Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 11 [12] November 2022 : 01-05

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD

ORIGINAL ARTICLE **OPEN ACCESS**

Identification of Mirror Repeats within the *BRCA2* Gene Using Simple Manual Bioinformatics Approach

Madhuri[1], Bhargava Anu[2], Dangi Namrata[3], Bhardwaj Vikash [4]*
[1,2.3.4*]Department of Zoology, Baba Mastnath University, Asthal Bohar Rohtak 124021, India
***Corresponding Author:** Bhardwaj Vikash
Email: vikashbhardwaj@gmail.com

ABSTRACT
Large portion of the genome of eukaryotes is formed of DNA repeat sequences. The heterochromatin region of the chromosome contains an accumulation of these sequences. DNA repeats can be divided into direct, inverted, everted, and mirror repeats depending on their symmetry and position. Mirror repeat sequence is a repetitive DNA sequence that has not been studied in Homo sapiens. This work explores mirror DNA sequences in the exons of BRCA2 (Breast Cancer gene2) using a simple manual bioinformatics approach. BRCA2 gene act as cell growth suppressor and mutations in this gene can lead to breast cancer. We have identified 170 mirror repeats within exons of BRCA2 gene. Highest density of mirror repeats is present in exon 11. Largest mirror repeat of length 46 bps is present in exon 11 and it is an IMR (Imperfect Mirror Repeat). To ascertain the function of these mirror repetitions in breast cancer, additional research will be needed.
Keywords: *BRCA 2, IMR, Mirror repeat, Repetitive DNA sequence.*

Received 06.10.2022 Revised 12.10.2022 Accepted 08.11.2022

INTRODUCTION

Repetitive sequences are important constituent of eukaryotic genome and may comprise up to 90% of the genome [8]. These are genomic fragments that repeats itself throughout the genome and are accumulated in the heterochromatin region of the chromosome [7, 11]. These sequences evolve by concerted evolution and play an important role in maintaining genomic stability, controlling expression of genes, DNA packaging etc. [11].

Repetitive DNA sequences replicate within the genome, cause mutations and their expansion results in genetic diseases including huntington disease, fragile X syndrome, frontotemporal dementia (FTD) and myotonic dystrophy [1, 14].

On the basis of their distribution, repetitive sequences can be classified as tandem repeats and interspersed repeats [3, 17]. Repetitive DNA sequences restricted at a specific location in the genome are called as tandem repeats, these comprise of mini and microsatellites whereas interspersed repeats such as LINEs and SINEs are randomly dispersed throughout the genome [3, 11]. Depending on the symmetry and position, repetitive DNA sequences can be classified into everted repeats, inverted repeats, and direct repeats. These repetitive sequences can form different types of non-B-DNA structures [15, 2]. Mirror repeat has also been reported in several genomes. A mirror repeat is defined as a DNA sequence which has homology with the other portion of same DNA sequence within a gene or genome [9]. For instance, the sequence TCGGTAATGGCT has a homology between one component, TCGGTA, and another part, ATGGCT of the same sequence. In highly super coiled DNA, homopurine and homopyrimidine mirror repeats can form intramolecular triplexes, or H-DNA [10, 12]. These H-DNA are reported to regulate the expression of diseases associated genes [16].

Breast cancer risk is increased by germline mutations in *BRCA1* and *BRCA2* (shah *et al*., 2018). *BRCA2* gene is located on chr13q, which act as cell growth suppressor and produce TSG (Tumor Suppressor Gene) proteins. Molecular analysis on exons of *BRCA 2* gene shows that mutations in the exons of *BRCA2* may result in breast cancer [6, 7]. Our current study identified mirror DNA sequences in the exons of *BRCA2* gene, using a simple manual bioinformatics approach. It may assist in determining whether these mirror sequences control the mechanism of breast cancer or not. Additional research will be needed to understand the role of these sequences in the human genome..

MATERIAL AND METHODS

Genome sequence of *BRCA2* (in FASTA format; Gene Id-675) of *Homo sapiens* was obtained from NCBI (National Center for Biotechnology Information). *BRCA2* gene consists of 27 exons. Exon sequences obtained from NCBI represent the query sequence. Using the Reverse Complement tool (https://www.bioinformatics.org/sms/rev comp.html), the parallel complements of these exons were generated. Subject sequences are represented by these parallel compliments. The FASTA format of exons (the query sequence) and its parallel complement (the subject sequence) were both aligned for the homology search using BLAST. At various expected thresholds, hits were obtained and the settings were set at a word size of seven. Mirror repeats were identified by using the threshold value at which the most hits were obtained. Mirror DNA sequences were identified in hits when the position number of query and subject sequence is exactly reversed. These mirror sequences were classified depending on the symmetry of mirror repeats as Perfect Mirror repeat (PMR) and Imperfect Mirror Repeats (IMR). The mirror DNA sequences were searched in the genome of *Homo sapiens, Mus musculus, Danio rerio* and *D. melanogaster* using mega BLAST tool.

RESULTS AND DISCUSSION

Various genetic, environmental and acquired factors are involved in causing breast cancer. It is a multistep process, which may occur due to pathogenic mutations in *BRCA1* and *BRCA2* gene. These genes undergo nonsense, frame shift, and splicing mutations to produce truncated proteins [7, 8, 13].

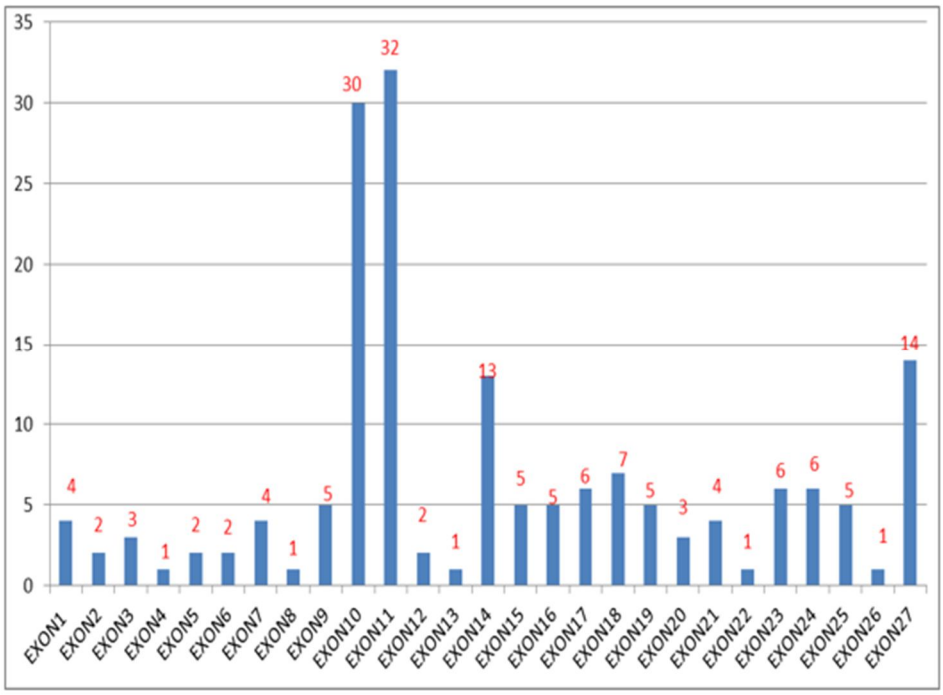

Fig.1: shows distribution of mirror repeats in different exons of *BRCA 2* gene .

Table 1: Representing the length of mirror repeats present in the different exons of the *BRCA2* gene.

Accession No.	Exon	7-12bp	13-18bp	19-24bp	25-50bp	Total MR's
NC_000013	Exon-1	4	0	0	0	4
	Exon-2	2	0	0	0	2
	Exon-3	3	0	0	0	3
	Exon-4	1	0	0	0	1
	Exon-5	1	1	0	0	2
	Exon-6	2	0	0	0	2
	Exon-7	4	0	0	0	4
	Exon-8	1	0	0	0	1
	Exon-9	3	2	0	0	5
	Exon-10	29	0	1	0	30
	Exon-11	19	4	2	7	32
	Exon-12	2	0	0	0	2
	Exon-13	1	0	0	0	1
	Exon-14	11	1	1	0	13
	Exon-15	5	0	0	0	5
	Exon-16	5	0	0	0	5
	Exon-17	6	0	0	0	6
	Exon-18	6	0	0	1	7
	Exon-19	3	2	0	0	5
	Exon-20	2	1	0	0	3
	Exon-21	3	0	1	0	4
	Exon-22	1	0	0	0	1
	Exon-23	6	0	0	0	6
	Exon-24	6	0	0	0	6
	Exon-25	5	0	0	0	5
	Exon-26	1	0	0	0	1
	Exon-27	13	0	0	1	14
	Total	145	11	5	9	170

Table 2: Classification and location of selected mirror repeats distributed in the exons of *BRCA2* gene

Exon	Mirror Repeat	Position of Mirror Start position	Stop position	Length of mirror repeat	Type Of Mirror
Exon5	ACAATGTACACATGTAACA	13	31	19	Perfect with single spacer
Exo10	AGAACAAAAGAAAGAAAGA	502	520	20	Imperfect mirror repeat
Exon 10	AAAGAAAGAAA	508	518	11	Perfect with single spacer
Exon 10	AGAAAGAAAGA	510	520	11	Perfect with single spacer
Exon-11	ACAAAAATCATCTCTCCGAAAAACA	3274	3298	25	Imperfect mirror repeat
Exon-11	GATAT-ATGTAAATGTAGTATAG	3957	3978	22	Imperfect mirror repeat
Exon-11	AAGTAAAAATGATACTGTTTGTATTC ATAAAGATGAA	2114	2150	37	Imperfect mirror repeat
Exon-11	CAACTAGTGACCTTCCAGGGA-CAAC	2595	2619	25	Imperfect mirror repeat
Exon-11	AAACGTAAAAATGGAAA	4655	4671	17	Imperfect mirror repeat
Exon-11	TAAAACGGAGCAAAAT	2327	2342	16	Imperfect mirror repeat
Exon-11	CCAAAGAGTCATTTAATAAAATTGTA AATTTCTTTGATCAGAAACC	2407	2452	46	Imperfect mirror repeat
Exon-11	TGAAAGAAAGT	2653	2573	21	Perfect with single spacer
Exon-11	AACGTAAAAATGGAAATTGGTAAAAC TGAAA	4656	4686	31	Imperfect mirror repeat
Exon-11	CCATAATTTAACACC	1391	1405	15	Imperfect mirror repeat
Exon-11	TTTGAAGGTACAGTTGAAATTAA-ACGGAAGTTT	1635	1667	33	Imperfect mirror repeat
Exon-11	ACAAAACGCAAGACA	3143	3152	15	Perfect mirror repeat
Exon-11	ATAAAGAAGCAA-AATGTAATAAGGAAAAACTA	154	185	31	Imperfect mirror repeat
Exon-11	AAAATATAAAA	2338	2348	11	Perfect with single spacer
Exon-11	AAAAATAAAAA	3552	3562	11	Perfect with single spacer
Exon - 14	GAAAACAGACAAAAG	281	295	15	Perfect with single spacer
Exon - 14	TTTAAAACTAAATCACATTTT	221	241	21	Imperfect mirror repeat
Exon - 18	ATAGAAGCAGAAGATCGGCTATAAAA --AAGATA	18	49	30	Imperfect mirror repeat
Exon - 21	ACCATCACGTGCACTAACA	23	41	19	Imperfect mirror repeat
Exon - 27	AATGAAA--- TTTCTCTTTTGGAAAGTAA	337	362	26	Imperfect mirror repeat

Table 3: Shows distribution of mirror repeats of size >13 bps in selected genera.

Sr. No.	Mirror Sequences	*Homo sapiens* (taxid:9606)	Mus *musculus* (taxid:10090)	*D.melanogas ter* (taxid:7215)	*Dani rerio* (taxid:7955)
1.	ACAATGTACACATGTAACA	+	+	+	+
2.	AGAACAAAAGAAAGAAAGA	+	+	+	+
3.	ACAAAAATCATCTCTCCGAAAAACA	+	+	+	+
4.	GATAT-ATGTAAATGTAGTATAG	+	+	+	+
5.	AAGTAAAAATGATACTGTTTGTATTCATAAAGATGAA	+	+	+	+
6.	CAACTAGTGACCTTCCAGGGA-CAAC	+	+	+	+
7.	AAACGTAAAAATGGAAA	+	+	+	+
8.	TAAAACGGAGCAAAAT	+	+	+	+
9.	CCAAAGAGTCATTTAATAAAATTGTAAAT	+	+	+	+
10.	TGAAAGAAAGT	+	+	+	+
11.	AACGTAAAAATGGAAATTGGTAAAACTGAAA	+	+	+	+
12.	CCATAATTTAACACC	+	+	+	+
13.	TTTGAAGGTACAGTTGAAATTAA-ACGGAAGTTT	+	+	+	+
14.	TTCATTACTT	+	+	+	+
15.	ACAAAACGCAAGACA	+	+	+	+
16.	ATAAAGAAGCAA-AATGTAATAAGGAAAAACTA	+	+	+	+
17.	TTATTTATT	+	+	+	+
18.	GAAAACAGACAAAAG	+	+	+	+
19.	TTTAAAACTAAATCACATTTT	+	+	+	+
20.	ATAGAAGCAGAAGATCGGCTATAAAA--AAGATA	+	+	+	+
21.	ACCATCACGTGCACTAACA	+	+	+	+
22.	AATGAAA---TTTCTCTTTTGGAAAGTAA	+	+	+	+

Here, a simple manual bioinformatic method is used to identify mirror DNA sequences within *BRCA2* gene (BReast CAncer gene2) of breast cancer. The present work identified a total of 170 mirror repeats within exons of *BRCA2.* These sequences are randomly distributed within the different exons of gene (Fig-1). The mirror sequences with shorter strech of nucleotides of size 7-12bps were most abundant whereas the larger mirror sequences were less abundant in the *BRCA2* gene (Table1). Mirror repeats having size ≥10bps can form non-canonical BDNA forms [4, 5, 12]. We found a total of 25 homopurine(polydA-dG)-homopyrimidine (poly dC-dT) sequences, of which three mirror repeats may adopt H-DNA structure (highlighted in green colour in table 2). Similarly, 15 poly (dT-dA) repeats were identified in *BRCA2* gene, out of which three can adopt cruciform structures (highlighted in blue colour in table 2).

We have identified various type of mirror DNA sequences in the exons of the *BRCA2* gene. These mirror repeats are classified depending on the presence or absence of spacer elements. Mirror sequences having one spacer are identified as Single Spacer Mirror Repeat (SSMR). Largest mirror repeat is of 46bps (present in exon 11) and smallest mirror repeat is of 7bps (present in several exons). Frequency of perfect mirror repeats were more as compared to imperfect mirror repeats within the *BRCA2* gene. Classification and location of some selected mirror repeats are represented in table 2. complete detail of identified mirror repeats is given in the supplementary file.

Non-.BDNA motif search tool was also used to identify mirror repeats within the exons of *BRCA2* gene. Using simple manual bioinformatic approach we identified 170 mirror repeats whereas Non-.BDNA motif search tool could not found mirror repeats in exons of *BRCA2* gene.

By comparing the number of mirror repeats obtained from both the tools, it can be concluded that our method is more efficient in finding mirror repeats, as non B-DNA motif search tool could not search mirror repeats in exons of *BRCA2* gene.

mega BLAST tool was used to search selected mirror repeats (having size in the range of >13 bps) in different genera's (Table-3). Sequences selected for mega BLAST analysis were present in the genome of *Homo sapiens, Mus musculus*; *Danio rerio* and *D. melanogaster*; as these sequences are conserved

throughout all domains of life, this suggests that they play an important role in the genome. However, what role will these identified mirror repeat play at cellular and genetic level in breast cancer is still unknown?

CONCLUSION

We have analysed *BRCA2* gene for identification of mirror repeats within the exons of *BRCA2* gene using two simple bioinformatics tools. BLAST tool have identified a total of 170 mirror repeats, 17 mirror repeats were identified as IMR (Imperfect Mirror Repeat), 23 as PMR (Perfect Mirror Repeat), and 130 as PWSS (Perfect With One Spacer) whereas, non B-DNA motif search tool could not identify mirror repeats within the exons of *BRCA2* gene. Further studies are required to find out the role of these mirror repeats in breast cancer.

FUNDING

This research received no external funding.

CONFLICTS OF INTEREST

Authors don't have any conflict of interest.

REFERENCES

1. Charlesworth,B, Sniegowski, P and Stephan, W. (1994).The evolutionary dynam of repetitive DNA in eukaryotes. Nature.371:215-20.
2. Guiblet, W. M., Cremona, M. A., Harris, R. S., Chen, D., Eckert, K. A., Chiaromonte, F., & amp; Makova, K. D. (2021). Non-B DNA: a major contributor to small-and large-scale variation in nucleotide substitution frequencies across the genome. Nucleic acids research, 49(3), 1497-1516.
3. Gurusaran, M, Ravella, Dand Sekar, K.2013. RepEx: repeat extractor for biological sequences. Genomics. 102(4):403-408.
4. 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.
5. 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.
6. Mehrgou ,A and Akouchekian,M.(2016). The importance of *BRCA1* and *BRCA2* genes mutations in breast cancer development.Med J Islam Repub Iran. Vol. 30:369.
7. 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.
8. Michelle, W. and Wong,B.2015 Prevalence of BRCA1 and BRCA2 germline mutations in patients with triple-negative breast cancer.Breast Cancer Research and Treatment. DOI 10.1007/s10549- 015-3293-7
9. Mirkin, S.M.(2001). DNA topology: fundamentals. Encyclopedia of Life Sciences.111.Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci. 2017 Nov 1;13(11):1387-1397. doi:10.7150/ijbs.21635. PMID: 29209143; PMCID: PMC5715522.
10. 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.
11. Paço, A., Freitas, R., & Vieira-da-Silva, A. (2019). Conversion of DNA sequences: From a transposable element to a tandem repeat or to a gene. Genes, 10(12), 1014.
12. Schroth, G.P., Ho, P.S., 1995. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA.Nucleic Acids Research. 23(11), 1977–1983.
13. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. (2017) .Risk Factors and Preventions of Breast Cancer. Int J Biol Sci.;13(11):1387-1397.
14. Usdin, K., House, N. C. M., & Freudenreich, C. H. (2015). Repeat instability during DNA repair: Insights from model systems. Critical Reviews in Biochemistry and Molecular Biology, 50(2), 142– 167.
15. Wang G, Vasquez KM. Effects of replication and transcription on DNA structure-related genetic instability. Genes. 2017; 8(1), 17. Available from: 10.3390/genes8010017.
16. Wang, G& Vasquez, K. M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America, 101(37), 13448–13453
17. 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 Xenopus tropicalis (Anura, Pipidae).Frontiers in genetics. 11,637.

CITATION OF THIS ARTICLE

Madhuri, B Anu, D Namrata, B Vikash. Identification of Mirror Repeats within The *BRCA2* Gene Using Simple Manual Bioinformatics Approach**.** Bull. Env.Pharmacol. Life Sci., Vol 11 [12] November 2022: 01-05