Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886| UGC CARE II
Vol. 19 Issue 05, May- 2024
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

In silico Evaluation of Mirror Repeats within *Transformer* Gene of *Drosophila melanogaster*

Kavita Saini

Department of Zoology, Govt. College for Women, Gohana (Sonipat)

Abstract

A variety of DNA repeats found in prokaryotes and eukaryotes. Recent studies on repeat DNA sequences demonstrates its importance to understand the expression and suppression of gene, and more specifically evolutionary concepts. The genomic sequences of eukaryotes and prokaryotes enrich with variety of repeats, these repeats are simple tandem repeats, satellite DNA or palindromic sequences. These repeats were also classified in different types, such as inverted, direct and mirror repeats. In this study, a computational approach was deployed to find mirror repeats in the *Drosophila melanogaster transformer* (*tra*) gene. This was performed by employing fast parallel complement blast (FPCB), an effective and quick bioinformatics approach to find repeats in gene and genome. In this analysis, we found 26 mirror repeats in complete *transformer* gene while 17 mirror repeats found considering only the exonic region of *tra*. In addition, MEGABLAST is performed to recognize the presence of mirror repeats across the species of *Drosophila* to elucidate their related functionalities. It was observed, mirror repeats from *Drosophila melanogaster* were also found in other species of *Drosophila*. D. *nasuta* and D. *bipectinata* genome showed presence of one or more identical mirror repeats with *D. melanogaster*. This indicates the conserved characteristics of specific mirror repeats in the genome of *Drosophila* that further signifies the evolutionary and functional significance of mirror repeats in the genome. *Drosophila* shares significant overlap with the human genome and used as model organism to test and bring therapeutic solution. This study opens avenue to understand the role of mirror repeats in *tra* gene of *Drosophila melanogaster* that can further be related with the human genome mirror repeats.

Keywords: Tra gene, Mirror repeats, Drosophila, FPCB

1. Introduction

Drosophila melanogaster known as common fruit fly, has 139.5 million base pairs that hold about 15, 682 genes as declared by Ensemble release in the earlier study (Nothiger, n.d.). 60% of the appeared genome is functional for non-protein coding DNA which is involved in gene expression control. Moreover, *Drosophila* genome is 60% homologous to the human genome with less redundancy and 75% of the genes that are responsible for human diseases were also found in these flies. In the account of these characteristics, *Drosophila* is suitable for the study of complex pathways that are highly relevant to biomedical research (Tolwinski, 2017). *Drosophila* is useful to study several human diseases such as neurodegenerative disorders, Huntingtons, spinocerebellar ataxia and Alzimers disease (Ugur *et. al*., 2016).

Drosophila melanogaster is used as a model organism in developmental biology and various human diseases, distinct genes regulating diverse phases of development were discovered in *Drosophila*. Many of the genes were homologous and were known to be involved in human development and diseases (Tolwinski, 2017). In the recent development the study of mitochondrial DNA related diseases, it was said that the gene of *Drosophila* can be manipulated by genetic tool (Chen *et al.*, 2019). Mitochondrial DNA of *Drosophila* has covalently closed duplex circle of about 19500 base pairs. It has 90% of coding capacity and 10% A-T rich region which varies in size and sequence between species and strains and are resistant to cloning (Maarten H. L. de Bruijn & 1983).

The small repetitive sequences are classified as direct, inverted, and mirror repeats. Inverted repeats are arranged in the reverse direction and were found to be a critical genetic element for genome instability. They play an important role for the regulation of transcription and translation. Even in bacteria, inverted repeats and their associated hairpin structure are frequently found as a part of the independent transcription terminal (Lillo & Spanò, 2007). Whereas mirror DNA refers to small repeats that are frequently observed in both coding and non-coding parts of the genome (Usha & Sandeep, 2021). It is the small repeat sequence with bilateral or central symmetry in the same strand of the genome sequence. For example, 5'AGTTCATTACTTGA3' consist of sequences AGTTCAT and TACTTGA that are mirror repeats of each other (Bhardwaj Vikash & Gupta Swapnil2, 2013).

Mirror repeats are classified into perfect, imperfect, mirror repeats with spacer nucleotide and mirror repeats without spacer nucleotide. Though imperfect repeats (IMRs) are common in protein-coding DNA region even they overlap each other frequently. Whereas short IMRs occur with the long IMRs and fewer times have different centres of symmetry, making it difficult to assess their relationship with other specific DNA and protein motifs (Lang, 2005). Mirror repeats play a critical role in ensuring the survival of hairpin structure, transcribed messenger RNA, and cruciform structure DNA. It has impact on important biological processes such as translation transcription, recombination, and chromatin structure formation. Moreover, the formation of H and Z DNA by short repetitive sequence can be intrinsically mutagenic in bacteria, yeast, mouse and mammalian cells (Wang & Vasquez, 2017).

In the previous studies mirror repeats were found to be involved in cell cycle, gene expression, chromosome structure, karyotypic evolution and convoluted in chromosomal relocations(Jurka *et al*., 2007). In addition, it was used to characterise the *Arabidopsis thaliana* flowering gene (Usha & Sandeep, 2021). The manual bioinformatics approach is used to identify mirror repeats where FPCB (FAST PARALLEL COMPLEMENT BLAST) is the most commonly used method for analysing or locating mirror repeats in various prokaryotes and eukaryotes. The FPCB consists of three steps: 1. Downloading sequence from the NCBI in FASTA format, 2. Making reverse complement of the sequence, 3. Align query sequence with the parallel sequence and perform BLAST to find the mirror repeats (Zubair *et al*., 2015).

The objective of the current study is to identify the mirror repeats in *Drosophila melanogaster* transformer (*tra*) sex determination gene as well as to study the wide distribution of identified mirror repeats in whole genome of *Drosophila melanogaster*. This study can be used as a precursor to investigate the role and function of mirror repeats in *tra* gene of *Drosophila melanogaster.*

2. Material and Methods

The nucleotide sequence of *tra* (transformer) gene of *Drosophila melanogaster* (Gene Id-39849) was retrieved from NCBI (National centre for Biotechnology information) in FASTA format. The 992 base pairs long *tra* gene was fragmented into 500 base pairs segments. The resultant fragmented sequences were treated as query sequence and the reverse complement

of the coding sequence is generated by reverse complement tool (https://www.bioinformatics.org/sms/revcomp.html) treated as subject sequence. Both the sequences were aligned for the similarity at the local regions using BLAST tool. Figure 1 illustrates the strategy of FPCB.

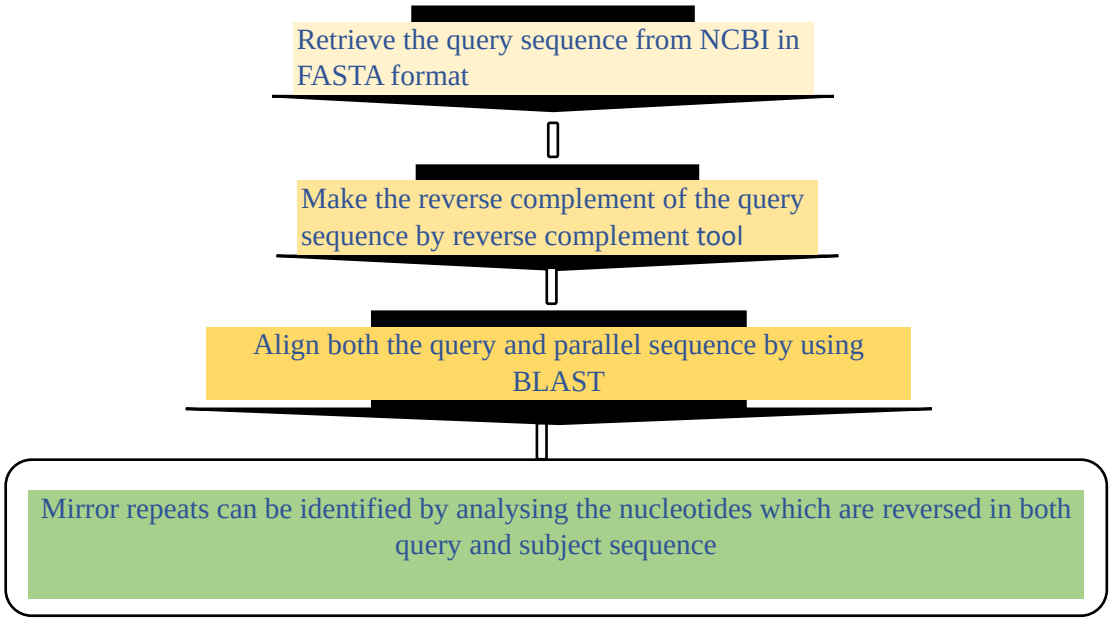

Figure 1.FPCB approach to find out mirror repeats

The program parameters were fixed, word size 7 was used in the alignment and the hits were observed as different expected threshold values (E-value). The maximum number of hits observed was used to identify mirror repeats.

Mirror repeats were identified in alignments where the position number in the subject and query sequences were same in reverse order. The repeats were classified based on the presence of spacer between subject and query sequences.

3. Results

In this study, a manual bioinformatics approach was used to find the mirror repeats (MRs) in the *Drosophila melanogaster.* Our study compromises of the (transformer) *tra* gene, a sex determination gene in *Drosophila*.

Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886| UGC CARE II
Vol. 19 Issue 05, May- 2024
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

3.1 Complete Gene Mirror Repeats

The studied query sequence was retrieved from the NCBI database in FASTA format, the total length of sequence was 992 base pairs. These were fragmented in sets of 500 base pair for the study of mirror repeats. Each fragmented set of base pairs were reverse complemented with the use of reverse complement tool. Afterwards, both the query and subject sequences were aligned using BLAST tool. The frequency of occurrence of mirror repeats with 10+ base pair sequences were higher. Further, 26 mirror repeats were identified in transformer gene (*tra*) at expected threshold (E-value) 20 where 17 mirror repeats were dispersed between the exons of the gene. The occurrence of perfect with one spacer mirror repeats type were present more in all the fragmented sequences, as shown in Table 1.

In the *tra* gene, out of 26 mirror repeats, 17 were perfect with one spacer, 5 were perfect and 4 were imperfect repeats. In the fragment of 501-992 we observed maximum number of hits 42 and corresponding mirror repeats were 14 with 3 perfect, 3 imperfects and 8 perfect with one spacer type.

Table 1. Number of mirrors repeats and their type at expected threshold E-value 20.

CDS	Expected threshold	No. of hits	No. of mirror repeat	Mirror repeat	Position of mirror repeats	Type and how many times present in gene
1-500	20	34	12	1.ACTTTTTGTTGttttttttCTA	97-119	Imperfect/1
				2. AGAGAAAGAGA	175-185	Perfect with one spacer/1
				3. GAGAAAGAG	326-334	Perfect with one spacer/2
				4. TGCCGACAGCAGT	19-31	Imperfect/1
				5. GCATTACG	213-220	Perfect/1
				6. AGACAGA	171-177	Perfect with one spacer/1
				7. AAGAGAA	180-186	Perfect with one spacer/2
				8. ACTCTCA	201-207	Perfect with one spacer/1
				9. AAGAGAA	330-336	Perfect with one spacer/2
				10. AGAAAGA	367-373	Perfect with one spacer/3
				11. GCAGACG	397-403	Perfect with one spacer/1

				12. CTCGCTC	467-473	Perfect with one spacer/1
501-992	20	42	14	1.CCGCCAGCGACCGCC	337-351	Perfect with one spacer/1
				2. ATCATCAACTACTA	158-171	Perfect/1
				3.GCCGCAATCGCAACCGAAGTC GCAGCAGTGAACGAAAACGCCG	39-81	Imperfect/1
				4.TACATTAAAATAATAAA–TACAT	457-478	Imperfect/1
				5.ATAAATACATATATA	469-483	Imperfect/1
				6. ATTTATTTA	433-441	Perfect with one spacer/1
				7. GCGGGGCG	386-393	Perfect/1
				8. TGTTTTGT	450-457	Perfect/1
				9. AAACAAA	220-226	Perfect with one spacer/1
				10. AATCTAA	232-238	Perfect with one spacer/1
				11. ACCGCCA	336-342	Perfect with one spacer/1
				12. TTACATT	239-445	Perfect with one spacer/1
				13. TATATAT	478-484	Perfect with one spacer/1
				14. TCAAACT	486-492	Perfect with one spacer/1

3.2 Exon Mirror Repeats

Further the similar study was performed on the exonic region of *tra* gene. The transformer gene consists four exons in it. Contrary to complete gene sequence, in the exons out of 17 mirror repeats, 13 repeats were found to be perfect with one spacer, 2 were imperfect and 2 were perfect mirror repeats as shown in Table 2.

Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886| UGC CARE II
Vol. 19 Issue 05, May- 2024
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

Table 2. Transformer exon mirror repeats and their types at expected threshold 10.

CDS Exon	Expected threshold	No. of hits	No. of mirror repeat	Mirror repeat	Position of mirror repeats	Type
8-47	10	1	1	1. TGCCGACAGCAGT	12-24	Imperfect
121-191	10	3	3	1. AGAGAAAGAGA	55-65	Perfect with one spacer
				2. AGACAGA	51-57	Perfect with one spacer
				3. AAGAGAA	60-66	Perfect with one spacer
296-699	10	19	10	1. ATCATCAACTACTA	363-376	Perfect
				2.GCCGCAATCGCAACCGAAGTCG CAGCAGTGAACGAAAACGCCG	244-286	Imperfect
				3. GAGAAAGAG	31-39	Perfect with one spacer
				4. AGAAAGA	72-78	Perfect with one spacer
				5. AAGAGAA	35-41	Perfect with one spacer
				6. GCAGACG	102-108	Perfect with one spacer
				7. CTCGCTC	172-178	Perfect with one spacer
				8. AGAGAGA	204-210	Perfect with one spacer
				9. GAGAGAG	205-211	Perfect with one spacer
				10. AGAAAGA	72-78	Perfect with one spacer
757-906	10	7	3	1.CCGCCAGCGACCGCC	81-95	Perfect with one spacer
				2. GCGGGGCG	130-137	Perfect
				3. ACCGCCA	80-86	Perfect with one spacer

4. Discussion

The *transformer* (*tra*) gene controls all aspects of somatic sexual differentiation in female *Drosophila melanogaster* but not in males (M Mckeown *et. al*., 1987). Tra regulates sex differences in body size in a separate pathway from the canonical sex determination pathway and other aspects of sexual dimorphism. Tra function in the fat body, on the other hand, regulates growth in a non-cell autonomous manner by regulating the brain's secretion of insulin-like peptides Elizabeth J. Rideout *et. al*., (2015).

Based on the preceding studies, we know that *tra* is a female specific gene in *Drosophila.* It is also detected that the *tra* gene of *Drosophila melanogaster* has many mirror repeats with varying base pair lengths. Which mostly occurred in 10+ base pair sequences in

the gene and genome. The study of mirror repeats aids evolutionary understanding for different genera, and their further biological approaches.

5. Conclusion

Our study concluded that mirror repeats are frequently distributed in the *Drosophila* genome. There were 26 and 17 mirror repeats identified in both *transformer* gene and *transformer* exon respectively. Amongst these mirror repeats, perfect with one spacer mirror repeats were found more repeatedly in *transformer* gene and exon as compared to perfect and imperfect mirror repeats.

Furthermore, the frequent occurrence of mirror repeats in gene was analysed and proposed for the further research to understand the specific role of mirror repeats in all genomes and genes of other organisms. These may also aid in the study of mutation or evolutionary history of all living organism. In addition, this study can contribute to the development of new tools in molecular and bioinformatics approaches that can help to elucidate the mirror repeats genomic relevance.

References

Bacolla, A., Cooper, D. N., & Vasquez, K. M. (2010). Non- B DNA Structure and Mutations Causing Human Genetic Disease. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed.). Wiley. https://doi.org/10.1002/9780470015902.a0022657

Bhardwaj Vikash, & Gupta Swapnil2. (n.d.). *FPCB : a simple and swift strategy for mirror repeat identification*.

Chen, Z., Zhang, F., & Xu, H. (2019). Human mitochondrial DNA diseases and *Drosophila* models. *Journal of Genetics and Genomics*, *46*(4), 201–212. https://doi.org/10.1016/j.jgg.2019.03.009

Cugusi, S., Kallappagoudar, S., Ling, H., & Lucchesi, J. C. (2015). The *Drosophila* Helicase Maleless (*MLE*) is Implicated in Functions Distinct From its Role in Dosage Compensation*. *Molecular & Cellular Proteomics*, *14*(6), 1478–1488. https://doi.org/10.1074/mcp.M114.040667

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genetics*, *7*(12), e1002384. https://doi.org/10.1371/journal.pgen.1002384

Elizabeth J. Rideout, Marcus S. Narsaiya, Savraj S. Grewal, (2015). The Sex Determination Gene *transformer* Regulates Male-Female Differences in *Drosophila* Body Size. *PLoS Genetics*, https://doi.org/10.1371/journal.pgen.1005683

Gadagkar, R., Nanjundiah, V., Joshi, N. V., & Chandra, H. S. (1982). Dosage compensation and sex determination in*Drosophila*: Mechanism of measurement of the X/A ratio. *Journal of Biosciences*, *4*(3), 377–390. https://doi.org/10.1007/BF02702750

Jurka, J., Kapitonov, V. V., Kohany, O., & Jurka, M. V. (2007). Repetitive Sequences in Complex Genomes: Structure and Evolution. *Annual Review of Genomics and Human Genetics*, *8*(1), 241–259. https://doi.org/10.1146/annurev.genom.8.080706.092416

Lang, D. M. (2005). Imperfect DNA mirror repeats in E. coli TnsA and other protein-coding DNA. *Biosystems*, *81*(3), 183–207. https://doi.org/10.1016/j.biosystems.2005.03.004

Lang, D. M. (2007). Imperfect DNA mirror repeats in the gag gene of HIV-1 (HXB2) identify key functional domains and coincide with protein structural elements in each of the mature proteins. *Virology Journal*, *4*(1), 113. https://doi.org/10.1186/1743-422X-4-113

Lillo, F., & Spanò, M. (2007). Inverted and mirror repeats in model nucleotide sequences. *Physical Review E*, *76*(4), 041914. https://doi.org/10.1103/PhysRevE.76.041914

Lovett, S. T. (2004). Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences: Replication slippage. *Molecular Microbiology*, *52*(5), 1243–1253. https://doi.org/10.1111/j.1365-2958.2004.04076.x

Maarten H. L. de Bruijn, & 1983. (n.d.). *Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code*.

McKeown, M., Belote, J. M., & Baker, B. S. (1987). A molecular analysis of transformer, a gene in Drosophila melanogaster that controls female sexual differentiation. *Cell*, *48*(3), 489–499. https://doi.org/10.1016/0092-8674(87)90199-1

Nothiger, R. (n.d.). *Genetics of sex determination: What can we learn from Drosophila?* 8.

Oliver, B., Kim, Y. J., & Baker, B. S. (1993). Sex-lethal, master and slave: A hierarchy of germ-line sex determination in *Drosophila*. *Development*, *119*(3), 897–908. https://doi.org/10.1242/dev.119.3.897

Sandeep, Y., Usha, Y., & Dinesh C., S. (2021). In-silico evaluation of 'Mirror Repeats' In HIV Genome. *International Journal of Pharma and Bio Sciences*, *11*(5). https://doi.org/10.22376/ijpbs/lpr.2021.11.5.L81-87

Tolwinski, N. (2017). Introduction: *Drosophila*—A Model System for Developmental Biology. *Journal of Developmental Biology*, *5*(3), 9. https://doi.org/10.3390/jdb5030009

Ugur, B., Chen, K., & Bellen, H. J. (2016). *Drosophila* tools and assays for the study of human diseases. *Disease Models & Mechanisms*, *9*(3), 235–244. https://doi.org/10.1242/dmm.023762

Usha, Y., & Sandeep, Y. (2021). Characterization of Flowering Genes of Arabidopsis thaliana for Mirror Repeats. *Biointerface Research in Applied Chemistry*, *12*(3), 2852–2861. https://doi.org/10.33263/BRIAC123.28522861

Wang, G., & Vasquez, K. (2017). Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes*, *8*(1), 17. https://doi.org/10.3390/genes8010017

Zubair, M. S., Muhammad, M. H., & Ahmad, R. G. (2015). *Bioinformatics Study On Mirror DNA In Mycobacterium Tuberculosis Using Fast Parallel Complement Blast Strategy.4*(12), 7.