Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 11 [12] November 2022 : 12-16

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

ORIGINAL ARTICLE **OPEN ACCESS**

Identification of mirror repeats within the *BRCA1* gene of breast cancer

Madhuri[1], Bhargava Anu[2] and Bhardwaj Vikash[3]*
[1,2,*3]Department of Zoology
Baba Mastnath University, Asthal Bohar Rohtak 124021, India
***Corresponding author: -** Bhardwaj Vikash
E mail: vikashbhardwaj@gmail.com

ABSTRACT
About 50% of the human genome is repetitive or repeat derived. Eukaryotic genomes contain various types of repetitive sequences, including transposons, retro transposons, interspersed repeats and tandem repeats. Repetitive sequence that is not well studied in Homo sapiens as well as in other organisms is mirror repeat sequence. DNA mirror repeat is a segment identified on the basis of center of symmetry on a single strand and has identical terminal nucleotides. This work explores mirror DNA sequences in the exons of BRCA1 (BReast CAncer gene1) using a simple manual bioinformatics method. BRCA1 is an autosomal dominant gene located on chr17q, which codes for TSG (Tumor Suppressor Gene) proteins, there is high probability of developing breast cancer if this gene is altered or mutated in any way. We have identified 141 mirror repeats within the exons of BRCA1 gene. These mirror repetitions are not only found in the BRCA1 gene but are also widely spread in the genomes of Homo sapiens, Mus musculus, Danio rerio, and D.melanogaster. To find the role of these mirror repeats, additional research will be needed.
Keywords: *BRCA 1, Mirror repeat, Repetitive DNA sequences, TSG.*

Received 09.10.2022 Revised 12.10.2022 Accepted 08.11.2022

INTRODUCTION

Repetitive sequences are genomic fragments that repeats itself throughout the genome [1]. These sequences evolve by concerted evolution and results in changes within the genome [11]. Large portion of the genome of eukaryotes is formed of repetitive DNA sequences. The behavior of repetitive sequences can result in mutations that cause genetic diseases and more than 30 inherited genetic diseases, including Huntington disease and myotonic dystrophy, are caused by the proliferation of these sequences. [1, 12].
These sequences comprise approximately 50% of the human genome [10]. Depending on their distribution, repetitive sequences can be categorized as interspersed repeats and tandem repeats [2, 14]. In contrast to interspersed repeats, which are randomly distributed across the genome, tandem repeats are next to one another. [2, 14] Repetitive DNA sequences create deletion mutations or insertion mutations if they are inserted into a gene [6]. DNA repeats can be divided into four categories based on their symmetry and position: direct repeats, everted repeats, inverted repeats, and mirror repeats. Numerous copies of a sequence on the same strand of DNA are referred to as direct repeats, while repeats in reverse direction on the opposite strand of the DNA molecule are referred to as inverted repeats. Everted repeats are DNA fragments with duplicated sequences on the opposite strand. In a gene or genome, a mirror repeat is a DNA sequence that shares homology with another portion of the same sequence [5]. As an illustration, consider the following sequence - TCGGTAATGGCT, One portion of the sequence, TCGGTA, is homologous to another portion. In highly super coiled DNA some mirror repeats can form intramolecular triplexes, or H-DNA [8]. These H-DNA are reported to regulate the expression of diseases associated genes [13].
The second largest cause of mortality for women globally is breast cancer, which is the most prevalent cancer among women [14]. *BRCA1* gene is located on chr17q, which codes for TSG (Tumor Suppressor Gene) proteins. According to molecular analysis, *BRCA1* exon mutations result in breast cancer [7, 8]. The objective of the current study is to identify mirror repeats in the exons of *BRCA1* gene using a simple manual bioinformatics method.

MATERIAL AND METHODS

Using the gene name *BRCA1* and accession number NM_007300.4, from the National Center for Biotechnology Information (NCBI) nucleotide sequence of the exons present in the *BRCA 1* gene (FASTA format) were retrieved. The exon sequences represent the query sequence. Using Reverse Complement tool, Parallel compliment of these exons were generated (http://www.bioinformatics. org/sms/rev_ comp_html). The parallel compliment of exons represents the subject sequence. Exon sequences retrieved from NCBI (query sequence) and parallel compliment of these exons were aligned for homology search using BLAST tool (Fig-1). Hits were obtained at various expected thresholds, and the parameters were set for a word size of 7. Mirror repeats were identified, using the threshold value at which the highest numbers of hits were obtained. Mirror repeats were identified in hits when the subject and query sequences had the reverse position number. On the basis of their symmetry, mirror repeats were divided into two categories: Perfect Mirror Repeats (PMR) and Imperfect Mirror Repeats (IMR). The mirror repeats were searched in the genome of *Homo sapiens, Mus musculus*, *Danio rerio* and *D. melanogaster* using mega BLAST tool.

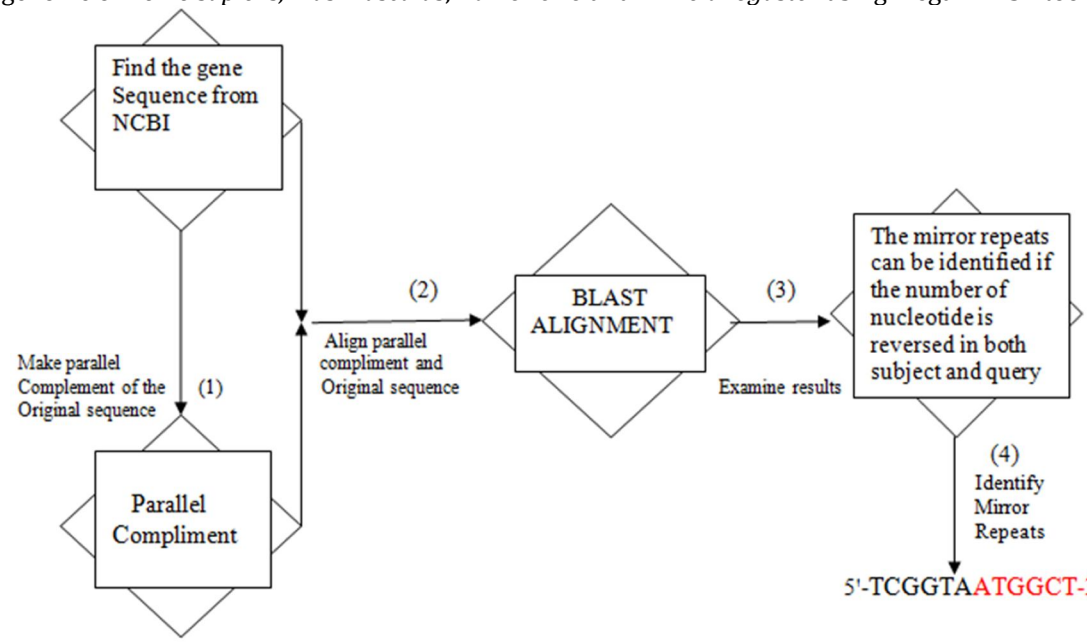

fig.1: Pictorial representation of methodology used to identify mirror repeats within *BRCA1* gene.

RESULTS AND DISCUSSION

The second most common reason for female fatalities is breast cancer. Pathogenic *BRCA1* and *BRCA2* mutations contribute to 5% of breast cancer cases. Breast cancer susceptibility gene (*BRCA1*) is located on chr17q and consists of 24 exon sequences. The risk of getting breast, ovarian, or prostate cancer increases when this gene is mutated. The 1863 amino acid *BRCA1* gene has been shown to include 300 disease-causing mutations. According to BIC (Breast Information Core), nonsense, frame shift, and splicing mutations in *BRCA1* and *BRCA2* genes result in truncated proteins 6, 7, 14].

In the current study, mirror repeats in the exons of *BRCA1* gene of breast cancer were identified using a simple manual bioinformatic method. We have also used nBMST tool for identification of mirror repeats. The results confirms a total of 141 mirror repeats within exons of *BRCA1* (Table-1). These mirror repeats are randomly distributed within the different exons of gene (Fig-2). The mirror repeats of size 7-12bps were most abundant whereas the larger mirror repeats were less abundant in the *BRCA1* gene (Table1). In the *BRCA1* gene, we also found those mirror repeats (≥10bps) that can create non-canonical BDNA forms [3, 4, 8]. We found a total of 24 homopurine (polydA-dG) and homopyrimidine (poly dC-dT)-rich sequences, of which two mirror repeats may adopt the H-DNA structure (represented in green colour in table 2).

We have identified different types of mirror repeats within the *BRCA1* gene and these are classified, based on the presence or absence of spacer elements at the centre of symmetry. Smallest mirror repeat found in the gene is of 7 bps and largest consists of 27 base pairs. The perfect mirror repeats were more common as compared to imperfect mirror repeats within the *BRCA1* gene. A few selected mirror repeats identified in the *BRCA1* gene are shown in Table 2 along with their location and classification. The complete detail of identified mirror repeats is given in the supplementary file. We also used Non-.BDNA motif search tool to identify mirror repeats within the exons of *BRCA1* gene. With this tool we could not found mirror repeats

in exons of *BRCA1* gene. By comparing the results obtained from both the tools, we can say that our method is more efficient in searching mirror repeats, because non B-DNA motif search tool could not search mirror repeats in exons of *BRCA1* gene.

Mega BLAST was used to examine repeats with sizes between 12 and 27 bps in order to study the presence of identified mirror repeats in various genera (Table 3). Here, a + sign indicates the presence of a mirror repeat and a - sign indicates its absence. Identified mirror repeats were present in the genome of *Homo sapiens, Mus musculus*; *Danio rerio* and *D. melanogaster,* this indicates universal existence of these mirror repeats**,** however, what role will these identified mirror repeats play at cellular and genetic level in breast cancer is still unknown.

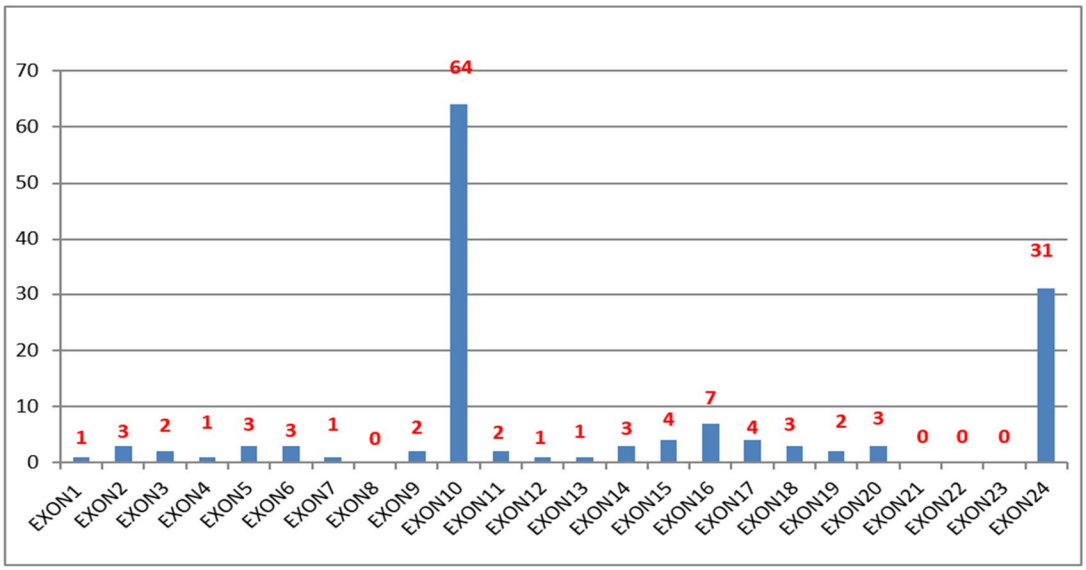

Fig. 2: shows distribution of mirror repeats in different exons of *BRCA 1* gene.

Table 1: Representing the length and location of mirror repeats present in the different exons of the *BRCA1* gene.

Accession No.	Exon	7-12bp	13-18bp	19-24bp	25-50bp	Total MR's
NM_007300.4	Exon-1	1	0	0	0	1
	Exon-2	3	0	0	0	3
	Exon-3	2	0	0	0	2
	Exon-4	1	0	0	0	1
	Exon-5	3	0	0	0	3
	Exon-6	2	1	0	0	3
	Exon-7	0	1	0	0	1
	Exon-8	0	0	0	0	0
	Exon-9	2	0	0	0	2
	Exon-10	57	3	3	1	64
	Exon-11	2	0	0	0	2
	Exon-12	1	0	0	0	1
	Exon-13	1	0	0	0	1
	Exon-14	3	0	0	0	3
	Exon-15	4	0	0	0	4
	Exon-16	5	1	1	0	7
	Exon-17	4	0	0	0	4
	Exon-18	3	0	0	0	3
	Exon-19	2	0	0	0	2
	Exon-20	3	0	0	0	3
	Exon-21	0	0	0	0	0
	Exon-22	0	0	0	0	0
	Exon-23	0	0	0	0	0
	Exon-24	30	0	1	0	31
	Total	129	6	5	1	141

Table 2: Classification of selected mirror repeats distributed in the exons of *BRCA1* gene.

Exon	Mirror Repeat	Position of Mirror Start position	Stop position	Length of mirror repeat	Type Of Mirror
Exon-6	AAACCGTGCCAAA	25	34	13	Perfect with single spacer
Exon-6	AAAAGGAAAA	92	104	10	Perfect mirror repeat
Exon-7	ACGTCTGTCTACA	85	97	13	Perfect with single spacer
Exon-10	TCCAAGAGAAGAAAAAGAAGAGAAACT	1508	1534	27	Imperfect mirror repeat
Exon-10	AAATGAACAGACAAGTAAA	1382	1400	19	Perfect with single spacer
Exon-10	AAAGATAATAGAAA	1788	1801	14	Perfect mirror repeat
Exon-10	GATTCAAACTTAG	3414	3426	13	Perfect with single spacer
Exon-10	AAAGGAAGAAAATCAAGGAAA	2045	2065	21	Imperfect mirror repeat
Exon-10	AAGGCAAAAACAGAA	1707	1721	15	Imperfect mirror repeat
Exon-10	AGAATCCTAGAGATACTGAAGA	415	436	22	Imperfect mirror repeat
Exon-10	AGTAATAATGA	54	64	11	Perfect with single spacer
Exon-10	CGGAAGAAGGC	726	736	11	Perfect with single spacer
Exon-10	ATGAAGAAGTA	2668	2678	11	Perfect with single spacer
Exon-10	ATGGAAGGTA	1335	1344	10	Perfect mirror repeat
Exon-10	CTCATTACTC	2223	2232	10	Perfect mirror repeat
Exon-16	TTGAAAGTTCCCCAATTGAAAGTT	123	146	24	Imperfect mirror repeat
Exon-16	AGCCTCTTCTCTGA	30	43	14	Imperfect mirror repeat
Exon-19	AAAGAAAGAAA	18	28	11	Perfect with single spacer
Exon-24	CGTTTTGTAAATGTTGTGC	1451	1469	19	Imperfect mirror repeat
Exon-24	ACTGGAGGTCA	762	772	11	Imperfect mirror repeat
Exon-24	GGAGGTGGAGG	908	918	11	Perfect with single spacer

Table 3: Shows distribution of mirror repeats of size >11bps 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.	AAACCGTGCCAAA	+	+	+	+
2.	ACGTCTGTCTACA	+	+	+	+
3.	TCCAAGAGAAGAAAAAGAAGAGAAACT	+	+	+	+
4.	AAATGAACAGACAAGTAAA	+	+	+	+
5.	AAAGATAATAGAAA	+	+	+	+
6.	GATTCAAACTTAG	+	+	+	+
7	AAAGGAAGAAAATCAAGGAAA	+	+	+	+
8	AAGGCAAAAACAGAA	+	+	+	+
9	AGAATCCTAGAGATACTGAAGA	+	+	+	+
10	TTGAAAGTTCCCCAATTGAAAGTT	+	+	+	+
11	AGCCTCTTCTCTGA	+	+	+	+
12	CGTTTTGTAAATGTTGTGC	+	+	+	+

CONCLUSION

In our current study, two simple bioinformatics methods were used to find mirror repeats in the *BRCA1* gene associated with breast cancer. A total of 141 mirror repeats have been observed through BLAST search whereas non B-DNA motif search tool could not identify mirror repeats within the exons of *BRCA1* gene. Our results shows different type of mirror repeats present within different exons of *BRCA1* gene. It will be highly interesting to know, the unknown function of these mirror repeats. Further molecular and biophysical studies on these mirror repeats are required to explain the function of these mirror repeats in breast cancer.

FUNDING

No external funding was used for this study.

CONFLICTS OF INTEREST

There are no conflicts 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. Gurusaran, M, Ravella, Dand Sekar, K. (2013). RepEx: repeat extractor for biological sequences. Genomics. 102(4):403-408.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. Treangen, T.J and Salzberg, S.L .(2012). "Repetitive DNA and next-generation sequencing :computational challenges and solutions". Nature Reviews Genetics.**13** (1): 36–46.
11. Ugarkovic, D. andPlohl, M. (2002). Variation in satellite DNA profiles - Causes and effects. EMBO Journal, 21(22), 5955–5959.
12. 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.
13. 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.
14. 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, Bhargava Anu and Bhardwaj Vikash. Identification of mirror repeats within the *BRCA1* gene of breast cancer. Bull. Env.Pharmacol. Life Sci., Vol 11 [12] November 2022: 12-16