Identification of mirror repeats within the *fibroblast growth factor*8 gene of *Xenopus tropicalis.*

DEEPTI YADAV, MAMTA DHANKHAR, VIKASH BHARDWAJ*

Department of Zoology, Baba Mastnath University, Asthal Bohar, Rohtak-124021, India

ABSTRACT

The *Xenopus* genus of African frogs, commonly known as the African clawed frog, comprises more than 20 species of frogs with varying polyploidy levels. *Xenopus* has unique features such as the absence of tongue, visible ear, and vocal sac. *X. tropicalis* is a suitable model for genetic studies because of its smaller genome size. The genome of *X. tropicalis* consists of various repetitive sequences like tandem repeats, interspersed repeats, transposons, and retrotransposons that can be utilized to understand their importance in genome function and organization. One such repeat which is not well studied in *X. tropicalis* is mirror repeat. The present study aims to determine the mirror repeats within the *fibroblast growth factor*8 gene of *X. tropicalis*. A simple manual bioinformatics approach was applied to identify the mirror repeats in the five exons and different regions of the *fgf*8 gene. We found 245 mirror repeats of significant length in the different regions of *fgf*8 gene and these mirror repeats are not only limited to the *fgf*8 gene of *X. tropicalis* but are also well distributed in the genome of *X. tropicalis*, *X. laevis, Caenorhabditis elegans* and *Drosophila melanogaster.* Further studies will help us determine the nature and function of these mirror repeats in the genome.

Keywords: Fibroblast growth factor, Mirror repeat, Repetitive sequences, *Xenopus tropicalis.*

DOI Number: 10.14704/nq.2022.20.9.NQ44159 **Neuro Quantology 2022; 20(9):1381-1391**

INTRODUCTION

DNA molecule serves as a central reservoir for genetic information. Various cellular functions are controlled by genetic information stored within the DNA. Apart from the natural B-DNA form, DNA may have multiple forms, which may evolve due to mutation, strand slippage, and unequal crossing over, resulting in the formation of repeated sequences in the genome (Gurusaran 2013; Jain et al. 2008; Zattera et al. 2020). Repetitive sequences are the DNA fragments present in multiple copies in the genome and are considered to be responsible for generating significant difference in the genomes of different organisms. They play an essential role in the regulation of cell cycle, gene expression, chromosome structure determination, karyotypic evolution and are also involved in chromosomal rearrangement events such as deletions, duplications, inversions, etc. (Jurka et al. 2007; Mehrotra and Goyal 2014; Zattera et al. 2020). Based on their distribution, repetitive sequences can be classified into two groups i.e. tandem repeats and interspersed repeats (Gurusaran 2013; Jurka et al. 2007; Pathak and Ali 2012; Zattera et al. 2020). Mini and micro-satellites are the most common tandem repeats lying adjacent to each other in a head-to-tail manner, while interspersed repeats are randomly arranged and dispersed throughout the various genome, including long and short interspersed elements (Gurusaran 2013; Pathak and Ali 2012; Zattera et al. 2020). DNA repeats may be classified into direct, inverted, everted, and mirror repeats on the basis of their position and symmetry. Direct repeats are the sequences that are precisely repeated multiple times on the same strand of DNA. Inverted repeats are arranged in reverse direction on the other strand (3'-5') of the DNA molecule. The DNA fragment with exactly duplicated sequences on the opposite strand is called everted repeats. Mirror repeats are the DNA fragments having similar base-pair arrangements around a center of symmetry (Bhardwaj et al. 2013; Mirkin 2001). For example, like in the given sequence-ATGGCTTCGGTA, one part (ATGGCT) share homology with other part of the same sequence. The mirror repeat sequences can

form non-canonical DNA structures accountable for causing various diseases like Fragile X syndrome, Huntington's disease. It is also found that homopurine and homopyrimidine mirror repeats are responsible for H- DNA and triplex DNA formation (Mirkin et al. 1987; Belotserkovskii et al. 1990; Scroth and Ho 1995; Pabis, 2021).

Xenopus tropicalis has a smaller genome size of 1.7 Gbps arranged on ten chromosomes. It has been used as a suitable model organism for various developmental, embryological, and immunological studies because of its good regeneration potential (Grainger 2012; Khokha 2012; Phipps et al. 2020; Robert and Ohta 2009). *X. tropicalis* shows a significant level of synteny with the human genome, as revealed through whole-genome sequencing and annotation analysis (Edholm and Robert 2013). The genome of *X. tropicalis* encodes over 20,000 protein-coding genes, having orthologs of around 1,700 human disease genes and consist of various repetitive sequences like tandem repeats, interspersed repeats, transposons, retrotransposons, etc. (Hellsten et al. 2010). Mirror repeats are abundantly present in the eukaryotic genome (Mehrotra and Goyal 2014) but are not extensively studied in the *X. tropicalis.* In this study we have identified mirror repeats within the *fibroblast growth factor*8 (*fgf*8) gene of *X. tropicalis* using a simple manual bioinformatics approach for the very first time. It may help us to determine whether these mirror repeats regulate the mechanism of limb generation in *X. tropicalis*. Further studies, will help us to determine the nature and function of these repeats found in *Xenopus* genome.

MATERIALS AND METHODS

The nucleotide sequence of *the fgf*8 gene (in FASTA format) of *X. tropicalis* (Gene Id-493525) was retrieved from the nucleotide database of the National Center for Biotechnology Information (NCBI). The 9,699 base pairs long *fgf*8 gene was divided into different regions of 500 base pairs each. The sequence obtained represents the query sequence. The parallel complement of the coding sequences of the *fgf*8 gene and its exons were generated using the Reverse Complement tool (https://www.bioinformatics.org/sms/rev_comp.html). The parallel complement depicts the subject sequence. Using BLAST tool, both query sequence and subject sequence were aligned for homology search. The program parameters were fixed for a word size limit of 7, and hits were observed at different expected threshold values (E-value). The threshold value at which maximum numbers of hits were observed was used to identify mirror repeats, as shown in Table 1. On alignments, where the position number gets reversed in the subject and query sequence, were identified as mirror repeat. Classification of mirror repeats was done on the basis of the spacer element present between the subject and query sequence. Using the MegaBlast tool, the mirror repeats were searched among the genome of *X. tropicalis, Caenorhabditis elegans, Xenopus laevis,* and *Drosophila melanogaster*. Algorithm parameters like word size expected threshold and maximum target sequence were adjusted accordingly during the search.

RESULTS AND DISCUSSION

Fibroblast growth factor (FGF) belongs to the largest growth factor family that controls the proliferation of cells involved in the regeneration and growth of soft tissues. Of the 20 members of FGFs; various kinds such as FGF-2,-4,-8, and -10 play a vital signaling role in vertebrate limb development. Two signaling centers, the apical ectodermal ridge and posterior zone of polarizing activity, help the limb buds to develop into complex structure called limbs. The *fgf*8 gene is expressed in the apical ectodermal ridge in developing limbs and plays a crucial role in initiating and maintaining limb outgrowth. During gastrulation, the *fgf*8 gene is expressed in the posterior region and contributes in the patterning of the anteroposterior axis of the limb bud in *Xenopus* (Christen and Slack 1997; Han et al. 2001; Keenan and Beck 2016, Phipps et al. 2020).

In the present study, a simple manual bioinformatics approach was applied initially to identify mirror repeats present in the *fgf*8 gene of *X. tropicalis*. The maximum number of hits was observed at E- value 100, so we identified mirror repeats at this value. The mirror repeats were distributed in all the regions of the *fgf*8 gene irrespective of their size. In total, 245 mirror repeats (Fig. 2) were

found in different regions of the *fgf*8 gene and out of which 65 mirror repeats were distributed among five exons of the gene. The exons are highlighted with different colors in figure 2; the star represents the relative position of mirror repeats within the gene. Exon 5, being the longest, contains thirty-one mirror repeats, whereas exon 2 is the smallest, contains only one mirror repeat. Mirror sequences with sizes ranging from 7-12 bps (Table 2) were abundantly present in the gene, and the frequency of smaller mirror repeats was found to be more than the larger mirror repeats (Table 3). We further selected those mirror repeats (≥10bps) that can adopt stable non-canonical DNA structures by using published experimental data (Hamada et al. 1982; Haniford and Pulleyblank 1983; Schroth and Ho, 1995). According to the data, total 32 homopurine (polydA-dG) – homopyrimidine (poly dC-dT) rich sequences were found within the *fgf*8 gene and two mirror repeats may form stable H-DNA structures. The gene also contains 46 poly (dT-dA) repeats, out of which, two may form cruciform structures while 17, mirror repeats with poly (dC-dA/dT-dG), may adopt Z-DNA conformations.

Our results revealed the presence of different types of mirror repeats among the different regions of the *fgf*8 gene. The gene *fgf*8 contains mirror repeats of varying lengths, ranging from 7 base pairs which were smallest to (35, 50) base-pair long, medium-sized repeat, and the largest mirror repeat of 91 base-pairs. The nomenclature of identified mirror repeats was based on the number of spacer elements found at the center of symmetry. The repeats with no spacer element and having identical sequences around a central axis are termed as perfect mirror repeats. A perfect mirror may have a spacer element of any size i.e. single, double, and multiple spacers. In imperfect mirror repeats, one can observe mismatches around the central axis. The frequency of perfect mirror repeats within the different regions of the *fgf*8 gene was more than the imperfect mirror repeats. Out of 245 mirror repeats, 45 perfect mirror repeats, 156 perfect mirror repeats with single spacer, and 44 imperfect mirror repeats were observed. The location and classification of selected mirror repeats in the different regions of the *fgf*8 gene were depicted in table 4. The complete detail of all mirror repeats present within the *fgf*8 gene is provided in the supplementary file.

Using the MegaBlast tool, the mirror repeats were searched among the genome of *X. tropicalis, C.elegans, X. laevis,* and *D. melanogaster*. The result confirms the universal existence of mirror repeat sequences of sizes 10-15 base pairs (Table 5) in all the organisms mentioned above. These sequences must have some vital role in the genome since they are conserved in all domains of life. The complete MegaBlast result of all mirror repeats is provided in the supplementary file.

CONCLUSION

In this study, we found that the *fgf*8 gene of *X. tropicalis* is enriched in mirror repeats. The gene consists of mirror repeats of various length and type. We have identified sequences that can form non-canonical B-DNA forms within the gene. These mirror repeats may participate in various activities occurring at cellular and genomic level and to identify this further study will help us determine whether they will play an essential role in regulating cellular metabolism, gene function, etc.

Funding

This research received no external funding.

Conflicts of interest

Authors don't have any conflict of interest.

Credit author statement

Vikash Bhardwaj: Conceptualization, Methodology, Software, Supervision, Reviewing, Editing

Deepti Yadav: Data Curation, Investigation, Writing–Original draft preparation

Mamta Dhankhar: Writing- Review and Editing.

References

Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M. and

Frank-Kamenetskii, M.D. (1990). Formation of intramolecular triplex in homopurine-

homopyrimidine mirror repeats with point substitutions. Nucleic Acids Research,

18(22), 6621–6624.

Christen, B. and Slack, J.M. (1997). FGF-8 is associated with anteroposterior patterning and

limb regeneration in Xenopus. Developmental Biology. 192(2):455-466.

Edholm, E.S. and Robert, J.(2013). Recent research progress and potential uses of the amphibian Xenopus as a biomedical and immunological model system. Resources. 2(3):167-183.

Grainger, R.M. (2012). *Xenopus tropicalis* as a model organism for genetics and genomics: past, present and future.Methods in Molecular Biology. 917:3-15.

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

Hamada, H. Petrino, M.G., Kakunga, T. (1982). A novel repeated element with Z-DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. Proceedings of National Academy of Sciences USA.79(21):6465-6469.

Han, M.J., An, J.Y., Kim, W.S. (2001). Expression patterns of *Fgf*-8 during development and limb regeneration of the axolotl.Developmental Dynamics. 220(1):40-48.

Haniford, D.B., Pulleyblank, D.E. (1983). Facile transition of poly [d (TG). d (CA)] into a left-handed helix in physiological conditions.Nature. 302(5909):632-634.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. Science.328:633-636.

Jain, A., Wang, G., Vasquez, K.M. (2008). DNA triple helices: biological consequences and therapeutic potential. Biochimie. 90(8):1117-11130.

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. 8:241-259.

Keenan, S.R.,Beck, C.W. (2016). *Xenopus* limb busd morphogenesis.Developmental Dynamics.245(3):233-243.

Khokha, M.K. (2012). *Xenopus* white papers and resources: folding functional genomics and genetics into the frog.Genesis.50: 133-142.

Mehrotra, S. and 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., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-

Kamenetskii, M.D. (1987). DNA H form requires a homopurine-homopyrimidine

mirror repeat. Nature, 330(6147): 495–497.

Mirkin, S.M. (2001). DNA topology: fundamentals. ELS 111.

Pabis, K. (2021). Triplex and other DNA motifs show motif- specific associations with mitochondrial DNA deletions and species lifespan.Mechanisms of Ageing and Development.194.

Pathak, D., Ali, S. (2012). Repetitive DNA: Atool to explore Animal Genomes/Transcriptomes. Functional genomics In Tech. 155-180.

Phipps, L.S., Marshall, L., Dorey, K., Amaya, E. (2020). Model systems for regeneration: *Xenopus*. Development. 147(6).

Robert, J. and Ohta, Y. (2009). Comparative and developmental study of the immune system in *Xenopus.* Developmental Dynamics. 238:1249-1270.

Scroth, G.P. and Ho, P.S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Research. 23(11):1977-1983.

Vikash, B., Swapnil, G., Sitaram, M., Kulbhushan, S. (2013). FPCB: A simple and swift strategy for mirror repeat identification.arXiv preprint arXiv: 1312.3869.

Wang, G. and Vasquez, K.M. (2004). Naturally occurring H- DNA- forming sequences are mutagenic in mammalian cells. Proceedings of National Academy of Science USA. 101(37):13448-13453.

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

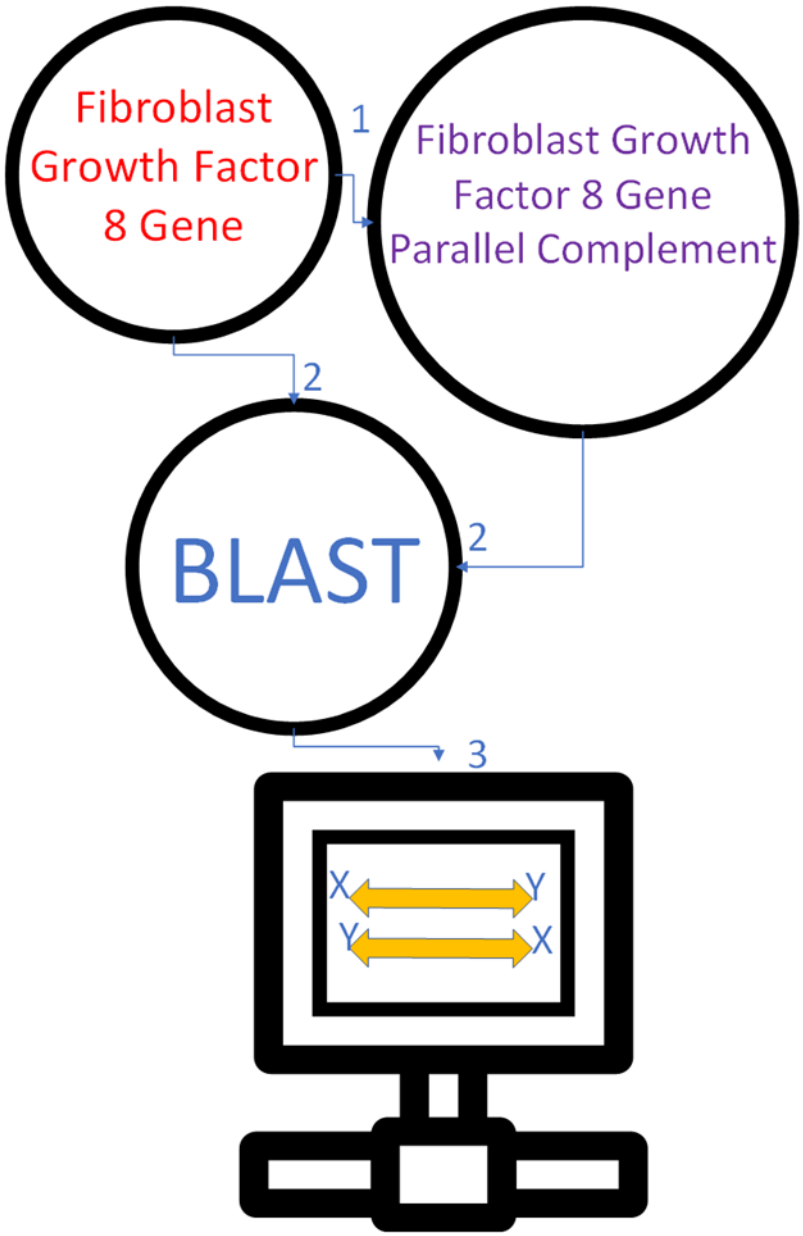

Fig.1: Pictorial representation of methodology used to characterize mirror repeats using FPCB approach involving (i) Retrieved the original sequence from NCBI (ii) Making its parallel complement using Reverse Complement Tool (iii) Analysis of BLAST result for identification of mirror repeats.

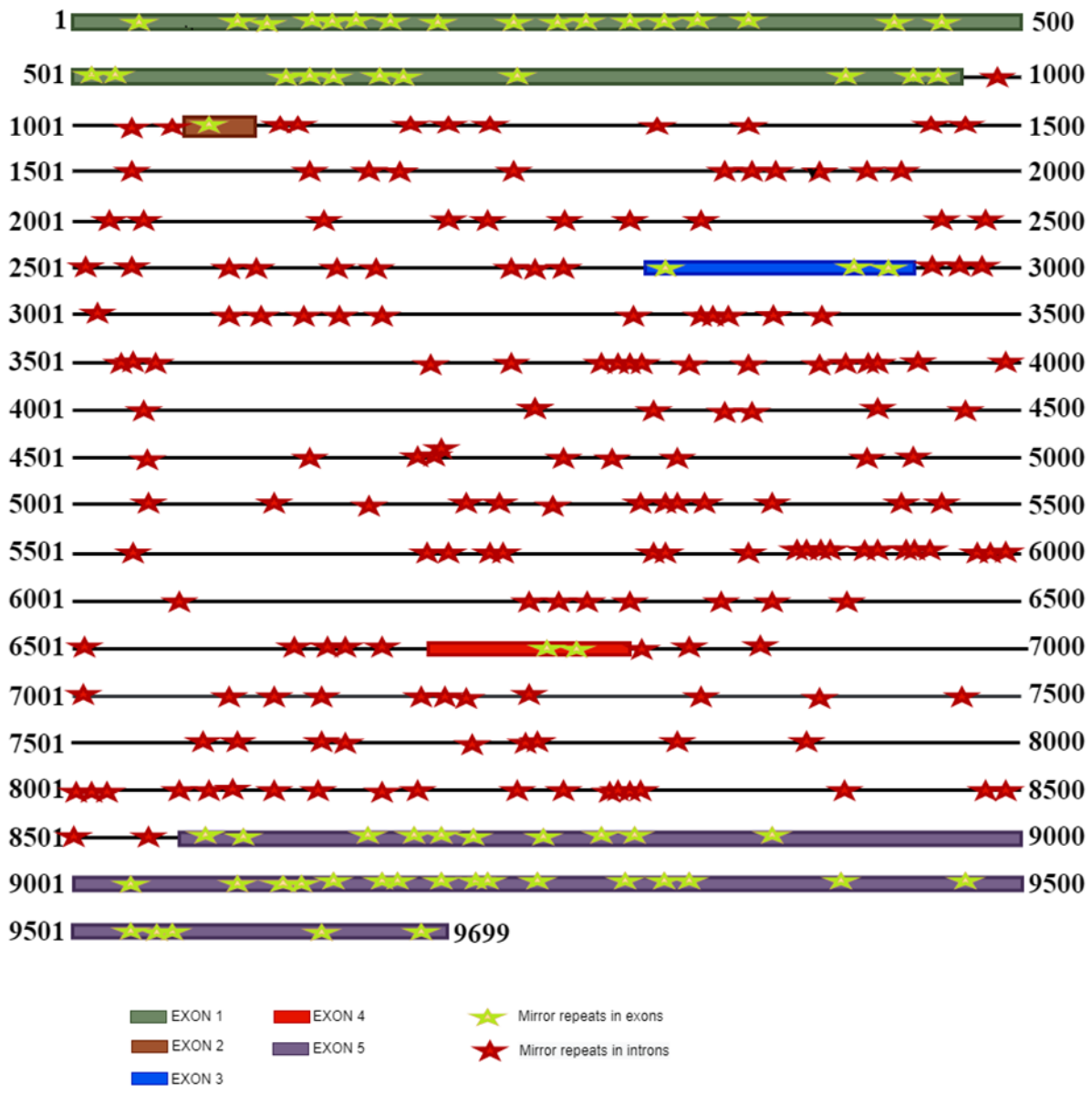

Fig 2: Diagrammatic representation of relative position of mirror repeats within the *fgf*8 gene of *Xenopus tropicalis.*

Table 1: Total number of hits observed in exons of *fgf*8 gene at expected threshold value 100 and word size 7. Maximum number of hits observed at E-Value 100 in all the exons.

Table 2: Representing the length and location of mirror repeats present in different regions of the *fgf*8 gene.

Exons (bps)	Length of exons (bps)	Expected threshold value	Number of hits	Number of mirror repeats
Exon 1 (1-973)	973	100	103	28
Exon 2 (1061-1097)	36	100	1	1
Exon 3 (2802-2942)	148	100	7	3
Exon 4 (6683-6789)	107	100	4	2
Exon 5 (8565-9699)	1135	100	186	31

Symbol/Gene ID	Gene Regions(bps)	7-12bp	13-18bp	19-24bp	>25bp	Total MR's
*fgf*8/ 493525	1-500	13	4	0	0	17
	501-1000	9	2	1	0	12
	1001-1500	8	3	1	0	12
	1501-2000	6	2	1	2	11
	2001-2500	9	1	0	0	10
	2501-3000	11	2	0	0	13
	3001-3500	11	0	0	1	12
	3501-4000	16	1	0	0	17
	4001-4500	7	1	0	0	8
	4501-5000	8	2	0	0	10
	5001-5500	12	0	1	0	13
	5501-6000	15	3	1	1	20
	6001-6500	5	2	1	0	8
	6501-7000	7	1	1	1	10
	7001-7500	9	2	0	0	11

	7501-8000	8	1	0	0	9
	8001-8500	18	1	0	0	19
	8501-9000	9	2	1	0	12
	9001-9500	13	3	0	0	16
	9501-9699	2	1	0	2	5
Total		**196**	**34**	**8**	**7**	**245**

Table 3: Table showing the mirror repeats that are repeated more than once within the *fgf*8 gene of *Xenopus tropicalis.*

Sequence	Query position		Repeat copy	Query position	
	From	To		From	To
TTCACTT	96	102	TTCACTT	1176	1182
CTGCGTC	669	675	CTGCGTC	8585	8591
CTACATC	947	953	CTACATC	6750	6756
ATAAATA	2264	2270	ATAAATA	9188	9194
TAATAAT	1459	1465	TAATAAT	9548	9554
ATTTTTA	1051	1057	ATTTTTA	1107	1113
				9312	9318
AAAGAAA	5112	5118	AAAGAAA	9032	9018
AAGGGAA	3992	3998	AAGGGAA	6270	6276
				8714	8720
TTTATTT	3913	3919	TTTATTT	5980	5986
TATAAATAT	2732	2740	TATAAATAT	3851	3859
AAATAAA	5813	5819	AAATAAA	7744	7750
ACGTGCA	7582	7588	ACGTGCA	9473	9479
CTTTTTC	3799	3805	CTTTTTC	5683	5689
CACCCAC	954	960	ACCTCCA	1839	2483
ACCTCCA	994	1000	CACCCAC	2477	1845
TTTCCTTT	2662	2669	TTTCCTTT	5161	5168
CTGTCTGTC	5437	5445	CTGTCTGTC	5938	5946
TGGGGGT	3544	3550	TGGGGGT	6503	6509
GGAGAGG	1728	1734	GGAGAGG	2536	2542

Table 4: Classification of selected Mirror Repeats distributed in the complete *fgf*8 gene of *Xenopus tropicalis*.

Symbol /Gene ID	Gene regions (bps)	Location of mirror repeats			Length (bps)	Type of mirror repeats
		Query Start Site	Sequence	Query End Site		
*fgf*8/ 493525	1-500 bps	315	AGCGAGAGAGCGA	327	13	Perfect mirror with single spacer
		319	AGAGAGCGAGAGA	331	13	Perfect mirror with single spacer
	501-1000 bps	622	TATCATTACTAT	633	12	Perfect mirror
		629	ACTATTATTGTTATTATCA	647	19	Perfect mirror with single spacer
		731	GTGGCTCGGTG	741	11	Perfect mirror with single spacer
		904	CGGGCAGCCCGACAGGC	920	17	Imperfect mirror
	1001-1500 bps	1314	TAGGGGTCACTGGGTAT	1330	17	Imperfect mirror
	1501-2000 bps	1881	CGAGGGGAGAGGGAAGC	1897	17	Imperfect mirror
	2501-3000 bps	2751	TTTATATATTT	2761	11	Perfect mirror with single spacer
		2927	GGCAGAAGACGG	2938	12	Perfect mirror
	3001-3500 bps	3019	TCCAGGGAGCAGAATGTGTGCAACAGGTGGGACCT	3053	35	Imperfect mirror repeat
		3098	GAAGGTGGAAG	3108	11	Perfect mirror with single spacer
	4001-4500 bps	4344	CCAGAGAGACC	4354	11	Perfect mirror with single spacer
		4474	TGCATCTACGT	4484	11	Perfect mirror with single spacer
	5501-6000 bps	5531	TACAGGGGACAT	5542	12	Perfect mirror
		5886	TCTGTCTGTCT	5896	11	Perfect mirror with single spacer
	6001-6500 bps	6246	TAAGACAGAAT	6256	11	Perfect mirror with single spacer
		6293	CTTTAAATTTC	6303	11	Perfect mirror with single spacer
	6501-7000	6631	ATTCCCCCTTA	6641	11	Perfect mirror with single spacer

	bps	6865	aaaaaaaaaaCgggggggAACACAGGGAAGTGAC-ATTGTAAAAACAAATTTTTGATCA-AAAGGCAGACAAGAGGTGGgaaaggaaaaa	6952	91	Imperfect mirror repeat
	8001-8500 bps	8107	CATTTATTTAC	8117	11	Perfect mirror with single spacer
		8411	AACCAATAAATACCCAA	8427	17	Imperfect mirror
	8501-9000 bps	8749	GACACCACACCACAG	8763	15	Perfect mirror with single spacer
	9001-9699 bps	9321	AGCTGCTTTTTCATCGA	9337	17	Imperfect mirror repeat
		9549	AATAATATTGTTATTATTTTTAAAAAAA	9576	28	Imperfect mirror repeat
		9630	TATGTGTACATATTAATATTAATGATGGAAATATTTATT-TAAATGTGTAT	9679	51	Imperfect mirror repeat

Table 5: **Distribution of selected mirror repeats among different genera's, here + sign shows presence of mirror and – sign shows the absence of mirror.**

S.No.	Mirror sequences	*X. tropicalis*	*X. laevis*	*C. elegans*	*D. melanogaster*
1	AGCGAGAGAGCGA	+	+	+	+
2	AGAGAGCGAGAGA	+	+	+	+
3	ACTATTATTGTTATTATCA	+	-	-	+
4	GTGGCTCGGTG	+	+	+	+
5	TCCAGGGAGCAGAATGTGTGCAACAGGTGGGACCT	+	-	-	-
6	AAAAGAAGAAAA	+	+	+	+
7	AGCTGCTTTTTCATCGA	+	-	+	-
8	TATCATTACTAT	+	+	+	+

9	CGGGCAGCCCGACAGGC	+	-	-	-
10	ATTCCCCCTTA	+	+	+	+
11	AATTATAATATCAA	+	+	+	+
12	GGTAGATATAGCTGG	+	-	-	+
13	CAGTTTTGAC	+	+	+	-
14	CAGATATTTATAAAC	+	-	+	+
15	TCACTTCACT	+	+	+	-