Bioinformatics Study On Mirror DNA In Mycobacterium Tuberculosis Using Fast Parallel Complement Blast Strategy.

Munzali S. Zubair, Musa Hassan.Muhammad, R.G. Ahmad, Umar A.A

Abstract: Many softwares that have been used for identification DNA mirror repeats were not simple and effective for finding mirror repeats, so a new method for the identification of mirror repeats within the gene sequences in both eukaryotic and prokaryotic organisms has been developed and now is available. The method is called FAST PARALLEL COMPLEMENT BLAST which is the easiest method for observing mirror repeats in bacterial genome and other eukaryotic genome. This simple method involves only three steps: Downloading sequence of mycobacterium tuberculosis gene using NCBI in FASTA format, Making a Parallel complement, and searching for mirror repeats using BLAST which shows the homology between the original sequence and the parallel complement sequence. We explored twenty genes (20) of Mycobacterium tuberculosis H37Rv for DNA mirror repeats and the result shows that mirror repeats of 18 genes were successfully found. Furthermore, nobody knows the exact role of mirror repeats yet and detailed mechanisms and functions of most repeats are still unknown.

Keywords: Fast Parallel Complement Blast, Bioinformatics, DNA, mirror repeats, mycobacterium tuberculosis H37Rv, and Gene Sequences.

INTRODUCTION

A DNA mirror repeats can be identified when a sequence shares homology with the other part of the sequence within a gene or genome. A DNA mirror repeat is a sequence segment delimited on the basis of its containing a center of symmetry on a single strand and identical terminal nucleotides (Dorothy M. Lang, 2007). This implies that one part of the sequence is a mirror image of the other sequence. For example topA DNA mirror repeat 5'CAGCGGCCGGCGAC3', the sequence in red is a mirror image to sequence in black sequence. Most different type of repeats are found in bacterial genomes such as direct repeats, inverted repeats, complement repeats, and reverse complement (palindromic) repeats. Direct repeats are two or more repeating DNA fragments appeared in the same orientation and on the same strand. For example, a sequence direct repeats fragmentTGGAA and another sequence fragment ACCTT and are direct to each other. Inverted repeats are two or more repeating DNA fragments appeared in inverted orientation and on the same strand. For example, a sequence fragment GGAATCGATCTT and another sequence fragment AAGATCGATTCC are reverse repeats. Palindromic repeats are also called Complement repeats are two or more repeating DNA fragments looked in the same orientation but on the complement strand, i.e. in complement orientation on the same strand. For example, a sequence fragment CAGCGGC and another sequence fragmentGTCGCCG are reverse repeats. Reverse complement repeats are two or more repeating DNA fragments appeared in inverted orientation but on the complement strand, i.e. in reverse complement orientation on the same strand.

- *Munzali S. Zubair, Musa Hassan.Muhammad, R.G. Ahmad, Umar A.A*
- *Phone Number:+2347068823187*
- *E-Mail: musachakla@yahoo.com*
- *Affiliation: Federal University Dutse, Department of Biological Science*

For example a sequence fragmentATGGC and another sequence fragment GCCAT are reverse complement repeats (Yo-Cheng Chang,2013).

5'TGGAATGGAA 3' direct repeats 3'ACCTTACCTT5'
5'GGAATCGATCTTAAGATCGATTCC 3' inverted repeats
3'CCTTAGCTAGAATTCTAGCTAAGG 5' 5'
CAGCGGCCGGCGAC3' DNA mirror repeats 5'
CAGCGGCCGGCGAC 3'

The aim of this work is to find the mirror repeats in the Mycobacterium Tuberculosis H37Rv genome.

Methodology

Gene Sequences

The gene sequences or coding sequences were downloaded in FASTA format from NCBI link: http://www.ncbi.nlm.nih.gov

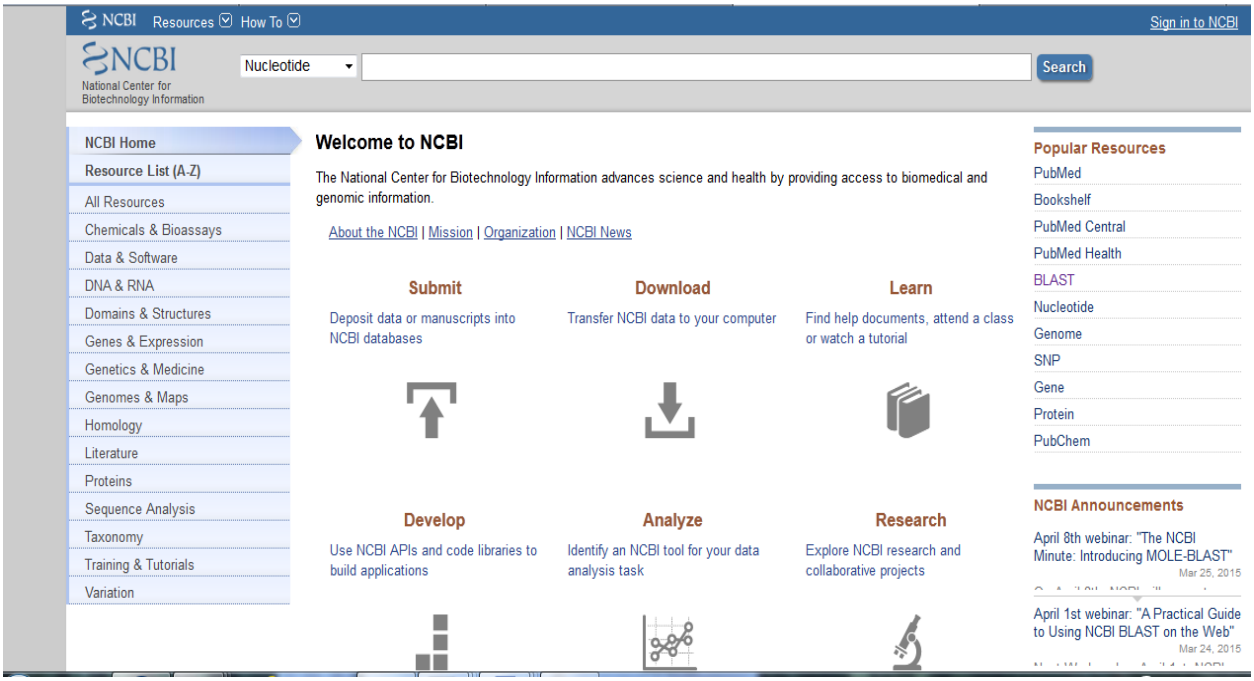


Figure 1: *shows home page*

Parallel Complement Sequence

Parallel complement sequences of the twenty (20) were downloaded from reverse Complement program (http://www.bioinformatics.org/sms/rev_comp.html)

Making Mirror Repeats

Mirror repeats of the downloaded sequences were found using BLAST _N. In which both FASTA format of Nucleotide sequence and its parallel complement were pasted and alignedfor BLAST homology search using BLAST tool: (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&QUERY=&SUBJECTS=)

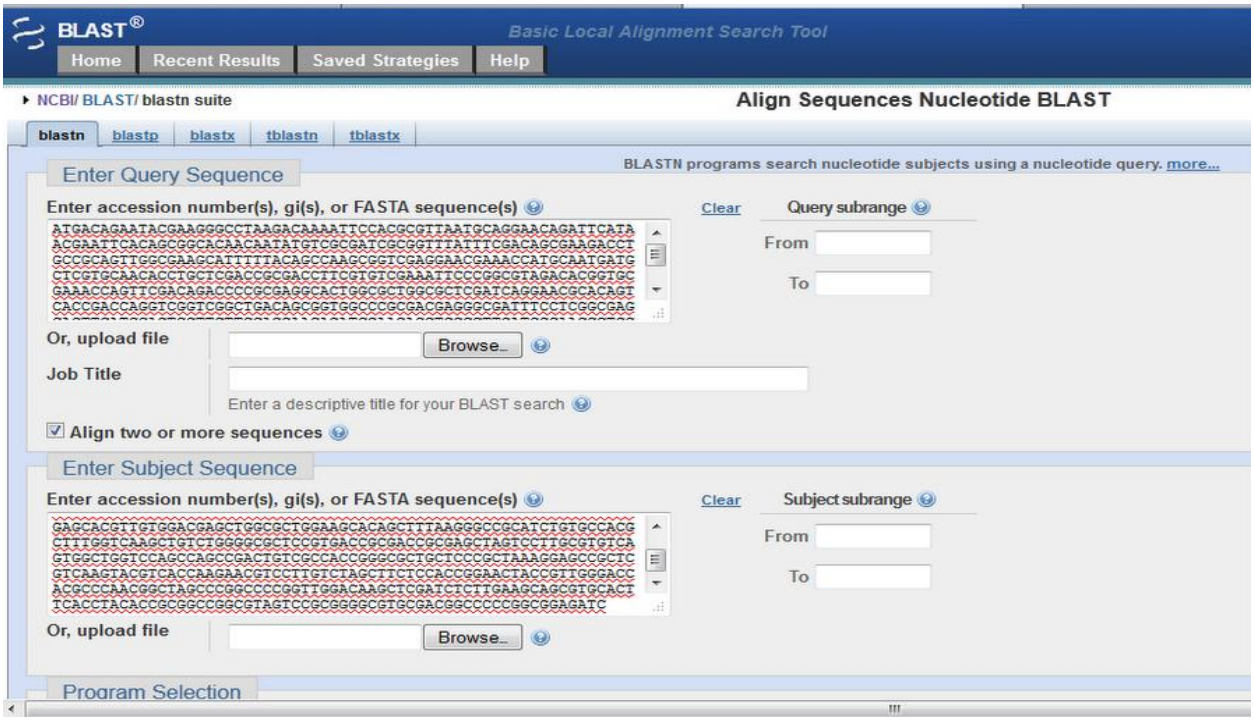
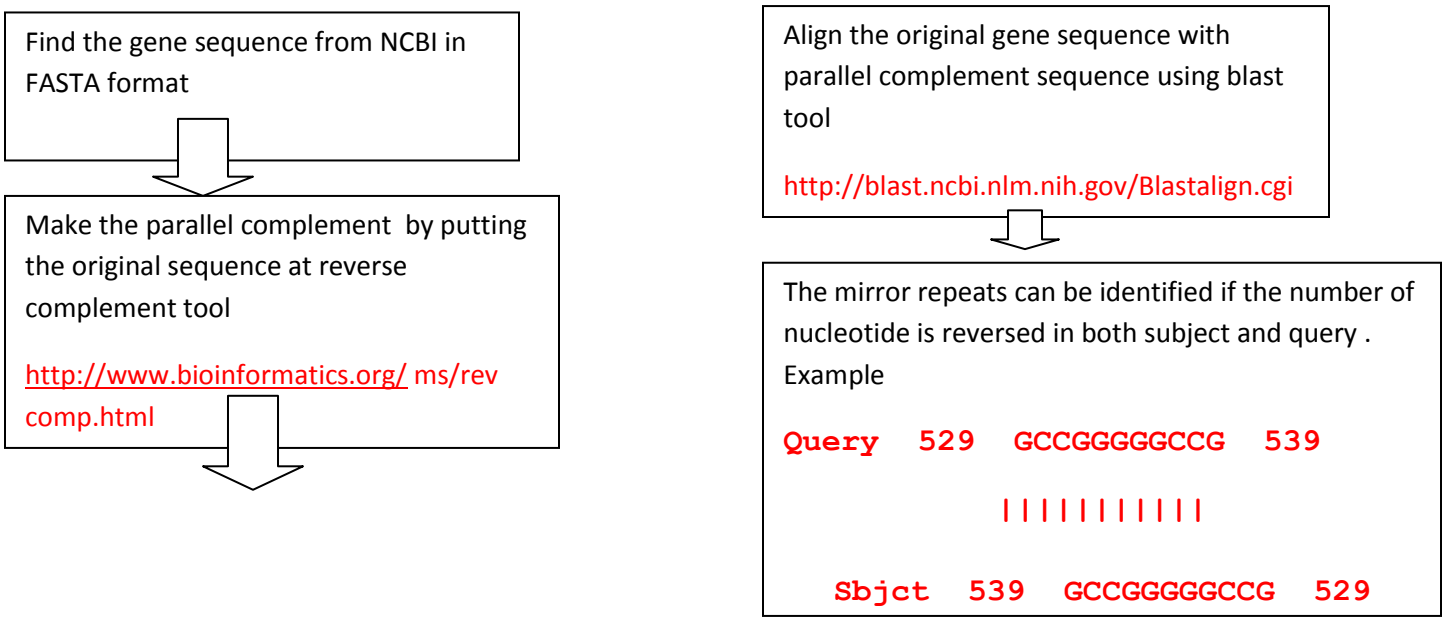


Figure 2: *shows NCBI BLAST*

>Mirror repeats analysis: It is mirror repeats when the numbers (equal) on subject and query reverse in opposite direction but it is not mirror repeat if the numbers on subject and query are different.



The figure 3: *Depicts the simple strategy for the identification of mirror repeat (Pattabiraman, 2014)*

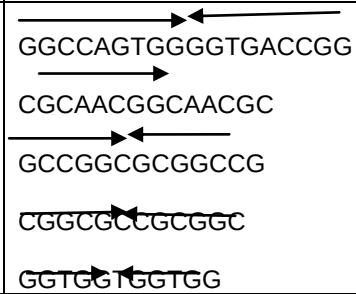
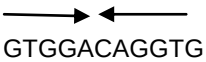
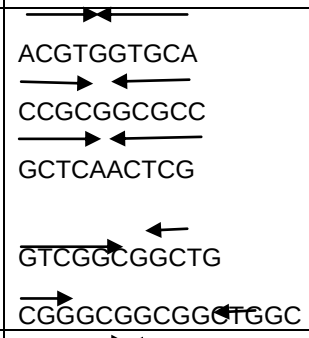
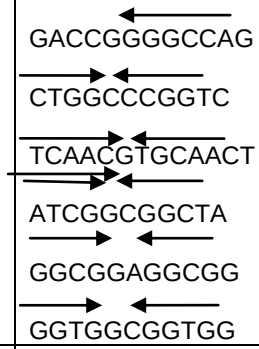
S/N	Gene's name	Gene ID
1	DNA topoisomerase I topA	Gene ID: 885608
2	bacterioferritin bfrB	Gene ID: 886176
3	iron-regulated lsr2	Gene ID: 885580
4	phosphoserine/ threonine phosphatase PstP	Gene ID: 887070
5	isocitrate lyase icl	Gene ID: 886291
6	membrane protein MmpS5	Gene ID: 888233
7	chromosomal replication initiator protein dnaA	Gene ID: 885041
8	acyl-CoA reductase acrA1	Gene ID: 887950
9	recN DNA repair protein	Gene ID: 885805
10	bacterioferritin bfrA Rv1876	Gene ID: 885767
11	transmembrane transport protein mmpL5	Gene ID: 888219
12	isochorismate synthase entC	Gene ID: 888824
13	katG catalase-peroxidase	Gene ID: 885638
14	lipoprotein LprG	Gene ID: 886700
15	beta-glucosidase BglS	Gene ID: 886780
16	rpsL 30S ribosomal protein S12	Gene ID: 888259
17	fabG1 3-oxoacyl-ACP reductase FabG1	Gene ID: 886551
18	DNA replication and repair protein recf	Gene ID: 887089
19	inhA NADH-dependent enoyl-[ACP] reductase	Gene ID: 886523
20	bacterioferritin bfrA	Gene ID: 885767

Table 1: *Various types of mirror repeats observed in Mycobacterium tuberculosis H37Rv using FPCB strategy.*

Result and Discussion

This study applies bioinformatic methods to the identification and function of mirror repeats in the M. tuberculosis H37Rv genomes. Mycobacterium tuberculosis, the causative agent of tuberculosis, is one of the most successful and deadly pathogens that kills millions of people. Mycobacterium tuberculosis, H37Rv strain was analyzed in order to improve our understanding of the biology of Mycobacterium tuberculosis in causing the disease and to help develop of new therapeutic interventions. Using computer applications, mirror repeats were found in mycobacterium tuberculosis H37Rv strain from 18 different genes out of 20 genes. It is known that a single gene may perform different function and a gene may be found in multiple species which may perform either the same or different function. Most of the genes (18 out of 20 genes) analyzed using FPCB strategy in this study were found to have the presence of the mirror repeats. It will be a high significant to find the role of mirror DNA at molecular level in respect to transcription and translation process occurring in a cell. Up till now the exact function of mirror repeats has not been elucidated. Future studies may be required to determine the exact role of mirror repeats.

MIRROR REPEATS OF MYCOBACTERIUM TUBERCULOSIS				
Sl/no.	Gene/cds	Sequence reference	No. of mirror repeats	Mirror repeats
1	DNA topoisomerase I topA	Gene ID: 885608	11	CAGCGGCCGGCGAC CGCGCCGCCGCCGAGC TGGAAGAAGGT CGCCGCGCCGC CGCGCCGGCGGCCGCGC ACCCGCAAGGTGAAGAACGCCCA GTCGGCGGCGAGGCTGACGATGCCGAGCGGCGACTG GGCAAGAACGG CTCAGCGACTC GGAGTGGACCCCGCCTCGGGTGAGG CCGGCAAAACGGCC
2	bacterioferritin bfrB	Gene ID: 886176	1	GCCGGGGGCCG
3	iron-regulated lsr2	Gene ID: 885580	1	GCCGGCGCGGCCG
4	phosphoserine/ threonine phosphatase PstP	Gene ID: 887070	1	CCGGCCTCGGCC
5	isocitrate lyase icl	Gene ID: 886291	2	GGGCGTTGCGGG TGCCACCACCGT
6	membrane protein	Gene ID: 888233	1	GAACTCTCAAG

	MmpS5			
7	acyl-CoA reductase acrA1	Gene ID: 887950	5	
8	chromosomal replication initiator protein dnaA	Gene ID: 885041	2	
9	stearoyl-CoA 9-desaturas desA3	Gene ID: 888821	1	
10	recN DNA repair protein	Gene ID: 885805	5	
11	transmembrane transport protein mmpL5	Gene ID: 888219	2	
12	beta-glucosidase BglS	Gene ID: 886780	7	
13	isochorismate synthase entC	Gene ID: 888824	2	
14	katG catalase-peroxidase	Gene ID: 885638	3	
15	fabG1 3-oxoacyl-ACP reductase FabG1	Gene ID: 886551	0	No mirror repeats

16	lipoprotein LprG	Gene ID: 886700	2	GGCTGCTCGTCGG CTGGGTGGGTC
17	rpsL 30S ribosomal protein S12	Gene ID: 888259	1	CCGAAGAAGCC
18	DNA replication and repair proteinrecf	Gene ID: 887089	0	No mirror repeats
19	bacterioferritin bfrA Rv1876	Gene ID: 885767	1	GGCG-GCCCACACCCGCGCGG
20	inhA NADH-dependent enoyl-[ACP] reductase	Gene ID: 886523	1	TTCGTATGCTT

Table 2: *Shows summary of different types of mirror repeats observed in various mycobacterium tuberculosis using FPCB strategy their NCBI number.*

The red sequences are mirror to the blue sequences.

References

[1] Pattabiraman N: Can the Double Helix Be Parallel? Biopolymers. 1986; 25(9): 1603–1606

[2] Dorothy M. Lang, (2007). Imperfect DNA mirror repeats in the gag gene of HIV-1 (HXB2)identify key functional domains and coincide with protein structuralelements in each of the mature proteins. Virology Journal 2007. (2007).

[3] Yo-Cheng Chang Chuan-Hsiung Chang, (2003). Common Repeat Sequences in Bacterial Genomes. Journal of Medical and Biological Engineering, 23(2): 65-72

[4] http://www.ncbi.nlm.nih.gov

[5] http://www.bioinformatics.org/sms/rev_comp.html

[6] http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAS_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&QUERY=&SUBJECTS=)