INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

In Silico Analysis of Structural Photosynthetic Genes of *Arabidopsis thaliana* for Unique Mirror Repeats

Usha Yadav[1], Sandeep Yadav[1], Dinesh C Sharma[1]*

1 School of Life Sciences, Starex University, Gurugram, India (122413) *

OPEN ACCESS

Received: 01.10.2021

Accepted: 13.01.2022

Published: 04.02.2022

Citation: Yadav U, Yadav S, 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. https://doi.org/10.17485/IJST/v15i3.1833

* **Corresponding author**.

ddcsharma@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: The underlying work explores mirror sequences in the photosynthetic genes of *Arabidopsis thaliana*. At present, these sequences are standing at the forefront to be explored for their origin, distribution and function in plants. **Methods**: FPCB, a recently developed bioinformatics approach was utilized for identification of mirror sequences. It is a three step strategy based on pattern matching of alignments, produced after aligning gene sequence and its complement using mega-BLAST. This algorithm was quick and efficient enough to characterize a range of mirror sequences. **Findings**: All the analyzed genes were reported to harbor great variety of mirror sequences at quite high frequencies. LHCA1 gene have the highest total count of these sequences and ATPB gene have lowest of all. A total of 401 unique mirror sequences of different lengths and compositions were reported in the twelve selected genes. Promoter motifs were found to be greatly enriched with these repeats. Eleven mirror sequences of significant lengths were also reported using the above approach. **Novelty**: This work is the very first attempt to characterize photosynthetic genes of *Arabidopsis thaliana* for mirror repeats. This will further aggravate efforts to develop fingerprinting techniques based on these unique mirror sequences, which are very powerful tools to study taxonomic and evolutionary relationships. Mirror sequences are also potential candidates as drug delivery systems and in molecular medicine.

Keywords: Mirror repeats; HDNA; FPCB; photosynthetic genes; Arabidopsis thaliana

1 Introduction

The DNA code is the template which directs the organization and function of a complete organism by directing synthesis of range of macromolecules[1,2]. Repetitive DNA sequences gradually gets accumulated into the genomes during the course of evolution because of non-repairable variations arising from events like mutations, slipped strand mispairing and unequal crossing[3,4]. The term repetitive DNA sequences collectively represent the DNA fragments which are present in multiple copies into the genomes[5]. Despite their conscious presence, they were regarded as junk for a while but now they have well-established roles in a number of basic genetic processes[6–8].

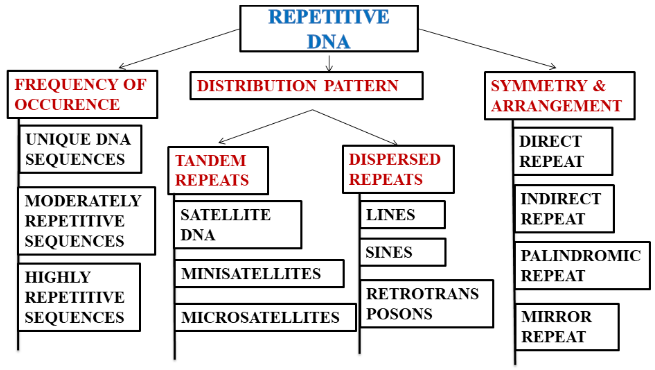

Fig 1. Categorization of repetitive DNA elements based on frequency of occurrence, distribution patterns and symmetry

Eukaryotic genomes are differentially enriched in a variety of repetitive DNA sequences[9]. When these repetitive elements are clustered together nearly at the same patch, they are termed as tandem repeats or if dispersed throughout the genome at irregular intervals, they are termed as dispersed repeats. Dispersed repeats being more common, can be further classified as LINES (Long Interspersed Elements, with length ≈1-7kbps) and SINES (Short Interspersed Elements, with length ≈100-400bps)[10,11].

If one considers symmetry or arrangement of nucleotides as a classifying criterion, three prominent types of DNA repeats can be observed namely, direct, indirect and mirror repeats. When a simple core repeat unit is duplicated downstream uninterruptedly, it represents a direct repeat. When genomic segments duplicated downstream but on the opposite strand, this represents an inverted repeat[12]. According to Mirkin and his coworkers, mirror repeats are DNA segments in which DNA bases that are equidistant from the symmetry center are identical to each other.

All of these symmetrical DNA repeats are capable of adopting a wide range of alternative or Non-B DNA conformations[13]. At inverted repeat motifs, cruciform and hair-pin like structures are formed, depending upon the DNA characteristics[14]. Direct repeats being the simplest ones are associated with formation of a diversity of alternative structures like Z-DNA at sequences of alternating purine-pyrimidine, slippage structures, H-DNA and G-quartet with varying compositions and topological conditions[15].

Homo-purine:homo-pyrimidine rich DNA tracks with mirror symmetries (or H-Palindromes) is reported to form a significant triplex structure known as H-DNA in the negatively supercoiled state[16]. H-Palindromes are over represented in eukaryotic genomes but occur at change frequencies in the prokaryotic genomes[9]. In the light of earlier investigations, it was hypothesized that H-DNA could act as a molecular switch to modulate the gene expression and many cellular proteins might be stabilizing them specifically[17]. These structures are known to act as regulatory elements in transcriptional, site-specific mutagenesis or recombination and replication and other DNA-metabolism events in a structural dependent manner[18]. They also act as mutational hotspots and repeat induced mutagenesis in one of the well exploited mechanism in evolution of species[19,20].

Above line of evidences supports the notion that H-DNA structures are functionally important to genome function, and hence maintained in the genome because of positive selection pressure during evolution. But genomic instability caused by them in the coding and regulatory regions results in the development of a number of disorders like Autosomal Dominant Polycystic Kidney Disease (ADPKD), Tuberous Sclerosis Complex (TSC), Lymphangioleiomyomatosis (LAM), Friedreich's ataxia, Follicular Lymphoma, Hereditary Persistence of Fetal Hemoglobin (HPFH)[21–23]. Hence, it would be tempting to speculate their origin and function in the genome. But on the very first front, mirror sequence motifs which can form H-DNA, need to be identified thoroughly in the genomes. This can also help in understanding the molecular mechanisms resulting in disease development in human beings.

DNA tracks with mirror symmetries were reported in many eukaryotic species ranging from yeast to humans[24,25], chloroplast genome of *Nicotiana tabaccum* plants[26–28] and recently in the deadly SARS-CoV-2 viral strain[29]. Our previous work has also reported the presence of mirror repeats in the flowering genes of *Arabidopsis thaliana*[30] and in the genome of HIV[31]. This was huge in terms of their ubiquitous presence across all eukaryotic kingdoms and their parasites also. The functional aspect of H-DNA is still not investigated in plants. This study is the very first attempt to unmask the story of mirror repeats in plants. For this purpose, the most beloved model plant for genetics and molecular bioinformatics *Arabidopsis thaliana* was chosen. This is the best match for above study since, it has small genome which is completely sequenced.

FPCB (Fast-Parallel Complement-Blast), a simple, accurate and swift manual bioinformatics based approach was deployed to carry out the present investigation. It is a three step strategy based on pattern matching of alignments, produced after aligning gene sequence and its complement using megaBLAST, to extract mirror repeats [32]. This work will prove very significant in order to further study the function and evolution of mirror repeats in plants and development of range of many fingerprinting technologies based on them. They may prove as key tools in taxonomic and evolutionary studies. If it would be possible in near future, to selectively modulate site specific gene expression and H-DNA formation, through these sequences, it would prove very significant in the field of molecular medicine and drug-delivery systems.

2 Material & Methods

A total of 12 structural genes involved in formation of major subunits of photosynthetic apparatus in *Arabidopsis thaliana* were selected, at least one from each subunit (PSII, PSI, ATP synthase, Cytochrome complex) [33].

S.N.	Gene Symbol	Gene Name	Description/ subunit involve	Function	Location	NCBI Gene ID	Size
1	psbA	photosystem II protein D1	Subunit of photosystem II	reaction centre protein of photosystem II	Chr-genome	844802	1.06kb
2	psbD	photosystem II protein D2	Subunit of photosystem II	reaction centre protein of photosystem II	Chr-genome	844775	1.06kb
3	LHCA1	chlorophyll a-b binding protein 6	photosystem I light harvesting complex gene 1	Encodes a component of the light harvesting complex associated with photosystem I.	Chromosome 3	824654	1.67kb
4	LHCB2.1	LIGHT-HARVESTING CHLOROPHYLL B-BINDING 2	photosystem II light harvesting complex protein 2.1	Lhcb2.1 protein encoding a subunit of the light harvesting complex II.	Chromosome 2	815058	1.49kb
5	atpA	ATP synthase CF1 alpha subunit	ATP synthase CF1 alpha subunit	Encodes a subunit of ATPase synthase machinery	Chr-genome	844790	1.52kb
6	atpB	ATP synthase CF1 beta subunit	ATP synthase CF1 beta subunit	Encodes a subunit of ATPase synthase machinery	Chr-genome	844757	1.50kb
7	ATPC1	ATPase, F1 complex, gamma subunit protein	ATPase, F1 complex, gamma subunit protein	One of two genes (with ATPC2) encoding the gamma subunit of Arabidopsis chloroplast ATP synthase.	Chromosome 4	825797	1.53kb
8	psaA	photosystem I P700 chlorophyll a apoprotein A1	photosystem I P700 chlorophyll a apoprotein A1	Encodes a subunit of photosystem I, transmembrane protein on thylakoid membrane	Chr-genome	844768	2.25kb
9	psaB	photosystem I P700 chlorophyll a apoprotein A2	photosystem I P700 chlorophyll a apoprotein A2	Encodes a subunit of photosystem I, transmembrane protein on thylakoid membrane	Chr-genome	844770	2.21kb
10	petA	cytochrome f	cytochrome f	Integral part of Z-scheme of photosynthesis	Chr-genome	844748	963bp
11	petB	cytochrome b6	cytochrome b6	Integral part of Z-scheme of photosynthesis	Chr-genome	844729	1.45kb
12	PETE1	plastocyanin 1	plastocyanin 1	One of two Arabidopsis plastocyanin genes. Expressed at 1/10th level of PETE2.	Chromosome 1	843942	1.05kb

*Chr-genome represents chloroplast genome

Above listed gene sequences were analyzed for presence of mirror repeats using FPCB approach. Originally developed by Vikash et al, FPCB stands for Fasta-Parallel Complement-BLAST. It is a simple three step bioinformatics based strategy for extracting both short and long mirror sequences through a pattern matching algorithm. The strategy is based on the principle that during parallel DNA PCR with a parallel primer in the reaction mixture, the resulting product is the original template DNA but with a reversed orientation. With this basic idea, if we align the original gene sequence with its parallel complement using BLAST analysis, many alignments produced were shown to have mirror symmetries [34]. Please refer to fig.2 for pictorial representation of methodology.

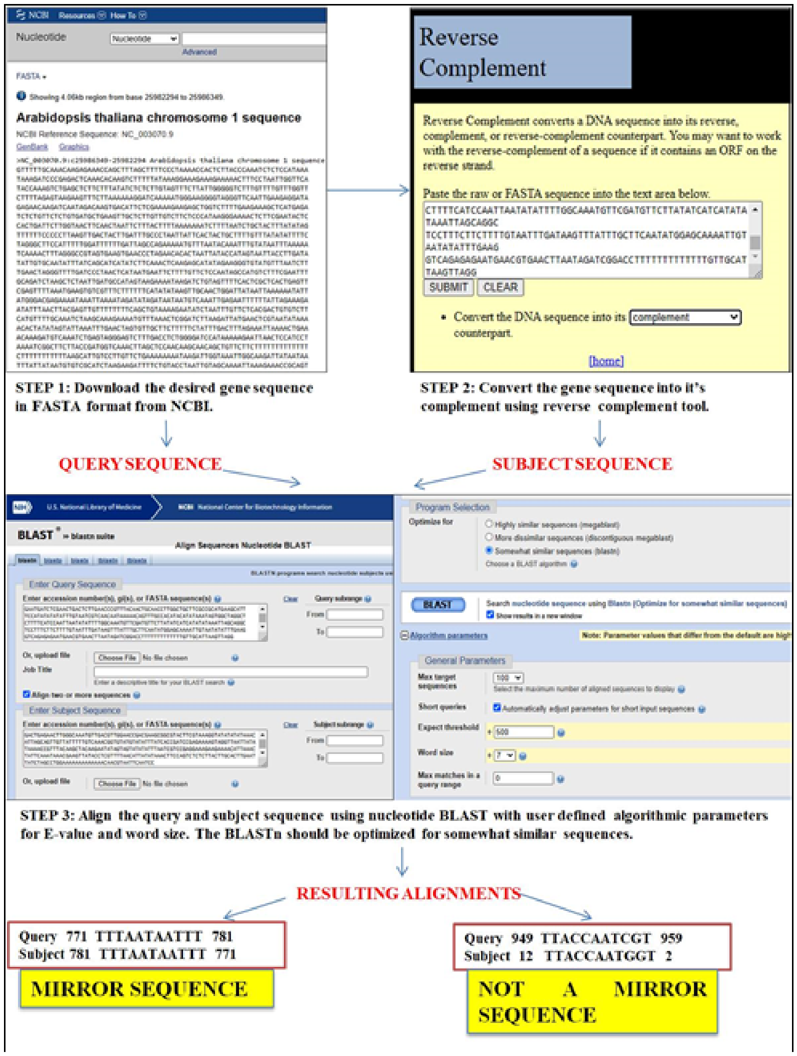

Fig 2. Screenshots of the methodology showing the exact workflow of algorithm

STEP 1: DOWNLOADING GENE SEQUENCE

The complete CDS of gene sequence was downloaded in FASTA format from the public domain NCBI (https://www.ncbi.nlm.nih.gov/) [35]. The complete gene was divided into smaller regions of 1000bps each, to extract maximum number of repeats. Single such region will be termed as query sequence in underlying paper.

STEP 2: CONVERTING QUERY SEQUENCE TO PARALLEL COMPLEMENT

The parallel complement of query sequence was then extracted using the online available bioinformatics Reverse Complement Tool (https://www.bioinformatics.org/sms/rev_comp.html) [36]. The parallel complement will then represent the subject sequence.

STEP 3: ALIGNMENT OF QUERY AND SUBJECT SEQUENCE

The query sequence and subject sequence stated in the above steps were aligned for homology through the most prominently used bioinformatics domain Align Sequences Nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABAS

E=n/a&QUERY=&SUBJECTS)[37]. For the present study, the program was optimized for somewhat similar sequences with word size 7 and expected frequency or E-value at 100.

2.1 Identification of mirror repeats

Identification of mirror repeats was carried out on the basis of Pattern Matching Algorithm: If the position number of alignments is exactly reversed in subject and query sequence, it will be a mirror sequence. The total number of mirror sequences present in the query sequence/gene sequence can also be confirmed through Dot Matrix Plot. In such plots, alignments were represented as dots over a matrix. The number of dots present on the diagonal depicts the total number of mirror alignments present in the analyzed sequence.

3 Results & Discussion

The present study is the very first attempt to study mirror repeat sequences in the coding regions of *Arabidopsis thaliana*, by utilizing a very simple recently developed bioinformatics strategy. Many previous studies have talked about predominance of mirror repeats in many organisms including viruses, animals and plants[38]. But none of them targeted the model plant organism, which is also a valuable bioinformatics resource. In the lieu of recent studies, this work also confirms the over representation of mirror repeats in the photosynthetic genes of Arabidopsis. Identification of such sequences will further paved the way to study their roles in the plant kingdom and how they are evolving continuously inside the genomes.

3.1 Distribution of Mirror repeats in the photosynthetic genes

The present decade have concretely established role of many different DNA repeats including the mirror repeats. Mirror repeats which are rich in purines and pyrimidine's are capable of adopting an alternative DNA secondary structure, the H-DNA. This very structure is proving to be a promising candidate for drug-delivery systems in molecular medicine because of its ability to modulate site-specific gene expression. Hence, the present investigation has worked upon identification and distribution of mirror repeats in the model plant to further carry out research on the functional aspects of mirror repeats in plants. *Arabidopsis thaliana* has a compact genome of ~135bps, which relatively lacks repetitive DNA segments as compared to other plant genomes[39]. But our study, has confirmed the overabundance of mirror repeats in all the studied genes of Arabidopsis. This might be indicating that how important these repeats are to the genome, as these are very well conserved during very high evolutionary pressure for compact genome size and shorter cell cycle in Arabidopsis. For the first time, we have reported the occurrence of a total of 401 unique mirror repeat sequences of varying lengths and composition in the photosynthetic genes of Arabidopsis. Studies have already documented that in Arabidopsis, the chloroplast circular DNA composed of 154,478 bp and have inverted repeats in it[40]. Since, many photosynthetic genes were majorly located on the chloroplast DNA itself, the present study confirms the presence of mirror sequences on it also.

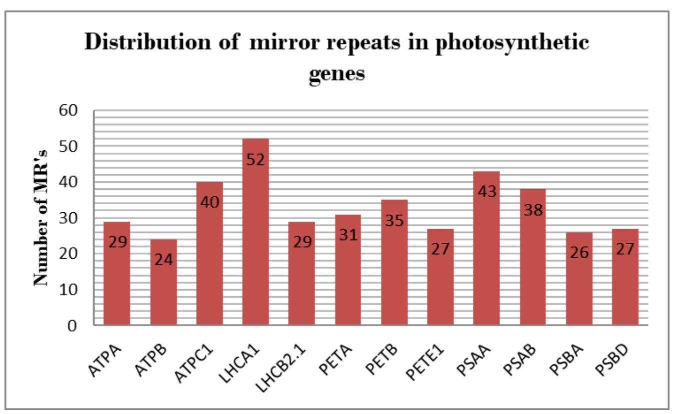

Fig 3. Distribution of MR's (mirror repeats) in all the selected genes

As per our analysis, LHCA1 harbors the maximum numbers of MR's and ATPB have the minimum count even having a similar gene size (the gene length of LHCA1is 1.6kb and that of ATPB is 1.5Kb). Hence, gene size is not a determining factor for abundance of MR's. But this might be because of their functional differences because LHCA1 is a very prominent protein subunit of light harvesting complex associated with photosystem 1 and ATPB encodes the beta subunit of ATPase synthase machinery. Irrespective of their locations, whether situated on chromosomes or on chloroplast DNA, all the genes have comparable number of mirror sequences. This was in line with the already stated results about enrichment of mirror sequences in the eukaryotic genomes. All the mirror sequences were randomly distributed with no definite pattern into the gene sequences.

Table 1. Depicting the frequencies of mirror sequences in the respective regions of the genes studied

S.N.	Symbol/Gene ID	Region	No. of mirror sequences in size range				TOTAL
			(7-12bp)	(13-18bp)	(19-24bp)	(≥25bp)	
1.	atpA/844790	1-1000bp	20	1	0	1	22
		1000-1517bp	6	1	0	0	7
2.	atpB/844757	1-1000bp	15	0	0	0	15
		1000-1497bp	7	1	0	1	9
3.	ATPC1/825797	1-1000bp	20	3	3	0	26
		1000-1528bp	11	1	2	0	14
4.	LHCA1/824654	1-1000bp	21	2	2	3	28
		1000-1673bp	20	0	2	2	24
5.		1-1000bp	19	0	0	1	20
	LHCB2.1/815058	1000-1491bp	6	2	0	1	9
6.	petA/844748	1-963bp	23	5	2	1	31
7.	petB/844729	1-1000bp	19	5	2	0	26
		1000-1452bp	8	0	0	1	9
8.	PETE1/843942	1-1048bp	23	2	2	0	27
9.	psaA/844768	1-1000bp	18	2	1	2	25
		1000-2000bp	9	2	0	1	12
		2000-2253	5	0	1	0	6
10.	psaB/844770	1-1000bp	18	2	0	0	20
		1000-2000bp	13	1	0	1	15
		2000-2205bp	3	0	0	0	3
11.	psbA/844802	1-1006bp	23	3	0	0	26
12.	psbD/844775	1-1006bp	22	5	0	0	27
TOTAL= 401							

The novel strategy deployed in exploring mirror symmetries, FPCB, have yielded a combination of both short and long mirror sequences with size ranging from 7-50 bps. These sequences were characterized on the basis of presence of spacer nucleotide in them. Imperfect mirror sequences were more common as compared to perfect sequences (41). Single spacer, double spacer and multi-spacer mirror sequences were a part of imperfect group.

Highest density of mirror symmetries was present in the regions of size ranging from 1-1000bp, which typically corresponds to the promoter domains of the genes. Their enrichment in the promoter regions of the genes supports the notion that they are involved in replication and transcriptional regulation (42). As we go downstream into the regions of gene, their abundance generally decreases.

Identified mirror sequences of significant lengths can be worked out for their taxonomic distribution and evolutionary relevance. Only Sequences of length ≥28bp were considered significance for further studies because they are taken according to the default parameters (word size value) of megablast analysis (43). Eleven such sequences were identified in six genes as depicted in Table 2.

Out of the twelve genes studied, only six genes namely atpA, LHCA1, LHCB2.1, petB, psaA, psaB were reported to have MR's of significant lengths. Remaining others has shorter sequences of length less than 28bps. The longest mirror sequence was identified in the LHCB2.1 gene with a length of 54bps. It was an imperfect mirror repeat with multi-spacer nucleotides. Perfect symmetries were common only among shorter mirror sequences. None of these identified sequences of significant lengths was a perfect mirror repeat. All of them are imperfect mirror repeats. These sequences can be effectively used for profiling genes and even whole genomes, which paves a way towards new criterion of classification of genomes. Many questions are still open to research like whether a particular mirror sequence is only located over a certain region or it is dispersed throughout the

Table 2. Representing the actual mirror sequence alignments of significant length*

S.N.	Genes	Mirror repeats of significant length (≥28bp)	Length (bp)
1.	atpA	867…AGATGTTTTTTATTTACATTCACGTCTTTTAGA…899	33
2.	LHCA1	461…TTTTACTC—TGTTTTGTTGCTCAGTTT…486	28
		50…ACAGAAGAAC-ATGATATCGAGAAGACA…76	28
		515…TGTTTTGATGACATGATGCAACATTGTAC—-GTTTTGT…550	40
		197…CAAGAAGCTCGAGGAATTGAAAGTTAAAGAGATC-AAGAAC…236	41
		487…AAATATGTGTACCTTATGAGCTTT—ATGTGTATCAA…581	37
3.	LHCB2.1	191…AACGGCTC-TCAAGC—CCTCCAACGAACTCCTCCGCAA…226	38
		89…ATTGGTGTTTTATGTGAGTTTGGTACTTAATGATTTT—-TATTTTGTGGGTA…138	54
4.	petB	207…GGTTATTCTTTAC-CTTGGGATCAAATTGGTTATTGG…242	37
5.	psaA	177…TACCAGTGATTTGGAGGAAATCTCTCGAAAAGTA-TTTAGTGCCCAT…222	47
6.	psaB	109…CGTTATATACCCATCACCAATATATTGC…136	28

genome or chromosomes? How is it distributed in the homologous genes or organisms related to plants and even in other higher eukaryotes also? What mechanisms are causing them to persist in constantly evolving genomes? How are they affecting plants? And many more.

4 Conclusion

The present work has reported the occurrence of mirror repeats in the photosynthetic genes of *Arabidopsis thaliana* for the very first time. Direct and inverted repeats were earlier noted in the genomes of Arabidopsis. The presence of all of these simple DNA repeats in the genome of Arabidopsis signifies their importance in the plant genomes. But by which mechanisms they are doing so, is still an unanswered question. The present *in-silico* analysis stated a total of 401 mirror sequences in the twelve photosynthetic genes of different lengths using FPCB technique. Amongst them, eleven mirror sequences of significant lengths were traced out in the six genes. Shorter repeats of length 7-12bp were quite abundant. Irrespective of the sizes, these sequences were present in each of the studied gene. The recently developed bioinformatics strategy FPCB, was quite efficient in terms of extracting a large number and varieties of mirror repeats. A new bioinformatics based software can be developed in the near future based on the above used algorithm. There predominance points out towards their conservative nature in genes and hence predicts their important roles in evolution of genome and chromosome architect. Some of these sequences which are enriched in purines or pyrimidine's might adopt H-DNA conformation and represents a plausible target for gene editing and drug deliveries. There are still many questions open to plant biologists relating to role of mirror sequences and H-DNA in plants. It is still to be explored, whether all of the above reported mirror sequences are capable of adopting the H-DNA conformation and under what specific conditions? How are they modulating gene function? This study opens a new forefront to plant biotechnologist to study about mirror repeats in *Arabidopsis thaliana*.

5 Acknowledgement

We would like to thank Shri Mohinder Singh ji (Hon'ble Chancellor, Starex University) for providing facility to perform present research. We also want to express our gratitude to Prof.(Dr.) M. M. Goel (Hon'ble Vice-Chancellor, Starex University) for providing all time courage and support. We also thank Dr.Vikash Bhardwaj for introducing the concept of FPCB to us.

References

1) Minchin S, Lodge J. Understanding biochemistry: structure and function of nucleic acids. *Essays in Biochemistry*. 2019;63(4):433–456. Available from: https://dx.doi.org/10.1042/ebc20180038.
2) Brazda V, Fojta M, Bowater RP. Structures and stability of simple DNA repeats from bacteria. *Biochemical Journal*. 2020;477(2):325–339. Available from: https://dx.doi.org/10.1042/bcj20190703.
3) Harhay GP, Harhay DM, Bono JL, Capik SF, DeDonder KD, Apley MD, et al. A Computational Method to Quantify the Effects of Slipped Strand Mispairing on Bacterial Tetranucleotide Repeats. *Scientific Reports*. 2019;9(1):18087–18087. Available from: https://dx.doi.org/10.1038/s41598-019-53866-z.
4) Gralak E, Faria MV, Figueiredo AST, Rizzardi DA, Neumann M, Mendes MC, et al. Genetic divergence among corn hybrids and combining ability for agronomic and bromatological traits of silage. *Genetics and Molecular Research*. 2017;16(2). Available from: https://dx.doi.org/10.4238/gmr16029643.
5) Berselli M, Lavezzo E, Toppo S. NeSSie: a tool for the identification of approximate DNA sequence symmetries. *Bioinformatics*. 2018;34(14):2503–2505. Available from: https://dx.doi.org/10.1093/bioinformatics/bty142.

6) Arancio W, Coronnello C. Repetitive sequences in aging. *Aging*. 2021;13(8):10816–10817. Available from: https://dx.doi.org/10.18632/aging.203020.
7) Carta A, Bedini G, Peruzzi L. A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytologist*. 2020;228(3):1097–1106. Available from: https://dx.doi.org/10.1111/nph.16668.
8) Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. *Biological Reviews*. 2005;80(2):227–250. Available from: https://dx.doi.org/10.1017/s1464793104006657.
9) Sergei MM. Discovery of alternative DNA structures: a heroic decade (1979-1989). *Frontiers in Bioscience*. 2008;13(13):1064–1064. Available from: https://dx.doi.org/10.2741/2744.
10) Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive Sequences in Complex Genomes: Structure and Evolution. *Annual Review of Genomics and Human Genetics*. 2007;8(1):241–259. Available from: https://dx.doi.org/10.1146/annurev.genom.8.080706.092416.
11) Bustos AD, Cuadrado A, Jouve N. Sequencing of long stretches of repetitive DNA. *Scientific Reports*. 2016;6(1):36665–36665. Available from: https://dx.doi.org/10.1038/srep36665.
12) Bissler JJ. DNA inverted repeats and human disease. *Frontiers in Bioscience*. 1998;3(4):d408–418. Available from: https://dx.doi.org/10.2741/a284.
13) Natale F, Scholl A, Rapp A, Yu W, Rausch C, Cardoso MC. DNA replication and repair kinetics of Alu, LINE-1 and satellite III genomic repetitive elements. *Epigenetics & Chromatin*. 2018;11(1):61–61. Available from: https://dx.doi.org/10.1186/s13072-018-0226-9.
14) Guiblet WM, Cremona MA, Harris RS, Chen D, Eckert KA, Chiaromonte F, et al. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Research*. 2021;49(3):1497–1516. Available from: https://dx.doi.org/10.1093/nar/gkaa1269.
15) Poggi L, Richard GF. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiology Molecular Biology Reviews*. 2020;85(1). Available from: https://doi.org/10.1128/MMBR.00110-20.
16) McKinney JA, Wang G, Mukherjee A, Christensen L, Subramanian SHS, Zhao J, et al. Distinct DNA repair pathways cause genomic instability at alternative DNA structures. *Nature Communications*. 2020;11(1):236–236. Available from: https://dx.doi.org/10.1038/s41467-019-13878-9.
17) Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. DNA H form requires a homopurine–homopyrimidine mirror repeat. *Nature*. 1987;330(6147):495–497. Available from: https://dx.doi.org/10.1038/330495a0.
18) Buske FA, Mattick JS, Bailey TL. Potential in vivo roles of nucleic acid triple-helices. *RNA Biology*. 2011;8(3):427–439. Available from: https://dx.doi.org/10.4161/rna.8.3.14999.
19) Shah KA, Mirkin SM. The hidden side of unstable DNA repeats: Mutagenesis at a distance. *DNA Repair*. 2015;32:106–112. Available from: https://dx.doi.org/10.1016/j.dnarep.2015.04.020.
20) Sproul JS, Barton LM, Maddison DR. Repetitive DNA Profiles Reveal Evidence of Rapid Genome Evolution and Reflect Species Boundaries in Ground Beetles. *Systematic Biology*. 2020;69(6):1137–1148. Available from: https://dx.doi.org/10.1093/sysbio/syaa030.
21) Tateishi-Karimata H, Sugimoto N. Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. *Nucleic Acids Research*. 2021;49(14):7839–7855. Available from: https://dx.doi.org/10.1093/nar/gkab580.
22) Khristich AN, Armenia JF, Matera RM, Kolchinski AA, Mirkin SM. Large-scale contractions of Friedreich's ataxia GAA repeats in yeast occur during DNA replication due to their triplex-forming ability. *Proceedings of the National Academy of Sciences*. 2020;117(3):1628–1637. Available from: https://dx.doi.org/10.1073/pnas.1913416117.
23) Zhang J, Fakharzadeh A, Pan F, Roland C, Sagui C. Atypical structures of GAA/TTC trinucleotide repeats underlying Friedreich's ataxia: DNA triplexes and RNA/DNA hybrids. *Nucleic Acids Research*. 2020;48(17):9899–9917. Available from: https://dx.doi.org/10.1093/nar/gkaa665.
24) Behe MJ. An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Research*. 1995;23(4):689–695. Available from: https://dx.doi.org/10.1093/nar/23.4.689.
25) Zain R, Sun JS. Do natural DNA triple-helical structures occur and function in vivo? *Cellular and Molecular Life Sciences*. 2003;60(5):862–870. Available from: https://dx.doi.org/10.1007/s00018-003-3046-3.
26) Bucher P, Yagil G. Occurrence of oligopurine.oligopyrimidine tracts in eukaryotic and prokaryotic genes. *DNA Sequence*. 1991;1(3):157–172. Available from: https://dx.doi.org/10.3109/10425179109020767.
27) Schroth GP, Ho PS. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Research*. 1995;23(11):1977–1983. Available from: https://dx.doi.org/10.1093/nar/23.11.1977.
28) Mehrotra S, Goyal V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics & Bioinformatics*. 2014;12(4):164–171. Available from: https://dx.doi.org/10.1016/j.gpb.2014.07.003.
29) Dawoudi MR. Mathematical Modeling Approaches to Understanding Severe Acute Respiratory Syn-drome Coronavirus 2 (SARSCoV-2) DNA Sequences Linked Coronavirus Disease (COVID-19) for Discovery of Potential New Drugs. *Open Access Journal of Biomedical Science*. 2020;2. Available from: https://dx.doi.org/10.38125/oajbs.000173.
30) Yadav U, Yadav S, Sharma CS. Characterization of Flowering Genes of Arabidopsis thaliana for Mirror Repeats. *Biointerface Research in Applied Chemistry*;12(3):2852–2861. Available from: https://doi.org/10.33263/BRIAC123.28522861.
31) Sandeep Y, Usha Y, C SD. In-silico evaluation of 'Mirror Repeats' In HIV Genome. *International Journal of pharma and Bio Sciences*. 2021;11(5):81–87. Available from: https://dx.doi.org/10.22376/ijpbs/lpr.2021.11.5.l81-87.
32) Vikash B, Swapni G, Sitaram M, Kulbhushan S. FPCB: a simple and swift strategy for mirror repeat identification. *eprint arXiv:13123869*. 2013. Available from: https://arxiv.org/abs/1312.3869.
33) Wójtowicz J, Gieczewska KB. The Arabidopsis Accessions Selection Is Crucial: Insight from Photosynthetic Studies. *International Journal of Molecular Sciences*. 2021;22(18):9866–9866. Available from: https://dx.doi.org/10.3390/ijms22189866.
34) Vikash B, Kulbhushan S. Parallel DNA Synthesis : Two PCR product from one DNA template. *arXiv preprint arXiv:13123869*. 2013. Available from: http://arxiv.org/abs/1309.3658.
35) National Centre for Biotechnology Information. 2021. Available from: https://www.ncbi.nlm.nih.gov/.
36) Reverse Complement Tool. 2021. Available from: https://www.bioinformatics.org/sms/rev_comp.html.
37) Nucleotide blast. 2021. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABASE=n/a&QUERY=&SUBJECTS.
38) Farrugia R, Portelli B, Grech I, Camilleri D, Casha O, Micallef J, et al. Air damping of high performance resonating micro-mirrors with angular vertical comb-drive actuators. *Microsystem Technologies*. 2019;p. 1–5. Available from: https://dx.doi.org/10.1007/s00542-019-04416-0.
39) Bevan M, Walsh S. The Arabidopsis genome: A foundation for plant research. *Genome Research*. 2005;15(12):1632–1642. Available from: https://dx.doi.org/10.1101/gr.3723405.

40) Sato S. Complete Structure of the Chloroplast Genome of Arabidopsis thaliana. *DNA Research*. 1999;6(5):283–290. Available from: https://academic.oup.com/dnaresearch/article-lookup/doi/10.1093/dnares/6.5.283.
41) Lang DM. Imperfect DNA mirror repeats in the gag gene of HIV-1 (HXB2) identify key functional domains and coincide with protein structural elements in each of the mature proteins. *Virology Journal*. 2007;4(1):113–113. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2211468/.
42) Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*. 1994;371(6494):215–220. Available from: https://dx.doi.org/10.1038/371215a0.
43) Shah N, Nute MG, Warnow T, Pop M. Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics*. 2019;35(9):1613–1614. Available from: https://dx.doi.org/10.1093/bioinformatics/bty833.