International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

In Silico Analysis and Characterization of Mirror Repeats in the Human GLUT4 (SLC2A4) Gene

Preeti Sharma

Research Scholar, Department of Zoology, Baba Mastnath University, Asthal Bohar, Rohtak 124021, Haryana, India

Parveen Kumar

Assistant Professor Department of Zoology, Baba Mastnath University, Asthal Bohar, Rohtak 124021, Haryana, India

pkzoologybmu20@gmail.com

Abstract

Mirror repeats are special nucleotide patterns in which the sequence is symmetrical about a central point. This study is the first to look for mirror repeats in the human (*Homo sapiens*) GLUT-4 gene. Researchers used nucleotide BLAST (BLASTn) to extract mirror repeats from the full gene sequence, which is 6,540 base pairs long. They then used the Non-B DNA Motif Search Tool to find and analyze these motifs in the GLUT-4 (SLC2A4) gene. In total, 152 mirror repeats were found. Of these, 80.3% are perfect mirror repeats, while 19.7% are imperfect, indicating that most of these repeats are highly conserved and symmetrical. This research aims to clarify the role of mirror repeats in the GLUT-4 gene, a component of the human insulin signaling system.

Keywords: *Homo sapiens*, perfect mirror repeats, BLASTn, Non-B DNA motif, Imperfect mirror repeats.

Introduction

GLUT (Glucose transporter) family 1 and 2 transporters form channels that allow glucose to move through the plasma membrane at the cellular level. When insulin is present, the GLUT-4 isoform shifts from intracellular storage to the plasma membrane (Figure 1). At the cell surface, GLUT-4 facilitates passive entry of glucose into fat and muscle cells, aided by membrane-permeant esters of phosphatidylinositol 3,4,5-trisphosphate (Schultz, C. *et al*., 1998). The human genome contains thirteen sugar transporter proteins, including GLUT-4, GLUT-1 through GLUT-12, and HMIT (Joost and Thorens, 2001; Wood and Tray-hurn, 2003). These proteins help hexose sugars cross cell membranes without using ATP (Hruz and Mueckler, 2001). DNA stores genetic information and controls many biological processes. DNA can adopt conformations beyond the usual B-DNA form because of mutations, strand breaks, or incomplete crossing over. This can lead to repeated sequences in the genome (Gurusaran *et al*., 2013; Jain *et al.*, 2008; Zattera *et. al*., 2020). Repetitive DNA sequences, which are similar DNA segments found in many copies in all organisms, play a key role in the cell cycle, gene expression, chromosome structure, and karyotype evolution (de

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Koning *et al*., 2011; Jurka *et al*., 2007; Mehrotra and Goyal, 2014; Zattera *et al*., 2020). These repeats affect important processes like transcription, translation, recombination, and chromatin structure. In bacteria, yeast, mice, and humans, short repetitive sequences can produce H- and Z-DNA, which may cause mutations (Pandya *et al*., 2021; Wang and Vasquez, 2017). Based on their sequence, small repetitive sequences can be straight, inverted, or mirror repeats (Gurusaran *et al*., 2013; Jiang *et al*., 2018; Mirkin, 2001). An inverted repeat is a DNA fragment repeated in reverse order on the same strand (Mirkin, 2001). These repeats are important for controlling transcription and translation (Varshney *et al*., 2020). In bacteria, transcription ends often have inverted repeats and the related hairpin structures (Brazda *et al*., 2020; Lillo and Spanò, 2007). Direct repeats are sequences repeated in the same direction on the same DNA strand (Mirkin, 2001).

Both direct and inverted repeats can cause genetic rearrangements that lead to duplication or deletion of genetic material (Lovett, 2004; Marie and Symington, 2022). Repeated DNA sequences are classified as either tandem or interspersed, depending on their distribution throughout the genome (Jurka *et al.*, 2007; Pathak and Ali, 2012; Zattera *et al.*, 2020). Tandem repeats include microsatellites, minisatellites, and satellite DNA. Interspersed repeats are divided into long interspersed nucleotide sequences (LINES) and short interspersed nucleotide sequences (SINES), based on the length of the DNA sequences (Mehrotra and Goyal, 2014; 2012; Zattera *et al*., 2020). According to Mirkin (2001), mirror DNA repeats are short, symmetrical sequences on the same strand, meaning one part of the sequence matches another part. In mirror repeats (MR), the parallel and antiparallel complements are identical, unlike in normal DNA. DNA repeats are found in both prokaryotes and eukaryotes. Many studies have shown that mirror repeats exist in a range of species, including bacteria, viruses, and plants (Yadav *et al.*, 2022).

Researchers have studied where mirror repeats (MRs) are found in many eukaryotic organisms. Saini *et al.* (2023a, 2023b) looked at the patterns of MRs in three fruit fly species: *Drosophila melanogaster*, *Drosophila nasuta* Lamb, and *Drosophila bipectinata*. Later, Dhankhar *et al*. (2024) examined MRs in several *Caenorhabditis elegans* genes and also included *Xenopus tropicalis* and *Drosophila melanogaster in their analysis*.

A recent study on repetitive DNA sequences highlights the importance of understanding gene expression and suppression, particularly in the context of evolutionary conservation. Simple tandem repeats, satellite DNA, and palindromic sequences are examples of repeats that add variety to genomes in both prokaryotes and eukaryotes. Glucose transporter 4 (GLUT-4) is the insulin-sensitive glucose transporter found in human and rodent skeletal muscle and fat tissue. It moves to the plasma membrane when insulin is present. GLUT-4 was genetically engineered in 1989 and identified as the main isoform responsible for increased glucose uptake in muscle and fat tissue (Birnbaum, 1989; Charron *et al*., 1989; Fukumoto *et al*., 1989; James *et al.*, 1989; Kaestner *et al*., 1989). One study found that parallel complement sequences are important for gene silencing and 3’ end processing of mRNA in E. coli and Drosophila (Szabat and Kiersek, 2016). However, there

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

is little information about mirror repeats in GLUT-4. This study aims to identify mirror repeats in the GLUT-4 gene, which are important in the insulin signaling pathway. Mirror repeats can be perfect or imperfect, depending on the sequence. These mirror sequences are responsible for non-canonical or Non-B DNA structures in the genome (Schroth & Ho, 1995). They can also form both parallel and antiparallel complementary strands. In this study, we will look at the role of GLUT-4 mirror repeats in the insulin signaling pathway.

MATERIALS AND METHODS

The human GLUT-4 gene sequence (NCBI Reference Sequence: NC_000017.11) was obtained in FASTA format from the National Center for Biotechnology Information **(**Bhardwaj *et al*., 2013). The 6,540-base-pair gene was divided into fragments of 1,000 base pairs each. These fragments served as query sequences, while the reverse complement of the coding sequence, created using a reverse-complement tool(https://www.bioinformatics.org/sms/revcomp.html), was used as the subject sequence. Both sequences were aligned across relevant regions using BLAST to assess homology. This method also helped search for mirror repeats in the breakpoint region, where the gene was split into 1,000-base-pair segments. The program's parameters were set, and a 7-word-size alignment took considerable time. Hits were recorded at different expected threshold values (E-values), and the E-value with the most hits was used to identify mirror repeats. Mirror repeats were found in alignments where the position numbers in the subject and query sequences matched in reverse order. These repeats were grouped by placing a spacer between the subject and query sequences. MEGA BLAST results were then used to search the entire genome for the identified mirror repeats. To increase the number of hits, parameters such as word size, expected threshold, and maximum target sequence length were adjusted during the search.

The Non-B-DNA Motif Search Tool (https://nonbabcc.ncifcrf.gov/apps/nBMST/default/) was used to find mirror repeats in the GLUT-4 gene. Figure 1 shows the flowchart outlining the steps for retrieving these mirror repeats.

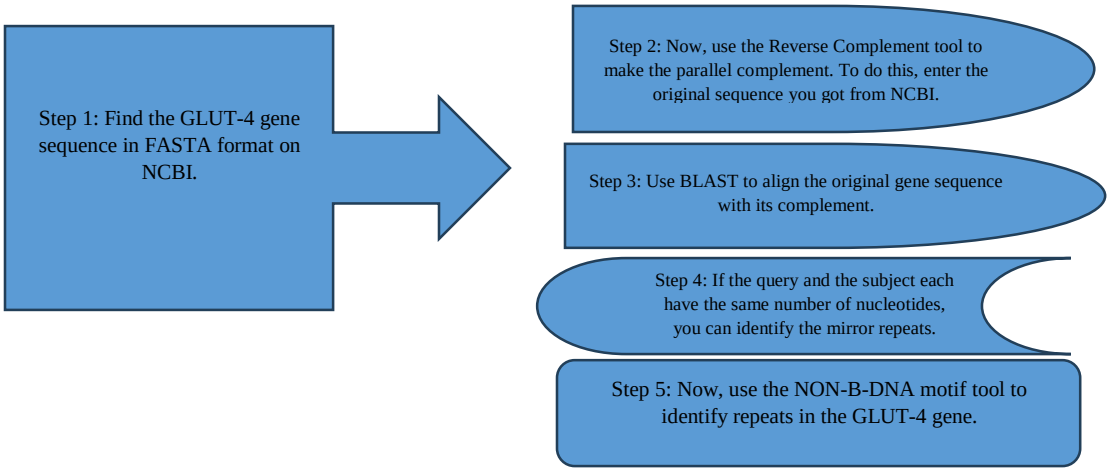

Figure 1: Shows the method used to retrieve the mirror repeats.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

RESULTS AND DISCUSSION

This summary outlines the main features and functions of the GLUT-4 gene, focusing on its genetic structure and role in the insulin signaling pathway. The GLUT-4 gene is located on human chromosome 17 (GRCh38.p14), with the NCBI reference sequence NC_000017.11. The full gene sequence, which is 6,540 base pairs long, was divided into seven segments of about 1,000 base pairs each: a) 1-1,000 b.p., b) 1,001-2,000 b.p., c) 2,001-3,000 b.p., d) 3,001-4,000 b.p., e) 4,001-5,000 b.p., f) 5,001-6,000 b.p., and g) 6,001-6,540 b.p.

Table 1 shows how we identified mirror repeats using a threshold of 70, which yielded the highest number of hits (171). After a frequency of 70, the number of hits stays the same, so we chose 70 as the threshold for the full gene and used this word range for the mirror repeat analysis.

Table 1: Explains why this frequency was chosen as the threshold:

S. No.	At frequency	Number of free hits	Mirror repeats
1.	0.05	No significant result	N.A.
2.	10	25	Range (1,2,4,8,10,11,20,22,23,25)
3.	20	59	Range (1,2,4,8,10,11,20,22,23,25,37,39,57,59)
4.	50	59	Range (1,2,4,8,10,11,20,22,23,25,37,39,57,59)
5.	60	59	Range (1,2,4,8,10,11,20,22,23,25,37,39,57,59)
6.	70	171	Range (1,2,4,8,10,11,20,22,23,25,38,40,60,62,93,117,122, 143,145,151,154,161,164)
7.	100	171	Range (1,2,4,8,10,11,20,22,23,25,38,40,60,62,93,117,122, 143,145,151,154,161,164)
8.	500	171	Range (1,2,4,8,10,11,20,22,23,25,38,40,60,62,93,117,122, 143,145,151,154,161,164)
9.	1,000	171	Range (1,2,4,8,10,11,20,22,23,25,38,40,60,62,93,117,122, 143,145,151,154,161,164)

The mirror repeats found in the gene were classified as IMR (imperfect mirror repeat), PMRSS (perfect mirror repeat with a single spacer), and PMR (perfect mirror repeat). In perfect mirror repeats, the sequences on either side of the central axis of symmetry are identical, while in imperfect mirror repeats, they do not match exactly. A perfect mirror repeat can have one or more spacer elements between the repeats. In total, 152 mirror repeats were identified in the gene. Of

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

these, 122 were perfect mirror repeats, and 30 were imperfect mirror repeats. Mirror repeats are symmetrical DNA sequences that play an important role in genome evolution.

The large number of short mirror repeats suggests that these sequences form easily during DNA replication, often through replication slippage and mutations. Because they do not significantly affect DNA structure, they are tolerated by evolution and contribute to genetic variation in the genome. Mirror repeats of moderate length (8–20 base pairs) can form secondary DNA structures that may affect DNA replication, recombination, and gene regulation.

These repeats can cause mutations or rearrangements that might help organisms adapt to changes in their environment, which supports evolution. The fact that there are very few long mirror repeats (over 20 base pairs) shows that these sequences are usually removed by natural selection.

Longer repeats can form stable non-B DNA structures, which may cause DNA breaks, replication errors, and genomic instability. Because of these harmful effects, natural selection removes most long repeats from the genome. As a result, the distribution of mirror repeats shows a balance between natural selection and mutation.

Long repeats are rare because they can destabilize the genome, whereas short repeats are retained because they promote diversity with lower risk. Mirror repeats play a key role in the evolution, adaptation, and stability of the genome. Research shows that long mirror repeats are rare, but short ones are very common. This points to an evolutionary balance in which long repeats are avoided due to instability, while short repeats are retained for variation. Mirror repeats are important for the genome's evolution and adaptability. Using a threshold frequency of 70, the dataset provides a detailed view of mirror repeats (MRs) in the coding sequence (CDS), ranging from 1 to 6540 base pairs. In total, 171 hits, or 134 mirror repeats, were found and distributed across different parts of the sequence.

Mirror repeats are not spread evenly across the CDS regions. Some segments, such as 1–1000, 3001–4000, and 5001–6000, have more repeats, indicating higher densities in these areas. The number of mirror repeats in each segment is fairly steady, usually between 15 and 25, indicating a moderate but consistent presence throughout the gene. Most repeats are PMSS (Perfect Mirror Repeat with Single Spacer), but PM (Perfect Mirrors) and I.P. (Imperfect Mirror repeats) are also found. This suggests that symmetric sequences are common in the CDS.

Most of the sequences are short, supporting the idea that shorter mirror repeats are more common and stable in the genome. Since mirror repeats are widely but unevenly distributed, they may not arise by chance but are likely retained for their functional or structural roles. The high number of PMSS-type repeats suggests a preference for highly symmetric DNA motifs, which can influence how secondary structures form.

Table 2 lists the number of hits and mirror repeats found in the full GLUT-4 gene. We saw different patterns of mirror repeats in each region. Region 1 had six imperfect mirror repeats, three perfect mirror repeats, and fifteen perfect mirror repeats with a single spacer. Region 2 had five imperfect

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

mirror repeats, four perfect mirror repeats, and fourteen perfect mirror repeats with a single spacer. In both regions, perfect mirror repeats with spacers were the most common, which suggests they are important for the structure and function of the gene.

A close comparison of the genomic regions showed clear differences in the number and structure of mirror repeat motifs. The results show that perfect mirror repeats with single spacer sequences (PMSS) are the most common type across all regions studied, suggesting they are important for genome evolution and structure. Region 3 was different because it had no typical perfect mirror repeats. Instead, it had seven imperfect mirror repeats and a high number of PMSS motifs, with fifteen in total.

The lack of perfect mirror repeats and the presence of many spacer-associated repeats suggest that these sequences often diverge and become more flexible in structure. This arrangement may help create unusual DNA shapes while still maintaining some symmetry by using spacers.

In the following region, researchers found 21 mirror repeats: 4 imperfect, 4 perfect, and 13 PMSS motifs. Compared to region 3, this area had a more balanced mix of perfect and imperfect mirror repeats, suggesting a transitional genomic structure with moderate sequence conservation. PMSS motifs remained the most common, underscoring their continued importance across regions. The 4,001 to 5,000 base-pair segments had the highest density of mirror repeats, with twenty-five motifs in total. Fourteen were PMSS motifs, seven were perfect mirror repeats, and four were imperfect repeats. The greater number of perfect and spacer-associated mirror repeats in this segment may indicate greater genomic stability and more sequence conservation.

Finding several types of repeats in a small genomic region suggests this area may play a role in structural interactions or regulatory functions. A similar pattern was seen in the nearby coding sequence (CDS) region, where researchers found twenty-two mirror repeats. This region had only two imperfect repeats, four perfect mirror repeats, and sixteen PMSS motifs.

The large number of PMSS motifs in the CDS region may reflect selective pressure to maintain interrupted symmetrical sequences rather than perfect repeats. This might help protect the coding sequence while still allowing for secondary structures or regulatory interactions at the nucleic acid level. In contrast, the terminal region, which is 540 base pairs long, had the fewest repeats because of its shorter length.

Although this region is smaller, it still contains many PMSS motifs, including eleven mirror repeats (five perfect and two imperfect). The strong presence of PMSS motifs in this short segment suggests that spacer-associated mirror repeats are the most stable and commonly maintained repeat structure in the genome we studied. When we compare the distribution of repeat motifs, it is clear that the genome prefers PMSS patterns over perfect or imperfect mirror repeats. This pattern may indicate that these motifs help the genome remain flexible, allow DNA to bend or form hairpin shapes, and support regulation, stability, or adaptation.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Table 2 shows the CDS region, threshold frequency, number of hits, number of mirror repeats, and types of mirror repeats found in the GLUT-4 gene. The mirror repeats are listed as PM for perfect mirror repeats, PMSS for perfect mirror repeats with a single spacer, and IP for imperfect mirror repeats.

Table 2 presents the distribution of mirror repeats in different CDS regions. In region 1 (positions 1 to 1,000), there are 6 imperfect and 18 perfect mirror repeats. Of these, 3 are perfect, and 15 are perfect mirror repeats with a single spacer. In region 2, we found 5 imperfect repeats, 14 perfect repeats with a single spacer, and 4 perfect mirror repeats. In region 3, there are 7 imperfect repeats, and all 15 perfect repeats have a single spacer. There are no completely perfect mirror repeats in this region.

In the next region, there are four imperfect mirror repeats, thirteen perfect mirror repeats with a single spacer, and four perfect mirror repeats. In the following region, we found four imperfect repeats, fourteen perfect repeats with a single spacer, and seven perfect mirror repeats. In region 6, there are 16 perfect mirror repeats with a single spacer, 4 perfect mirror repeats, and 2 imperfect mirror repeats. In the last region, we observed two imperfect repeats, two perfect repeats, and eleven perfect mirror repeats with a single spacer.

Table 2: Table 2 shows the CDS region, threshold frequency, number of hits, and the number and types of Mirror repeats in the GLUT-4 gene:

CDS:	**Threshold Frequency**	**Number of hits**	**Number of mirror repeats**	**Mirror Repeats**	**Type of Mirror repeats**
1-1000	70	171	24	1. CCTGCGCGTCC 2. GCGGCGCGGCG 3. GGCCCCCACCCCCACTCCTGG 4. AGGCGGCGCGGCGCTCCGGA 5. CCCTCCTCCC 6. ACGGCCCTCCCTGCA 7. TCAGAGACT 8. GACTGTCAG 9. CCGCCAGCGAGCGGCCACC 10. GAGGGGGAG 11. AGGCCGGA 12. CCGTCGGGCTTCC 13. TGGTTGGT	PMSS PMSS I.P. I.P. PM I.P. PMSS PMSS I.P. PMSS PM I.P. PM

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

				14. CCCTCACACTACC 15. CCCGCCC 16. GCCTCCG 17. TATTTAT 18. GGTCTGG 19. ACGGGCA 20. CGGAGGC 21. TTCCCTT 22. GCCACCG 23. GGGAGGG 24. AGATAGA	I.P. PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS
1001 - 2000	70	134	23	1. GCTTGGAAGGTTCG 2. ACACTCTCACA 3. CTCACAGAGGACACTC 4. TTCCTCCTTCCTCATT 5. AGAGGGGAGGGAAGA 6. TCACAACACT 7. TAATAATGCACGTGTTAAT 8. GGGCCCGGG 9. CCCGGGCCC 10. GGGTACAACATTGG 11. GGGGTGGGG 12. CCTGGTCC 13. GGTGGTGG 14. ACCCCCA 15. TTACATT 16. CTGTGTC 17. GGGTGGG 18. TGGGGGT 19. AGGAGGA 20. GGACAGG 21. GGAAAGG 22. GACCCAG 23. CCTCTCC	PM PMSS I.P. I.P. I.P. PM PMSS PMSS PMSS I.P. PMSS PM PM PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

2001 - 3000	70	94	22	1. CGCTCAACCAACTGGC 2. CCCGAGAGCCC 3. GTCCTGGCGGTGCTG 4. CTCTCTTCTTCCCTC 5. GGCCATTATCGG 6. CCAGGAAAAGGGCC 7. CACGGGCAC 8. CCACTCACC 9. CCTGCGGGGCGCCC 10. CTCCGCCTC 11. CTTCCCTTC 12. GTCGTGCTG 13. ACCT-CTACATCATCCA 14. TCTGTCT 15. TTCTCTT 16. CATGTAC 17. GCCCCCG 18. AGAAAGA 19. GCCTCCG 20. GTCCCTG 21. TGTATGT 22. GGGAGGG	I.P. PMSS I.P. I.P. I.P. I.P. PMSS PMSS I.P. PMSS PMSS PMSS I.P. PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS
3001 - 4000	70	151	21	1. CTGGTGTGGTC 2. CCATCACTTCTTCTTCTCCCCCACC 3. CCCGTGCCC 4. TCCGGACCAGGACT 5. CGACCAGC 6. CTGGGGTC 7. AGGTG-TTGTTGGTGGA 8. GGAGCGGGCGGGG 9. GCGGGGCG 10. GAGTTGAG 11. CCGAGCC 12. CACCCAC 13. CCCACCC 14. GTCTCTG	PMSS I.P. PMSS I.P. PM PM I.P. I.P. PM PM PMSS PMSS PMSS PMSS PMSS

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

				15. ACACACA 16. GAGGGAG 17. GACCCAG 18. TTCACTT 19. AAGGGAA 20. CCCTCCC 21. GAGCGAG	PMSS PMSS PMSS PMSS PMSS PMSS PMSS
4001 - 5000	70	142	25	1. CACCTCTTTCTCCAC 2. CTAAGAGTGTTTGGGGTTCAGAGAA TC 3. CCCAGAGACCC 4. TCTTCCTTCT 5. CCAGCTCTCTCTACC 6. ACCCGGCCCA 7. CCAGGGACC 8. CCCCGCCCC 9. TCAGGGACT 10. GACTCTCCTCTGAG 11.CCCTCCCCACCTCACTCCGTCAAC ACC T—CTTTCTCCACCTGTCCC 12. GTCCTCCTG 13. CTTGGTTC 14. GGAGGAGG 15. CTGCCGTC 16. CCCTTCCC 17. AGGCCGGA 18. CCCGCCC 19. ACGAGCA 20. CCACACC 21. ACAAACA 22. TTCCCTT 23. GTCCCTG 24. CTTCTTC 25. GGGTGGG	PMSS I.P. PMSS PM I.P. PM PMSS PMSS PMSS I.P. I.P. PMSS PM PM PM PM PM PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

5001 - 6000	70	147	22	1. AGTTCTCTCTTGA 2. AGAAGGGGGATTGGAGGGAAGA 3. GTCTCGCTCTG 4. TGGGAAGCTGGGGGAAGGGT 5. GGACGGCAGG 6. GTGTTGTG 7. CCACCACC 8. GGAAAAGG 9. TATTTAT 10. GGTTTGG 11. GGGTGGG 12. GACTCAG 13. CTTCTTC 14. ACTTTCA 15. GCCTCCG 16. CCACACC 17. TTTATTT 18. ATTTTTA 19. GCCACCG 20. CCGCGCC 21. CTCCCTC 22. CCTCTCC	PMSS I.P. PMSS I.P. PM PM PM PM PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS
6001 - 6,540	70	50	15	1. CCCCTCCTCCCC 2. ATATTTTTATA 3. GAAACCAAAG 4. TAAATGTATATATATAGAT 5. CTGTG-AGGGAGGTGTC 6. ATAGATA 7. GTCACTG 8. TGGAGGT 9. CAGGGAC 10. CTCCCTC 11. GTAAATG 12. CTCACTC 13. CAGGGAC 14. GACACAG 15. TAGCGAT	PM PMSS PM IP IP PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS PMSS

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Figure 2 shows the alignment patterns and sequence similarity among the analyzed sequences in a dot plot generated with the NCBI-BLAST program. With a word size of 7 and a frequency threshold of 70, this method identified conserved sequence regions and local homologous alignments with higher sensitivity and greater resolution.

The plot displays many short, scattered diagonal segments, which indicate short mirror repeats spread throughout the sequence. There are no long, continuous off-diagonal lines, suggesting the absence of long repeat regions. This indicates that the DNA sequence has many short repeats but few or no long repeats.

The scattered dots indicate that these repeats are distributed throughout the sequence rather than grouped. The diagram also suggests the sequence contains many short repeats that add variation. Since there are few long repeats, the risk of structural instability is lower.

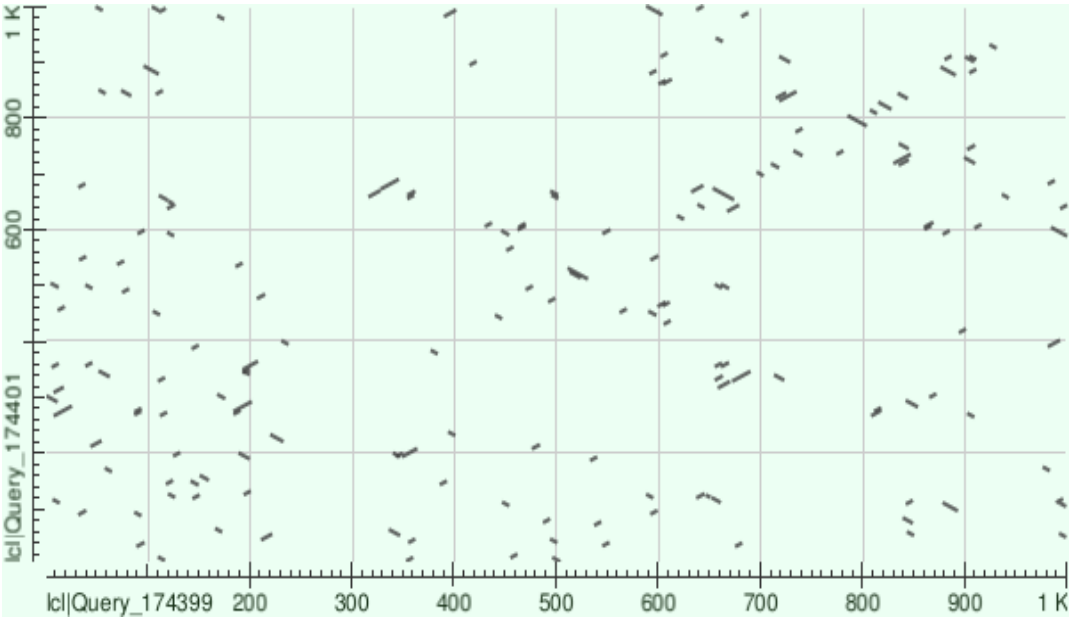

Figure 2: Dot-plot part-1 at frequency 70:

Table 3 lists the total number of perfect and imperfect mirror repeats by length. Most mirror repeats are short, with 72 repeats between 3 and 7 base pairs. There are 38 repeats of 8 to 10 base pairs, 37 repeats of 11 to 20 base pairs, and just 5 repeats longer than 20 base pairs.

Table 3: Total number of perfect and imperfect mirror repeats of different lengths.

Region	3-7 b.p	8-10 b.p.	11-20 b.p.	Above 20 b.p.	Perfect M.R.	Imperfect M.R.
1-1,000	10	06	07	1	18	06
1,001-2,000	10	06	07	Nill	18	05

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

2,001-3,000	09	05	08	Nill	15	07
3,001-4,000	11	05	04	01	17	04
4,001-5,000	08	11	04	02	21	04
5,001-6,000	14	04	03	01	20	02
6,001-6,540	10	01	04	Nill	13	02
Total	72	38	37	05	122	30

This distribution gives us insight into evolution. Most mirror repeats are short, with almost half in the 3-7 base-pair range. As repeats get longer, they become much less common. There are a moderate number between 8 and 20 base pairs, but repeats longer than 20 are rare. In general, short mirror repeats appear far more often than long ones.

The pie chart in Figure 3 shows that the 3–7 base-pair group accounts for the largest share, indicating that short mirror repeats are the most common in this dataset. The 8–10 and 11–20 base-pair groups each account for a moderate, similar share, which shows a steady drop as repeat length increases. Repeats longer than 20 base pairs are very rare.

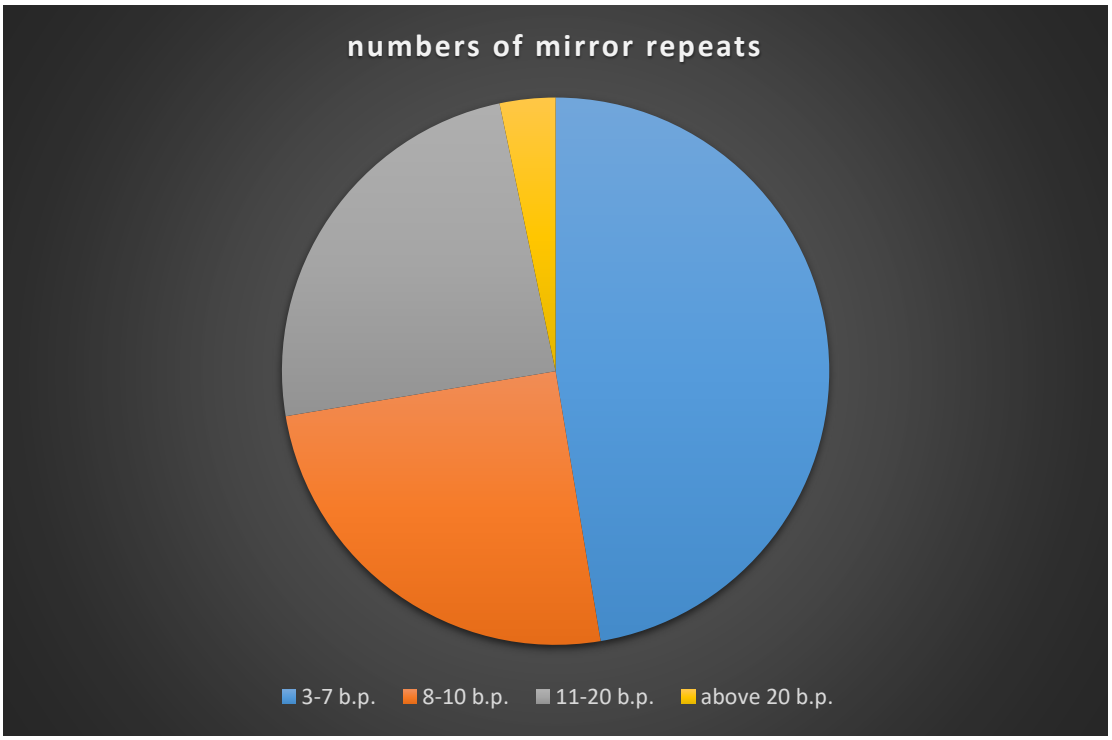

Figure 3: Representing the relative numbers of mirror repeats in the form of a pie chart.

The chart shows that mirror repeats are more common when they are shorter, but become rare as their length increases. This pattern suggests that mirror repeats might help maintain genome stability and support molecular diversity.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open access journal
Impact Factor: 8.3 www.ijesh.com ISSN: 2250 3552

Figure 4 shows a line chart comparing seven groups (labeled 1 to 7) across four parameters: number of hits, total number of mirror repeats (m.r.), number of mirror repeats, and number of perfect mirror repeats.

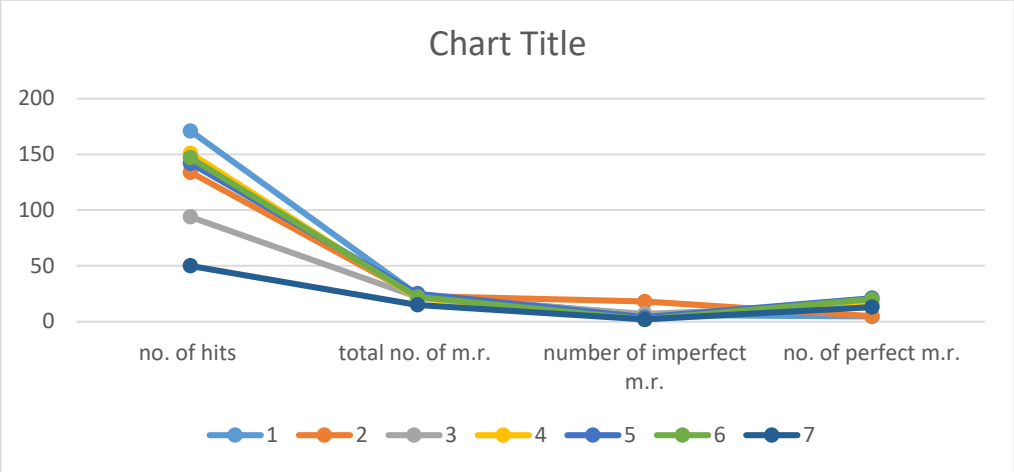

Figure 4: Representing the zones of gene GLUT-4 for mirror repeat positions:

The graph shows that each group begins with the most hits, although the exact numbers vary. After this initial peak, the number of mirror repeats drops quickly and then remains roughly the same across all groups.

As the number of mirror repeats keeps going down, only a small part of the hits are mirror repeats. The number of perfect mirror repeats increases slightly compared to the previous group, but it remains low. All groups show the same pattern: a drop and then a small increase. So, there are many hits at first, but only a few are mirror repeats, and even fewer are perfect mirror repeats. Each step gets more specific.

References:

1. Schultz, C. *et al*., 1998. Membrane-permeant esters of phosphatidylinositol 3,4,5-trisphosphate. *Journal of Biological Chemistry*, 273(49), pp.33265–33268.
2. Gurusaran, M., Ravella, D., Sekar, K. (2013). Rep Ex: repeat extractor for biological sequences. Genomics. 102(4):403-408.
3. Jain, A., Wang, G., and Vasquez, K. M. (2008). *DNA triple helices: Biological consequences and therapeutic potential*. Biochimie,**90**(8), 1117–1130.
4. 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
5. 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. https://doi.org/10.1371/journal.pgen.1002384

6. 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
7. Mehrotra S, Goyal V. (2014). Repetitive Sequences in Plant Nuclear DNA: Types, distribution, evolution and function. Genomics, Proteomics and Bioinformatics 12(4): 164-171
8. Pandya, G., Kirtonia, A., Sethi, G., Pandey, A. K., & Garg, M. (2021). The implication of epithelial–mesenchymal transition in cancer progression and metastasis. Cells, 10(10), 2619
9. Wang, Y., and Vasquez, K. M. (2017). Effects of replication stress on DNA damage and genome stability. Genes, 8(2), 46
10. Jiang, C., Zhao, Z., & Mutational Mechanisms Group. (2018). The origin and evolution of short tandem repeats in the human genome. Genomics, Proteomics & Bioinformatics, 16(1), 1–9. https://doi.org/10.1016/j.gpb.2017.11.003
11. Mirkin, S.M. (2001). DNA topology: fundamentals. ELS 111
12. Yadav S, Yadav U and Sharma DC (2022) In silico Approach for the Identification of Mirror Repeats in Selected Operon Genes of Escherichia coli strain K-12 Substrain MG1655. Biomedical and Biotechnology Research Journal 6(1):9397.
13. Saini K, Dangi N, Bhardwaj V. (2023a). Mirror repeats in the Intersex gene of Drosophila melanogaster Meigen. Indian Journal of Entomology 85(1): 40–45.
14. Saini K, Dangi N, Yadav D, Bhardwaj V. (2023b). Identification of mirror repeats within the Maleless (mle) gene of Drosophila melanogaster Meigen. Indian Journal of Entomology 85(1): 14–18.
15. Dhankhar M, Yadav D, Bhardwaj V. (2024). In silico approach for the identification of mirror repeats within the ced-3 gene of Caenorhabditis elegans. Indian Journal of Science and Technology 17(22): 2255–2262.
16. Varshney, M., Parel, J. T., Raizada, N., and Sarin, S. K. (2020). Initial psychological impact of COVID-19 and its correlates in Indian community: An online (FEEL-COVID) survey. PLoS ONE, 15(5), e0233874.
17. Brázda, V., Haronikova, L., Liao, J.C.C. and Fojta, M., (2020). DNA and RNA quadruplex-binding proteins. International Journal of Molecular Sciences, 21(11), p.4172.
18. Lillo, F. and Spanò, A. (2007). DNA supercoiling and its role in transcriptional regulation. International Journal of Medical Microbiology, 297(2), pp.65–81.
19. Susan T. Lovett (2004). Encounters between DNA replication and transcription complexes cause instability of repeated DNA sequences in bacteria. Molecular Microbiology, 52(5), 1243–1253. https://doi.org/10.1111/j.1365-2958.2004.04045.

20. Marie, L., and Lorraine S. Symington (2022). *Mechanisms of gross chromosomal rearrangements*. Nature Reviews Molecular Cell Biology, 23(8), 529–543. https://doi.org/10.1038/s41580-022-00465-8.
21. Pathak, D., Ali, S. (2012). Repetitive DNA: A tool to explore Animal Genomes/Transcriptomes. Genomics in Tech. 155-180. Functional genomics in Tech. 155-180
22. Birnbaum, M. J. (1989). Identification of a novel gene encoding an insulin-responsive glucose transporter protein. *Cell, 57*(2), 305–315.
23. Charron, M. J., Brosius, F. C. III, Alper, S. L., and Lodish, H. F. (1989). A glucose transport protein expressed predominantly in insulin-responsive tissues. *Proceedings of the National Academy of Sciences USA, 86*, 2535–2539.
24. Fukumoto, H., Kayano, T., Buse, J. B., Edwards, Y., Pilch, P. F., Bell, G. I., & Seino, S. (1989). Cloning and characterization of the major insulin-responsive glucose transporter expressed in human skeletal muscle and other insulin-responsive tissues. *Journal of Biological Chemistry, 264*, 7776–7779.
25. James, D. E., Strube, M., and Mueckler, M. (1989). Molecular cloning and characterization of an insulin-regulatable glucose transporter. *Nature, 338*, 83–87.
26. Kaestner, K. H., Christy, R. J., McLenithan, J. C., Braiterman, L. T., Cornelius, P., Pekala, P. H., and Lane, M. D. (1989). Sequence, tissue distribution, and differential expression of mRNA for a putative insulin-responsive glucose transporter in mouse 3T3-L1 adipocytes. *Proceedings of the National Academy of Sciences USA, 86*, 3150–3154.
27. Szabat, M., Gudanis, D., Kotkowiak, W., Gdaniec, Z., Kierzek, R. and Pasternak, A., 2016. Thermodynamic features of structural motifs formed by β-L-RNA. *PLoS ONE*, 11(2), p.e0149478.
28. Bhardwaj, B., Swapni, G., Sitaram, M. and Kulbhushan, S. (2013). FPCB: A simple and swift strategy for mirror repeats identification. arXiv preprint arXiv:1312.3869
29. Joost, H.G. and Thorens, B., 2001. The extended GLUT-family of sugar/polyol transport facilitators: Nomenclature, sequence characteristics, and potential function of its novel members. *Molecular Membrane Biology*, 18(4), pp.247–256.
30. Wood, I.S. and Trayhurn, P., 2003. Glucose transporters (GLUT and SGLT): Expanded families of sugar transport proteins. *British Journal of Nutrition*, 89(1), pp.3–9.
31. Hruz, P.W. and Mueckler, M.M., 2001. Structural analysis of the GLUT1 facilitative glucose transporter. *Molecular Membrane Biology*, 18(3), pp.183–193.