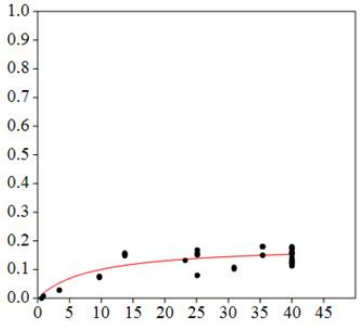
CG4004



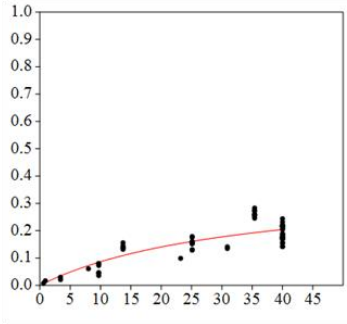
CG4404



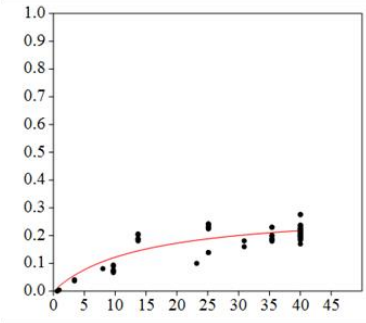
CG5180



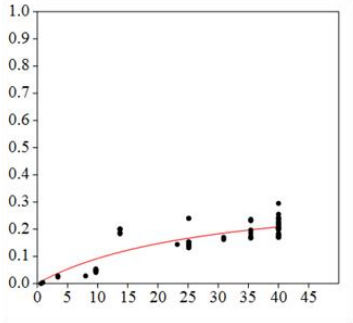
CG5953



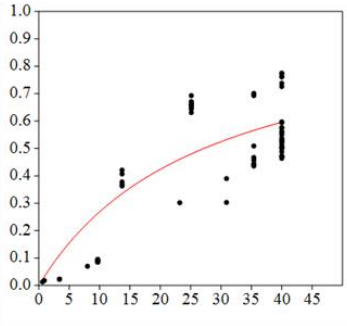
CG6163



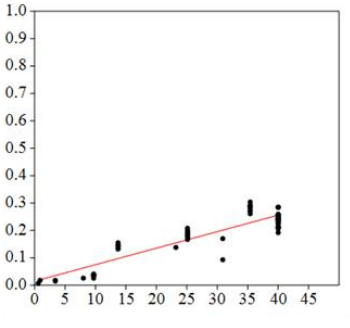
CG6175



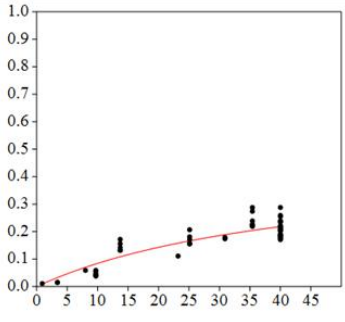
CG6276



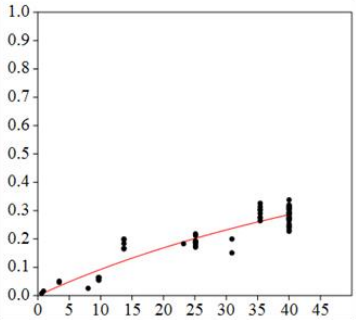
CG6683



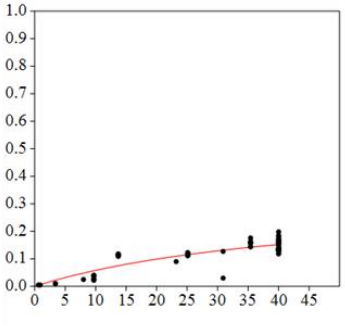
CG7745



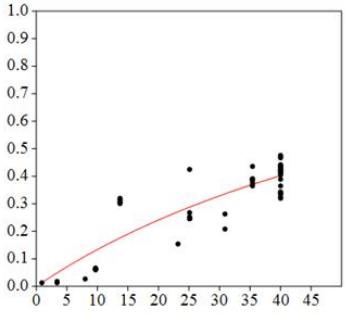
CG8281



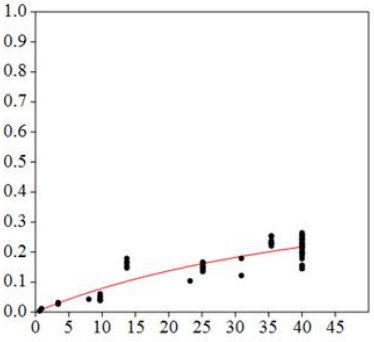
CG8765



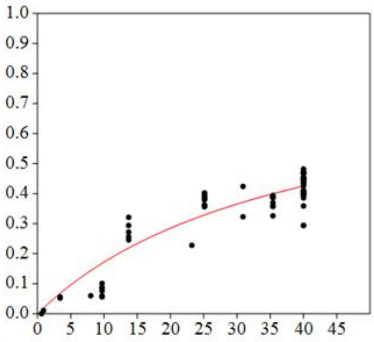
CG8944



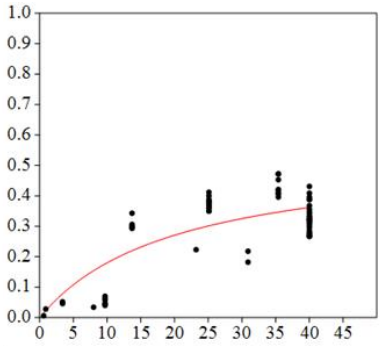
CG9948



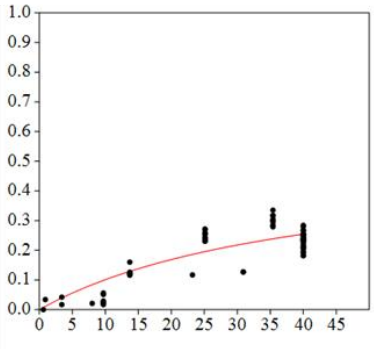
CG10151



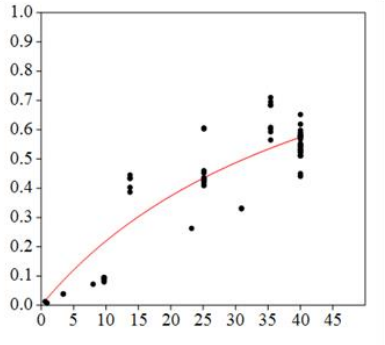
CG10904



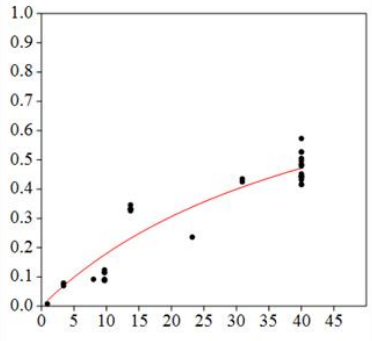
CG10949



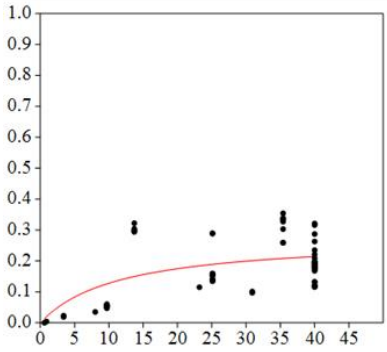
CG11504



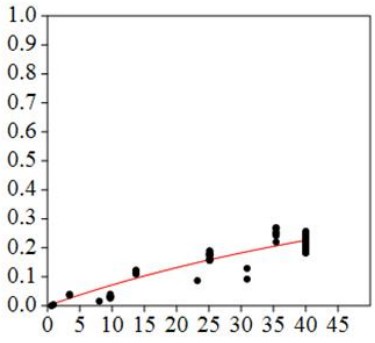
CG11723



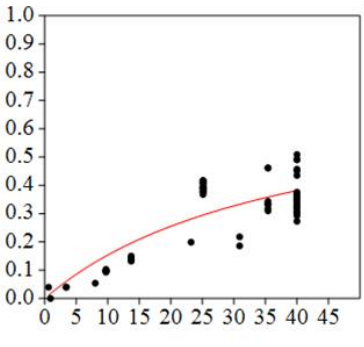
CG12609



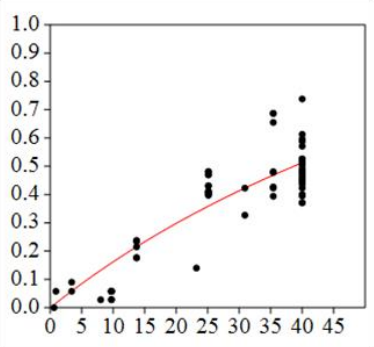
CG12768



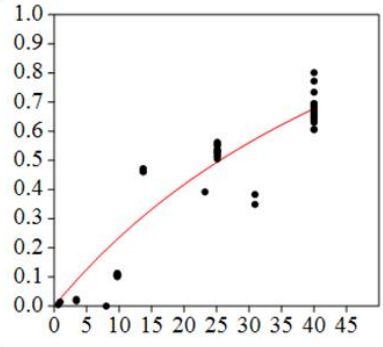
CG13204



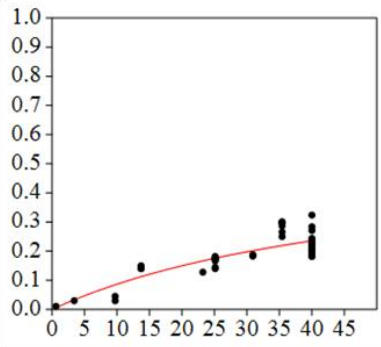
CG15601



CG30403



CG31627



CG45071

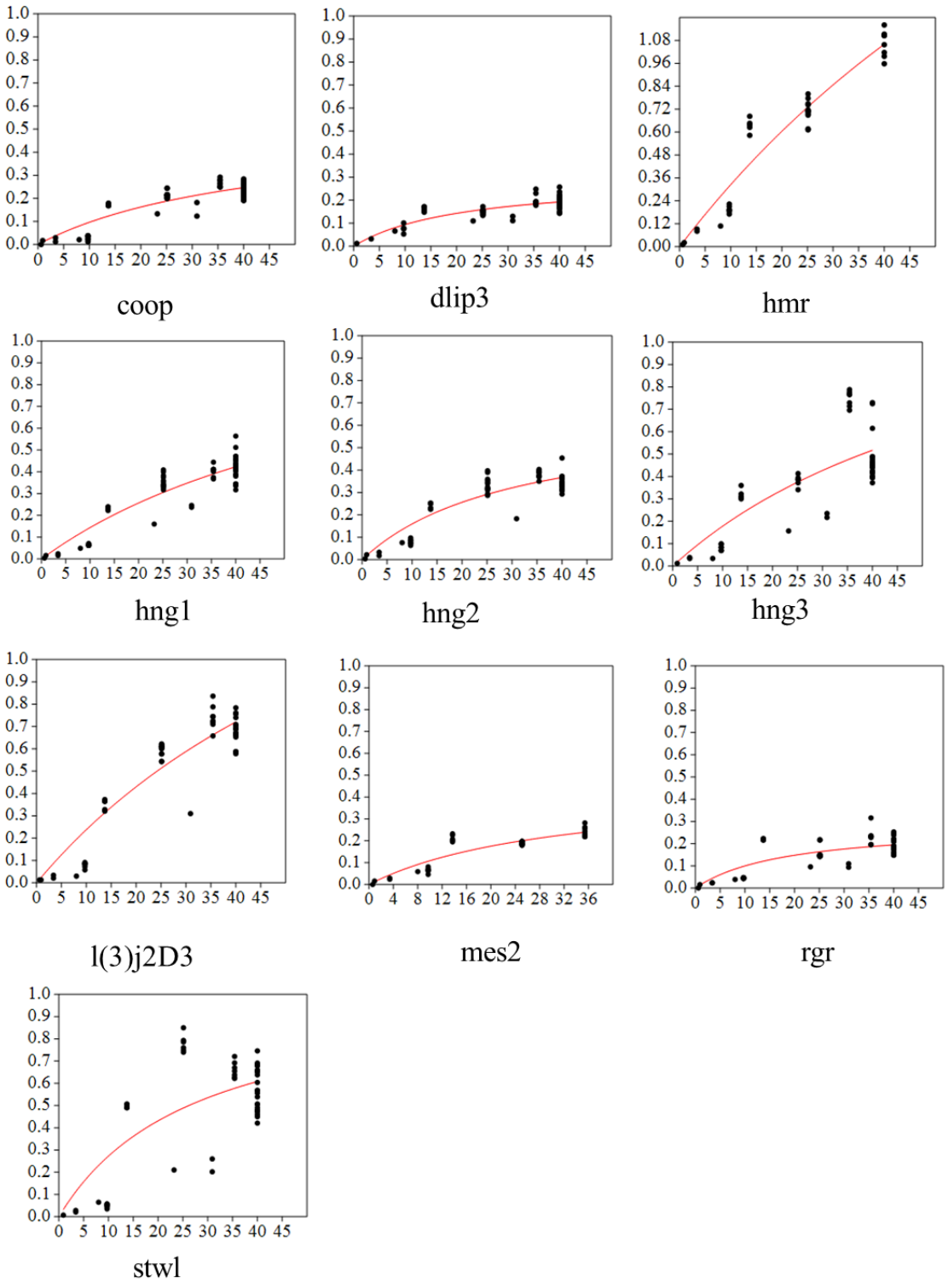
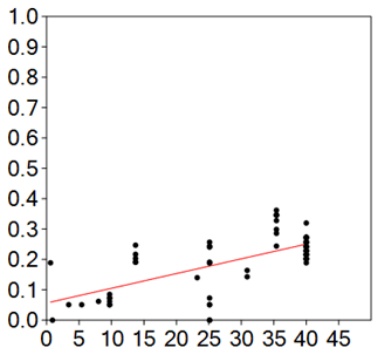
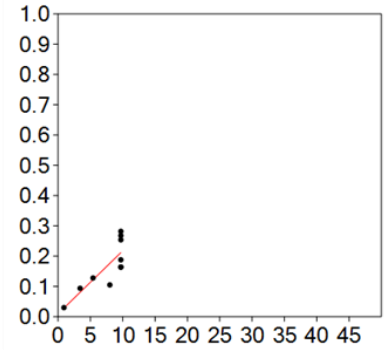


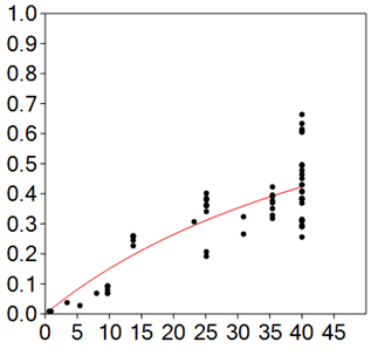
Figure 2.7. Divergence with time for MADF domain containing genes in 12 Drosophila species. X-axis is the divergence in million years for 12 species and Y-axis is the genetic distance calculated using the nucleotide substitution model for the specific gene. ADD1, CG3838, CG8119, CG12155, CG33017 and jigr-1 show linear relationship with time while the remaining genes are saturated.



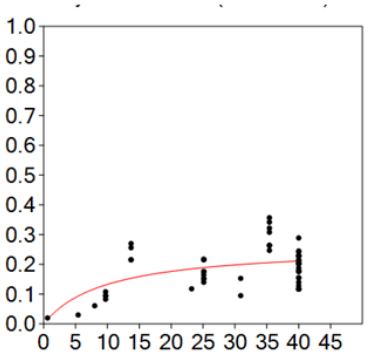
ravus



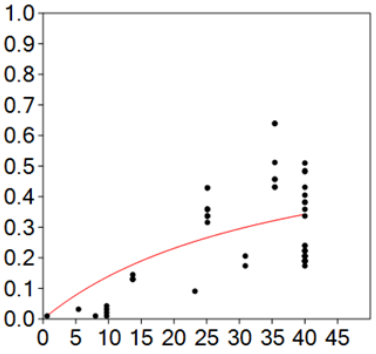
CG8119



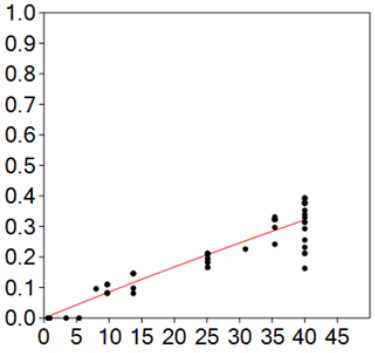
BEAF-32



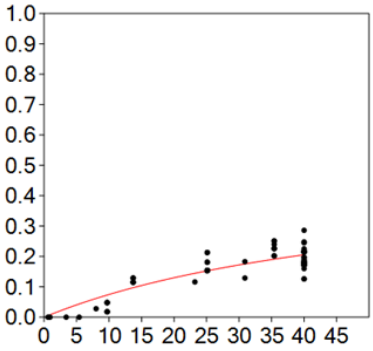
CG3838



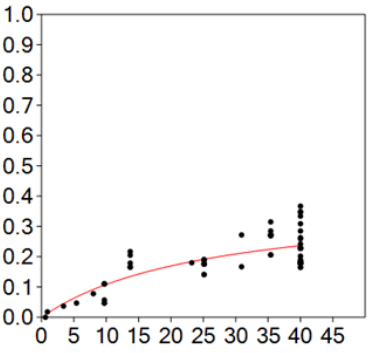
CG3919



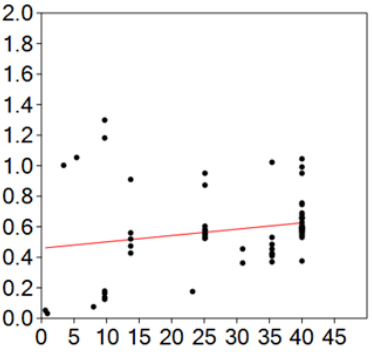
CG4404



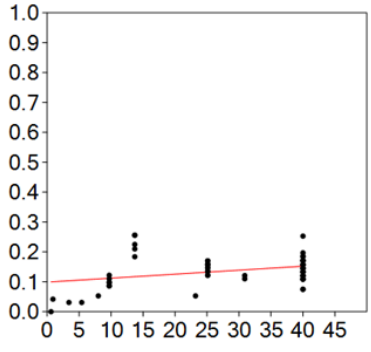
CG6276



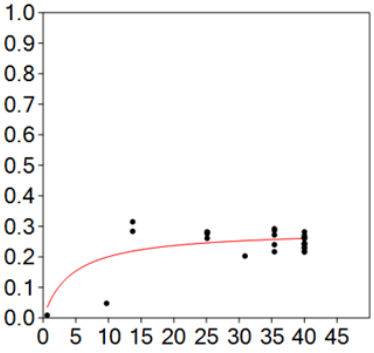
CG10209



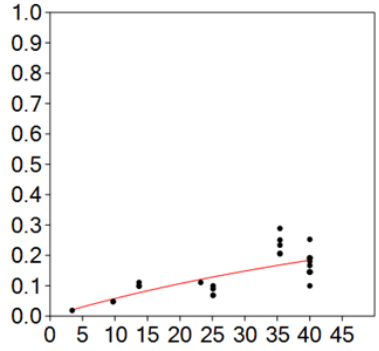
CG11723



CG13204



CG34149



CG45071

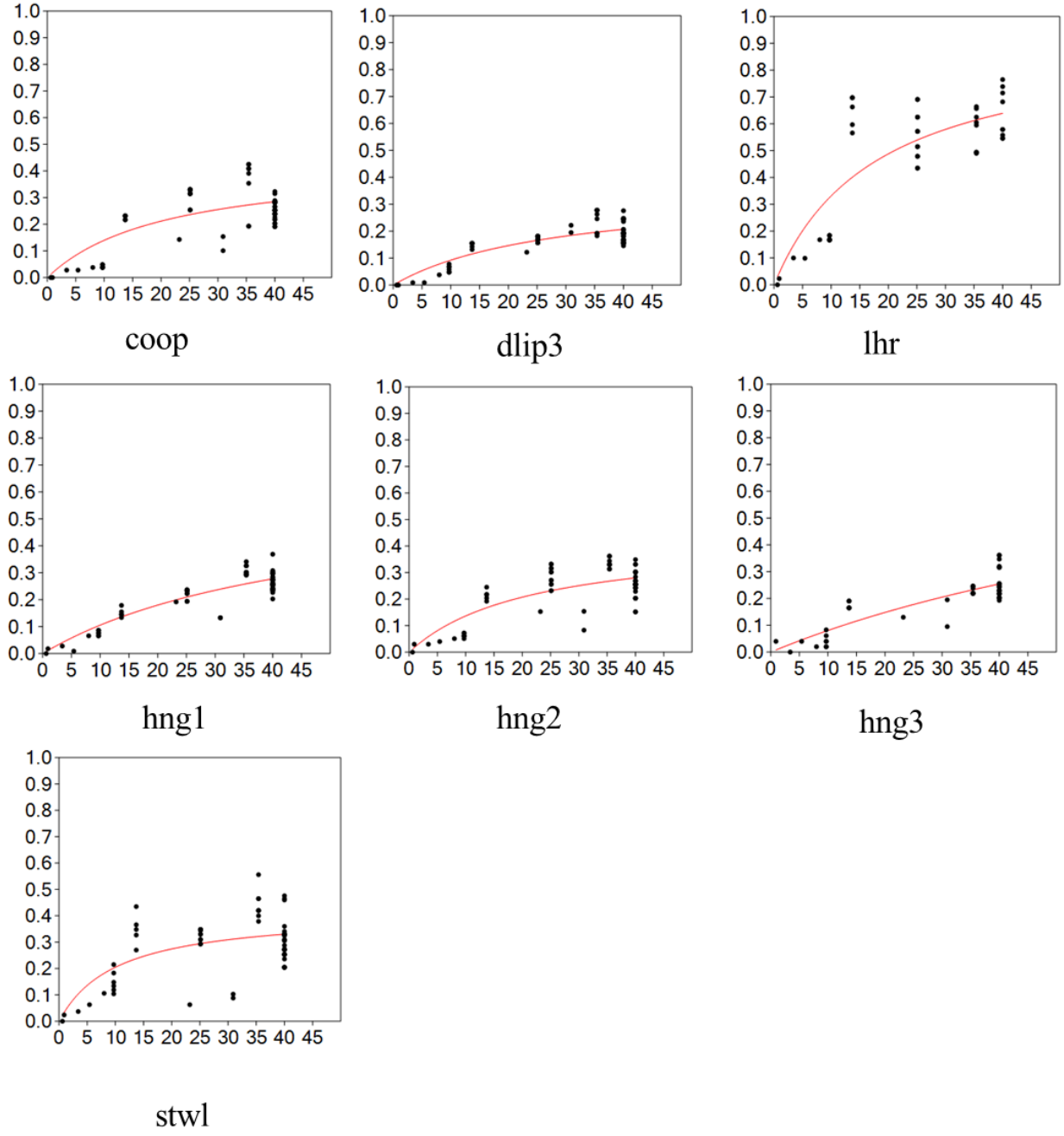


Figure 2.8. Divergence with time for BESS domain containing genes in 12 Drosophila species. X-axis is the divergence in million years for 12 species and Y-axis is the genetic distance calculated using the nucleotide substitution model for the specific gene. Ravus and CG8119 show linear relationship and the rest are saturated with time.

2.5 Discussion

Our data seems to agree with a model where the MADF-BESS genes duplicated more than 40 million years ago and were subfunctionalized. The exception is CG8119, which is not represented in all the 12 Drosophila species. In fact, CG8119 is represented only in the melanogaster group (*D. simulans, D. sechellia, D. melanogaster, D. yakuba, D. Ananassae*) suggesting that it originated about 15 Mya. The *stwl* gene is also an outlier in this family for the following reasons. *stwl* is the longest protein the family with the largest distance between the MADF and BESS domains. Although the protein does not have signatures for additional domains in the intervening sequences, based on current methods, there is a possibility of novel binding motifs in that region. *stwl* is also the only protein of the family implicated in epigenetic modification of chromatin. If this feature is related to the MADF and BESS domain function then the other MADF-BESS family members described here may be epigenetic regulators whose functions have not been discovered. *stwl* is also the nearest gene to CG3919, another MADF-BESS gene that codes for a small protein (302 amino acids) and appears to be closely related in terms of identity of MADF and BESS domains to *stwl*. However functional studies show that CG3919 is not functionally similar to *stwl* which suggests that divergence between the two genes is consequence of differences in the regulatory regions.

The subfunctionalization of a family can happen at three levels. Post duplication, the coding region of a gene can gain mutations which may perturb function at the level of a single polypeptide, but in combination the protein products of a family will retain the ability to perform the original function. This is best explained as the duplication-complementation model (DDC); (Force *et al.* 1999a; Hahn 2009). The DDC model emphasizes that mutations facilitate rather than hinder gene duplication. The second element of subfunctionalization is the change in spatio-temporal expression of the duplicate genes, as compared to the parent, allowing division and diversification of expression domains. The third element of subfunctionalization is the dose of the active species (Hahn 2009). Since members of the MADF-BESS family have the potential to hetero-dimerize via the BESS domain, common functionality of the family may be dependent on the formation of a heterodimer that may be the functional transcriptional regulator. Differential expression of genes may regulate formation and concentrations of hetero and homodimers and therefore regulate function.

The MADF-BESS gene family appears to retain functional redundancy and has also retained similar amino acid sequence in the two 'functional' domains. The MADF and BESS domain families independently have ~50 and ~25 members respectively, being in the same ballpark with the largest family in flies, the trypsin gene family, with 111 members (Zhang *et al.* 2003). It is interesting to note that in spite of tens of millions of years of evolution, the 16 MADF-BESS genes have retained the N/C architectural positioning of the two domains as well as have conserved the domain sequences.

We used IQ Tree software to predict the nucleotide substitution pattern each gene. Interestingly, we discover that more than 50% of MADF and BESS genes follow K2P. K2P(Kimura 2 Parameter) was first described and coined by Kimura (Kimura 1980). K2P is one of the simplest nucleotide substitution models. This denotes that all nucleotides are present at the same frequency in the gene. Transitions and Transversions happen at a different rate hence 2parameters. This shows us that most of these genes are evolving in a simple way. Another, interesting observation is that most of the BESS genes compared to MADF(only 5) have more invariant sites (Posada and Buckley 2004) which depicts that nucleotide sites in the BESS coding region do not change.

MADF and BESS genes duplicated and expanded at the time of fly evolution. After gene duplication the duplicates are either maintained as original or they diverge. We analyzed the nucleotide sequences of all MADF and BESS genes to predict the degree of divergence that has happened between the duplicates. We see that with the exception of few genes in both MADF and BESS genes most of these genes are saturated with respect to the substitutions that have taken place. This is an observed phenomenon after years of gene duplication(Smith and Smith 1996). Interestingly, we see that CG8119 which has both MADF and BESS domains is a recent member added to this family only when the melanogaster group emerged. CG8119 shows a linear curve which shows that the nucleotides are undergoing substitutions in this gene.

2.6 Uncovering redundancy using directed reverse genetic screens.

In order to discover if members of the MADF-BESS family have diversified in the last 40 years or have maintained their ancestral function, we decided to conduct a directed genetic screen using publically available RNAi lines from National Institute of Generics (NIG), Mishima; the Vienna Drosophila RNAi Collection (VDRC), Vienna and lines available in

Bloomington Drosophila Stock Centre (BDSC), Indiana. These lines were procured and based on the expression patterns of these genes and the ease of monitoring phenotypes we chose to reduce transcript levels of all MADF-BESS genes using five Gal4 drivers. The first driver chosen, *Da-Gal4* reduced transcript levels ubiquitously in cells and our data indicated that with the exception of a few genes, the knockdown of these genes one at a time was not lethal. Tissue specific drivers (Table 2.4) uncovered roles for individual genes in different tissues. One striking feature of the assay was that, in general, the MADF-BESS genes did not all together affect the development of a single tissue. Some MADF-BESS genes showed phenotypes in the wing (*stwl, Dip3, CG9437, CG8359, CG13897*) but the same set did not affect ovary or eye development (Figure 2.9).

Our targeted screen suggested that different members of the MADF-BESS family affect different stages of *Drosophila* development. An exception was the early *Drosophila* embryo (0-1 Hours; pre Maternal-to-Zygotic transition), where most of the MADF-BESS genes were known to be deposited as Maternal transcripts (Table 2.1) and where at knockdown of least four genes perturbed development.

Table 2.4: A targeted UAS-Gal4 reverse genetics screen. Genetic perturbations caused by single gene knockdown of MADF-BESS genes in different tissues of *Drosophila melanogaster*.

Gene	**Wing (MS1096 Gal4)**	**Eye (Eyeless Gal4)**	**Ovary (Nanos Gal4)**	**Embryo (Nanos Gal4)**
	Blade or Hinge Phenotype	*Nail Polish Assay; Eye Facet morphology*	*Morphology, DAPI staining*	*% Hatching of Embryos (<80%)*
adf-1	Line Not Available	Line Not Available	No	No
brwl	No	No	Yes	Yes
CG3919	No	No	No	No
CG4404	No	No	No	Yes
CG6276	No	No	Line Not Available	Maternal RNAi line not available
CG8119	No	No	No	No
CG11723	No	No	No	No
CG13204	No	No	No	No
CG30403	Line Not Available	Line Not Available	Line Not Available	Maternal RNAi line not available

CG45071	Line Not Available	Line Not Available	Line Not Available	Maternal RNAi line not available
coop	No	No	Line Not Available	Maternal RNAi line not available
dlip3	Yes	Yes (Strong phenotype)	No	Not done
CG9437/hng1	Yes	Yes (Mild phenotype)	No	Yes
CG8359/hng2	Yes	No	Line Not Available	Maternal RNAi line not available
CG13897/hng3	Yes	No	No	Yes
stwl	Yes	No	Yes	Yes

Notes: Wild type flies show 3-7% lethality in terms of hatching.

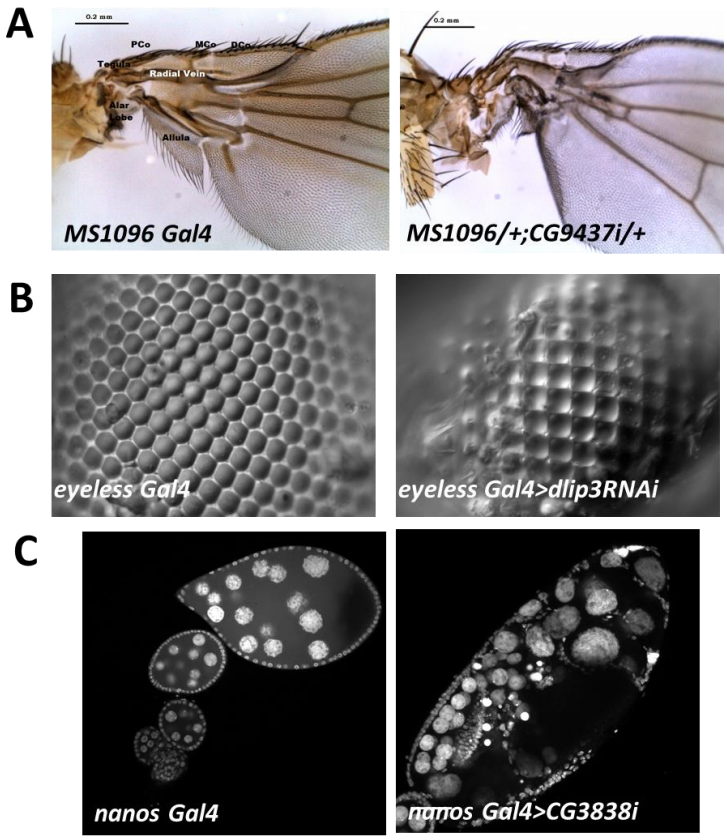


Figure 2.9. Targeted screen to knockdown MADF-BESS transcripts causes wing eye and ovarian defects. (A) CG9437 knockdown in dorsal wing leads to wing-hinge defects. **(B)** dlip3 knockdown in eye

precursor gives rise to tetragonal facets. **(C)** CG3838 knockdown in GSCs shows a repertoire of ovarian defects.

The screen allowed us to focus on studying redundancy and maintenance of subsets on MADF-BESS genes in context of different developmental domains. The methodology for uncovering redundancy was as follows:

(i) Discover a developmental domain where knockdown of any one of the MADF-BESS genes gave a mild or a weak phenotype.

(ii) Knockdown the other MADF-BESS genes in combination with the one gene that gave a phenotype. An enhancement of the phenotype would suggest that both genes were contributing to patterning/development.

(iii) Once a list of MADF-BESS genes are generated that contribute to the phenotype; attempt to rescue phenotype generated by knockdown of one gene by overexpression of the other.

(iv) If the double knockdown and rescue experiments suggest that multiple MADF-BESS genes are contributing to a single developmental domain/pathway then dissect out features of redundancy and overlap between the genes.

(v) Finally, uncover mechanism by which these genes are contributing to a common developmental function.

In the next two chapters, we apply the above ideas to two developmental domains where MADF-BESS genes contribute to *Drosophila* development. One, the *Drosophila* wing (Chapter 3) and second the Ovary (Chapter 4).

CHAPTER 3

The *Drosophila melanogaster* wing-hinge is patterned by a subset of MADF-BESS genes

---

3.1 Summary

In this chapter, we find that a subset of MADF-BESS family genes contribute to wing-hinge development. Three of the genes we have discovered, namely CG9437, CG8359 and CG13897 did not previously have an assigned function. These genes, based on our studies have been christened, based on their phenotype on loss of function, as *hinge* genes – *hinge1*, *hinge2* and *hinge3* respectively. All three genes appear to show significant functional redundancy in terms of their contribution to wing-hinge development. We demonstrate that the *hinge* genes, along with *stonewall* have critical roles in patterning the wing hinge by modulating Wingless (Wg) and Homothorax (Hth) expression, while other MADF-BESS genes in the family either play supporting roles or retain some functional aspects of the core, hinge-patterning function.

3.2 Introduction

The *Drosophila* **adult** wings are derived from a pair of primordial wing imaginal disc that are located dorsally in second thoracic segment. The cells that make up the disc are derived from the embryonic ectoderm (Williams *et al.* 1994). At the age of 24hrs, the embryo contains approximately 50 imaginal disc cells. The imaginal disc is compartmentalized (REFs; Making of the fly). The Disc has anterior/posterior boundary and a dorsal/ventral boundary that is aligned with the respective body axis. These boundaries subdivide the disc into four major compartments (Figure3.1). The disc is also subdivided into a proximal a distal compartments (Lawrence and Morata 1993). The proximal-distal axes divide the wing discs into regions that give rise to half of the notum (proximal), the hinge region (proximal) of the fly wing and the distal wing blade.

Of the three body axes that pattern the wing, namely the anterior/posterior axis, the dorso-ventral and the proximal/distal axis, the anterior-posterior axes is organized by engrailed, hedgehog and decapentaplegic (Guillen *et al.* 1995; Strigini and Cohen 1999). The dorso-ventral axis that is regulated by wingless and vestigial (Couso *et al.* 1993; Williams *et al.* 1994; Strigini and Cohen 1999) and finally, the proximal-distal axis structured by wingless and distaless forming the wing blade (distal) and hinge structures (proximal) (Cohen1 and Jurgens1 1989; Ng *et al.* 1995; Klein and Arias 1998).

wingless (wg) is the best characterized gene in the Wnt family of glycoproteins which is required in numerous developmental events for proliferation and patterning of different tissues at different times. In imaginal discs, during appendage formation different enhancers are required to activate in different expression regions which are controlled by different signaling pathways (Couso *et al.* 1993). In second instar larvae *wg* under Hedgehog signaling is expressed in ventral/anterior cells (Rodríguez Dd *et al.* 2002). Later it is confined to the cells which will form the wing margin this is controlled by Notch pathway (Neumann and Cohen 1996a). *wg* is also expressed in pouch in two concentric rings, which define and form the wing hinge of the wing. Wing hinge joins the thorax and the wing blade is an important structure for the flight of the the fly. Neumann and Cohen 1996 identified a 1.2kb enhancer region 5' of wg promoter which is required for the expression of wg in the hinge region. Deletion of this DNA fragment affects the wing hinge considerably and these mutants are called wg^{spade} mutants.

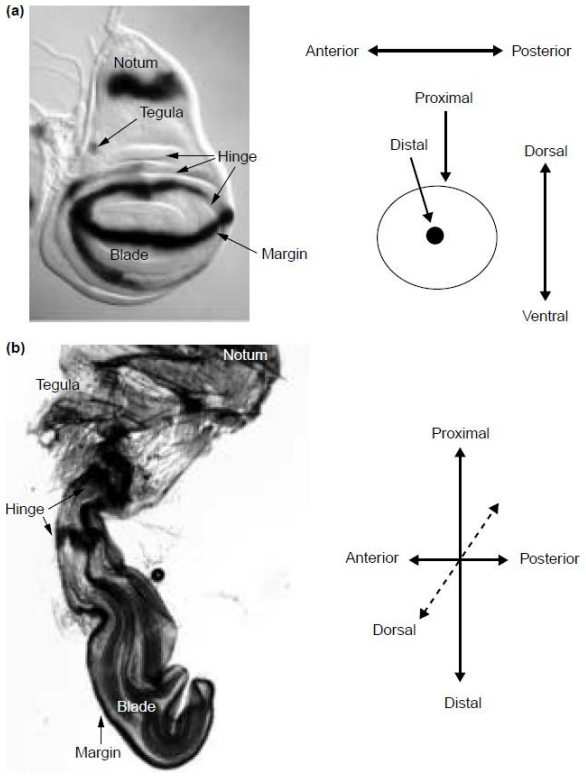


Figure 3.1 The wing disc of *Drosophila* gives rise to the adult wing. (A) wingless expression in the third instar wing disc. Wg is expressed in two ring-like domains in the hinge region and along the dorsoventral compartment boundary (D-V boundary) dividing the wing blade. The inner ring-like domain frames the wing blade. (B) the three dimensional adult wing formed from the two dimensional wing disc after pupation. Adapted from (Klein 2001).

Teashirt (zinc-finger protein) and *Homothorax* (meis-family homeobox protein) are two transcription factors that are responsible for wing hinge patterning (Casares and Mann 2000; Soanes *et al.* 2001). Both *tsh* and *hth* are expressed in all the cells of wing imaginal disc of the first instar larvae (Azpiazu and Morata 2000; Wu and Cohen 2002). In second instar larvae, morphogens like wg and dpp together, are responsible for repression of tsh and hth (Azpiazu and Morata 2000; Wu and Cohen 2002; Zirin and Mann 2004). *tsh* repression begins shortly after wg expression in ventral/anterior cells in early second instar. However what maintains this repression in the remains unknown (Wu and Cohen 2002). hth repression begins after *tsh* repression and wg and dpp together repress *hth* in the pouch(Zirin and Mann 2004). In this chapter we see how MADF-BESS genes are responsible wing-hinge development by genetically interacting with the important players like *wg, tsh,* and *hth* in the wing.

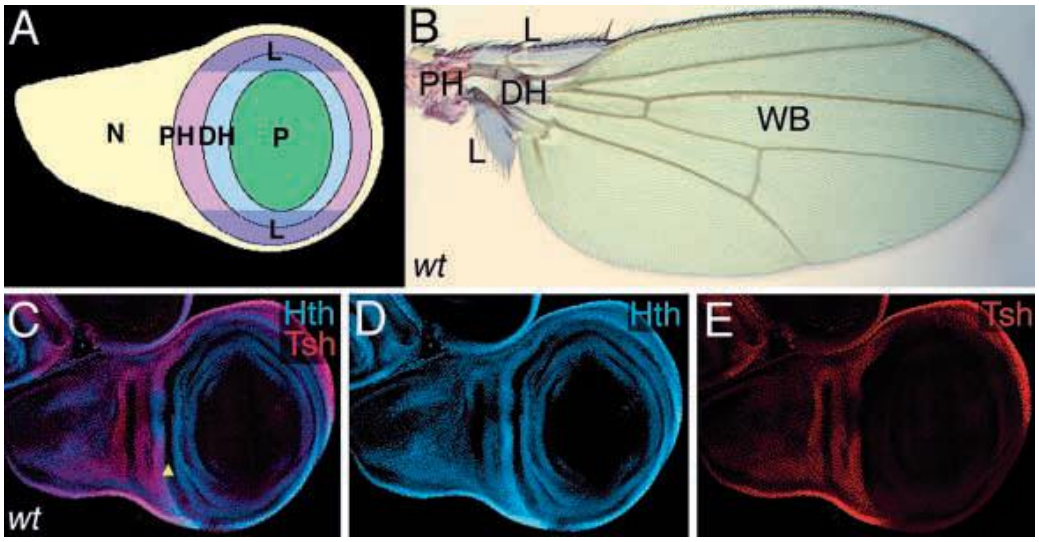


Figure 3.2 Proximo-distal axis is determined by Homothorax and Teashirt expression. (A) cartoon of third instar disc, anterior is up and posterior is down, pouch (P), proximal (PH) and distal (DH) hinge, lateral hinge (L) and notum (N). **(B)** Same regions in the adult wing (C-E) third instar imaginal disc staining for tsh (red) and hth (blue). Adapted from (Zirin and Mann 2004).

3.3 Material and Methods

3.3.1 *Drosophila* husbandry

All flies were raised at 25 ºC in standard corn meal agar. Crosses were set up at 25 ºC in a Sanyo Incubator. The females of the F1 progeny were screened for the phenotypes in all cases.

3.3.2 Fly Lines

MADF-BESS RNAi lines were procured from the Vienna Drosophila RNAi Center (VDRC), and the Transgenic RNAi Project (TRiP) lines were procured from Bloomington Drosophila Stock Center (BDSC) and the National Institute of Genetics (NIG) in Japan. Line numbers 4278,

30141, 3204R-2, 8119R-2, and 39733 had 19, 10, 3, 2, and 1 off-targets based on parameters defined in NEXT-RNAi (Horn *et al.* 2010). Lines that do not show off-target effects were used for our primary experiments.

VDRC: The lines obtained from the VDRC were the following: *UAS-CG9437i (hng1), UAS-CG8359i (hng2), UAS-CG13897i (hng3), UAS- stwli, UAS- coopi, UAS-CG3838i, UAS- adf-1i UAS-dip3i, UAS-CG11723i, UAS-CG6276i, UAS-CG4404i, UAS-CG3919i*

NIG: The lines obtained from the NIG were the following: *UAS-CG13204i* and *UAS-CG8119i*
BDSC: The lines obtained from the BDSC were the following*: UAS-CG3919i, UAS-mirrori, UAS-CG11723i, UAS-teashirti, UAS-extradenticlei, UAS- jingi, UAS-tiptopi, UAS-dppi, UAS-rotund* and *UASmCherry* . Deficiencies for *hng1* and *hng2* and mutants for *vestigial*, *pangolin* and nubbin were also obtained from the BDSC. *UAS-jing* and *UAS-hth* were obtained from the BDSC, the National Centre for Biological Sciences (Bangalore, India) (NCBS) stock center, and the Shashidhara Lab (Indian Institute of Science Education and Research, Pune, India), respectively.
Gal4 drivers: *MS1096 Gal4,* which expresses strongly in the dorsal region of the wing disc and weakly in the ventral regions, was used for most of the experiments. MS1096 is a Gal4 P-element insertion in the Beadex/dLMO Enhancer (P{GawB}BxMS1096) (Guillen et al. 1995; Milán et al. 2004). In addition, we utilized *da-Gal4, vg-Gal4, ap-Gal4, Sd-Gal4, omb-Gal4, MS209-Gal4 and ptc-Gal4* that express in the wing disc for characterization of our lines.
For detailed information about the fly lines please refer to Appendix 1

3.3.3 Wing measurements and statistical analyses

The wings were detached from the flies and mounted in clove oil. The images were captured using Leica Microsystems Light Microscope. Five wings for every genotype were used to mark the alula and wing boundaries and subsequently measured using Image J software. Wing area includes the proximal and distal wing together. For wing area measurements, wings with folds were avoided, and, when unavoidable, the area of the folded section was added to the total. All graphs were made in Sigma Plot (Manufacturer), and statistical analysis was done using Student's t-test in Graph Pad.

3.3.4 Immunostaining and in situ hybridization

Wing discs were dissected in PBS and fixed with 4% paraformaldehyde in 1X PBS for 20 minutes at room temperature. They were blocked in 1X PBS, 2% BSA, and 0.1% Triton for 1 hr; incubated with the primary antibody (anti-Wg 1:1000, anti-Hth 1:500 and anti-GFP 1:1000) overnight at 4C; washed 4 times for 10 min in 1X Phosphate Buffer Saline containing 0.1% Triton 3 and incubated with the appropriate fluorescent secondary antibody for 1 hr at room

temperature in the dark. The wing discs were then washed and mounted in Antifade. Anti-Hth was kindly provided by L. S. Shashidhara. Anti-Wg was purchased from the Developmental Studies Hybridoma Center. Anti-GFP (A11122) was obtained from Invitrogen. Images were taken on Zeiss 710 LSM confocal microscope at 20X and subsequently processed using Image J software. In situ hybridization in larvae was carried out as described in Kraut et al. (2001). Digoxigenin-labeled sense and antisense probes for *hng1* and hng2 were generated against 300- to 524-bp and 350- to 570-bp genomic regions, respectively. Anti-DIG (full name) was obtained from Roche and used at a dilution of 1:1000. Detection was done using nitroblue tetrazolium salt/5-bromo-4-chloro-3-indolyl phosphate (NBT-BCIP) stock solution.

3.4 Results

3.4.1 Members of the MADF-BESS family pattern the wing hinge

To understand the functions of the members of the MADF-BESS family in Drosophila, we decided on a loss-of-function approach, using the UAS-Gal4 system (Brand and Perrimon 1993; Duffy 2002) to reduce the transcript levels of every member of this family individually using double-stranded RNAi (Zamore *et al.* 2000; Kennerdell and Carthew 2000; Dietzl *et al.* 2007). *Drosophila* lines available in public stock centers were procured (see Materials and Methods), and a reverse genetic screen was conducted using wing-specific drivers. The primary result of this screen was that knockdown of three of the MADF-BESS domain genes produced a phenotype in the wing hinge (Figure 3.3& 3.6) with multiple wing-specific Gal4 Driver lines. The phenotype was similar in most cases – namely an affect in the wing hinge area that lead to wings being held-out (Figure 3.3). These concerned genes were CG9437, CG8359, and CG13897. Based on the phenotype, we have named these genes *hinge1 (hng1), hinge2 (hng2),* and *hinge3 (hng3),* respectively, and will refer to them as such in the subsequent text. A fourth gene from the family, *stonewall* (*stwl*), also produced a hinge phenotype (Brun *et al.* 2006). For the remainder of the text, we also abbreviate the RNAi line specific for the gene by adding an "i" at the end of the gene name. For example, a UAS-*hinge1* RNAi line will be abbreviated as *hinge1i or hng1i.*

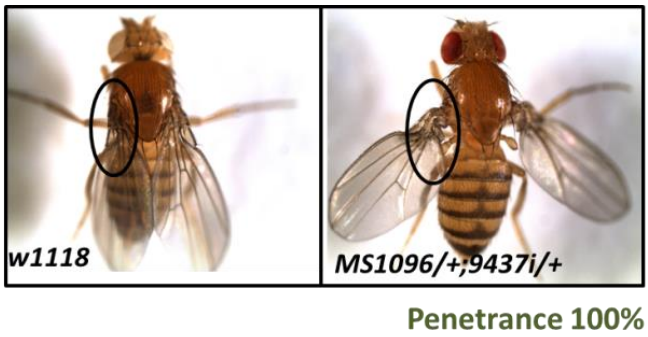


Figure 3.3. *hng1* knockdown using MS1096-Gal4 shows held out wings defect: Expression of a UAS-RNAi line driven under the control of MS1096-Gal4 at 25^0C resulted in a wing-hinge phenotype for CG9437i (*hng1i*). This affected the flight of the fly and also its ability to flap the wings.

The *hng1i* fly tends to keep its wings apart (Figure 3.3) and cannot fold its wings over the abdomen. The wings thus remained "held out," away from the body. The wing hinge functions to connect the wing blade to the thorax and has essential roles in fluttering of the wing during flight and in flexing the wing over the abdomen at rest. Experiments using a high-speed camera indicated that the *hng1*i females (MS1096/+; UAS-*hng1i*/+) could flap their wings but not fly, based on experiments using a cylinder drop assay. *hng2i* and *hng3i* knockdown flies were also found to be flightless.

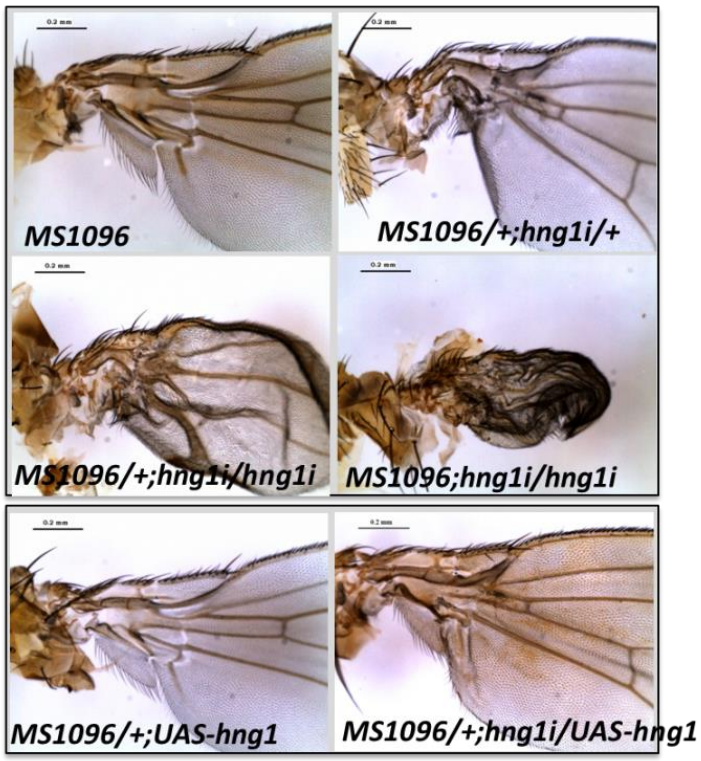


Figure 3.4 *hng1* knockdown using MS1096- Gal4 shows wing-hinge defects: Reduction of *hng1* transcripts in the wing-imaginal disc, by expressing UAS-*hng1* RNAi line, in the MS1096 expression domain leads to wing-hinge defects in the adult fly. Defects include a reduced/mispatterned alula, a bent hinge, and a disorganized costa region.

The phenotype is 100% penetrant and is dose dependent with an increased knockdown leading to a stronger phenotype that also affects the more distal wing blade (Figure 3.4). The defects can be rescued by co-expression of UAS-*hng1* in the same expression domain (Figure 3.4) suggesting that the RNAi line specifically affects *hng1* transcripts. Expression of UAS-*hng1* by itself does not affect normal wing development. Similar phenotypes were seen with other wing-specific Gal4 drivers such as *ptc-Gal4*, *sd-Gal4*, vg-Gal4, and *sal-Gal4* (Figure 3.5).

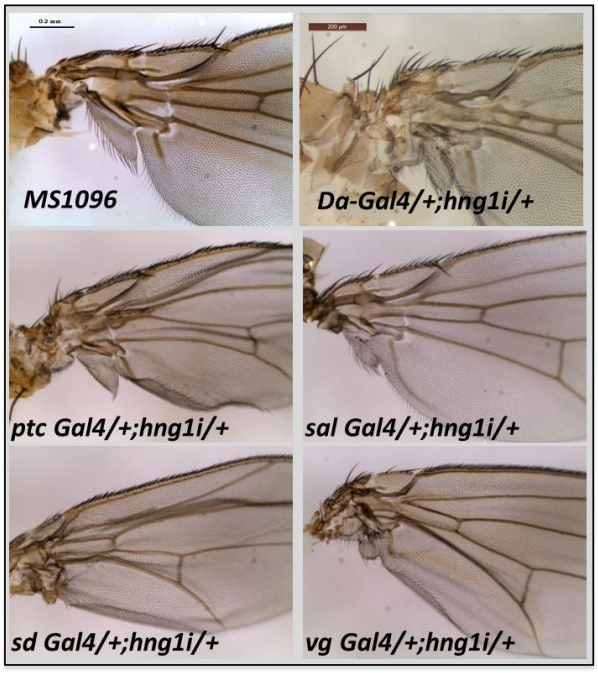


Figure 3.5. *hng1* shows wing-hinge defects with other wing specific drivers: *hng1*i when driven by da , ptc, sal, sd and vg Gal4 lines gives hinge phenotypes. A number of drivers like engrailed-Gal4, omb-Gal4 and hh-Gal4 however do not give these phenotypes.

The primary phenotype with MS1096-Gal4 was a bend in the costa region of the wing hinge and a reduction in the size of the hinge, with a dramatic effect on the patterning and size of the alula with respect to wild type. The phenotype was dose dependent, with drastic reduction in alula and wing-blade size with an increase in UAS or Gal4 dosage (Figure 3.4). When we used a more ubiquitous driver such as *daughterless-Gal4* (*da-Gal4*), the primary phenotype in the

animal was still a hinge defect (Figure 3.5), indicating specific roles for *hng1* in proximal wing development.

The *hng1*i phenotype can be rescued to a significant extent by expressing UAS-*hng1* in the same domain (MS1096) where *hng1* is knocked down (Figure 3.4). The rescue of the phenotype by UAS-*hng1* also demonstrates that *hng1* can directly affect the hinge phenotype, making it very unlikely that there are significant off-target RNAi effects or effects due to the insertion of the RNAi lines close to some other gene affecting wing development. We checked for the Gal4 dilution effect by crossing MS1096/+; *hng1*i/+ to UAS-mCherry, but the *hng1*i phenotype was unchanged. Overexpression of UAS-*hng1* by itself did not perturb normal wing patterning (Figure 3.4). Phenotypes similar to *hng1*i were seen with knockdowns of *hng2* and *hng3* (Figure 3.6) using gene-specific RNAi lines expressed in the *MS1096* expression domain. *hng1*, *hng2*, and *hng3* thus appeared to have critical roles in wing-hinge development. If these three genes are indeed functionally equivalent, as suggested by similar protein sequences in the MADF and BESS domains, it was expected that double knockdowns would enhance the initial phenotype. Simultaneous knockdown of *hng2* and *hng1* or *hng3* and *hng1* leads to enhanced wing-hinge phenotypes (Figure 3.6).

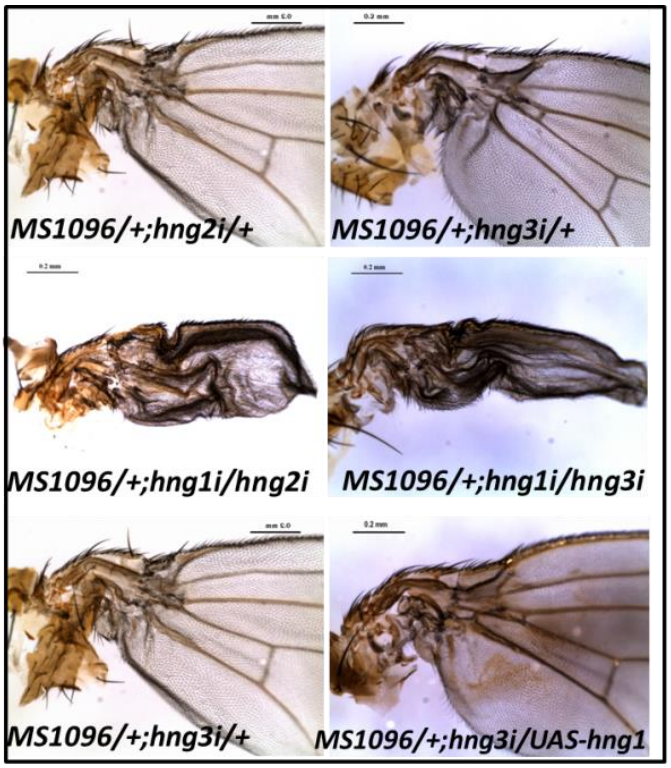


Figure 3.6. *hng2* and *hng3* knockdown show wing hinge defects: Two other genes (CG8359/*hng2* and CG13897/*hng3*) in the family also show a similar phenotype on knockdown. Double knockdowns of *hng1+hng2* or *hng1+hng3* in a single dose each mimic the phenotypes seen in an increase dose of *hng1* knockdowns. Over expression of *hng1* in *hng3i* background does not rescue the *hng3i* phenotype.

We further tested the effect of alternate reagents, namely deficiencies (Df) in the *hng1* and *hng2* loci, to support our observations. Specific deficiencies in 2R (57C3-57C7) and 3L (85B1-85C2) genomic regions completely remove *hng1* and *hng2*, respectively. Interestingly, heterozygous Df (*hng2*)/+ flies show a wing-hinge phenotype (Figure 3.7), with a penetrance of 68% (38/56 animals), strongly supporting the RNAi loss-of-function phenotype for hng2. The Df (*hng2*)/+ hinge phenotype closely resembles the hng phenotypes with a disorganized, mis-patterned hinge and a reduced alula (compare Figure 3.7 to Figure 3.4). This result strongly supports the conclusion derived from the RNAi experiments that *hng2* is indeed required for normal hinge development. When combined with MS1096; UAS-*hng1*i, Df (*hng1*) and Df (*hng2*) enhanced the wing-hinge phenotype (Figure 3.7).

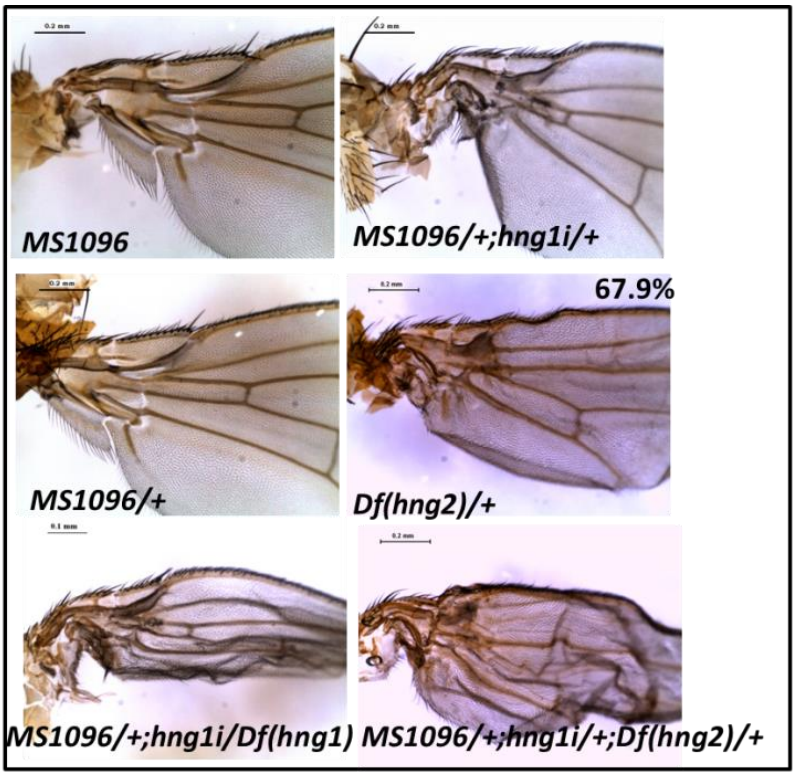


Figure 3.7. Deficiencies in *hng1* and *hng2* genomic locus show wing-hinge defects. Deficiency in the 85B1-85C2 genomic regions removes *hng2* completely, and 68% of flies show the wing-hinge phenotype.

This, when combined with *hng1* (MS1096/+; UAS-*hng1*-RNAi/+) knockdown, gives enhancement in the Deficiency phenotype. Deficiency in the 57C3-57C7 genomic region removes *hng1* completely and also shows enhancement when combined with *hng1* (MS1096/+; *UAS-hng1*-RNAi/+) knockdown.

We measured the size (Figure 3H) of the alula and wing blade (including the proximal hinge). These measurements allowed us to quantitatively assess the enhancement and suppression of phenotypes in the wing hinge and the wing blade.

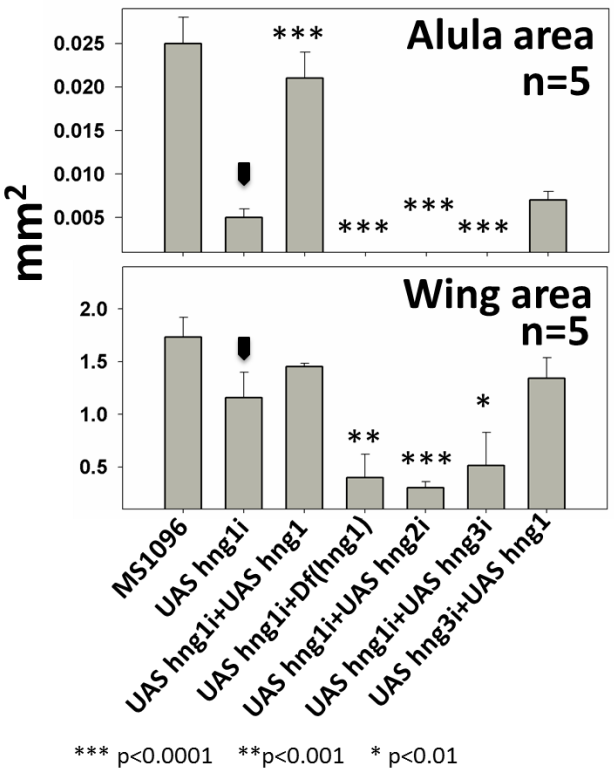


Figure 3.8. Quantitative data representation Parameters such as wing size (mm2) and alula size (mm2) are measured to quantify the phenotype in the above experiments. Arrowhead indicates the MS1096/+; UAS-*hng1*/+ line used as a control for statistical analyses. *p <0.05, **p<0.01, and ***p<0.001.

3.4.2 Expression of *hng* genes in wing imaginal discs

To confirm expression of the hinge genes in the wing imaginal disc, which gives rise to the adult wing, we visualized transcripts of *hng1* and *hng2* using in situ hybridization (Figure 3.9). *hng1* and *hng2* are expressed in the third instar larval wing disc, including regions that form the wing hinge. A GFP enhancer trap line for CG13897 (*hng3*) shows expression in the hinge (yellow arrows) as well as in the wing pouch (Figure 3.9).

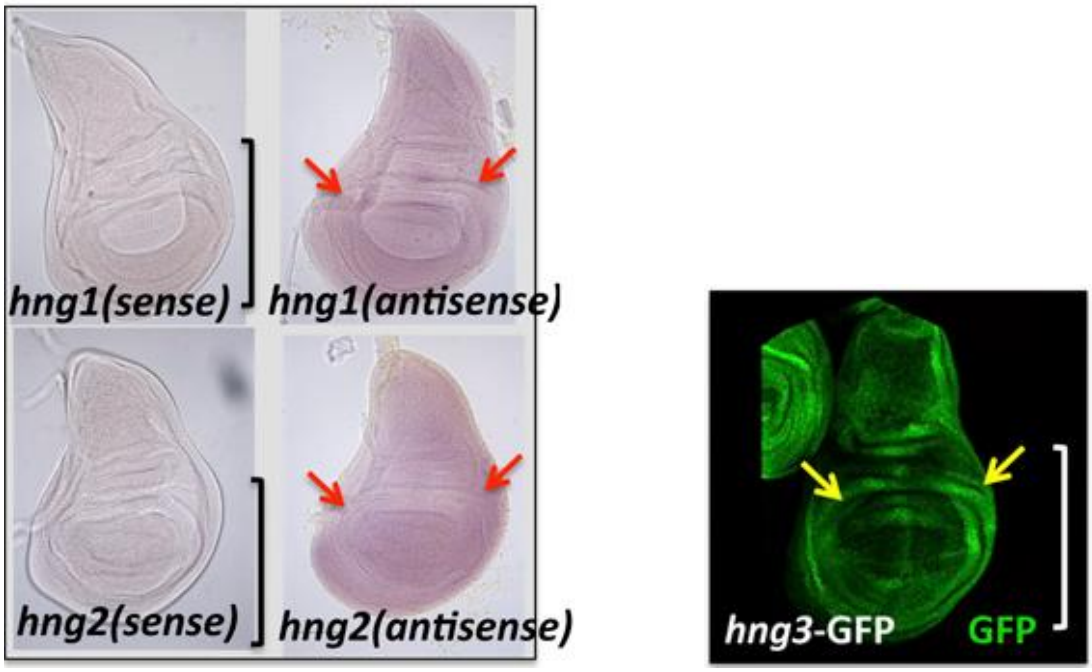


Figure 3.9. *In situ* hybridization against *hng1* and *hng2* transcripts *hng1* and *hng2* are indeed expressed in wing imaginal discs. The brackets mark the wing pouch with the red arrows marking part of the hinge-forming region. Expression pattern of *hng3* as shown by anti-GFP staining of an enhancer trap line (YB0086DE) in wing imaginal discs.

3.4.2 *hinge1* interacts with genes of wing-hinge gene regulatory network.

Since flies lacking *hng1* have a wing-hinge defect, we tested for genetic interactions between *hng1* and genes that play important roles in wing-hinge development. A central pathway involved in wing-hinge development consists of *teashirt (tsh)*, *homothorax (hth)*, and *extradenticle (exd)*(Mann and Abu-Shaar 1996; Casares and Mann 2000; Azpiazu and Morata 2000; Wu and Cohen 2002). *tsh* acts like an activator of *hth*, and binding of Hth is necessary for nuclear localization of Exd. The Hth:Exd complex then activates downstream targets that pattern the wing hinge. A double knockdown of *tsh* and *hng1* rescued the hinge defect and also the size and patterning of alula, indicating that *hng1* is a negative regulator of *tsh* function (Figure 3.10A&D). A similar, though a less dramatic rescue, was seen upon simultaneous *hth* and *hng1* knockdown. Based on these results, it was predicted that Hth activity is upregulated in *hng1*i. Indeed, an increase in Hth expression (Figure 3.10B) along with an expansion of the Hth expression domain was observed in *hng1*i wing discs. Based on this result we predicted that a further increase in Hth expression using UAS-*hth* in the *hng1*i background would dramatically

enhance the *hng1*i phenotype, and this was indeed observed (Figure 3.10A). Jing is a zinc-finger transcription factor implicated in repression of *tsh* and *hth* in the wing hinge(Culi *et al.* 2006). We tested if *jing* and *hng1* interact genetically. *jing* and *hng1* double knockdown rescued the hinge defect while *Jing* overexpression in *hng1* knockdown in animals enhanced the *hng1*i defect (Figure 3.10C and D). This indicates that jing negatively regulates *hng1*.

A major player in wing-hinge development is Wg, which has roles in patterning by restricting the tsh-hth network to the wing hinge. Wg staining in the third instar wing disc marks an outer (wg-OR) and inner ring (wg-IR)(Couso *et al.* 1993) (Neumann and Cohen 1996b) (Russell 2000) (Rodríguez Dd *et al.* 2002) with a gap in between. The two rings are critical regions for wing-hinge development with the members of the wing-hinge gene regulatory network (GRN) interacting with or regulating Wg or being regulated by expression of Wg. wg-IR, regulated by the Wg^{spade}-flag enhancer(Neumann and Cohen 1996b) patterns a major section of the region of the hinge that is affected in the *hng1* knockdown. Wg-IR also drives intercalary proliferation, generating the gap region between the rings (Zirin and Mann 2004). When *hng1* is knocked down in the MS1096 expression domain (MS1096/+;UAS-*hng1*i/UAS-*hng1*i), a broadening of the Wg expression (Figure 3.11A) domain with an intrusion into the gap region was observed. A similar broadening was observed for MS1096/+; Df(*hng1*i)/ UAS-*hng1*i (Figure 3.11C). Knockdown of *hng1* in the patched (ptc-Gal4) expression domain (ptc-Gal4/+; UAS-*hng1*i/+) also leads to derepression of Wg in the gap region. Interestingly, although ptc-Gal4 also expresses where the anterior/posterior boundary cuts the dorsal/ventral boundary, Wg expression was not de-repressed or broadened at the D/V boundary, indicating specific roles for *hng1* regulation in the wing hinge. The $Wg^{spadeflag}$ animal lacks expression of wg-IR, resulting in a wing-hinge phenotype (Neumann and Cohen 1996b). Since *hng1*i causes de-repression of Wg in the wg-IR, we predicted that a heterozygous combination of *hng1*i with $wg^{spadeflag}$ would lead to a mild rescue of the *hng1*i phenotype due to a decrease in Wg in the wg-IR. Indeed, MS1096-Gal4/+; UAS-*hng1*i/$Wg^{spadeflag}$ animals show a rescue of the wing hinge when compared to *hng1*i (Figure 3.11D).

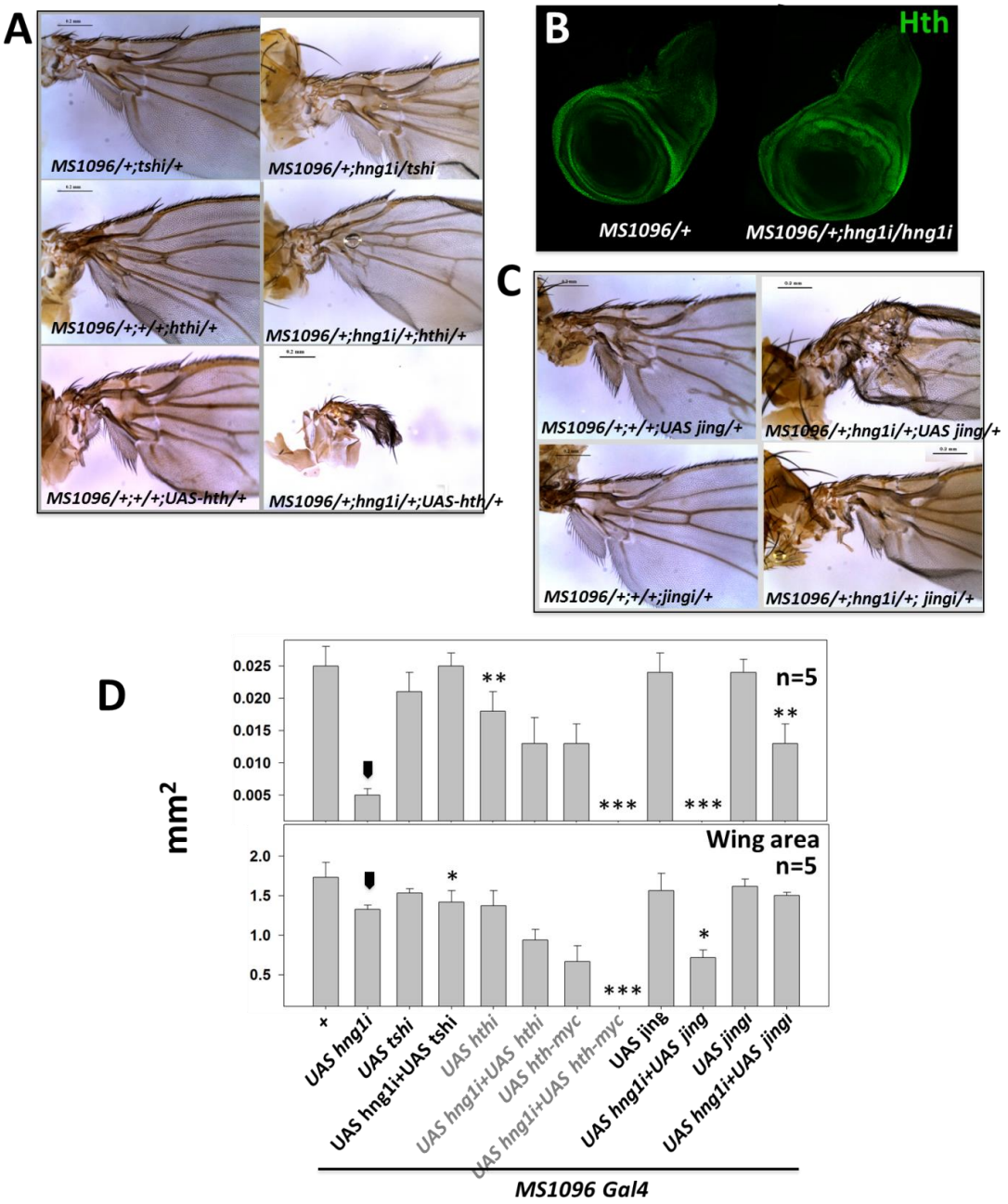


Figure 3.10 *hinge* genes are part of the GRN that patterns the wing hinge. (A) A knockdown of *tsh* and *hth* in the MS1096/+; *hng1i*/+ animal rescues the *hng1* phenotype, whereas overexpression of hth enhances the *hng1* phenotype severely. The RNAi and the UAS lines used to alter transcript levels for *hth* and *tsh*, by themselves, have mild hypomorphic effects. **(B)** Hth is broadened/derepressed in and around the gap region when *hng1* expression is reduced in the MS1096-Gal4 expression domain in the wing imaginal discs. **(C)** A knockdown of jing in the MS1096/+; *hng1*i/+ animal leads to a rescue of the *hng1i* phenotype while co-expression of UAS-*jing* leads to an enhancement of the proximal wing phenotype **(D)** Wing size (mm2) and alula size (mm2) are measured for genetics interactors of *hng1* with known wing-hinge GRN genes. Arrowhead indicates the MS1096/+; UAS-*hng1i*/+ line used as a control for statistical analyses. *p< , 0.05, **p<, 0.01, and ***p< , 0.001.

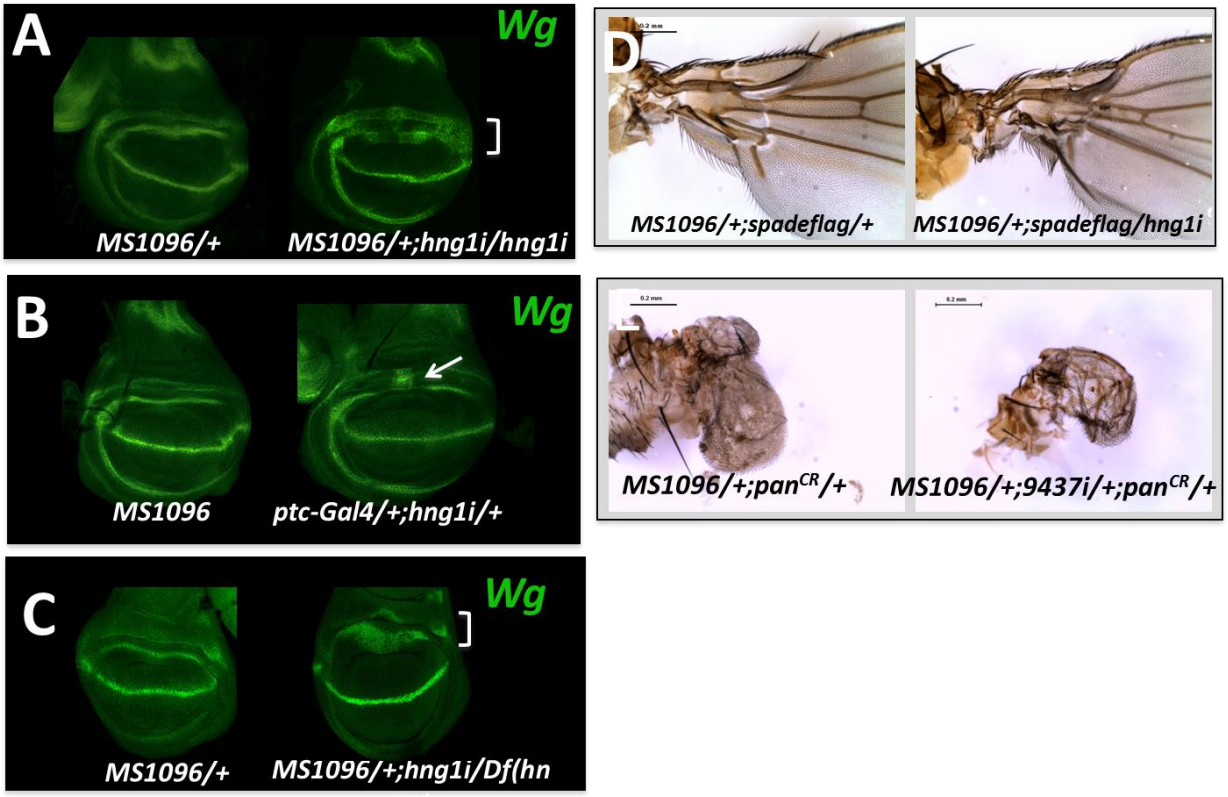


Figure 3.11. *hinge* genes are part of the GRN that patterns the wing hinge. (A) Wg is derepressed in and around the gap region between the IR and the OR in the MS1096/+; UAS-*hng1*i/UAS-*hng1*i wing imaginal disc. **(B)** Wg is derepressed in a small stripe in the gap region when *hng1* expression is reduced in the ptc-Gal4 expression domain. **(C)** Wg is derepressed in and around the gap region between the IR and the OR in the MS1096/+; UAS-*hng1*i/Df(*hng1*) wing imaginal disc. **(D)** *hng1* knockdown in heterozygous spadeflag background in the MS1096-Gal4 domain mildly rescues the hinge defect. **(E)** Knockdown of *hng1i* appears to weakly rescue the wing phenotype of MS1096---Gal4/+;UAS *panCR/+*.

Since loss of *hng1* in the wing disc showed upregulation of wingless expression, we checked for increased cell proliferation in these animals. We did not, however, detect any significant change in cellular proliferation (Figure 3.12).

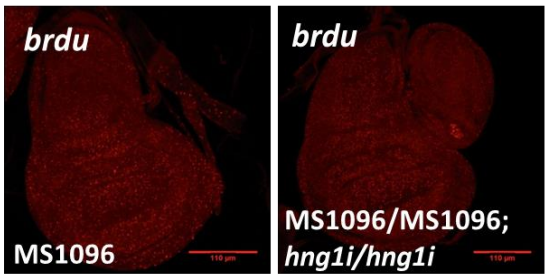


Figure 3.12: *hng1* knockdown does not show any change in cell proliferation. Brdu staining in the MS1096 Gal4/+ wing discs and MS1096/MS1096; *hng1*i/*hng1*i wing discs did not show any change which indicates that defects manifested in the adult wing are because of mis-patterning and not because of changes in cell proliferation during development.

Wg expression in wg-IR is driven by two independent mechanisms: Nubbin, vg, and rotund are required for wg expression in the wg-IR in the early third instar stages. The second mechanism involves hth function (Rodríguez Dd *et al.* 2002). We knocked down *hng1* in *vg* and *nubbin* mutant background (Figure 3.13A); however, we did not find any significant interaction. Overexpression of rotund in MS1096/+; *hng1*i background also did not show any change in the phenotype. Dpp expression in the posterior compartment of the wing disc has been shown to be responsible for the patterning of the proximal wing; the alula and dpp transcription is particularly mirror dependent(Foronda *et al.* 2009). dpp knockdown in MS1096/+; *hng1*i/+ background resulted in the enhancement of the *hng1* phenotype (Figure 3.13B). However, we do not see any interaction with mirror.

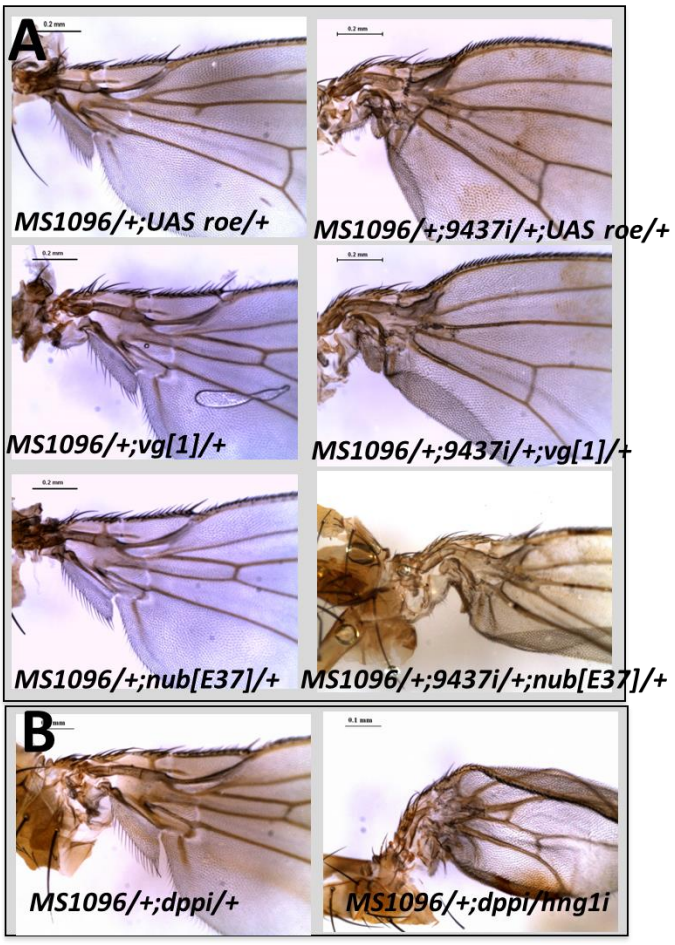


Figure 3.13. *hng1* does not genetically interact with genes that regulate wg-IR expression (A) Knockdown of *rotund, vestigial* and a *nubbin* mutant in an *hng1i* knockdown background do not modify the *hng1i* phenotype. (B) Knockdown of a weak *UAS dppi* line in the *hng1i* background enhances the *hng1i* phenotype.

3.4.3 Testing redundancy using double gene knockdowns

hng1, hng2, and *hng3* appear to be genes with similar or equivalent roles in the wing hinge. Other MADF-BESS knockdowns, with the exception of *stwl,* do not appear to give a hinge phenotype in single knockdown experiments. It is possible that these genes have a partially redundant function in the wing hinge. One method of testing this would be to simultaneously knock down each of these genes, along with *hng1,* and check for an enhancement of the *hng1* phenotype. Any enhancement would also provide evidence for a functional role for the other MADF-BESS genes, even in the absence of gene expression data. Double knockdowns on all remaining MADF-BESS genes in the background of *hng1* RNAi indicated that *Dip3, Coop, CG3838, CG11723, and CG4404* (Figure 3.14, A and B) had roles in wing-hinge patterning as they enhanced the *hng1* phenotype, while *CG15845, CG8119, CG30403, and CG13204* did not. This indicates that a substantial fraction of the MADF-BESS family (9 of 16) plays a direct or a supporting role in wing-hinge development. The lack of interaction with some of the tested genes may indicate weak RNAi lines that do not reduce transcripts significantly or that these genes do not express in the wing hinge or that these non-interacting genes do not have a role in wing-hinge development. It is also feasible that all members of the family may be involved in wing-hinge patterning.

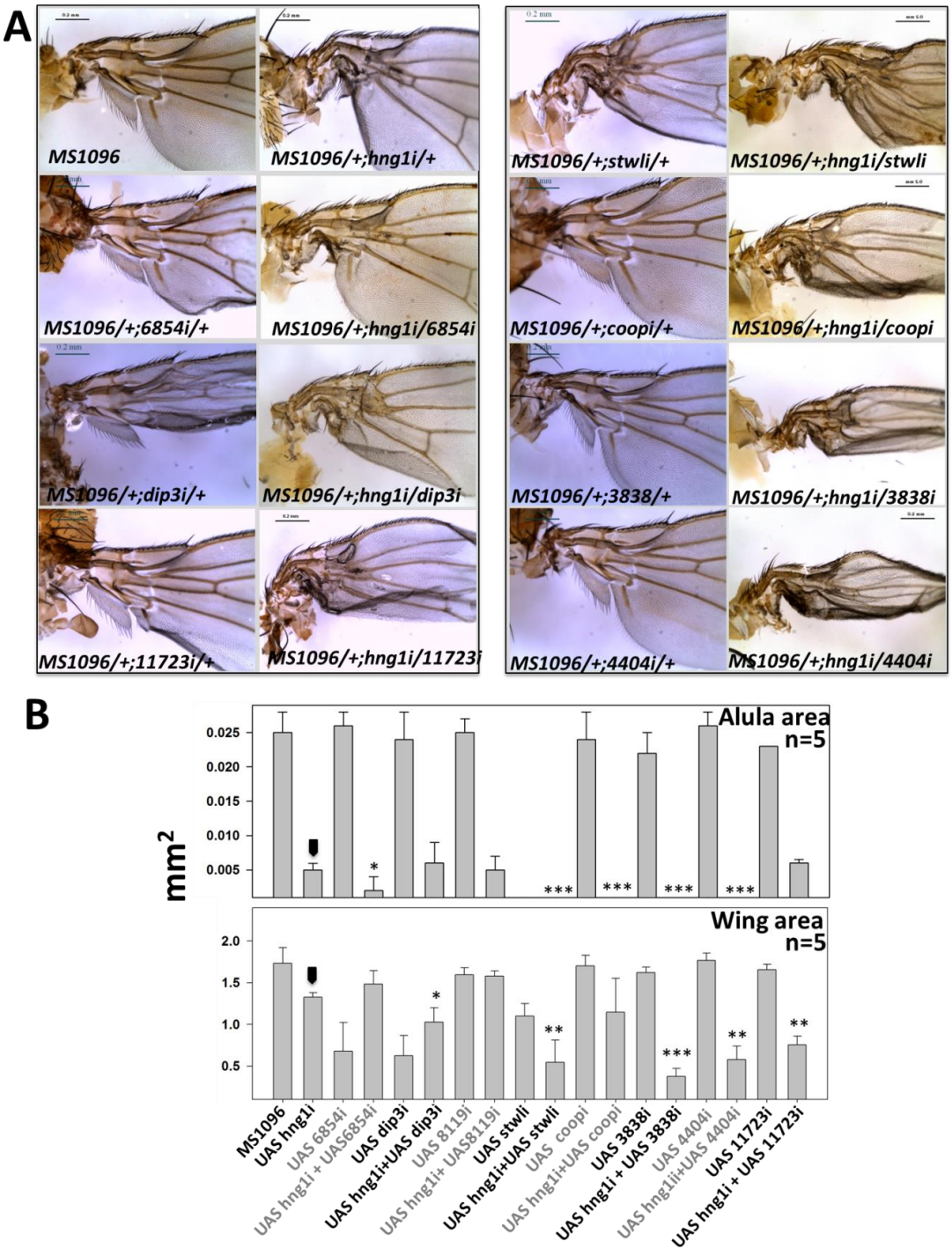


Figure 3.14. *hng1* phenotype is enhanced by knockdown of other MADF-BESS genes. (A) CG6854, stwl, coop, CG3838, CG4404, and CG11723 knockdown in the *hng1* background enhance the MS1096/+; UAS-*hng1*i/+ phenotype. The knockdown of these genes by themselves, with the exception of stwl, using RNAi lines does not affect the hinge significantly. The enhancement indicates that these genes are expressed in the cells that pattern the wing hinge and may have partially redundant roles in the hinge-mediated development of the wing hinge. (B) Wing size (mm2) and alula size (mm2) are measured for

genetic interactions of *hng1* with other MADF-BESS family genes. Arrowhead indicates the MS1096/+; UAS-*hng1*/+ line used as a control for statistical analyses. *p<, 0.05, **p<, 0.01, and ***p<, 0.001.

3.4.4 Confirmation of redundancy/equivalence by rescue experiments

We next tested the ability of the MADF-BESS genes to rescue the *hng1* phenotype. As shown in Figure 3.15A, in addition to *hng1* itself (Figure 3.4), expression of *CG11723 and CG13204* could rescue the *hng1*i phenotype to a significant extent (Figure 3.15A and B). Dip3 overexpression, on the other hand, enhanced the phenotype of *hng1*i. The data for Dip3 is reminiscent of data for Dip3 in *Drosophila* eye development where both loss and gain of Dip3 function shows similar phenotypes (Duong *et al.* 2009). *hng1* overexpression could not (Figure3.6) rescue the *hng3i* phenotype, indicating that there are at least some functional differences in the protein products of *hng1* and *hng3*. These data raise the possibility of Hng1 and Hng3 proteins being functionally diverse and also the possibility of the Hng1:Hng3 dimer being the functional entity for hng family function. This possibility is discussed in the next section and incorporated in a model for Hng activity.

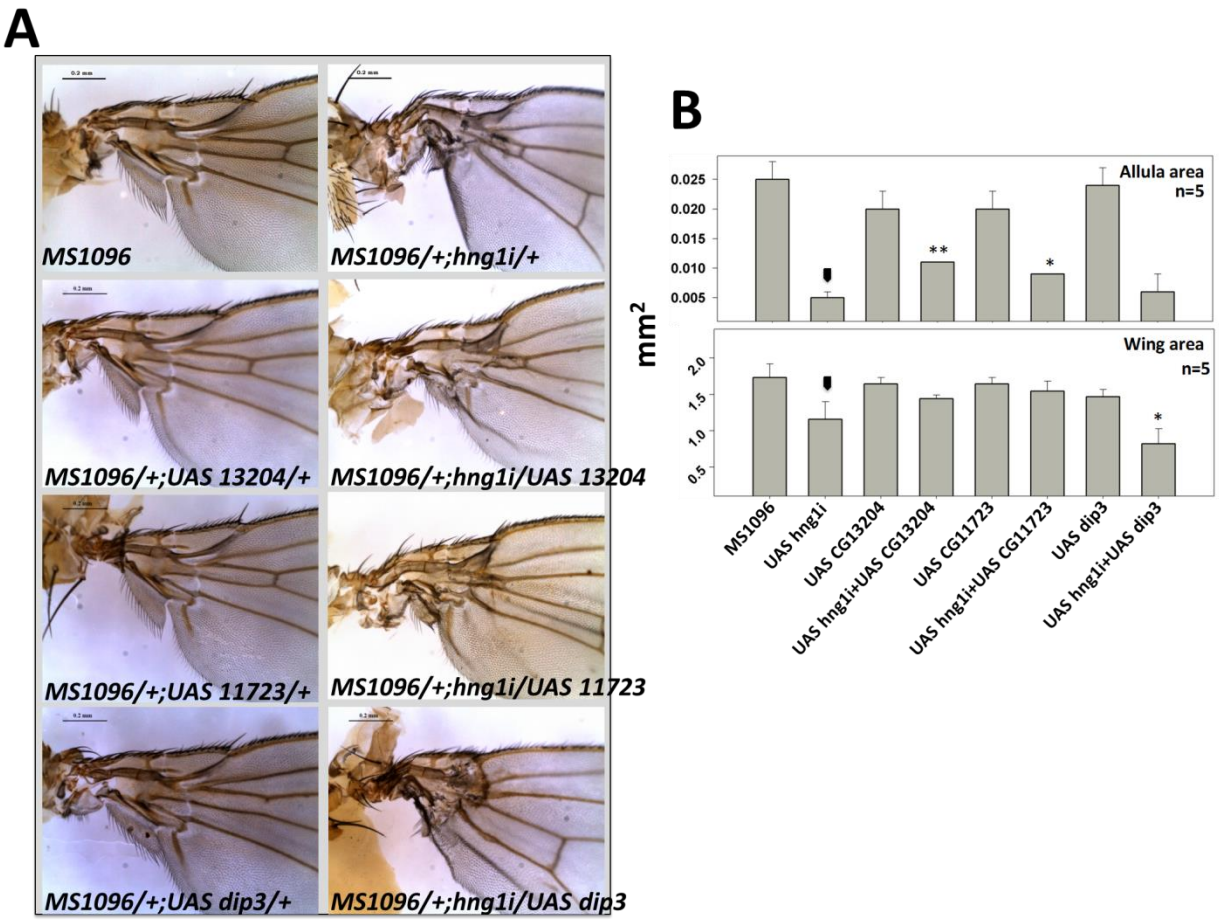


Figure 3.15. *hng1* phenotype is rescued by expression of other MADF-BESS genes. (A) G13204 and CG11723 expression rescue the MS1096/+; UAS*hng1i*/+ phenotype. This indicates that MADF-BESS proteins retain similar biological activity and equivalence in terms of their protein function. (B) Wing size (mm2) and alula size (mm2) are measured for rescue of *hng1* phenotype. Arrowhead indicates the MS1096/+; UAS-*hng1*/+ line used as a control for statistical analyses. *p 0.05 and **p < 0.01.

3.5 Discussion

3.5.1 Are *hinge* genes redundant?

The MADF-BESS gene family appears to retain functional redundancy and has also retained similar amino acid sequence in the two "functional" domains. The MADF and BESS domain families independently have 50 and 25 members, respectively, being in the same ballpark with the largest family in flies, the trypsin gene family, with 111 members (Zhang 2003). It is interesting to note that, in spite of tens of millions of years of evolution, the 16

MADF-BESS genes have retained the N-terminal/C-terminal architectural positioning of the two domains as well as have conserved the domain sequences. The data collected in our study indicate that at least 9 of 16 members participate in the development of the wing hinge. The hinge genes, including *stwl*, appear to have hinge development as a primary function while five other genes appear to retain at least partial function in the hinge. Knockdown experiments where *hng1*, *hng2*, and *hng3* transcripts were reduced singly and in combination indicate that the genes are not genetically redundant in the classical sense of where a single knockdown of a gene has no effect, but a double knockdown has a drastic effect. In our example of the MADF-BESS domain proteins, a subset of these proteins can be said to be partially redundant at the genetic level, but since they all affect the same function and many seem to be equivalent at the protein level, we might consider them as redundant at the protein network level. Another point to note is that our experiments are done using an inbred population grown under a single set of laboratory conditions. It is quite possible that, under different conditions of temperature and diet or in a different genetic background, the genes may be demonstrated to be completely redundant genetically. Based on a lack of a phenotype on single gene knockdown, four other genes in the family do not show the ability to function as the hinge genes, but do enhance *hng1* phenotypes. These genes are probably expressed in wing-hinge development and retain some activity equivalent to the hng genes, but have diverged enough not to be core hinge genes. It is also possible that these are hng genes, but the RNAi lines used are not efficient enough to give phenotypes. In summary, a subset of the members of the family fit a broad definition of redundant genes.

3.5.2 Regulation of Wg/Hth expression is critical for normal wing development

Wg/Wnt is a member of a family of secreted molecules with conserved signaling pathways in animals (Cadigan and Nusse 1997; Wodarz and Nusse 1998; Cadigan 2002; Swarup and Verheyen 2012). Wnt/Wg signaling is activated by binding to receptors such as Frizzled/Arrow, which leads to translocation of stabilized b-catenin to the nucleus and subsequent activation of Wnt target genes. In the Drosophila wing, the dose and spatiotemporal expression of Wg is critical for normal patterning and growth of the wing (Couso *et al.* 1993). Wg is required for proliferation in the first instar and later for patterning in the third instar. During first instar, *tsh* and *hth* are expressed throughout the disc (Zirin and Mann 2004).

Repression of *tsh* by Wg and Dpp in second instar from the pouch is required for the proper development of the wing blade. This is followed by repression of hth and its confinement to the hinge region of the disc (Wu and Cohen 2002). *hth* expression in late third instar is driven by the Wg expression. Combined signals from *vestigial*, *nubbin*, and *rotund* are required for the *wg*-IR expression (Rodríguez Dd *et al.* 2002). *wg*-IR is also driven by an independent mechanism involving a feedback loop with hth in late third instar (Casares and Mann 2000; Rodríguez Dd *et al.* 2002). Wg signaling is regulated at multiple levels, and our data point to roles for the MADF-BESS family as fail-safe regulators for maintaining robust Wg expression. Our data indicate that there appears to be increased activation of the Wg/Tsh/Hth pathway in *hng1*i animals, which in turn leads to the hinge phenotype. In the absence or reduction of *hng1*, Wg is derepressed in the gap region and the Wg spatiotemporal domain broadens. Ectopic expression of Wg earlier was shown to lead to a mispatterned proximal wing (Russell 2000), and our phenotype appears to be of a similar nature. Coop (Song *et al.* 2010), a Pangolin-interacting protein and a member of the MADF-BESS family, has been shown to be a negative regulator of Wnt/Wg signaling, regulating Distalless at the D/V boundary. Our data on the other hand show that *hng1* regulates Wg expression at the presumptive hinge region, with Coop, which is expressed in the wing hinge playing a secondary role. Prima facie, *hng1* appears to regulate Wg expression in the regions that form the wing hinge. At the late third instar, because Wg and hth expression in the wg-IR are dependent on each other, *hng1* may regulate expression of one or both .*hng1* also interacts genetically with a constitutively repressed variant of pan (UAS-TCF-DN), weakly rescuing the pangolin phenotype (Figure 3.11E), indicating similarity to Coop function, and providing evidence for *hng1* regulating Wg signaling. In addition to demonstrating Coop as a negative regulator of Wg signaling, (Song *et al.* 2010) tested Adf1 and CG6854 and showed that these proteins also negatively regulate Wg signaling. This lends support to the idea that MADF-BESS family members in general are involved in regulating Wg expression and or signaling.

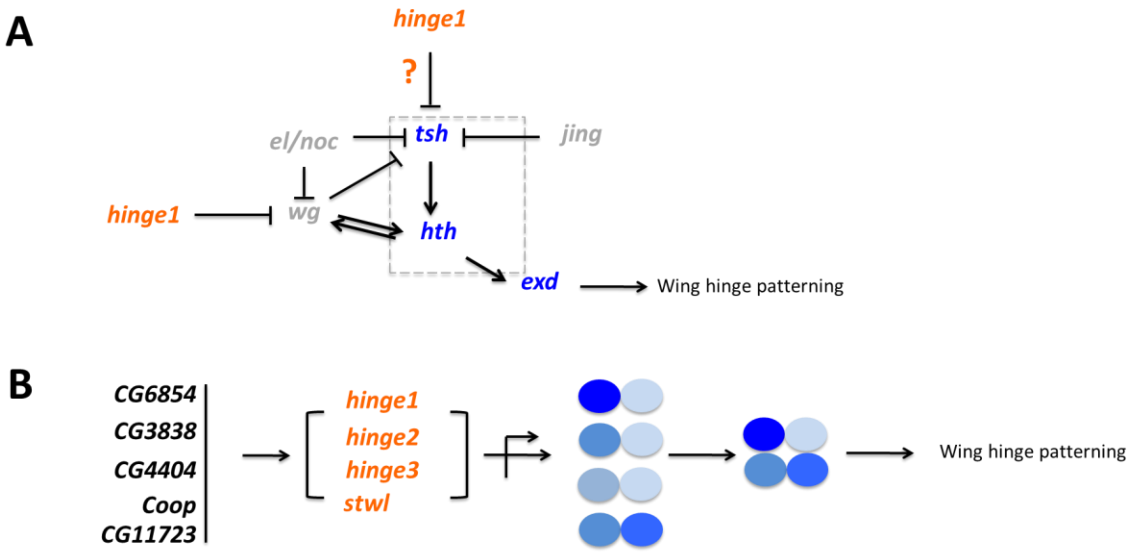


Figure 3.16. Model for MADF-BESS function in the wing-hinge. (A) The GRN for wing-hinge development includes *wg*, *tsh*, *hth*, and *exd* as major patterning genes. In the *hng1* loss of function, our data indicate an increase in activity of Tsh/Hth/Exd. *hng1* appears to negatively regulate the Wg/Hth-positive autoregulatory loop. *hng1* also negatively regulates tsh, possibly acting downstream of jing. (B) The three *hng* genes along with stwl appear to be functionally equivalent and are part of the GRN that patterns the wing hinge. Five additional genes retain, at least partially, functions of the hng family of genes and can replace, to an extent, hng function. The four hinge genes code for proteins (blue circles), which we hypothesize may be part of a dimer/tetramer that is the active transcriptional regulator. Function could be regulated by increasing/decreasing the concentration of the Hng proteins, with the concentration of the functional polymer dependent on spatiotemporal expression and also the levels of the *hng* genes.

CHAPTER 4

Functional redundancy of *brickwall* with *stonewall* in *Drosophila* ovary

---

4.1 Summary

In this chapter, I discuss the discovery of functional roles of one MADF-BESS gene, namely CG3838, in ovary development. I find that loss of function of CG3838 leads to an age dependent fused ovariole phenotype, followed by the loss of the oocyte. Mutants show increase in apoptosis and derailed mitosis. Tissue specific removal of CG3838 from the germarium leads to decrease in the number of germ stem cells and subsequent tumorous ovary. Interestingly, these ovarian defects can be rescued by over-expression of *stonewall,* another MADF-BESS gene, suggesting epistasis or redundancy.

4.2 Introduction

The *Drosophila* adult female has two ovaries, each of which is made up of 15-20 strands of developing egg chambers called as the ovarioles (Figure 4.1). The anterior end of each ovariole contains the germline stem cells (GSCs) enclosed in 'germarium'. As the oocyte develops, it extends towards the posterior end (Figure 4.1). The anterior tip of the germarium comprises of 2-3 germline stem cells which are surrounded by 8-10 terminal filament cells, 5-7 cap cells and inner sheath cells. These cells are responsible for providing the microenvironment or the niche for the maintenance of the GSCs. Each GSC divides asymmetrically forming one germline stem and one daughter cell/cystoblast. The cystoblast undergoes 4 rounds of incomplete mitosis. Since this mitosis occurs with an incomplete cytokinesis, this results in the formation of a 16-cell cyst. One cell out of 16-cell cyst is the future oocyte, and the rest form the 15 nurse cells. As the cyst moves into the middle of the germarium it is surrounded by a monolayer of epithelial cells that arise from the 2-3 somatic stem cells present in the 2b region of the germarium (Song and Xie 2003). Once the encapsulation of the cyst is over (3a region of germarium Figure 4.1), it is pinched off from the germarium forming an egg chamber. This egg

chamber then undergoes 14 stages of maturation, to form a developed oocyte at the posterior end of the ovariole.

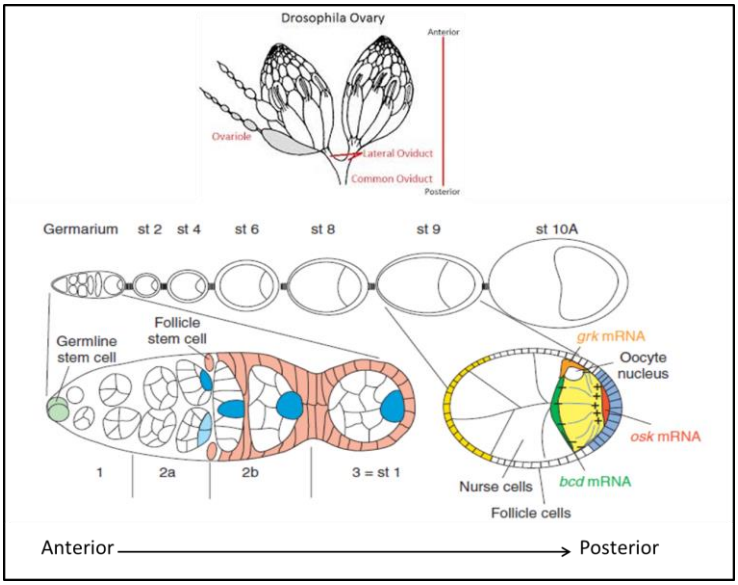


Figure 4.1. Adult Drosophila ovary. An adult female contains two ovarioles each is made up of ovarioles which are developing egg chambers connected by stalk cells. Germarium is located at the anterior end of the ovariole contains germ stem cells which divides to form the 16-cell cyst which when enclosed by a sheath of follicle cells form an egg chamber. At the posterior end of the ovariole is a later stage egg chamber which has the developing oocyte and 15 nurse cells surrounded by a monolayer of follicle cells. **Adapted from** (Roth and Lynch 2009)**.**

During the process of oogenesis as the cyst enters the region 2a (Figure 4.1) it divides mitotically to form a 16-cell cyst. One of these cells starts accumulating transcripts that include *BicD, egl* and *orb*; responsible for the specification of the oocyte (Lantz *et al.* 1994; Mach and Lehmann 1997; Huynh and St Johnston 2000). The egg chamber is mitotically active until stage 6; post which it enters the endocycle state. This transition is brought upon by Notch signaling pathway (Assa-Kunik *et al.* 2007). During endocycle the nurse cells undergo several rounds of replication without cell division. Also, in the later stages the nurse cells transport material to the developing oocyte via ring canals which are important for embryogenesis.

Drosophila germarium is a powerful system to study stem cell maintenance and regulation. One important molecule for GSC renewal and maintenance is *dpp* (decapentaplegic), a member of Bone Morphogenetic Protein (BMP) superfamily (Xie and Spradling 1998). Loss of *dpp* in the niche leads to the decrease of germ stem cells while the overexpression leads to germline tumors (Xie and Spradling 1998). Dpp ligand is produced at the anterior tip of the

germarium by terminal cells (and escort cells) and transduces its effects via the Thickveins (Tkv) and Saxophone (Sax) receptors on the GSCs which results in high expression levels of *daughters against dpp* (*dad*) in GSCs. High levels of dpp in germ stem cells prevents expression of *bam* (bag of marbles) that is required for differentiation of stem cells to a cystoblast lineage (McKearin and Ohlstein 1995). Loss of *bam* results in germline tumors with undifferentiated cells while overexpression leads to early differentiation of GSCs (McKearin and Ohlstein 1995).

As discussed earlier the cystoblast undergoes four rounds of mitosis with incomplete cytokinesis as a result of which the cystoblast/daughter-cells are interconnected by intracellular bridges known as fusomes (Lin *et al.* 1994). Fusomes are made up of proteins like alpha-spectrin and *hu-li-tao-shao* (*hts*) and are derived from the spectrosomes in the GSC (Lin *et al.* 1994). As the cells in cyst divides to form the 16-cell cyst, the fusome also grows to form a branched structure connecting the compartments (see yellow in Figure 4.2). At the end, the cell with most fusome material is specified as the developing oocyte (Cuevas *et al.* 1997; Grieder *et al.* 2000). Ovaries deficient in *spectrin* and *hts* fail to specify oocyte in the early stages of oogenesis (Lin *et al.* 1994; Cuevas *et al.* 1997).

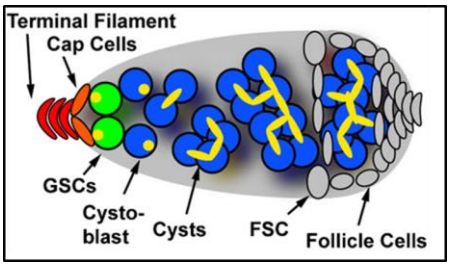


Figure 4.2. *Drosophila* female germarium: Anterior tip of the germarium has the terminal filament cells (in red) and cap cells (in orange) that provide the niche for the GSCs (in green). GSCs asymmetrically divide to form cystoblast (in blue) which undergo four rounds of mitosis to form the 16-cell cysts. The cells in the cysts are interconnected by intracellular bridges called as fusomes (branched structures in yellow). Figure adapted from (Herzig *et al.* 2014).

stonewall (stwl) is a MADF-BESS gene, and essential for GSC development (Clark and McKearin 1996). *stwl* mutants affect oocyte differentiation and GSC division (Clark and McKearin 1996; Akiyama 2002). Stwl maintains GSCs by silencing the differentiation promoting genes, *nanos* and *pumilio*; possibly by interacting with histone modifiers (Maines *et al.* 2007). In this study we focus on another MADF-BESS gene *CG3838* which upon knockdown

shows various ovarian defects and affects oocyte development. We also discuss dynamics between *CG3838* and *stwl.*

4.3 Material and Methods

4.3.1 Drosophila husbandry

All flies were raised at 25ºC in standard corn meal agar. Crosses were set up at 25 ºC. The females of the F1 progeny were dissected for the phenotypes in all cases.

4.3.2 Fly Lines

For ovary and maternal knockdowns, Valium 20 and Valium 22 Transgenic RNAi Project (TRiP) lines procured from Bloomington Drosophila Stock Center (BDSC) were used.
Lines used from BDSC*: UASCG3838i, UAS-CG3919i, UAS-stwli, UAS-CG11723i, UAS-CG8119i, UAS-CG13204i, UAS-adf-1i,* and *UAS-9437i.*
Gal4 drivers used: *nanos-Gal4, maternal alpha-tubulin-Gal4 and GR1 Gal4.*
For detailed information about the fly lines please refer to Appendix 1

4.3.3 Immunostaining and Imaging

Adult ovaries were dissected in chilled 1X PBS and fixed with 4% paraformaldehyde in 1X PBS with 0.3% Triton-X for 20 min at room temperature. They were blocked in 2% BSA, and 0.3% Triton for in 1XPBS for 1 hour; incubated with the primary antibody overnight at 4 ºC; washed 4 times for 10 min in 1X PBS containing 0.1% Triton X and incubated with the appropriate fluorescent secondary antibody for 1 hour at room temperature in the dark. The ovaries were then washed and mounted in Antifade. DAPI (Invitrogen Molecular Probes) was used to stain the nuclei at 1:1000. Primary antibodies used are mouse anti-alpha spectrin 3A9 (1:50), mouse anti-orb 4H8 (1:10), rat anti-vasa (1:50), anti-hnt 1G9 (1:10), anti B-gal 40-1a (1:10) from DSHB, rabbit anti-phospho-histone S10 (1:150), rabbit anti-activated caspase (1:150) from Cell signaling, mouse anti-alpha tubulin (1:500) from Sigma. Mouse anti-C3G (1:500) was kindly provided to us by Prof. Scott Hawley. All secondary antibodies used were from Invitrogen Molecular Probes. Phalloidin 568 and 488 (Invitrogen Molecular Probes) were used to stain actin at 1:100 dilutions. Images were taken on Zeiss 710 LSM confocal microscope and Leica SP8 confocal microscope at 40X and 63X and subsequently processed using Image J software. GraphPad was used for making graphs and for statistical analysis.

4.4 Results

4.4.1 CG3838 is required for ovary development in an age dependent manner

To understand the role of MADF-BESS genes in ovary development, we used the UAS-Gal4 system (Brand and Perrimon 1993; Duffy 2002) to knockdown these genes one by one in the ovary. Valium 20 and Valium 22 lines available in Bloomington Fly were used for this study (see Appendix 1). CG3838 upon knockdown with *nanos*-Gal4 gave a fused ovary phenotype and other oocyte defects (Figure 4.3A). Based on the phenotype and the fact that mutants of one another MADF-BESS gene known as *stonewall* shows similar defects (Clark and McKearin 1996; Akiyama 2002) we named CG3838 gene *brickwall* (*brwl*). About 72% of the total ovarioles (50/69) screened showed the phenotype. Interestingly, this phenotype is age dependent; the ovaries of newly emerged flies do not show this phenotype and flies aged seven days or beyond show this phenotype. This phenotype also enhances as the flies age. Figure 4.3 (B) clearly shows minor to drastic defects from 3^{rd} to 18^{th} day of fly eclosion. CG3838 knockdown ovaries dissected on 10^{th} and 18^{th} day old flies show a drastic change in the ovariole architecture; under-developed individual egg chambers, loss in oocyte specification and apoptotic nuclei are seen in 18-day fly ovaries. Gal4 drivers like maternal alpha tubulin Gal4 and E22C- Gal4 that express in the GSCs show a similar phenotype. Gal4 drivers that express in other somatic cells like GR1-Gal4 (in follicle cells) and c205-Gal4(in cap cells) failed to show such defects which clearly indicates that *brwl*/*CG3838* has tissue specific roles in GSCs.

Knockdowns for other MADF-BESS genes like *CG13897, CG4404, CG9437, CG8119, CG3919, adf-1, CG13204* and *CG11723* using the above Gal4 drivers did not present any phenotype. This shows that *brwl* has a specific role in early development of ovary. Hence, *stwl* and *brwl* are the two genes that are involved in ovary development and the rest of the study focuses on them.

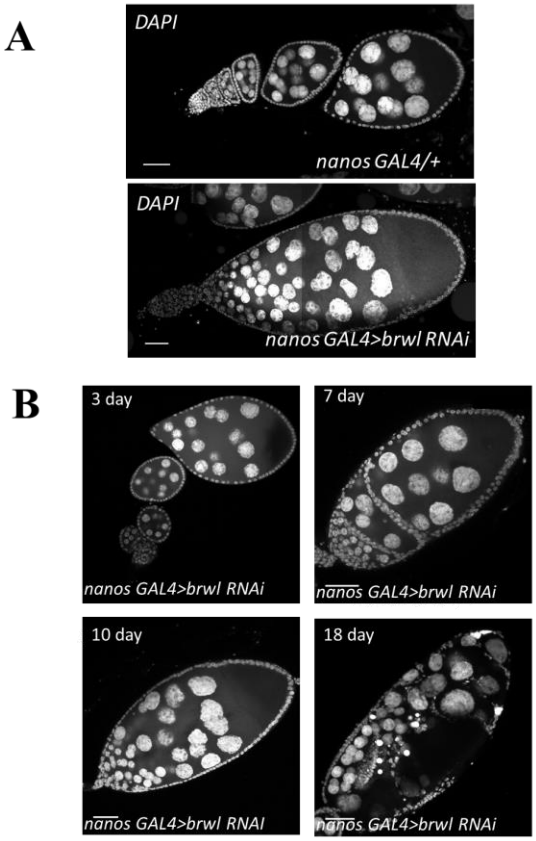


Figure 4.3. *brickwall* knockdown in ovary: DAPI stained ovarioles (A) Reduction of *brwl* transcripts in the ovary of the adult fly by expressing UAS-*brwl* RNAi line, using nanos-Gal4 leads to a repertoire of ovarian defects in the adult female. Defects include a fused ovariole and failure of oocyte specification (B) Observed ovarian defects are age dependent defects are manifested in the adult female after 7 days of fly eclosion and the phenotype enhances as the fly grows older.

In order to substantiate the RNAi phenotypes we searched for loss of function reagents in CG3838 genomic locus. Two P-element insertions, available in BDSC, were obtained as shown in Figure 4.4 (A). The P-element $brwl^{KG00824}$ is inserted at the N-terminal of *CG3838* and is a part of BDGP Gene Disruption Project collection (Bellen *et al.* 2004). The P-element $brwl^{MI054561}$ is inserted at the C-terminal of the gene and is the part of transposon Minos-mediated integration cassette (MiMic) collection (Venken *et al.* 2011). When tested, both $brwl^{KG00824}$ and $brwl^{MI054561}$ females show ovary defects when homozygous (Figure 4.4B). Notably these defects appear more drastic than those seen with the *brwl*RNAi. $brwl^{MI054561}$ exists over a balancer in the stock with very few homozygotes in the population while $brwl^{KG00824}$ is a homozygous stock. Less than 5% of the total ovarioles show the ovary phenotype in heterozygous $brwl^{KG00824}$ /+ and $brwl^{MI054561}$ /+

females (Figure 4.4D). *brwl*^{MI054561} shows a stronger phenotype than *brwl*^{KG00824} (compare second and third panel Figure 4.4B) at the same stage. A strong phenotype is seen in ovaries of trans-heterozygous *brwl*^{KG00824} / *brwl*^{MI054561} females. Based on this result, and to reduce artefactual effect of accumulated background mutations, *brwl*^{KG00824} / *brwl*^{MI054561} genotype is used for all further studies and for characterization of the phenotype.

As in the RNAi experiments, *brwl* mutants also show age dependent phenotype (Figure 4.4C) however in the case of the insertional mutants, the phenotype was seen after 10 days of fly eclosion as opposed to 7 days for that of *nanos-brwlRNAi* (Figure 4.3B). By 18th day the defects are so drastic in case of the mutants that the whole ovarian architecture is lost and the ovarioles seems to be full of undifferentiated cells and appear like tumorous ovarioles (Figure 4.3B). *brwl*^{KG00824} / *brwl*^{MI054561} ovaries when stained with phalloidin show loss of ring canals in 10th and 18th day old ovarioles; which also disturbs the process of oogenesis. DAPI stained nuclei in *brwl*^{KG00824} / *brwl*^{MI054561} ovaries resemble apoptotic/pycnotic nuclei suggesting probable cell death in the tissue. This is reminiscent of *stwl* which has been shown to be involved in maintenance of balance between euchromatin and heterochromatin to prevent the accumulation of DNA damage in presence of replicative stress (Yi *et al.* 2009).

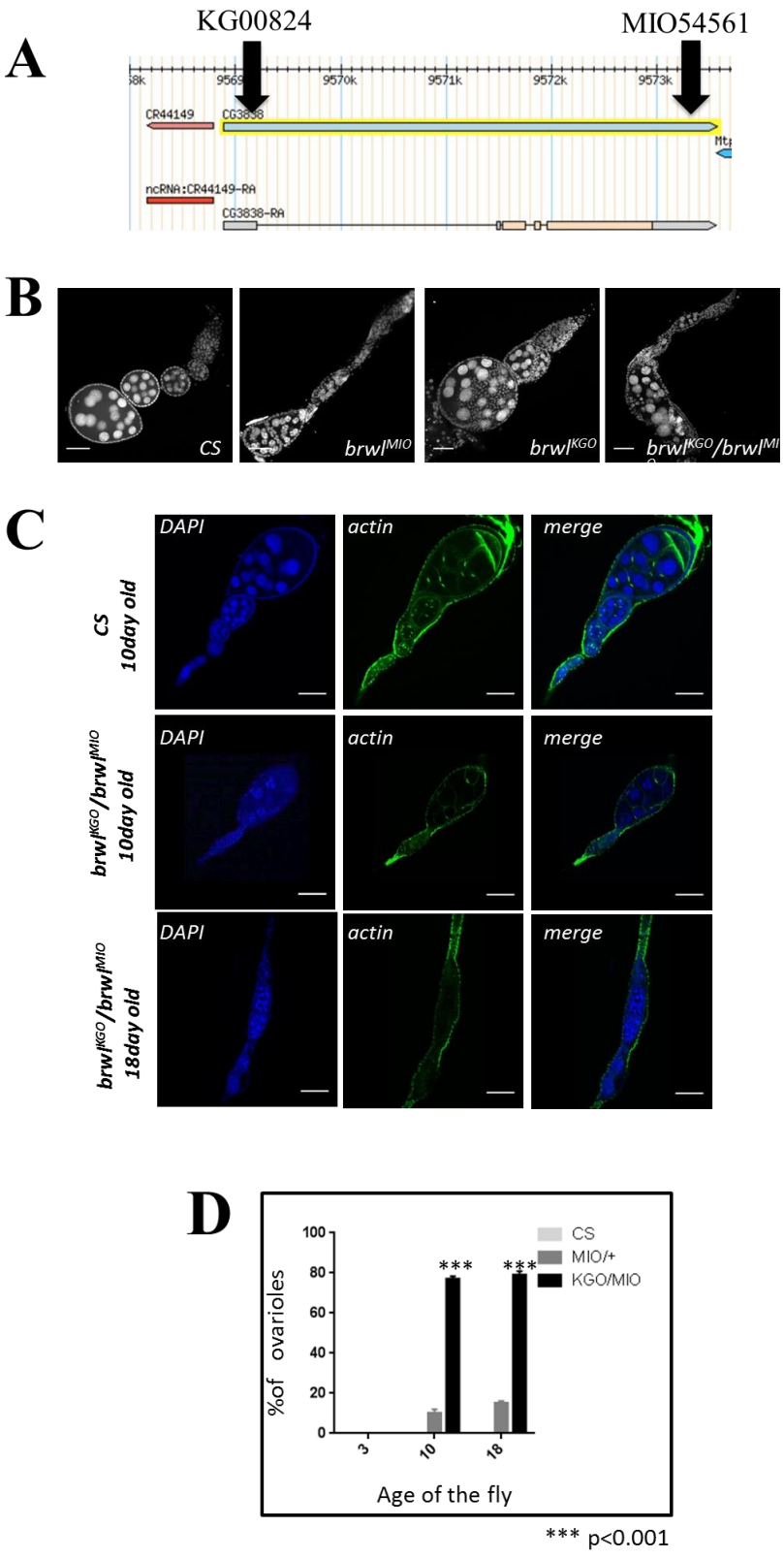

A
KG00824
MIO54561
9569k
9570k
9571k
9572k
9573k
CR44149
CG3838
ncRNA:CR44149-RA
CG3838-RA
B
CS
brwlMIO
brwlKGO
brwlKGO/brwlMI
C
DAPI
actin
merge
CS
10day old
brwlKGO/brwlMIO
10day old
brwlKGO/brwlMIO
18day old
D
%of ovarioles
100
80
60
40
20
0
***
***
CS
MIO/+
KGO/MIO
3
10
18
Age of the fly
*** p<0.001

Figure 4.4. *brickwall* mutants show ovary defects: (A) A cartoon of physical insertion of the P-elements in the *CG3838* gene locus (taken from Flybase). (B) DAPI stained ovarioles. Both $brwl^{KGO}$ and $brwl^{MIO}$ mutants show defective development of ovarioles. A trans-heterozygote of both the mutants shows tumorous like ovarioles at advanced age of the fly. (C) DAPI(blue) stained nuclei and phalloidin (green) stained actin .The age dependent defects are manifested in the adult female after 10 days of fly eclosion and the phenotype enhances as the fly grows older (18day old) and appears like tumorous ovarioles. (D) Quantitative data representing the percentage of ovarioles that show the phenotype on third, tenth and eighteenth day after the fly eclosion.

4.4.2 *brwl* mutants show oocyte defects

Oocyte is specified in the germarium and in its mature form is present at the posterior end of the egg chamber, with 15 nurse cells, both being surrounded by follicle cells. Oocyte develops until stage 14 of oogenesis after which it is secreted in the fallopian tube for fertilization. During the course of oogenesis, nurse cells are tasked with providing the mRNAs and raw materials that are needed for first three hours of development post- fertilization. After the oocyte is secreted the nurse cells are degraded. Proteins like Orb, Egl and BicD that accumulate in the prospective oocyte are essential for oocyte specification and also axis formation. *orb* is required in ovary development at multiple levels- for formation of 16-cell cyst, differentiation of egg chamber and polarity establishment (Lantz *et al.* 1994). Orb is expressed in a crescent shaped pattern in the posterior end of the developing oocyte.

Here $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries were stained with oocyte specific marker like *orb* and they show loss of oocyte specification, mis-orientated oocytes and sometimes loss of oocytes altogether. 79.6% of the total ovarioles showing the phenotype on 10^{th} day, show mis-orientated oocyte and 20.83% show no oocytes (Figure 4.5A & B). By the 18^{th} day 63.33% of the ovarioles showing the phenotype, show no oocytes, while around 36.67% show mis-oriented oocytes (Figure 4.5A &B). This clearly indicates the enhancement of the phenotype and deterioration of the tissue.

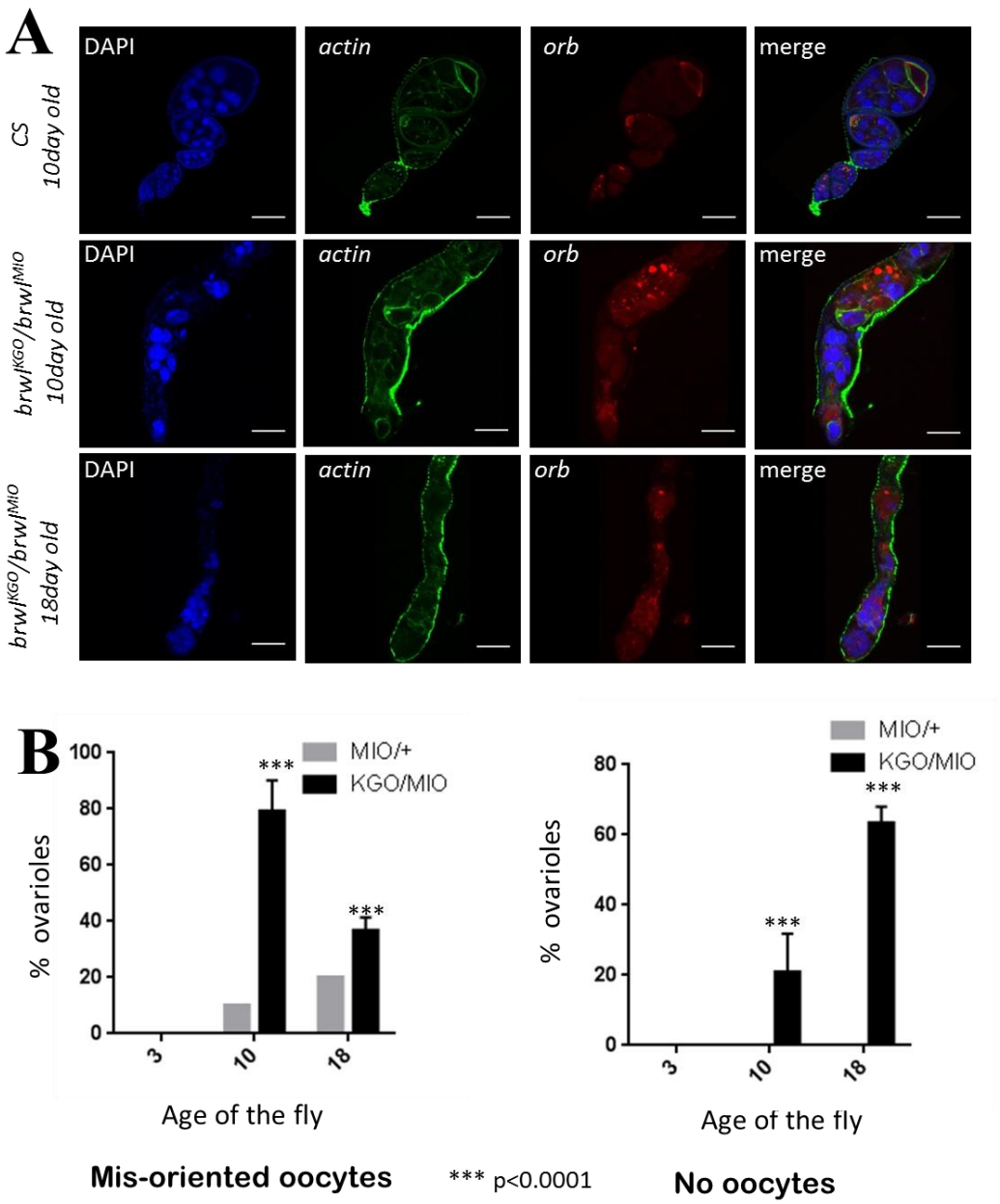


Figure 4.5. *brickwall* mutants show oocyte defects: (A) DAPI (blue), actin (green) and orb (red) stained ovarioles. *brwl*^{KG00824} / *brwl*^{MI054561} mutants display defective ovarioles. These defects were characterized in two major categories mis-oriented oocytes and no-oocytes .These defects are age dependent defects and the phenotype enhances as the fly grows older. 18day old fly shows more number of ovarioles that lacks oocytes. (B) Quantitative data representing the percentage of ovarioles that show the phenotype on third, tenth and eighteenth day after fly eclosion.

4.4.3 *brwl* mutants show caspase activation

DAPI staining of *brwl*^{KG00824} / *brwl*^{MI054561} mutants showed pycnotic nuclei that suggest cell death. To further confirm this finding, *brwl*^{KG00824} / *brwl*^{MI054561} ovaries were stained with the

apoptotic marker *caspase-3*. *brwl* mutants ovaries undergo cell death as shown by staining with the apoptotic marker *caspase-3* (Figure 4.6A). 88.88% (n=39) of ovarioles showing the phenotype are positive for caspase activity on the 10th day after fly eclosion while all the ovarioles showing the phenotype are positive for caspase activity on the 18th day. This drastic increase in cell death most likely contributes to oogenesis arrest in these mutants.

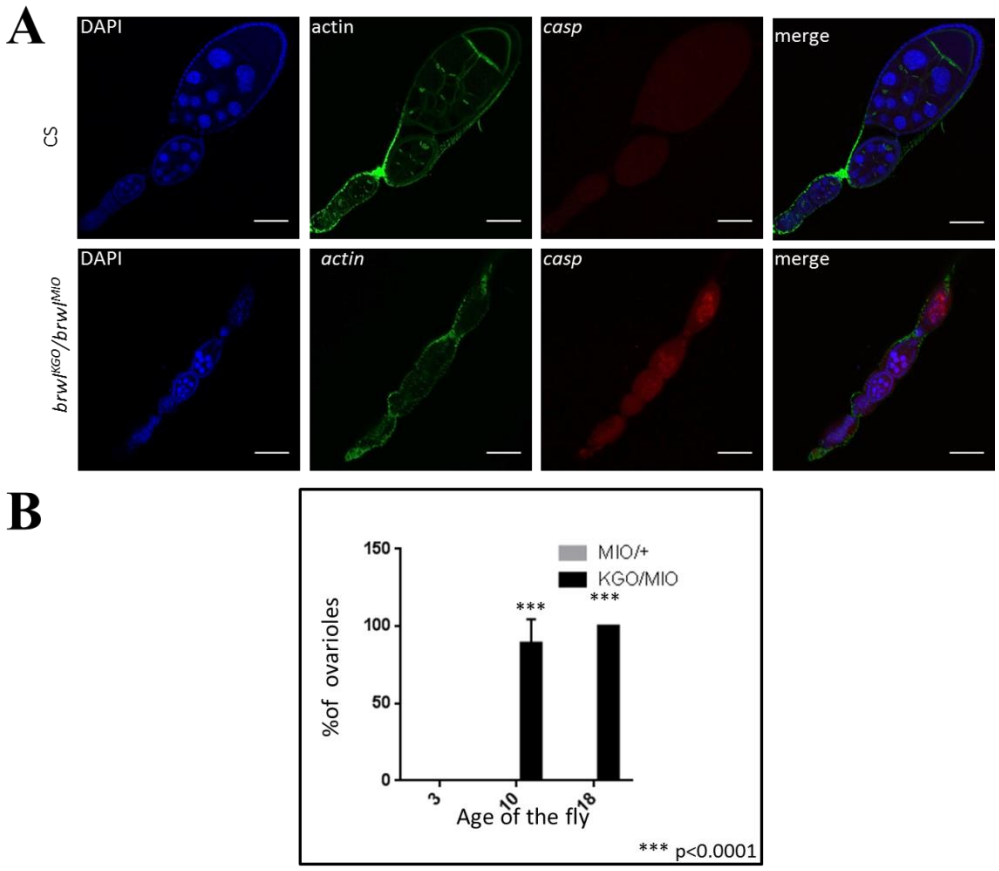


Figure 4.6. *brickwall* mutants show caspase activation: (A) DAPI (blue), actin (green) and activated caspase-3 (red) stained ovarioles. *brwl*^{KG00824} / *brwl*^{MI054561} mutants are positive for caspase activation as stained for apoptotic marker caspase-3. These defects are age dependent defects and the phenotype enhances as the fly grows older (18day old) shows 100% of ovarioles showing phenotype with caspase activation. (B) Quantitative data representing the percentage of ovarioles that show the phenotype on third, tenth and eighteenth day after the fly eclosion.

4.4.4 Mitosis is affected in *brwl* mutants

Mitosis in *Drosophila* ovaries is tightly regulated at multiple levels by different pathways. Mitosis in germ stem cells is regulated by *fused* gene that encodes for a serine-threonine kinase which is required for proper division and differentiation of the GSC. *fused*

mutants show accumulated germline tumors that have undergone five mitotic divisions instead of usual four (Narbonne-Reveau *et al.* 2006). Follicle cell divisions are regulated by Wg/Hedgehog Pathway (Song 2003) (Nystul and Spradling 2010) in somatic stem cells and later by Notch/Delta pathway (Assa-Kunik *et al.* 2007). These follicle cells undergo mitosis until stage 6 of oogenesis after which they enter endocycle and this transition is mediated by Notch/Dleta pathway (Deng *et al.* 2001) (Shcherbata *et al.* 2004). Wild type ovaries when stained for phospho-histone-3 S10 (PH3) show mitotically active cells until stage 6 (white arrows Figure 4.7 top panel) after which they enter endocycle. Interestingly, in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries this pattern of mitosis is lost. In 10 days old ovaries very few mitotically active cells are observed (Figure 4.7 middle panel) whereas by 18th day old random mitotically active cells are seen throughout the ovariole. We also confirmed this by staining the ovaries with *cyclin E* antibody which also shows similar aberrant pattern and is indicative of severe cell cycle defects in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries (Figure 4.8).

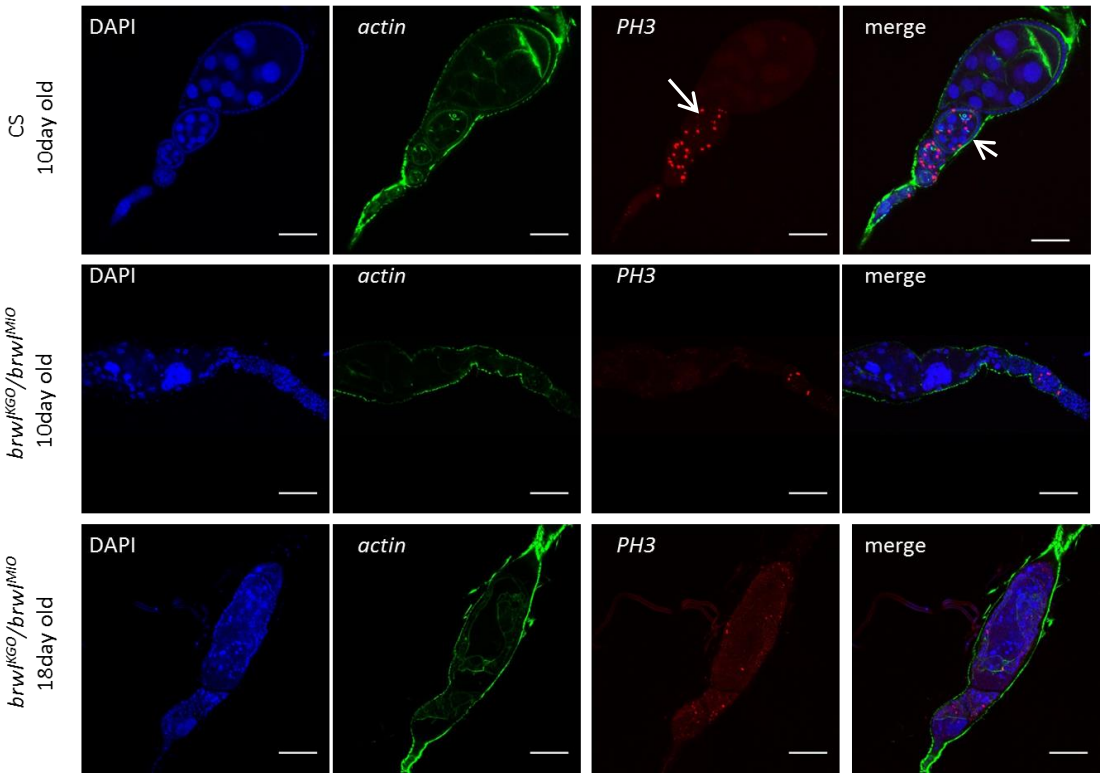


Figure 4.7 *brickwall* mutants affect mitotic cell division: DAPI (blue), actin (green) and phosphor-histone-3 S10 (PH3) (red) stained ovarioles. $brwl^{KG00824}$ / $brwl^{MI054561}$ mutants show decrease in the number of mitotically active cells in 10 day old ovaries. This pattern becomes abrupt and random by 18th day where mitotic cells are seen randomly situated in the ovariole. 100% of ovarioles show this phenotype.

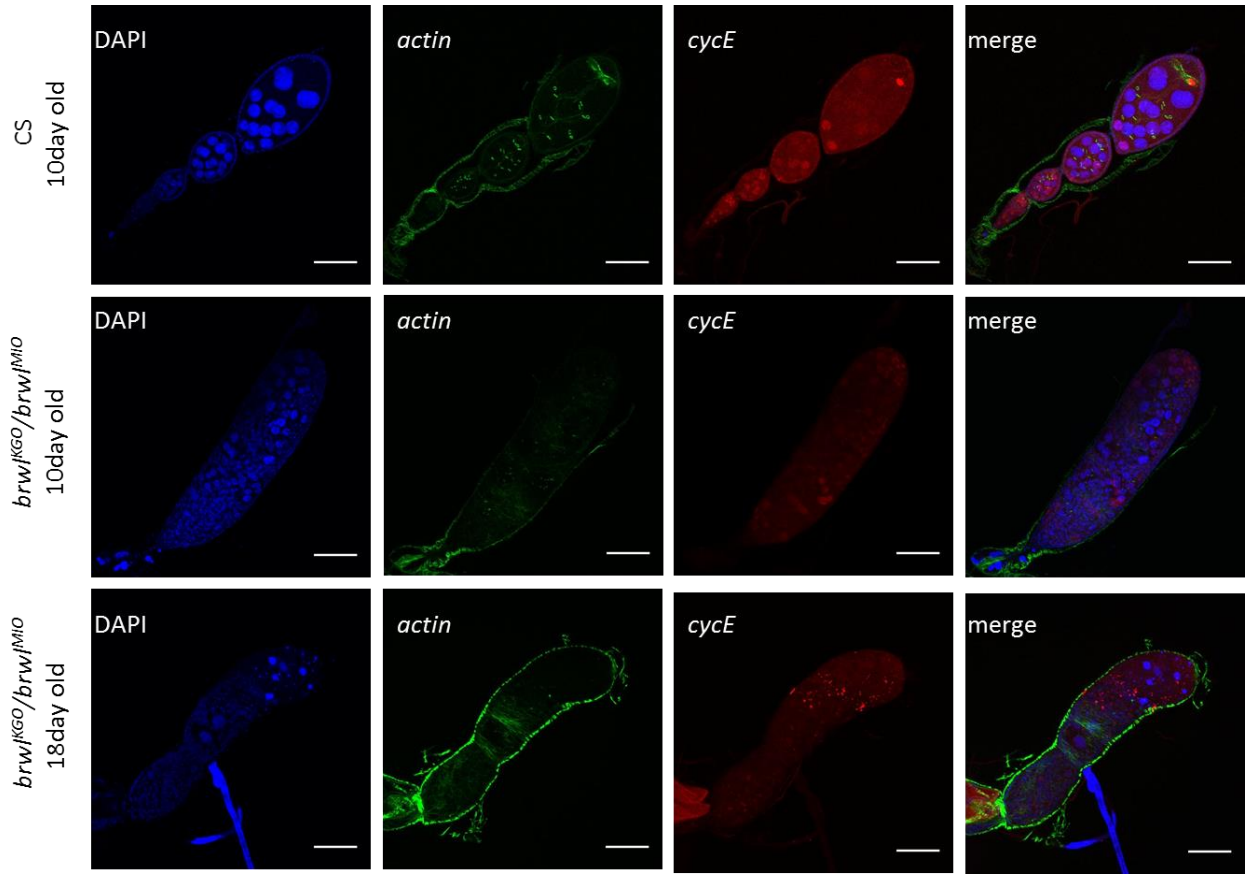


Figure 4.8 *brickwall* mutants affect mitosis as seen by *cyclinE*: DAPI (blue), actin (green) and *cyclinE* (red) stained ovarioles. $brwl^{KG00824}$ / $brwl^{MI054561}$ mutants show decrease in the number of mitotically active cells in 10 day old ovaries. This pattern becomes abrupt and random by 18th day where mitotic cells are seen randomly situated in the ovariole. 100% of ovarioles show this phenotype.

4.4.5 *brwl* mutants affect cyst formation and proper differentiation in the early stages of oogenesis

brwl mutants show severe defects in the germarium and *brwl*-RNAi manifests the phenotype only when driven by Gal4 drivers expressing in the GSCs so we took a closer look at the germarium. Germarium, as described earlier (Figure 4.2), is present, anterior of the ovariole with the GSC at the anterior tip. After asymmetric division, the daughter cell that moves away from the niche gets differentiated and forms the cystoblast which undergoes four rounds of mitosis to form cysts. Fusomes are germline specific organelles that play an important role in cyst formation. In germ stem cells and cystoblast fusomes are round and spherical in shape often

referred to as Spectrosomes (Lin *et al.* 1994). Fusomes are made up of membrane skeletal proteins such as alpha and beta spectrin and adducin-like protein called as *hu-li tao shao (hts)*. As the cells in the cysts divide with incomplete cytokinesis it distributes the fusome material among its daughter cells, as result it forms a more branched structure (Figure 4.2). To see if there are any defects in the cysts formation we stained the $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries with *alpha-spectrin.* Next, we counted the number of spectrosomes and fusomes. In wild type usually there are 2-3 spectrosomes present, however in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries either one spectrosome was present or totally absent in 10 day old flies (Figure 4.9A & B). Similar result was observed for 18th day old ovaries (Figure 4.9 A &B). This depicts decrease in the number of germ stem cells and cystoblast. Number of fusomes present in the wild type is 9-11, however $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries did not show any significant change in the number on the 10th day (6-9) (Figure 4.9 A & B). $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries also showed that the cysts are not organized in a linear pattern in the germarium as compared to the wild type. Instead of 2, 4,8,16 cell cyst organization they were randomly distributed in the germarium. There was a significant decrease in the number of fusomes in 18th day old flies (3-5 fusomes) (Figure 4.9 A & B).

vasa belongs to maternal effect posterior group gene and is localized in the cytoplasm of the pole cells in the embryo. In male and female it is present in the pronurse cell nuclei throughout the oogenesis. It is localized perinuclearly in these nuclei and after stage 8 of oogenesis it is transported to the oocyte (Lasko and Ashburner 1990). Vasa staining in 10 day old $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show that the cysts have lost their organization and are distributed randomly in the germarium (Figure 4.9 A second panel). Vasa stained cysts show considerable decrease by 18th day as the number of cysts decrease (as shown by alpha spectrin staining). DAPI stained 18th day $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show presence of huge nuclei (Figure 4.9A see white arrows third panel). Presence of vasa staining around these nuclei suggests these are cysts might have division defects.

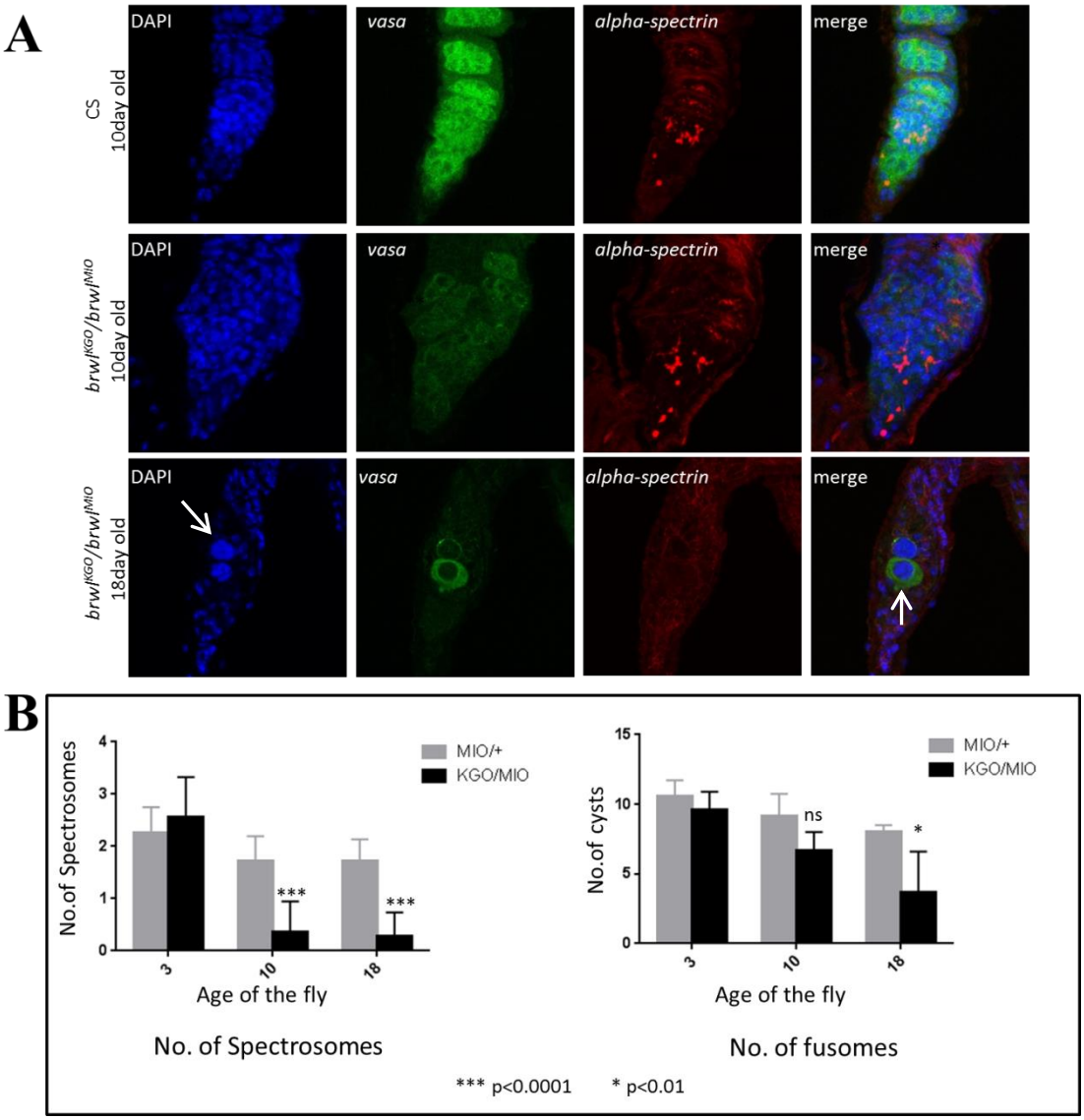


Figure 4.9 *brickwall* mutants affect affect the cyst formation: DAPI (blue), *vasa* (green) and *alpha-spectrin* (red) stained ovarioles. (A) $brwl^{KG00824}$ / $brwl^{MI054561}$ mutants show decrease in the number of spectrosomes in 10 day old ovaries. By18th day the number of fusomes is significantly less. (B) Quantitative data representing number of spectrosomes and fusomes on third, tenth and eighteenth day after the fly eclosion.

4.4.6 *brwl* mutants show decrease in the number of GSCs

Stem cells are responsible for replenishing new cells to replace the old ones. However, germ stem cells like any other cell are subjected to replicative and oxidative stress which over a period of time leads to decline in their activities or numbers. Each *Drosophila* ovariole contain 2-3 germ stem cells which after 21days of fly eclosion start to decrease and by 63 days only 50% of germ stem cells are left (Pan *et al.* 2007). Since $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show drastic

decrease in the number of spectrosomes and cysts as the fly ages a germ stem cell marker *dad-lacZ* was used to stain the GSCs. *daughters against dpp (dad)* is a target gene of dpp and is expressed in GSCs only. In cystoblast *dad* expression is lowered as it moves away from the dpp niche. $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show 50% decrease in the number of *dad-lacZ* positive cells in 10 day old flies (Figure 4.10 A & B).

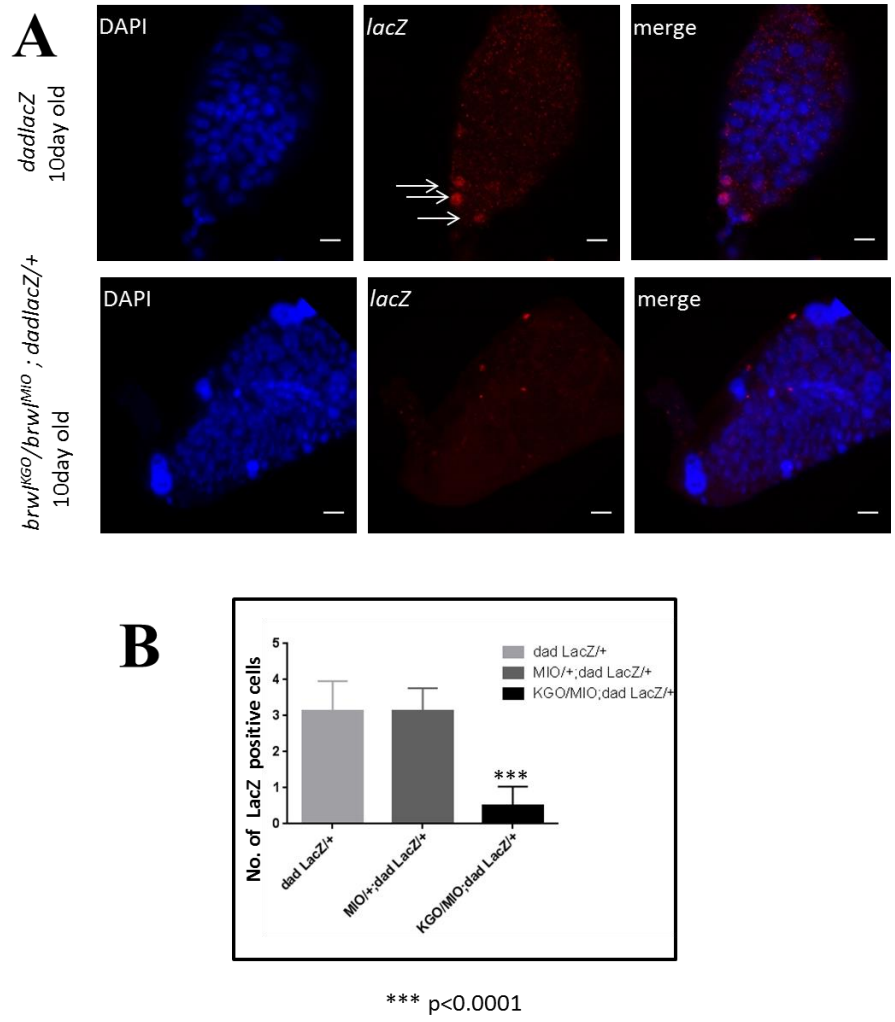


Figure 4.10 *brickwall* mutants show decrease in germ stem cells: DAPI (blue) and *dad-LacZ* (red) stained ovarioles. $brwl^{KG00824}$ / $brwl^{MI054561}$ mutants show decrease in the number of *dad-LacZ* positive cells. (B) Quantitative data representing the number of *dad-lacZ* positive cells 10days after fly eclosion.

4.4.7 *brwl* mutant phenotype is rescued by over-expression of *stwl* in germ stem cells

stwl is a MADF-BESS gene shown to be essential for germ cell development (Clark and McKearin 1996). *stwl* mutants block oocyte differentiation and cause the presumptive oocyte to

develop as nurse cell. It is also required for proper cystoblast differentiation into cysts. *stwl* mutants, similar to *brwl* mutants also show a decrease in the number of germ stem cells and cysts. Overexpression of stwl with nanos-Gal4 in $brwl^{MI054561}$/ $brwl^{MI054561}$ background leads to the rescue of the ovarian phenotype. Egg chambers are formed with 15 nurse cells and a developing oocyte (Figure 4.11 A &B). 32.8% of total $brwl^{MI054561}$/ $brwl^{MI054561}$*; nanos-Gal4/UAS-stwl-HA* ovarioles show ovary defects which is almost 50% reduction with respect to $brwl^{MI054561}$/ $brwl^{MI054561}$ ovarioles. $brwl^{MI054561}$/ $brwl^{MI054561}$*; nanos-Gal4/UAS-stwl-HA* ovarioles show proper arrangement of ring canals in the egg chambers (Figure 4.11A). Overexpression of *stwl* in $brwl^{MI054561}$/ $brwl^{MI054561}$ background rescues the oocyte defects depicted by orb staining (Figure 4.12). Overexpression of stwl with nanos Gal4 by itself does not show any phenotypic defects. Mitotic defects are also rescued in $brwl^{MI054561}$/ $brwl^{MI054561}$*; nanos-Gal4/UAS-stwl-HA* ovarioles as shown by anti-*cyclinE* (Figure 4.13 A) and anti-*PH3 S10* (Figure 4.13B) staining. All the above experiments were performed in 10day old flies. Interestingly, overexpression of *stwl* in $brwl^{MI054561}$/ $brwl^{MI054561}$ background did not decrease the apoptotic activity that $brwl^{MI054561}$/ $brwl^{MI054561}$ ovarioles show (Figure 4.14). This raises the question whether this rescue is just the delay and that accumulation of the ovary defects is slow than the usual (10days in $brwl^{MI054561}$/ $brwl^{MI054561}$ ovaries).

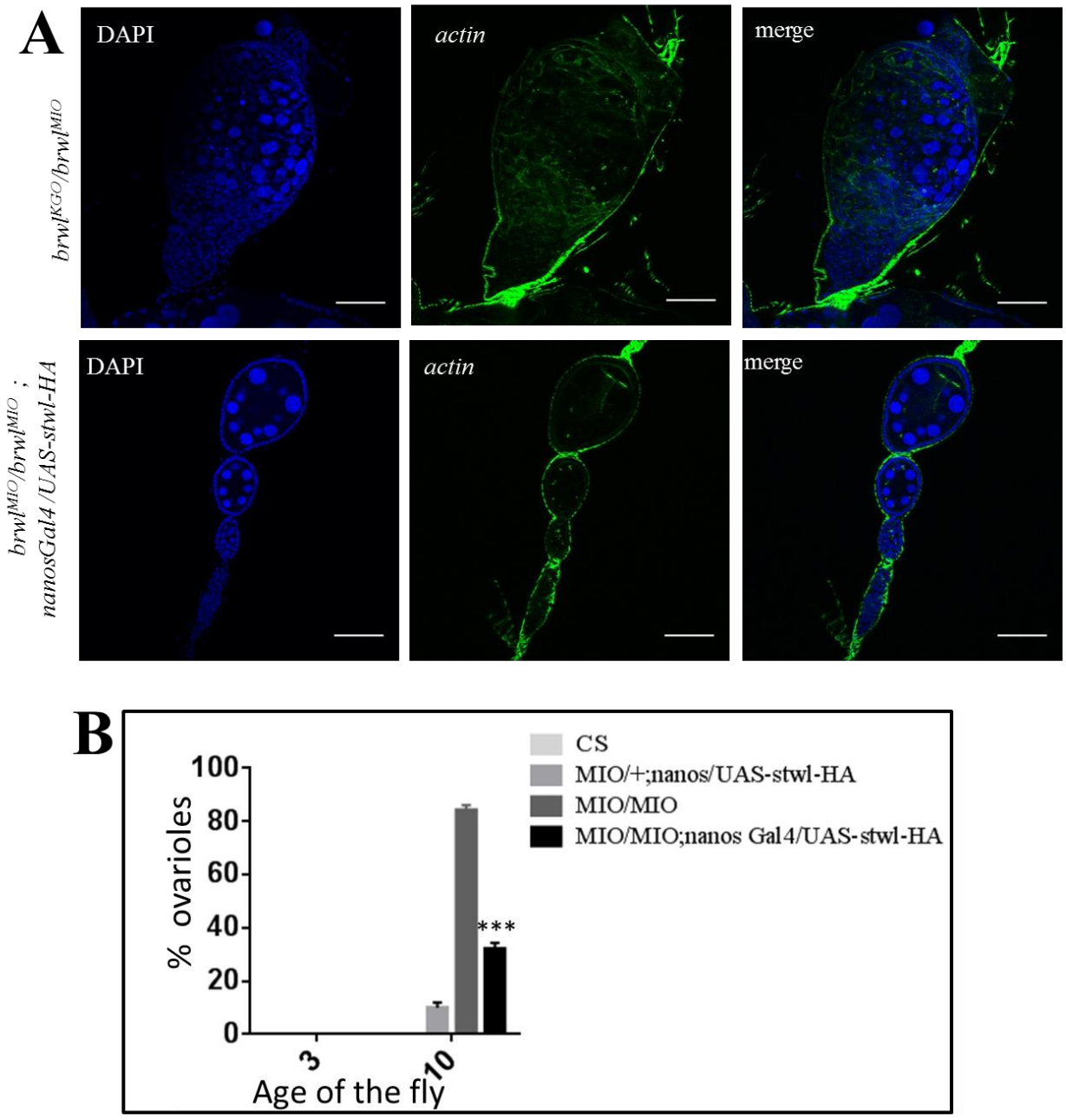


Figure 4.11 *brickwall* mutant phenotype is rescued by over-expression of *stonewall* by nanos Gal4: DAPI (blue) and *actin* (green) stained ovarioles.(A) $brwl^{MI054561}$/ $brwl^{MI054561}$ mutants show ovary defects which are rescued in $brwl^{MI054561}$/ $brwl^{MI054561}$; nanosGal4/ UAS-stwl-HA. (B) Quantitative data representing the number of ovarioles showing the phenotype on 3rd and 10th day after fly eclosion.

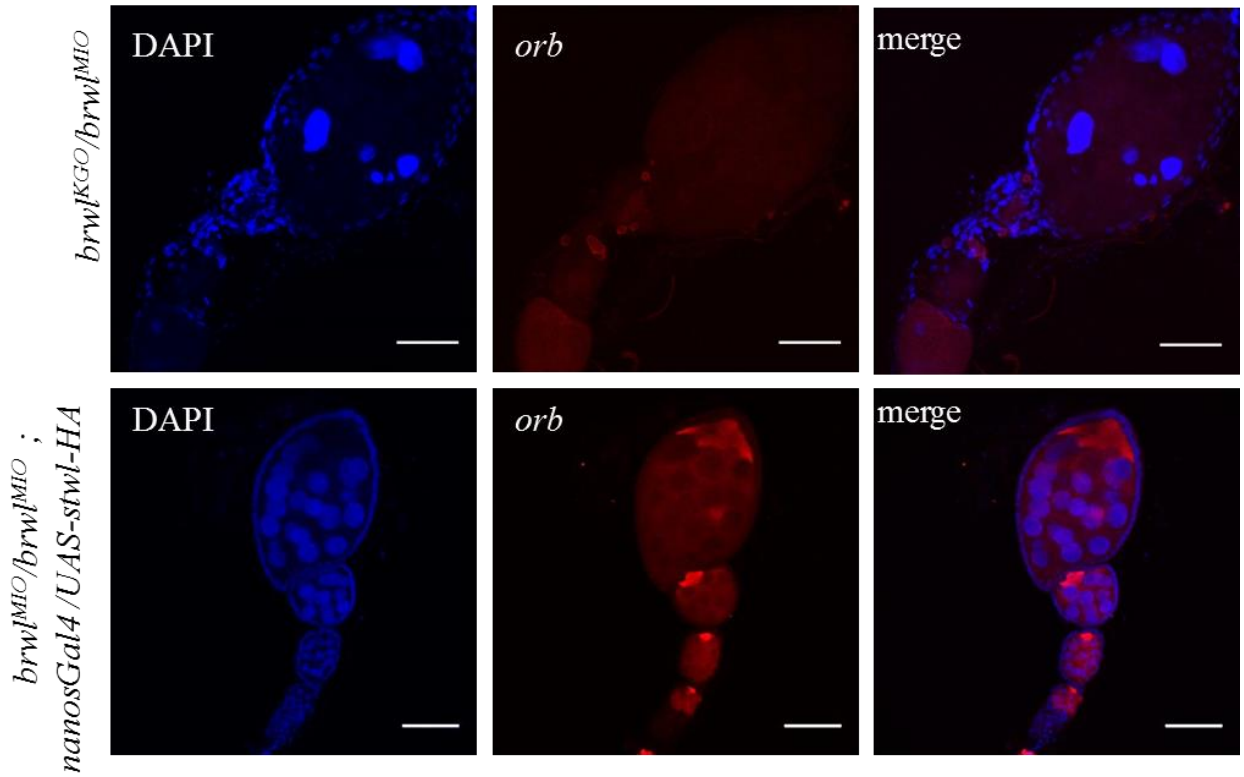


Figure 4.12 Oocyte defects in *brickwall* mutants are rescued by over-expression of *stonewall* by nanos Gal4: DAPI (blue) and *orb* (red) stained ovarioles. $brwl^{MI054561}$/ $brwl^{MI054561}$ mutants show oocyte defects which are rescued in $brwl^{MI054561}$/ $brwl^{MI054561}$; *nanosGal4/ UAS-stwl-HA* ovarioles.

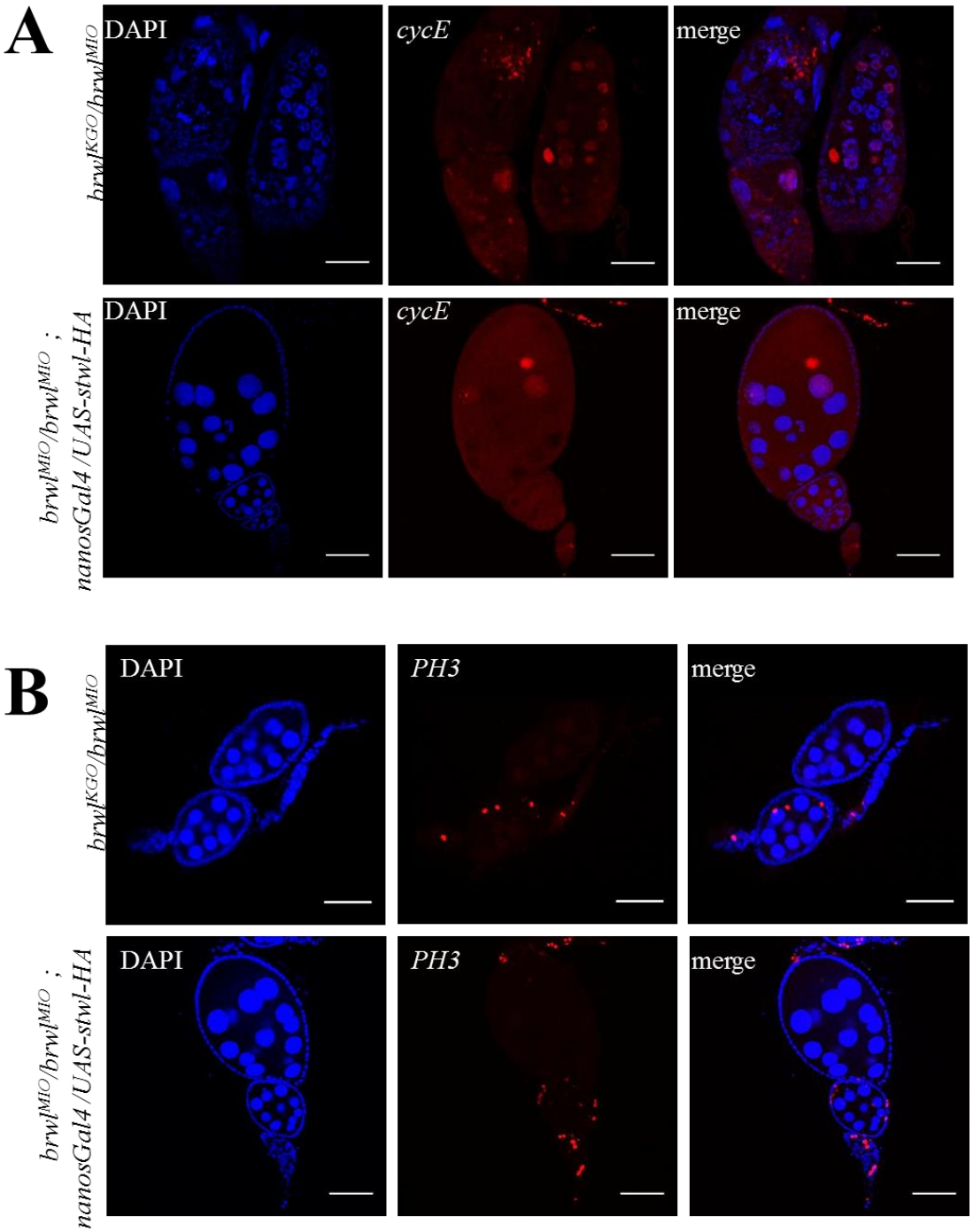


Figure 4.13 Mitotic defects in *brickwall* mutants are rescued by over-expression of *stonewall* by nanos Gal4: DAPI (blue) and (A) *cyclinE* (red) (B) PH3 S10 (red) stained ovarioles.(A) $brwl^{MI054561}$/ $brwl^{MI054561}$ mutants show mitotic arrests and defects which are rescued in $brwl^{MI054561}$/ $brwl^{MI054561}$; *nanosGal4/ UAS-stwl-HA* as shown by *cyclinE* and *PH3* stainings.

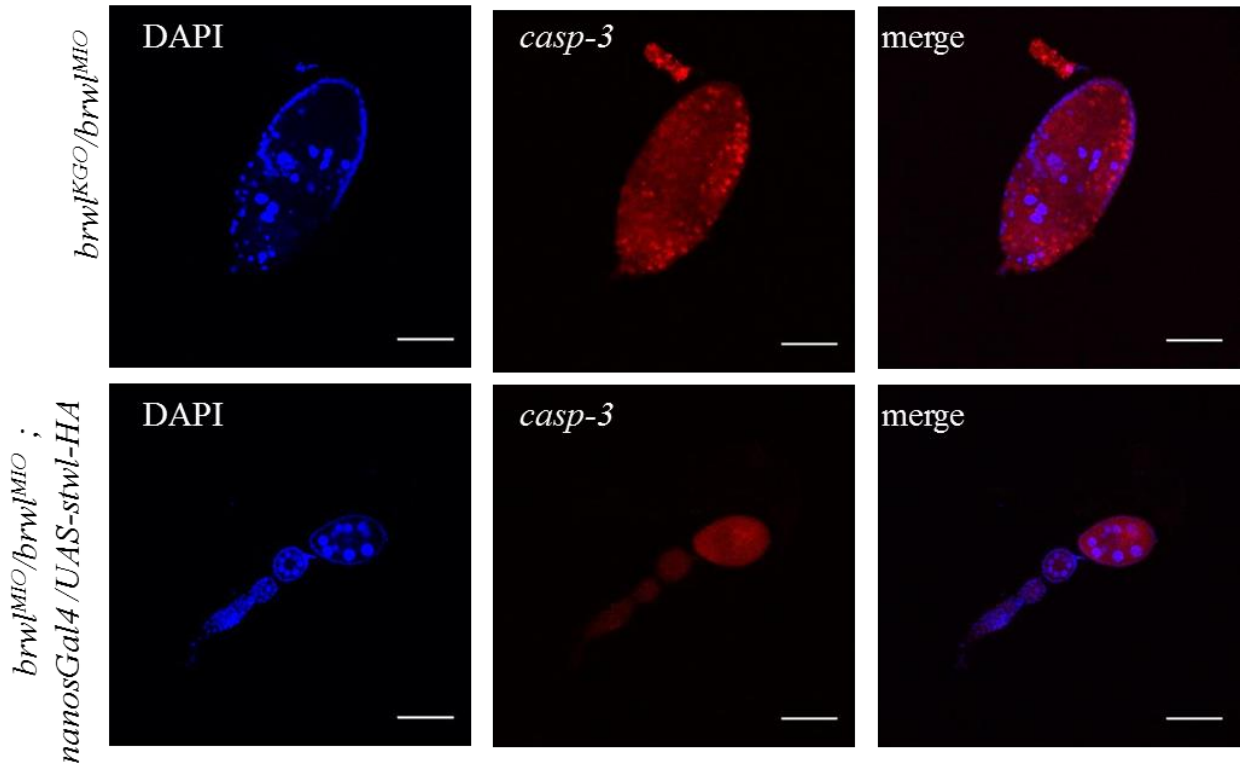


Figure 4.14 Apoptosis in *brickwall* mutants is not rescued by over-expression of *stonewall* by nanos-Gal4: DAPI (blue) and *caspase-3* (red) stained ovarioles. (A) $brwl^{MI054561}$/ $brwl^{MI054561}$ mutants show apoptosis which are not rescued in 10 day old $brwl^{MI054561}$/ $brwl^{MI054561}$; *nanosGal4/ UAS-stwl-HA* ovarioles.

4.5 Discussion

4.5.1 *brickwall* is required for GSC development

Stem cells play a critical role in tissue homeostasis in every organism. Reduction in their number or activity leads to reduction in numbers of critical cells. Stem cells like any other cells are subject to replicative and oxidative stresses. A number of studies in mice and *Drosophila* indicate that stem cell aging has an effect on life span of an organism (Ryu *et al.* 2006; Biteau *et al.* 2010). The intestinal epithelium undergoes degeneration due to overproliferation of intestinal stem cells in aging flies (Biteau *et al.* 2010).

Each *Drosophila* ovary contains 15-20 ovarioles each of which has germ stem cells encased in the germarium at its anterior tip and developing oocyte at its posterior end. Germarium pockets 2-3 germ stem cells surrounded by terminal filament cells and cap cells which are somatic cells. These somatic cells provide a niche for the maintenance of germ stem cells. Germ stem cells undergo asymmetric division and form two daughter cells, the one that moves away from the niche forms the cystoblast which then undergoes four rounds of mitosis and differentiates to form the cysts. These cysts develop to form the egg chambers which are comprised of 15 nurse cells and a developing oocyte enveloped by a monolayer of follicle cells (Spradling *et al.* 2011). Terminal filament cells and cap cells secrete Dpp and Gbb ligands forming the niche for stem cell renewal. The niche also prevents expression of *bam* (*bag of marbles*) which is responsible for stem cell differentiation (McKearin and Ohlstein 1995; Margolis and Spradling 1995; Xie and Spradling 1998). In wild type, the number of germ stem cells decreases by 50% 63 days after eclosion while the cap cells decrease by almost 20% (Pan *et al.* 2007). GSC reduction also leads to decrease in the number of cysts by 30% 63 days after eclosion. This study also show that decrease in the number of cap cells leads to the reduction of Dpp and Gbb secretion as a result decrease in the GSCs and eventually cysts. *dpp* mutants shows 50% reduction in germ stem number 35 days after eclosion (Casanueva and Ferguson 2004; Pan *et al.* 2007). $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries show a 60% reduction in the germ stem cells 10 days after eclosion. The total number of cysts does not show any significant change by 10^{th} day. $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries from 18 day old flies show 40% decrease in the number of cysts most of which are randomly arranged. Failure of proper cyst formation leads to defects in the oocyte specification which is seen in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries. Most of the ovarioles show ectopic Orb expression or mis-localized oocytes. By 18^{th} day most of the ovarioles are devoid of oocytes.

$brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries also show mitotic defects as a result; later stages of oogenesis are arrested and degenerated. Follicle cells undergo cell proliferation until stage 6 of oogenesis. This is important as in the later stages as the egg chamber grows the follicle cells organize around the chamber and a cross talk between the follicle cells and the nurse cells is important for the oocyte axis formation. Most of the $brwl^{KG00824}$ / $brwl^{MI054561}$ ovarioles show disorganized arrangement of the follicle cells around the egg chambers. Presence of pcynotic nuclei and degenerated chambers indicates cell death which was confirmed by the activated

caspase-3 staining in $brwl^{KG00824}$ / $brwl^{MI054561}$ ovaries. This study shows that that MADF-BESS gene *brickwall* plays an essential role in ovary development and reduction of *brickwall* in ovary leads to tissue degeneration.

4.5.2 *stonewall* and *brickwall* genetically interact for proper ovary development.

MADF-BESS family of genes is a family of transcription factors which contain an N-terminal DNA binding domain MADF (**M**yb **Adf**-1 like) domain and a C-terminal protein-protein interacting BESS (**BE**AF-1, **S**tonewall, **S**u (var)) domain. So, far very few members of this family are characterized. Adf-1and Dlip3 function as transactivators (England *et al.* 1992; Cutler *et al.* 1998; Bhaskar and Courey 2002; Ratnaparkhi *et al.* 2008) while Coop, Hng1, Hng2, Hng3 function as repressors for proper wing-hinge development (Song *et al.* 2010; Shukla *et al.* 2014). *stwl* is required for germ cell development and prevents differentiation by epigenetically regulating the differentiation factors (Maines *et al.* 2007). It is required for cystoblast maturation into nurse cells and oocyte specification (Clark and McKearin 1996). Our study finds that similar to *stonewall*, *brickwall* is also required for germ cell maintenance, proper cyst formation and oocyte specification. *brickwall* mutant defects can be rescued by over-expressing *stonewall* using a germ cell driver; this indicates that *brickwall* is a positive regulator and a genetic interactor of *stonewall* in ovary development. The BESS domain can homodimerize or heterodimerize with other proteins with BESS domain. This leads to the possibility of physical interaction between *stonewall* and *brickwall. stwl* knockdown or overexpression of *stwl* mutant in the wing leads to wing defects (Yi *et al.* 2009; Shukla *et al.* 2014). Our previous study shows that knockdown of *brwl* in *hng1* knockdown background enhances the *hng1* knockdown phenotype and shows severe wing defects (Figure 3.14). This clearly indicates that *brwl* and *stwl* have overlapping roles in wing-hinge development. However, the mechanism explaining the interaction between the two genes is yet to be understood. Understanding the interaction between *stwl* and *brwl* in ovary and wing development can provide us with insights into how duplicate genes might have undergone subfunctionalization and are evolutionary stable in the population.

APPENDIX 1

List of *Drosophila* transgenic lines used over the course of this thesis.

Table 1. List of RNAi Lines

S. No.	Name	Genotype <Other Information>	Chr (X/I/II)	Source	Stock No.
1	*UAS-CG9437i(hng1i)*	P{KK103642}VIE-260B	II	VDRC	100101
2	*UAS-CG8359i(hng2i)*	P{KK103584}VIE-260B	II	VDRC	105177
3	*UAS-CG13897i(hng3i)*	P{KK111648}VIE-260B	II	VDRC	108487
4	*UAS-CG13897i(hng3i)*	W[1118];{GD7674}v39733	III	VDRC	39733
5	*UAS-stwli*	P{KK105453}VIE-260B	II	VDRC	102848
6	*UAS-coopi*	w[1118];P{GD6554}v14692	III	VDRC	14692
7	*UAS-CG3838i*	P{KK112405}VIE-260B	II	VDRC	106551
8	*UAS-CG8119i*		III	NIG	8119R-2
9	*UAS-adf-1i*	w[1118]; P{GD1358}v4278	III	VDRC	4278
10	*UAS-dip3i*	P{KK111529}VIE-260B	III	VDRC	107803
11	*UAS-CG13204i*		II	NIG	13204R-2
12	*UAS-CG11723i*	w[1118]; {GD7373}v39723	III	VDRC	39723
13	*UAS-CG6276i*	w[1118];P{GD11876}v30141	II	VDRC	30141
14	*UAS-CG4404i*	w[1118]; P{GD16746}v48183	III	VDRC	48183
15	*UAS-CG3919i*	w[1118];P{GD11108}v38975/TM3	III	VDRC	10978
16	*UAS-tshi*	y^1 sc^* v^1; P{TRiP.HMS01443}attP2	III	BDSC	35030
17	*UAS-hthi*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF02733}attP2/TM3, Sb[1]	III	BDSC	27655
18	*UAS-exdi*	y1 sc v1; P{TRiP}attP2	III	BDSC	34897
19	*UAS-jingi*	y1 sc v1; P{TRiP}attP2	III	BDSC	27024
20	*UAS-tiptopi*	y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMS01527}attP2	III	BDSC	35812
21	*UAS-mirrori*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.JF02196}attP2	III	BDSC	31907

22	*UAS-dppi*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00011}attP2	III	BDSC	33618
23	*UAS-CG3838i*	y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00521} attP2/TM3, Sb[1]	III	BDSC	36785
24	*UAS-CG13897i*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00721} attP2/TM3, Sb[1]	III	BDSC	42765
25	*UAS-stwli*	y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00337}attP2	III	BDSC	35415
26	*UAS-CG4404i*	y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GLC01645} attP2/TM3, Sb[1]	III	BDSC	50527
27	*UAS-CG3919i*	y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00226} attP2	III	BDSC	33355
28	*UAS-CG11723i*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMJ02079}attP40	II	BDSC	42514
29	*UAS-CG13204i*	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL01179}attP2	III	BDSC	42806
30	*UAS-CG8119i*	y[1] sc[*] v[1]; P{y[+t7.7]TRiP.GL00454}attP2/ TM3, Sb[1]	III	BDSC	38893
31	*UAS-CG9437i*	Injected in y[1] v[1] P{y[+t7.7]=nos-phiC31\int.NLS}X; P{y[+t7.7]=CaryP}attP40 <Cloning Vector Used Valium 22>	II	In House	

Note:

1. RNAi lines are abbreviated with an "i" following the name of the gene. For example, *gene-i*.

2. Lines 23-31 were used for ovary and maternal knockdowns

Table 2. Over-expression lines.

S. No.	Symbol	Genotype <Other Information>	Chr #	Source
1	*UAS-Homothorax*	w⁻; UAS-hth/sb	III	LSS Lab
2	*UAS-CG9437*	w⁻; UASp-CG9437	II,III	Transgenic/In-House
3	*UAS- CG9437 mcherry*	w⁻;UASp-CG9437:mcherry	II,III	Transgenic/In-House
4	*UAS-CG13204*	w⁻;UAS-CG13204-HA-FLAG/Sb	III	Transgenic/In-House
5	*UAS-CG11723*	w⁻;UAS-CG11723-HA-FLAG	II	DPiM
6	*UAS- CG3838*	w⁻;UAS-CG3838-HA-FLAG	III	DPiM
7	*UAS-Dip3*	w⁻;UAS-Dip3-HA-FLAG	II	DPIM
8	*UAS-CG3919*	w⁻;UAS-CG3919-HA-FLAG	III	DPiM
9	*UAS-Exd*	w⁻;UAS-exd-HA-FLAG/sb	III	DPIM
10	*UAS-Jing*	UAS-Jing	II	NCBS/KVR Lab
11	*UAS-CG3838*	M{UAS-3838.ORF.3xHA.GW}ZH-6Fb	III	FlyORF
12	*UAS-Stwl*	UAS-stwl-HA-FLAG	III	DPiM

Table 3: Mutants and Deficiencies

S.No.	Gene Name	Genotype	Chr No.	Source	No.#
1	*CG9437*	Df(BSC484)	II	BDSC	24988
2	*CG8359*	Df(BSC306)	II	BDSC	25010
3	*nubbin*	nub[E37]/CyO, P{ry[+t7.2]=ftz/lacB}E3	II	BDSC	8856
4	*vestigial*	vg[1]	II	BDSC	432
5	*pangolin*	y[1] w[1118]; P{w[+mC]=UAS-pan.dTCFDeltaN}4	II	BDSC	4784
6	*CG3838*	y[1] w[67c23]; P{y[+mDint2] [BR.E.BR]=SUPor-	II	BDSC	12901

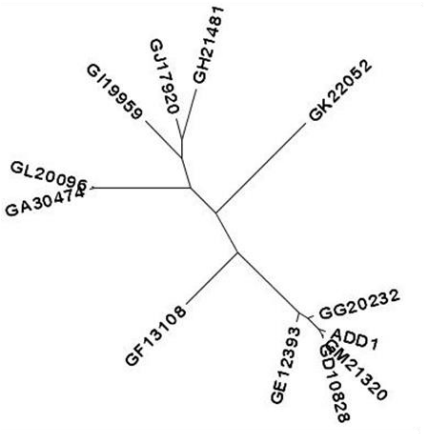
		P}CG3838[KG00824]			
7	*CG3838*	y[1] w[*]; Mi{y[+mDint2]=MIC}CG3838[MI05456]	II	BDSC	42326
8	*stwl*	y[1] w[*]; P{w[+mC]=lacW}stwl[j6C3], l(3)j6C3[j6C3]/TM3, Sb[1]	III	BDSC	12087

Table 4: Gal4 drivers Used

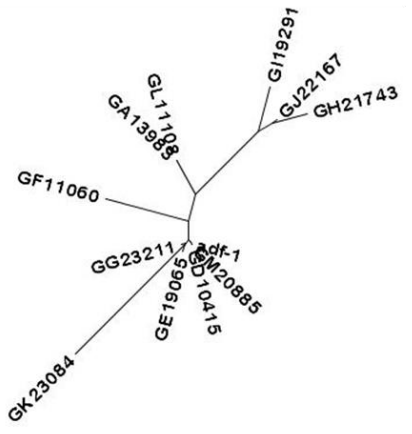
S.No.	Symbol	Genotype	Chromosome	Source	No.#
1.	*MS1096*	P{GawB}BxMS1096	X	BDSC	8860
2.	*daughterless*	w[*]; P{w[+mW.hs]=GAL4-da.G32}2; MKRS/TM6B, Tb[1]	III	BDSC	55851
3.	*patched*	w[*]; P{w[+mW.hs]=GawB}ptc[559.1]	II	BDSC	2017
4.	*scalloped*	w[*] P{w[+mW.hs]=GawB}sd[SG29.1]	X	BDSC	8609
5.	*vestigial*	w*; P{GAL4-vg.M}2; TM2/TM6B, Tb1	III	BDSC	6819
6.	*apterous*	y[1] w[1118]; P{w[+mW.hs]=GawB}ap[md544]/CyO	II	BDSC	3041
7.	*omb*	y[1] w[1118] P{w[+mW.hs]=GawB}bi[md653]	X	BDSC	3045
8.	*nanos*	P(Gal4::VP16-nos.UTR)CG6325[MVD1])	III	BDSC	31777
9.	*GR1*	P{GawB}GR1	III	BDSC	36287
10.	*maternal alpha tubulin*	y w; P(mat-tub-Gal4)mat67; P(mat-tub-Gal4)mat15 (line 2318)	II,III	Daniel St. Johnston	15.4 & 13.4

APPENDIX II

Figure 1. Phylogenetic Trees for all MADF domain containing genes: Maximum Likelihood Method Based Phylogenetic Trees for MADF domain genes in *Drosophila* constructed using IQTree Software (Trifinopoulos *et al.* 2016).

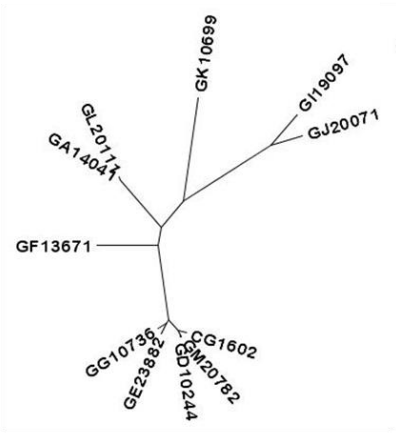


ADD1

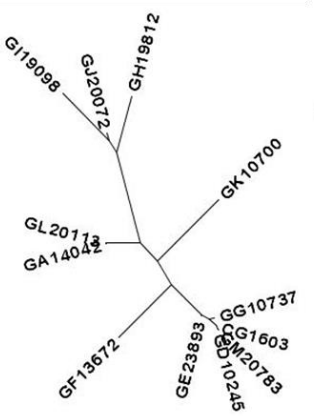


adf-1

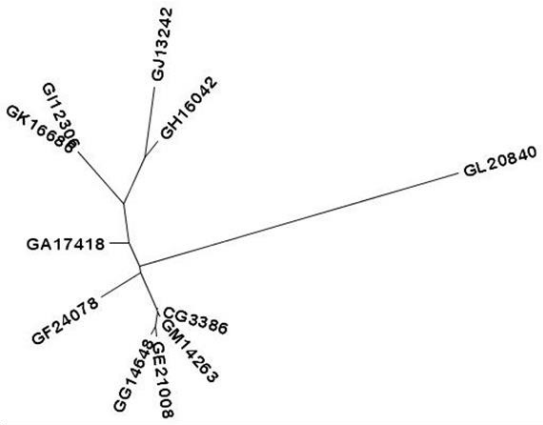
az2



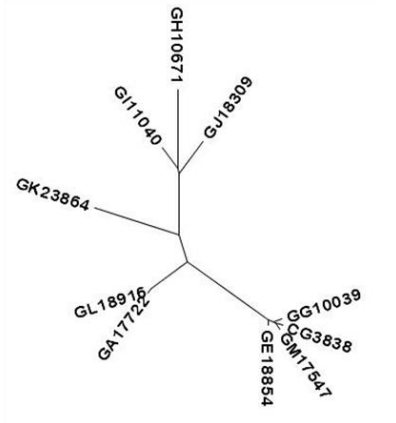
CG1602



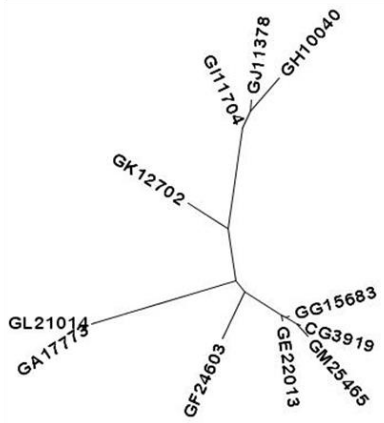
CG1603



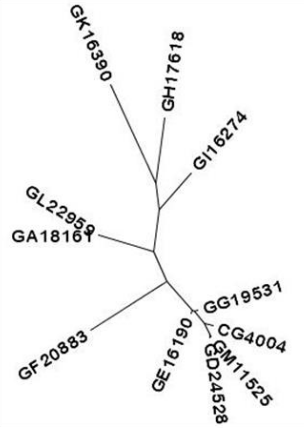
CG3386



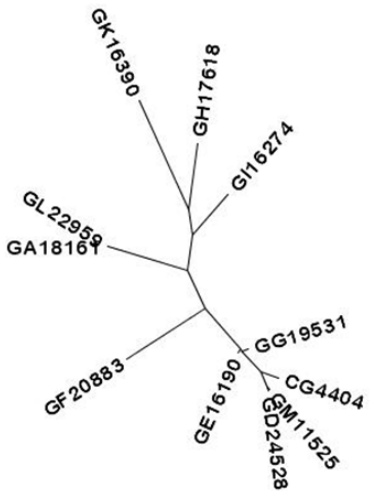
CG3838



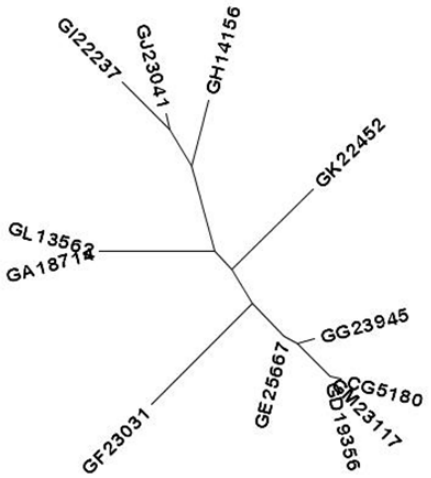
CG3919



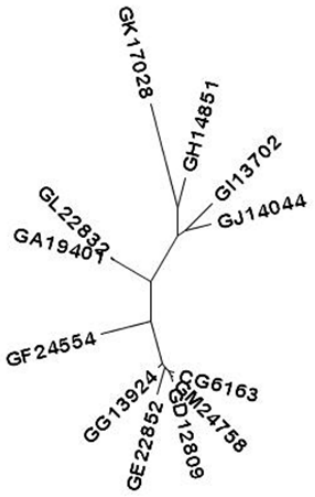
CG4004



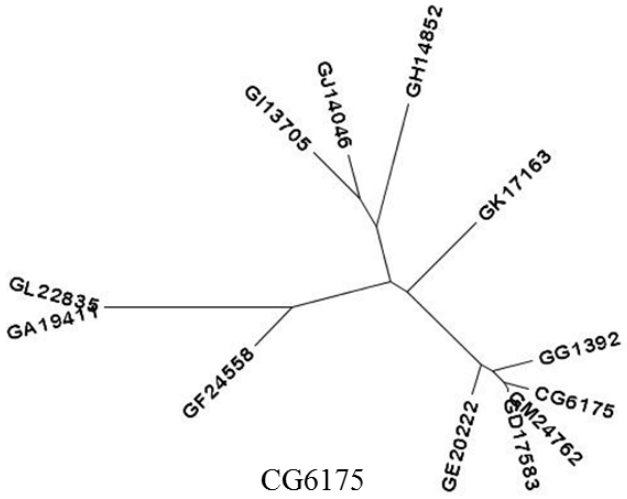
CG4404



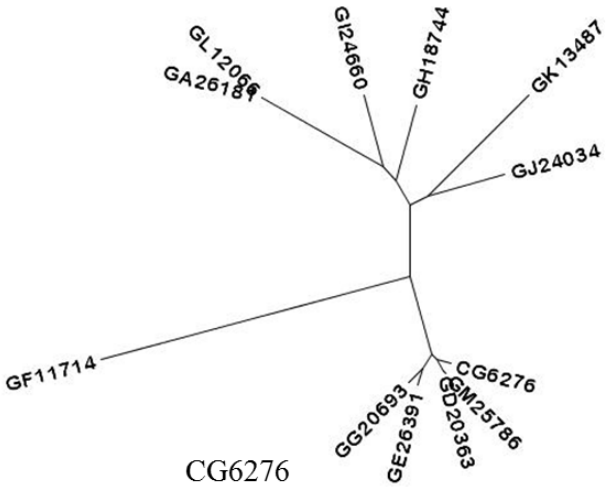
CG5180



CG6163

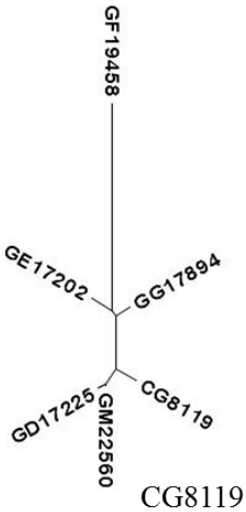


CG6175

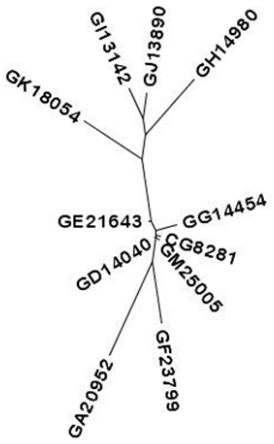


CG6276

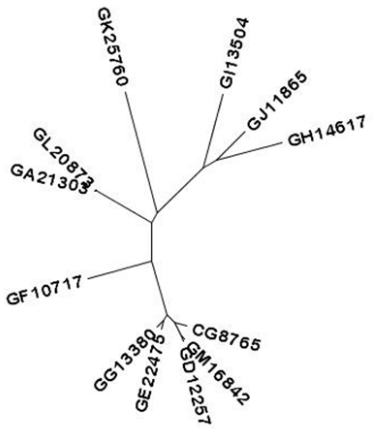
CG7745



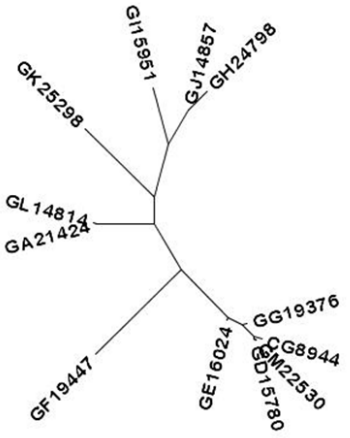
CG8119



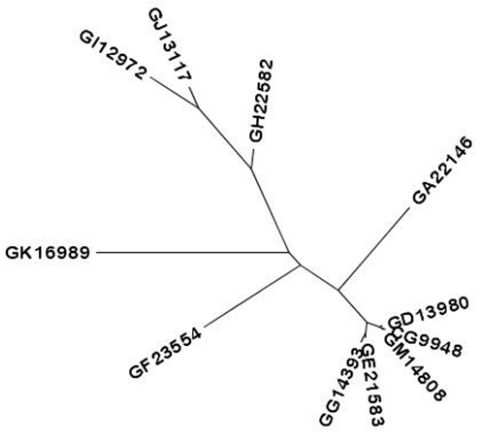
CG8281



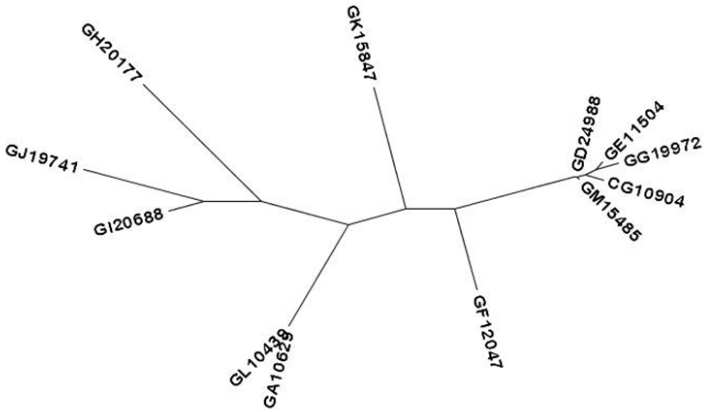
CG8765



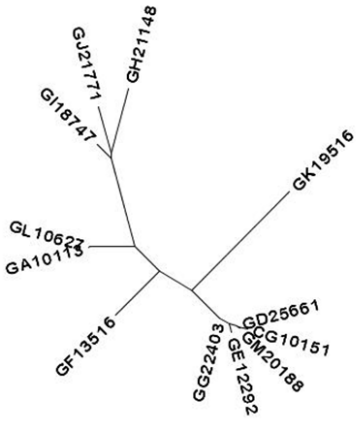
CG8944



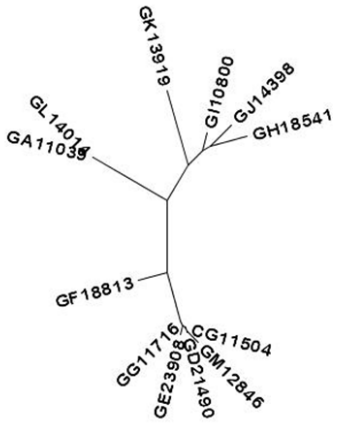
CG9948



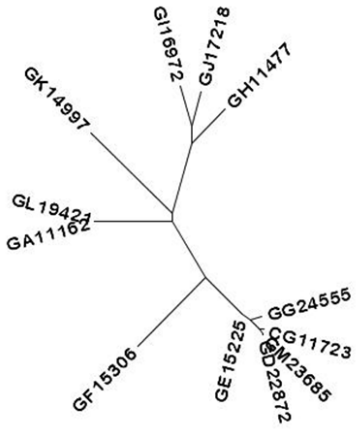
CG10904



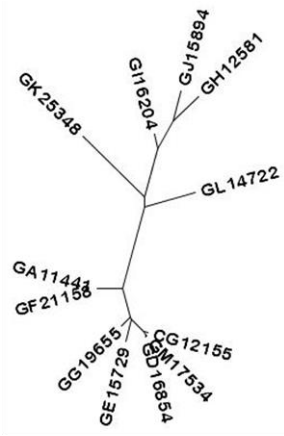
CG10151



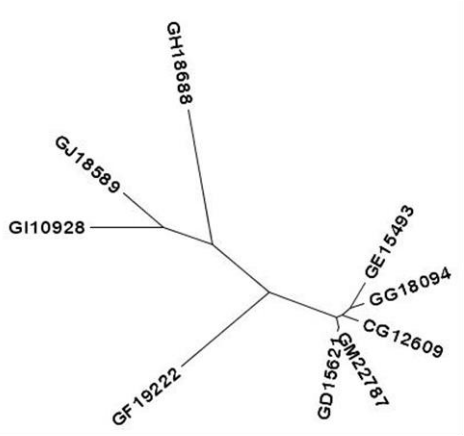
CG11504



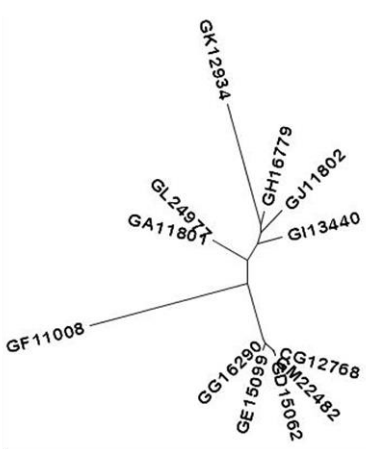
CG11723



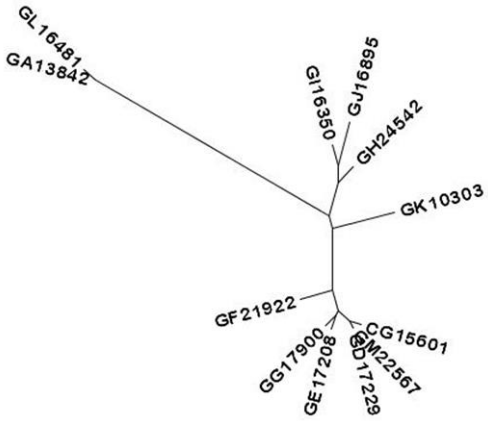
CG12155



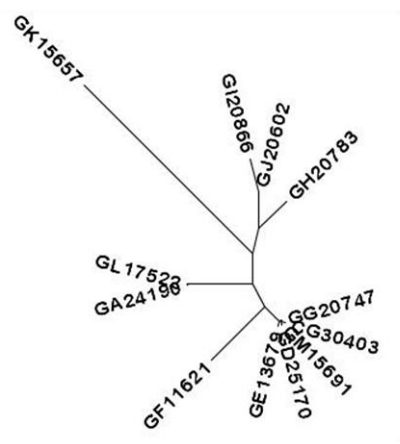
CG12609



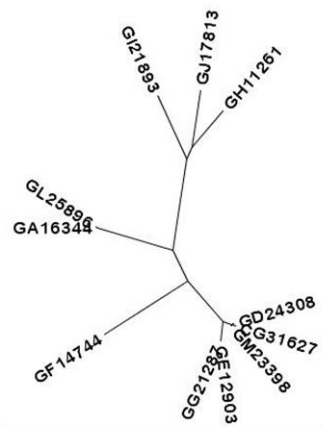
CG12768



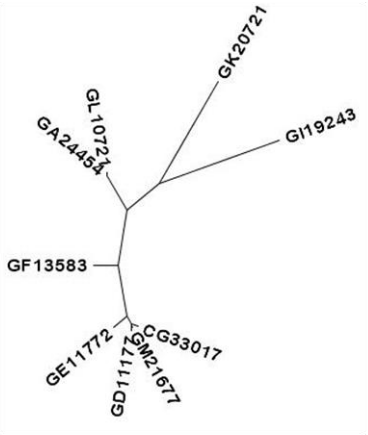
CG15601



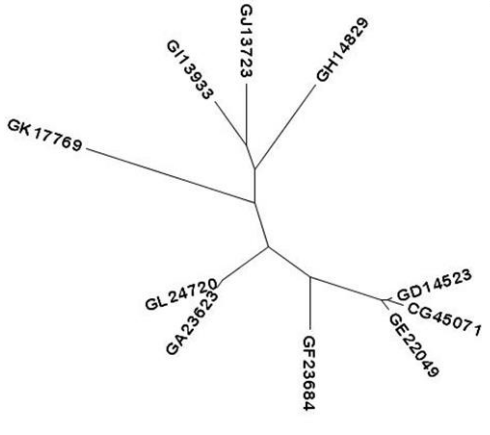
CG30403



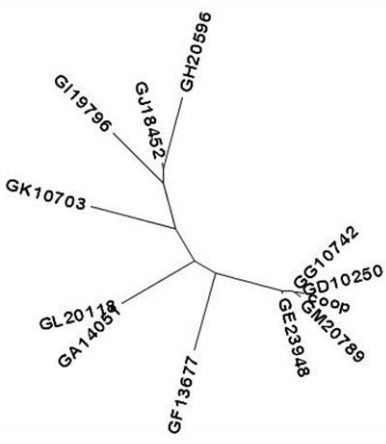
CG31627



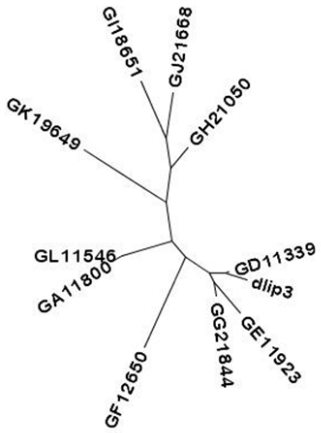
CG33017



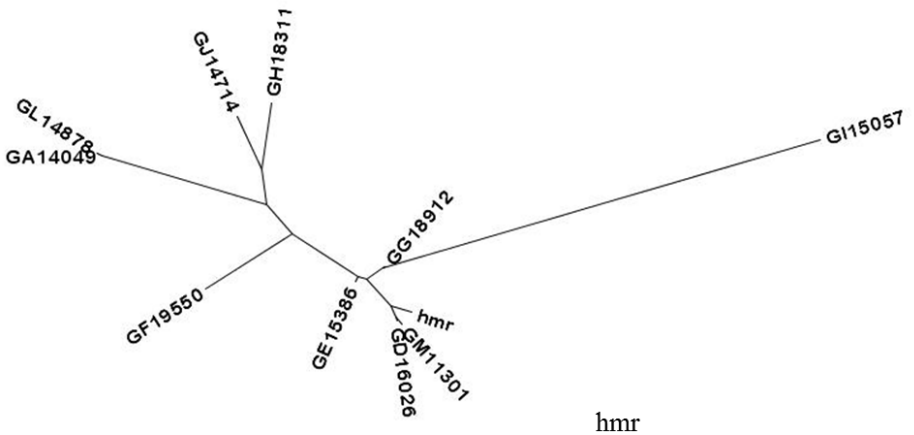
CG45071



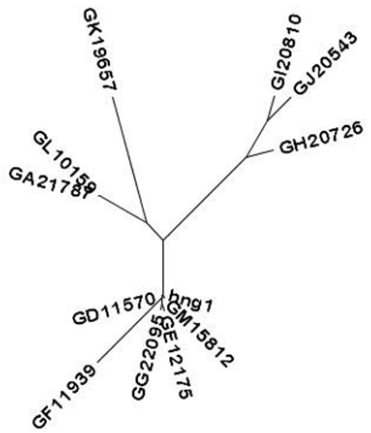
ccop



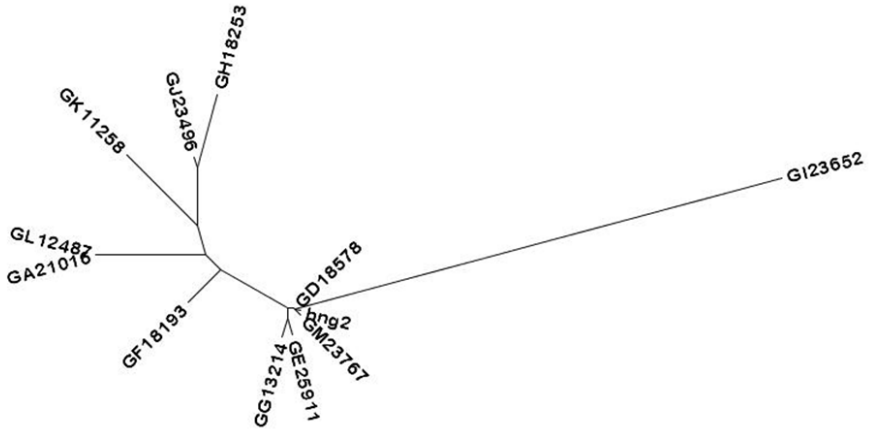
dlip3



hmr

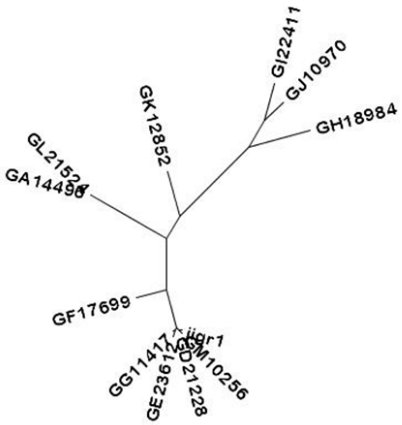


hng1

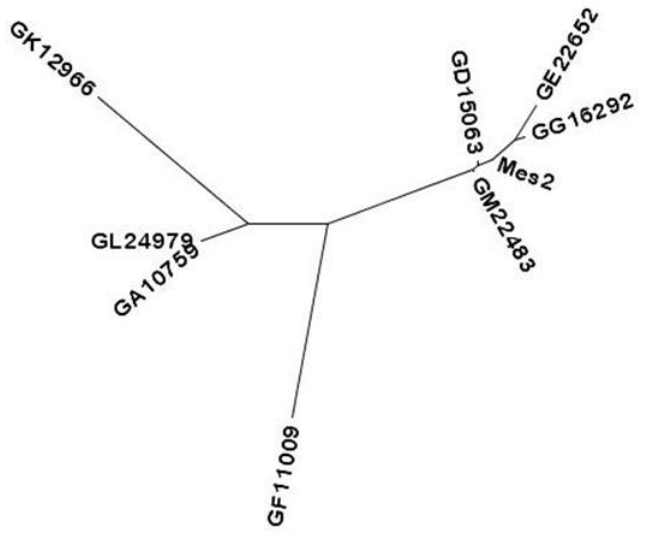


hng2

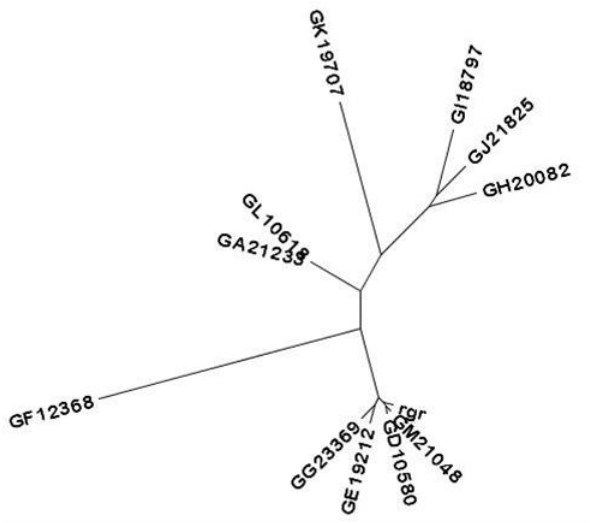
hng3



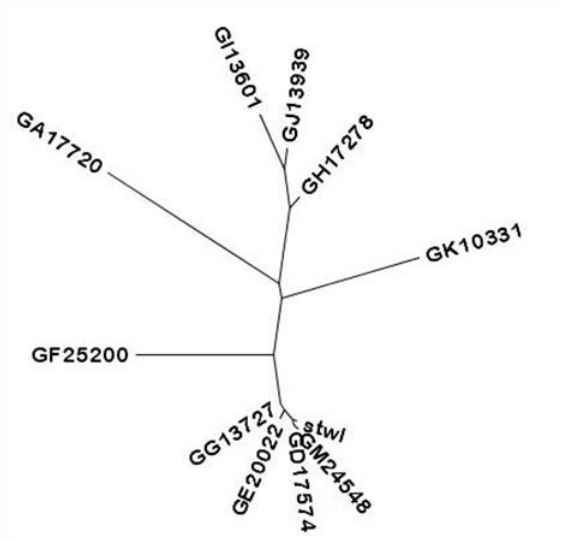
jigr-1



mes2

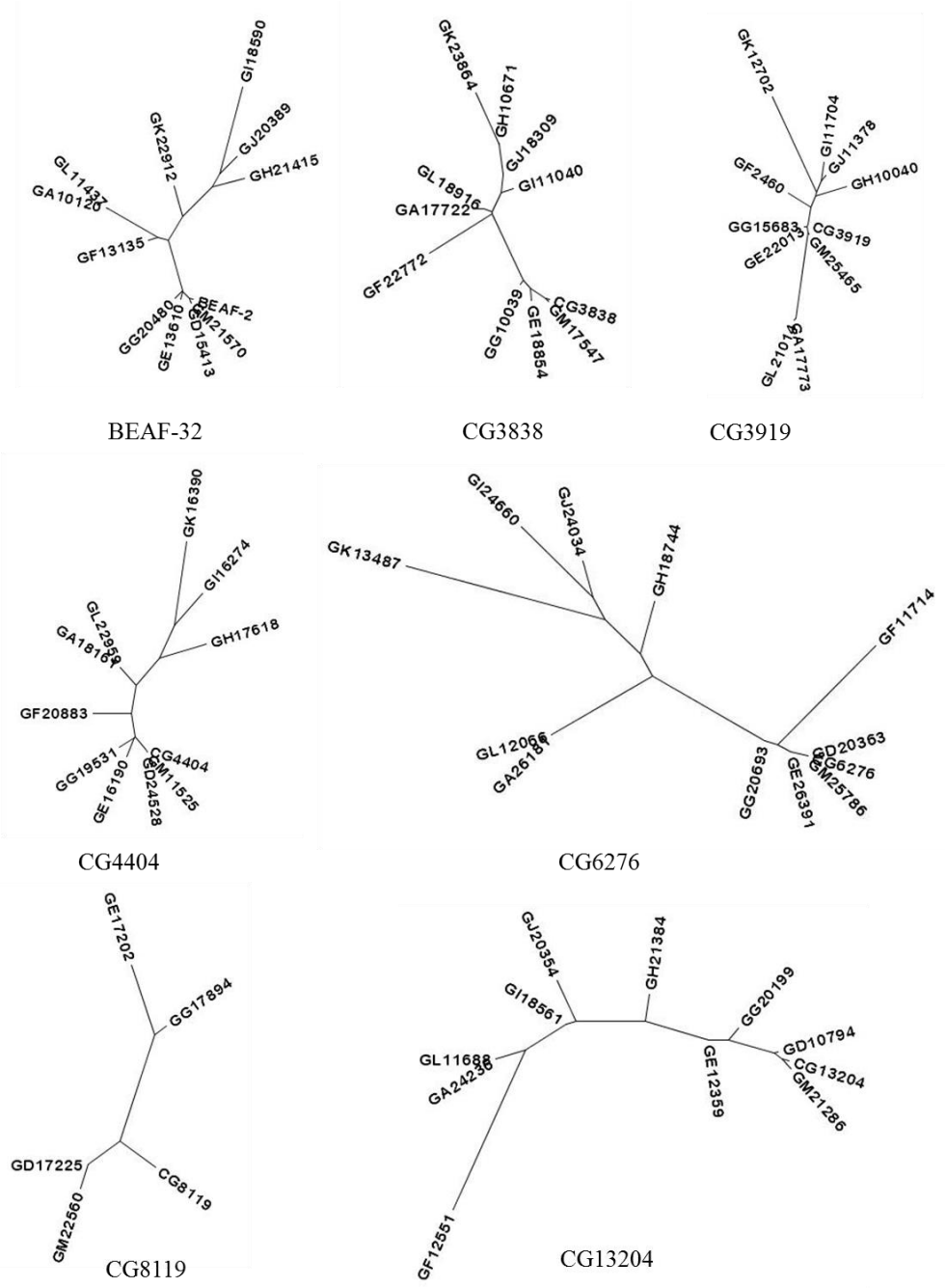


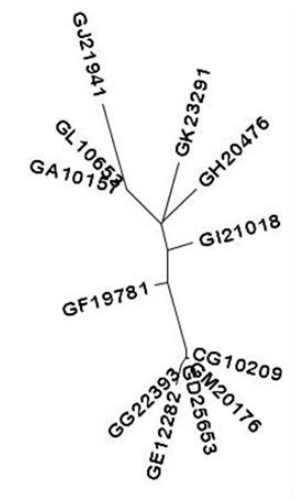
rgr



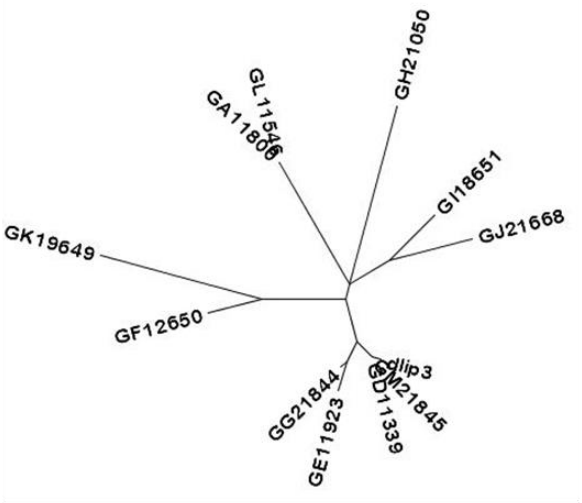
stwl

Figure 2. Phylogenetic Trees for All BESS domains in *Drosophila*. Maximum Likelihood Method Based Phylogenetic Trees for BESS domain genes constructed using IQTree Software (Trifinopoulos *et al.* 2016).

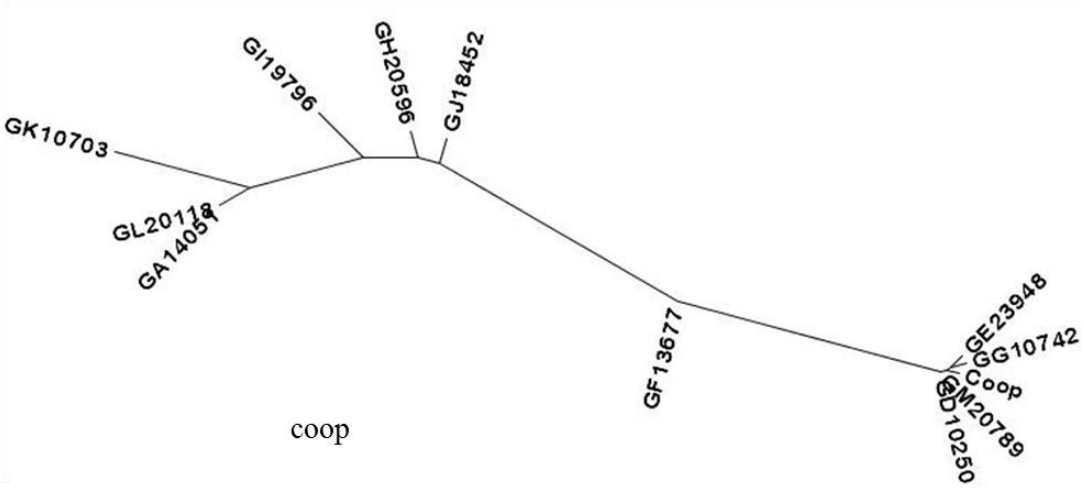




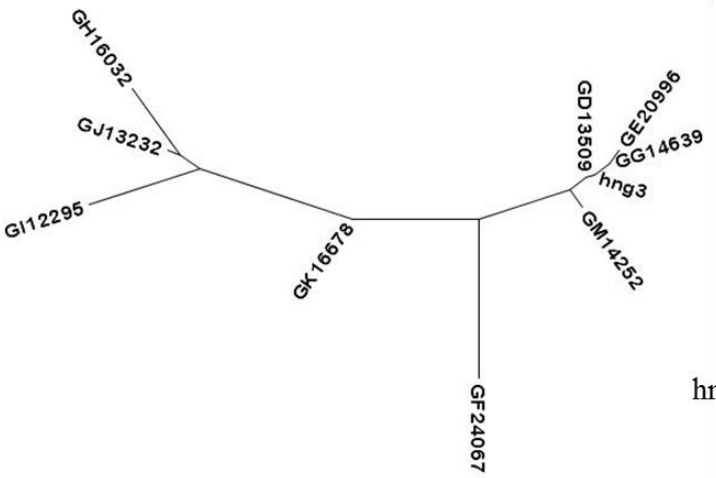
CG10209



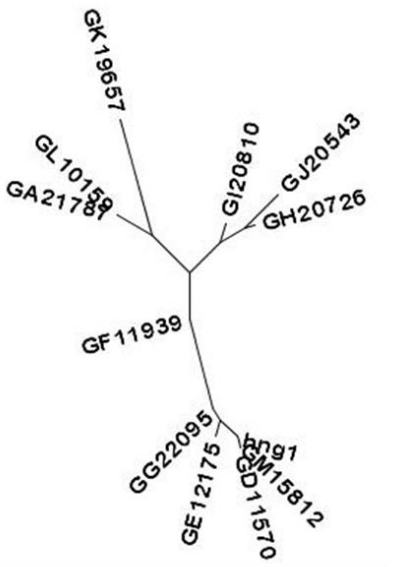
dlip3



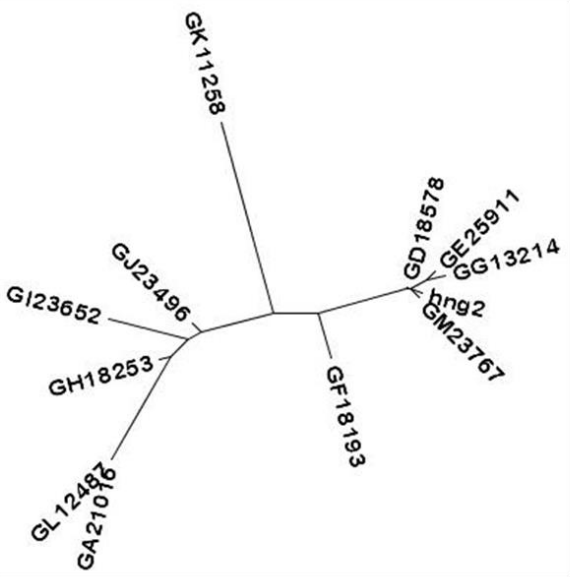
coop



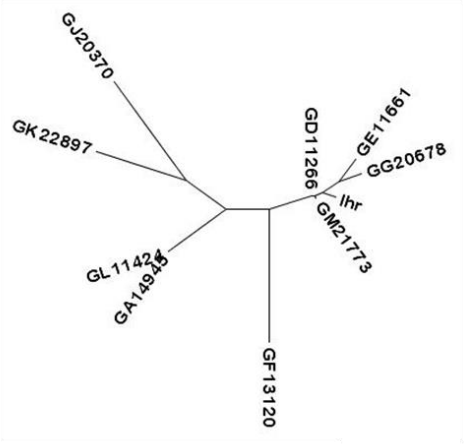
hng3



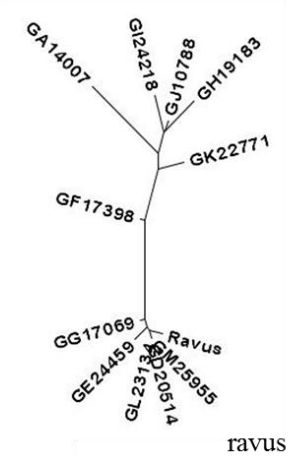
hng1



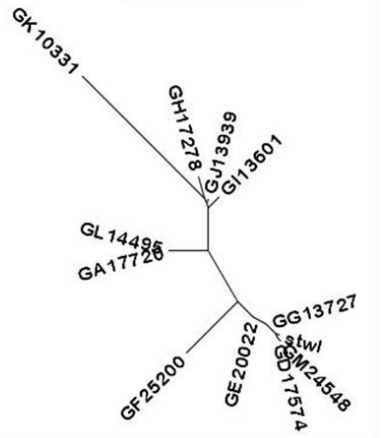
hng2



lhr



ravus



stwl

References

Akiyama T., 2002 Mutations of stonewall disrupt the maintenance of female germline stem cells in Drosophila melanogaster. Dev. Growth Differ. 44: 97–102.

Assa-Kunik E., Torres I. L., Schejter E. D., Johnston D. S., Shilo B.-Z., 2007 Drosophila follicle cells are patterned by multiple levels of Notch signaling and antagonism between the Notch and JAK/STAT pathways. Development 134: 1161–9.

Azpiazu N., Morata G., 2000 Function and regulation of homothorax in the wing imaginal disc of Drosophila. Development 127: 2685–93.

Bailey J. A., Yavor A. M., Massa H. F., Trask B. J., Eichler E. E., 2001 Segmental duplications: organization and impact within the current human genome project assembly. Genome Res. 11: 1005.

Bellen H. J., Levis R. W., Liao G., He Y., Carlson J. W., *et al.*, 2004 The BDGP gene disruption project: Single transposon insertions associated with 40% of Drosophila genes. Genetics 167: 761–781.

Bhaskar V., Courey A. J., 2002 The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. Gene 299: 173–84.

Biteau B., Karpac J., Supoyo S., DeGennaro M., Lehmann R., *et al.*, 2010 Lifespan extension by preserving proliferative homeostasis in Drosophila (SK Kim, Ed.). PLoS Genet. 6: 1–15.

Bowers J. L., Chapman B. A., Rong J., Paterson A. H., 2003 Unraveling angiosperms genome evolution by phylogenitc analysis of chromosomal duplications events. Nature 422: 433–438.

Brand A. H., Perrimon N., 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118: 401–15.

Bridges C. B., 1935 Salivary Chromosome Maps. J Hered 26 60-64. 26: 60–64.

Bridges C. B., 1936 THE BAR "GENE" A DUPLICATION. Sci. 83: 210–211.

Brodie E. D., 2010 How an ancient genome duplication electrified modern fish. Proc. Natl. Acad. Sci. U. S. A. 107: 21953–21954.

Brun S., Rincheval-Arnold a, Colin J., Risler Y., Mignotte B., *et al.*, 2006 The myb-related gene stonewall induces both hyperplasia and cell death in Drosophila: rescue of fly lethality by coexpression of apoptosis inducers. Cell Death Differ. 13: 1752–62.

Brunet F. G., Crollius H. R., Paris M., Aury J. M., Gibert P., *et al.*, 2006 Gene loss and

evolutionary rates following whole-genome duplication in teleost fishes. Mol. Biol. Evol. 23: 1808–1816.

Cadigan K. M., Grossniklaus U., Gehring W. J., 1994 Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. Proc. Natl. Acad. Sci. U. S. A. 91: 6324–8.

Cadigan K. M., Nusse R., 1997 Wnt signaling: a common theme in animal development. Genes Dev. 11: 3286–3305.

Cadigan K. M., 2002 Regulating morphogen gradients in the Drosophila wing. 13: 83–90.

Casanueva M. O., Ferguson E. L., 2004 Germline stem cell number in the Drosophila ovary is regulated by redundant mechanisms that control Dpp signaling. Development 131: 1881–1890.

Casares F., Mann R. S., 2000 A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identities in Drosophila. Development 127: 1499–508.

Casola C., Lawing a M., Betrán E., Feschotte C., 2007 PIF-like transposons are common in drosophila and have been repeatedly domesticated to generate new host genes. Mol. Biol. Evol. 24: 1872–88.

Cavodeassi F., Rodríguez I., Modolell J., 2002 Dpp signalling is a key effector of the wing-body wall subdivision of the Drosophila mesothorax. Development 129: 3815–23.

Celniker S. E., L Dillon L. A., Gerstein M. B., Gunsalus K. C., Henikoff S., *et al.*, 2009 Unlocking the secrets of the genome. Nature. June 18: 927–930.

Christoffels A., Koh E. G. L., Chia J. M., Brenner S., Aparicio S., *et al.*, 2004 Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol. Biol. Evol. 21: 1146–1151.

Clark K. A., McKearin D. M., 1996 The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development 122: 937–950.

Cohen1 S. M., Jurgens1 G., 1989 Proximal -distal pattern formation in Drosophila: cell autonomous requirement for Distal-less gene activity in limb development. EMBO J. 8: 2045–2055.

Conant G. C., Wolfe K. H., 2008 Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9: 938–50.

Couso J. P., Bate M., Martinez Arias A., 1993 A wingless-dependent polar coordinate system in

Drosophila imaginal discs. Science (80-.). 259: 484–489.

Cuevas M. De, Lilly M. A., Spradling A. C., 1997 Germline Cyst Formation in Drosophila. Annu. Rev. Genet. 31: 405–428.

Culi J., Aroca P., Modolell J., Mann R. S., 2006 jing is required for wing development and to establish the proximo-distal axis of the leg in Drosophila melanogaster. Genetics 173: 255–66.

Cutler G., Perry K. M., Tjian R., 1998 Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol. Cell. Biol. 18: 2252–61.

Deng W. M., Althauser C., Ruohola-Baker H., 2001 Notch-Delta signaling induces a transition from mitotic cell cycle to endocycle in Drosophila follicle cells. Development 128: 4737–46.

Dickerson J. E., Robertson D. L., 2012 On the origins of mendelian disease genes in man: The impact of gene duplication. Mol. Biol. Evol. 29: 61–69.

Dietzl G., Chen D., Schnorrer F., Su K.-C., Barinova Y., *et al.*, 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151–6.

Duffy J. B., 2002 GAL4 System in Drosophila : A Fly Geneticist ' s Swiss Army Knife. Genesis 15: 1–15.

Duncan, I., Montgomery G., 2002 From cis-trans test to the genetic control of development. Genetics 161(1): 1--10. Genetics 161(1):1-1.

Duong H. A., Nagaraj R., Wang C. W., Ratnaparkhi G., Henry Y., *et al.*, 2009 Non-Cell-Autonomous Inhibition Of Photoreceptor Development by Dip3. Dev. Biol. 323: 105–113.

Dykhuizen D., Hartl D. L., 1980 Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background. Genetics 96: 801–817.

Edgar R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–7.

England B. P., Admon a, Tjian R., 1992 Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc. Natl. Acad. Sci. U. S. A. 89: 683–7.

Fiston-Lavier A. S., Anxolabehere D., Quesneville H., 2007 A model of segmental duplication formation in Drosophila melanogaster. Genome Res. 17: 1458–1470.

Force A., Lynch M., Pickett F. B., Amores A., Yan Y. L., *et al.*, 1999a Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.

Force a, Lynch M., Pickett F. B., Amores a, Yan Y. L., *et al.*, 1999b Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–45.

Foronda D., Pérez-Garijo A., Martín F. a, 2009 Dpp of posterior origin patterns the proximal region of the wing. Mech. Dev. 126: 99–106.

Genomes D. 12, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–18.

Grieder N. C., Cuevas M. de, Spradling A. C., 2000 The fusome organizes the microtubule network during oocyte differentiation in Drosophila. Development 127: 4253–4264.

Gu Z., Cavalcanti A., Chen F.-C., Bouman P., Li W.-H., 2002 Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. Mol. Biol. Evol. 19: 256–262.

Gu Y., Xing S., He C., 2016 Genome-Wide Analysis Indicates Lineage-Specific Gene Loss during Papilionoideae Evolution. Genome Biol. Evol. 8: evw021.

Guillen I., Mullor J. L., Capdevila J., Sanchez-Herrero E., Morata G., *et al.*, 1995 The function of engrailed and the specification of Drosophila wing pattern. Development 121: 3447–3456.

Hahn M. W., 2009 Distinguishing among evolutionary models for the maintenance of gene duplicates. J. Hered. 100: 605–617.

Heger A., Ponting C. P., 2007 Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. 17: 1837–1849.

Herzig B., Yakulov T. A., Klinge K., Gunesdogan U., Jackle H., *et al.*, 2014 Bällchen is required for self-renewal of germline stem cells in Drosophila melanogaster. Biol. Open 3: 510–521.

Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B. C., *et al.*, 1996 Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. Nucleic Acids Res. 24: 4420–49.

Horn T., Sandmann T., Boutros M., 2010 Design and evaluation of genome-wide libraries for RNA interference screens. Genome Biol. 11: R61.

Hunter S., Jones P., Mitchell A., Apweiler R., Attwood T. K., *et al.*, 2012 InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40: D306-12.

Huynh J. R., St Johnston D., 2000 The role of BicD, Egl, Orb and the microtubules in the restriction of meiosis to the Drosophila oocyte. Development 127: 2785–2794.

Innan H., Kondrashov F., 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11: 97–108.

Jia F., Lo N., Ho S. Y. W., 2014 The impact of modelling rate heterogeneity among sites on phylogenetic estimates of intraspecific evolutionary rates and timescales. PLoS One 9.

Kellis M., Birren B. W., Lander E. S., 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. TL - 428. Nature 428 VN-: 617–624.

Kennerdell J. R., Carthew R. W., 2000 Heritable gene silencing in Drosophila using double-stranded RNA. Nat. Biotechnol. 18: 896–8.

Kimura M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Kimura M., 1983 Rare variant alleles in the light of the neutral theory. Mol. Biol. Evol. 1: 84–93.

Klein T., Arias a M., 1998 Different spatial and temporal interactions between Notch, wingless, and vestigial specify proximal and distal pattern elements of the wing in Drosophila. Dev. Biol. 194: 196–212.

Klein T., 2001 Wing disc development in the fly: the early stages. Curr. Opin. Genet. Dev. 11: 470–5.

Klenk H. P., Clayton R. A., Tomb J. F., White O., Nelson K. E., *et al.*, 1997 The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. Nature 390: 364–70.

Kondrashov F. a, Rogozin I. B., Wolf Y. I., Koonin E. V, 2002 Selection in the evolution of gene duplications. Genome Biol. 3 Res. 3: 1–9.

Krakauer D. C., Nowak M. a, 1999 Evolutionary preservation of redundant duplicated genes. Semin. Cell Dev. Biol. 10: 555–559.

Kumar S, Stecher G and T. K., 2013 MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., *et al.*, 2001 Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Lantz V., Chang J. S., Horabin J. I., Bopp D., Schedl P., 1994 The Drosophila orb RNA-binding protein is required for the formation of the egg chamber and establishment of polarity. Genes Dev. 8: 598–613.

Lasko P. F., Ashburner M., 1990 Posterior localization of vasa protein correlates with, but is not

sufficient for, pole cell development. Genes Dev. 4: 905–921.

Lawrence P. a, Morata G., 1993 A no-wing situation. Nature 363: 210–211.

Lespinet O., Wolf Y. I., Koonin E. V, Aravind L., 2002 The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12: 1048–59.

Lewis E. B., 1978 A gene complex controlling segmentation in Drosophila. Nature 276: 565–570.

Li W. H., Gu Z., Wang H., Nekrutenko A., 2001 Evolutionary analyses of the human genome. Nature 409: 847–849.

Liberles D. A., Kolesov G., Dittmar K., 2010 Understanding Gene Duplication Through Biochemistry and Population Genetics. In: *Evolution after Gene Duplication*, John Wiley & Sons, Inc., pp. 1–21.

Lin H., Yue L., Spradling A. C., 1994 The Drosophila fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. Development 120: 947–56.

Lynch M., Conery J. S., 2001 The Evolutionary fvate and Consequences of Duplicate Genes. Science (80-.). 290: 1151–1155.

Lyne R., Smith R., Rutherford K., Wakeling M., Varley A., *et al.*, 2007 FlyMine : an integrated database for Drosophila and Anopheles genomics.

Mach J. M., Lehmann R., 1997 An egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in Drosophila. Genes Dev. 11: 423–435.

Magrane M., Consortium U., 2011 UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011: bar009.

Maines J. Z., Park J. K., Williams M., McKearin D. M., 2007 Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development 134: 1471–1479.

Mann R. S., Abu-Shaar M., 1996 Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. Nature 383: 630–3.

Marchler-Bauer A., Derbyshire M. K., Gonzales N. R., Lu S., Chitsaz F., *et al.*, 2015 CDD : NCBI ’ s conserved domain database. Nucleic Acids Res. 43: 222–226.

Margolis J., Spradling a, 1995 Identification and behavior of epithelial stem cells in the Drosophila ovary. Development 121: 3797–807.

Marygold S. J., Leyland P. C., Seal R. L., Goodman J. L., Thurmond J., *et al.*, 2013 FlyBase:

improvements to the bibliography. Nucleic Acids Res. 41: D751–DD757.

McKearin D., Ohlstein B., 1995 A role for the Drosophila bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. Development 121: 2937–2947.

Megy K., Emrich S. J., Lawson D., Campbell D., Dialynas E., *et al.*, 2012 VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. 40: D729-34.

Miller M. A., Pfeiffer W., Schwartz T., 2011 Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.

Monson R. K., 2003 Gene Duplication, Neofunctionalization, and the Evolution of C_4 Photosynthesis. Int. J. Plant Sci. 164: S43–S54.

Narbonne-Reveau K., Besse F., Lamour-Isnard C., Busson D., Pret A. M., 2006 fused regulates germline cyst mitosis and differentiation during Drosophila oogenesis. Mech. Dev. 123: 197–209.

Nei M., 1969 Gene duplication and nucleotide substitution in evolution. Nature 221: 40–42.

Neumann C. J., Cohen S. M., 1996a Sternopleural is a regulatory mutation of wingless with Both Dominant and Recessive Effects on Larval Development of Drosophila melanogaster. Genetics: 1147–1155.

Neumann C. J., Cohen S. M., 1996b Distinct mitogenic and cell fate specification functions of wingless in different regions of the wing. Development 122: 1781–9.

Ng M., Diaz-Benjumea F. J., Cohen S. M., 1995 Nubbin encodes a POU-domain protein required for proximal-distal patterning in the Drosophila wing. Development 121: 589–99.

Nowak M. A., Boerlijst M. C., Cooke J., Maynard Smith J., 1997 Evolution of genetic redundancy. Nature 388: 167–71.

Nowick K., Hamilton A. T., Zhang H., Stubbs L., 2010 Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. Mol. Biol. Evol. 27: 2606–2617.

Nystul T., Spradling A., 2010 Regulation of epithelial stem cell replacement and follicle formation in the Drosophila ovary. Genetics 184: 503–15.

Oakley T. H., Ostman B., Wilson A. C. V, 2006 Repression and loss of gene expression outpaces activation and gain in recently duplicated fly genes. Proc. Natl. Acad. Sci. U. S. A. 103: 11637–11641.

Ohno S., 1970 *Evolution by gene duplication.*

Pan L., Chen S., Weng C., Call G., Zhu D., *et al.*, 2007 Stem Cell Aging Is Controlled Both Intrinsically and Extrinsically in the Drosophila Ovary. Cell Stem Cell 1: 458–469.

Pfeiffer W., Stamatakis A., 2010 Hybrid MPI/Pthreads parallelization of the RAxML phylogenetics code. 2010 IEEE Int. Symp. Parallel Distrib. Process. Work. Phd Forum 978: 1–8.

Piatigorsky J., O'Brien W. E., Norman B. L., Kalumuck K., Wistow G. J., *et al.*, 1988 Gene sharing by delta-crystallin and argininosuccinate lyase. Proc. Natl. Acad. Sci. U. S. A. 85: 3479–83.

Posada D., 2001 Selecting the Best-Fit Model of Nucleotide Substitution. Syst. Biol. 50: 580–601.

Posada D., 2003 jModelTest : Phylogenetic Model Averaging. : 2001–2004.

Posada D., Buckley T. R., 2004 Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst. Biol. 53: 793–808.

Ratnaparkhi G. S., Duong H. a, Courey A. J., 2008 Dorsal interacting protein 3 potentiates activation by Drosophila Rel homology domain proteins. Dev. Comp. Immunol. 32: 1290–300.

Rodríguez Dd D. D. A., Terriente J., Galindo M. I., Couso J. P., Díaz-Benjumea F. J., 2002 Different mechanisms initiate and maintain wingless expression in the Drosophila wing hinge. Development 129: 3995–4004.

Roth S., Lynch J. A., 2009 Symmetry breaking during Drosophila oogenesis. Cold Spring Harb. Perspect. Biol. 1: a001891–a001891.

Rouquier S., Blancher a, Giorgi D., 2000 The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. U. S. A. 97: 2870–2874.

Rubin G. M., Yandell M. D., Wortman J. R., Gabor Miklos G. L., Nelson C. R., *et al.*, 2000 Comparative genomics of the eukaryotes. Science 287: 2204–15.

Russell S., 2000 The Drosophila dominant wing mutation Dichaete results from ectopic expression of a Sox-domain gene. Mol. Gen. Genet. 263: 690–701.

Ryu B.-Y., Orwig K. E., Oatley J. M., Avarbock M. R., Brinster R. L., 2006 Effects of aging and

niche microenvironment on spermatogonial stem cell self-renewal. Stem Cells 24: 1505–11.

Schultz J., Milpetz F., Bork P., Ponting C. P., 1998 SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U. S. A. 95: 5857–64.

Shcherbata H. R., Althauser C., Findley S. D., Ruohola-Baker H., 2004 The mitotic-to-endocycle switch in Drosophila follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions. Development 131: 3169–3181.

Shukla V., Habib F., Kulkarni A., Ratnaparkhi G. S., 2014 Gene duplication, lineage-specific expansion, and subfunctionalization in the MADF-BESS family patterns the Drosophila wing hinge. Genetics 196: 481–96.

Smith J. M., Smith N. H., 1996 Synonymous nucleotide divergence: What is "saturation"? Genetics 142: 1033–1036.

Soanes K. H., Mackay J. O., Core N., Heslip T., Kerridge S., *et al.*, 2001 Identi ® cation of a regulatory allele of teashirt (tsh) in Drosophila melanogaster that affects wing hinge development . An adult-speci ® c tsh enhancer in Drosophila. 105: 145–151.

Song X., Xie T., 2003 Wingless signaling regulates the maintenance of ovarian somatic stem cells in Drosophila. Development 130: 3259–3268.

Song X., 2003 wingless signaling regulates the maintenance of ovarian somatic stem cells in Drosophila. Development 130: 3259–3268.

Song H., Goetze S., Bischof J., Spichiger-haeusermann C., Kuster M., *et al.*, 2010 Coop functions as a corepressor of Pangolin and antagonizes Wingless signaling. Genes Dev. 24: 881–886.

Spradling A., Fuller M. T., Braun R. E., Yoshida S., 2011 Germline stem cells. Cold Spring Harb. Perspect. Biol. 3.

Strigini M., Cohen S. M., 1999 Formation of morphogen gradients in the Drosophila wing. Semin. Cell Dev. Biol. 10: 335–44.

Strimmer K., Haeseler A. von, 2009 Genetic distances and nucleotide substitution models. phylogenetic Handb. A Pract. approach to phylogenetic Anal. hypothesis Test.: 111–125.

Sturtevant A. H., 1925 The Effects of Unequal Crossing over at the Bar Locus in Drosophila. Genetics 10: 117–147.

Sugino R. P., Innan H., 2006 Selection for more of the same product as a force to enhance

concerted evolution of duplicated genes. Trends Genet. 22: 642–644.

Swarup S., Verheyen E. M., 2012 Wnt/Wingless signaling in Drosophila. Cold Spring Harb. Perspect. Biol. 4: a007930.

Taylor J. S., Raes J., 2004 Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38: 615–43.

The Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815.

Tischler J., Lehner B., Chen N., Fraser A. G., 2006 Combinatorial RNA interference in Caenorhabditis elegans reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol. 7: 1–13.

Tomancak P., Beaton A., Weiszmann R., Kwan E., Shu S., *et al.*, 2002 Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. 3: RESEARCH0088.

Tomancak P., Berman B. P., Beaton A., Weiszmann R., Kwan E., *et al.*, 2007 Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. 8: R145.

Trifinopoulos J., Nguyen L.-T., Haeseler A. von, Minh B. Q., 2016 W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res.: 1–4.

Venken K. J. T., Schulze K. L., Haelterman N. a, Pan H., He Y., *et al.*, 2011 MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat. Methods 8: 737–743.

Wagner A., 1999 Redundant Gene Functions and Natural Selection. J. Evol. Biol. 12: 1–16.

Wakeley J., 1994 Substitution-rate variation among sites and the estimation of transition bias. Mol. Biol. Evol. 11: 436–442.

Waterhouse R. M., Tegenfeldt F., Li J., Zdobnov E. M., Kriventseva E. V, 2013 OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41: D358-65.

Williams J. A., Paddock S. W., Vorwerk K., Carroll S. B., 1994 Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature 368: 299–305.

Wodarz a, Nusse R., 1998 Mechanisms of Wnt signaling in development. Annu. Rev. Cell Dev. Biol. 14: 59–88.

Wolfe K. H., Shields D. C., 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.

Wu J., Cohen S. M., 2002 Repression of Teashirt marks the initiation of wing development. Development 129: 2411–8.

Xia X., 2013 DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30: 1720–1728.

Xie T., Spradling A. C., 1998 decapentaplegic is essential for the maintenance and division of germline stem cells in the Drosophila ovary. Cell 94: 251–260.

Yang Z., Goldman N., 1993 Comparison of Models for Nucleotide Substitution Used in Maximum-Likelihood Phylogenetic Estimation.

Yang Z., 1996 Amon-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11: 367–372.

Yi X., Vries H. I. De, Siudeja K., Rana A., Lemstra W., *et al.*, 2009 Stwl Modifies Chromatin Compaction and Is Required to Maintain DNA Integrity in the Presence of Perturbed DNA Replication. Mol. Biol. Cell 20: 983–994.

Zamore P. D., Tuschl T., Sharp P. A., Bartel D. P., 2000 RNAi : Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell 101: 25–33.

Zhang J., Rosenberg H. F., Nei M., 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. U. S. A. 95: 3708–3713.

Zhang P., Gu Z., Li W. H., 2003 Different evolutionary patterns between young duplicate genes in the human genome. Genome Biol 4: R56.

Zhang J., 2003 Evolution by gene duplication: An update. Trends Ecol. Evol. 18: 292–298.

Zirin J. D., Mann R. S., 2004 Differing strategies for the establishment and maintenance of teashirt and homothorax repression in the Drosophila wing. Development 131: 5683–93.