Indian Journal of Entomology 85(1): 40-45 (2023) DoI. No.: 10.55446/IJE.2022.979

# MIRROR REPEATS IN THE *INTERSEX* GENE OF *DROSOPHILA MELANOGASTER* MEIGEN

**Kavita Saini, Namrata Dangi and Vikash Bhardwaj***

Department of Zoology, Baba Mastnath University, Asthal Bohar Rohtak 124021, Haryana, India
*Email: vikashbhardwaj@gmail.com (corresponding author): ORCID ID 0000-0001-6569-665X

**ABSTRACT**

**Genomes of many organisms, including both prokaryotes and eukaryotes, contain numerous repeat elements. Among these DNA repeats, mirror repeats have been studied rarely. Hence, this study on the mirror repeats within the *intersex* gene of *Drosophila melanogaster*, which reveal 11 mirror repeats. Out of these mirror repeats, ten mirror repeats are perfect, and six are within a single exon. The presence of mirror repeats within the intersex gene raises further questions concerning their origin, evolution, and biological function.**

**Key words:** *Drosophila melanogaster, intersex* gene, mirror repeats, DNA repeats, parallel complement, *Drosophila* spp., origin, evolution, function

*Drosophila melanogaster*, also known as the fruit fly, is a model organism with widespread use in classical and modern genetics and biomedical research (Wangler et al., 2015; Mirzoyan et al., 2019; Pandey and Nichols, 2011; Tolwinski, 2017). Its quick generation time and cheap maintenance cost make it appropriate for studying extensive pathways (Mirzoyan et al., 2019). Its genome has been annotated with 17,726 genes, 13,907 of which are protein-coding genes that encode 21,953 distinct polypeptides (Kaufman 2017). Over 75% of disease-related genes have the same function in humans and flies, enabling researchers to conduct studies on flies and adapt their results to human biology (Reiter et al., 2001; Aquilina and Cauchi, 2018; Evangelakou et al., 2019). The nucleotide base sequences in the genome of any organism are methodically organized, information-rich patterns composed mainly of repeated DNA sequences; and wide variance in genome size is the consequence of repeated DNA sequences and these sequences aid in explaining genome function and structure (Hartl et al., 1994; Kapitonov and Jurka, 2003; Gurusaran et al., 2013). Repeated DNA sequences are genomic segments that repeat themselves throughout the genome and can range from a few to several thousand nucleotide sequences that are incorporated into the genomes of higher animals (Britten and Kohne, 1968; Jurka et al., 2007; Biscotti et al., 2015).

Since publishing whole-genome data from many organisms, the measurement and categorization of repeat elements have been an area of interest in computational biology research (Volfovsky et al., 2001; Treangen and Salzberg, 2012). There are several forms of repetitive elements present in the DNA of different organisms, including inverted repeats, direct repeats, tandem repeats, interspersed DNA repeats (LINES and SINES), and mirror repeats (Jurka et al., 2007; Treangen and Salzberg, 2012). To the best of our knowledge, mirror repeats have never been investigated within the genome of *D. melanogaster*. An essential objective in the study of eukaryotic developmental genetics is to understand the processes that determine sexuality. The somatic sexual differences in *D. melanogaster* are controlled by a hierarchical network of regulatory genes (Burtis and Baker, 1989; Baker, 1989; Christiansen et al., 2002; Peng et al. 2021). For the female sex differentiation, the *intersex* gene operates in conjunction with the *doublesex* gene, which is located at the very bottom of the hierarchy of sex determination(Acharyya and Chatterjee 2002; Garrett-Engele et al. 2002a; Garrett-Engele et al. 2002b; Cavaliere et al. 2009; Arunkumar and Nagaraju 2011). Consequently, the current study uses a manual computational approach to locate mirror repeats within the *intersex* gene that play a role in sex determination in *D. melanogaster*.

## MATERIALS AND METHODS

The whole gene sequence for the *intersex* gene (Gene ID 45881) was retrieved from the NCBI and saved in the FASTA format. Further, the sequence was split, generating two distinct sequences consisting of 500 and 227 base pairs, respectively (the total number of base pairs in the *intersex* gene is 727). A parallel complement sequence of each component

was retrieved using the Reverse Complement tool, which can be found at https://www.bioinformatics.org/sms/rev comp.html. The original sequence and its parallel complement sequence were aligned for homology searches using the BLAST software, which may be found at https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE TYPE=BlastSearch&LINK LOC=blasthome. The program selection was changed to a nucleotide sequence pairwise, and the word size was set to 7. Both sequences were initially aligned at different threshold values, as shown in Table 1. The instances where the number of hits was maximum were selected, and then searched for mirror repeats in each hit. The hits were examined, and where the nucleotide position number was precisely reversed in both the subject and the query sequences were marked as mirror repeats. As shown in Table 2, the detected mirror repeats were divided into two categories: perfect and imperfect. Identified mirror repeats were also explored in the genomes of two other species, *D. nasuta and D. bipectinata*. Fig. 1 explains the complete methodology that is used in this study.

## RESULTS AND DISCUSSION

The identification of mirror repeats in the *intersex* gene of *D. melanogaster* was performed using a manual bioinformatics computational approach. Only one exon is included in the *intersex* gene, with a total sequence length of 727 base pairs. It was observed that the number of hits also increases with increasing E value, and once the E value reaches a certain threshold, the number of hits saturates and stops increasing further. Mirror repeats were also searched in all hits, as indicated in Table 1, and when the number of hits stops going further after a specific E value, it was fixed as threshold E value. After going through this procedure, 11 mirror repeats were retrieved in the complete gene sequence. The gene was split into two parts and found the mirror repeats on the individual segment. The same number of mirror repeats were found in both cases, and there are 11 mirror repeats in the entire gene, and there are 11 mirror repeats cumulatively in both parts. This led to the conclusion that the number of mirror repeats extracted is the same in both cases. Considering the findings presented in Table 1, 20 such repeats were selected, the

Table 1. No. of mirror repeats in exon, gene parts and complete gene at different E values

| Expected threshold (E value) | No. of hits for Exon | Mirror repeats in Exon | No. of hits for Part-1 1-500 | Mirror repeats in Part-1 1-500 | No.of hits for Part-2 501-727 | Mirror repeats in Part-2 501-727 | No. of hits for complete sequence | Complete gene |
|---|---|---|---|---|---|---|---|---|
| 0.05 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 10 | 16 | 3 | 13 | 3 | 8 | 6 | 27 | 8 |
| 15 | 16 | 3 | 13 | 3 | 8 | 6 | 27 | 8 |
| 20 | 29 | 6 | 27 | 5 | 8 | 6 | 27 | 8 |
| 40 | 29 | 6 | 27 | 5 | 8 | 6 | 47 | 11 |
| 100 | 29 | 6 | 27 | 5 | 8 | 6 | 47 | 11 |
| Total mirror repeats | | 6 | | 5 | | 6 | | 11 |

Table 2. List of mirror repeats in *intersex* gene of *D. melanogaster*

| S. No. | Mirror repeat gene ID- 45881 | Position of mirror repeats | Type | Length |
|---|---|---|---|---|
| 1. | AAAAGAAAA | 13-21 | Perfect with single spacer | 9 |
| 2. | GGACCCCCAGG | 151-161 | Perfect with single spacer | 11 |
| 3. | CCAGGACC | 157-164 | Perfect mirror | 8 |
| 4. | TCGCGCT | 305-311 | Perfect with single spacer | 7 |
| 5. | ATCACTA | 470-476 | Perfect with single spacer | 7 |
| 6. | ACATACA | 575-581 | Perfect with single spacer | 7 |
| 7. | TGATTAGT | 608-615 | Perfect | 8 |
| 8. | TTGATAGTT | 642-650 | Perfect with single spacer | 9 |
| 9. | TTTATTATTT | 666-675 | Perfect mirror | 10 |
| 10. | TAAATAAAT | 699-707 | Perfect with single spacer | 9 |
| 11. | AATAAATATTCTTAGATATAA | 701-721 | Imperfect mirror | 21 |

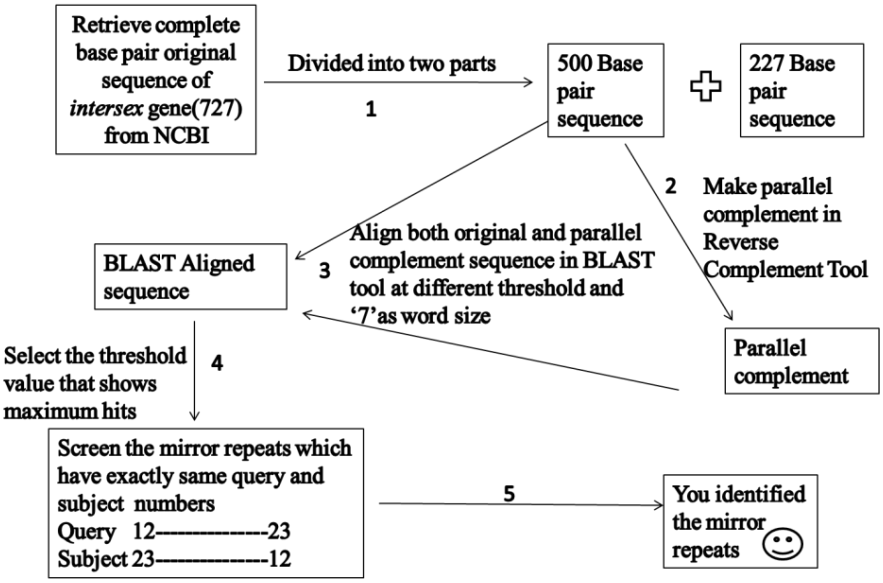


Fig. 1. Steps in extracting of mirror repeats

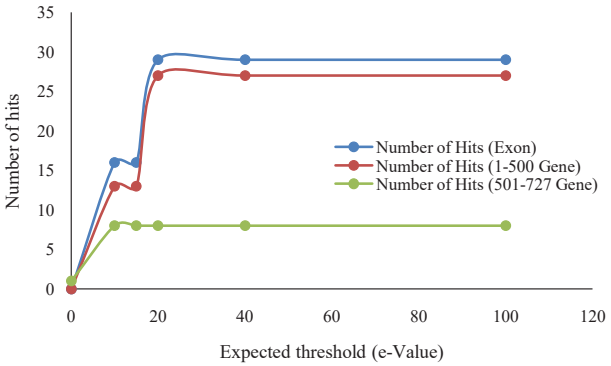


Fig. 2. Number of hits vs. expected threshold (E-value) for *intersex* gene of *D. melanogaster*

expected threshold for exons and two other sections of the gene because after this point, the number of hits and mirror repeats stops increasing further. Alternatively, 40 were selected as the expected threshold for the entire gene sequence. Here, the number of hits are directly correlated with the expected threshold shown in Fig. 2. The observation specifies that the number of hits increases as the E value increases, and once the E value reaches a definite threshold, the number of hits stops and shows a saturation`.

Table 2 demonstrates mirror repeats prevalent in the *intersex* gene of *D. melanogaster*. Subsequently, we classified mirror repeats as perfect or imperfect. Perfect mirror repeats have a 100 per cent symmetrical (CCAGGACC) arrangement of nucleotides around their centre of symmetry. Imperfect mirror repeats are not ideally 100% symmetrical (AATAAATATTCTTAGATATAA). There are 11 mirror repeats within the *intersex gene*; 3 are perfect mirror repeats without any spacer, while 7 are perfect with spacer, with only a single imperfect mirror repeat. As shown in Table 2, the *intersex* gene has many short mirror repeat sequences. This table also includes information about their location, type, and length. The mirror repeats found in the intersex gene of *D. melanogaster* range in length from 7 to 21 base pairs, with the largest repeat being 21 base pairs in length. Fig. 2 depicts the relative location of mirror repeats in the complete sequence (1-727) and two portions of the *intersex* gene. In this figure, two mirror repeats (157-164, 701-721) starts between the other two mirror repeats (151-161,699-707), overlapping each other. Table 3 highlights the mirror repeat sequences to indicate the precise location of mirror repeats within the *intersex* gene of *Drosophila melanogaster*. In this table, it is evident that two mirror repeats overlap with two other mirror repeats, but the remaining seven mirror repeats are dispersed differently.

In the *intersex* gene of *D. melanogaster*, there is just one exon. Exon locations in the *intersex* gene range is from 79 to 645 base pair sequence. There are 6 mirror repeats out of total 11 mirror repeats found in the gene's exon. 4 mirror repeats identified in an exon are perfect with single spacer mirror repeats and 2 are perfect mirror repeats and all with a maximum sequence length of 11 base pairs and a minimum sequence length of 7 base pairs. Table 4 shows the mirror repeats found in the exon of the *intersex* gene. Fig. 3 and 4 depicts the relative location of mirror repeats throughout the *intersex* gene. Table 5 highlights the mirror repeats identified

Table 3. Distribution and sequence of mirror repeats in *intersex* gene of *D. melanogaster*

| CDS | Number of mirror repeats | Sequence with highlighted mirror repeats |
|---|---|---|
| 1-727 | 11 | TTTTTCTTCACAAAAAGAAAACAATTCGCGGCTGTTCAATATTTTT<br>CTCCCCCAATATCCATCGATTTCGAGTGCCAAATGAATCCCAACA<br>TGAACATGATGCCCATGTCTGGGCCACAAATGATGCAGGTAATGC<br>AATCCTCGCCATCGGGACCCCCAGGACCAGTGCAGCATCAACAGC<br>AGCAGCCTCCACAGCCACTGCAGCAGCAGCAGCAGGCCGAAAAA<br>TTGGACAACATTTCCAGGGTGAAGAGTTTGCTGGGACCACTGCGG<br>GAGTCCATGTTCCTCACCATCCGGTCGAGCGCCTTCGCGCTGCAG<br>CAAAACAATCTCGCGGACAACTTAAAGAGGGACACGGGTGCCCA<br>CCATGTTCCGCGGTTCGACAAGCACTTGGAAGACTTTTACGCCTG<br>TTGCGACCAGATCGAGATCCACTTGAAGACGGCGATGCAGTGCCT<br>CCAGCAGCAGAACTCCTCCAATCACTATCTCCCCGGTCCGGTGAC<br>TCCCATGCGCATGGAGACCTTTATGCCGGACAACGCCGGCCCCAT<br>TTCGTATCCCACTTACTTGAACACGGTCCGCGTTCACATACAGTCC<br>GCCAAGGATATACACGACACTCTGATTAGTGCCGCGCAGAACATT<br>TCGCAGGCTGATTGATAGTTGTAGTAGCATTAGGTTTTATTATTTC<br>ACACGCATGCACTTAAGTTAAGTAAATAAATATTCTTAGATATAA<br>GATAAC |

Table 4. Shows mirror repeats in exon of *intersex* gene of *D. melanogaster*

| EXON | Expected threshold | No. of mirror repeats | Mirror repeat | Position in exon | Type |
|---|---|---|---|---|---|
| 79-645 | 20 | 6 | 1. GACCCCCAGG | 73-83 | Perfect with single spacer |
| | | | 2. CCAGGACC | 79-86 | Perfect mirror |
| | | | 3. TCGCGCT | 227-233 | Perfect with single spacer |
| | | | 4. ATCACTA | 392-398 | Perfect with single spacer |
| | | | 5. ACATACA | 497-503 | Perfect with single spacer |
| | | | 6. TGATTAGT | 530-537 | Perfect mirror |

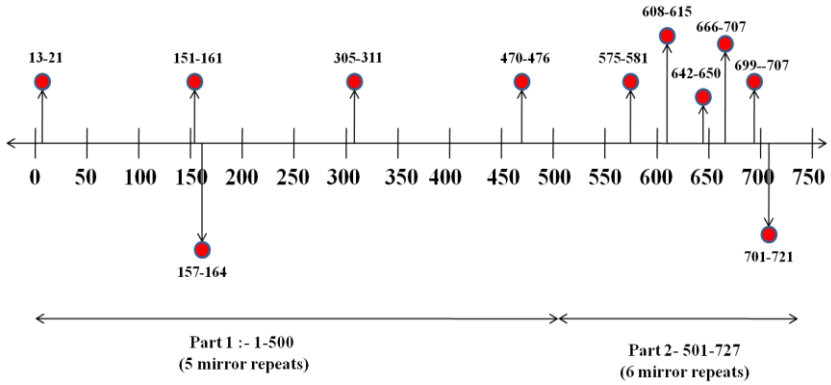


Fig. 3. Relative position of mirror repeats in CDS-1-727 *intersex* gene of *D. mealnogaster*

within the *intersex* gene's exon. This allowed us to observe that only one-mirror repeat overlaps with the another, and the remaining mirror repeats are scattered separately. Comparative observations of mirror repeats in genomes of *Drosophila* spp. (*D. nasuta* and *D. bipectinate*) through mega BLAST tool are given in Table 6; *D. nasuta* Genome possessed 9 mirror repeats in the *intersex* gene, but *D. bipectinata* had just 5. Due to their evolutionary conservation, several mirror repeat sequences identified in both *D.nasuta and D.bipectinata* may play a significant role in these genomes. The mirror repeats denoted by the – sign (indicates absence within that genome) are due to the limitations of the BLAST server tool in handling short input sequences.

Table 5. Distribution of mirror repeats with sequence in exon of *intersex* gene.

| EXON | No. of mirror repeats | Sequence with highlighted mirror repeats |
|---|---|---|
| Exon 1 | 6 | ATGAATCCCAACATGAACATGATGCCCATGTCTGGGCCACAA<br>ATGATGCAGGTAATGCAATCCTCGCCATCGGGACCCCCAGGA<br>CCAGTGCAGCATCAACAGCAGCAGCCTCCACAGCCACTGCAG<br>CAGCAGCAGCAGGCCGAAAAATTGGACAACATTTCCAGGGT<br>GAAGAGTTTGCTGGGACCACTGCGGGAGTCCATGTTCCTCAC<br>CATCCGGTCGAGCGCCTTCGCGCTGCAGCAAAACAATCTCGC<br>GGACAACTTAAAGAGGGACACGGGTGCCCACCATGTTCCGCG<br>GTTCGACAAGCACTTGGAAGACTTTTACGCCTGTTGCGACCA<br>GATCGAGATCCACTTGAAGACGGCGATGCAGTGCCTCCAGCA<br>GCAGAACTCCTCCAATCACTATCTCCCCGGTCCGGTGACTCCC<br>ATGCGCATGGAGACCTTTATGCCGGACAACGCCGGCCCCATT<br>TCGTATCCCACTTACTTGAACACGGTCCGCGTTCACATACAGT<br>CCGCCAAGGATATACACGACACTCTGATTAGTGCCGCGCAGA<br>ACATTTCGCAGGCTGATTGA |

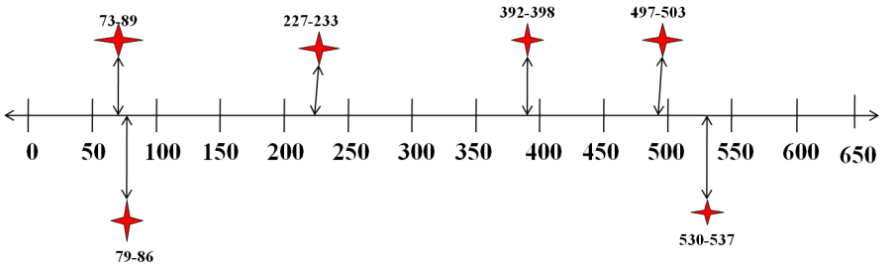


Fig. 4. Relative position of mirror repeats in CDS-73-645 *intersex* gene of *D. mealnogaster*

Table 6. Comparative analysis of mirror repeats of the *intersex* gene of *D. melanogaster* with other *Drosophila* species (*D. nasuta* and *D. bipectinata*) using mega BLAST

| S. No. | Mirror repeats | Mirror repeats in *intersex* gene of *D. melanogaster* | Genome of *D. melanogaster* | Genome of *D. nasuta* | Genome of *D. bipectinata* |
|---|---|---|---|---|---|
| 1. | GGACCCCCAGG | + | + | - | + |
| 2. | AAAAGAAAA | + | - | + | + |
| 3. | CCAGGACC | + | - | + | - |
| 4. | TCGCGCT | + | - | + | - |
| 5. | ATCACTA | + | - | + | - |
| 6. | AATAAATATTCTTAGATATAA | + | + | - | - |
| 7. | TTTATTATTT | + | - | + | + |
| 8. | TTGATAGTT | + | - | + | + |
| 9. | TAAATAAAT | + | - | + | + |
| 10. | TGATTAGT | + | - | + | - |
| 11. | ACATACA | + | - | + | - |

This study applied a manual computational approach to identify mirror repeats with*in the intersex* gene of *D. melanogaster*. A single gene may perform different functions in different cellular environments. Moreover, a gene may be found in multiple species that perform either the same or different functions. This study observed the presence of mirror repeats in the exon and gene sequences of an *intersex* gene. A total of 11 mirror repeats were identified in the complete gene sequence and out of which 6 mirror repeats are present in the exon of this gene. It will be highly significant to find the role of mirror repeats at the molecular level concerning transcription and translation processes within the cell. However, to date, the exact function of mirror repeats has not been elucidated. Further studies may be required to determine the exact role of mirror repeats.

**FINANCIAL SUPPORT**

This research received no external funding.

**AUTHOR CONTRIBUTION STATEMENT**

KS and VB formulated the study, KS performed all the bioinformatics analysis, KS and ND drafted the manuscript, and VB supervised the project and reviewed the manuscript.

**CONFLICT OF INTEREST**

The authors do not have any conflict of interest.

**REFERENCES**

Acharyya M, Chatterjee R N. 2002. Genetic analysis of an *intersex* allele ( *ix* $^{5}$ ) that regulates sexual phenotype of both female and male *Drosophila melanogaster*. Genetical Research 80(1): 7-14.

Aquilina B, Cauchi R J. 2018. Modelling motor neuron disease in fruit flies: Lessons from spinal muscular atrophy. Journal of Neuroscience Methods 310: 3-11.

Arunkumar K P, Nagaraju J. 2011. *Drosophila intersex* orthologue in the silkworm, Bombyx mori and related species. Genetica 139(1): 141-147.

Baker B S. 1989. Sex in flies: the splice of life. Nature 340(6234): 521-524.

Biscotti M A, Olmo E, Heslop Harrison J S. 2015. Repetitive DNA in eukaryotic genomes. Chromosome Research 23(3): 415-420.

Britten R J, Kohne D E. 1968. Repeated Sequences in DNA: Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161(3841): 529-540.

Burtis K C, Baker B S. 1989. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56(6): 997-1010.

Cavaliere D, Di Cara F, Polito, Lino C, Digilio F A. 2009. Cloning and functional characterization of the intersex homologous gene in the pest lepidopteron Maruca vitrata. The International Journal of Developmental Biology 53(7): 1057-1062.

Christiansen A E, Keisman E L, Ahmad S M, Baker B S. 2002. Sex comes in from the cold: the integration of sex and pattern. Trends in Genetics 18(10): 510-516.

Evangelakou Z, Manola M, Gumeni S, Trougakos I P. 2019. Nutrigenomics as a tool to study the impact of diet on aging and age-related diseases: the *Drosophila* approach. Genes & Nutrition 14(1): 1-12.

Garrett Engele C M, Siegal M L, Manoli D S, Williams B C, Li H, Baker B S. 2002a. *intersex* , a gene required for female sexual development in *Drosophila* , is expressed in both sexes and functions together with *doublesex* to regulate terminal differentiation. Development 129(20): 4661-4675.

Gurusaran M, Ravella D, Sekar K. 2013. RepEx: Repeat extractor for biological sequences. Genomics 102(4): 403-408.

Hartl D L, Nurminsky D I, Jones R W, Lozovskaya E R. 1994. Genome structure and evolution in Drosophila: applications of the framework P1 map. Proc. Natl. Acad. Sci. U.S.A. 91(15): 6824-6829.

Jurka J, Kapitonov V V, Kohany O, Jurka M V. 2007. Repetitive Sequences in Complex Genomes: Structure and Evolution. Annu. Rev. Genom. Hum. Genet 8(1): 241-259.

Kapitonov V, Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. U.S.A. 100(11): 6569-6574.

Kaufman T C, 2017. A short history and description of *Drosophila melanogaster* classical genetics: chromosome aberrations, forward genetic screens, and the nature of mutations. Genetics 206(2): 665-689.

Mirzoyan Z, Sollazzo M, Allocca M, Valenza A M, Grifoni D, Bellosta P. 2019. *Drosophila melanogaster*: A model organism to study cancer. Frontiers in Genetics 10: 51.

Pandey U B, Nichols C D. 2011. Human Disease Models in *Drosophila melanogaster* and the Role of the Fly in Therapeutic Drug Discovery. Pharmacol Rev 63(2): 411-436.

Peng Q, Chen J, Pan Y. 2021. From *fruitless* to sex: On the generation and diversification of an innate behavior. Genes, Brain and Behavior 20(8).

Reiter L T, Potocki L, Chien S, Gribskov M, Bier E. 2001. A Systematic Analysis of Human Disease-Associated Gene Sequences In *Drosophila melanogaster*. Genome Research. 11(6): 1114-1125.

Tolwinski N. 2017. Introduction: *Drosophila*—A Model System for Developmental Biology. JDB 5(3): 9.

Treangen T J, Salzberg S L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13(1): 36-46.

Volfovsky N, Haas B J, Salzberg S L. 2001. A clustering method for repeat analysis in DNA sequences. Genome Biol 2(8): 1-11

Wangler M F, Yamamoto S, Bellen H J. 2015. Fruit Flies in Biomedical Research. Genetics 199(3): 639-653.

(Manuscript Received: June, 2022; Revised: December, 2022;
Accepted: December, 2022; Online Published: December, 2022)
Online First in www.entosocindia.org and indianentomology.org Ref. No. e22600