ISSN (Online): 2091-2609
DOI Prefix: 10.3126/ijasbt

International Journal of Applied Sciences and Biotechnology (IJASBT)

Research Article

In-silico approach for the identification of mirror repeats within Glucagon, Sox17, Gapdh and PKLR diabetic genes of *Rattus norvegicus*

Babli[1], Parveen Kumar[1]*

[1]Department of Zoology, Baba Mastnath University, Asthal Bohar, Rohtak- 124021, Haryana, India

Article Information

Received: 23 December 2025
Revised version received: 21 March 2026
Accepted: 23 March 2026
Published: 20 April 2026

Cite this article as:
Babli and P. Kumar (2026) Int. J. Appl. Sci. Biotechnol. Vol 14(1): 28-43. DOI: 10.3126/ijasbt.v14i1.92261

***Corresponding author**
Parveen Kumar,
Department of Zoology, Baba Mastnath University, Asthal Bohar, Rohtak- 124021, Haryana, India.
Email: pkzoologybmu20@gmail.com

Peer reviewed under authority of IJASBT

Abstract

Repetitive sequences are immensely distributed in the genomes of all viruses, bacteria, plants and animals. Repetitive sequences throughout the genome in both coding and non-coding regions are a major source of endogenous DNA damage, due to the propensity of many of them to form alternative non-B DNA structures that can interfere with replication, transcription and DNA repair The alternative non-canonical structures include triple-helices, hairpins, cruciforms and slipped structures and they are more likely to form at particular repetitive sequences such as mirror repeats, inverted repeats, direct repeats, and short tandem repeats. Although mirror DNA repeats were discovered in genomic DNA,the ability of a subset of lengthy homopurine-homopyrimidine mirror repetitions (H-motifs) to form a triple-helical DNA secondary structure was the only functional property that was thoroughly identified. Over the past decade, biological databases have expanded significantly due to developments in genomics, computational biology, and bioinformatics. A computational method was used in this study to identify mirror repetitions in the *Rattus norvegicus*'s diabetic genes. The Fast Parallel Complement Blast (FPCB) tool, a rapid and efficient bioinformatics method for identifying gene and genome repeats, was used to do this. Mirror repeats were found 199 in Glucagon, 135 in Sox17, 116 in Gapdh and 167 in PKLR genes during this study. Through this study, we can better understand the function of mirror repeats in the *Rattus norvegicus*'s genes, which may be further linked to mirror repeats in the human genome.

Keywords: *Rattus norvegicus*; Repetitive sequences; Mirror repeat; FPCB; Blast; Diabetic genes.

Introduction

DNA is both a dynamic, three-dimensional molecule that frequently deviates from the traditional double helix and a sequence of letters that may be broken down linguistically. Since 1953, B-DNA, the classic right-handed DNA double-helical structure, has been identified. Although B-DNA is the most common arrangement within cells, over 20 alternative non-canonical structures have been identified (Ghosh and Bansal, 2003). Alternative structures like cruciforms, hairpins, slipped structures and triple helices are more likely to occur at specific repeated sequences like short tandem repeats, mirror repeats, inverted repeats, and direct repetitions (Wells, 2007). Based on in vitro investigations of prokaryotic (Todd and Glickman 1982; Hoede *et al.,* 2006) and eukaryotic cells (Biffi *et al.*, 2013; Bacolla *et al.*, 2016; Kouzine *et al.*, 2017), non-canonical secondary structures are linked to higher mutability. The rate at which complimentary strands re-associate after being heated and disassociated should be inversely related to the amount of haploid DNA present in each cell. This pattern is true for bacteria and viruses, but it was found that in eukaryotes, large portions of the genome reassociated far

more quickly than their cellular DNA quantity would have indicated. That insight led to the startling discovery that repetitive DNA made up more than half of the genomes of many eukaryotes. From a few to several thousand nucleotide sequences, repeated DNA sequences are genomic segments that recur throughout the genome and are found in the genomes of higher animals (Britten and Kohne, 1968; Jurka *et al.,* 2007; Biscotti *et al.,* 2015).

Reassociation kinetics led to the initial discovery of repetitive DNA, which was categorized as "highly" and "middle" repetitive sequences, roughly corresponding to tandem and interspersed repeats Interspersed repeats are DNA fragments that are introduced into host DNA more or less randomly and have a maximum size of 20–30 kb. On the other hand, tandem repeats are collections of DNA segments that are arranged head-to-tail next to one another. Interspersed repeats are transposable elements (TEs) that have been inserted into genomic DNA; they are primarily inactive and frequently incomplete copies. A sufficiently long repeat may cause the results of mutational events to preferentially change from contractions to expansions. Strand slippage and/or hairpin formation during replication or repair (Ruggiero and Topal, 2004), replication fork stalling that varies by strand orientation (Gerhardt *et al*., 2016; Rastokina *et al*., 2023), impairments in Okazaki fragment maturation (Tsutakawa *et al*., 2017), homologous recombination (Tang *et al*., 2011), error-prone DNA repair (Du et al., 2012; Ezzatizadeh *et al.,* 2014) and challenges processing single-stranded DNA nicks or gaps within H-DNA structures (Li *et al.,* 2024; Masnovo *et al.,* 2024) are some of the known mechanisms that cause DNA repeat instability. Even though these structures are crucial to human physiology, it is well known that their unusual forms make it difficult for DNA machinery to process them.

Two-dimensional gel-electrophoresis, which shows the change in writhe on formation of these non-B DNA structures, can be used to monitor the superhelical stress-induced transition for purine-pyrimidine conformations. A regionally underwound DNA structure changes the molecule's topology, making it less compact and hence as a result there is decrease in mobility in an agarose gel (Collier and Wells, 1990). The magnitude of this mobility shift can reveal how many helical translocally unwound helices and how much superhelical stress are needed to induce the transition to the alternative structure. Non-B motifs do not have a uniform thermodynamic capacity to form secondary structures. Studies using biophysical simulations and experiments indicate that hairpin formation is optimal at certain spacer and arm lengths (Nag and Petes, 1991; Goddard *et al.,* 2000). In particular, for spacer and arm lengths that are particularly suited to hairpin/cruciform formation, we would expect to observe higher mutability if the physical formation of a secondary structure affected mutability (Sinden *et al.*, 1991). Spacer-to-arm mutation enrichment is indeed variable for different spacer sizes and various arm lengths, for example, structure-forming repeats have been shown to stall replication (Dahan *et al*., 2018). Intramolecular triple helix, H-DNA, which requires a sequence of all purines on one strand and all pyrimidines on the other strand arranged in mirror symmetry (Mirkin *et al.*, 1987). This subset of mirror repeats (H-motifs) forms a triplex when a strand from one half of the repeat (whether purine or pyrimidine) folds back to bond with the duplex half of the repeat, while its complement remains single-stranded (Hisey *et al*., 2024). To this day, formation of H-DNA is the only known functional property of mirror repeats.

Notably, H-DNA is a target for multiple DNA repair enzymes (Khristich and Mirkin, 2020; Wang and Vasquez, 2023) and develops in vivo (Matos-Rodrigues *et al.*, 2022), limiting important genetic events by halting transcription and DNA replication. The length of the repeat motif causes an exponential rise in H-DNA stability (Lyamichev *et al*., 1989). As a result, stable H-DNA structures significantly increase H-motif repeat expansion and induce genomic instability (Wang and Vasquez, 2023; Hisey *et al*., 2024). This phenomenon has been associated with several human disorders because sufficiently lengthy H-motifs affect gene activity (Depienne and Mandel, 2021). The first known H-motif in disease, $(GAA)_n$ repeats in Friedreich's ataxia, has been joined by the recent discoveries of $(CCCTCT)_n$ repeat expansions in X-linked dystonia parkinsonism (Bragg *et al.*, 2017), $(AAGGG)_n$ repeat expansions in CANVAS disease (Rafehi *et al.,* 2019), and $(GAA)_n$ repeats in FGF14-related ataxia (Pellerin *et al*., 2023).

Expansions of H-motif STRs were also observed in several human cancers (Bacolla *et al.,* 2016).In these known examples, carriers of a relatively long H-motif transmit an even longer allele to their offspring, and this allele continues to expand somatically throughout life, progressing the disease. Therefore, knowing how H-motifs modify to lengthen is essential. Slipped-strand structures with looped out bases can result from misalignment of direct repeat motifs, which are made up of two copies of the repeating unit separated by a non-repetitive spacer. (Sinden *et al.,* 2007).

Inverted repeat motifs, which consist of repetitive sequences complementary to each other (e.g., 5′-GACTGC and GCAGTC-3′) separated by a non-repetitive spacer, are capable of forming hairpins and cruciform structures (Nag and Petes 1991). Mirror repeat motifs that consist of stretches of homopurines: homopyrimidines arranged in a mirrored fashion, separated by a spacer, can form triple-helix (H-DNA) structures (Htun and Dahlberg, 1988). Finally, Z-DNA motifs, which consist of alternating pyrimidines and purines, such as $(CG:CG)_n$ or $(CA:TG)_n$, can form left-handed zig-zag DNA structures (Wang et al.

1979; Singleton *et al*., 1982). The frequency of different repetitions in bacterial genomes and organelles can be either random or enriched for direct or inverted tandem repeats. Nonetheless, eukaryotic genomes have an excess of mirror repeats of homopurines and homopyrimidines.

Whereas it has long been known that direct repeats are abundant in eukaryotes, the dramatic overrepresentation of such mirror repeats as H palindromes was a surprise. In the human genome, the majority of mirror repeats are also direct repeats. Up to 85% of all H palindromes are direct repeats in the worst scenario. Just 20% of inverted repeats are also straight repeats, for contrast. Similar ratios were found in yeast DNA, with AT-rich regions showing the largest overlap between mirror and direct repetitions. On the other hand, prokaryotic DNA rarely coincides mirror and direct repetitions. While the percentage of the genome that corresponds to coding sequences varies greatly, ranging from 67% in yeast DNA to 4% in a sample human sequence, the overrepresentation for eukaryotic genomes is identical for all of them each structure for overlap with the GenBank annotation's coding features.

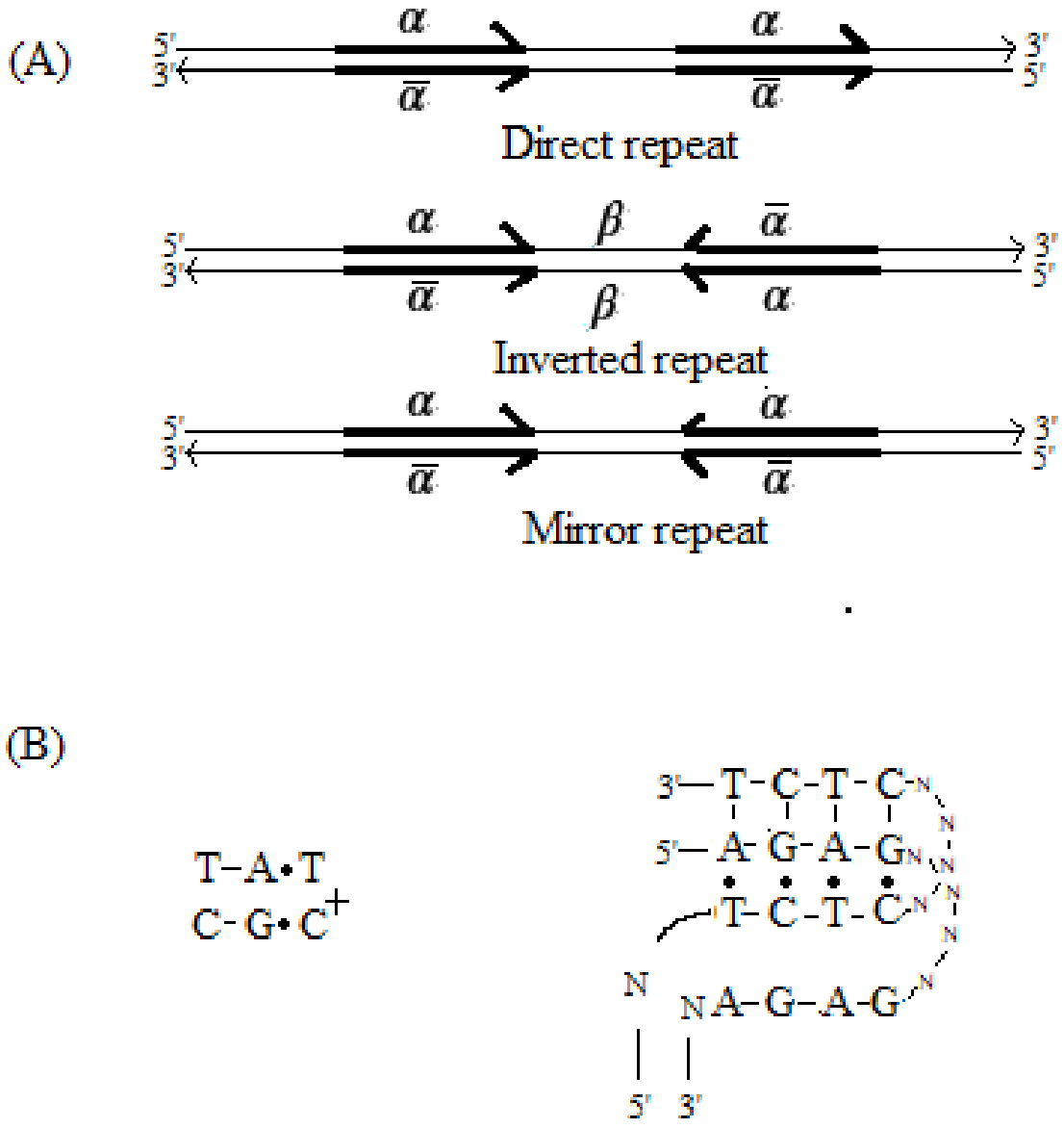


Fig. 1: (A) Different type of mirror repeats: Direct repeat, inverted repeat and Mirror repeat.

(B) H-DNA (intramolecular triplex) structure formed by homopurine-homopyrimidine structure mirror repeats.

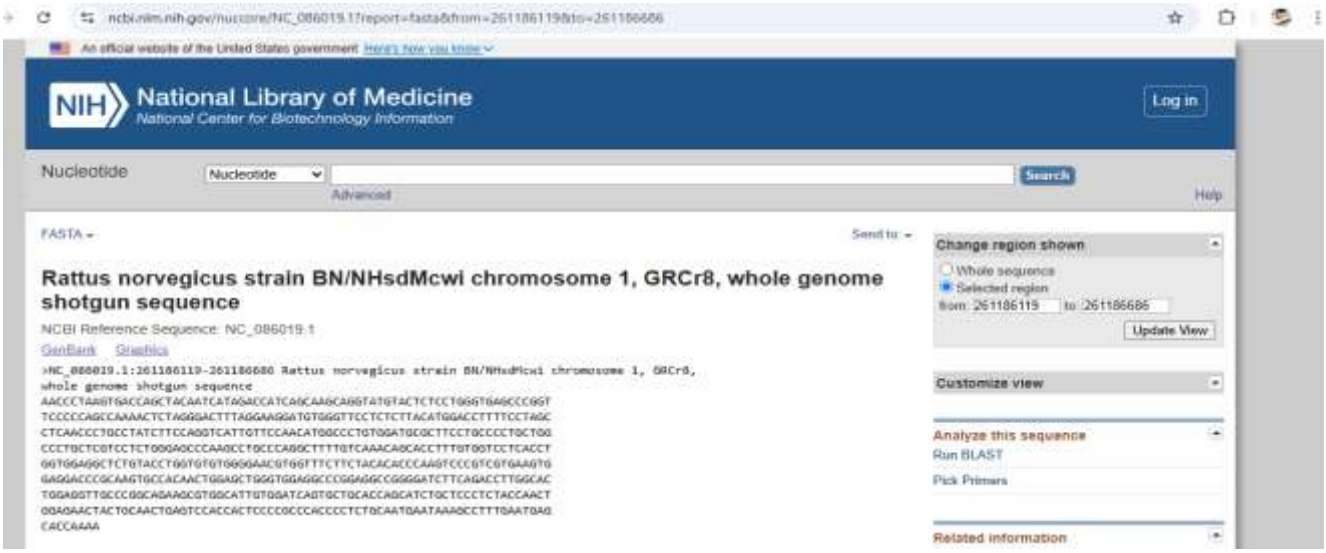
Diabetes is a chronic noncommunicable disease (Unnikrishnan *et al., 2016)* defined by chronic hyperglycemia caused by abnormalities in the secretion and function of the pancreatic hormone insulin (Kharroubi *et al.,* 2015). Diabetes mellitus is classified into type 1, type 2, and gestational diabetes depending on its cause and clinical aspects (*Rayburn et al., 1997)*. Animal models can be developed through two main mechanisms: disease induction or genetic modification. *Rattus norvegicus* is one of the most well-established experimental model organisms, with use of the species dating back to the mid-19th century (Modlinska and Pisula, 2020). The long-standing use of *Rattus norvegicus* in the laboratory as a model organism has led to a multitude of discoveries, providing insight into human physiology, behaviour and disease. For sequence data, one may use DNA, RNA or Protein data depending on your interest. All the completed sequences can be retrieved from the NCBI public website: http://www.ncbi.nlm.nih.gov/. The database for the DNA sequence is called GenBank. The manual bioinformatics approach is used to identify mirror repeats where FPCB (FAST PARALLEL COMPLEMENT BLAST) is the most commonly used method for analyzing or locating mirror repeats in various prokaryotes and eukaryotes In addition, it was used to characterize maleless (mle) gene, a sex determination gene in Drosophila melanogaster (Saini *et al*., 2023). The existence of highly divergent repeats could be associated with the presence of a single-type triplet periodicity in various genes or with the packing of bacterial DNA into a nucleoid (Korotkov *et al*., 2023).

Mammalian Glucagon (Gcg) is primarily expressed in the alpha cells of the pancreas, L-cells of the intestine, and some neurons of the brainstem (Lafferty *et al.,* 2021). Sox17 is expressed at high levels in the human fetal pancreas and at lower levels in adult islets (McDonald *et al.,* 2009). Embryos from diabetic rats and embryos cultured in high glucose concentrations (Deveze-Alvarez *et al.,* 2001) showed decreased activity of Gapdh (by 40–60%) and severe dysmorphogenesis on gestational days 10.5 and 11.5. Pyruvate kinase is a key glycolytic enzyme. Isoforms that are expressed in the red cell, liver, pancreatic beta-cells, small intestine, and proximal renal tubule are encoded by the 12 exons of the PKLR gene, which maps to chromosome 1q23 (Yamada and Noguchi, 1999).

Methodology

The coding sequence of the Glucagon, Sox 17, Gapdh and PKLR genes of *Rattus norvegicus* retrieved from NCBI website (National Centre for Biotechnology Information) in FASTA format (https://www.ncbi.nlm.nih.gov/). The nucleotides sequences of gene were splitted into different regions of 500 nucleotides. Using the reverse complement tool, a parallel complement of each region of gene was contrived. Using the BLAST tool, the original query sequence and its parallel complement were aligned for homology by selecting a word size limit of 7, and expected thresholds (E value) where maximum numbers of hits were observed. Mirror repeats can be identified easily as those hits where position number of the query sequence and subject sequence are exactly reversed. The identified mirror repeats were classified as perfect mirror repeats, perfect mirror repeats with one spacer and imperfect mirror repeats (Fig. 2) (Bhardwaj *et al*., 2013).

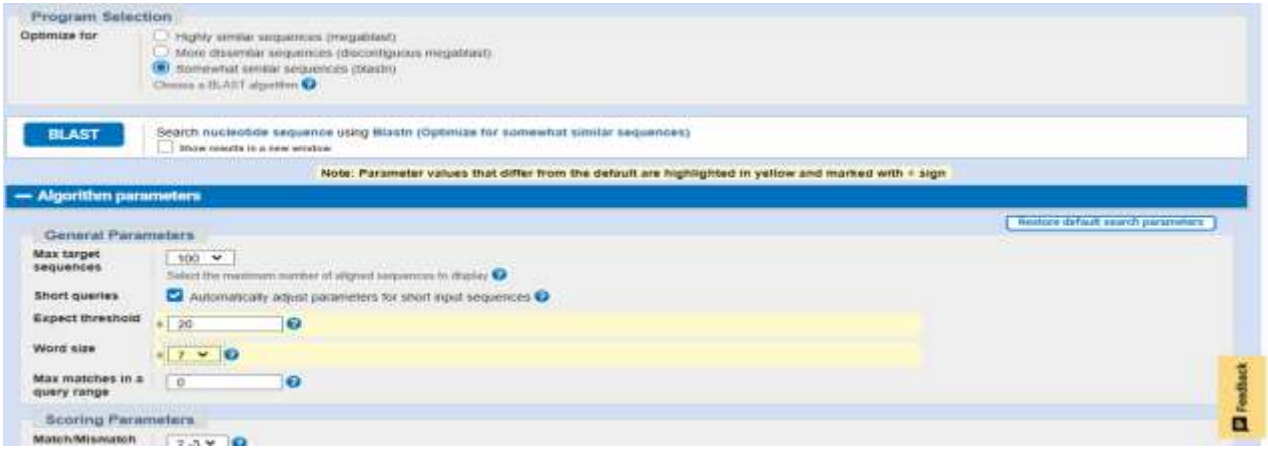
Step 1. Retrieve the original nucleotide sequence from NCBI, divide the whole sequence into fragments (1-500), that represents Query sequence.



Step 2. Using the Reverse Complement Tool, make the Parallel Complement of Query sequence. The Parallel Complement represents Subject sequence.

Step 3. Align the both Query and Subject sequence using Blast Tool.

Set program selection - Somewhat similar sequences, Algorithm parameters – expected threshold 20 and word size limit 7.



Step 4. Mirror repeats can be detected by examining the nucleotide which are reversed in Query and Subject sequence.

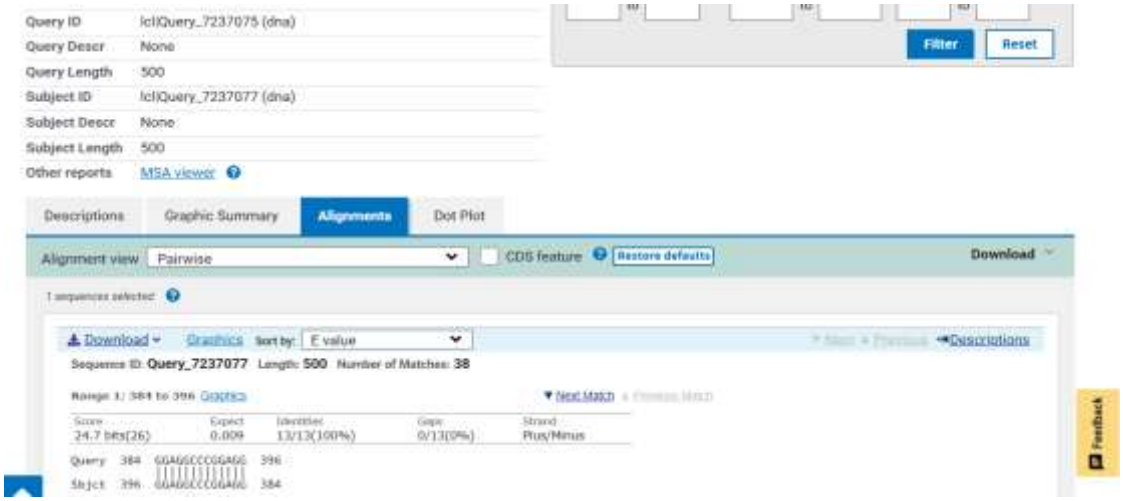
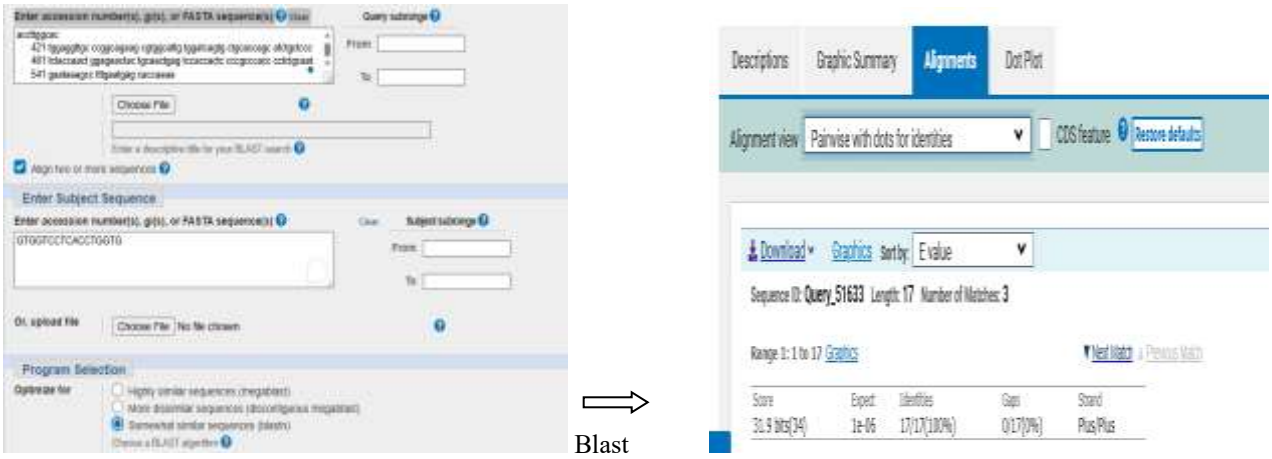


Fig. 2: FPCB approach to find out mirror repeats.

Further, identified mirror repeats were again searched within the complete gene and genome of *Rattus norvegicus* andand Homo sapiens genome using MegaBLAST tool. Distribution of perfect mirror repeats of gene is studied. Firstly, perfect mirror repeat was examined within complete gene by program selection (somewhat similar sequences) Blast at expected threshold 20 and word size was set to 7.



Secondly, perfect mirror repeats examined within *Rattus norvegicus* and Homo sapiens genome by program selection (highly similar sequences) MegaBLAST at threshold limit 20 and word size was set to 16.

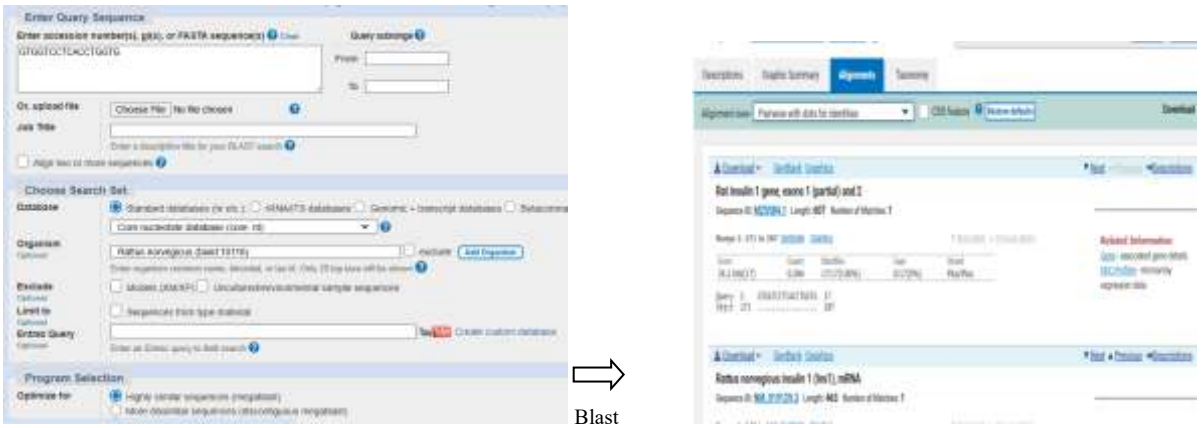


Fig. 3: MegaBLAST tool used to find distribution of mirror repeats in different genomes.

Results and Discussion

Mirror repetitions are found in numerous organisms, including bacteria, viruses, and plants, according to several previous studies (Yadav *et al*., 2022; Yadav *et al*., 2022). But no one selected the model organism *Rattus norvegicus* for their study for identification of mirror repeats. So, we have first time identified mirror repeats within the diabetic Glucagon (Gene Id: 24952), Sox17 (Gene Id: 312936), Gapdh (Gene Id: 24383) and PKLR (Gene Id: 24651) genes of *Rattus norvegicus*. In the present research work, a simple bioinformatic approach, was used to contrive mirror repeats present within the genes of *Rattus norvegicus* and its exons and introns. This recent developed method is very efficient to characterize mirror repeats. At the E-value of 20, a maximum number of hits were observed, so we fixed this value for identification of mirror repeat. Identified mirror repeats have different length varying from 7 base pairs to the largest mirror repeat of 58 base pairs. Further, the classification of identified mirror repeats was done on the basis of number of spacer element and the arrangement of nucleotides. Perfect mirror repeats have identical sequence around center of symmetry whereas imperfect mirror repeats have mismatch around center of axis. Perfect mirror repeats were further divided on the basis of spacer elements single spacer, double spacer and multispacer. Similarly, imperfect mirror repeats may also have a spacer or no spacer elements. The frequency of occurrence of imperfect mirror repeats within *Rattus norvegicus* gene was less than the frequency of perfect mirror repeats. A total of 199 mirror repeats were found in different regions of the Glucagon gene (Fig. 4) out of 199 mirror repeats, 122 were perfect with one spacer, 50 were perfect and 27 were imperfect repeats. A maximum number of hits 60 observed in fragment 4001-4500 and the corresponding mirror repeats were 13 with 6 perfect, 1 imperfect and 6 perfect with one spacer type.

We obtained 23 mirror repeats within exons and 79 mirror repeats within introns and of *Rattus norvegicus*' s Glucagon gene. The exon 1st, being the largest, contain 19 mirror repeats, whereas intron 2nd and 3rd, the largest have 23 mirror repeats as shown in Table 1.

Complete gene mirror repeats in the Sox17 gene:

A total of 199 mirror repeats were found in different regions of the Glucagon gene (Fig. 4) out of 199 mirror repeats, 122 were perfect with one spacer, 50 were perfect and 27 were imperfect repeats. A maximum number of hits 60 observed in fragment 4001-4500 and the corresponding mirror repeats were 13 with 6 perfect, 1 imperfect and 6 perfect with one spacer type..

We obtained 27 mirror repeats within exons and 35 mirror repeats within introns and of *Rattus norvegicus*'s Sox17 gene. The exon 5th, being the largest, contain 14 mirror repeats, whereas intron 3rd, the largest one has 16 mirror repeats as shown in Table 2.

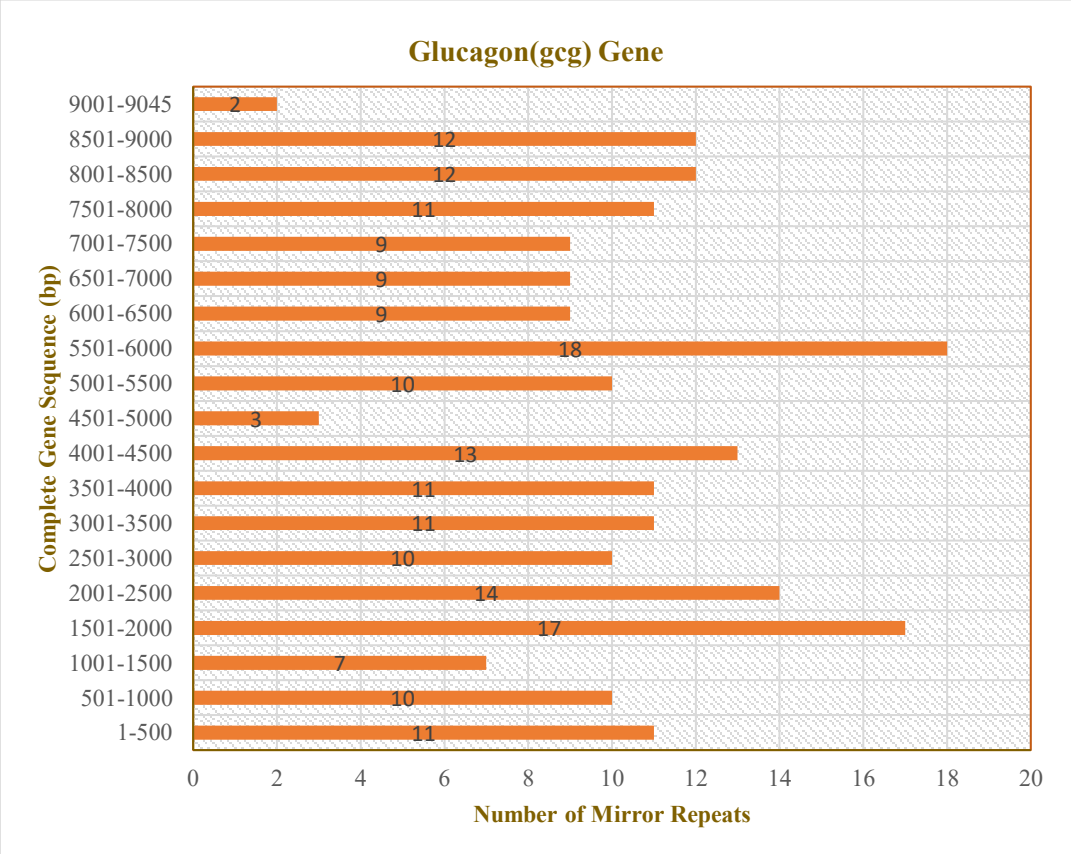


Fig. 4: Distribution of MR's (mirror repeats) in Glucagon gene.

Table 1: Total number of mirror repeats observed within the exons and introns of Glucagon gene

Exons	Length of Exons	Number of mirror repeats	Introns	Length of Introns	Number of mirror repeats
Exon1 (63-3134 bp)	3072	19	Intron 1(1-62 bp)	62	1
Exon2 (4886-5047 bps)	162	2	Intron 2 (3135-4885 bp)	1751	23
Exon 3(6616- 6753 bp)	138	0	Intron 3 (5048-6615 bp)	1568	23
Exon 4(7724- 7867 bp)	144	2	Intron 4 (6754-7723 bp)	970	19
Exon 5(8594-9045 bp)	452	0	Intron 5 (7868-8593 bp)	726	13

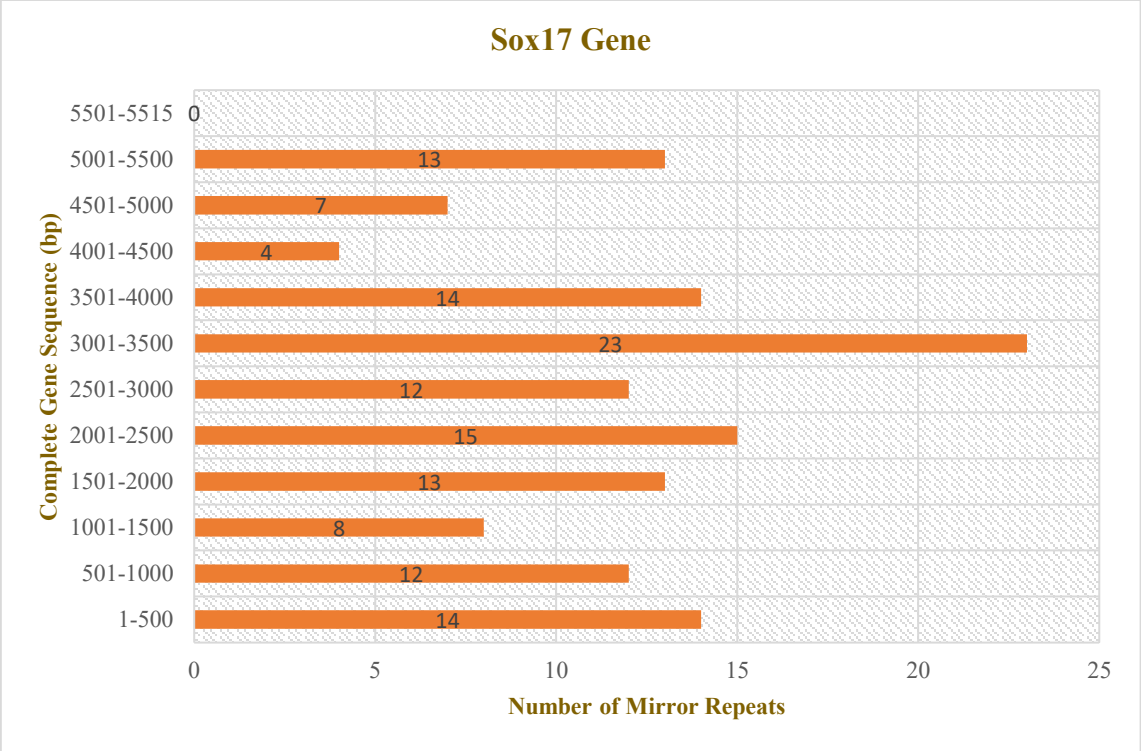


Fig. 5: Distribution of MR's (mirror repeats) in Sox17 gene.

Table 2. Total number of mirror repeats observed within the exons and introns of Sox 17 gene

Exons	Length of Exons	Number of mirror repeats	Introns	Length of Introns	Number of mirror repeats
Exon 1(1-140 bp)	140	0	Intron 1 (141-305 bp)	165	3
Exon 2 (306-677 bp)	372	0	Intron 2 (678-1008 bp)	331	5
Exon 3(1009-1296 bp)	288	0	Intron 3(1297-2967 bp)	1671	16
Exon 4(2968-3358 bp)	391	13	Intron 4(3359-3795 bp)	437	11
Exon 5(3796-5513 bp)	1718	14			

Complete Gene Mirror Repeats in The Gapdh Gene

A total of 116 mirror repeats were found in different regions of the Gapdh gene (Fig. 6). Out of 116, 70 mirror repeats were perfect with one spacer, 29 were perfect and 17 were imperfect repeats. A maximum number of hits 52 was observed in fragment 1501-2000 and the corresponding mirror repeats were 14 with 2 perfect, 2 imperfect and 10 perfect with one spacer type.

We obtained 26 mirror repeats within exons and 62 mirror repeats within introns and of *Rattus norvegicus*' s Gapdh gene. The exon 8th, being the largest, contain 9 mirror repeats, whereas intron 3rd, the largest one have 24 mirror repeats as shown in Table 3.

Complete gene mirror repeats in the PKLR gene:

A total of 167 mirror repeats were found in different regions of the PKLR gene (Fig.7). Out of 167, 131 mirror repeats were perfect with one spacer, 22 were perfect and 16 were imperfect repeats. A maximum number of hits 47 were observed in fragment 3501-4000 and the corresponding mirror repeats were 13 with 3 perfect, no imperfect and 10 perfect with one spacer type.

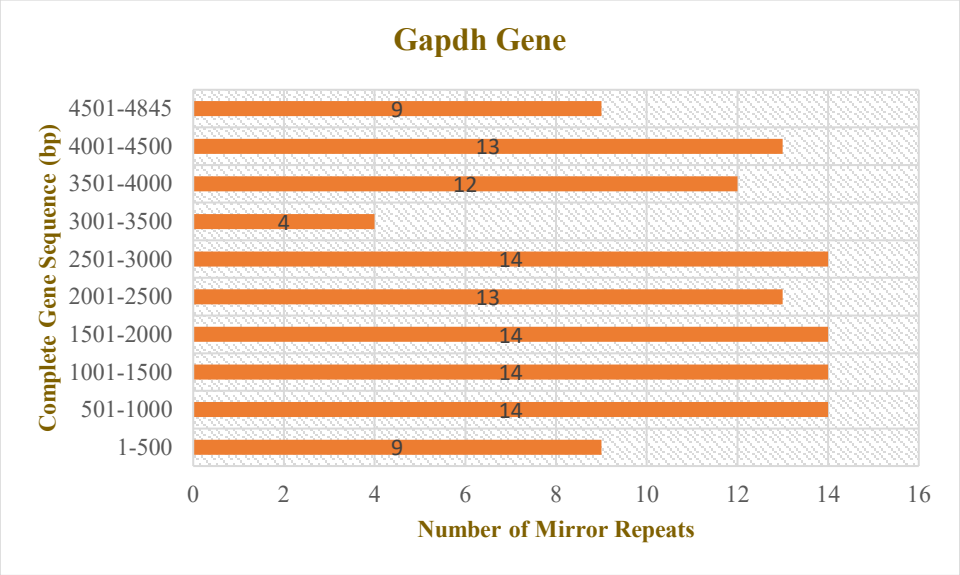


Fig. 6: Distribution of MR's (mirror repeats) in Gapdh gene.

Table 3: Total number of mirror repeats observed within the exons and introns of Gapdh gene

Exons	Length of Exons	Number of mirror repeats	Introns	Length of Introns	Number of mirror repeats
Exon 1(1-57 bp)	57	0	Intron 1 (58-972bp)	915	20
Exon 2 (973-1172 bp)	200	0	Intron 2 (1173-1260 bp)	88	2
Exon 3 (1261-1301 bp)	41	0	Intron 3(1302-3180 bp)	1879	24
Exon 4(3181-3387 bp)	207	3	Intron 4(3388-3467 bp)	80	1
Exon 5(3468-3558 bp)	91	2	Intron 5 (3559-3654 bp)	96	4
Exon 6 (3655-3852 bp)	198	4	Intron 6(3853-3964 bp)	112	1
Exon 7(3965-4195 bp)	231	7	Intron 7 (4196-4284 bp)	89	7
Exon 8(4285-4466 bp)	182	9	Intron 8 (4467-4576 bp)	110	3
Exon 9(4577-4845 bp)	269	1			

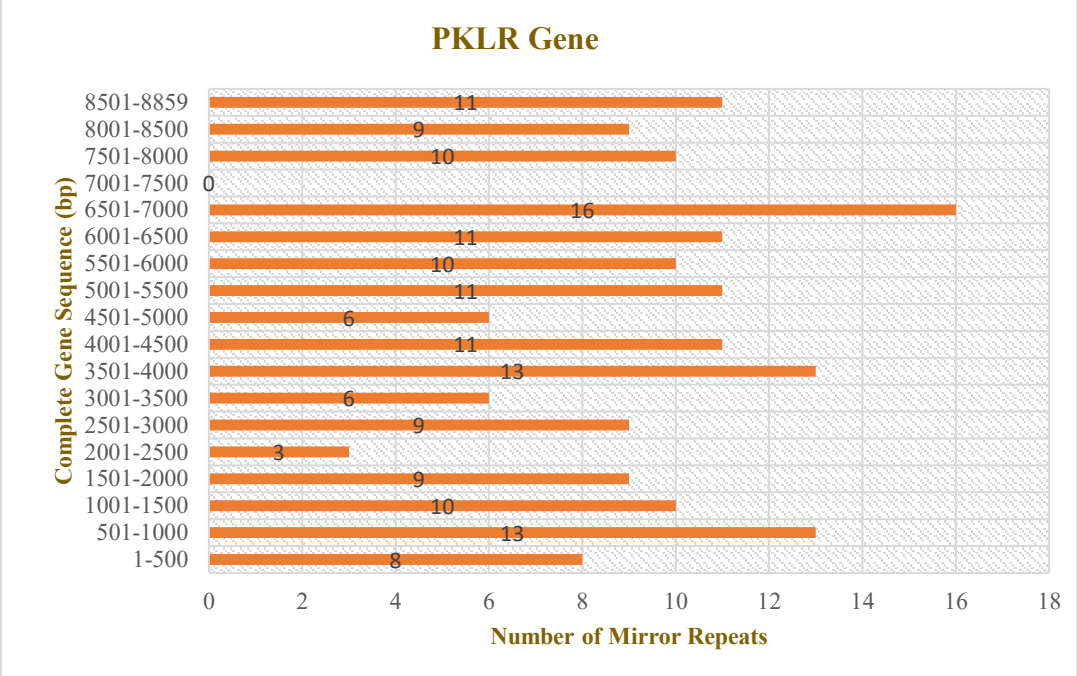


Fig. 7: Distribution of MR's (mirror repeats) in Gapdh gene.

Table 4. Total number of mirror repeats observed within the exons and introns of PKLR gene

Exons	Length of Exons	Number of mirror repeats	Introns	Length of Introns	Number of mirror repeats
Exon 1(1-134 bp)	134	0	Intron 1 (135-1091bp)	957	23
Exon 2(1092-1274 bp)	183	3	Intron 2(1275-3238 bp)	1964	12
Exon 3(3239-3330 bp)	92	0	Intron 3(3331-3471 bp)	141	2
Exon 4(3472-3603 bp)	132	1	Intron 4(3604-3737 bp)	134	5
Exon 5(3738-3924 bp)	187	3	Intron 5(3925-4240 bp)	316	13
Exon 6(4241-4511 bp)	271	2	Intron 6(4512-4623 bp)	112	2
Exon 7(4624-4774 bp)	151	1	Intron 7(4775-4986 bp)	214	1
Exon 8(4987-5139 bp)	153	2	Intron 8(5140-5232 bp)	93	3
Exon 9(5233-5399 bp)	167	4	Intron 9(5400-6619bp)	1220	24
Exon 10 (6620-6801 bp)	182	6	Intron 10(6802-7477 bp)	676	15
Exon 11 (7478-8859 bp)	1382	11			

We obtained 33 mirror repeats within exons and 100 mirror repeats within introns and of *Rattus norvegicus*'s PKLR gene. The exon 11th, being the largest, contain 11 mirror repeats, whereas intron 9th, the largest one have 24 mirror repeats as shown in Table 4.

Megablast is intended for comparing a query to closely related sequences and works best if the target percent identity is 95% or more. Using MegaBLAST tool, we further investigated identified mirror repeats among the genome of *Rattus norvegicus* and Homo sapiens of gene Glucagon in Table 5, Sox17 in Table 6, Gapdh in Table 7 and PKLR in Table 8. This result confirms the ubiquitous distribution of mirror repeats of size 10-14 base pairs in the genera mentioned. These repeat sequences must have some important role in the genome; therefore, they are being maintained during course of evolution. Here we wish to mention that we were unable to look distribution of mirror repeats having a length less than 7 or equal to 7 bps and even more than 7 within *Rattus norvegicus* genome as well as in another genome.

Table 5: Distribution of perfect mirror repeats of Glucagon gene among different genera, here +ve sign depicts presence of mirror repeats and –ve sign depicts absence of mirror repeats.

CDS length	Mirror repeats	Within Gene	Within *Rattus norvegicus* genome	Within *Homo sapiens* genome
1-500	TTTAATTT	+	-	-
	CCATTACC	+	-	-
501-1000	AAAACCAAAA	+	-	-
1001-1500	TTTTAATTTT	+	-	-
	TTATTTTATT	+	-	-
	AAAGGAAA	+	-	-
1501-2000	CTGTTTTGTC	+	-	-
	TGGAAGGT	+	-	-
	AGAAAAGA	+	-	-
	AAACCAAA	+	-	-
	TGAGGAGT	+	-	-
2501-3000	GATAAAATAG	+	-	-
	TATAATAT	+	-	-
3001-3500	ACCCCCCA	+	-	-
4001-4500	TAATTTTAAT	+	-	-
	TGTTTTTTGT	+	-	-
	CATTTTAC	+	-	-
	TTTAATTT	+	-	-
	ATTAATTA	+	-	-
4501-5000	ATGAAGTA	+	-	-
5501-6000	AACCTTCCAA	+	-	-
	GGAAAAGG	+	-	-
	AAAGGAAA	+	-	-
6001-6500	CTTGCCGTTC	+	-	-
	AGTCCTGA	+	-	-
	.TATCCTAT	+	-	-
-6501-7000	CTTGCCGTTC	+	-	-
	AGTCCTGA	+	-	-
	TATCCTAT	+	-	-
	TGAAGTT	+	-	-
7001-7500	AATAATAA	+	-	-
7500-8000	TTTTAGTTGATTTT	+	+	+

	CGATTAGC	+	-	-
8001-8500	TTAAAATT	+	-	-
	AATTTTAA	+	-	-
8501-9000	ACAGAAGACA	+	-	-
	GTCTTCTG	+	-	-
	TTGAAGTT	+	-	-
	AGTGGTGA	+	-	-

Table 6: Distribution of perfect mirror repeats of Sox17 gene among different genera, here +ve sign depicts presence of mirror repeats and –ve sign depicts absence of mirror repeats.

CDS length	**Mirror repeats**	**Within Gene**	**Within *Rattus norvegicus* genome**	**Within *Homo sapiens* genome**
1-500	GAGAAGAG	+	-	-
	GTAAAATG	+	-	-
501-1000	TTCCAACCTT	+	-	-
	TAAGGAAT	+	-	-
1501-2000	GTCTTTTCTG	+	-	-
	TCTGGTCT	+	-	-
2001-2500	TATTTTAT	+	-	-
	TTTAATTT	+	-	-
	AGTGGTGA	+	-	-
3000-3500	CCGGGTTGGGCC	+	+	+
	GCGGGGCG	+	-	-
	ACACCACA	+	-	-
3501-4000	CCCTTCCC	+	-	-
4001-4500	CCGGCCGGCC	+	-	-
	CAGGGGAC	+	-	-
5001-5500	GTTAATTG	+	-	-
	GTGGGGTG	+	-	-

Table 7: Distribution of perfect mirror repeats Gapdh gene among different genera, here +ve sign depicts presence of mirror repeats and –ve sign depicts absence of mirror repeats.

CDS length	Mirror repeats	Within Gene	Within *Rattus norvegicus* genome	Within *Homo sapiens* genome
1-500	CCCCGGGGCCCC	+	+	+
	GGCCCCGG	+	-	-
	CTTTTTTC	+	-	-
501-1000	CTCCCCTC	+	-	-
	GAGCCGAG	+	-	-
	GCGGGGCG	+	-	-
1001-1500	GAGGCCGGAG	+	-	-
	TCGGGGCT	+	-	-
	GACCCCAG	+	-	-
1501-2000	CCGCCGCC	+	-	-
	CTTCCTTC	+	-	-
2001-2500	CCCTTTTCCC	+	-	-
	TCTGGGGTCT	+	-	-
	GGGATTAGGG	+	-	-
	CCCAACCC	+	-	-
2501-3000	TTAAAGGAAATT	+	+	+
	ACAAAACA	+	-	-
3001-3500	TGTCCTGT	+	-	-
3501-4000	TCAGAAGACT	+	-	-
	AAGGGGAA	+	-	-
4001-4500	GCCTTCCG	+	-	-
	TGTTTTGT	+	-	-
	TATGGTAT	+	-	-
	TTTGGTTT	+	-	-
	GGACCAGG	+	-	-
4501-4845	CTGAGGAGTC	+	-	-
	AGAAAAGA	+	-	-
	GGTGGTGG	+	-	-

.

Table 8: Distribution of perfect mirror repeats of PKLR gene among different genera, here +ve sign depicts presence of mirror repeats and –ve sign depicts absence of mirror repeats

CDS length	Mirror repeats	Within Gene	Within *Rattus norvegicus* genome	Within *Homo sapiens* genome
1-500	CACGGGGCAC	+	-	-
	GTAGGATG	+	-	-
501-1000	TGGAAGGT	+	-	-
1001-1500	ATGTTGTA	+	-	-
1501-2000	GTTTTG	-	-	-
2501-3000	GAACCAAG	+	-	-
	GAAGGAAG	+	-	-
	AGGAAGGA	+	-	-
3001-3500	GTACCATG	+	-	-
3501-4000	TCCCAAAACCCT	+	+	+
	CTTTGGTTTC	+	-	-
	CACCCCAC	+	-	-
	CTCCCCTC	+	-	-
6001-6500	GAGAAGAG	+	-	-
	AGGTTGGA	+	-	-
6501-7000	TTCCCCTT	+	-	-
	AGTGGTGA	+	-	-
8001-8500	CCCCAACCCC	+	-	-
8501-8859	CAGGGGAC	+	-	-
	TCCCCCCT	+	-	-
	AATAATAA	+	-	-

These identified mirror repeats have any role in within the genome of *Rattus norvegicus* is still mysterious. However, to understand the function, our future goal is to identify proteins that bind to these identified mirror repeats. We strongly believe identification of mirror repeat binding proteins will open a new chapter of molecular biology.

Summary and Conclusion

Repetitive sequences play crucial roles in biological processes with both functional and non-functional implications. Certain repeats, like promoters and enhancer repeats, regulate gene expression, by acting as a binding site for regulatory proteins. They also serve as structural elements, such as centromeres and telomeres, which are vital for genome stability and cell division. By using a simple computational approach, we have identified 199 mirror repeats in Glucagon, 135 in Sox17, 116 in Gapdh and 167 in PKLR genes of *Rattus norvegicus*. These identified mirror repeats are not restricted only in the genome of *Rattus norvegicus*, but also scattered among the genome of *Homo sapiens*. The frequent existence of mirror repeats (shorter as well as larger length) in analyzed genes gives a hint about their significant roles in the genes or genomes of *Rattus norvegicus* and *Homo sapiens*. These types of studies will be exploring new trends and tools of molecular biology research as well as development of new concept for mirror repeat identification.

Author Contribution Statement

Parveen Kumar: Conceptualization, Methodology, Software, Supervisor, Reviewing, Editing.

Babli: Data curation, Investigation, Writing-Original draft preparation, Writing-Review Editing.

Conflict of Interest

Authors declare that they have no conflict of interest.

References

Bacolla A, Tainer JA, Vasquez KM, andCooper DN (2022). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic acids research* **44**: 5673-5688. DOI: 10.1093/nar/gkw261

Bhardwaj V, Swapni G, Sitaram M andKulbhushan S (2013). FPCB: A Simple and Swift Strategy for Mirror Repeat Identification *Preprint arXiv*: 1312.3869.

Biffi G, Tannahill D, McCafferty J and*Balasubramanian S (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem **5**: 182–186.* DOI: 10.1038/nchem.1548

Biscotti MA, Olmo E, andHeslop-Harrison JS (2015). Repetitive DNA in eukaryotic genomes *Chromosome Res.* **23**(3): 415-20.

Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R andOzelius LJ. (2017). Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proceedings of the National Academy of Sciences* **114**: E11020-E11028.

Brendel V, Beckmann JS Trifonov EN (1986) Linguistics of nucleotide sequences: morphology and comparison of vocabularies. *J Biomol Struct Dyn.* **4**:11–21.

Britten RJ and, Kohne DE. Repeated sequences in DNA (1968) Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**: 529-40.

Collier DA and Wells RD (1990) Effect of length, supercoiling, and pH on intramolecular triplex formation. Multiple conformers at pur.pyr mirror repeats. *Journal of Biological chemistry* **265**: 10652-10658.

Dahan D, Tsirkas I, Dovrat D, Sparks MA, Singh SP, Galletto R and Aharoni A (2018) Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res.* **46**(22): 11847-11857.

Depienne C and Mandel J L (2021). 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges?. The American Journal of Human *Genetics* **108**(5): 764-785.

Deveze-Alvarez M, Garcia-Soto J and Martinez-Cadena G (2001) Glyceraldehyde-3-phosphate Dehydrogenase is negatively regulated by ADP-ribosylation in fungus *Phycomycesblakesleeanus*. *Microbiology* **147**:2579-2584.

Du J, Campau E, Soragni E, Ku S, Puckett JW, Dervan PB and, Gottesfeld JM (2012) Role of mismatch repair enzymes in GAA· TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. *Journal of Biological Chemistry* **287**: 29861-29872.

Ezzatizadeh V, Sand, C, Sandi M, Anjomani-Virmouni S, Al-Mahdawi S and, Pook MA (2014). MutLα heterodimers modify the molecular phenotype of Friedreich ataxia. *PLoS One* **9**(6): e100523.

Gerhardt J, Bhalla A D, Butler JS, Puckett JW, Dervan PB, Rosenwaks Z and, Napierala M (2016) Stalled DNA replication forks at the endogenous GAA repeats drive repeat expansion in Friedreich's ataxia cells. *Cell reports* **16**:1218-1227.

Ghosh A and, *Bansal M (2003) A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr **59**: 620–626.*

Goddard NL, Bonnet G, Krichevsky O and*, Libchaber A. (2000) Sequence dependent rigidity of single stranded DNA. Phys Rev Lett **85**: 2400–2403.*

Hisey JA, Masnovo C and, Mirkin SM (2024) Triplex H-DNA structure: the long and winding road from the discovery to its role in human disease. *NAR Molecular Medicine* **1**(4): ugae024.

Hoede C, Denamur E and, Tenaillon O (2006) Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. *PLoS genetics* **2**: e176.

Htun H and, Dahlberg JE (1988) Single strands, triple strands, and kinks in H-DNA. *Science* **241**: 1791-1796.

Jurka J, Kapitonov VV, Kohany O and, Jurka MV (2007) Repetitive Sequences in Complex Genomes: Structure and Evolution. *Annual Review of Genomics and Human Genetics* **8**: 241-259.

Kharroubi AT and, *Darwish HM (2015) Diabetes mellitus: the epidemic of the century. World J Diabetes **6**:850–867.*

Khristich AN and, Mirkin SM (2020). On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *Journal of Biological Chemistry* **295**(13): 4134-4170.

Korotkov E, Suvorova Y, Kostenko D and, Korotkova M (2023) Search for Dispersed Repeats in Bacterial Genomes Using an Iterative Procedure. *Int J Mol Sci.* **24**:10964.

Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R and*, Przytycka TM (2017) Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. Cell Syst. **4**: 344–356.*

Lafferty RA, O'Harte FP, Irwin N and, Gault V A (2021). Flatt P R. Proglucagon-derived peptides as therapeutics. *Frontiers in endocrinology* **12**:689-678.

Li L, Scott WS, Khristich AN, Armenia JF and, Mirkin SM (2024) Recurrent DNA nicks drive massive expansions of (GAA)n repeats. *Proceedings of the National Academy of Sciences* **121**(49): e2413298121.

Lyamichev VI, Mirkin SM, Kumarev VP, Baranova LV, Vologodskii AV and, Frank-Kamenestskii MD (1989). Energetics of the B–H transition in supercoiled DNA carrying d (CT) xd (AG) x and d (C) nd (G) n inserts. *Nucleic acids research* **17**(22): 9417-9423.

Masnovo C, Paleiov Z, Dovrat D, Baxter L K, Movafaghi S, Aharoni A and, Mirkin S M (2024) Stabilization of

expandable DNA repeats by the replication factor Mcm10 promotes cell viability. *Nature Communications* **15**(1): 10532.

Matos-Rodrigues G, Van Wietmarschen N, Wu W, Tripathi V, Koussa NC, Pavani R and, Nussenzweig A (2022). S1-END-seq reveals DNA secondary structures in human cells. *Molecular cell* **82**(19):3538-3552.

McDonald E, Krishnamurthy M, Goodyer CG and, Wang R (2009) The emerging role of SOX transcription factors in pancreatic endocrine cell development and function. *Stem Cells Dev.* **18**:1379–1388.

Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA and, Frank-Kamenetskii MD (1987). DNA H form requires a homopurine–homopyrimidine mirror repeat. *Nature* **330**(6147): 495-497.

Modlinska K and, Pisula W (2020) The Norway rat, from an obnoxious pest to a laboratory pet. *Elife* **9**: e50651.

Nag DK and, *Petes TD (1991) Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in Saccharomyces cerevisiae. Genetics **129**: 669–673.*

Pellerin D, Danzi MC, Wilke C, Renaud M, Fazal S, Dicaire M J and, Brais B (2023) Deep intronic FGF14 GAA repeat expansion in late-onset cerebellar ataxia. New England *Journal of Medicine* **388**(2): 128-141.

Rafehi H, Szmulewicz DJ, Bennett MF, Sobreira NL, Pope K, Smith KR and, Lockhart PJ (2019). Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS. *The American Journal of Human Genetics* **105**(1): 151-165.

Rastokina A, Cebrián J, Mozafari N, Mandel NH, Smith CE, Lopes M and, Mirkin SM (2023) Large-scale expansions of Friedreich's ataxia GAA• TTC repeats in an experimental human system: role of DNA replication and prevention by LNA-DNA oligonucleotides and PNA oligomers. *Nucleic acids research* **51**(16): 8532-8549.

*Rayburn WF (1997). Diagnosis and classification of diabetes mellitus: highlights from the American Diabetes Association. J Reprod Med. **42**(9):585–586.*

Ruggiero BL and, Topal MD (2004). Triplet repeat expansion generated by DNA slippage is suppressed by human flap endonuclease 1. Journal of Biological Chemistry **279**(22): 23088-23097.

Saini K, Dangi N, Deepti Y and, Bhardwaj V (2023) Identification of mirror repeats within the maleless(mle)gene of Drosophila melanogaster meigen. Indian Journal of Entomology **85**(1): 14-18.

Sinden RR, Zheng GX, Brankamp RG and, *Allen KN (1991) On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. Genetics **129**: 991–1005.*

Singleton CK, Klysik J, Stirdivant SM and, Wells RD (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* **299**: 312-316.

Tang W, Dominska M, Greenwell PW, Harvanek Z, Lobachev KS, Kim HM and, Petes TD (2011) Friedreich's Ataxia (GAA) n•(TTC) n repeats strongly stimulate mitotic crossovers in Saccharomyces cerevisae. *PLoS genetics* **7**: e1001270.

Todd P A and, Glickman BW (1982) Mutational specificity of UV light in Escherichia coli: indications for a role of DNA secondary structure. Proceedings of the National Academy of Sciences; **79** (13): 4123-4127.

Tsutakawa SE, Thompson MJ, Arvai AS, Neil A J, Shaw SJ, Algasaier SI and, Tainer JA (2017). Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. *Nature communications* **8**(1): 15855.

Unnikrishnan R, Anjana RM and, *Mohan V (2016) Diabetes mellitus and its complications in India. Nat Rev Endocrinol. **12**(6):357–370.*

Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, Van Boom JH, Van Der Marel G and, Rich A (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature **282**: 680-686.

Wang G andVasquez KM (2023) Dynamic alternative DNA structures in biology and disease. Nature Reviews Genetics; **24**(4): 211-234.

*Wells RD (2007) Non-B DNA conformations, mutagenesis and disease. Trends Biochem Sci **32**: 271–278.*

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):93–97.

Yadav U, Yadav S and Sharma DC (2022) In Silico Analysis of Structural Photosynthetic Genes of Arabidopsis thaliana for Unique Mirror Repeats. Indian Journal of Science and Technology **15** (3):127–135.

Yamada K and Noguchi T (1999) Nutrient and hormonal regulation of pyruvate kinase gene expression. *Biochem J.* **337**:1–11.