Journal of Biology and Nature

Volume 16, Issue 1, Page 39-52, 2024; Article no.JOBAN.12024
ISSN: 2395-5376 (P), ISSN: 2395-5384 (O), (NLM ID: 101679666)

In-silico Evaluation of Mirror Repeats in Some Selected Genes of *Candida albicans*

Barkha Sehrawat [a], Priya Yadav [a], Mustak Sarjeet [a], Sakshi Yadav [a], Vidhi Yadav [a], Nupur Goyal [a], Parvej Alam [a], Meghali Ahlawat [a] and Sandeep Yadav [a++*]

[a] School of Life Sciences, Starex University Gurugram, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.56557/JOBAN/2024/v16i18639

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://prh.ikprress.org/review-history/12024

Original Research Article

Received: 04/02/2024
Accepted: 08/04/2024
Published: 12/04/2024

ABSTRACT

All cellular processes in a living cell are controlled by its genetic material. DNA in majority of the domains acts as a regulatory molecule by controlling various vital functions. The genetic makeup of DNA or RNA (in viruses) is unique in all the domains. Their uniqueness is determined by the presence of various types of repetitive patterns of bases. These includes inverted repeats, tandem repeats, VNTR's, palindromes etc. Among many repetitive pattern types, Mirror repeats (MR) found to be dispersed throughout the genes or genomes. These sequences are associated with various functional features like their involvement in H-DNA formation, in replication & transcription, nervous system related diseases development etc. The major focus of this investigation is to identify MR sequences from some selected genes of Candida albicans using a bioinformatics based pipeline. The approach refers to as FPCB which utilized some manual steps to extract out MR sequence

[++] Assistant Professor;
**Corresponding author: E-mail: yadavs.p117@gmail.com;*

from any targeted gene or genome. The current study find out that the identified Mirror repeats found to be dispersed throughout the selected genes along with variable length. Among them the maximum & minimum MR sequences were reported in the gene FAS2 (108) & HIS1 (15) respectively. The present study will be helpful to provide a new insight for molecular as well as computational based studies of *Candida albicans* as well as other related fungal groups.

Keywords: NCBI; Mirror Repeat (MR); BLAST; Candida albicans; FASTA.

1. INTRODUCTION

What makes a specific organism different from one another is their genetic material which is acts as a blueprint of life [1]. Genetic material could be either DNA (in case of bacteria, animals, fungi, plant etc.) or RNA (in case of some viruses like HIV, flu virus like corona virus etc.) It composed of repeated nucleotides units which also form a unique pattern of sequences that makes all organisms different from others. The sequence pattern which is composed of base pairs shows diverse group of repeated sequences [2].These repetitive sequences are differentiated into two types on the basis of their arrangement in genes or genomes. One includes tandem repeat which is present adjacent to each to other whereas another type is interspersed repeat which is scattered throughout the genome [3-4]. Tandem repeat further classified into microsatellite & minisatellite DNA [5]. Similarly Interspersed repeat consist of SINE & LINE sequences and retrotransposons [6-8].These repeats perform variety of functions like help in gene expression, function as evolutionary markers, makeup the promoters & enhancers parts on DNA and also associated with many diseases in human [9-11]. Repeat sequences are always attracting researchers because of their unique roles at molecular level. Among these repeats a special sequence pattern refers to as Mirror repeat also reported. Mirror repeat is basically that segment of genome or gene which shares a center of symmetry, forming exact mirror image of each other on the same strand. For example, in this sequence TGACGGCATTACGGCAGT, TGACGGCAT shares center of symmetry or shows homology with rest of its part. Due to mutation in polypyrimidine and polypurine mirror repeat it causes genomic instability that results in genetic, neurological diseases [12-15]. Mirror repeats have been identified in various phyla including bacteria, animal & plant viruses, as well as in human insulin gene [16-19]. The present research will focus on identification of mirror repeats in some selected genes of fungi *Candida albicans* using an *in silico* methodology (FPCB) [20]. FPCB stands for FASTA Parallel Complement BLAST, a bioinformatics based manual pipeline to find out mirror repeats in any targeted gene or genome sequence [20-22]. This study will be helpful in analysing the functioning as well as molecular prospective of mirror repeats sequence in fungal as well as other living domains.

2. MATERIALS AND METHODS

To identify mirror repeats in selected *Candida albicans* genes (Table-1), a bioinformatics based manual pipeline refers to as FPCB [20] were utilized. This approach can be applied on any targeted gene or genome family to extract MR sequences. It involve four different steps, in each step public domain databases were utilized which are freely accessible. The steps involve in this process (shows in Fig. 1) are followings: –

Table 1. List of genes targeted for mirror repeat identification

S. No.	Gene Name	Gene Functions	Gene ID
1	ADE 2	Nucleotide synthesis	3635259
2	FAS 2	Fatty acid synthase subunit, lipid synthesis	3635222
3	FTR 1	High-affinity iron permease	3643318
4	HEM 3	Haemosynthesis	3643940
5	HIS 1	Amino acid biosynthesis	3636501
6	MIG 1	DNA binding protein, glucose metabolism	3637812
7	LEU 2	Amino acid synthesis	3638034
8	SNF 1	Derepression of catabolite repression, carbon source utilization	3642643
9	NMT 1	Myristoyl-CoA:protein*N*-myristoyltransferase, lipid biosynthesis	3635624
10	URA 3	Nucleotide synthesis	3636649

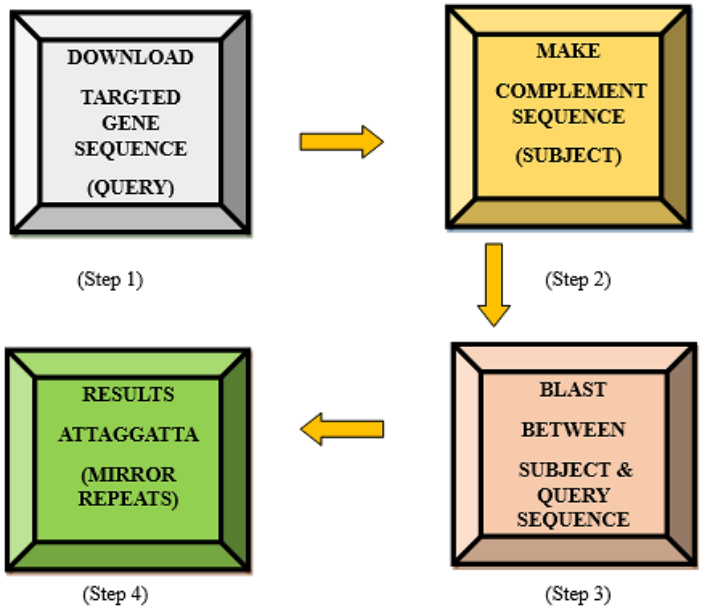

Fig. 1. Depicted flow model of FPCB

- The targeted gene sequence of *C. albicans* was downloaded from NCBI [23]
- The downloaded sequence can be processed through Reverse complement tool
- Subject & Query sequence align with each other using BLAST tool
- Result & Analysis

3. RESULTS

By using FPCB (a bioinformatics-based approach) we extract mirror repeats for selective genes of *Candida albicans*. The identified mirror repeats in all the 10 genes are of different lengths. The minimum length of mirror repeats were observed is of 7 nucleotide which was found in almost all genes but the maximum length of mirror repeat (MR No. 9) is of 54 nucleotide found in gene MIG 1. The graph shown below (Fig. 2) represents the total number of identified mirror repeats in each selected gene. The maximum & minimum numbers of MR sequences were reported in FAS2 (108) & HIS1 (15) gene respectively.

The complete gene wise distribution of mirror repeats sequence in selected genes of *Candida albicans* along with their length, position in the regions and sequences are enlisted in Tables 2 to 11.

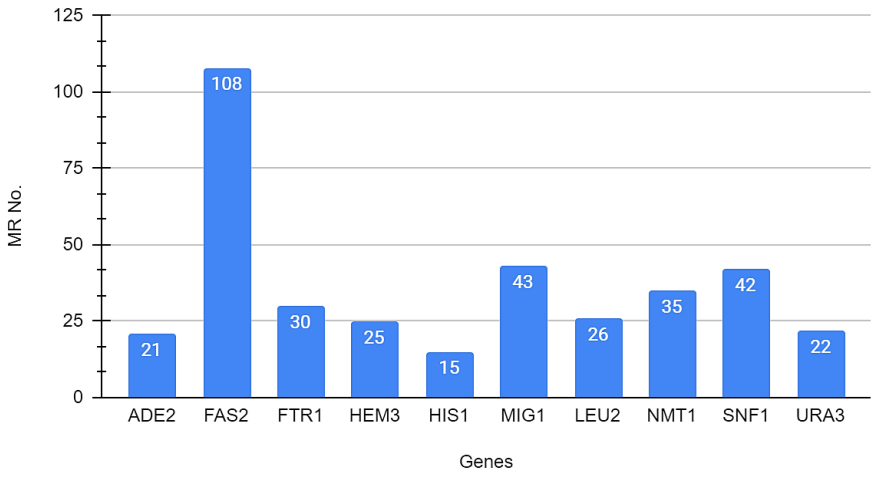

Fig. 2. Shows graphical representation of gene wise distribution of mirror repeats in selected genes of *Candida albicans*

Table 2. Shows Mirror repeat sequence pattern, their length & position in ADE2 gene

MR No	Sequence	Length	Position
	Region 1 1-500		
MR 1	TTCAAGAGAAATT	13	251
MR 2	TAAAGTTGAAAT	12	267
MR 3	TCCATTACCT	10	282
MR 4	GTTACACATTG	11	402
MR 5	GATGGTAG	8	463
	Region 2 501-1000		
MR 6	AATTTTTAA	9	18
MR 7	GAAGGGGAAG	10	101
MR 8	GGTGTGG	7	246
MR 9	TGTTCTTGT	9	270
MR 10	CACTTCAC	8	358
MR 11	TCGTGCT	7	382
MR 12	AAATAAA	7	484
	Region 3 1001-1707		
MR 13	AGACCAGA	8	60
MR 14	GTTATATATTG	11	130
MR 15	ATTCCTAAATCTTTA	15	156
MR 16	GGGAATAATAATGG	14	197
MR 17	TGCTCGT	7	245
MR 18	ACTCCTCA	8	306
MR 19	GGTATGG	7	399
MR 20	TAACAAT	7	530
MR 21	GAAGTATTAGGTAAAGCAGAAACTTT-AGAAGA-AATTGGAT-ATGAAG	47	633

Table 3. Shows mirror repeat sequence pattern, their length & position inFAS2 gene

MR No	Sequence	Length	Position
	Region 1 1-1000		
MR 1	GAACAAG	7	16
MR 2	TTGTTAACTGAATTGTT	17	37
MR 3	AACTCAA	7	90
MR 4	GGTATGG	7	163
MR 5	AACTATCAA	9	177
MR 6	TGCAACGT	8	221
MR 7	TTGTGTT	7	235
MR 8	GAGTACCCCATCAG	14	321
MR 9	TTGCACAAAAATTAAAGAAACCTT	24	467
MR 10	AATGGTAA	8	532
MR 11	GGAAAGG	7	574
MR 12	TACTTCAT	8	678
MR 13	CATCACTAC	9	729
MR 14	TTTGGTTT	8	763
MR 15	GGTTTGG	7	766
MR 16	TTCTGTCTT	9	789
MR 17	TTTCTTT	7	858
MR 18	AAATACGCATCAA	13	880
MR 19	GCCTCCG	7	922
MR 20	GTGGTGGTG	12	947
MR 21	GTTGTTG	7	958

MR No	Sequence	Length	Position
MR 22	TGATAGT	7	963
	Region 2 1001-2000		
MR 23	AAGGAAAAGGAA	12	81
MR 24	TTAGATT	7	117
MR 25	CAACCAAC	8	165
MR 26	AATATAA	7	373
MR 27	GTCAACAATTGATTGACAACTG	22	412
MR 28	AGTTTTGA	8	440
MR 29	TATGTGTAT	9	567
MR 30	TCAACCAACT	10	611
MR 31	GATTTAG	7	630
MR 32	TTGAAAGTT	9	724
MR 33	CCAACAACC	9	789
MR 34	ACAACCCCAACA	12	792
MR 35	CTTCTTC	7	829
MR 36	TCAAACT	7	905
MR 37	TTCTCTT	7	974
	Region 3 2001-3000		
MR 38	AACTTTCAA	9	19
MR 39	TTTAATCAGTGGTGGTGCCAAAGTT	25	91
MR 40	GTTATTG	7	113
MR 41	TATGTAT	7	166
MR 42	CTGGGTC	7	189
MR 43	TTGTTGTT	8	204
MR 44	AAAGAAA	7	274
MR 45	GGTTT-GGGTTGGGATTTGG	19	284
MR 46	AATGGTAA	8	335
MR 47	TGGTAATGGT	10	337
MR 48	TCTAAATCT	9	362
MR 49	GATTGTTAG	9	408
MR 50	TTTTGGTTTT	10	493
MR 51	TACTCTGAA-TCTAAAATCTCAT	22	512
MR 52	TCTAAAATCT	10	521
MR 53	CTTATTC	7	541
MR 54	TGTTTGT	7	586
MR 55	CAACAAC	7	637
MR 56	AAAGGAAA	8	691
MR 57	CAAGAAGAAC	10	749
MR 58	TGGTGGT	7	778
MR 59	AGCTATCGA	9	880
	Region 4 3001-4000		
MR 60	AAACAAA	7	1
MR 61	TGGAAGGT	8	20
MR 62	TCCATGGGGTAACT	14	78
MR 63	TGGAAGGT	8	134
MR 64	CTCAAACTC	9	233
MR 65	GAAGAAG	7	274
MR 66	AAGAAGAA	8	275
MR 67	TTAGATT	7	302
MR 68	AAAGAAA	7	342
MR 69	AAACAAA	7	349
MR 70	AAGTTGTTGTTCAA	14	365
MR 71	GTTCAACACGATTTAGAACCATTTG	25	373

MR No	Sequence	Length	Position
MR 72	AAAGAAA	7	406
MR 73	ACAAACA	7	425
MR 74	GAAATTTTTGAAATTGA-AGA-AAGTGGTGAATACACAGTTAGAATCTTGAAAG	52	
MR 75	GTTCCGAAAGCTTTG	15	517
MR 76	TGGTCAAATTCCAACTGGT	19	552
MR 77	CCTATGGTATCC	12	584
MR 78	TATCCCAGAAGACACTAT	18	591
MR 79	GTTGGTTG	8	639
MR 80	GGTATGG	7	742
MR 81	GTCTCTG	7	754
MR 82	AGATAGA	7	780
MR 83	ATAGATA	7	782
MR 84	TGTTGTTGTTGT	12	860
	Region 5 4001-5665		
MR 85	AAGAAGAA	8	1
MR 86	AGAACACCAAAGGAAATGTCAAGA	24	72
MR 87	GTCAAGACCAACCACTACTACCAGAAATG	29	89
MR 88	AAGAAGAAGCCGAGTTG-GCTAAAGAAGAA	29	406
MR 89	TGGCTAAAGAAGAATTTGGT	20	421
MR 90	AGTTCTTGA	9	460
MR 91	AAGAGAGAA	9	469
MR 92	TGAAGAAGT	9	479
MR 93	AATGGGGTAA	10	529
MR 94	TTCAACTT	8	594
MR 95	GTTGCATCCTTCCATGGTA-CTTCCACCGTTG	31	621
MR 96	TCTGGTCT	8	813
MR 97	TTTGGTTT	8	939
MR 98	TGTTGTTGT	9	968
MR 99	TTGTTGTT	8	970
MR 100	ACAAAACA	8	1207
MR 101	TCAAAAAACT	10	1250
MR 102	TAAACAAAT	9	1279
MR 103	TGTAGATGT	9	1304
MR 104	TTGAAAGAAACTT	13	1354
MR 105	AAGTTGAA	8	1378
MR 106	TATTGAAATCACTCGTGACGTTAATG--GTGCTC-CTAAAGTAAT	42	1502
MR 107	TTTAAGTGAATTT	13	1640
MR 108	AATTTTAA	8	1648

Table 4. Shows Mirror repeat sequence pattern, their length & position inFTR1 gene

MR No.	Sequence	Length	Position
	Region 1 1 – 500		
MR 1	TTGTTGTT	8	65
MR 2	GGGGCGGGG	9	160
MR 3	TAGTCTGTCTTAT	13	179
MR 4	TTATTATT	8	188
MR 5	TCTATGGGAAGGTATCT	13	258
MR 6	TATGGGTAT	9	306
MR 7	AAAGAAA	7	340
MR 8	AGATAGA	7	399
MR 9	TTAAATT	7	407

MR No.	Sequence	Length	Position
MR 10	ATTGGGTTA	9	411
MR 11	TTCATT—TTACCATTCATTACTT	22	442
MR 12	CTGTTGTC	8	485
	Region 2 501- 1146		
MR 13	GTTGTTG	7	71
MR 14	TTATTTCCACCTGTATT	17	129
MR 15	TGGTGGT	7	220
MR 16	TCTAAATCT	9	269
MR 17	GTCAACTG	8	287
MR 18	TCTGTCATTTCCTACAACATTTACTGGCT	29	371
MR 19	TGTTTTGT	8	412
MR 20	TTGTTGTT	8	416
MR 21	AACCCAA	7	488
MR 22	AAGAAGAA	8	512
MR 23	GAAGAAG	7	514
MR 24	AAACAAA	7	529
MR 25	AACTTCAA	8	561
MR 26	CAACAAC	7	566
MR 27	AACAACAA	8	567
MR 28	AAACAAA	7	613
MR 29	AAAGAAA	7	632
MR 30	AAACAAA	7	637

Table 5. Shows Mirror repeat sequence pattern, their length & position inHEM3 gene

MR No	Sequence	Length	Position
	Region 1 1 – 500		
MR 1	AGAAGGTTATT-GAAGA	17	95
MR 2	AACTTGTCATGTTCAA	16	121
MR 3	ATAAAGTACAAACCCAACCATTATATA	27	161
MR 4	TGGTGGT	7	192
MR 5	AAGAGAA	7	339
MR 6	TAGAGAT	7	351
MR 7	TACTTCAT	8	426
MR 8	TAGAAGAT	8	438
MR 9	AATTAATTAA	10	452
	Region 2 501- 1023		
MR 10	AACCAGATTAAACAAATTGGACCAA	25	7
MR 11	ACCAACCA	8	27
MR 12	TTATATT	7	54
MR 13	AATCAGAT—TAGGTTTGGGTCATAGAATAA	31	76
MR 14	TATGTAT	7	121
MR 15	GTATTATG	8	124
MR 16	GGAAAGG	7	165
MR 17	AAAGAAA	7	196
MR 18	TAGTGAT	7	295
MR 19	GATAATAG	8	337
MR 20	AGTTAGTCCTGATGGA	16	343
MR 21	TGAACAAGT	9	412
MR 22	AAATTAAA	8	440
MR 23	AAAGAAA	7	461
MR 24	CCAGACC	7	504
MR 25	AGTGTGA	7	517

Table 6. Shows mirror repeat sequence pattern, their length & position inHIS1 gene

MR No	Sequence	Length	Position
MR 1	GATTTAG	7	4
MR 2	CATTTAC	7	16
MR 3	TTTACCAGACCGTTT	15	18
MR 4	TCCTAAAAAGGGCAGATTATACG—AAAAATGCT	34	45
MR 5	GATATAG	7	127
MR 6	CGAAAAGC	8	339
MR 7	AAAGAAA	7	363
MR 8	CATTTAC	7	380
MR 9	TGAAAGT	7	507
MR 10	AAAGTGGTGAAA	12	509
MR 11	AAAGAAA	7	588
MR 12	AATATAA	7	616
MR 13	GGACAAACA-CAGCGACGACGAAGAGG	26	753
MR 14	AAAGGAAA	8	810
MR 15	GGTAATGTAATGG	13	820

Table 7. Shows mirror repeat sequence pattern, their length & position inMIG1 gene

MR No	Sequence	Length	Position
	Region 1 1- 500		
MR 1	TCCACACCT	9	10
MR 2	AGAAAAAGA	9	65
MR 3	GAATAAG	7	72
MR 4	AGATTAGA-ACATCAAACTAGACATATTAGA	31	124
MR 5	TTGTGTT	7	195
MR 6	TTAAGAATT	9	238
MR 7	AAAAATAAGAATCAAAA	17	274
MR 8	GGTATGG	7	340
MR 9	CTATTCCATTTTCTATAGATCGTAATGGAAATC ATGTATATCATCAACCTTATC	54	398
	Region 2 501- 1000		
MR 10	AACTCAA	7	1
MR 11	TGAAAGT	7	232
MR 12	ATTAAATTA	9	239
MR 13	CTAAATC	7	321
MR 14	ACATTACA	8	386
MR 15	TTACATT	7	389
MR 16	TTGGGTT	7	398
MR 17	CCTGGGTCC	9	427
	Region 3 1001-1713		
MR 18	TAATAAT	7	2
MR 19	TTGAGTT	7	39
MR 20	CTTCTTC	7	67
MR 21	TACTAATAATAAT	13	158
MR 22	TAATAATAAT	10	161
MR 23	ATAATAATA	9	163
MR 24	ATAATATGA--CAAAACCAAGTATAATA	26	166

MR No	Sequence	Length	Position
MR 25	ACAAAACCA-AGTATAATACCAAAACA	26	174
MR 26	CATCTTCTAC	10	202
MR 27	CTACATC	7	208
MR 28	TTAAATT	7	216
MR 29	CAACAACAAC	10	246
MR 30	ACAACAACA	9	248
MR 31	AGAATTATCACACTCATAAGA	21	256
MR 32	AACCCCAA	8	302
MR 33	CTAATAATC	9	337
MR 34	TATAATAT	8	350
MR 35	ATCAACCACCAACTA	15	418
MR 36	GGTAATGG	8	519
MR 37	AACACAA	7	527
MR 38	ACCACCA	7	536
MR 39	TTACTTCATT	10	562
MR 40	TAATGGTAAT	10	659
MR 41	TAATAAT	7	665
MR 42	AATTCAAATGCTAAACCAATGAGTTTAACCAAT TTATTAAGTTAA	45	669
MR 43	TTTAACCAATTT	12	692

Table 8. Shows mirror repeat sequence pattern, their length & position inLEU2 gene

MR No	Sequence	Length	Position
	Region 1 1 – 500		
MR 1	TGTCTGT	7	2
MR 2	TCTGTTAAAACCAAAACCATTACTGTCT	28	4
MR 3	TGGTGGT	7	147
MR 4	TCTTGAAAGTGCT	13	195
MR 5	CTAAATC	7	206
MR 6	ATGGGGTA	8	252
MR 7	GAACAAG	7	277
MR 8	TTGC-CAGTGACTCGTT	17	341
MR 9	AATTTAA	7	400
MR 10	TGGTGGTATATATTTTGGT	19	426
MR 11	CAAGAAC	7	451
MR 12	AUGGCAAGAAAA	11	452
MR 13	AAGAAGAA	8	458
MR 14	AAGTGAA	7	465
	Region 2 501- 1122		
MR 15	AGAAAGA	7	125
MR 16	TGTCTGT	7	168
MR 17	CCAAAACC	8	214
MR 18	AACCCAA	7	218
MR 19	TTTGGGTTT	9	304
MR 20	AAGTGAA	7	411
MR 21	ATTGTGTTA	9	468
MR 22	GAAGAAG	7	494
MR 23	AGGTTTTGGA	10	510
MR 24	AAAATTTTAAAA	12	599
MR 25	TAAAACAAAAT	11	606
MR 26	AATACATAA	9	614

Table 9. Shows mirror repeat sequence pattern, their length & position in SNF1 gene

MR No	Sequence	Length	Position
	Region1 1 – 500		
MR 1	GTCTCTG	7	25
MR 2	TAAAGTGAAAT	11	195
MR 3	AAAAGTTGCTTTGAAAA	17	231
MR 4	AGTCTGA	7	272
MR 5	AGAGAGA	7	298
MR 6	CATATTTAAGATTATTGAGACACCCACATA-TCATT—AAATTATAC	47	308
MR 7	GTTATTG	7	385
MR 8	TTATATT	7	417
MR 9	GTAAAATG	8	434
MR 10	CAGAAGAC	8	443
MR11	AGAAGAC-GAAGCCAGAAGA	19	444
	Region 2 501-1000		
MR 12	ATTGTTA	7	43
MR 13	GTTAAAATTG	10	65
MR 14	TTGGGTT	7	81
MR 15	TGTATGT	7	219
MR 16	GTAGATTACCATTCGATG	18	237
MR 17	TTGTTGTT	8	354
MR 18	ACCGCCA	7	439
MR 19	CAAAAAC	7	454
	Region 3 1001-1857		
MR 20	GTTAATGTAATTG	13	39
MR 21	CAAACAAAC	9	58
MR 22	AACAAACAA	9	60
MR 23	AGTAATGA	8	108
MR 24	TATTTAT	7	129
MR 25	AAGAGAA	7	142
MR 26	AAAGACTTGAAAAAGAGTAAA---TCAGAAA	28	162
MR 27	ATTCTTA	7	200
MR 28	CCACCACC	8	216
MR 29	ACCACCA	7	218
MR 30	CTCGGCTC	8	254
MR 31	CGTTACCAAATTCGACCATTGC	22	319
MR 32	ACCAACCA	8	347
MR 33	TCCTC-TGCAACAACATGCTCCT	22	407
MR 34	CTCCTCCTC	9	424
MR 35	CCTCCTCC	8	426
MR 36	GGCGCGG	7	531
MR 37	GAAGAAG	7	555
MR 38	AAGAAGAA	8	556
MR 39	ACAGACA	7	751
MR 40	TTTAGTTGATTT	12	680
MR 41	TTACATT	7	801
MR 42	ATTAATTA	8	818

Table 10. Shows mirror repeat sequence pattern, their length & position inNMT1 gene

MR No	Sequence	Length	Position
	Region 1 1 – 500		
MR 1	AAAGGAAA	8	108
MR 2	AAGTGAA	7	159
MR 3	GTCACTG	7	169
MR 4	GAAGAAG	7	175
MR 5	CCATTACC	8	226
MR 6	TTACCATT	8	229
MR 7	TAAATTATTATAT	13	300
MR 8	TTAGATT	7	344
MR 9	AGAAAAGA	8	400
MR 10	TTGGGGTT	8	416
MR 11	TTAAATT	7	479
MR 12	TAAATTAAATAAATCAAAT	19	480
	Region 2 501-1356		
MR 13	TTATGTATT	9	29
MR 14	ATTA-GCCCCCGTATTA	16	61
MR 15	AAAGAAA	7	80
MR 16	GGGTTAATAAACAAAACATTTGG	23	96
MR 17	CATTTAC	7	224
MR 18	ACCTCCA	7	229
MR 19	AAATCAAACTAAA	13	235
MR 20	CCTAATAATCC	11	275
MR 21	AATCCTAA	8	281
MR 22	TATAAATAT	9	347
MR 23	TATCAAGAACGAT	13	353
MR 24	GAAGAAG	7	389
MR 25	AAGAAGAA	8	390
MR 26	AATTTAA	7	396
MR 27	GGTTAAAAGTTATGTAGTTGAAGA—TGAAAATGGG	34	445
MR 28	TTACCATT	8	509
MR 29	TTTGTTT	7	562
MR 30	AAACCAAA	8	593
MR 31	CATTAATTAC	10	636
MR 32	TAAAAAAT	8	649
MR 33	TTCAATTGTTTAACTT	16	671
MR 34	TTTGGTAGTGGTGATGGTTT	20	722
MR 35	TTTTTAAATTATTATCTTTTT	21	740

Table 11. Shows mirror repeat sequence pattern, their length & position in URA3 gene

MR No.	Sequence	Length	Position
MR 1	TGACAGT	7	2
MR 2	GAGAGAG	7	28
MR 3	TAAGGAAT	8	129
MR 4	TATGTAT	7	166
MR 5	ATTATTA	7	234
MR 6	ACTTTCA	7	243
MR 7	AAGATAGAA	9	275
MR 8	GTCACTG	7	376
MR 9	AAAGAAA	7	418
MR 10	AGAAACCACCACCAACCAAGA	21	420

MR No.	Sequence	Length	Position
MR 11	AACCACCACCAA	12	423
MR 12	GTTATTG	7	450
MR 13	TTCTCAAAAAACTGTT	16	504
MR 14	CTAAATC	7	527
MR 15	TAAGGAAT	8	537
MR 16	TTGTTGTT	8	545
MR 17	GTTGTTG	7	547
MR 18	AAGAAGAA	8	584
MR 19	GAAGAAG	7	586
MR 20	GGATTAGG	8	646
MR 21	TGATATTATTATTGT	15	696
MR 22	TAGAGAT	7	765

4. DISCUSSION

The previous studies on mirror repeats also support the current one. According to the recent study done on human insulin gene were reported a total number of 210 MR sequences in insulin gene [21]. Similar like this other studies on animal & plant viruses (19), bacterial genes [16], model plant [17] & some animal domains [24-26] also provide an insight on the frequent occurrence of Mirror repeats in genes/genomes. The output of current study on selected genes of *Candida albicans* also shows that MR sequences frequently present in all the selected genes. These sequences show diversity at the level of their length as well as provide a hint on their specific occurrence in a particular domain. Due to their unique nature of occurrence they will be utilized for many prospective like as a biomarker for testing, in evolution based studies, in context of clinical uses, new *in silico* tools development [27-29] etc. The modern era of computational biology can be used them for phylogeny prospective or to classify the domains due to their conservative nature.

5. CONCLUSION

In-silico analysis of selected genes of *Candida albicans* concluded that the frequent occurrence of mirror sequencesgives many insights like these sequences will be the integral part of the genes & will be play crucial roles at genetic level in the said fungus. Thesesequences will be utilized for many purposes like in evolutionary studies, clinical & therapeutic based studies as well as in computational genomics or proteomics based research.

DISCLAIMER

This paper is an extended version of a preprint document of the same author.

The preprint document is available in this link: https://www.biorxiv.org/content/10.1101/2024.01.05.574287v1
[As per journal policy, preprint article can be published as a journal article, provided it is not published in any other journal]

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Rich A, Watson JD. Some relations between DNA and RNA. Proceedings of the National Academy of Sciences. 1954 Aug;40(8):759-64.
2. Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. Annu. Rev. Genomics Hum. Genet. 2007 Sep 22;8:241-59.
3. Usdin K. The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. Genome Research. 2008 Jul 1;18(7):1011-9.
4. Shammas MA. Repetitive sequences, genomic instability, and Barrett's esophageal adenocarcinoma. Mobile Genetic Elements. 2011 Sep 1;1(3): 208-12.
5. Mehrotra S, Goyal V. Repetitive sequences in plant nuclear DNA: Types, distribution, evolution and function. Genomics, Proteomics & Bioinformatics. 2014 Aug 1; 12(4):164-71
6. Smit AF. The origin of in6terspersed repeats in the human genome. Current Opinion in Genetics & Development. 1996 Dec 1;6(6):743-8.
7. Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. Targeted

identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. The Plant Cell. 2011 Sep;23(9):3117-28.
8. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, San Miguel PJ, Bennetzen JL. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genetics. 2009 Nov 20;5(11): e1000732.
9. Wahls WP. 2 meiotic recombination hotspots: Shaping the genome and insights into hypervariable minisatellite DNA change. Current Topics in Developmental Biology. 1997 Jan 1;37: 37-75.
10. Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Research. 1999;27(2): 573-80.
11. Hancks DC, Kazazian Jr HH. Active human retrotransposons: Variation and disease. Current Opinion in Genetics & Development. 2012 Jun 1;22(3):191-203.
12. Belland RJ. H-DNA formation by the coding repeat elements of neisserialopa genes. Molecular Microbiology. 1991 Oct; 5(10):2351-60.
13. LeProust EM, Pearson CE, Sinden RR, Gao X. Unexpected formation of parallel duplex in GAA and TTC trinucleotide repeats of Friedreich's ataxia. Journal of Molecular Biology. 2000 Oct 6;302(5): 1063-80.
14. Bissler JJ. Triplex DNA and human disease. Front Biosci. 2007 May 1;12: 4536-46.
15. Mirkin SM, Frank-Kamenetskii MD. H-DNA and related structures. Annual Review of Biophysics and Biomolecular Structure. 1994 Jun;23(1):541-76.
16. Yadav S, Yadav U, Sharma DC. 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 (BBRJ). 2022 Jan 1;6(1):93.
17. Yadav U, Yadav S, Sharma DC. Characterization of flowering genes of arabidopsis thaliana for mirror repeats. Bio interface Research in Applied Chemistry. 2021;12(3):2852-61.
18. Yadav S, Yadav U, Sharma DC. *In-silico* evaluation of 'mirror repeats' in HIV genome. (2021). Int. J. Life Sci. 2022 Sep 7;11(5):81-87.
19. Yadav P, Kumari J, Yadav P, Yadav R, Yadav S, Sharma D, Singh A, Sehrawat B, Yadav M, Yadav S. Identification of Mirror Repeats in Viral Genomes using FPCB Analysis. bioRxiv 2023:2023-04.
20. Vikash B, Swapni G, Sitaram M, Kulbhushan S. FPCB: A simple and swift strategy for mirror repeat identification. Arxiv Preprint Arxiv:1312.3869; 2013 Dec 13.
21. Yadav M, Sehrawat B, Yadav P, Kumari J, Yadav P, Yadav S, Yadav R, Singh A, Yadav DS. In-Silico analysis of Human Insulin Gene for Mirror repeat sequences. bioRxiv. 2023:2023-07.
22. Yadav U, Yadav S, Sharma DC. *In silico* analysis of structural photosynthetic genes of arabidopsis thaliana for unique mirror repeats. Indian Journal of Science and Technology. 2022 Feb 4;15(3):127-35.
23. National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 1988.
Available:https://www.ncbi.nlm.nih.gov/
24. Yadav D, Dhankhar M, Bhardwaj V. Identification of mirror repeats within the fibroblast growth factor8 gene of Xenopus tropicalis. Neuro Quantology. 2022;20(9): 1381.
25. Bhargava A, Madhuri ND. Identification of mirror repeats within the Trp53 (P53) gene of mus musculus. Journal of Survey in Fisheries Sciences. 2023 Apr 10;10(4S): 2163-71.
26. Dangi N, Saini K. Identification of mirror repeats within the pdgf-aa gene of danio rerio. Journal of Pharmaceutical Negative Results. 2023 Jan 31:2231-40.
27. Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud Z, Islam MM, Afrin T, Karim N. Identification of potential drug targets by subtractive genome analysis of Escherichia coli O157: H7: An in silico approach. Advances and Applications in Bioinformatics and Chemistry. 2015 Dec 8:49-63.
28. Berselli M, Lavezzo E, Toppo S. NeSSie: A tool for the identification of approximate DNA sequence symmetries. Bioinformatics. 2018 Jul 15;34(14): 2503-5.

29. Kim KR, Kim HY, Lee YD, Ha JS, Kang JH, Jeong H, Bang D, Ko YT, Kim S, Lee H, Ahn DR. Self-assembled mirror DNA nanostructures for tumor-specific delivery of anticancer drugs. Journal of Controlled Release. 2016 Dec 10;243:121-31.

Peer-review history:
The peer review history for this paper can be accessed here:
https://prh.ikprress.org/review-history/12024