



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)

www.iasir.net

EVOLUTIONARY STUDY OF X-LINKED GENES IN DROSOPHILA

Shagun Shukla

CSIR- Indian Institute of Toxicology Research, Lucknow,
Uttar Pradesh, INDIA

Abstract: To understand gene evolution in depth, studying evolution of both exon and intron in terms of their size, position, number and GC content is important. Till one and half decades, introns were considered as junk DNA and were out of focus from studies involving gene expression. However, a major breakthrough on introns, as playing significant role in splicing mechanism ultimately regulating gene expression, motivated researchers to explore more. Intron length depends on a number of factors like presence of RNA genes, regulatory elements, insertion of transposable elements, frequency and size of deletions etc. The genome of drosophila is 70% homologous to humans and hence human genes can be matched against drosophila counterparts. Therefore, studies on drosophila models will help in development of effective drugs. The major goal of this study is to explore the exon-intron architecture of X-linked and autosome-linked genes of *Drosophila* species and among their closely related species.

Keywords: *Drosophila melanogaster*; introns; exons; X-linked genes; *Anopheles gambiae*.

I. Introduction

Studying exon-intron architecture becomes the prime focus among evolutionary biologists to understand the gene evolutionary path in depth. To understand the gene evolution, studying the evolution of both exon and intron in term of their size, position, number and GC content is equally important. Till one and half decades, the introns were considered as the junk DNA and were out of focus from studies involving gene expression. However, a major breakthrough on introns, as playing significant role in splicing mechanism ultimately regulating gene expression, motivate researcher to explore more on their number, size and distribution across genes There are earlier reports about the exon and intron size variation across genome at inter-species level, however, very little is known about the pattern of structural variation as they are regulated by many factors e.g. recombination rate, inversion polymorphism, gene density, G and C content and gene function and also, the mechanism behind the same. The major goal of this study is to explore the exon-intron architecture of X-linked and autosome-linked genes of *Drosophila* species and among their closely related species to understand whether they are following a similar trend/pattern of variation, which may provide some clue to understand the gene/protein evolution and the underlying mechanism in higher eukaryotes including human.

Here, the in- silico analysis of the intron and exon size variation are presented with respect to high and low recombination region, with or without presence of inversion polymorphism (especially in *Drosophila melanogaster*) and also according to the gene density region in 2nd chromosome and X-chromosome of *Drosophila melanogaster* and X-chromosome of *Tribolium castaneum* and *Anopheles gambiae*. Further the genes selected under this study can be individually examined with respect to their structure and function. The study can be further extended to wet laboratory to observe the intron variation across different *Drosophila* species which may provide any clue to understand in detail the exon-intron length evolution and the evolutionary force(s) which might be acting selectively according to the demand of the gene function.

II. Literature Review

Initially, the sex was not determined by the genetic material of the organism. The temperature determines the sex of the offspring in reptiles. When mammals evolved from reptiles, millions of years ago, a normal pair of autosome evolved into sex chromosomes (X and Y). Y chromosome has homologous genes in X chromosome, suggesting the derivation of Y chromosome from X-like chromosome. Through the years of evolution, X and Y chromosomes stopped recombining during gamete formation. Addition of inversion in chromosome prevents their alignment and possible recombination. Building up of deleterious mutations on non-recombining Y chromosome lead to degeneration and shrinking of Y chromosome. But X chromosome remembered its genetic integrity and size, as it could recombine with the other X in females [1]. Studies have suggested that sex-linked genes evolve faster in comparison to non sex-linked genes. Also In some studies it was shown that, the male biased genes were relatively more divergent than female biased genes. This means the female biased genes take longer to evolve [2, 3, 4].

The argument continues on the topic whether nuclear introns were present in eukaryotes from the beginning (introns-early) or invaded them later in evolution (introns-late). The introns-first hypothesis is not supported by much evidence as there is no sign that any prokaryote has ever possessed a spliceosome or introns. However exon-intron structure of protein-coding genes seems to have evolved throughout and introns were a major factor of evolution all through the history of eukaryotes [5].

Exon and intron length vary among different genomes and also within a genome. The exon and intron numbers and lengths also varies depending on the species and the chromosome region. The reasons for intra- and interspecific heterogeneity of exon and intron lengths remain unknown [6]. Exon codes for a particular function or expression and are widely studied. The introns play a critical role in transcription of the mature mRNA from pre-mRNA through splicing mechanism and could possibly affect the translational form of proteins as a result of which multiple proteins are encoded from a single gene [7, 8]. The current opinion is that the exon-intron type of gene structure facilitates the emergence of new combinations of exons (and corresponding functional protein domains) and novel genes in evolution. Some intron characteristics, e.g., length, asymmetric position in the protein-coding sequence, relation to the exon length, intron density, splice sites, etc., are used in molecular evolution studies. Precise identification of the intron length and phase and their effects on alternative splicing is of great importance. The presence of introns forms a background for alternative splicing, which increases the diversity of gene products [6]. So introns are no more considered the junk DNA [9]. Earlier studies on exon-intron architecture revealed that the splice sites are recognised across the exons by the length of introns. In addition to that intron size also intelligently influence on the likelihood of an exon splicing type; whether constitutively or alternatively and play a major role on exon skipping [7, 10]. Intron length depends on a number of factors like presence of RNA genes, regulatory elements, insertion of transposable elements, frequency and size of deletions etc. A supposed correlation between the intron length and the genome size has not been confirmed. It is quite possible that the lengths of introns depend on their function, which justifies the interest in the length variation and other features of exons and introns across different chromosomes and genomes. The available databases and literature contain extensive information concerning the features of exons and introns. A number of publications report associations between intron length and various characteristics of genes and genomes. For example, in the human genome, gene introns in GC rich chromosome regions are shorter and are more closely spaced than in regions with a lower GC content. The mere presence of introns, as well as their location, influences some gene characteristics [6].

In *Drosophila melanogaster* and humans, the introns at telomeric and centeromeric ends are longer where recombination rate is low, and introns are shorter at the middle where recombination rate is high. Few models were suggested that might explain the above observation. At the two ends of chromosome are the conserved sequences that are naturally selected [11]. The completion of whole genome projects of several organisms especially 12 species of *Drosophila* and of human have raised many evolutionary questions regarding the structural and functional pattern of genes, their encoded proteins and the regulatory mechanism [12]. *Drosophila melanogaster* is a versatile model organism for studying structural and functional genomics, and on which evolution can be easily studied. It has a short life span and a number of generations can be studied on in lesser time period. It has a simple diet and can be maintained inexpensively in labs. It has many contrasting traits in the population making it easier to study. It also has a large polytene chromosome, whose band patterns allow to map genes accurately [13, 14]. The genome of drosophila is 70% homologous to humans and hence human genes can be matched against drosophila counterparts. Therefore, diseased genes can be studied on drosophila models which help in development of effective drugs. Many basic physiological, biological, and neurological properties are conserved between mammals and *D. melanogaster*. Nearly 75% of human disease-causing genes are believed to have a functional homolog in drosophila. Various human diseases (Neurodegenerative disorders, cardiovascular diseases, metabolic disorders, etc.) are vastly studied in drosophila [15]. They are also helpful to understand evolutionary mechanism behind intron-exon variation.

The African malaria vector, *Anopheles gambiae*, has become widely studied as a system for exploring models of speciation and is characterized by multiple polymorphic chromosomal inversions. Some studies conclude that genes located on the sex chromosome may be the major force driving speciation between these chromosomal forms of *A. gambiae*. Species divergence, and the evolution of reproductive isolation among populations within a species, can be mediated by several mechanisms. One such mechanism is through chromosomal inversions. A paracentric chromosome inversion is the result of chromosome breakage and repair where a chromosome fragment that does not include the centromere, is re-inserted in reverse orientation, resulting in an inverted gene orientation. This type of inversion can have both structural effects (e.g. alteration of gene expression, chromosome anomalies generated during meiosis) and/or genetic effects (e.g. reduced recombination). Reduced recombination associated with paracentric inversion heterozygotes can result in an accumulation in local populations of alleles that confer a fitness advantages or that are deleterious. This occurs because inversions protect gene regions in diverging groups from the homogenizing effects of gene flow. The subspecific chromosomal forms of *A. gambiae* are well-defined by the arrangement of five paracentric inversions on the right arm of chromosome 2 (2Rj, b, c, d and u) and one on the left arm of chromosome 2 (2La) [16].

The genome of the red flour beetle, *Tribolium castaneum*, has recently been sequenced and is currently being annotated. *Tribolium* has enjoyed a long history as a model for population genetics, and the recent development of genetic and genomic tools has contributed to its current status as a powerful genetic model organism for studies in pest biology as well as comparative studies in developmental biology. It is the first coleopteran genome to be sequenced [17].

Inversions are recognised to suppress recombination in heterozygotes, as a consequence of which the genes existing within the chromosomes leak out in establishing new allelic combinations. There is also increasing evidence that chromosomal inversions may facilitate speciation event by increasing genetic variation, species specific adaptation and ultimately leading to reproductive isolation. Therefore, chromosomal inversions could be used as an evolutionary marker in studying population genetics such as gene flow, demographic history and in establishing the phylogenetic relationship among *Drosophila* species [18].

The major focus of this study to explore more on the intron –exon size variation along different regions of X-chromosome and autosomes with respect to low and high recombination in *Drosophila* and its closely relatives to understand more on the intron-exon variation and the underlying mechanism which may provide some clue in understanding gene evolution.

III. Databases and Tools

- A. **Flybase-** FlyBase is a chief depository of molecular and genetic data of *Drosophilidae* insect family. A broad range of data is present for *Drosophila melanogaster*, it being the most widely studied species. The types of data available include mutant phenotypes, cytological maps, anatomical images, molecular characterization of mutant alleles, wild-type expression patterns, transgenic insertions and constructs, sequence-level gene models and molecular classification of protein functions. Information for other *Drosophila* species is growing and is expected to increase drastically in near future with the completion of genome sequences for other eleven species of *Drosophila* [19].
- B. **VectorBase-** VectorBase (<https://www.vectorbase.org>) is a Bioinformatics Resource Center (BRC) which is focused on invertebrate vectors of human disease. VectorBase is one of the Bioinformatics Resource Centers funded by NIAID to provide web-based resources to the scientific community conducting basic and applied research on organisms considered probable agents of bioterrorism or bio warfare or causing evolving or recurring diseases. This project has been funded with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services. As a requirement of this funding all VectorBase resources are freely available to the research community under the GNU General Public License. On this site information can be accessed about specific organisms, also Data resources for those organisms can be accessed, and can use a variety of tools to analyze these data sets.
- C. **BeetleBase-** BeetleBase (<http://beetlebase.org>) is a comprehensive sequence database and significant community resource for *Tribolium* genetics, genomics and developmental biology. Beetlebase house genomic sequence scaffolds mapped to 10 linkage groups, genetic linkage maps, the official gene set, Reference Sequences from NCBI (RefSeq), predicted gene models, ESTs and whole-genome tiling array data representing several developmental stages. The existing version of Beetlebase is made on the *Tribolium castaneum* 3.0 Assembly (Tcas 3.0) released by the Human Genome Sequencing Centre at the Baylor College of Medicine.
- D. **NCBI-** The National Centre for Biotechnology Information (NCBI) is assigned to create automated systems to store and analyze information on biochemistry, genetics and molecular biology, enabling the use of the databases and software. It gathers biotechnology information internationally and performs research in various methods of computer-based information processing and analyzing the function and structure of biologically significant molecules [20].
- E. **Mapviewer-** Mapviewer provides ability to browse maps and sequences assembled for a set of organisms. One can search an organism's complete genomic sequences, map views, and zoom for more details for a particular region of interest [21].
- F. **Genscan-** Genscan is a bioinformatics tool the main function of which is to acquire a DNA sequence and find open reading frames that corresponds to the genes. The program is also used to determine a specific sequence using percentage C+G content, identify disease severity and predict gene structure to analyze its function, expression level, disease and mutation [22].

IV. Conclusion

Comparative genomics studies among different species may help to understand the underlying evolutionary forces bringing these changes in both exons and introns in chromosome-specific and lineage-specific manner. In Eukaryotes, the mechanism behind the architectural changes of introns and exons are still unclear as they are regulated by several factors e.g. transposable elements, RNA genes, deletion and recombination. Comparative genomics tool provides a platform for studying gene and protein evolution among organisms. By comparing the genomes of different organisms, one can better understand the structure and function of genes and their evolutionary

changes which may help to develop new strategies to implement in human welfare related to agriculture and health (Shukla et al. 2014).

Recombination and chromosomal inversions are increasingly being recognized as important in influencing patterns of genetic variation and adaptation. But significant amount of structural variation have been reported in genes while present within the inversion, demanding more research to understand the evolution of genes at individual level. Also intronization (transformation into intron sequences of previously exonic sequences) to generate new intron-containing genes are found to occur at a faster rate, may help speciation event to occur.

V. References

- [1] D. Charlesworth, B. Charlesworth and G Marais, "Steps in the evolution of heteromorphic sex chromosomes," *Heredity*, 2005, 95: 118-128.
- [2] M. Proschel, Z. Zhang and J. Parsch, "Widespread Adaptive Evolution of *Drosophila* Genes with Sex-Biased Expression," *Genetics Society of America*, 2006, 174: 893-900.
- [3] Z. Zhang, T. M. Hambuch, and J. Parsch, "Molecular Evolution of Sex-Biased Genes in *Drosophila*" *Molecular Biology and Evolution*, 2004, 21: 2130-2139.
- [4] M. Parisi, R. Nuttall, D. Naiman, G. Bouffard, J. Malley, J. Andrews, S. Eastman and B. Oliver, "Paucity of Genes on the *Drosophila* X Chromosome Showing Male-Biased Expression," *Science*, 2003, 299: 697-700.
- [5] J.S. Mattick, "Introns: evolution and function," *Current Opinion in Genetics & Development*, 1995, 4: 823-831.
- [6] Sh. A. Atambayeva, V. A. Khailenko, and A. T. Ivashchenko, "Intron and Exon Length Variation in *Arabidopsis*," *Rice, Nematode, and Human. Molecular Biology*, 2008, 42(2): 312-320.
- [7] L. Zhu, Y. Zhang, W. Zhang, S. Yang, J. Chen and D. Tian, "Patterns of exon-intron architecture variation of genes in eukaryotic genomes," *BMC Genomics*, 2009, 10(47): 1471-2164.
- [8] T. E. Koralewski and K. V. Krutovsky, "Evolution of Exon-Intron Structure and Alternative Splicing," *Plos one*, 2011, 6(3): 1-10.
- [9] A. F. Palazzo and T. R. Gregory, "The Case for Junk DNA," *PLOS Genetics*, 2014, 10(5): 1-8.
- [10] D. C. Presgraves, "Intron Length Evolution in *Drosophila*," *Molecular Biology and Evolution*, 2006, 23(11): 2203-2213.
- [11] H. Li, G. Liu and X. Xia, "Correlations between recombination rate and intron distributions along chromosomes of *C. elegans*," *Progress in Natural Science*, 2009, 19(4): 517-522.
- [12] G. Clark, et al., "Evolution of genes and genomes on the *Drosophila* phylogeny," *Nature*, 2007, 450: 203-218.
- [13] Q. Zhou, D. Bachtrog, "Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in *Drosophila*," *Science*, 2012, 6092: 341-345.
- [14] M.W. Young and B. H. Judd, "Nonessential Sequences, Genes and the Polytene Chromosome Bands of *Drosophila melanogaster*," *Genetics*, 1978, 88(4): 723-742.
- [15] U. Pandey and C. D. Nichols, "Human Disease Models in *Drosophila melanogaster* and the Role of the Fly in Therapeutic Drug Discovery," *Pharmacological Reviews*, 2011, 63(2): 411-436.
- [16] Y. Lee, T. C. Collier, M. R. Sanford, C. D. Marsden, A. Fofana, A. J. Cornel and G. C. Lanzaro, "Chromosome Inversions, Genomic Differentiation and Speciation in the African Malaria Mosquito *Anopheles gambiae*," *Plos one*, 2013, 8(3): 1-12.
- [17] S. Wang, M. D. Lorenzen, R. W. Beeman and S. J. Brown, "Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome," *Genome Biology*, 2008, 9(3): R61.
- [18] N. Gulati and S. Mohanty, "Chromosomal Inversion Polymorphism in Studying Phylogenetics," *Indo Global Journal of Pharmaceutical Sciences*, 2014, 4(3): 172.
- [19] R. A. Drysdale, M. A. Crosby and The FlyBase Consortium, "FlyBase: genes and gene models," *Nucleic Acids Research*, 2005, 33: 390-395.
- [20] Gene Help, National Center for Biotechnology Information, url: <http://www.ncbi.nlm.nih.gov>.
- [21] NCBI Map Viewer, National Center for Biotechnology Information, url: <http://www.ncbi.nlm.nih.gov/mapview>.
- [22] C. Burge GenScan: Bioinformatics Software for Structure Prediction and Analysis of Gene, url: <http://genes.mit.edu/GENSCAN.html>.
- [23] S. Shukla, G. Chandrabose, D. Srivastava and S. Mohanty, "Analysing Intron and ExonSize Variation along X-Chromosome of *Drosophila melanogaster*," *International Journal of basic and Applied Biology*, 2014, 2(3): 169-172.