MOLECULAR POPULATION GENETICS OF GENES OF SPERMATOGENESIS IN *DROSOPHILA*

BY

SUJEETHA ANGEL RAJAKUMAR

A Thesis submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of:

MASTER'S OF SCIENCE

Department of Biochemistry and Medical Genetics
University of Manitoba
Winnipeg, Manitoba, Canada
March 2005

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Molecular Population Genetics of Genes of Spermatogenesis in *Drosophila*

BY

Sujeetha Angel Rajakumar

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree

of

MASTER OF SCIENCE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ddH_20	Distilled deionized water
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetate
$Mgcl_2$	Magnesium Chloride
ml	Millilitre
μl	Microlitre
PCR	Polymerase Chain Reaction
Tris Hcl	Tris hydrochloric acid

ACKNOWLEDGEMENTS

I wish to place on record my deep sense of gratitude, profound thanks to my supervisor, Dr. Alberto Civetta for his valuable guidance, encouragement and support towards this project. I sincerely thank Barb for helping me in my bench work. I thank Eric for preparing food for my flies.

I wish to express my sincere thanks to my committee members Dr. Dan Gietz and Dr. David Merz for their valuable comments and suggestions during my course of study. I wish to thank my external examiner Dr. Steve Whyard for readily accepting to be my examiner and for his valuable inputs. Also, I would like to thank my head of the department Dr. Jane Evans.

Thanks to Tuntun, Jan, Lil and all the office staff of Biochemistry and Medical Genetics for their help during my study.

Special thanks to all my friends for their wishes for me. My deepest regards to Suzy, my aunt Ida and uncle Jeyaraj and their families for upholding me in their prayers throughout….

I am forever indebted to my parents and brother for their constant encouragement, moral support, prayers and unconditional love.

ABSTRACT

Hybrid males resulting from crosses between closely related species of *Drosophila* are sterile. The F1 hybrid sterility phenotype is mainly due to defects occurring during sperm maturation. From this perspective, it is believed that genes controlling sperm maturation may be subjected to selective diversification between species but may also experience selective constraints that are typical of developmental genes. We compared the molecular evolutionary pattern of *don juan* (*dj*), *always early* (*aly*) and *bag of marbles* (*bam*), three genes playing a role at different stages during the sperm developmental pathway in *Drosophila*. *Don juan* is a late gene, expressed postmeiotically during spermiogenesis in elongated spermatids and also in mature sperms; *aly* is a meiotic arrest gene regulating entrance into meiotic division and *bam* gene regulates the progression through the early steps of the male and female germ cell lineage. The complete coding region of these genes was sequenced in different strains of *Drosophila melanogaster* and *D. simulans*. Estimates of proportion of nonsynonymous and synonymous intraspecific polymorphism and interspecific divergence suggest that purifying selection constrains the accumulation of random mutations in these genes. Selective constraints are stronger in regions that define function such as the nuclear localization domains in *aly* and *dj*. Nucleotide polymorphism and divergence were found to be within average values for developmental genes in *Drosophila* but Tajima's D and Fu and Li's test of neutrality suggest some form of purifying selection or positive selection within *D. melanogaster* populations. All the coding regions in *bam, aly* and *dj* showed higher rate of

interspecific than intraspecific nonsynonymous to synonymous substitutions. The sign of positive selection driving divergence from *D. simulans* is evident for the *dj* gene in African populations and exon two of *aly* gene in both African and Non-African populations. However, a significant result for *aly* might be biased by a high transition/transversion ratio indicative of mutation bias.

1. INTRODUCTION

1.1 Spermatogenesis in *Drosophila*

Spermatogenesis in *Drosophila* provides an excellent model system to study genes controlling development at the molecular level. Spermatogenesis is initiated at the apical tip of the testis in the germinal proliferation centre where the germ line stem cells are surrounded by specialized somatic cells called the hub [1]. A pair of cyst progenitor cells encloses each germline stem cell. The division of a germ line stem cell produces two daughters of which one stays as the stem cell (self-renewal cells) and the other differentiates into a gonial cell (gonialblast). The gonial cell enters an amplification stage consisting of a series of mitotic divisions, four in the case of *D. melanogaster*, and results in a cyst of 16 primary spermatocytes [1]. The primary spermatocytes grow and genes with a role in spermatogenesis and genes with functions in different tissues during the life cycle get transcribed. Gene expression stops at the mature primary spermatocyte stage. The primary spermatocyte then undergoes meiosis I and meiosis II resulting in 64 haploid spermatids [2]. The last step is spermatid differentiation (spermiogenesis) marked by a series of morphological changes where remodeling of subcellular components take place in the form of growth of the long flagellar axoneme, elongation of specialized mitochondrial derivative for motility, DNA condensation and nuclear shaping [2] (Figure 1).

There are three major steps that can be recognized during spermatogenesis in *Drosophila,* with different genes controlling the pathway. The steps include the mitotic divisions resulting in 16 mature primary spermatocytes, meiotic divisions and spermatid differentiation [2]. Mutations in different genes that play a role in the regulation of these three stages (mitotic divisions, meiotic divisions and sperm maturation) arrest spermatogenesis and causes male sterility.

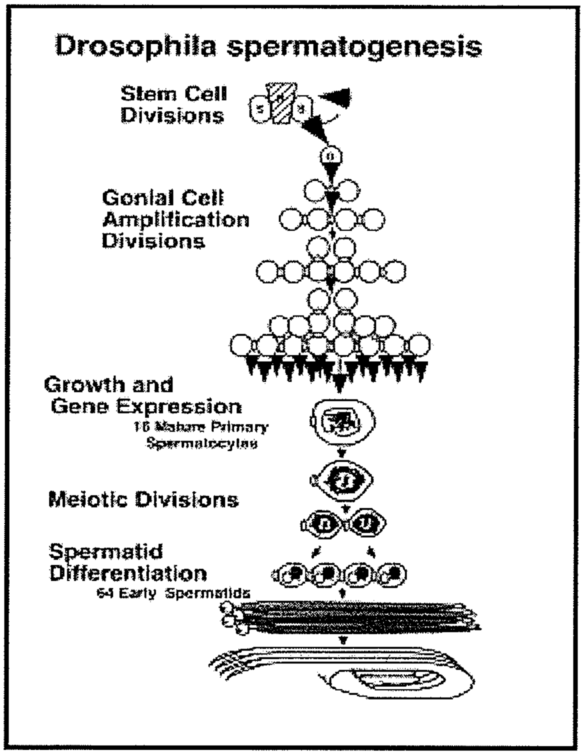

Drosophila spermatogenesis
Stem Cell Divisions
Gonial Cell Amplification Divisions
Growth and Gene Expression
16 Mature Primary Spermatocytes
Meiotic Divisions
Spermatid Differentiation
64 Early Spermatids

Figure 1. Spermatogenesis in *Drosophila* (From Fuller. M.T [2])

1.2 Hybrid male sterility

Male hybrids resulting from crosses between different *Drosophila* species are sterile in accordance with Haldane's rule [3]. Haldane's rule states "When in the F1 offspring of two different animal races one sex is absent, rare or sterile, that sex is the heterozygous (heterogametic) sex" [4]. In *Drosophila,* males are the heterogametic sex (X and Y chromosome) and females are the homogametic sex (two X chromosome). Sterility can be due to the faster evolution of genes in the X chromosome versus autosomes (faster X hypothesis) [5] or due to faster evolution of genes with a male-specific pattern of expression (faster males hypothesis) [6]. Faster X hypothesis could be explained by the accumulation of favourable recessive alleles which get fixed leading to hybrid breakdown due to accumulation of disproportionate number of substitutions [5]. In a recent study that tested for faster X evolution in *Drosophila* [7], the rates of sequence divergence for X-linked and autosomal loci were analyzed. It was found that X-linked and autosomal loci evolve at the same rates. Moreover, it was also found that the genes with sex-limited expression on the X chromosomes and autosomes evolve at similar rates. Contrary to what is seen in *Drosophila*, genes encoding mammalian X-linked sperm proteins evolve faster than genes on the autosomes [8].

Sex related genes are genes involved in mating behavior, spermatogenesis, fertilization and sex determination and they have been shown to evolve at a faster rate than other genes [9,10]. Particularly genes with a role in fertilization show faster rate of evolution between species and signs of positive selection [11,12]. It is possible that hybrid male sterility observed between closely related species is at least partially a consequence of rapid divergence of sex related genes. Microscopy analysis of six different F1 hybrid genotypes resulting from crosses between *D. simulans*, *D. sechellia* and *D. mauritiana* showed two distinct classes of spermatogenic defects namely premeiotic and postmeiotic defects. Out of the six interspecific hybrid genotypes, four showed defects that were postmeiotic [13]. Gene expression studies show that genes expressed in late stages of sperm development are differentially expressed in sterile hybrids [14]. The differential expression is further shown to result more from downregulation than upregulation of genes in the hybrids [14].

1.3 Genes of Spermatogenesis

Although several genes have been identified to play different roles during sperm development (Table 1), this study focuses on *bag of marbles* (*bam*), *always early* (*aly*) and *don juan* (*dj*). These genes play a role among the major steps of the spermatogenesis pathway (Figure 2). Mutations in these genes affect progression of spermatogenesis in male germ cell differentiation. *bam* regulates the progression through the early steps of both the male and female germ line cell lineage, *aly* is a meiotic arrest gene regulating entrance into meiotic divisions in males and *dj* is a late expressed gene in spermatogenesis exclusively expressed in males.

Gene	Chromosome	Accession No	Mutations/Tissue of expression
achi	2R(49A13-B1)	AE003822	Partially male sterile
Acp29AB	2L(29C1)	AE003621	Male accessory gland
aly	3L(63A3)	AE003476	G2/M transition, Recessive male sterile, Recessive meiotic
bam	3R(96C7-8)	AE003751	Cystoblast cell division
bgcn	2R(60A4)	AE003462	Germ cell development
bob	82D3-8	AQ026432	Nebenkern formation
bol	3L(66F5-6)	AE003553	Nebenkern, Primary spermatocyte
cdc2	2L(31D11)	AE003628	G2/M transition of cell cycle
chic	2L(26A5-B2)	AE003612	Nebenkern
comr	2R(58A3)	AE003455	G2/M transition, Interacts with aly
crl	X(14F1)	AE003502	Spermatid, meiotic cycle, recessive male sterile
dbf	(32A1-2)	N/A	Spermatid, nebenkern
DnaJ-60	2R(60C1)	AE003463	Spermatogenesis
ego	N/A	N/A	Male germ-line stem cell division
fbl	3L(77B9-C1)	AE003591	Spermatid, Nebenkern, Recessive male sterile
fzo	3R(94E6)	AE003742	Nebenkern, spermatid, Recessive male sterile
gdl	(71D3)	N/A	Testes
ifc	2L(26B2)	AE003612	Nebenkern, spermatid, spermatocyte, Recessive male sterile
lectin-21Ca	2L(21E2)	AE003588	Spermatogenesis
lectin-21Cb	2L(21E2)	AE003588	Spermatogenesis
lectin-22C	2L(22C1)	AE003584	Spermatogenesis
lectin-24A	2L(24C1)	AE003579	Spermatogenesis
lectin-24Db	2L(24D8)	AE003577	Spermatogenesis
lectin-28C	2L(28D2)	AE003619	Spermatogenesis
lectin-29Ca	2L(29C1)	AE003621	Spermatogenesis
lectin-30A	2L(30A6-7)	AE003624	Spermatogenesis
Meics	3L(70C7)	AE003536	Male meiosis
Msi	3L(75A2)	AE003523	Spermatid, spermatocyte, spermatozoon
pelo	3L(30C5)	AE003625	Nebenkern, spermatocyte, spermatid, Recessive male sterile

Gene	**Chromosome**	**Accession No**	**Mutations/Tissue of expression**
Rb97D	3R(97D5)	AE003758	Axoneme, Recessive male sterile
Samuel	2L(32C5-D1)	AE003630	Spermatogenesis
shk	(82C1-5)	N/A	Spermatid, Nebenkern, recessive male sterile
Taf12L	2L(25A3)	AE003575	Meiosis, spermatid differentiation
tho	(86E2-20)	N/A	Spermatid, spermatozoon, recessive male sterile
twe	2L(35F1)	AE003650	Primary spermatocyte cyst, spermatocyte
vis	2R(A12-13)	AE003822	Spermatogenesis
Spermiogenesis Genes			
Act5C	X(5C7)	AE003435	Actin filament
Ance	2l(34E2)	AE003641	Spermatid nuclear differentiation, male sterile
Bruce	3R(86A7-8)	AE003686	Sperm individualization
cbx	2R(46B13-C1)	AE003831	Sperm individualization
chc	X(13F5-7)	AE003500	Spermatozoon
dhod	3R(85A5)	AE003679	Spermatocytes, Spermatid
dj	3R(84B2)	AE003673	Sperm individualization
dud	(21-60)	N/A	Spermatozoon, nebenkern, recessive male sterile
Ecr	2R(42A9-13)	AE003784	Sperm individualization
janB	3R(99D3)	AE003772	Translational control in spermiogenesis
mlt	(46F)	N/A	Sperm individualization
Mst98Ca	3R(98C3)	AE003764	Translational control, spermiogenesis
Mst98Cb	3R(98C3)	AE003764	Translational control, spermiogenesis
nkg	(61-100)	N/A	Nebenkern, Sperm individualization
po	2L(28D11-E1)	AE003619	Sperm individualization

Table 1. Spermatogenesis and spermiogenesis genes of *Drosophila.* Ref: http://flybase.bio.indiana.edu/

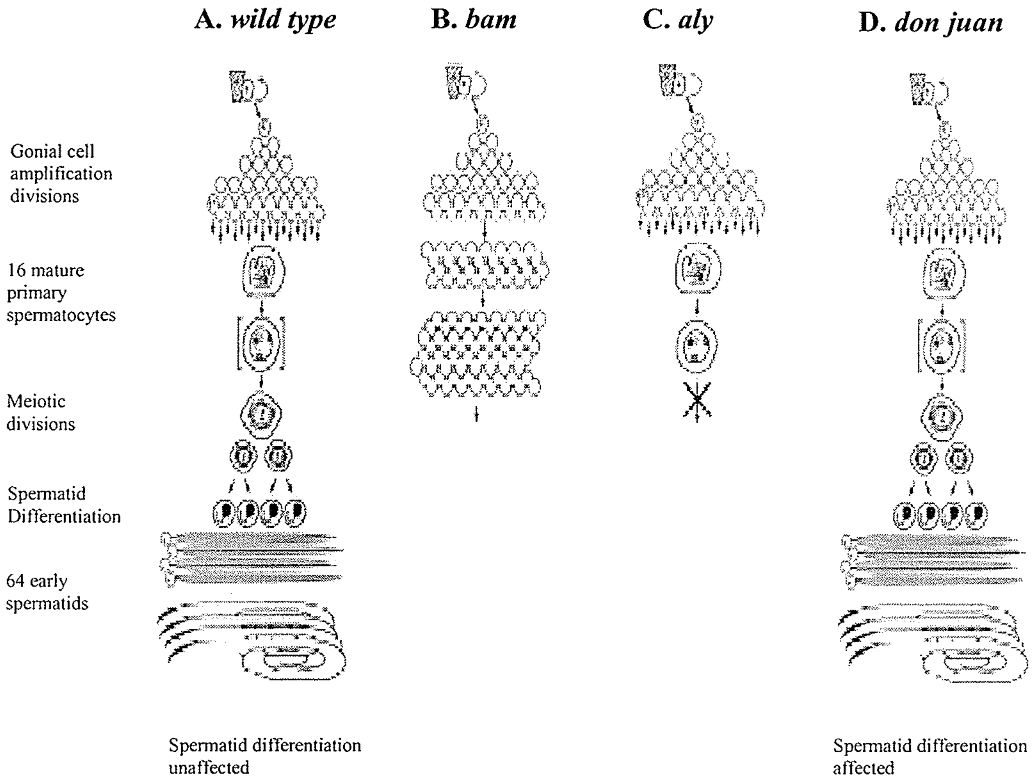

A. *wild type*
B. *bam*
C. *aly*
D. *don juan*
Gonial cell amplification divisions
16 mature primary spermatocytes
Meiotic divisions
Spermatid Differentiation
64 early spermatids
Spermatid differentiation unaffected
Spermatid differentiation affected

Figure 2. <u>Overview of events in the spermatogenesis pathway of *Drosophila* and mutation in genes affecting its progression.</u> A. Wild type showing the normal spermatogenesis pathway. B. Mutations in *bam* results in cysts of early germ cells and does not differentiate into primary spermatocytes. C. Mutations in *aly* arrest meiotic cell cycle progression. D. Mutations in *don juan* affect spermatid differentiation and individualization. Adapted from Fuller. M.T [2].

1.3a *bag of marbles (bam)*

bam regulates cystoblast divisions during male and female gametogenesis. In both male and female germline, stem cell divides to form a cyst of 16 inter connected cystocytes in which gene transcription and translation occurs. In males, this results in 25-fold increase in size of the 16 primary spermatocytes. Mutations in *bam* cause abnormal cysts that cannot develop into gametes as they fail to enter meiosis [15] (Figure 2.B). *bam* mutant spermatocytes are found to contain abnormal excessive number of small cells and the testis look like a bag of marbles. *bam* is thought to act downstream of other genes to offer male-specific or female specific functions as male and female *bam* transcripts are the same [15]. In females *bam* functions at two steps in differentiation where it specifies the germ line stem cell divisions to follow cystoblast (equivalent to gonialblast in males) fate and also to cease mitotic division and initiate meiotic division. In males *bam* is required only to cease mitotic division and begin meiotic division [2].

Bam encodes a 442 amino acid protein with a weak similarity to *Drosophila* ovarian tumor *Otu*, a gene required for germ cell differentiation. Mutations in *otu* gene produce tumorous egg chambers. The C terminus of bam protein (positions 404-432) matches the consensus for PEST (Proline, Glutamic acid, Serine, Threonine) domains [15]. PEST sequences are related to protein instability where the presence of these sequences can lead to the degradation of the proteins containing them [16]. Therefore the presence of a PEST domain in bam protein suggests bam produces an unstable protein product. A potential proteolytic cleavage site is found around positions 234-235 in *bam* [15]. Bam protein is localized to two different cellular compartments. They are the fusome (BamF) and the cytoplasm (BamC). Fusome is a germ cell specific organelle which has an elongated branched structure connecting the cells in the cystocyte. Bam is required as a switch from stem cells to cystoblast in females and males thereby promoting incomplete cytokinesis and activating fusome growth. After the fourth cystocyte division, BamC is degraded which blocks fusome growth and the cystocyte withdraws from the mitotic cycle [17].

The *bam* gene is 1454 base pairs long and includes three exons and two introns.

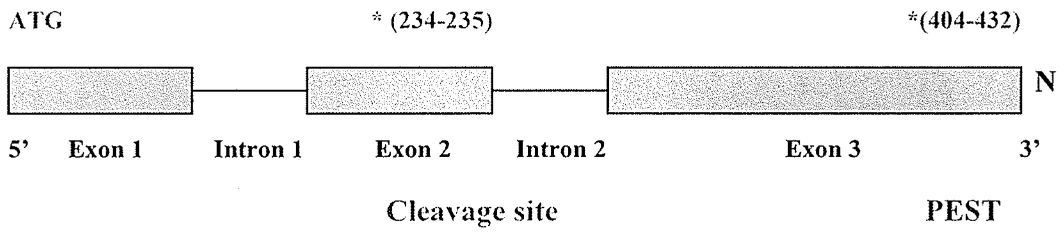
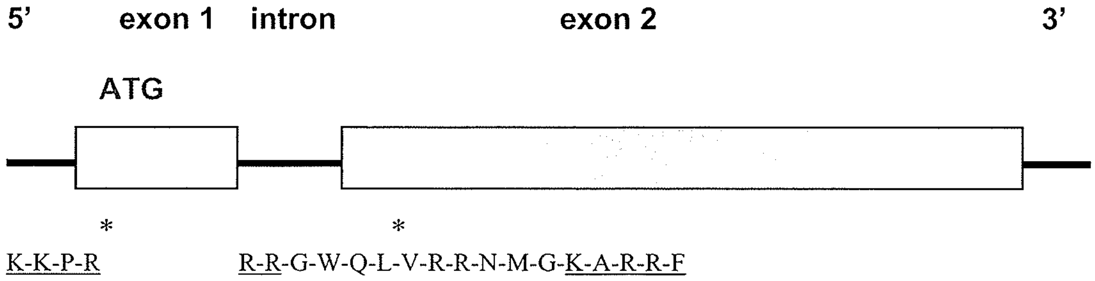


Figure 2a. Diagrammatic representation of *bam* gene showing exons and introns

1.3b *always early (aly)*

Meiotic arrest genes are divided into aly class and can class based upon its function. *aly* and *cookie monster* gene (*comr*) genes belong to aly class while the can class genes include *cannonball* (*can*), *meiosis I arrest* (*mia*) and *spermatocyte arrest I* (*sa*) [18]. The meiotic arrest genes are necessary for the G2/M transition of spermatogenesis and thereby they control the transcription of genes required for meiosis and spermiogenesis [19]. Gene *aly* acts upstream of *can, mia* and *sa* to coordinate the onset of meiosis with spermatid differentiation. Gene *aly* regulates the transcription of other genes such as *twine* and also two meiotic regulators *cyclin B* and *boule* [18]. Genes *can*, *mia* and *sa* are required for the accumulation of twine protein. In mutants of *can*, *mia* and *sa*, spermatid differentiation is arrested and spermatogenesis stops with the formation of mature primary spermatocytes [20]. In mutants of *aly*, spermatogenesis is also arrested at the mature primary spermatocyte stage [19] (Figure 2.C). Gene *aly* acts as a global regulator and the wild type function of *aly* is required for the accumulation of other proteins needed for entrance into meiosis and spermatid differentiation [2] (Figure 3). Gene *aly* plays an important role in modifying chromatin structure and triggering spermatogenesis specific gene transcription [20]. The aly protein is synthesized in the cytoplasm and transported to the nucleus of primary spermatocytes. This translocation represents an important control point (Figure 4). The translocation to the nucleus is aided by two predicted nuclear localisation signals (NLS) within *aly* [21]. The NLS of the aly protein is essential for its function because mutations within NLS result in the accumulation of aly protein in the cytoplasm [18]. Nuclear localization of *aly* is also mutually dependent on the wild type function of *comr*, an aly class meiotic arrest gene [18]. The comr protein is 68kDa in size and aly interacts with comr during the transport to the nucleus through the nuclear pore complex. The aly-comr complex interacts with the chromatin and controls its conformation. Mutant versions of *aly* and *comr* remain in the cytoplasm and are phenotypically similar in that they fail to transcribe other genes needed during spermatogenesis [18]. Genes *aly* and *comr* alter the chromatin structure so that other transcription factors can bind to the chromatin leading to transcriptional regulation in the mature primary spermatocytes [18]. Gene *aly* has no similarity to any DNA binding domain or transcriptional activators even though it is essential for transcription of other spermatogenesis genes [21].

The aly gene is located on chromosome three of *Drosophila*. It is made up of two exons and an intron with a length of 1.85 kb.



* Location of nuclear localization signals

Figure 2b. Diagrammatic representation of *aly* gene showing exons and introns

1.3c. *don juan (dj)*

During spermatogenesis in *Drosophila,* the *dj* gene is exclusively expressed in the male germ line. It encodes a basic lysine rich protein of 29kDa in size with structural similarities to histone H1. The *dj* gene product participates in the process where the spermatids become individualized and differentiated into motile sperm (figure 2.D). The carboxy terminal part of the protein is marked by a special feature of eight times direct repeated hexapeptide sequence (DPCKKK) [22]. The high lysine rich basic characteristic feature of dj resembles other basic structural proteins such as

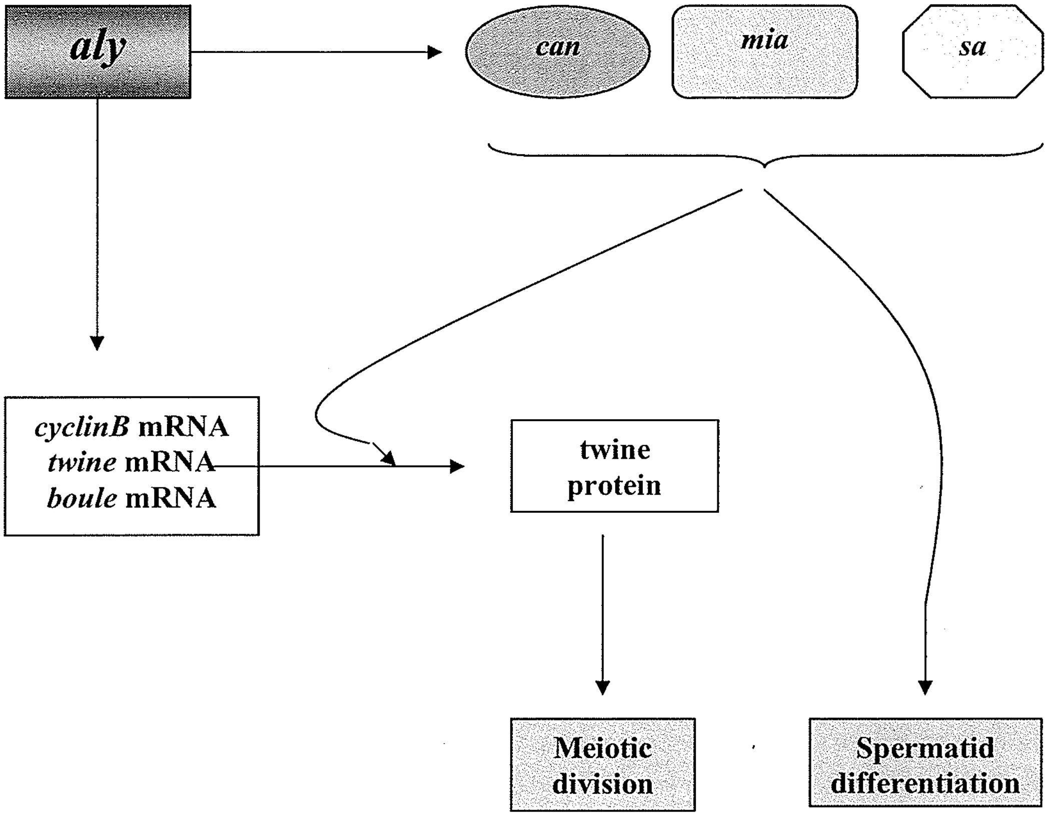


Figure 3. **Model showing control of meiotic division and spermatid differentiation by *aly* gene acting upstream of *can, mia, sa* and other genes.** *aly* acts through *can mia* and *sa* and controls spermatid differentiation genes. It also controls transcription of twine mRNA thereby controlling the meiotic divisions. *Can*, *mia* and *sa* individually or together control the translation or stabilization of twine protein.

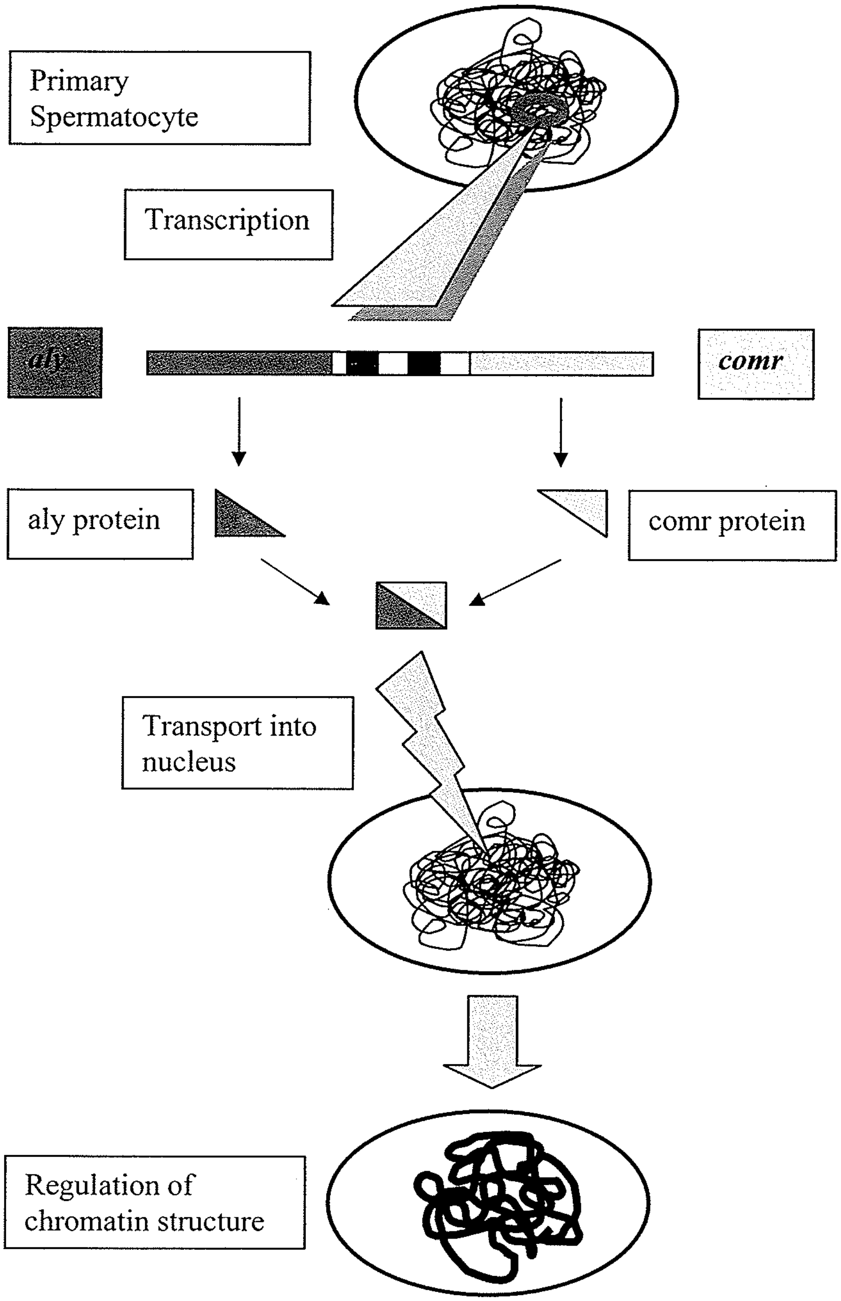


Figure 4. Mutually dependent aly class genes (*aly* and *comr*) are transported from the cytoplasm into the nucleus and regulates chromatin structure. It is required for meiosis and spermiogenesis.

mammalian cyclicins and calicin [23, 24, 25]. Except for the basic lysine rich content, dj does not show any sequence similarity to those proteins. Gene *dj* is expressed postmeiotically during spermiogenesis in elongated spermatids and in mature sperm. Gene *dj* is transcribed in primary spermatocytes and the mRNA remains translationally repressed until chromatin condensation occurs in spermatids. Translation of dj mRNA first appears during chromatin condensation in the nuclei of spermatids and then along the flagellum in the mitochondrial derivatives [22, 26]. Including *dj*, other spermiogenesis genes that are translationally repressed are the *Mst(3)CGP* gene family [27, 28], the *dihydroorotate dehydrogenase* (*dhod*) gene [29] and the *janus B* (*janB*) gene [30]. 5' untranslated regions (5'UTRs) are responsible for the translational repression in all these genes. In *dj*, the regulatory element is a translational repression element named TRE which is located 60 nucleotides upstream of the translational start site in the 5' untranslated region [31]. The dj protein is thought to be involved in the maturation of elongated spermatids during spermiogenesis [22]. Gene *dj* has a dual function: it is found to be expressed in the sperm tail but sequence comparisons suggest that it may play a role as a chromatin component [22]. It is a nuclear-encoded mitochondrial protein [26] localized to the nucleus of sperm heads during chromatin condensation and to the mitochondrial derivatives during sperm individualization. The dj protein has an internal mitochondrial localization site at the N-terminus end of the protein next to a single predicted protein cleavage site that plays a role in localization of the dj protein along the sperm flagellum [26]. Since dj resembles histone H1 protein in its basic nature it is thought to function as a DNA binding protein [22]. Therefore dj could be a transition protein during the final phase of chromatin condensation in spermatids [26].

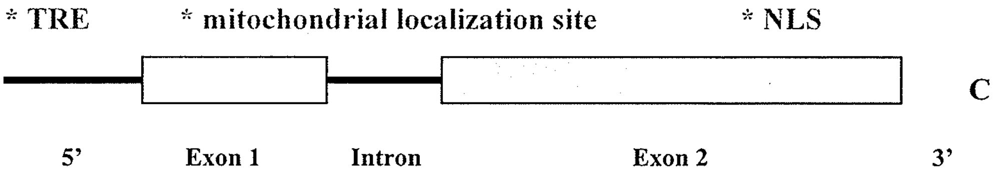


Figure 2c. Diagrammatic representation of *dj* gene showing exons and introns

1.4. Developmental genes and sex-related genes

The coding region of developmental genes has been widely conserved during the course of evolution. Developmental genes are genes that are involved in the body pattern formation of adult *Drosophila*. These genes are expressed at different levels in different tissues and between stages of development. The differential expression is caused by transcription factors interacting with genes and regulatory regions that are also conserved [32]. Sex-related genes are those that are involved in mating behavior, spermatogenesis, fertilization and sex determination. Sex-related genes are rapidly evolving between closely related species and there is a lack of selective constraints on their evolutionary pattern [9].

Spermatogenesis genes are sex-related genes controlling developmental process. Comparing the nucleotide sequences of spermatogenesis genes in *D. melanogaster* and *D. simulans* that control the pathway at different stages from early to late development of sperm will help understand the role of selection or neutrality during sequence gene evolution. This pattern can be compared to other developmental genes in different *Drosophila* species as developmental genes which are expressed at different parts of the body are thought to be highly conserved due to its role in development.

1.5. Sequence analyses of genes of spermatogenesis in *Drosophila*

Studies of molecular evolution analyze nucleotide variation in a gene within (Polymorphism) and between species (Divergence). Nucleotide changes can be broadly classified as synonymous and non-synonymous. Synonymous changes are nucleotide changes in the coding part of the gene that do not result in a change in the amino acid sequence of the encoded protein. Non-Synonymous substitutions are nucleotide changes that result in an amino acid change.

Under the hypothesis of neutral evolution, the ratio of non-synonymous to synonymous changes within species is expected to be equal to the ratio of non-synonymous to synonymous substitutions between species [33]. When a mutation is advantageous it gets selected and sweeps through a population. Thus the divergence of a species is increased when a particular mutation is selected and gets fixed in one species and not in the other species. Non-synonymous substitution can modify the protein structure and an increase in interspecific non synonymous over synonymous substitutions is suggestive of adaptive evolution driven by positive selection. When a mutation is disadvantageous, it gets selected against within species [33]. Non-synonymous and synonymous substitutions vary largely from gene to gene. Genes with a role in development show low rates of polymorphism and divergence.

Sex related genes show a higher rate of non-synonymous substitutions per non-synonymous site (Ka) to synonymous substitutions per synonymous site (Ks) between closely related species of *Drosophila* [9]. *Bam*, *aly* and *dj* are genes that affect the spermatogenesis pathway at different stages of development. Sequence comparisons of these three genes are likely to show patterns of positive selection as these genes are sex-related and also the selective constraints due to their developmental role. Within the sperm developmental pathway, evolutionary conservation of early versus late genes is expected.

1.6 Objectives and Aim

It is known that only male hybrids between closely related species of *Drosophila* are viable and sterile and that genes with a role in reproduction show a common pattern of rapid evolution between species. Nothing is known about the differentiation and interspecies divergence experienced by genes that control sperm development in *Drosophila*. Some specific questions are:

1. Are there differences in pattern of molecular evolution of genes that play a role at early, mid and late stages of sperm development ?
2. What is the role of selection in shaping within species polymorphism and interspecific divergence ?
3. How do genes of spermatogenesis compare to other developmental genes in levels of polymorphism and divergence within *D. melanogaster* ?

2. MATERIALS AND METHODS

2.1 *Drosophila* stocks

Samples were collected from sixteen isofemale *Drosophila melanogaster* strains from Winnipeg (Established by Dr. A. Civetta). Nine *Drosophila melanogaster* strains from Zimbabwe (Africa) and five *Drosophila simulans* strains from California were kindly provided by Dr. A.G. Clark (Cornell University). Flies were maintained on standard cornmeal molasses agar media (Table 2) and transferred to new vials every 14 days. This time interval produces a new generation of flies and the next culture is started.

Ingredient	**Quantity**
Cornmeal	65 g
Yeast	13 g
Agar	6.5 g
Cold water	170 mL
Boiling water	760 mL
Molasses	45.5 mL
99% Propionic acid	5 mL
10% Tegosept (50g methyl hydroxybenzoate per 500 ml 95% ethanol)	20 mL

Table 2: Cornmeal molasses agar medium recipe

Mix cornmeal, yeast and agar in cold water. Add the mix to boiling water and stir well. Then add molasses and mix well. Let the mix cool at room temperature to 60°C. Add tegosept and propionic acid.

2.2 DNA Extraction

Genomic DNA was extracted from each *Drosophila* strain using a standard DNA extraction protocol. Briefly, 5 to 10 flies were macerated in 100 μl of homogenizing buffer (0.1 M TrisHcl, 0.1 M EDTA, 1% sodium dodecyl sulphate in ddH_20) placed on ice. The mixture was incubated at 70°C for 30 minutes. Next, 14 μl of 8 M potassium acetate were added to the mixture and it was left on ice for 30 minutes followed by centrifugation at 4°C for 20 minutes at 14,000 rpm. The supernatant containing the DNA was transferred to a new microcentrifuge tube, 50 μl of 100% isopropanol were added and incubated at room temperature for 10 minutes. The tubes with the sample were spun for 10 minutes at room temperature (14,000 rpm) and the supernatant was discarded leaving the DNA pellet intact. The DNA pellet was washed two times with 40 μl of 70% ethanol and air dried for 30 minutes. The DNA was resuspended in 40 μl of nuclease free water (ddH_20) and stored at -20°C.

2.3 PCR amplification

PCR amplifications were carried out in a MJ Research PTC-200 Peltier Thermal Cycler for the *Drosophila* genes *bam*, *aly* and *dj*. Primers were designed using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3www.cgi).

The entire coding region of the *Drosophila bam, aly* and *dj* gene was PCR amplified using the following forward and reverse primer pairs which were designed using the *D. melanogaster* GenBank sequence entry X56202 (*bam*) (Table 3a), NT_037436 (*aly*) (Table 3b), NT_033777 (*dj*) (Table 3c).

```
5'
TTCTGGGACTCGACATGATATCGATACGTTAACAACAAAGAGTCTGGACGCCATCATTCTT
CCTCTTTCTCCTGAATTCGCAGACAGCGTGGCGTCAGGCATTTCAAACGGTAAAAAGAACC
TGGCGATAAGGAAAGATTTAAAAGGCAAAAATCGAGTGATTTGTGTGATTTAACTTAAGA
ATAATGCTTAATGCACGTGACGTGTGTCCTGAGGGCAACGACGACCAGCAGTTGGACCAC
AATTTTAAGCAGATGGAGGAGCATTTGGCCTTAATGGTGGAAGGCAATGAAAACGAAGAT
CCGAGGAAAGCCACTTGTGAGTACGAGGATACGAACGAAGATGGTGCAACCTGCACATCG
GGCGTTTTATCCGAAATCCAGGAGAACTTCGGTAGACTCCGGTTGTGTGACGTTACTGCAC
CACTCCTCGAATTCCACGGTTTGGATTGCTTGCAACAGATTCAAAAGCGCTCGCGCCATTT
TGCATTCGACGGTTCTCCGGCCAAGAAGTCGCGATCCGGAGGCGTGTTGGTCACCGGGCCA
AAGCAGAAGCAACTGCAGAAGGAAAATGTGTGGAACCGGAAGAGTAAAGGCTCTGCGTC
CGCGGATAATATTGAGAAACTGCCCATAACTATTGAGAAACTGCATATGATTGGTCTGCAC
GGCGATTGGTGAGTCTTCTGGAGTATA TCCCAAATAT ATCACATAATAAAAAGCTC
CTTATCTAAAC AATAGCTTAGAGCACAACGCCGTGCTGCGTTTGATGAATCTGTTCA
GATCCCTGCATGATCACCTGACCGCCGATTTGGGCTTCTCGCGCCAAAACTCAATGCCCTC
GGACTATCTGTTCGATATGCCGGTGAAGAGCACGATGCCTAAGAGCTTGAATGTGCGCTAC
CAACTGCAGGTGCTGTGCACCAAAGTAGAGCGCTTCCTTGTCCAGCAGCGCCGCACCTTGG
AGGCGAATCGCCACTTCGATTTCGAGAAATACGACGAGTGTGACAAGTTGCTTAAGGGTTT
CGCATCCTATTTGGACAACTTCAAACTGCTTTTAAAGCCCAAAATGCGCAATCGAAACGGA
AACTCGGGGAGCAATGCGGACAAGTGTAAGCTGTAGATTTGCAAGCAACCATTCAGCT
ATTCCTGCAACGATTTTATTAT TTACAGTCCATACTCAGCGCATGGAGAGATTGCTA
ATTGGTCTGCGCGATTGGATCAAGGCTGCGCATCTCAGTGTGCACGTATTTAACTGGGAAA
TGGATCTGGAGCACCGCTACTCCGGGGCCATGACCGAAAGCCACAAGTCGTTGAACGAGC
GGGCCATCCTTTTGTCCGGTGCCGAGCTAAGGGCGGCCGAAGCGCGTGGAATCAGTGCGG
AGGATCTGTTCATCGCCCAGAGATACAAACTGGGAGGTCCGATCTATTGCGTTCTGGAGCA
GCATGAGTTCCTCTCCGCTCTGATCGCCAATCCAGAGACCTATTTCCCGCCCAGTGTTGTCG
CCATTTGCGGGCCACAGAAGCTTGGCGCAGTGAGCATGGAGCAGCCGTCAGCGTCGGAGG
AGGAGTTTGAGGAGACCGAGGAAGTGCCATCATCGCCACCTCGTCACACCGGACGTGTAC
CTCGCTTCAGAAGCTAAACTAATGCTGTGCACATCGATAAAAGAATGACAGCAAATATGC
AATTTAAAAAAGCTACTCTTCTCATGGGAAGCAATAATTTCGTAAAGTAAACATATCTATA
GTGTAAGATATATTTGTCCAATAGTGCGGACTCCATATTTGTATTCGTGAATAAGCTTATAT
AAGCTTTTTAAAAATATTTATCAAATCGATACAAACAAAATCAAAATGAAAACGATTTATT
ACCCCTGTTTTGAGATTGATAACAAATTTATATAAGTTTAACTGTGTTTACATTTATTTGGC
AAAACTACAAATGTGTTTGCTTTTCACTTTTATAAATCTGTATTTTACTTAAACTTTAGAAA
TAAGAAATCCTTTAGTGCCTGAATTTATTTTGCAACTACGTTTTATTTGTATGAGGAACTTA
CCAGTTTTTCTTATTTGCTTTGCATTTGTATTTTGAAAGTCAAATAAATATTTACGATTTGTG
TTTGGACT 3'
```

Table 3a: PCR primers for the gene *bam*. Translational start codon and stop codon are highlighted in yellow and green respectively. Introns are marked with bolded pink font. Forward and reverse *bam* PCR primers are highlighted in blue.

```
5'
ATTTCAGCATGTCAGTTGATCCACTATCAATCGATAATTTCACAATCCAATCTGAGATCTGC
GAAGAAAACGAGTTTCTGGCAAATATAGGATTACTATCTACGACAACGTAATATTTTCTCT
TTTCTATCAAATATTAAAATAGAACTATAACTGTGCTACCATCGGTAGAATGTCGCGTCAT
CAATTGAAGAAACCCAGAAAGATGGTGGCGGCATGGCAAAACGATGAATTATTTATTAAA
CGCCCAAATTTCGCCCCGCGTATTAGGATTTCCGAAAAGCCAGAGATCCAGGGAAGAATT
AAACCAGGCGTGGCGTCCAAAAGGACTGAGAACTTTACAAAGAAGCCGTCCAATATATCT
GTAGATGTTTCGGAGGACGAGAAAGCGAAGGAAAAGGAAAAGGAGCAGGATCCCTACTC
CAATGACTTTATACTTGGCAAGAGGTTCGTAATGGGAAAGATTCCCTAGAGATCCCTTT
AAGTGCTTAATTGTTCATTCTCCTTTAGATTGTACAATTTCCTGAAGTATCTCAGCTCTC
ACCGTTGGATTTGGTGTGAGTTCGTCGACTCCTTCCTGGACAAGCCGACCCTGACCATGG
GCTACGATATGAAGCGCTTCATAGCGGAGTACTGTCCGCTCCTGCACTCTTGCTTCATGCC
CCGCAGAGGATGGCAATTGGTACGTCGGAATATGGGGAAGGCGCGTCGATTTTCGGCCGC
CTTCATCGAGCTGGAACGCGAAGAATTGGAGTGCCAGCGCCGCATTGTGCGCCAGTTGCA
GCAGCATAAGTTCAATCCCAAGGAGAACGTGGGCTACTTGGACCAGATACCCAAGCGTGT
GCCCCTGCCACTGGCCAAGGATGCCACGGTCAGCAGTTTTCTGCACGGAAACTCCTTTGAG
GGCATCGTCAATGGCACTGTCATGGGCTACGATCCGCAGGACTACACCTATCTGGTTCGAT
TCAATAGAAACGACAATGCAGTCGTGCTCAGTCTTCCGGATTCACAGCTCTATTCCGACGA
GGAAACCGCGGCGGTTCCCTTGTCAATTATTATGCGCGGCAACAAATCGTCCTCGGTTATT
TCGGAGAGCGCCAAGACCGAGAAGTTCGGAAACAAGAGGTACACCAAGGAACTTCTGGA
ATCAGTGCTAAGGGTTGGTAAACTACAGGATGTCAAGCACAAGATCCTCATGGACTTGGC
CCGAATGAATGAGGATTTCGAGACATTCAAGGAGATTGGTTCTTCAAGTAGTCGTCGCGAT
GCCAAGGTCACACCTCAGCGTGAGAATCTCCAGCGTCGCTATTCGGCCAGCATGATAACGC
TGCACCGAGTGAACGCTGATATCCTTGAACCGCTGCGCATCCTGCACGACTACCTGGTCGA
GTATCAGAAGCAGGACGAGGAGGAGGAGTCCAAAAGAGGTCGTCCCGCCAGCGAAGTCT
ATCAGAAGTGTCGCATGCAGGCGGAACAGGACCTCAAGACTGCCGCGGATGAGAAATTCC
TGAAGATAGAATCGGATCGCACGCAGGAGTTCGTCCGCAACCTTCACACCATACTGTATCT
CAATGGAAAGCTGGGGCGCGAGAACAGCTCCCATTTGGAGACGATTATCGCTGATCTGGT
TACCCACATGGTGGACAACATCCAGCCATCGCTGGGCCGGAAATTAAAAGATGGCGTCGA
TTCCCTGGAGCCTCTGCGTCAGCAGGTGGTGCAAATATTTAAAGACGTCAAAAAACCAGA
GCGCTTCCAAATCACCCAGCAGGCTCCGATGCAAACCGAGGATGGTATCTACAACTTTGTG
GTCGAGGCACAGCCGGATACTCCCAGCTAAACACACTACCTACTGGCCCTTTGGAATACTG
AAATAAAGCCTCGCTCTTATTTATGGCTCAATTAGGAGGAGTGTCATGTGCATTGGGAGTT
TGCCGGCAGAGGAGCTTAACGAAGTTTCTTGTGGCTGCTACCTTGTAGAGCTCTTTTGGTA
CTTACGCGAAGGGGTAGTTGGGGGGGGGGGTACGTGGTTCTAGGATTTATTTCAAGTTTC
CAGTGGCACGTTCCACAGGAAGTCAGCCGTAAAACAGTTGACAGATTGGCAACCAAATAA
GTAATGCGTATGTGAACGAAAACTTTAAGGGGGCACTCAAAAAAGGATAATCACTGGGGA
ATGTTTGCTTTTTCTAAAGGGCCTGTAAATGTGTTACAACTTTGATGATAGAAAGTGTAGTT
TATTAAATACTTAAAAATGATTAAGAAAAACTCTTACTATAACTCTCTAAGTACTAATAAA
CCCTGCAAAAAGGAGGTATGGAAATTTTCTCTTTTTTTGAGGCTAAATTTTGTGCAGCAAT
GTTGCATCCCCACATAAATCAGCATTTACGAGTAAAGTGTTTTCTATTTTTTAATTACTAGA
TGGATGCACTACGCTCGTAAATTGTGCGGGTGAGCCAGTTCCGCCGGGGAATTTGCCAACG
CCCATGTGGCCGCCATAAAGACATCATTAATATGCCTCCGGATGCACTGCGTCGCTCTGTC
CCCCGGGA 3'
```

Table 3b: PCR primers for the gene *aly*. Translational start codon and stop codon are highlighted in yellow and green respectively. Intron is marked with bolded pink font. aly1 forward and aly1 reverse primers are highlighted in red. aly2 forward and aly2 reverse primers are highlighted in blue. aly3 forward primer is underlined and aly3 reverse primer is in grey.

```
5'
CTTTGCAATTCGTTTTATTTATTTCTAGCAGTCAATTAAGTTCTTTTGGATCTAAGAGTTTCG
TGGGAGACGATAAATTTCTCATTAGATTGATTTTGATCTGATGATCTGAGATAATAATGTC
AGTTAAACTTGTATAGTTTTGGGGGCAGGTTAGATCTCAGATTCAGTTTAGATCCTGATTCC
ACAGACAAATAGTCTCCAGCTGTGGTTTTTTCAAAATTCTTTGTAAAACTTTTGGTACAAA
ATTTAAAAATTTTTCTCGAAATGTTTAAGAGAACCGCTTTAATTTTACGTCGGTGCTTTCAG
CCCACTTTTATACGGCCTCACCACATCAATGTCCTTGAGAACTTTAAGGAAGGTATGCAGT
GAACTCATATGCCTGGTTACCACTTGTTACTGTTAATACTACTTCACAACCGATGACC
TTCCCAATCAGGGGCAAGCAAAATTTGTCGATGTCTCTATTCACGATCCGCAACACATTC
GTTCTGCACTCGTCAGTCCAATGCAACGAAAGTTCTTGCAAGACCTGGAGCAGCAACAGA
CTGTTAGGATCAAGTGGTTTAAGGAAGGGAATCAGGATGAACTTGAAAACATGAAAAATG
AATGCCGGAGGCTAGCTCTAGAAATCATCATGGCTGCTAAAGGTGGCGACATCAAAAAAG
CCTGCAAGGAACTGGCTGAAAAAGAAAAGTGCAAGCAGATAGAACTGAAAAAGAAATGC
AAGGAATTGGAGAAGAAGACGAAGTGCGCGAAGAAAGACCCTTGCAAAAAGAAAGATCC
TTGCAAAAAGAAAGATCCCTGCAAAAAGAAAGATCCTTGCAAAAAGAAAGATCCTTGCAA
AAAGAAAGATCCCTGCAAAAAGAAAGATCCTTGCAAAAAGAAAGATCCGTGCAAAAAAA
AGGGTGGGGACCTAAAAAAGAAGTGCAAAAAATTGGCCGAAAAGGAAAAGTGCAAAAAA
CTGGCCAAAAAAGAAAAAATGAAAAAGTTGCAGAAAAAGTGCAAAAAAATGGCTCAGAA
GGAAAAATGCAAGAAAATGGCTAAAAAAGACAAATGCAAGAAAAAGTGAAGCTTTCGCG
GATTATTCAATGAAATACATACGTACCTGGTTTAATTCATTCAGCTCTGTTCAACGCGGCTT
TATCTAAAATATGGTTTTTTCATAATATACAATACGGCATTTTACCGAAAAATTAGATTTTA
TTTATTTAAAAAAATAACAAGGGGGAAAACAGTTAATGAGCATGTAACCCCCAGCTTTCG
AGTAATGAGTCCGTGGCAAGATTTCGTCGTTCATACGGACAGACTGATAGTCAGACGGTCA
TGTTTGTTAAGGAATCTATCTATATATATATATATAGATTTATATATCATATAAATAAA 3'
```

Table 3c: PCR primers for the gene *dj*. Translational start codon and stop codon are highlighted in yellow and green respectively. Intron is marked with bolded pink font. dj1 forward and reverse primers are highlighted in grey. dj2 forward and reverse primers are highlighted in pink.

PCR reactions for *bam* were carried out in 200μl PCR tubes with 1.5μl of 50mM $MgCl_2$ (3mM concentration), 0.6 μl of 10mM primers (each of forward and reverse), 2.5 μl of 10X buffer, 0.6 μl of 10mM dNTPs, 0.2 μl of Taq polymerase (5U/μl), 2 μl of DNA sample and brought up to 25 μl with double distilled H_20. Reactions were carried out for 30 cycles of 1 minute at 95°C for denaturation, 2 minutes at 63°C for annealing and 2 minutes at 72°C for extension. This was followed by a final extension at 72°C for 3 minutes. For *D.simulans*, the PCR conditions were the same as above except for $MgCl_2$ concentration of 4mM and an annealing temperature of 49°C for 3 minutes.

PCR reactions for *aly* were as described for *bam*. For *D. melanogaster* samples, the reaction was carried out in two different thermocycling profiles using; (i) aly1 Forward and reverse primers, aly2 Forward and reverse primers (Table 3) with 95°C for 4 minutes, 95°C for 1 minute, 65°C for 1 minute (-1/cycle), 72°C for 1.5 minutes (14 cycles) followed by 36 cycles of 1 minute at 95°C for denaturation, 1 minute at 55°C for annealing and 1.5 minutes at 72°C for extension. This was followed by a final extension at 72°C for 2 minutes. (ii) aly3Forward and reverse primers (Table 3) with 95°C for 2 minutes, 95°C for 1 minute, 65°C for 1 minute, extension at 72°C for 2 minutes (31 cycles) and a final extension of 72°C for 3 minutes. For *D. simulans* samples, the reaction was carried out in the following thermocycling profile; 95°C for 2 minutes, 95°C for 1 minute, annealing at 65°C for 1 minute, extension at 72°C for 2 minutes (31 cycles) and a final extension of 72°C.

PCR reactions for *dj* were as described for *bam* except for a 4.5 mM MgCl2 concentration. Reactions were carried out using dj1 forward & reverse dj2 forward & reverse primers (table 3) for 30 cycles of 1 minute at 95°C for denaturation, 2 minute at 54°C for annealing and 2 minutes at 72°C for extension. This was followed by a final extension at 72°C for 3 minutes. *D.simulans* samples were amplified using dj1 forward & reverse, dj2 forward & reverse primers (table 3) with the same concentrations and conditions except for an annealing temperature of 53.9°C for 2 minutes (dj1 forward & reverse) and 51.5°C for 2 minutes (dj2 forward & reverse).

In order to confirm the presence of a single amplification product of *bam*, *aly* and *dj*, the PCR reactions were subjected to electrophoresis at 120 volts in a 1% agarose gel containing 3.75 µl of ethidium bromide.

2.4 PCR product cleaning and quantification

In order to remove primer dimers that can interfere with the sequencing reaction, the PCR products were cleaned using the "Wizard SV Gel and PCR Clean-Up System kit" by Promega (cat no: A9281, Madison, U.S.A). The quick protocol provided by the kit was followed for cleaning the PCR products. An equal volume of membrane binding solution was added to the PCR reaction. The prepared PCR product was transferred to the SV mini column inserted into the collection tube and incubated for 1 minute followed by centrifugation at 10,000 × g for 1 minute. The flowthrough was discarded and the mini column was reinserted into the collection tube. 700 µl of membrane wash solution was added and centrifuged at 10,000 × g for 1 minute. The membrane was washed again with 500 µl of membrane wash solution. Next, the mini column was transferred to a clean 1.5 ml microcentrifuge tube. The DNA was eluted by adding 50 µl of nuclease free water to the column followed by one minute incubation and centrifuged at 10,000 × g for 1 minute. The cleaned DNA was stored at -20 °C.

The amount of amplified PCR product to be added in the sequencing reaction was quantified against a low mass ladder (Invitrogen) using the Quantity One software, Biorad. Three µl of cleaned DNA was run along with the low mass ladder in a 1% agarose gel. Using the low mass DNA ladder of known concentrations, *bam*, *aly* and *dj* PCR product concentrations in the cleaned reaction was estimated. Sequence reaction requires 25 to 100 fmoles of double stranded DNA. Depending on the length of the PCR product and the total amount of DNA in the cleaned reactions, the concentration that has to be added for sequencing is determined by using information in "Table for estimating the dsDNA concentration" (CEQ 2000 Dye Terminator Cycle Sequencing with Quick Start Kit manual) as reference.

2.5 Sequencing reaction

Following quantification, sequencing reaction was carried out in a 20 µl volume. Two µl of DCTS quick start master mix (dNTP mix, ddNTP dye terminators and polymerase enzyme), 1.2 µl sequencing buffer (Beckman Quick start kit product no: 608120), 1 µl forward or reverse primer (Table 4) and DNA aliquot according to the quantification of different samples were added [34]. Sterilized ddH_20 was added to bring up the reaction to a 20 µl volume. The reactions were carried out in the thermocycler.

For *bam* (Table 4a) , *aly* (Table 4b) and *dj* (Table 4c) the reactions were carried out using the following sequencing primers.

5'
GTTCTGGGACTCGACATGATATCGATACGTTAACAACAAAGAGTCTGGACGCCATCATTCT
TCCTCTTTCTCCTGAATTCGCAGACAGCGTGGCGTCAGGCATTTCAAACGGTAAAAAGAAC
CTGGCGATAAGGAAAGATTTAAAAGGCAAAAATCGAGTGATTTGTGTGATTTAACTTAAG
AATAATGCTTAATGCACGTGACGTGTGTCCTGAGGGCAACGACGACCAGCAGTTGGACCA
CAATTTTAAGCAGATGGAGGAGCATTTGGCCTTAATGGTGGAAGGCAATGAAAACGAAGA
TCCGAGGAAAGCCACTTGTGAGTACGAGGATACGAACGAAGATGGTGCAACCTGCACATC
GGGCGTTTTATCCGAAATCCAGGAGAACTTCGGTAGACTCCGGTTGTGTGACGTTACTGCA
CCACTCCTCGAATTCCACGGTTTGGATTGCTTGCAACAGATTCAAAAGCGCTCGCGCCATT
TTGCATTCGACGGTTCTCCGGCCAAGAAGTCGCGATCCGGAGGCGTGTTGGTCACCGGGCC
AAAGCAGAAGCAACTGCAGAAGGAAAATGTGTGGAACCGGAAGAGTAAAGGCTCTGCGT
CCGCGGATAATATTGAGAAACTGCCCATAACTATTGAGAAACTGCATATGATTGGTCTGCA
CGGCGATTGGTGAGTCTTCTGGAGTATATCCCAAATATATCACATAATAAAAAGCTC
CTTATCTAAACAATAG CTTAGAGCACAACGCCGTGCTGCGTTTGATGAATCTGTTCAG
ATCCCTGCATGATCACCTGACCGCCGATTTGGGCTTCTCGCGCCAAAACTCAATGCCCTCG
GACTATCTGTTCGATATGCCGGTGAAGAGCACGATGCCTAAGAGCTTGAATGTGCGCTACC
AACTGCAGGTGCTGTGCACCAAAGTAGAGCGCTTCCTTGTCCAGCAGCGCCGCACCTTGGA
GGCGAATCGCCACTTCGATTTCGAGAAATACGACGAGTGTGACAAGTTGCTTAAGGGTTTC
GCATCCTATTTGGACAACTTCAAACTGCTTTTAAAGCCCAAAATGCGCAATCGAAACGGAA
ACTCGGGGAGCAATGCGGACAAGTGTAAGCTGTAGATTTGCAAGCAACCATTCAGCTA
TTCCTGCAACGATTTTATTAT TTACAG TCCATACTCAGCGCATGGAGAGATTGCT
AATTGGTCTGCGCGATTGGATCAAGGCTGCGCATCTCAGTGTGCACGTATTTAACTGGGAA
ATGGATCTGGAGCACCGCTACTCCGGGGCCATGACCGAAAGCCACAAGTCGTTGAACGAG
CGGGCCATCCTTTTGTCCGGTGCCGAGCTAAGGGCGGCCGAAGCGCGTGGAATCAGTGCG
GAGGATCTGTTCATCGCCCAGAGATACAAACTGGGAGGTCCGATCTATTGCGTTCTGGAGC
AGCATGAGTTCCTCTCCGCTCTGATCGCCAATCCAGAGACCTATTTCCCGCCCAGTGTTGTC
GCCATTTGCGGGCCACAGAAGCTTGGCGCAGTGAGCATGGAGCAGCCGTCAGCGTCGGAG
GAGGAGTTTGAGGAGACCGAGGAAGTGCCATCATCGCCACCTCGTCACACCGGACGTGTA
CCTCGCTTCAGAAGCTAAACTAATGCTGTGCACATCGATAAAAGAATGACAGCAAATATG
CAATTTAAAAAAGCTACTCTTCTCATGGGAAGCAATAATTTCGTAAAGTAAACATATCTAT
AGTGTAAGATATATTTGTCCAATAGTGCGGACTCCATATTTGTATTCGTGAATAAGCTTAT
ATAAGCTTTTAAAAATATTTATCAAATCGATACAAACAAAATCAAAATGAAAACGATTTA
TTACCCCTGTTTTGAGATTGATAACAAATTTATATAAGTTTAACTGTGTTTACATTTATTTG
GCAAAACTACAAATGTGTTTGCTTTTCACTTTTATAAATCTGTATTTTACTTAAACTTTAGA
AATAAGAAATCCTTTAGTGCCTGAATTTATTTTGCAACTACGTTTTATTTGTATGAGGAACT
TACCAGTTTTTCTTATTTGCTTTGCATTTGTATTTTGAAAGTCAAATAAATATTTACGATTTG
TGTTTGGACT **3'**

Table 4a : <u>Sequencing primers for the gene *bam*</u>. Translational start codon and stop codon are highlighted in yellow and green respectively. Introns are marked with bolded pink font. *bam* forward primer (F2) and reverse primer (R2) are highlighted in pink. *bam* forward primer (F3) and reverse primer (R3) are highlighted in violet. *bam* forward primer (F4) and reverse primer (R4) are highlighted in dark yellow. *bam* forward primer (F6) used in *D. simulans* strains is highlighted in grey.

```
5'
ATTTCAGCATGTCAGTTGATCCACTATCAATCGATAATTTCACAATCCAATCTGAGATCTGC
GAAGAAAACGAGTTTCTGGCAAATATAGGATTACTATCTACGACAACGTAATATTTTCTCT
TTTCTATCAAATATTAAAATAGAACTATAACTGTGCTACCATCGGTAGAATGTCGCGTCAT
CAATTGAAGAAACCCAGAAAGATGGTGGCGGCATGGCAAAACGATGAATTATTTATTAAA
CGCCCAAATTTCGCCCCGCGTATTAGGATTTCCGAAAAGCCAGAGATCCAGGGAAGAATT
AAACCAGGCGTGGCGTCCAAAAGGACTGAGAACTTTACAAAGAAGCCGTCCAATATATCT
GTAGATGTTTCGGAGGACGAGAAAGCGAAGGAAAAGGAAAAGGAGCAGGATCCCTACTC
CAATGACTTTATACTTGGCAAGAGGTTCGTAATGGGAAAGATTCCCTAGAGATCCCTTT
AAGTGCTTAATTGTTCATTCTCCTTTAGATTGTACAATTTCCTGAAGTATCTCAGCTCTC
ACCGTTGGATTTGGTGTGAGTTCGTCGACTCCTTCCTGGACAAGCCGACCCTGACCATGGG
CTACGATATGAAGCGCTTCATAGCGGAGTACTGTCCGCTCCTGCACTCTTGCTTCATGCCCC
GCAGAGGATGGCAATTGGTACGTCGGAATATGGGGAAGGCGCGTCGATTTTCGGCCGCCT
TCATCGAGCTGGAACGCGAAGAATTGGAGTGCCAGCGCCGCATTGTGCGCCAGTTGCAGC
AGCATAAGTTCAATCCCAAGGAGAACGTGGGCTACTTGGACCAGATACCCAAGCGTGTGC
CCCTGCCACTGGCCAAGGATGCCACGGTCAGCAGTTTTCTGCACGGAAACTCCTTTGAGGG
CATCGTCAATGGCACTGTCATGGGCTACGATCCGCAGGACTACACCTATCTGGTTCGATTC
AATAGAAACGACAATGCAGTCGTGCTCAGTCTTCCGGATTCACAGCTCTATTCCGACGAGG
AAACCGCGGCGGTTCCCTTGTCAATTATTATGCGCGGCAACAAATCGTCCTCGGTTATTTC
GGAGAGCGCCAAGACCGAGAAGTTCGGAAACAAGAGGTACACCAAGGAACTTCTGGAAT
CAGTGCTAAGGGTTGGTAAACTACAGGATGTCAAGCACAAGATCCTCATGGACTTGGCCC
GAATGAATGAGGATTTCGAGACATTCAAGGAGATTGGTTCTTCAAGTAGTCGTCGCGATGC
CAAGGTCACACCTCAGCGTGAGAATCTCCAGCGTCGCTATTCGGCCAGCATGATAACGCTG
CACCGAGTGAACGCTGATATCCTTGAACCGCTGCGCATCCTGCACGACTACCTGGTCGAGT
ATCAGAAGCAGGACGAGGAGGAGGAGTCCAAAAGAGGTCGTCCCGCCAGCGAAGTCTAT
CAGAAGTGTCGCATGCAGGCGGAACAGGACCTCAAGACTGCCGCGGATGAGAAATTCCTG
AAGATAGAATCGGATCGCACGCAGGAGTTCGTCCGCAACCTTCACACCATACTGTATCTCA
ATGGAAAGCTGGGGCGCGAGAACAGCTCCCATTTGGAGACGATTATCGCTGATCTGGTTA
CCCACATGGTGGACAACATCCAGCCATCGCTGGGCCGGAAATTAAAAGATGGCGTCGATT
CCCTGGAGCCTCTGCGTCAGCAGGTGGTGCAAATATTTAAAGACGTCAAAAAACCAGAGC
GCTTCCAAATCACCCAGCAGGCTCCGATGCAAACCGAGGATGGTATCTACAACTTTGTGGT
CGAGGCACAGCCGGATACTCCCAGCTAAACACACTACCTACTGGCCCTTTGGAATACTGAA
ATAAAGCCTCGCTCTTATTTATGGCTCAATTAGGAGGAGTGTCATGTGCATTGGGAGTTTG
CCGGCAGAGGAGCTTAACGAAGTTTCTTGTGGCTGCTACCTTGTAGAGCTCTTTTGGTACTT
ACGCGAAGGGGTAGTTGGGGGGGGGGGGTACGTGGTTCTAGGATTTATTTCAAGTTTCCA
GTGGCACGTTCCAC 3'
```

Table 4b : <u>Sequencing primers for the gene *aly*</u>. Translational start codon and stop codon are highlighted in yellow and green respectively. Intron is marked with bolded pink font. *aly* forward (aly1. f2) and *aly* reverse (aly1.r2) sequencing primers are highlighted in pink. *aly* forward (aly2. f1) and *aly* reverse (aly2. r1) sequencing primers are highlighted in blue. *aly* forward (aly3. N1f) and aly reverse (aly3. N1r) are highlighted in grey. *aly* reverse (aly2. N2r) is highlighted in teal.

5′ CTTTGCAATTCGTTTTATTTATTTCTAGCAGTCAATTAAGTTCTTTTGGATCTAAGAGTTT
CGTGGGAGACGATAAATTTCTCATTAGATTGATTTTGATCTGATGATCTGAGATAATAATG
TCAGTTAAACTTGTATAGTTTTGGGGGCAGGTTAGATCTCAGATTCAGTTTAGATCCTGATT
CCACAGACAAATAGTCTCCAGCTGTGGTTTTTTCAAAATTCTTTGTAAAACTTTTGGTACAA
AATTTAAAAATTTTTCTCGAAATGTTTAAGAGAACCGCTTTAATTTTACGTCGGTGCTTTCA
GCCCACTTTTATACGGCCTCACCACATCAATGTCCTTGAGAACTTTAAGGAAGG**TATGCAG**
TGAACTCATATGCCTGGTTACCACTTGTTACTGTTAATACTACTTCACAACCGATGAC
CTTCCCAATCAGGGGCAAGCAAAATTTGTCGATGTCTCTATTCACGATCCGCAACACATT
CGTTCTGCACTCGTCAGTCCAATGCAACGAAAGTTCTTGCAAGACCTGGAGCAGCAACAG
ACTGTTAGGATCAAGTGGTTTAAGGAAGGGAATCAGGATGAACTTGAAAACATGAAAAAT
GAATGCCGGAGGCTAGCTCTAGAAATCATCATGGCTGCTAAAGGTGGCGACATCAAAAAA
GCCTGCAAGGAACTGGCTGAAAAAGAAAAGTGCAAGCAGATAGAACTGAAAAAGAAATG
CAAGGAATTGGAGAAGAAGACGAAGTGCGCGAAGAAAGACCCTTGCAAAAAGAAAGATC
CTTGCAAAAAGAAAGATCCCTGCAAAAAGAAAGATCCTTGCAAAAAGAAAGATCCTTGCA
AAAAGAAAGATCCCTGCAAAAAGAAAGATCCTTGCAAAAAGAAAGATCCGTGCAAAAAA
AAGGGTGGGGACCTAAAAAAGAAGTGCAAAAAATTGGCCGAAAAGGAAAAGTGCAAAAA
ACTGGCCAAAAAAGAAAAAATGAAAAAGTTGCAGAAAAAGTGCAAAAAAATGGCTCAGA
AGGAAAAATGCAAGAAAATGGCTAAAAAAGACAAATGCAAGAAAAAGTGAAGCTTTCGC
GGATTATTCAATGAAATACATACGTACCTGGTTTAATTCATTCAGCTCTGTTCAACGCGGCT
TTATCTAAAATATGGTTTTTCATAATATACAATACGGCATTTTACCGAAAAATTAGATTTT
ATTTATTTAAAAAAATAACAAGGGGGAAAACAGTTAATGAGCATGTAACCCCCAGCTTTC
GAGTAATGAGTCCGTGGCAAGATTTCGTCGTTCATACGGACAGACTGATAGTCAGACGGTC
ATGTTTGTTAAGGAATCTATCTATATATATATATATAGATTTATATATCATATAAATAAA **3’**

Table 4c : Sequencing primers for the gene *dj.* Translational start codon and stop codon are highlighted in yellow and green respectively. Intron is marked with bolded pink font. dj1 forward and dj1 reverse primers are highlighted in grey. dj2 forward and dj2 reverse primers are highlighted in pink. dj3 forward primer is highlighted in blue.

By the end of the thermal cycling program of the sequence reactions for *bam*, *aly* and *dj*, 5 μl of stop solution was prepared by using 0.4 μl of 0.5 M EDTA, 2 μl of 3 M NaOAc, 1.6 μl of sterilized ddH_2O and 1.0 μl of glycogen added last. Five μl of stop solution is added to each of the sequence reaction mix. To precipitate the DNA, 60 μl of cold 95% ethanol was added, mixed and centrifuged at 14,000 rpm at 4°C for 15 minutes. The supernatant was removed and 100 μl of 70 % ethanol was added, centrifuged at 14,000 rpm at 4°C for 5 minutes and the supernatant was removed again and the pellet was washed twice with 70% ethanol. The DNA pellet was air dried for 45 minutes. The pellet was resuspended in 40 μl of sample loading solution. Samples were transferred to corresponding wells of the sample plate and covered with one drop of mineral oil. Separation buffer was added into the buffer plate (CEQ Beckman Coulter). The sample plate, buffer plate, gel cartridge and capillary array were installed in the CEQ 2000XL DNA Analysis System. The sample plate was run using the following system conditions; capillary temperature of 50°C, denature temperature of 90°C for 120 seconds, injection time of 27 seconds at 2.0 kV and separation time of 85.0 minutes at 4.2 kV. Raw sample data is collected and converted to the nucleotide base sequences by the CEQ system analysis software.

2.6 Sequence data analysis

The sequences obtained using different sequencing primers of each gene for each strain were individually aligned to the *D. melanogaster* Genbank sequence. Each alignment was manually inspected and mismatches due to sequence error were eliminated by inspecting sequences obtained from at least a forward and reverse sequencing primer from two independent PCR products. Four independent sequencing reactions were performed for each template. A single sequence for each strain was obtained as a consensus from these alignments. DNA sequences of all the strains were multiple aligned using the Clustal X program version 1.82 (ftp://ftp-igbmc.ustrasbg.fr/pub/ClustalX/) [35]. Following the sequence alignment, the analyses were performed using the programs DNAsp 4.0 (DNA Sequence Polymorphisms) (http://www.ub.es/dnasp/) [36] and MEGA2 (www.megasoftware.net) [37].

3. RESULTS

Sequencing was carried out on the three genes *bam*, *aly* and *dj*. The gene *bam* was sequenced in twenty four strains of *D. melanogaster* (sixteen strains from Winnipeg, North America and eight strains from Zimbabwe, Africa) and five strains of *D. simulans*. The gene *aly* was sequenced in twenty four strains of *D. melanogaster* (fifteen strains from Winnipeg, North America and nine strains from Zimbabwe, Africa) and six strains of *D. simulans*. The gene *dj* was sequenced in twenty five strains of *D. melanogaster* (sixteen strains from Winnipeg, North America and nine strains from Zimbabwe, Africa) and five strains of *D. simulans*. Analysis was carried out on the *bam* gene for exons one (499 bp), two (372 bp) and three (473 bp), introns one (66 bp) and two (62 bp). Analysis on *aly* gene includes the 5' upstream region (89 bp), exons one (275 bp), two (1336 bp) and an intron (69 bp). Analysis on *dj* include the 5' upstream region (70 bp), exon one (94 bp), two (644 bp) and an intron (98 bp).

3.1a Polymorphic and fixed sites

Polymorphic and fixed sites were observed for all the three genes *bam*, *aly* and *dj* for *D. melanogaster* and *D. simulans* strains in the 5' upstream (*dj*), coding and intron regions (*bam, aly* and *dj*) (Tables 5, 6 and 7).

In *bam,* 7 intraspecific polymorphisms were found within exon one of *D. melanogaster* with three non-synonymous and four synonymous polymorphisms. Introns one and two have one and two silent polymorphisms respectively. Exon two has two non-synonymous and seven synonymous intraspecific polymorphisms. Exon 3 has one non-synonymous and three synonymous polymorphisms (Table 5).

For *bam* in *D. simulans*, three non-synonymous and four synonymous polymorphisms were detected in exon one. Exon one has a nine base pair insertion and a twelve base pair deletion in *D. simulans*. Introns one and two have zero polymorphisms within *D. simulans*. Exon two has three non-synonymous and two synonymous polymorphisms. Exon three has seven non-synonymous and ten synonymous polymorphisms. A three base pair insertion polymorphism is found in exon three of *D. simulans* (Table 5). When a cross species mega blast in the trace archives of NCBI (http://www.ncbi.nlm.nih.gov) was carried out with the *Drosophila simulans bam* gene that was sequenced in this study, it gave an average percentage (%) identity of 97.13%. This confirms that the *bam* gene in *Drosophila simulans* isolated and sequenced was right.

bam

```
Position                            1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5
          1 2 3 3 4 4 4 6 6 8 9 9 9 2 2 2 3 3 3 4 6 6 8 2 3 3 3 7 8 8 9 9 9 0 0 2 4 4 4 4 4 5 5 5 5 5 5 6 7 1 1 1 2 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 7 8 8 8 0
Strains   9 2 2 9 0 6 9 3 8 9 0 7 8 2 3 4 4 6 9 8 1 5 8 2 1 2 3 7 2 8 7 8 9 4 9 1 1 3 4 5 9 0 1 2 3 4 8 9 8 1 2 5 2 4 6 7 8 9 0 1 2 3 4 5 6 7 0 2 6 6 3 7 9 8
Wpg1      G T G C G C T T A T A G A G G A G G T A G A C G T A C A A C T T A T C C G _ _ _ _ _ _ _ _ _ T G G G A A G C G A T A A T A T T G A G A G C A T C G T
Wpg2      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg3      . . . . . . . . . . . . . . . . C . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg4      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg5      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg6      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg7      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg8      . . . . . . . . . . . . . . . . C . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg9      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg10     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg11     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg12     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg13     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . A . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg14     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg15     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg16     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . A . . . . . . . . . . .
Zim35     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ C . . . . . . . . . . . . . . . A . . A . . . . . . . .
Zim18     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . A . . . . . . . . . . .
Zim22     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . A . . . . . . . . . . .
Zim10     . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim32     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim5      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim7      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim49     . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . _ _ _ _ _ _ _ _ _ . . . . . . . . . . . . . . . . . . . . . . . . . . . .

sim4      A G A G A G C C T C C A T T T C G C C T G C T A G T C C G A G C T T A A C G C A G A C T T C T G G A C T A G _ _ _ _ _ _ _ _ _ _ _ _ G A T G T A T C
sim6      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . _ _ _ _ _ _ _ _ _ _ _ _ . . . . . . . .
sim3      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . _ _ _ _ _ _ _ _ _ _ _ _ . . . . . . . .
sim7      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . _ _ _ _ _ _ _ _ _ _ _ _ . . . . . . . .
sim1      . . . . . . . . . . . . . . . . . . . . . . . . . G G . . . . . . A . C . . . . . . . . . . . . A . . . . . _ _ _ _ _ _ _ _ _ _ _ _ . . C . C . . .

R/S       R R R R R R S S R R R R R R R R R R R R R S R S S R R R S S R S S R R S R S S S S S S S S S R S S S S S R R                 S     S S R S R S R S S
F/P       F F F F F F F F F F F F F F F F P F F F P F F F F F P F F F F F P P F F F F F F F F F F F F P P P F F F F F                 P     P P F F F P F F F
                                                            P                   P                                                                 P
```

```
bam                                                                                               1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Position  5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3
          3 3 4 5 5 5 5 5 5 7 8 8 8 0 2 4 7 8 9 0 0 1 1 1 3 6 7 7 9 9 0 2 2 5 6 6 8 9 2 5 7 7 8 9 9 0 1 2 4 1 4 5 6 6 8 8 9 0 0 1 2 2 3 3 3 3 4 4 5 6 6 6 6 8 8 9 2 2 3
Strains   4 8 6 1 2 3 5 7 9 1 4 5 9 1 6 3 0 6 4 2 9 2 6 8 2 3 2 3 4 7 7 0 3 2 8 9 4 3 4 5 1 7 9 3 4 6 3 9 1 3 5 5 1 7 1 9 1 7 8 5 1 8 5 6 7 9 0 4 6 0 2 3 5 4 5 3 0 2 8
Wpg1      T A A C C T A C A A G T G T C C G A G C C G T G A C C C G A T A C T C A T A G G C C T A T A G A C C A C T C G G A G C T C C A C A A C G A C T T T C T G G C T
Wpg2      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg3      . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg4      . . . . . . . . . . . . . . . . . . . . T . . . . . . . T . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg5      . . . . . . . . . . . . . . . . . . . . T . . . . . A . T . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg6      . . . . . . . . . . . . . . . . . . . . T . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg7      . . . . . . . . . . . . . . . . . . . . T . . . . . . . T . . . . . . . . . . . . A A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg8      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg9      . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg10     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg11     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg12     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg13     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg14     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg15     . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . .
Wpg16     . . . T . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim35     . . . T . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . C . . . . . . .
Zim18     . . . . . . . . . . . . . . . . . . . . T . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . C . . . . . . .
Zim22     . . . . . . . . . . . . . . . . . . . . T . . . . G . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . C . . . . . . .
Zim10     . . . . . . . . . . . . . . . . . . . . T . . . . G . . T . . . . . . . . . . . . A . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . G .
Zim32     . . . . . . . . . . . . . . . . . . . . T . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim5      . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim7      . . . . . . . . . . . . . . T G . . . . T A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Zim49     . . . . . . . . . . . . . . . G . . . . T A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

sim4      C G C C T C T G T G A C T T C C G T C T A G A A A C C A G A T G T C G T C G T T T C T C A G T G C G C G T T A A T T T G G A T T G A C G G T C C A T A A A C T
sim6      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . C T A . . . . . . . . . . .
sim3      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . A . . . . . . . . . . . . . . . . . . . . . . . . . . ? . . C T A . . . . . . . . . . A
sim7      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . T . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ? C ? . . A . . . . . . . . . . .
sim1      . . . . . . . . . . . . . C . . C . . . . . . . C . . . . T . . . . . A . . . . . . . . . . . . . . A . . . . . A . . . . C ? C . C T A . . . . . . . . . . .

R/S       S S S S S S S S S S R R S S S S S R S R S S R S R S S R R R R S S R R R S R R S S S S S S S R R S S S R S S S R R R R R R R R S S S S S R R S S S R S R S S R S
F/P       F F F P F F F F F F F F F P P P P F F F F P F F P P P F P P P F F F F F F F F F F P F F F F F F P F F F P F F F F F F F F F F F F F P F P P F F F P F F F F F P P
                                                  P                             P                             P           P         P   P   P
```

bam	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Position	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4
	6	6	7	7	7	7	8	9	0	0	0	1	3	4	4
Strains	1	7	2	3	4	8	2	7	2	4	6	2	4	7	8
Wpg1	C	G	_	_	_	C	C	A	T	T	A	C	A	G	G
Wpg2	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg3	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg4	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg5	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg6	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg7	.	.	_	_	_	.		.	.	.	.	.	.	.	.
Wpg8	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg9	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg10	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg11	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg12	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg13	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg14	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg15	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Wpg16	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim35	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim18	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim22	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim10	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim32	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim5	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
Zim7	.		_	_	_	.	.	.	.	.	.	.	.	.	.
Zim49	.	.	_	_	_	.	.	.	.	.	.	.	.	.	.
sim4	G	?	C	A	G	G	A	C	A	C	A	T	G	A	T
sim6	.	?	.	.	.	.	.	.	.	.	.	.	.	.	.
sim3	.	G	.	.	C	.	.	.	.	T	.	.	.	C	?
sim7	.	A	.	.	C	.	.	A	.	.	T	.	.	.	?
sim1	.	G	_	_	_	.	C	.	T	T	.	?	.	.	.
R/S	R	S	S	S	S	R	R	S	S	S	R	S	S	S	S
F/P	F	P	F	F	F	F	F	F	F	F	P	F	F	F	F
							P	P	P	P				P	

Table 5: Polymorphic sites in **exon one**, **intron one**, **exon two**, intron two and **exon three** of the *bam* gene in *D. melanogaster* and *D. simulans*. Dots represent similarities between strains. R- replacement site, S-silent site, F-fixed site, P- polymorphic site. "–" indicates gaps. Only nucleotide positions where there are changes between strains in the exons and introns of the gene are shown.

The *aly* gene of *D. melanogaster* shows five non-synonymous and six synonymous intraspecific polymorphisms in exon one. There is a six base pair deletion in Zim30 strain in exon one. The intron has six silent polymorphisms and a two base pair deletion in Zim30 and Zim35 strains. Exon two has fifteen non-synonymous and forty synonymous intraspecific polymorphisms (Table 6).

The sequences of *aly* in *D. simulans* show ten non-synonymous and one synonymous polymorphisms in exon one. Intron one has four silent polymorphisms within *D. simulans*. A two base pair deletion in all *simulans* strains and a two base pair deletion in sim6 strain are found in the intron region of *D. simulans*. Exon two has six non-synonymous and fifteen synonymous polymorphisms. A six base pair addition is found in exon two of *D. simulans* (Table 6). When a cross species mega blast in the trace archives of NCBI (http://www.ncbi.nlm.nih.gov) was carried out with the *Drosophila simulans aly* gene that was sequenced in this study, it gave an average percentage (%) identity of 96.26 %. This confirms that the *aly* gene in *Drosophila simulans* isolated and sequenced was right.

aly	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
Position	0	0	3	4	5	5	5	6	6	6	6	6	6	7	8	8	9	9	9	0	0	0	0	1	1	1	1	2	3	3	3	4	5	5	5	6	6	7	7	7	8	9	9	9	0	0	0	0	0	0
Strains	2	9	3	5	1	2	9	1	2	3	5	8	9	8	0	9	3	8	9	2	3	4	8	3	4	6	8	6	1	4	9	7	2	6	9	4	5	0	1	6	3	1	4	5	2	3	4	5	6	7
Wpg1	G	G	A	A	A	A	C	C	C	G	G	T	T	C	A	A	C	G	A	A	A	T	A	G	C	T	G	A	C	A	T	G	C	C	T	C	T	A	T	C	C	C	A	G	G	A	A	A	A	G
Wpg2	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Wpg3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg4	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg8	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Wpg9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg11	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Wpg12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg13	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg14	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg15	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg16	.	A	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Zim35	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Zim18	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Zim22	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim10	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Zim32	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	G	.	.	.	.	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
Zim5	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T	.	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim7	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim49	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim30	.	.	G	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	_	_	_	_	_	_
sim3	G	G	G	T	C	G	A	A	A	T	G	G	C	T	A	C	A	T	G	G	A	C	G	A	G	G	T	A	C	C	T	A	T	C	A	T	C	C	C	T	C	A	T	G	_	_	_	_	_	_
sim6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	A	.	.	.	A	.	.	.	_	_	_	_	_	_
sim2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	.	.	.	_	_	_	_	_	_
sim4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	A	.	.	.	.	.	.	.	_	_	_	_	_	_
sim1	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	A	.	.	.	.	.	.	C	_	_	_	_	_	_
sim5	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	A	.	.	C	.	.	.	.	_	_	_	_	_	_
R/S	R	R	S	R	S	R	R	R	R	S	R	R	S	S	R	R	R	R	S	S	R	R	S	R	S	R	R	S	R	R	R	S	R	S	R	R	S	R	S	R	R	R	R	R						
F/P	P	P	P	F	F	F	F	F	F	F	P	F	P	F	P	F	F	F	F	F	P	F	F	F	P	F	F	P	P	F	P	F	F	P	P	P	F	F	P	F	P	F	F	P						
																	P																P				P			P										

																																																		1	1	1	1	1
aly	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	6	7	7	7	8	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	9	0	0	0	0	0
Position	3	5	5	6	6	6	7	7	7	7	7	8	9	9	0	1	2	7	7	0	2	3	6	7	8	9	9	1	2	7	0	1	1	5	5	6	7	9	0	0	0	0	1	2	2	7	8	8	8	0	0	1	2	3
Strains	1	0	8	3	4	5	2	4	6	7	9	9	0	5	5	5	7	2	5	5	9	8	5	4	7	5	8	6	6	3	6	6	7	1	4	4	7	9	0	2	7	8	7	1	8	6	1	3	7	2	8	3	2	7
Wpg1	C	T	A	A	A	A	C	A	A	G	T	T	A	C	C	T	T	C	C	C	C	G	C	C	T	C	G	A	G	C	T	A	C	G	G	G	C	A	A	A	T	T	C	A	C	A	A	C	A	A	C	T	A	A
Wpg2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg8	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	C	.	.	.	.	.	.	G	.	.	.	.	.	C	.
Wpg10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg11	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg14	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	G	.	C	C	.	.	.	.	.	.	.	T	.	.	.	C	.
Wpg15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Wpg16	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim35	.	.	.	G	.	.	.	.	_	_	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	T	C	.	.	T	.	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim22	.	G	.	G	.	G	.	.	.	.	.	.	.	.	.	C	.	.	G	.	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	G	.	.	.	.	.	.	.	.	.	C	.
Zim10	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.
Zim32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	G	T	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim5	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim7	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	G	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim49	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Zim30	.	.	.	G	.	.	.	.	_	_	.	.	.	.	.	C	.	.	G	T	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
sim3	T	T	G	G	A	A	A	T	_	_	C	T	A	T	C	T	C	C	G	C	C	G	C	T	T	C	C	T	C	C	T	A	C	C	C	T	G	A	C	C	T	C	C	A	T	G	G	T	A	T	T	C	C	G
sim6	.	.	.	.	.	.	.	.	_	_	.	_	_	.	.	.	.	T	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
sim2	.	.	.	.	.	.	.	.	_	_	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A
sim4	.	.	.	.	.	.	C	.	_	_	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A
sim1	.	.	.	.	.	.	.	.	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	.	.	.	.	.	?	?	.	.	.	?	.	.	.	.	.	.	A
sim5	.	.	.	.	G	.	.	.	_	_	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	?	.	.	.	.	.	.	A
R/S	S	S	S	S	S	S	S	S			S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	S	R	S	S	S	R	R	S	R	R	S	S	S	R	R	R	S	S	R	R	R	S	S	S
F/P	F	P	F	P	P	P	F	F			F	P	P	F	P	P	F	P	P	P	P	P	P	P	P	P	P	P	F	P	P	P	P	F	F	F	F	P	F	P	P	F	P	P	F	F	P	F	P	F	F	F	P	P
							P				P																																											

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
aly	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	6
Position	5	7	8	8	8	8	9	9	1	1	1	1	2	2	6	6	7	7	8	1	2	3	6	6	9	0	0	2	4	4	7	9	9	9	0	1	1	3	3	4	4	5	1	2	3	4	4	5	6	6	7	8	8	0
Strains	8	4	3	4	6	9	3	7	6	7	8	9	0	1	4	9	1	5	4	7	7	5	7	9	5	4	7	1	6	9	7	3	4	5	9	2	5	0	1	5	8	7	6	1	6	1	7	6	0	2	7	8	9	6
Wpg1	A	A	T	C	T	G	T	G	_	_	_	_	_	_	A	A	G	T	A	G	A	T	G	A	T	G	T	A	C	A	C	A	C	C	G	G	A	G	T	T	C	A	A	C	G	C	G	T	A	C	G	G	A	A
Wpg2	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg3	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg4	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg6	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg7	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg8	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg9	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg10	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg11	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg12	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg13	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg14	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	.	.
Wpg15	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Wpg16	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim35	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.
Zim18	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	A	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
Zim22	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	.	.	.
Zim10	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Zim32	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim5	.	G	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Zim7	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	G	T	A	.	.	.	G	T	A	.	.	.
Zim49	.	.	.	.	.	.	.	.	_	_	_	_	_	_	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	?	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim30	.	.	.	.	.	.	.	.	_	_	_	_	_	_	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sim3	C	A	G	A	G	A	C	A	A	C	C	G	A	G	G	G	A	C	A	G	C	C	A	G	T	A	C	T	C	A	C	T	T	C	G	C	G	A	G	C	A	A	A	C	G	C	G	C	A	C	G	G	G	A
sim6	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sim2	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sim4	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sim1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
sim5	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
R/S	S	R	R	R	R	R	R	S							R	S	R	S	S	S	R	S	R	R	S	S	S	R	S	S	R	R	S	S	S	R	S	S	R	S	S	S	R	S	R	S	S	S	R	S	S	R	S	R
F/P	F	P	F	F	F	F	F	F							P	F	F	F	P	P	F	F	F	F	P	F	F	F	P	P	P	F	F	P	P	F	P	P	F	F	F	P	P	P	P	P	P	P	P	P	P	P	F	P
			P																		P																																	

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
aly	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7	7	7	7	8	8
Position	1	1	1	2	2	3	5	7	8	8	8	9	0	0	0	0	1	2	3	3	3	4	4	6	8	9	4	4
Strains	4	5	6	3	5	8	8	9	0	1	2	1	0	3	4	7	5	9	2	3	7	2	9	0	7	9	2	7
Wpg1	A	A	T	A	G	C	C	C	C	G	G	A	C	T	T	C	C	T	T	G	A	T	G	A	T	G	A	C
Wpg2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Wpg3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Wpg4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.
Wpg7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.
Wpg8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.
Wpg10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
Wpg11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg14	.	.	.	.	C	.	.	.	A	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Wpg16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.
Zim35	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Zim18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Zim22	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	C	.	.
Zim10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
Zim32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	.	.	.
Zim5	.	C	.	.	.	.	T	.	A	.	.	T	.	.	.	G	.	A	.	.	C	C	.	.	.	C	.	.
Zim7	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	.	.	A
Zim49	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.
Zim30	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.
sim3	A	A	C	C	G	T	C	A	A	A	A	A	T	T	G	C	C	T	G	T	A	C	G	G	T	C	T	C
sim6	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.
sim2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.
sim4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
sim1	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
sim5	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R/S	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**R**	**R**	**R**	**S**	**R**	**R**	**S**	**R**	**S**	**R**	**S**	**S**	**S**	**R**	**S**
F/P	**P**	**P**	**F**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**F**	**P**	**P**	**P**	**F**	**P**	**P**	**P**	**F** **P**	**F**	**P**	**P**	**P**	**P**	**P**	**P**	**F**	**P**

Table 6: Polymorphic sites in **exon one**, **an intron** and **exon two** of the *aly* gene in *D. melanogaster* and *D. simulans*. Dots represent similarities between strains. R- replacement site, S-silent site, F-fixed site, P- polymorphic site. "–" indicates gaps. Only nucleotide positions where there are changes between strains in the exons and intron of the gene are shown.

In the *dj* gene of *D. melanogaster*, one silent polymorphism was present in the 5' untranslated region. One non-synonymous and one synonymous intraspecific polymorphisms were found in exon one. The intron has two silent polymorphisms and exon two has two non-synonymous and three synonymous intraspecific polymorphisms (Table 7).

D. simulans of *dj* has zero silent polymorphisms in the 5' untranslated region which is located sixty nucleotides upstream of the translational start site. Zero non-synonymous and zero synonymous polymorphisms were found in exon one. The intron has four silent polymorphisms within *D. simulans*. Exon two has five non-synonymous and twenty nine synonymous polymorphisms. Eighteen base pair addition is found in exon two of sim4 strain. A three base pair deletion is found in all the *D. simulans* strains (Table 7). When a cross species mega blast in the trace archives of NCBI (http://www.ncbi.nlm.nih.gov) was carried out with the *Drosophila simulans dj* gene that was sequenced in this study, it gave an average percentage (%) identity of 97.11%. This confirms that the *dj* gene in *Drosophila simulans* isolated and sequenced was right.

dj						1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
Position	3	5	5	6	7	0	4	5	7	9	9	0	0	1	1	3	4	6	8	9	9	9	2	6	6	7	9	0	1	1	1	3	4	5	6	6	9	9	9	0	1	2	3	3	5	6	6	6	7	7
Strains	8	0	5	5	4	9	8	1	4	6	8	0	2	5	9	8	9	2	8	3	6	7	0	1	9	8	3	8	0	4	9	2	1	2	2	5	3	5	8	6	0	4	6	7	8	4	8	9	0	1
Wpg1	C	A	G	C	T	T	T	C	G	C	T	G	T	A	C	G	C	G	C	C	A	A	G	C	T	G	G	T	A	C	A	G	T	T	T	A	G	A	G	A	A	T	A	G	G	G	_	_	_	_
Wpg2	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	_	_	_
Wpg3	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg4	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	_	_	_
Wpg5	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg6	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg7	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg8	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg9	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	_	_	_
Wpg10	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg11	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg12	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg13	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg14	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg15	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Wpg16	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim35	.	.	A	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim18	.	.	A	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim22	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim10	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim32	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim5	.	.	A	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim7	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim49	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
Zim30	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	_	_	_	_
sim1	T	G	C	G	C	A	T	A	A	A	A	A	G	T	C	G	A	A	C	C	C	G	A	A	T	G	A	G	A	T	A	G	C	C	A	G	A	A	G	G	A	A	G	A	A	C	_	_	_	_
sim5	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	_	_	_	_
sim7	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	_	_	_	_
sim3	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	T	G	.	.	.	.	A	.	.	.	.	.	G	.	.	.	.	.	.	.	.	_	_	_	_
sim4	.	.	.	.	.	.	.	.	.	C	G	G	T	.	.	.	.	.	.	.	.	.	.	C	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	G	.	.	T	G	C	G
R/S	S	S	S	S	R	R	S	R	S	S	S	S	S	S	S	S	S	S	S	R	R	S	R	R	S	S	S	S	R	S	R	S	S	R	S	S	R	S	S	R	R	R	R	S	S	S				
F/P	F	F	F	F	F	F	P	F	F	F	F	F	F	F	P	P	F	F	P	P	F	F	F	F	P	P	F	F	P	F	P	F	F	F	F	F	F	P	P	F	P	F	F	P	F	P				
			P					P		P	P	P	P	P													P					P				P														

dj	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	9	9
Position	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	0	0	0	1	1	2	2	3	3	4	5	7	9	1	3	4	4	4	5	7	0	1	3	7	8	8	0	0
Strains	2	3	4	5	6	7	8	9	0	1	2	3	4	5	9	0	1	4	0	3	6	2	8	1	4	6	9	2	7	5	3	1	5	5	6	7	1	2	7	6	1	0	2	5	0	5
Wpg1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	G	C	G	G	C	T	C	G	T	T	C	T	C	C	T	T	C	T	A	G	G	G	C	T	A	A	G	G	A	A	G	G
Wpg2	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Wpg3	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg4	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Wpg5	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg6	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg7	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg8	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Wpg9	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg10	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Wpg11	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg12	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg13	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg14	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
Wpg15	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.
Wpg16	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim35	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim18	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim22	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim10	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim32	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim5	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim7	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim49	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
Zim30	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
sim1	_	_	_	_	_	_	_	_	_	_	_	_	_	_	G	C	G	G	C	A	T	A	T	T	C	T	C	C	C	T	T	G	G	_	_	_	T	T	G	G	G	G	G	G	A	A
sim5	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	_	_	_	.	.	.	.	.	.	A	A	.	.
sim7	_	_	_	_	_	_	_	_	_	_	_	_	_	_	A	A	A	.	.	.	.	G	.	C	.	.	.	.	T	C	C	.	A	_	_	_	.	.	.	.	.	.	A	A	.	.
sim3	_	_	_	_	_	_	_	_	_	_	_	_	_	_	.	.	.	A	.	T	C	.	.	.	.	C	A	T	.	C	.	T	.	_	_	_	.	C	.	.	.	.	A	A	.	.
sim4	C	G	A	A	A	A	A	A	G	A	C	C	C	T	A	A	A	A	T	T	C	.	C	.	T	.	.	.	T	.	.	.	**A**	_	_	_	.	.	.	.	.	A	A	A	.	.
R/S															**R**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**				**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**
F/P															**P**	**P**	**P**	**P**	**P**	**F**	**F**	**F**	**P**	**P**	**P**	**P**	**P**	**P**	**F**	**P**	**F**	**F**	**F**				**F**	**P**	**F**	**F**	**P**	**P**	**F**	**F**	**P**	**F**
																				P	**P**	**P**							**P**		**P**	**P**	**P**										**P**	**P**		

Table 7: Polymorphic sites in 5', **exon one**, **an intron** and **exon two** of the *dj* gene in *D. melanogaster* and *D. simulans*. Dots represent similarities between strains. R- replacement site, S-silent site, F-fixed site, P- polymorphic site. "–" indicates gaps. Only nucleotide positions where there are changes between strains in the exons and intron of the gene are shown.

3.1b Transitions (Ts) / Transversions (Tv)

Transition is the substitution of a purine base by another purine base or a pyrimidine base by another pyrimidine. Transversion is the substitution of a purine base by a pyrimidine or vice-versa. Transition / Transversion ratio (R) was calculated for the *D. melanogaster* genes *bam*, *aly* and *dj* using Kimura 2- parameter pair-wise comparison (Table 8). Assuming that transitions and transversions are equally likely, R should equal 0.5.

Genes **(*D. melanogaster*)**		**Transitions / Transversions (R)**
bam	Coding	0.489
	Non-Coding	0.101
	Exon 1	0.482
	Intron 1	N/A
	Exon 2	0.486
	Intron 2	0.004
	Exon 3	N/A
aly	Coding	2.666
	Non-Coding	0.964
	Exon 1	2.337
	Intron	0.964
	Exon 2	2.501
dj	Coding	0.798
	Non-Coding	0.523
	Exon 1	0.104
	Intron	0.523
	Exon 2	0.925

Table 8: Summary of Transition (Ts) and Transversion (Tv) polymorphisms for the *D. melanogaster* genes *bam*, *aly* and *dj*. N/A values due to more changes at the same site.

Within the coding region of the three genes, *aly* shows the highest R value (2.666) and it is much greater than the non-coding region (0.964). There might be a higher occurrence of transitions or a lower occurrence of transversions in the coding region of *aly*. Synonymous and silent substitutions more often result in transitions than transversions [38]. Therefore, a high proportion of transitions might be indicative of selection against non-synonymous substitutions.

3.2 **Highly conserved gene regions**

In the carboxy terminal part of the *dj* gene (exon two) where the nuclear localization signal is present, there are no non-synonymous substitutions both within and between species. The eight times direct repeated signal DPCKKK remains conserved both within and between species with zero non-synonymous polymorphisms (Figure 5). There are ten synonymous polymorphisms in the nuclear localization signals within *D.simulans*. Also, the mitochondrial localization signal, IRPHHI in exon one of *dj* shows the absence of non-synonymous substitutions both within and between species. The absence of non-synonymous substitutions in these localization signals suggests strong selective constraints due to its functional importance. The 5' untranslated region of *dj* gene which has the translational repression element (TRE), has one polymorphism within *D. melanogaster* and no polymorphisms within *D.simulans*. This lack of polymorphism shows that selective constraints have kept the translational repression sequence signal of *dj* free of mutation accumulation.

There are two nuclear localization regions in *aly* gene. The first region is present in exon one with the amino acid pattern KKPR or KPRK. This region has zero non-synonymous and zero synonymous substitutions both within and between species. The second region of nuclear localization is present in exon two with the amino acid pattern RRGWQLVRRNMGKARRF. This region has four synonymous intraspecific substitutions and zero non-synonymous substitutions.

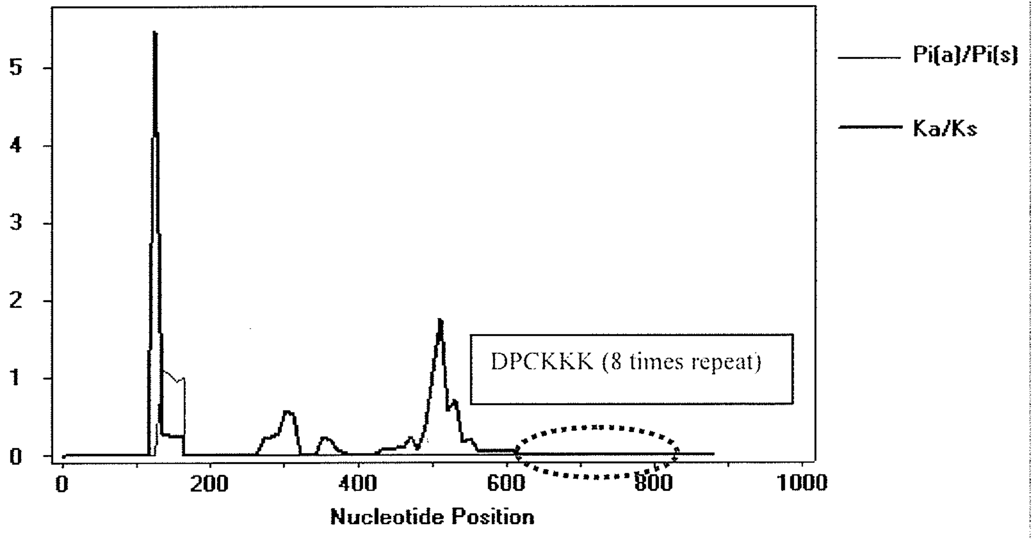


Figure 5: Graph showing the position of Nuclear Localization Signal in *dj* that remains conserved in both within and between species. Pi(a)/Pi(s) indicates non-synonymous over synonymous polymorphisms within species. K(a)/K(s) indicates non-synonymous over synonymous substitution between species.

3.3 Population subdivision

Two different populations (African Zimbabwe and Non-African Winnipeg) of *D. melanogaster* were compared for genetic differentiation. The populations were tested for significant genetic differentiation in order to know whether the analysis of the populations could be pooled together. The test is a permutation based statistical test which is applied using Ks where Ks is the weighted average of the average number of nucleotide differences between sequences from within locality Winnipeg and Zimbabwe [39]. Ks* and Kst* takes account of the number of nucleotide differences between different haplotypes (two or more liked sites) and weighting to large numbers of differences between sequences are not given much importance. Ks* and Kst* values for *bam*, *aly* and *dj* are significant between the African and non-african populations (Table 9). Therefore the two populations cannot be pooled together for within species polymorphic analysis.

Gene	Sample Size	Ks*	Kst*	P-value
bam	N1 = 8; N2 =16	1.32579	0.04800	0.0050**
aly	N1 = 9; N2 =15	2.50636	0.05199	0.00***
dj	N1 = 9; N2 =16	0.87793	0.13429	0.00***

Table 9: Genetic differentiation test for the African (Zimbabwe N1) and Non-African (Winnipeg N2) populations of the genes *bam*, *aly* and *dj*. Ks* and Kst* are nucleotide sequence-based statistics. Kst = 1-(Ks/Kt) where Kt is the average number of differences between sequences regardless of the locality [39].

3.4 Within species sequence polymorphism

Polymorphisms within species for the African, Non-African populations of *D. melanogaster* and *D. simulans* were calculated by comparing π (pi) and θ (theta) values. π is the average number of differences between all pairs of sequences in a sample of n sequences. θ is the number of variable positions in a sample of n sequences.

Genes	Non-African (Winnipeg)	Non-African (Winnipeg)	African (Zimbabwe)	African (Zimbabwe)	*D.simulans*	*D.simulans*
	π_{Total}	θ_{Total}	π_{Total}	θ_{Total}	π_{Total}	θ_{Total}
bam	0.00169	0.00208	0.00398	0.00425	0.00724	0.00802
aly	0.00595	0.00590	0.01128	0.01356	0.00697	0.00820
dj	0.00247	0.00276	0.00133	0.00168	0.01837	0.01929

Table 10: Polymorphisms in *bam*, *aly* and *dj* of Non-African, African and *D. simulans* populations.

π and θ values are calculated from all the sites in the gene. From π and θ values, it is observed that there is a higher level of polymorphism for *bam* and *aly* gene in African population compared to Non-African population. The level of polymorphism is much higher in *aly* (African versus Non-african) compared to *bam* which is again higher than the average values of *D. melanogaster* genes (π_{Total} = 0.00402, θ_{Total} = 0.00403). *D. simulans* shows higher level of polymorphism in *dj* gene when compared to other genes (Table 10).

Theta values are calculated only for synonymous sites in the coding region for the genes *bam, aly*, *dj* in order to compare to estimates from other genes. Data for other genes were obtained from Andolfatto [40] (Figure 6).

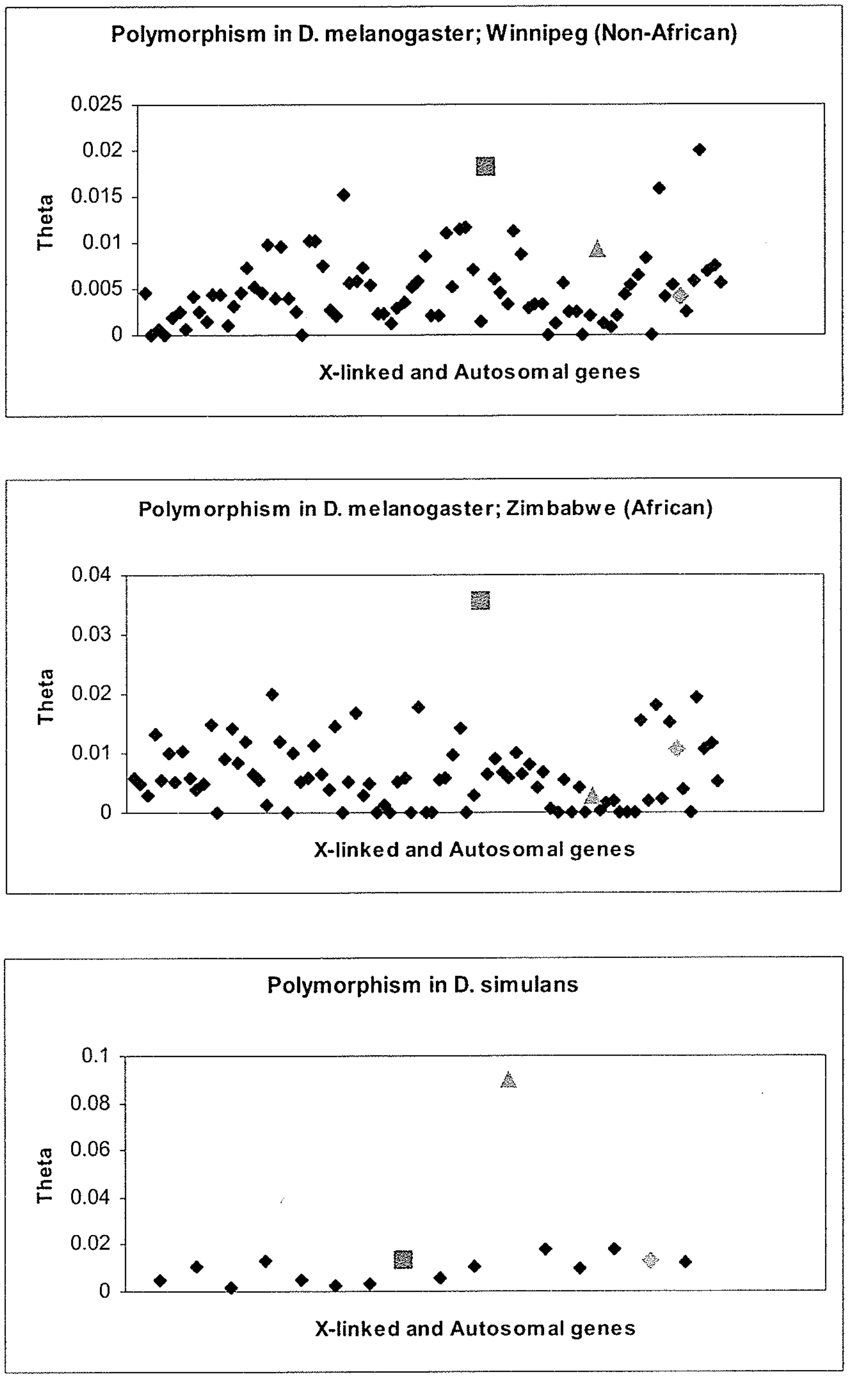


Figure 6: Synonymous polymorphism in *aly*, *dj* and *bam* compared to other X-linked and autosomal genes. Blue dots represent polymorphism represented by theta values in other genes. Pink squares represent polymorphism represented by theta values in *aly*, orange triangles show polymorphism in *dj* and green diamonds show polymorphism in *bam*.

3.5 Test of selection based on polymorphism data

Tajima's D test of neutrality and Fu and Li's test of neutrality were used on the genes *bam*, *aly* and *dj*. Tajima's D test is based on the differences between the number of segregating sites and the average number of nucleotide differences [41]. Fu and Li use the statistical properties of the numbers of external and internal mutations and their relationships to detect departures from neutrality [42]. In the geneology of a random sample of genes in a population, external mutations are ones that occurred in the external branches and internal mutations are ones that occurred in the internal branches. Internal branches are the older part of the geneology (branches connecting ancestors in a phylogenetic tree) while external branches are the younger part of the geneology (branches connecting ancestors to present taxa). External mutations get affected when there is selection while internal mutations stay neutral [42]. Tajima's D and Fu and Li are both based on the hypothesis that all substitutions at a locus are neutral. Tajima's D and Fu and Li's tests were calculated in *D. melanogaster* (Winnipeg and Zimbabwe) and *D. simulans* populations for *bam*, *aly* and *dj* (Table 11). For Tajima's D even though there are no significant values for the genes *bam*, *aly* and *dj*, the tests are mostly negative which shows an excess of rare or recent mutations. An excess of rare polymorphisms (in low frequency) might be due to purifying selection where most new mutations are eliminated to preserve the function of the protein/DNA sequence or positive selection where new favorable mutations are selected to fixation. Both processes lead to only very few recent mutations contributing to population polymorphism.

		h	Hd	Tajima's D	Fu and Li's D	Fu and Li's F
bam	Winnipeg	11	0.933	-0.70786 ns	-0.78628 ns	-0.87942 ns
	Zimbabwe	8	1.000	0.00052 ns	-0.09150 ns	-0.07779 ns
	D.simulans	5	1.000	-0.72278 ns	-0.72278 ns	-0.77662 ns
aly	Winnipeg	15	1.000	0.03535 ns	-0.53186 ns	-0.43104 ns
	Zimbabwe	9	1.000	-0.86194 ns	-1.02343 ns	-1.10558 ns
	D.simulans	4	1.000	-0.95426 ns	-0.92027 ns	-1.01423 ns
dj	Winnipeg	10	0.917	-0.12732 ns	-0.43915 ns	-0.40673 ns
	Zimbabwe	4	0.694	-0.55157 ns	-0.72564 ns	-0.75852 ns
	D.simulans	5	1.000	-0.35721 ns	-0.35721 ns	-0.38563 ns

Table 11: Tests of Neutrality for *bam, aly* and *dj*. Number of haplotypes denoted as h and estimates of haplotype diversity as Hd. ns denotes non-significant values.

3.6 Test of Selection using Polymorphism and Divergence

McDonald-Kreitman test (MK test) is used to test the relationship between levels of polymorphism and divergence at any given gene [33]. The test uses within species number of polymorphism and interspecies divergence to test for the occurrence of selection. The null hypothesis of MK test states that if both polymorphism and substitutions are neutral, the ratio of replacement to synonymous polymorphism within a species should be the same as the ratio of replacement to synonymous substitutions between two species. A neutrality index (N.I) can be computed that indicates the extent to which the levels of amino acid variation within species depart from the neutral model. N.I > 1 suggests maintenance of large proportions of amino acid polymorphism within species due to balancing selection. N.I < 1 suggests polymorphic replacements are transitory in the population and get fixed between species due to their adaptive value. MK test was carried out for the genes *bam*, *aly* and *dj* (Table 12).

		African (Zimbabwe)	**African (Zimbabwe)**	**Non-African (Winnipeg)**	**Non-African (Winnipeg)**
		Divergence	**Polymorphism**	**Divergence**	**Polymorphism**
bam	Non-synonymous substitutions	54	15	52	15
	Synonymous substitutions	28	16	28	12
		N.I = 0.486 ns		N.I = 0.673 ns	
aly	Non-synonymous substitutions	40	35	40	23
	Synonymous substitutions	28	45	29	32
		N.I = 0.544 ns		N.I = 0.521 ns	
dj	Non-synonymous substitutions	9	7	8	9
	Synonymous substitutions	9	25	10	28
		N.I = 0.280 P value 0.04255*		N.I = 0.402 ns	

Table 12: McDonald Kreitman test for the genes *bam*, *aly* and *dj*. N.I indicates Neutrality Index. ns indicates non-significance.

MK test shows significant value for the African population of *dj* with a neutrality index of 0.280 indicating that non-synonymous polymorphisms are transitory in the population and get fixed between species by positive selection. *bam* and *aly* also have a N.I values less than one but non-significant. Exons were also analyzed separately for all the three genes to test for signs of selection acting on particular regions of the gene. As a result, *bam* did not show any significant value for exons one, two or three. *Dj* did not show any significant value for exon one and two but the N.I value was always less than one. In *aly*, the MK test turned significant for exon two in both the African and Non-African populations with a N.I value less than one (Table 13). This result suggests that in exon two non-synonymous polymorphisms are transitory in the population and thus get fixed between species due to positive selection.

aly	**African (Zimbabwe)**	**African (Zimbabwe)**	**Non-African (Winnipeg)**	**Non-African (Winnipeg)**
Exon 2	**Divergence**	**Polymorphism**	**Divergence**	**Polymorphism**
Non-synonymous substitutions	24	22	24	10
Synonymous substitutions	18	40	19	27
	N.I = 0.413 Pvalue: 0.02889*		N.I = 0.293 Pvalue: 0.00868**	

Table 13: McDonald Kreitman test for exon two of *aly*. N.I indicates Neutrality Index

Polymorphism and divergence were also calculated by comparing ratios of non-synonymous substitutions (Pi(a)) versus synonymous substitutions (Pi(s)) within species (polymorphism) and non-synonymous substitutions (Ka) versus synonymous substitutions (Ks) between species (divergence). When carried out for the three genes, African and Non-African populations of *bam* and *aly* show a higher K(a)/K(s) than Pi(a)/Pi(s). In dj, African population shows a higher K(a)/K(s) than Pi(a)/P(s) suggesting positive selection whereas Non-African population shows a lower K(a)/K(s) than Pi(a)/P(s) (Table 14).

		Polymorphism			**Divergence**		
		Pi(a)	**Pi(s)**	**Pi(a)/Pi(s)**	**K(a)**	**K(s)**	**K(a)/K(s)**
bam	**African**	0.0010	0.0105	**0.103**	0.0625	0.1094	**0.552**
	Non-African	0.0009	0.0027	**0.343**	0.0607	0.1041	**0.565**
aly	**African**	0.0058	0.0324	**0.178**	0.0404	0.1147	**0.334**
	Non-African	0.0019	0.0203	**0.095**	0.0382	0.1041	**0.314**
dj	**African**	0.0000	0.0017	**0.000**	0.0194	0.1443	**0.123**
	Non-African	0.0013	0.0098	**0.134**	0.0187	0.1562	**0.108**

Table 14: Polymorphism and divergence showing ratios of Pi(a)/Pi(s) and K(a)/K(s) for *bam*, *aly* and *dj*

Exons for all the three genes were analyzed separately to find the pattern of evolution within genes. When analyzed for *dj*, Ka/Ks and Pi(a)/Pi(s) rate is higher in exon one than exon two in Non-African populations. It shows that exon two is more conserved than exon one. Ka/Ks is always higher than Pi(a)/Pi(s) in exon one and two in both African and Non-African populations suggesting positive selection (Table 15).

dj							
		Pi(s)	**Pi(a)**	**Pi(a)/Pi(s)**	**Ks**	**Ka**	**Ka/Ks**
Winnipeg	Exon1	0.0062	0.0070	**1.129**	0.0251	0.0345	**1.386**
Winnipeg	Exon 2	0.0105	0.0004	**0.046**	0.1805	0.0164	**0.080**
Zimbabwe	Exon 1	0.0000	0.0000	**0.000**	0.0000	0.0405	N/A
Zimbabwe	Exon 2	0.0020	0.0000	**0.000**	0.1711	0.0164	**0.085**

Table 15: Polymorphism and divergence showing ratios of Pi(a)/Pi(s) and K(a)/K(s) for exon one and two of *dj* gene.

aly shows higher Pi(a)/Pi(s) and Ka/Ks rates in exon one than exon two in both African and Non-African populations. It shows that exon two is more conserved than exon one. Ka/Ks is always higher than Pi(a)/Pi(s) in exon one and two in both African and Non-African populations suggesting positive selection (Table 16).

aly								
		Pi(s)	**Pi(a)**	**Pi(a)/Pi(s)**	**Ks**	**Ka**	**Ka/Ks**
Winnipeg	Exon1	0.0360	0.0052	**0.143**	0.1943	0.0977	**0.465**
Winnipeg	Exon 2	0.0010	0.0010	**0.060**	0.0920	0.0258	**0.268**
Zimbabwe	Exon 1	0.0271	0.0065	**0.236**	0.1781	0.0971	**0.512**
Zimbabwe	Exon 2	0.0316	0.0040	**0.126**	0.0940	0.0276	**0.280**

Table 16: Polymorphism and divergence showing ratios of Pi(a)/Pi(s) and K(a)/K(s) for exon one and two of *aly* gene.

3.7 Codon usage bias

The MK test assumes that synonymous substitutions are neutral. However synonymous mutations, which were assumed to be neutral are now shown to be affected by codon usage bias. Effective number of codons (ENC) is a measure of codon usage bias [43]. ENC is calculated on a scale of 20 (If each amino acid is coded by only one codon) to 61 (If there is equal and random usage of all synonymous codons). Therefore values close to 20 depict a high codon usage bias and values close to 61 depict a low codon usage bias. Genes with high codon usage bias in *D. melanogaster* generally have G and C at silent positions. Therefore there will be a high G+C content especially at the third position of the codon as changes in the third position are most often synonymous. ENC, G+C content and G+C content at the third position were calculated for *D. melanogaster* genes *bam*, *aly* and *dj* (Table 17).

On analyzing the overall coding regions, *bam* and *aly* have low codon usage bias (*bam* 52.771 & 53.087; *aly* 53.134 & 53.413) and high G+C content at the third codon position. Exon two shows higher codon bias with higher G+C content at the third codon position than other exons in both *bam* and *aly*. Also, G+C content is higher in coding regions than non-coding regions in *bam* and *aly*. *Dj* shows higher codon usage bias (*dj* 48.198 & 47.591) and lower G+C content than *bam* and *aly*. G+C content in *dj* is low in coding regions than non-coding regions (Table 17).

Genes		**ENC**	**G+C**	**G+C 3rd Position**
bam **(Winnipeg)**	Exon 1	61	0.505	0.579
	Intron 1	N/A	0.311	N/A
	Exon 2	44.45	0.516	0.721
	Intron 2	N/A	0.357	N/A
	Exon 3	51.726	0.575	0.645
	Total (coding)	**52.771**	**0.533**	**0.643**
bam **(Zimbabwe)**	Exon 1	61	0.505	0.579
	Intron 1	N/A	0.309	N/A
	Exon 2	46.277	0.513	0.718
	Intron 2	N/A	0.357	N/A
	Exon 3	51.718	0.576	0.647
	Total (coding)	**53.087**	**0.532**	**0.643**
aly **(Winnipeg)**	5’	N/A	0.334	N/A
	Exon 1	59.412	0.453	0.525
	Intron	N/A	0.385	N/A
	Exon 2	50.822	0.529	0.672
	Total (coding)	**53.134**	**0.517**	**0.645**
aly **(Zimbabwe)**	5’	N/A	0.342	N/A
	Exon 1	60.771	0.457	0.537
	Intron	N/A	0.400	N/A
	Exon 2	51.037	0.528	0.668
	Total (coding)	**53.413**	**0.517**	**0.645**
dj **(Winnipeg)**	5’	N/A	0.272	N/A
	Exon 1	N/A	0.416	0.469
	Intron	N/A	0.459	N/A
	Exon 2	49.540	0.391	0.477
	Total (coding)	**48.198**	**0.401**	**0.474**
dj **(Zimbabwe)**	5’	N/A	0.273	N/A
	Exon 1	N/A	0.415	0.467
	Intron	N/A	0.454	N/A
	Exon 2	49.439	0.392	0.481
	Total (coding)	**47.591**	**0.395**	**0.477**

Table 17: Codon usage bias, G+C content, G+C content at third position of the *D. melanogaster* genes *bam*, *aly* and *dj* calculated by DNAsp software.

3.8 Spermatogenesis and Developmental genes

Under the hypothesis of neutral evolution, the ratio of replacement to synonymous fixed differences between species [R/S(F)]should be the same as the ratio of replacement to synonymous polymorphisms within species [R/S(P)]. On comparing spermatogenesis, sex-related and developmental genes, it can be noted that the ratio of replacement to synonymous fixed differences between species [R/S(F)] is lower than replacement to synonymous polymorphisms within species [R/S(P)] in developmental genes (Figure 7). Developmental genes, which are genes involved in body pattern formation, are thought to be highly conserved across species. Sex-related genes, which include sex determination genes, mating behavior, fertilization and spermatogenesis genes are rapidly evolving between closely related species. Sex-related and spermatogenesis genes have a higher ratio of [R/S(F)] between species versus [R/S(P)] within species which suggests that polymorphisms get fixed between species due to selection.

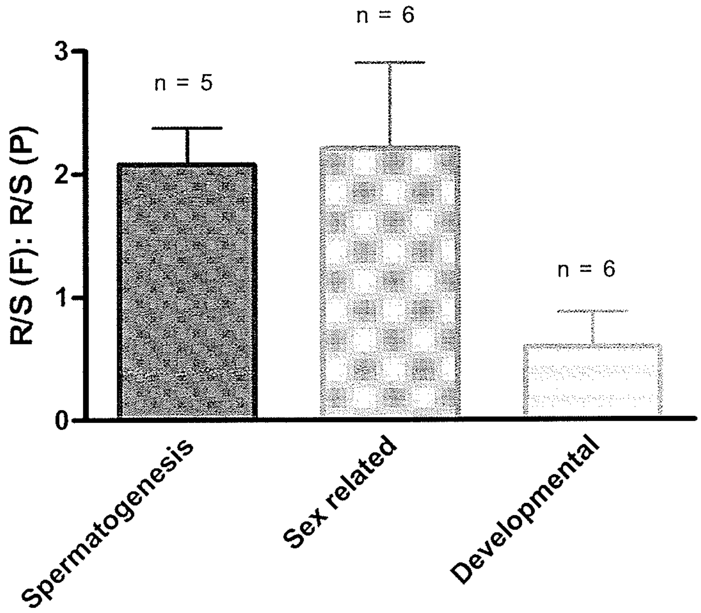


Figure 7: Genes of spermatogenesis, sex related and development showing ratios of replacement to synonymous substitutions between species R/S (F) versus within species R/S (P). R denotes replacement ; S denotes synonymous. F denotes Fixed ; P denotes polymorphism. Data for spermatogenesis genes, sex related genes and developmental genes are taken from different citations [44, 45, 46, 47, 48, 49]. Spermatogenesis genes also include genes sequenced in this study (*bam*, *aly* and *dj*).

4. DISCUSSION

D. melanogaster and *D. simulans* are native of Africa and spread worldwide as human commensals 10,000 years ago [50]. Genetic variation is found to be lower in non-african populations than African populations in both species [51]. There are different models proposed to explain the difference in genetic variation between African and Non-African populations. One possibility is that population bottlenecks that is reductions in population size in Non-African population during colonization would have occurred leading to decreased genetic variation. The other possibility would be "local adaptation hypothesis", where a change in a habitat outside Africa would have led to different adaptation for populations [52]. While population bottleneck is expected to affect the entire genome, local adaptation might differentially affect genes leading to a non-uniform pattern of polymorphism. Studies on nucleotide variation between African and Non-African populations show a higher rate of π and θ for many genes in African populations than Non-African populations [53, 54]. Therefore *D. melanogaster* is thought to have experienced founder effect during its dispersal from Africa. Therefore African populations are more variable than Non-African populations. *D. simulans* on the other hand has the same history of *D. melanogaster* but it has experienced less severe founder effects and so it is more variable than *D. melanogaster.*

The levels of polymorphism within *D. melanogaster* were calculated for the genes *bam*, *aly* and *dj* in African and Non-African populations by comparing π (pi) and θ (theta) values from all the sites in the gene. As noted from table 10, π_{Total} =

0.01128 and θ_{Total} = 0.01356 for *aly* gene in African populations are higher than non-african populations (*aly* π_{Total} = 0.00595 and θ_{Total} = 0.00590) within *D. melanogaster*. The total polymorphism in *aly* African population is much higher than the average values observed for *D. melanogaster* genes (African and Non-African).
D. melanogaster genes have an average value of π_{total} = 0.00402 & θ_{total} = 0.00403 [38].

The level of polymorphism for *bam D. melanogaster* population is also higher in the African population than the Non-African population but is less than the average values observed for *D. melanogaster* genes (African and Non-African).

Both *bam* and *aly* show the increased nucleotide variation expected in African population compared to Non-African population. However, *dj* does not show any increased nucleotide difference between African and Non-African populations within *D. melanogaster*. Moreover, when π and θ values were calculated from all the sites in the gene, it shows values that are less than the average values of *D. melanogaster* genes (African and Non-African) (Table 10).

In order to further compare the pattern of polymorphism, theta values were calculated only for synonymous sites of coding regions for the genes *bam*, *aly* and *dj*. The values were compared to data from many other genes [40]. From the graph (Figure 6), it is observed that *aly* has a higher rate of synonymous polymorphism in African population than Non-African population. Gene *aly* in both African and Non-African population has values higher than other *Drosophila* genes (Figure 6). Gene *bam* also shows an increase in the rate of synonymous polymorphism in African population than Non-African population but not as high as *aly*. On the contrary, *dj* shows a lower rate of synonymous polymorphisms in African population than Non-African population of *D.melanogaster*.

On the whole, *bam* and *aly* show higher total and synonymous polymorphism in African than Non-African populations. This result is expected if nucleotide variation is depleted due to population bottleneck during the spread of *D. melanogaster* from Africa.

The explanation for the decreased variation in African population of *dj* gene could be due to the chromosomal location [40]. The chromosomal location of *dj* is 3R (third chromosome right arm). Chromosome inversions suppress crossing over when heterozygous and thus reduce polymorphism levels in a population [55]. African populations are more often found with autosomal inversions compared to Non-African populations. Inversion frequencies could have changed recently in African population where the nucleotide changes of chromosomes that are inverted will be less than the standard chromosomes [55].

Within species polymorphism was further studied by calculating the Transition / Transversion (R) ratio in the three *D. melanogaster* genes *bam*, *aly* and *dj*. Under the assumption of neutral evolution, transitions and transversions occur by random mutation and the R value will be approximately 0.5 since the occurrence of transitions is expected to be half that of transversions. The R value is always higher in the coding than the non-coding regions of all the three genes (*bam* coding 0.489; non-coding 0.101, *aly* coding 2.666 ; non-coding 0.964 and *dj* coding 0.798 ; non-coding 0.523). Transversions in the coding region are expected to result in non-synonymous substitutions more often than transitions [38]. Therefore, the reason for the higher occurence of transitions than transversions in the coding regions could be due to some level of selection against non-synonymous substitutions. Non-synonymous substitutions will change the amino acid composition of a protein becoming deleterious whereas synonymous substitutions do not affect the amino acid composition.

When the R values of the coding regions of the three genes (Non-African) are compared, only *aly* (2.666) shows a higher R ratio in the coding region than *bam* and *dj* (0.489 and 0.798 respectively) whose values are closer to expectations under neutrality.

In order to unfold the reason behind the high R value in *aly*, a more detailed analysis of R within the coding region was carried out. *aly* has a similar R value in exon two (*aly* 2.501) and exon one (*aly* 2.337). Therefore *aly* seems to have an overall mutation bias to either more transitions or less transversions. *dj* shows a much higher R value in exon two (*dj* 0.925) than exon one (0.104). The higher R value in exon two than exon one is interesting, if such difference is the result of increase in transitions. Exon two is expected to have less non-synonymous substitutions than exon one.

Accordingly Pi(a)/Pi(s) rate is higher in exon one than exon two which shows that exon two has lower non-synonymous polymorphisms than exon one within species. Therefore the high R value is more likely due to increase in transitions reflecting selection against non-synonymous substitutions in exon two.

Interestingly, nuclear localization signals are present in exon two of *dj*. There is a recognizable nuclear localization signal in exon two of *dj* (sites 412 to 555). It consists of an eight times direct repeat of hexapeptide sequence (DPCKKK). This region remains highly conserved both within *D. melanogaster* and between species showing absence of non-synonymous substitutions. This region is very much essential for the localization of the dj protein into the nucleus and is therefore important for the function of the protein. The conservation of exon two due to functional reasons such as the presence of the nuclear localization signal is further supported by the fact that when exons were analyzed separately in *dj* gene of Non-African population, both exon one and two showed higher Ka/Ks ratio than Pi(a)/Pi(s) ratio suggesting positive selection (Table 15).

The general assumption that synonymous changes are free from selection does not always hold. Codon bias affects synonymous substitutions and such bias might be due to selection which improves the efficiency of protein synthesis by enhancing the translational process [56, 57]. The opposite of codon bias due to selection is mutational bias which can also be responsible for the unequal usage of synonymous codons. In

Drosophila, mutational bias is towards A+T substitutions [58]. In *Drosophila*, genes with high codon bias have an increased G+C content at silent positions and most particularly at the third position [58]. A high G+C content shows the action of selection pressure to overcome mutation bias towards A+T substitutions. G+C content is lower in non-coding than coding regions in *bam* and *aly* which further shows the mutational bias towards A+T in the non-coding regions.

The gene *bam* in *D. melanogaster* shows a low codon bias in the overall coding region and when exons are analyzed seperately (Table 17).

The gene *aly* in *D. melanogaster* also shows a low codon usage bias (ENC Winnipeg 53.134, Zimbabwe 53.413) (Table 17) in the overall coding region. Exon two shows a slightly higher codon usage bias and G+C content at third codon position.

The gene *dj* in *D. melanogaster* shows the highest codon usage bias (ENC Winnipeg 48.198; Zimbabwe 47.591) of all three genes in the overall coding region. This higher codon bias is associated with high G+C content at the third codon position. It is important to note that when *dj* is compared to the average ENC of *D. melanogaster* genes, *dj* codon usage bias is within average values. Different reasons might contribute to the higher codon usage bias in *dj* than *bam* and *aly*: short genes in terms of nucleotide sequence length are a smaller sample than long genes and this small sample size can introduce bias in estimates [58]. The length of *dj* (833 base pairs) is relatively smaller than *aly* (1680 base pairs) and *bam* (1470 base pairs). The effect of mutations to nonoptimal codons is relatively higher in smaller genes than longer genes as a nonoptimal codon requires twice as long to incorporate an amino acid as does an optimal codon [58]. High codon bias in *dj* could also be due to translational efficiency, that is the preferential usage of certain codons that can base pair with the most abundant tRNA's in the cell. The high codon bias on exon two of *aly* and *dj* (Table 17) might be due to the functional importance of exon two as nuclear localization signals are present in these exons. This is further supported by the low rates of non-synonymous polymorphisms in exon two than exon one within species (Table 16 and 15).

Statistical test of selection based on within species polymorphism data did not show any significant value for all the three genes in any of the populations (Table 11). This can be due to small sample sizes [38]. For example, Tajima's D test using 11 and 6 alleles per locus at the *Adh* gene of *D. melanogaster* and *D. simulans* respectively, gave non-significant results while a sample of 99 alleles from the *Adh* gene in *D. pseudoobscura* gave significant Tajima's D [38]. Selection might be weak for Tajima's and Fu and Li's test to detect. The values of Tajima's D and Fu and Li's test are mostly negative for all the three genes *bam*, *aly* and *dj* which indicates an excess of rare or recent mutations which could be due to some form of purifying selection where deleterious mutations segregate at low frequency or positive selection where advantageous alleles get fixed recently.

Because MK test combines polymorphism and divergence information, it is possible to test whether polymorphisms have been selected due to their adaptive value and fixed between species. The null hypothesis of MK test states that if both synonymous polymorphism and substitutions are neutral, the ratio of replacement to synonymous polymorphism within a species should be the same as the ratio of replacement to synonymous substitutions between two species. The MK test was non-significant for *bam* and *aly* while results for d*j* showed a significant deviation from neutrality in African populations (Table 12). The deviation is in the direction of a higher rate of non-synonymous to synonymous substitutions between species versus within species suggesting that polymorphic replacements are transitory in the population and get fixed between species due to their adaptive value.

Interestingly, exon two of *aly* showed significant *P* values for both African (0.02889) and Non-African (0.00868) populations (Table 13) suggesting that there is an increased rate of replacement substitutions only between species but a low rate of intraspecific replacement polymorphisms. Given the identification of bias in transition/transversion ratio of substitutions within species for *aly*, it is likely that the proportion of synonymous changes within this gene do not reflect what would be expected under random mutation.

Addressing the aims specifically, there appear to be differences in the pattern of evolution of genes at early versus late stages of sperm development. *bam* which is a early spermatogenesis gene seems to be more conserved with lower within species polymorphisms and interspecies divergence. *dj* which is a spermiogenesis gene shows higher rate of divergence and conservation of functionally important sites in the gene. Selection can modify the amount of genetic variation found in a population by eliminating deleterious mutations (purifying selection) or fixing advantageous mutations (positive selection). All three genes show some form of purifying selection acting upon but *dj* in particular shows signs of positive selection in African population as *dj* being a late spermiogenesis gene plays an important role in sperm maturation. Although the sample size is small, the rate of fixed replacements to polymorphisms is higher for spermatogenesis genes analyzed in this study than previously reported rates on other developmental genes.

5. CONCLUSIONS

- *bam* and *aly* show higher rates of polymorphism in African than non-african population when total polymorphism from all the sites in a gene and polymorphism from theta values of synonymous sites were observed. This pattern is consistent with bottleneck effects or local adaptation.

- The level of polymorphism in *aly* is much higher than the average values observed for *D. melanogaster* genes. In *bam* the polymorphism is lower than the average values of *D. melanogaster* genes.

- *Dj* gene shows a lower rate of polymorphism in African than non-african populations which might be due to autosomal inversion polymorphisms.

- *Dj* and *aly* have a higher R value in exon two than exon one suggesting more transitions reflecting selection against non-synonymous changes is much stronger in exon two due to the presence of nuclear localization signals.

- C terminal end of *dj* gene with the nuclear localization signals remains selectively constrained with the absence of non-synonymous substitutions within and between species. Mitochondrial localization signal in exon one has no non-synonymous substitutions within and between species.

Nuclear localization signals in *aly* has no non-synonymous substitutions and it remains conserved both within and between species.

- There is high codon bias in exon two of *aly* and *dj* which might be due to the functional importance as nuclear localization signals are present.

- Tajima's test and Fu and Li's test of neutrality though not significant are mostly negative for *bam*, *aly* and *dj* indicating some form of purifying selection or positive selection. McDonald Kreitman test for *dj* gene in African population is significant. McDonald Kreitman test for exon two of *aly* gene in both African and Non-African population is significant. It indicates adaptive diversification between species due to positive selection.

- Pi(a)/Pi(s) and Ka/Ks in *aly* and *dj* genes are higher in exon one than exon two showing conservation of exon two due to functional conservation.

REFERENCES

1. Fuller MT. Spermatogenesis. *The Development of Drosophila*: Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1993;71-147.

2. Fuller MT. Genetic control of cell proliferation and differentiation in Drosophila spermatogenesis. *Semin Cell Dev Biol* 1998;**9**(4):433-44.

3. Bock I. Interspecific hybridization in the genus *Drosophila. Evol Biol* 1984;**18**:41-70.

4. Haldane JBS. Sex ratio and unisexual sterility in hybrid animals. *J.Genet* 1922;**12**:101-109.

5. Charlesworth B, J.A. Coyne and N.Barton. The relative rates of evolution of sex chromosomes and autosomes. *Am.Nat* 1987:113-146.

6. Wu CI, N.A. Johnson and M.F.Palopoli. Haldane's rule and its legacy: Why are there so many sterile males? *TREE* 1996;**11**:281-284.

7. Betancourt AJ, Presgraves DC, Swanson WJ. A test for faster X evolution in Drosophila. *Mol Biol Evol* 2002;**19**(10):1816-9.

8. Torgerson DG, Singh RS. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. *Mol Biol Evol* 2003;**20**(10):1705-9.

9. Civetta A, Singh RS. Sex-related genes, directional sexual selection, and speciation. *Mol Biol Evol* 1998;**15**(7):901-9.

10. Civetta A, Singh RS. Broad-sense sexual selection, sex gene pool evolution, and speciation. *Genome* 1999;**42**(6):1033-41.

11. Civetta A. Positive selection within sperm-egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. *Mol Biol Evol* 2003;**20**(1):21-9.

12. Glassey B, Civetta A. Positive selection at reproductive ADAM genes with potential intercellular binding activity. *Mol Biol Evol* 2004;**21**(5):851-9.

13. Kulathinal R, Singh RS. Cytological characterization of premeiotic versus postmeiotic defects producing hybrid male sterility among sibling species of the *Drosophila melanogaster* complex. *Evolution* 1998;**52**(4):1067-1079.

14. Michalak P, Noor MA. Genome-wide patterns of expression in Drosophila pure species and hybrid males. *Mol Biol Evol* 2003;**20**(7):1070-6.

15. McKearin DM, Spradling AC. bag-of-marbles: a Drosophila gene required to initiate both male and female gametogenesis. *Genes Dev* 1990;**4**(12B):2242-51.

16. Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 1986;**234**(4774):364-8.

17. McKearin D, Ohlstein B. A role for the Drosophila bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* 1995;**121**(9):2937-47.

18. Jiang J, White-Cooper H. Transcriptional activation in Drosophila spermatogenesis involves the mutually dependent function of aly and a novel meiotic arrest gene cookie monster. *Development* 2003;**130**(3):563-73.

19. White-Cooper H, Schafer MA, Alphey LS, *et al.* Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in Drosophila. *Development* 1998;**125**(1):125-34.

20. Lin TY, Viswanathan S, Wood C, *et al.* Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in Drosophila males. *Development* 1996;**122**(4):1331-41.

21. White-Cooper H, Leroy D, MacQueen A, *et al.* Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* 2000;**127**(24):5463-73.

22. Santel A, Winhauer T, Blumer N, *et al.* The Drosophila don juan (dj) gene encodes a novel sperm specific protein component characterized by an unusual domain of a repetitive amino acid motif. *Mech Dev* 1997;**64**(1-2):19-30.

23. Hess H, Heid H, Franke WW. Molecular characterization of mammalian cylicin, a basic protein of the sperm head cytoskeleton. *J Cell Biol* 1993;**122**(5):1043-52.

24. Hess H, Heid H, Zimbelmann R, *et al.* The protein complexity of the cytoskeleton of bovine and human sperm heads: the identification and characterization of cylicin II. *Exp Cell Res* 1995;**218**(1):174-82.

25. von Bulow M, Heid H, Hess H, *et al.* Molecular nature of calicin, a major basic protein of the mammalian sperm head cytoskeleton. *Exp Cell Res* 1995;**219**(2):407-13.

26. Santel A, Blumer N, Kampfer M, *et al.* Flagellar mitochondrial association of the male-specific Don Juan protein in Drosophila spermatozoa. *J Cell Sci* 1998;**111 (Pt 22)**:3299-309.

27. Kuhn R, Schafer U, Schafer M. Cis-acting regions sufficient for spermatocyte-specific transcriptional and spermatid-specific translational control of the Drosophila melanogaster gene mst(3)gl-9. *Embo J* 1988;**7**(2):447-54.

28. Kuhn R, Kuhn C, Borsch D*, et al.* A cluster of four genes selectively expressed in the male germ line of Drosophila melanogaster. *Mech Dev* 1991;**35**(2):143-51.

29. Yang J, Porter L, Rawls J. Expression of the dihydroorotate dehydrogenase gene, dhod, during spermatogenesis in Drosophila melanogaster. *Mol Gen Genet* 1995;**246**(3):334-41.

30. Yanicostas C, Lepesant JA. Transcriptional and translational cis-regulatory sequences of the spermatocyte-specific Drosophila janusB gene are located in the 3' exonic region of the overlapping janusA gene. *Mol Gen Genet* 1990;**224**(3):450-8.

31. Blumer N, Schreiter K, Hempel L*, et al.* A new translational repression element and unusual transcriptional control regulate expression of don juan during Drosophila spermatogenesis. *Mech Dev* 2002;**110**(1-2):97-112.

32. Wilkins AS. Conserved Genes and Functions in Animal Development. *The Evolution of Developmental Pathways*: Sinauer Associates, Inc.,, 2002;127-168.

33. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 1991;**351**(6328):652-4.

34. Danielson PB. Capillary Electrophoresis Sequencing: Maximum Read Length at Minimal cost. *Biotechniques* 2002;**32**(1):24-28.

35. Thompson JD, Gibson TJ, Plewniak F, *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;**25**(24):4876-82.

36. Rozas J, Sanchez-DelBarrio JC, Messeguer X, *et al.* DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003;**19**(18):2496-7.

37. Kumar S, Tamura K, Jakobsen IB, *et al.* MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001;**17**(12):1244-5.

38. Moriyama EN, Powell JR. Intraspecific nuclear DNA variation in Drosophila. *Mol Biol Evol* 1996;**13**(1):261-77.

39. Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. *Mol Biol Evol* 1992;**9**(1):138-51.

40. Andolfatto P. Contrasting patterns of X-linked and autosomal nucleotide variation in Drosophila melanogaster and Drosophila simulans. *Mol Biol Evol* 2001;**18**(3):279-90.

41. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;**123**(3):585-95.

42. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993;**133**(3):693-709.

43. Wright F. The 'effective number of codons' used in a gene. *Gene* 1990;**87**(1):23-9.

44. Walthour CS, Schaeffer SW. Molecular population genetics of sex determination genes: the transformer gene of Drosophila melanogaster. *Genetics* 1994;**136**(4):1367-72.

45. Tsaur SC, Ting CT, Wu CI. Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of Drosophila: II. Divergence versus polymorphism. *Mol Biol Evol* 1998;**15**(8):1040-6.

46. Parsch J, Meiklejohn CD, Hauschteck-Jungen E, *et al.* Molecular evolution of the ocnus and janus genes in the Drosophila melanogaster species subgroup. *Mol Biol Evol* 2001;**18**(5):801-11.

47. Balakirev ES, Balakirev EI, Ayala FJ. Molecular evolution of the Est-6 gene in Drosophila melanogaster: contrasting patterns of DNA variability in adjacent functional regions. *Gene* 2002;**288**(1-2):167-77.

48. Baines JF, Chen Y, Das A, *et al.* DNA sequence variation at a duplicated gene: excess of replacement polymorphism and extensive haplotype structure in the Drosophila melanogaster bicoid region. *Mol Biol Evol* 2002;**19**(7):989-98.

49. Ayala FJ, Hartl DL. Molecular drift of the bride of sevenless (boss) gene in Drosophila. *Mol Biol Evol* 1993;**10**(5):1030-40.

50. Lachaise D, Cariou M, David J, *et al.* Historical biogeography of the *Drosophila melanogaster* species subgroup. In: MK H, B W, GT P, eds. *Evolutionary Biology*. New York: Plenum Press, 1988;159-225.

51. Mousset S, Derome N. Molecular polymorphism in Drosophila melanogaster and D. simulans: what have we learned from recent studies? *Genetica* 2004;**120**(1-3):79-86.

52. David JR, Capy P. Genetic variation of Drosophila melanogaster natural populations. *Trends Genet* 1988;**4**(4):106-11.

53. Begun DJ, Aquadro CF. African and North American populations of Drosophila melanogaster are very different at the DNA level. *Nature* 1993;**365**(6446):548-50.

54. Begun DJ, Aquadro CF. Molecular variation at the vermilion locus in geographically diverse populations of Drosophila melanogaster and D. simulans. *Genetics* 1995;**140**(3):1019-32.

55. Navarro A, Barbadilla A, Ruiz A. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. *Genetics* 2000;**155**(2):685-98.

56. Akashi H. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics* 1994;**136**(3):927-35.

57. Akashi H, Eyre-Walker A. Translational selection and molecular evolution. *Curr Opin Genet Dev* 1998;**8**(6):688-93.

58. Powell JR, Moriyama EN. Evolution of codon usage bias in Drosophila. *Proc Natl Acad Sci U S A* 1997;**94**(15):7784-90.