Indian Journal of Entomology 85(1): 14-18 (2023) DoI. No.: 10.55446/IJE.2022.983

IDENTIFICATION OF MIRROR REPEATS WITHIN THE *MALELESS* (*MLE*) GENE OF *DROSOPHILA MELANOGASTER* MEIGEN

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

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

DNA repeats present in prokaryotes and eukaryotes genomes that include simple tandem repeats, satellite DNA, or palindromic sequences are classified as inverted, direct, and mirror repeats (MRs). Out of these, MRs are not well studied in *Drosophila melanogaster*. In this study, manual bioinformatics approach was used to find MRs in *D. melanogaster* maleless *(mle*) gene. In this analysis, 123 MRs were found within the complete *mle* gene, while 78 MRs were found in the exonic region of the *mle* gene. MegaBLAST was performed to elucidate the presence of identified MRs across the genomes of *D. melanogaster, D. nasuta* and *D. bipectinate*. These demonstrated the conserved characteristics of specific MRs in *Drosophila* genome and the evolutionary and functional significance of MRs in diverse genomes. This study establishes a link between the presence of MRs in *mle* gene of *D. melanogaster* and MRs in human genome.

Key words: DNA repeats, palindromic/ gene sequence, inverted/ direct repeat, mega BLAST, repetitive DNA, perfect/ imperfect mirror repeats, parallel complement, sex determination gene, exons

The DNA molecule serves as a central repository for genetic data. Genetic data held within the DNA regulates a wide range of biological functions. In addition to its normal B-DNA form, DNA can acquire a variety of shapes as a result of mutations, strand breakage, and inadequate crossing over. This can result in the formation of repetitive sequences within the genome (Gurusaran et al., 2013; Jain et al., 2008; Zattera et al., 2020). Any organism possesses homologous DNA segments in multiple copies, which are referred to as repetitive DNA sequences, which can be essential for regulating the cell cycle, gene expression, chromosome structure, and karyotypic evolution (de Koning et al., 2011; Jurka et al., 2007; Mehrotra and Goyal, 2014; Zattera et al., 2020). These repeats impact important biological processes such as translation, transcription, recombination, and chromatin structure formation. Moreover, the formation of H and Z DNA by short repetitive sequences can be intrinsically mutagenic in bacteria, yeast, mouse and mammalian cells (Pandya et al., 2021; Wang and Vasquez, 2017). The small repetitive sequences on the basis of the alignment of nucleotide sequences can be classified as direct, inverted, and mirror repeats (Gurusaran et al., 2013; Jiang et al., 2018; Mirkin, 2001). When a repeating DNA fragment is inverted on the same strand, it is called an "inverted repeat" (Mirkin, 2001). They play an important role in the regulation of transcription and translation (Varshney et al., 2020). Even in bacteria, inverted repeats and their associated hairpin structure are frequently found as a part of the independent transcription terminal (Brazda et al., 2020; Lillo and Spanò, 2007). Direct repeats are the sequences that are repeated multiple times on the same strand of DNA in the same orientation (Mirkin, 2001). Both direct and inverted repeats can stimulate the genetic rearrangements that cause deletion and duplication of genetic material (Lovett, 2004; Marie and Symington, 2022).

On the other hand, repetitive DNA sequences are divided into two groups based on their distribution within the genome: tandem repeats and interspersed repeats (Jurka et al., 2007; Pathak and Ali, 2012; Zattera et al., 2020). Long interspersed elements (LINES) and short interspersed elements (SINES) are two forms of interspersed repeats based on the length of DNA sequences, while microsatellite, minisatellite, and satellite DNA are types of tandem repeats (Mehrotra and Goyal, 2014; Pathak and Ali, 2012; Zattera et al., 2020). Mirror DNA repeats are small repeat sequences with bilateral or central symmetry in the same strand (Mirkin, 2001). It means that one section of the sequence fragment is an exact replica of another. In the case of mirror repeats (MR), parallel complement and anti-parallel complement are the same, while in normal DNA, parallel complement and anti-parallel complement are not identical.

The model organism *D. melanogaster* is employed

for the study of numerous human disorders because it contains distinct genes that regulate various developmental processes. Numerous homologous genes are known to play a role in human development and disease (Mirzoyan et al., 2019; Perveen, 2018; Tolwinski, 2017). *Drosophila* is useful for studying several human diseases such as neurodegenerative disorders viz., Huntington's, spinocerebellar ataxia and Alzheimer's disease (Bolus et al., 2020; Ugur et al., 2016). The mechanisms of determining *Drosophila's* sexual differentiation have been revealed through genetic, developmental and molecular investigations (Chen et al., 2019; Cugusi et al., 2015; Gadagkar et al., 1982; Oliver et al., 1993) The ratio of X chromosomes to sets of autosomes (X:A), which is a chromosomal signal, was used to control a small number of regulatory genes, which in turn direct differentiation to produce the morphological, physiological, and behavioural distinctions that differentiate males and females (Nöthiger and Steinmann-Zwicky, 1987). Maleless (*mle*) is required in both somatic and germ cells for male *D. melanogaster*. *Mle* is essential for X-chromosome dosage compensation in somatic cells (Lv et al., 2019). Unknown is the function of *mle* in the germline (Cugusi et al., 2015; Rastelli and Kuroda, 1998). The current study aims to identify the mirror repeats in the *D. melanogaster* maleless (*mle*) sex determination gene as well as to study the wide distribution of identified mirror repeats within the whole genome of *D. melanogaster* and other related species of *Drosophila*.

MATERIALS AND METHODS

The nucleotide sequence of the *mle* (maleless) gene of *D. melanogaster* (gene Id- 35523) was retrieved from NCBI (National Centre for Biotechnology Information) in FASTA format. The 5922-base pair long *mle* gene was fragmented into 500 base pair segments. The resultant fragmented sequences were treated as query sequences, and the reverse complement of the coding sequence was generated by a reverse complement tool (https://www.bioinformatics.org/sms/revcomp.html) and treated as the subject sequence. Both the sequences were aligned for similarity in the local regions using the BLAST tool. Fig. 1 illustrates the strategy used to identify mirror repeats. Using the strategy, the search was also conducted for mirror repeats within the break point region (the point where it was fragmented out of the gene into 500 nucleotide base pairs). https://nonb-abcc.ncifcrf.gov/apps/nBMST/default/ Non-B DNA Motif Search Tool was used to search for mirror repeats within the *mle* gene.

The programme parameters were fixed, word size 7 was used in the alignment and hits were observed at different expected threshold values (E-values). The E-value at which the greatest number of hits was observed was used to identify mirror repeats. Mirror repeats were identified in alignments where the position number in the subject and query sequences was the same in reverse order. The mirror repeats were classified based on the presence of a spacer between the

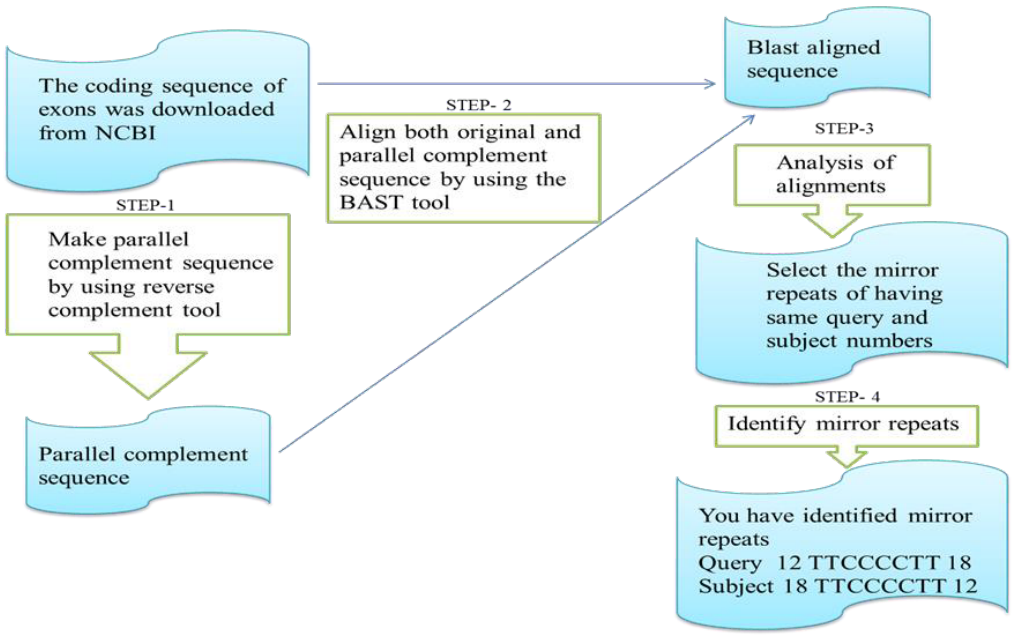

Fig. 1. Schematic representation of the methodology used to identify mirror

subject and query sequences. Here, Mega BLAST was used to search for the identified mirror repeats in the entire genome. Eventually, a phylogenetic analysis was performed to determine the conserved characteristics of these mirror repeats in different genomes like *D. nasuta* and *D. bipectinate*. During the search, algorithm parameters including word size, expected threshold, and maximum target sequence were adjusted to maximize the hits.

RESULTS AND DISCUSSION

In this study, a manual bioinformatics approach was used to find the mirror repeats within the *D. melanogaster's* maleless *mle* gene, a sex determination gene in *Drosophila*. A straightforward bioinformatics approach was employed to locate mirror repeats within the *mle* gene of *Drosophila melanogaster,* which is involved in the sex determination of *Drosophila*. The total nucleotide sequence length of *mle* gene is 5922 nucleotide base pairs with 6 exons. 123 mirror repeats were identified in the *maleless* gene (*mle*) at an expected threshold (E-value) 20. Of those 78 mirror repeats were dispersed between the six exons of the *mle* gene. As the *mle* gene (taking 500 nucleotide base pairs) was divided into 12 segments, a total 11 breakpoint regions were created. Moreover, mirror repeats were also searched within each break point region. Out of 11 breakpoint regions, only one breakpoint region was able to detect mirror repeats. Consequently, in the mle gene, 124 mirror repeats were identified. The repeats known as "perfect mirror repeats" are those that have identical sequences aligned around a central axis. A perfect mirror repeats may have a spacer element in between them, which was termed a perfect mirror with one spacer. Mismatches can be seen in imperfect mirror repeats around the central axis. The occurrence of the perfect with one spacer mirror repeat type was present more in all fragmented sequences, as shown in Table 1. In the maleless gene, out of 124 mirror repeats, 110 were perfect mirror repeats and 14 were imperfect repeats.

In the *mle* gene, 88% of identified mirror repeats are perfect mirror repeats, as shown in Fig. 2. In the fragment 1501-2000 of the *mle* gene, the maximum number of mirror repeats was observed (19 with one perfect, six imperfect, and twelve perfects with one spacer type). Thus, this region within the *mle* gene is rich in mirror repeats. However, the minimum number of mirror repeats of 5 was found in the region of 5001–5500 with five perfect with one spacer type in the *mle* gene, as shown in Table 1. The composition of perfect mirror repeats (P) and imperfect mirror repeats (IP) in complete genes and in exons is shown in Fig. 2.

Table 1. Mirror repeats in *mle* gene and their types at the expected threshold E-value 20

CDS	Expected threshold	Hits	Mirror repeats	Types of mirror repeats
1-500	20	46	14	P- 4 IP-1 POS-9
501-1000	20	30	10	P-1 IP-3 POS-6
1001-1500	20	28	10	IP-1 POS-09
15001- 2000	20	65	19	P-1 IP-6 POS-12
2001-2500	20	23	09	P-2 IP-2 POS-05
2501-3000	20	21	09	IP-1 POS-08
3001-3500	20	48	12	P-04 POS-08
3501-4000	20	42	08	P-01 POS-07
4001-4500	20	28	08	P-03 POS-05
4501-5000	20	18	06	POS-06
5001-5500	20	24	05	POS-05
5501-5922	20	55	13	P-04 POS-09

P stands for perfect, IP stands for imperfect and POS stands for perfect with one spacer mirror repeats

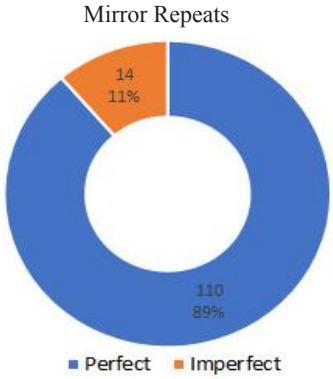

Fig. 2. Composition of perfect mirror repeats (P) and imperfect mirror repeats (IP) in (a) complete *mle* gene of *D melanogaster*

To further investigate the distribution of identified mirror repeats across the species, Mega BLAST tool was used. It was observed that the distribution of 25 mirror repeats has a length of 10+ base pairs by using

the Mega BLAST tool. As a result of the Mega BLAST tool's restrictions, which include a word size limit of 16, it was unable to detect all small-scale mirror repeats. Table 2 contains the entire Mega BLAST result for the large mirror repeat found in various parts of the mle gene. Out of 25 repeats, two repeat sequences were not found in the genome of *D. melanogaster*, which clearly indicates there is a need to upgrade BLAST/ Mega BLAST tools for efficient search for small nucleotide sequences. The ancillary file contains the entire Mega BLAST result for all mirror repeats. Lastly, non-B DNA motif search tool (nBMST) repeats were used within the mle gene. Only one mirror repeat "TATTGGAGGAGGATATGGAA ATAATGCAGGAGGTTAT" having a 17 nucleotide base pair spacer was found using this tool, the result of non-B DNA search tool is shown in Fig. 3. It was observed that there is a need to develop bioinformatics tools that can identify mirror repeats in a gene, clearly and quickly.

Using a simple manual computational approach, it was possible to detect 124 mirror repeats within the mle gene of *D. melanogaster*. This study can contribute to the development of new tools in molecular and bioinformatics approaches that can help elucidate the mirror repeats' genomic relevance. This kind of study has not been done before. Furthermore, molecular and biophysical research could determine how to comprehend the molecular function of identified mirror repeats.

Sequence_name	Source	Type	Start	Stop	Length	Score	Strand	Repeat	Spacer	Permutations	Subset
seq	ABCC	Mirror_Repeat	5370	5406	37	NA	+	10	17	1	0

Composition	Sequence
3A/0C/4G/3T	tattggaggaggatatggaaataatgcaggaggttat

Fig. 3. Result of non-B DNA motif search tool

Table 2. Distribution of selected mirror repeats (10+ bp)

Identified mirror repeats in *mle* gene of *D. melanogaster*	*D. melanogaster* (genome)	*D. nasuta* (genome)	*D. bipectinata* (genome)
ACCTATATTGGTGTG GGTAATTCCA	+	-	-
ACAAGCAGACGAA CA	-	-	-
GAACATTTACAAG	+	-	+
GCTTCCTTTTCAACTTT–CTTCG	+	-	-
TGTCGGGCTGT	+	-	+
ATGGCGGTGACGGC GGTA	+	-	-
CGCCGGCAGAACGA—-TGTGGTACGAT GCTCCGAACTACA	+	-	-
—AGAACCACGTCC AGCTTCCTTGCGTG GGGTCTGTAAGCAA AACGGACGC			
TCTGTCGGCTGGCT	+	-	-
TTATCAGACAAAGA CTATT	+	-	-
TATATTATTATAT	+	-	+
TTTAGAGATTT	+	-	+
CTAAAATCTGCGTCTTTAATC	+	-	-
TTAGCTAAATGTTG TATACCGGTT	+	-	-
CGAAACAACAATGC	+	+	+
TACTTTATAAA	+	-	+
TCTATATATAGCT TTAAATT	+	-	-
AAAATGTAGTGTAC GATGTAAAA	+	-	-
GGGCTCCTCCTCAG G	+	-	+
TAAGCTCGAAT	+	-	+
GTGGCCGAGCGCGTAGCT-CGCGAG-CGG TG	-	-	-
CATTAAATTAC	+	-	+
CTTGATATAGGTC	+	-	-
AAAAACAAAAA	+	+	+
ATACATACATA	+	+	+
TACATACATATAT	+	+	+

+ denotes presence and – denotes absence

ACKNOWLEDGEMENTS

The authors thank the Department of Zoology, Baba Mastnath University, Asthal Bohar Rohtak for providing the space and resources.

FINANCIAL SUPPORT

This research received no external funding

AUTHOR CONTRIBUTION STATEMENT

Vikash Bhardwaj: Conceptualization, methodology, software, supervision, reviewing, and editing; Kavita Saini: Data curation, investigation, and writing-original draft preparation; and Namrata Dangi: Writing- review and editing.

CONFLICTS OF INTEREST

The authors do not have any conflicts of interest.

REFERENCES

Bolus H, Crocker K, Boekhoff Falk G, Chtarbanova S. 2020. Modeling neurodegenerative disorders in *Drosophila melanogaster*. IJMS 21(9): 3055.

Brazda V, Fojta M, Bowater R P. 2020. Structures and stability of simple DNA repeats from bacteria. Biochemical Journal 477(2): 325-339.

Chen Z, Zhang F, Xu H. 2019. Human mitochondrial DNA diseases and *Drosophila* models. Journal of Genetics and Genomics 46(4): 201-212.

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 and Cellular Proteomics 14(6): 1478-1488.

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.

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.

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

Jain A, Wang G, Vasquez K M. 2008. DNA triple helices: Biological consequences and therapeutic potential. Biochemie 90(8): 1117-1130.

Jiang J, Wang Y, Sušac L, Chan H, Basu R, Zhou Z H, Feigon J. 2018. Structure of Telomerase with Telomeric DNA. Cell 173 (13): 1179-1190.

Jurka J, Kapitonov V V, Kohany O, Jurka M V. 2007. Repetitive sequences in complex genomes: structure and evolution. Annual Review Genomics and Human Genetics 8(1): 241-259.

Lillo F, Spanò M. 2007. Inverted and mirror repeats in model nucleotide sequences. Physical Review E 76(4): 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.

Lv M, Yao Y, Li F, Xu L, Yang L, Gong Q, Xu Y Z, Shi Y, Fan Y J, Tang Y. 2019. Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in *Drosophila*. Nucleic Acids Research 47(6): 3142-3157.

Marie L, Symington L S. 2022. Mechanism for inverted-repeat recombination induced by a replication fork barrier. Nature Communications 13(1): 32.

Mehrotra S, Goyal V. 2014. Repetitive Sequences in Plant Nuclear DNA: Types, distribution, evolution and function. Genomics, Proteomics and Bioinformatics 12(4): 164-171.

Mirkin S M, 2001. DNA Topology: Fundamentals. John Wiley and Sons, Ltd (ed.), ELS. Wiley.

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.

Nöthiger R, Steinmann Z M. 1987. Genetics of sex determination: what can we learn from *Drosophila* ? Development 101: 17-24.

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.

Pandya N, Bhagwat S R, Kumar A. 2021. Regulatory role of Non-cnonical DNA Polymorphisms in human genome and their relevance in cancer. Biochimica et Biophysica Acta (BBA)- Reviews on Cancer 1876(1): 188594.

Pathak D, Ali S. 2012. Repetitive DNA: A tool to explore animal genomes/ transcriptomes. Meroni G (ed.), Functional Genomics. InTech. 212

Perveen, F K. 2018. Introduction to *Drosophila*. Perveen F K (ed.), *Drosophila melanogaster* - model for recent advances in genetics and therapeutics. InTech. 268

Rastelli L, Kuroda M I. 1998. An analysis of maleless and histone H4 acetylation in *Drosophila melanogaster* spermatogenesis. Mechanisms of Development 71(1-2): 107-117.

Tolwinski N. 2017. Introduction: Drosophila- a model system for developmental biology. Journal of Developmental Biology 5(3): 9.

Ugur B, Chen K, Bellen H J. 2016. *Drosophila* tools and assays for the study of human diseases. Disease Models and Mechanisms 9(3): 235-244.

Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. 2020. The regulation and functions of DNA and RNA G-quadruplexes. Naure Reviews Molecular Cell Biology 21(8): 459-474.

Wang G, Vasquez K. 2017. Effects of replication and transcription on DNA structure-related genetic instability. Genes 8(1): 17.

Zattera M L, Gazolla C B, Soares A de A, Gazoni T, Pollet N, Recco P S M, Bruschi D P. 2020. Evolutionary dynamics of the repetitive DNA in the karyotypes of *Pipa carvalhoi* and *Xenopus tropicalis* (Anura, Pipidae). Frontiers in Genetics 11: 637.

(Manuscript Received: August, 2022; Revised: December, 2022;
Accepted: December, 2022; Online Published: December, 2022)
Online First in www.entosocindia.org and indianentomology.org Ref. No. e22850