American-Eurasian Journal of Scientific Research 11 (4): 250-264, 2016
ISSN 1818-6785

DOI: 10.5829/idosi.aejsr.2016.11.4.22899

Gene Amplification of Isolates from Molecular Data Through Estimated Phylogenetic Tree and Dendrogram Analysis

[1]Rajashree Nanda, [2]Banani Pattanayak and [1]N.K. Dhal

[1]DRWA(ICAR), Opp. Kalinga Studio, Baramunda, Bhubaneswar, Odisha, India
[2]CSIR-IMMT, Bhubaneswar, India

Abstract: Recent advances in molecular biology techniques have opened the door for culture-independent diagnostic methods. Immunological detection and identification by use of distinctive metabolites and nucleic acid probes are the tools most often used for diagnosis. One such technique is PCR, which has been shown to be useful for the culture-independent diagnosis of various microbial infections. Sequence similarity tools, such as BLAST, seek sequences most similar to a query from a database of sequences. They return results significantly similar to the query sequence and that are typically highly similar to each other. Most sequence analysis tasks in bioinformatics require an exploratory approach, where the initial results guide the user to new searches. However, diversity has not yet been considered an integral component of sequence search tools for this discipline. We evaluate the effectiveness of the targeted sequence with its similarity by BLAST results. It has been demonstrated that the utility of dendrograms at representing the essential features of the hierarchical structure of the isosurfaces for molecular line data cubes. The conclusions from this study were obtained by analyzing the gene sequence from the molecular analysis through different sequence analyzer and techniques.

Key words: PCR · Alignement · BLAST · Phylogenetic · Dendrogram

INTRODUCTION

Biologists are nowadays confronted with two main problems, namely the exponentially growing volume of biological data of high variety, heterogeneity and semi-structured nature and the increasing complexity of biological applications and methods afflicted with an inherent lack of biological knowledge. As a result, many and important challenges in biology and genomics are challenges in computing and here especially in advanced information management and algorithmic design. The currently most widely used and accepted tool for conducting similarity searches on gene sequences is BLAST (Basic Local Alignment Search Tool) [1]. BLAST comprises a set of similarity search programs that employ heuristic algorithms and techniques to detect relationships between gene sequences and rank the computed 'hits' statistically. An essential problem for the biologist is currently the processing and evaluation of BLAST query results, since a BLAST search yields its result exclusively in a textual format (e.g., ASCII, HTML, XML). This format has the benefit of being application-neutral but at the same time impedes its direct analysis. In this paper, we describe a new powerful tool, called BlastQuest, for managing BLAST results stemming from multiple individual queries. This tool provides the biologist with interactive and Web-enabled query, analysis and visualization facilities beyond what is possible by current BLAST interfaces. In particular, BLAST results from multiple queries are imported, structured and stored in a relational database to support a series of built-in analysis operations that can be used to select, filter, group and order these data efficiently and without referring to the original BLAST result files. In addition, users have the option to interact with the data through a forms-based query interface. BlastQuest is being supported by the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida and is used by campus researchers and their collaborators across the United States.

Corresponding Author: Rajashree Nanda, DRWA(ICAR), Opp. Kalinga Studio, Baramunda, Bhubaneswar, Odisha, India.

Sequence similarity search is one of the earliest and most commonly employed tools of bioinformatics for molecular biologists. In current sequence search tools, the results retrieved from the database are typically also highly similar to each other. For many bioinformatics tasks, the result set needs to be diversified to produce a subset of results containing sequences well aligned with the query but sufficiently different from each other. This need is apparent in the use of non-redundant databases such as the nr database used in BLAST [1,2]. Sequence similarity search is an area that would benefit from more diverse results rather than just top-similar results. Identifying all functional domains of a query sequence, which may be comprised of separate homologous domains in different sequences, can only be established by an approach whose main purpose is to cover most of the query sequence rather than finding the most similar sequence. In this paper, we formalize the problem of diversification and investigate methods to post-process results from the commonly employed search tools to remove redundancy from the results and enable exploratory browsing. An example of such searches is to find proteins with different functions but similar enough to the query sequence. Different segments of the primary structure may correspond to different functional domains. Tools such as BLAST incorporate a domain identification step and present the identified domains to the user in addition to the query results. However, domain identification is limited to known, characterized domains and novel domains in the query sequence will be overlooked by this approach. Such novel domains may be shared by some of the database sequences and a diverse search may identify these regions. For this purpose, finding a diverse set of regions with similar segments would be a more appropriate approach than simply investigating the top-similar sequences. With our proposed method, we are also able to control the effect of diversification based on the dissimilarity of biological functions of sequences.

Sequence alignment is utilized to arrange the sequences of DNA, RNA, or amino acid sequences to identify regions of similarity. Global alignment follows a general similarity measure and attempts to align each residue in every sequence using gaps and local alignment focuses on determining similar sub-regions. Sequence search tools such as BLAST [1,2] and FASTA [3] seek similar sequences to a given query in large sequence databases. Our proposed approach is applicable to post-process the results of any sequence similarity search tool. However, for our experiments, we focus on BLAST and on one of its popular variants, Position-Specific Iterated BLAST (PSI-BLAST) [2], which seeks locally similar sequences on protein databases by using profiles updated dynamically in iterations.

BLAST compares nucleotide or protein sequences to large sequence databases, calculates the statistical significance of matches and returns the results with attributes such as query coverage, total score, max score, e-value and maximal identity. Total score is the sum of the scores of all high scoring pairs(HSPs) from the same database sequence [1]. Unlike BLAST, PSI-BLAST profiles are built by considering evolutionary relationships and using them enables detection of a protein's distant relatives. As diversification is applied for blastp, which is original BLAST for protein-protein search and for PSI-BLAST, it could be possible to use for all versions of BLAST including nucleotide-nucleotide BLAST, blastn and translation BLAST types such as blastx, tblastx and tblastn. The translation models may compare nucleotides to amino acids or vice versa.

Diversification of search results aims to produce results similar to the query but different from each other. Although there is no prior work on diversification in sequence search, the notions of diversity and novelty are present in the context of information retrieval and recommendation systems. Carbonell and Goldstein [4] were the first to introduce Maximal Marginal Relevance (MMR) for text retrieval and summarization. This method builds a result set by maximizing the query relevance and minimizing the similarity between documents in the result set. It uses a parameter (λ) that specifies the proportions of relevancy and diversity.

Phylogenetic analysis is a powerful tool to study the relationships among sequences. It provides insights of gene or sequence which includes relationship, origins, evolution and possible changes in structural and functional properties. These evolutionary relationships are represented by a branching diagram, or tree, with branches joined by nodes and leading to terminal at the tips of the tree. Following terminologies should be known for phylogenetic analysis. Terminology used for phylogenetic study:

- Node: a node represents a taxonomic unit. This can be at axon (an existing species) or an ancestor (unknown species: represents the ancestor of 2 or more species).
- Branch: defines the relationship between the taxa in terms of descent and ancestry.
- Topology: is the branching pattern.

- Branch length: often represents the number of changes that have occurred in that branch.
- Root: is the common ancestor of all taxa.
- Distance scale: scale which represents the number of differences between sequences (e.g. 0.1 means 10 % differences between two sequences). DNA sequences are preferable for phylogenetic analyses of closely related species. On the other hand, the amino acid sequences are used for phylogenetic analyses of more distant relationships. The sequences can be analysed using many computer programs. The methods most often used as following example;

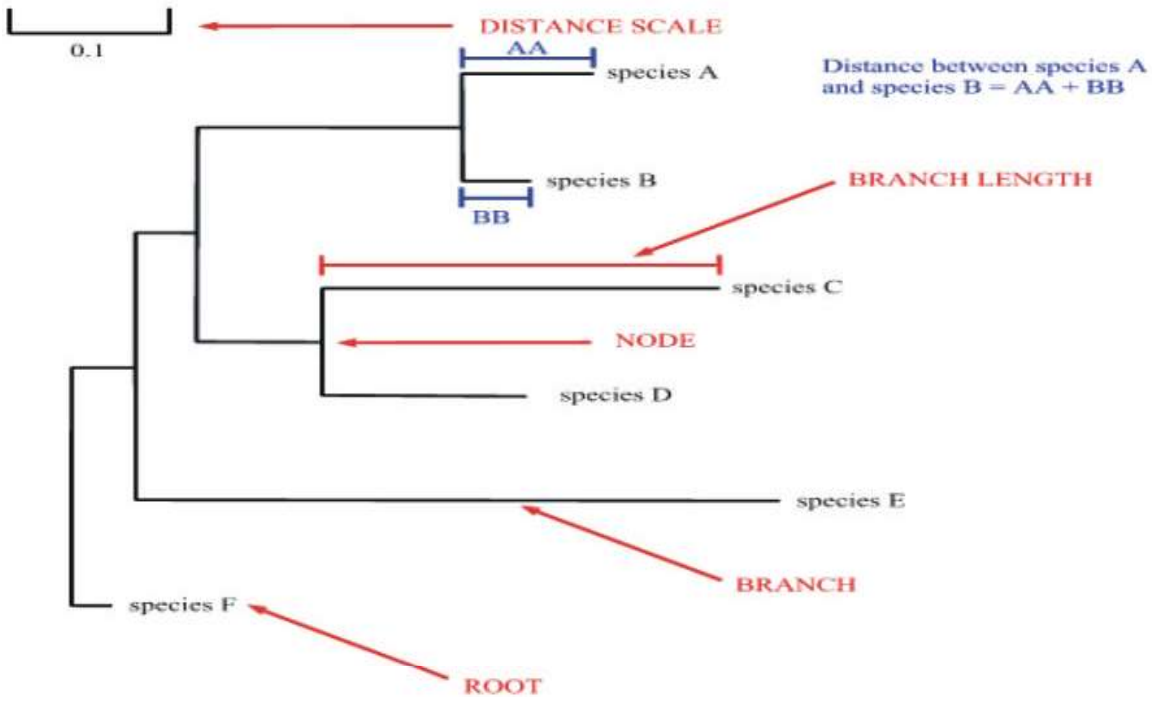

Fig. 1: Terminologies used for phylogeny

DNA sequences are preferable for phylogenetic analyses of closely related species. On the other hand, the amino acid sequences are used for phylogenetic analyses of more distant relationships. The sequences can be analysed using many computer programs. The methods most often used for phylogenetic analysis are neighbor-joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference.

Phylogenetic analysis Programs and its availability

Software	Usage	Availability
ClustalX	Performing MSA	Free
DAMBE	Trimming external gaps	Free
MEGA	For construction of phylogenetic tree using NJ method	Free
PHYLIP	For construction of phylogenetic tree using MP method	Free

After that Neighbor-joining tree construction by MEGA and Maximum Parsimony tree construction by Phylip method are followed for the alignment.

Practically, the dendrogram of an N-dimensional intensity image is constructed by first identifying the local maxima that will comprise the top level of the dendrogram hierarchy. Then, the data are contoured with a large number of levels. For each contour value beginning with the maximum level, the dendrogram algorithm checks whether each pair of previously distinct regions have merged together. If so, the contour level and which surfaces merged are recorded and the next contour level is considered. We enforce binary mergers: if three or more distinct objects merge into a single object between one contour level and the next, we refine the separation between contour levels so each merger involves only two objects. The dendrogram (tree diagram) is constructed by drawing vertical segments corresponding to contour levels where the topology of the surfaces are unchanged and connecting corresponding branches at the levels where isosurfaces merge.

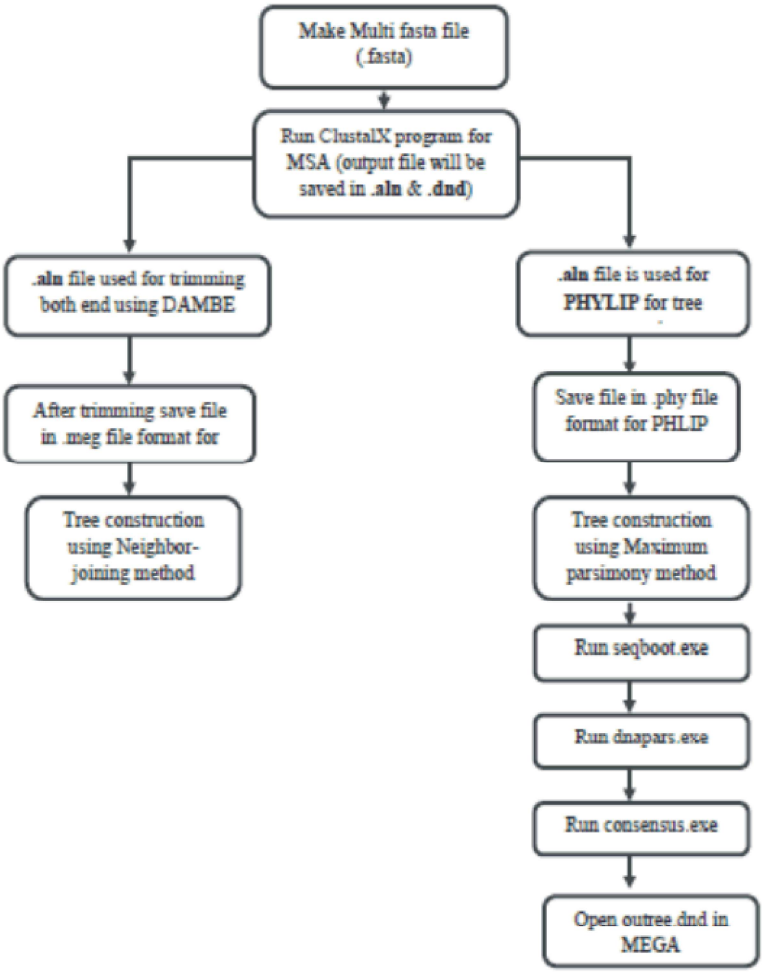

Fig. 2: Flowchart of phylogenetic analysis to determine the evolutionary relationship of bacteria with its neighbours.

MATERIALS AND METHODS

Molecular Taxonomy of the Potent Fungal Isolates

DNA Extraction from Pure Fungal Cultures: Approximately 50 gm of mycelium (Wet weight) from a fresh culture plate was suspended in 500 µl of lysozyme solution and incubated at 37°C until the cells become translucent. Then 250 µl of 2%SDS was added followed by the addition of protease K(4µl) and shaken well and kept at 55°C for nearly one hour by shaking thoroughly at regular intervals.250 µl of neutral phenol-chloroform Solution.(Phenol 25: chloroform 24:Isoamylalcohol 1/v/v) was added and thoroughly mixed for 30 sec, then spinned for 5-10 minutes in the microcentrifuge (at 4000 rpm) and the supernatant was removed, leaving the white interface behind. This step was repeated was repeated twice or until the absence of interface 0.1 µl of 3 M sodium acetate (P^H – 4.8) was added and mixed well; then 1 ml isopropanol was added and mixed well, again then 1 ml isopropanol was added and mixed well again. The thoroughly mixed sample was KOH at 0 to 5°C for 15 minutes. Then sample was centrifuged at 4°C for 6 months. The supernatant was carefully poured off . 600 µl of 75% ethanol was added and again centrifuged at 10, 000 rpm for 2 minute at 4°C until it was used [5].

PCR Amplification of 18Sr DNA: All PCR reactions were performed in a Robocycler [6] using PCR Master Mix [7]. Fragments of about 500 bp in size, corresponding to either the ITS1 or ITS2 region of the ribosomal-DNA ITS region, were first amplified with the primer pairs ITS1 forward and ITS2 reverse, or ITS3 and ITS4, according to [8] White *et al*. Primer pair EF3RCN [9] with the reverse primer ITS4 (as above) was also tested. All reactions were carried out in a 25-mL volume containing 12.5 pmol of each primer and 1.5–2.5 mL DNA as template, following

the manufacturer's instructions for PCR Master Mix. The thermocycling program was as follows: 5 min denaturation at 94 1C, followed by 30 cycles of 1 min denaturation at 94 1C, 1 min annealing at 60 1C and 1 min extension at 72 1C. Five minutes at 72 1C were used as a final extension step. For DGGE analysis, a nested PCR was performed in a total volume of 100 mL (2 50 mL reaction size), each with 3.5 mL of PCR product from the first amplification as template DNA. The same forward and reverse primers were used, with a 37-base GC-clamp attached to either 50 -end of the forward primer, to stabilise the melting behaviour of the DNA fragments (Muyzer *et al*., 1993). The cycling scheme was as follows: 5 min denaturation at 94 1C, followed by 35 cycles of 1 min denaturation at 94 1C, 1 min annealing at 58 1C for ITS1-GC primer or 62 1C for ITS2-GC primer, 1 min extension at 72 1C with a final extension step at 72 1C. PCR products were analysed by electrophoresis in a 2% (w/v) agarose gel.

Denaturing Gradient Gel Electrophoresis (DGGE): For the genetic fingerprint of the amplified ITS regions, 50-100 mL of PCR products containing the GC-clamp were precipitated with 96% ethanol at 20 1C overnight, resuspended in 15 mL double-distilled H2O and separated by DGGE. Gel electrophoresis was performed as previously described [10] in 0.5 TAE (20 mM Tris, 10 mM acetate, 0.5 mM Na2EDTA; pH 7.8) with 8% (w/v) acrylamide gels containing a gradient of denaturants in a D-GENE-System (Bio-Rad, Munich, Germany). The gradient of denaturants, as well as running conditions, was optimised in this study and is given in the results section. After completion of electrophoresis, gels were stained in an ethidium bromide solution and documented with a UVP documentation system (Bio-Rad Quantity-One, Munich, Germany).

18S r DNA Sequencing: The purified fragments were directly sequenced bidirectionally. The reaction mixture (10µl) contains 4 µl of big dye terminator ready reaction mix, 1 µl of template (100 ng/µl), 2 µl of primer(10Pmol/λ), 3 µl of Milli Q Water and the PCR condition is of 1 min at 96°C in initial denaturation step(25 cycles).The data was analysed using applied biosystem DNA editing and assembly software and sequence comparisons were obtained using the microseq software.

Sample Identification: A distance matrix is generated using the Jukes-Cantor corrected distance model. When generating the distance matrix, only alignment model positions are used, alignment inserts are ignored and the maximum comparable position is 200. Identification was done on the pair-wise alignment algorithms and phylogenetic tree.

Phylogenetic Analysis: The sequence analysis was carried out using bioinformatic tool BLAST of NCBI based on maximum identity score first few sequences were selected and aligned using multiple sequence alignment software ClustalW2 Dendrogram was constructed. Sequence similarity search was made for the 18SrDNA sequence of the desired fungal strains. 18SrDNA gene sequence of the strain was manually aligned with species members of the same genus family obtained from the Gene Bank Data Base. The Evolutionary tree was inferred by using neighbour-joining method [11]. The CLUSTAL X Package was used for multiple alignment and identification of the Strain.

Molecular Identification of Potential Bacterial Isolates: Two Cr resistant potential bacterial isolates has been selected on the basis of their Chromium and Cadmium tolerance and reduction capacity like i.e.C.W.B2 and C.W.B4 and two Cadmium resistant Bacterial isolates like i.e. CD.W.B2 and CD.W.B6. After Biochemical and Enzymatic Characterization, these potential Bacterial isolates was identified by 16SrRNA analysis.

Molecular Identification of C.W.B2 Strain: Genomic DNA isolation: The highest yield of DNA to be obtained was preserved with TE buffer, yields of 25 µl total DNA from the 1.5ml broth. The extracted DNA was electrophoresed on 1% Agarose-ethidiumbromide. The bands were observed under UV-Transilluminator.

RESULT AND DISCUSSION

PCR Primers and Amplification Conditions: The DNA extraction provided a good yield of DNA, which could be used in 16 S genes amplification. With using Bacterial universal primers, 518F (50-CCAGCAGCCGCGGTAATACG-30) and 800R (5'-TACCAGGGTATCTAATCC-30). Satisfactory amplification products were obtained after adding large quantities of Taq DNA Polymerase .The amplified Sample and 100 bp DNA ladder was electrophoresed on 2% agarose gel and compared. The bands were observed under UV-Transilluminator.

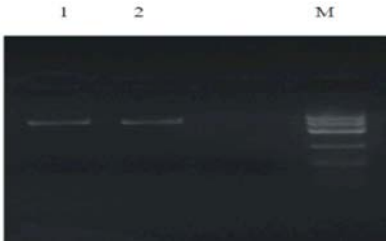

Fig. 3: Agarose gel electrophoresis of PCR products

Fig. 4: Amplicon photograph

Amplification of DNA extracted from the bacterial strain. Lane 1, PCR marker (numbers on the left are in kilobases); lane 2, results obtained with primer pair 16sF and 16sR.

16S rRNA Based Identification: The amplified product of 1330 bp was sequenced by using the ABI 3600 system. Sequencing was carried out. Later, nucleotide sequence data were deposited in the Gen-Bank sequence database. Purified PCR products were sequenced using internal overlapping primers (Pidiyar V, 2002). Sequence was initially analyzed at NCBI server (http://www.ncbi.nlm.nih.gov/) using BLAST (blastn) tool and corresponding sequences were downloaded. Similarity matrix was prepared using DNA dist program in PHYLIP analysis package using Jukes Cantor corrections. Phylogenetic tree was constructed by the neighbor-joining method using the MEGA package.

Contig Sequence:

>CRW.B2

ATGCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGGACGGGTGA
GTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGAGCTAATAC
CGGATAACATTTTCTCTTGCATAAGAGAAAATTGAAAGATGGTTTCGGCTATCACTTACAG
ATGGGCCCGCGGTGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCAT
AGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTACAAGAGTA
ACTGCTTGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGC
GGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGG
GGAACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATGCGTAGAG
ATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTAACTGACGCTGAGGCGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGT
GCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCT
GGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTG
ACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGACAGGTGGTGCATGGTTG
TCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTT
AGTTGCCAGCATTTAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGT
GGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA
TGGTACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTCTCAGTTC
GGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCA
TGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGT
AACACCCGAAGTCGGTGGGGTAACCTTTATGGAGCCAGCCG

Blast Alignment Result
>CRW.B2

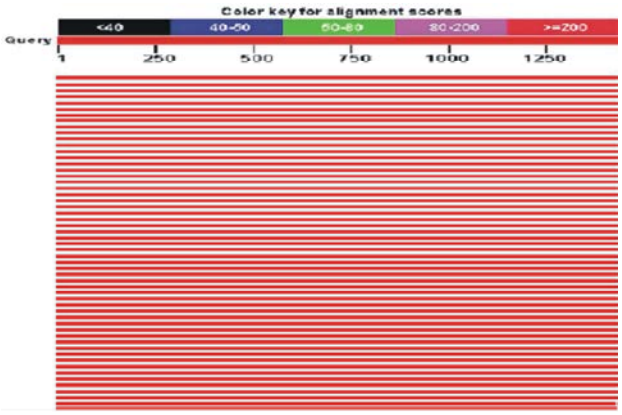

Fig. 5: The graphic summary of BLAST

This is a sample blast result from BLAST on the NCBI site, using a fungal sequence as a query. The Graphic Summary (Fig. 5) shows alignments (as colored boxes) of database matches to our Query sequence (solid red bar under the color key). As its name suggests, BLAST (Basic Local Alignment Search Tool) is designed to identify local regions of sequence similarity. This means that BLAST may report multiple discrete regions of sequence similarity between a query sequence and a subject sequence in a database. For example, if a spliced (mature) mRNA sequence is aligned to the unknown genomic sequence, we would expect to see multiple alignment blocks (many of which likely correspond to transcribed exons) in our BLAST output. Regions of the genomic sequence without significant alignment that fall between these exons are likely to be introns. The color of the boxes corresponds to the score (S) of the alignment, with red representing the highest alignment scores. Generally, the higher the alignment score, the more significant the hit. In this case, we notice that the top one hits match much better to our sequence than the remaining BLAST hits. We also see that these three database matches span almost the entire length of our Query sequence.

Multiple sequence Alignment

>CRW.B2

```
CLUSTAL 2.1 multiple sequence alignment

gi-631252602-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 79
gi-298239721-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 59
gi-392514580-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 99
gi-538283896-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 56
gi-723266040-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 58
gi-669636729-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 85
gi-353260321-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 100
gi-722093272-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 82
gi-340026005-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 100
CRWB2                       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 52
gi-507481883-B.flexus       GCAAGTCGAGCGAACTGATTAGAAGCTTGCTTCTATGACGTTAGCGGCGG 54
                            **************************************************

gi-631252602-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 129
gi-298239721-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 109
gi-392514580-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 149
gi-538283896-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 106
gi-723266040-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 108
gi-669636729-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 135
gi-353260321-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 150
gi-722093272-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 132
gi-340026005-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 150
CRWB2                       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 102
gi-507481883-B.flexus       ACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTCCG 104
                            **************************************************

gi-631252602-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 179
gi-298239721-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 159
gi-392514580-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 199
gi-538283896-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 156
gi-723266040-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 158
gi-669636729-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 185
gi-353260321-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 200
gi-722093272-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 182
gi-340026005-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 200
CRWB2                       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 152
gi-507481883-B.flexus       GGAAACCGGAGCTAATACCGGATAACATTTTCTCTTGCATAAGAGAAAAT 154
                            **************************************************

gi-631252602-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 229
gi-298239721-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 209
gi-392514580-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 249
gi-538283896-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 206
gi-723266040-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 208
gi-669636729-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 235
gi-353260321-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 250
gi-722093272-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 232
gi-340026005-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 250
CRWB2                       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 202
gi-507481883-B.flexus       TGAAAGATGGTTTCGGCTATCACTTACAGATGGGCCCGCGGTGCATTAGC 204
                            **************************************************
```

```
gi-631252602-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 279
gi-298239721-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 259
gi-392514580-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 299
gi-538283896-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 256
gi-723266040-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 258
gi-669636729-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 285
gi-353260321-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 300
gi-722093272-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 282
gi-340026005-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 300
CRWB2                      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 252
gi-507481883-B.flexus      TAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCATAGCCGACCTGA 254
                           **************************************************

gi-631252602-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 329
gi-298239721-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 309
gi-392514580-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 349
gi-538283896-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 306
gi-723266040-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 308
gi-669636729-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 335
gi-353260321-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 350
gi-722093272-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 332
gi-340026005-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 350
CRWB2                      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 302
gi-507481883-B.flexus      GAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG 304
                           **************************************************

gi-631252602-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 379
gi-298239721-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 359
gi-392514580-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 399
gi-538283896-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 356
gi-723266040-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 358
gi-669636729-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 385
gi-353260321-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 400
gi-722093272-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 382
gi-340026005-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 400
CRWB2                      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 352
gi-507481883-B.flexus      AGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAAC 354
                           **************************************************

gi-631252602-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 429
gi-298239721-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 409
gi-392514580-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 449
gi-538283896-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 406
gi-723266040-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 408
gi-669636729-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 435
gi-353260321-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 450
gi-722093272-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 432
gi-340026005-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 450
CRWB2                      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 402
gi-507481883-B.flexus      GCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA 404
                           **************************************************
```

```
gi-631252602-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 479
gi-298239721-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 459
gi-392514580-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 499
gi-538283896-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 456
gi-723266040-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 458
gi-669636729-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 485
gi-353260321-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 500
gi-722093272-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 482
gi-340026005-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 500
CRWB2                      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 452
gi-507481883-B.flexus      AGAACAAGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCAGAA 454
                           ************************************************

gi-631252602-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 529
gi-298239721-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 509
gi-392514580-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 549
gi-538283896-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 506
gi-723266040-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 508
gi-669636729-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 535
gi-353260321-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 550
gi-722093272-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 532
gi-340026005-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 550
CRWB2                      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 502
gi-507481883-B.flexus      AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAG 504
                           ************************************************

gi-631252602-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 579
gi-298239721-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 559
gi-392514580-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 599
gi-538283896-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 556
gi-723266040-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 558
gi-669636729-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 585
gi-353260321-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 600
gi-722093272-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 582
gi-340026005-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 600
CRWB2                      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 552
gi-507481883-B.flexus      CGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGT 554
                           ************************************************

gi-631252602-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 629
gi-298239721-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 609
gi-392514580-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 649
gi-538283896-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 606
gi-723266040-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 608
gi-669636729-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 635
gi-353260321-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 650
gi-722093272-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 632
gi-340026005-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 650
CRWB2                      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 602
gi-507481883-B.flexus      CTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGG 604
                           ************************************************
```

```
gi-631252602-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 679
gi-298239721-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 659
gi-392514580-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 699
gi-538283896-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 656
gi-723266040-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 658
gi-669636729-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 685
gi-353260321-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 700
gi-722093272-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 682
gi-340026005-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 700
CRWB2                      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 652
gi-507481883-B.flexus      AACTTGAGTGCAGAAGAGAAAAGCGGAATTCCACGTGTAGCGGTGAAATG 654
                           **************************************************

gi-631252602-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 729
gi-298239721-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 709
gi-392514580-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 749
gi-538283896-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 706
gi-723266040-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 708
gi-669636729-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 735
gi-353260321-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 750
gi-722093272-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 732
gi-340026005-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 750
CRWB2                      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 702
gi-507481883-B.flexus      CGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTA 704
                           **************************************************

gi-631252602-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 779
gi-298239721-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 759
gi-392514580-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 799
gi-538283896-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 756
gi-723266040-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 758
gi-669636729-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 785
gi-353260321-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 800
gi-722093272-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 782
gi-340026005-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 800
CRWB2                      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 752
gi-507481883-B.flexus      ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCT 754
                           **************************************************

gi-631252602-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 829
gi-298239721-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 809
gi-392514580-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 849
gi-538283896-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 806
gi-723266040-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 808
gi-669636729-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 835
gi-353260321-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 850
gi-722093272-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 832
gi-340026005-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 850
CRWB2                      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 802
gi-507481883-B.flexus      GGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCC 804
                           **************************************************
```

```
gi-631252602-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 879
gi-298239721-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 859
gi-392514580-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 899
gi-538283896-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 856
gi-723266040-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 858
gi-669636729-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 885
gi-353260321-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 900
gi-722093272-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 882
gi-340026005-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 900
CRWB2                       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 852
gi-507481883-B.flexus       CTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTC 854
                            **************************************************

gi-631252602-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 929
gi-298239721-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 909
gi-392514580-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 949
gi-538283896-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 906
gi-723266040-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 908
gi-669636729-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 935
gi-353260321-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 950
gi-722093272-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 932
gi-340026005-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 950
CRWB2                       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 902
gi-507481883-B.flexus       GCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG 904
                            **************************************************

gi-631252602-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 979
gi-298239721-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 959
gi-392514580-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 999
gi-538283896-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 956
gi-723266040-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 958
gi-669636729-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 985
gi-353260321-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 1000
gi-722093272-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 982
gi-340026005-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 1000
CRWB2                       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 952
gi-507481883-B.flexus       CATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACAT 954
                            **************************************************

gi-631252602-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1029
gi-298239721-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1009
gi-392514580-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1049
gi-538283896-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1006
gi-723266040-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1008
gi-669636729-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1035
gi-353260321-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1050
gi-722093272-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1032
gi-340026005-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1050
CRWB2                       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1002
gi-507481883-B.flexus       CCTCTGACAACTCTAGAGATAGAGCGTTCCCCTTCGGGGGACAGAGTGAC 1004
                            **************************************************
```

```
gi-631252602-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1079
gi-298239721-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1059
gi-392514580-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1099
gi-538283896-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1056
gi-723266040-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1058
gi-669636729-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1085
gi-353260321-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1100
gi-722093272-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1082
gi-340026005-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1100
CRWB2                      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1052
gi-507481883-B.flexus      AGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGT 1054
                           **************************************************

gi-631252602-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1129
gi-298239721-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1109
gi-392514580-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1149
gi-538283896-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1106
gi-723266040-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1108
gi-669636729-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1135
gi-353260321-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1150
gi-722093272-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1132
gi-340026005-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1150
CRWB2                      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1102
gi-507481883-B.flexus      CCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCATTTAGTTGGGCA 1104
                           **************************************************

gi-631252602-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1179
gi-298239721-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1159
gi-392514580-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1199
gi-538283896-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1156
gi-723266040-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1158
gi-669636729-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1185
gi-353260321-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1200
gi-722093272-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1182
gi-340026005-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1200
CRWB2                      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1152
gi-507481883-B.flexus      CTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA 1154
                           **************************************************

gi-631252602-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1229
gi-298239721-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1209
gi-392514580-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1249
gi-538283896-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1206
gi-723266040-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1208
gi-669636729-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1235
gi-353260321-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1250
gi-722093272-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1232
gi-340026005-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1250
CRWB2                      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1202
gi-507481883-B.flexus      AATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGG 1204
                           **************************************************
```

```
gi-631252602-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1279
gi-298239721-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1259
gi-392514580-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1299
gi-538283896-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1256
gi-723266040-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1258
gi-669636729-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1285
gi-353260321-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1300
gi-722093272-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1282
gi-340026005-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1300
CRWB2                      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1252
gi-507481883-B.flexus      TACAAAGGGCTGCAAGACCGCGAGGTCAAGCCAATCCCATAAAACCATTC 1254
                           ************************************************** 

gi-631252602-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1329
gi-298239721-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1309
gi-392514580-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1349
gi-538283896-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1306
gi-723266040-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1308
gi-669636729-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1335
gi-353260321-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1350
gi-722093272-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1332
gi-340026005-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1350
CRWB2                      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1302
gi-507481883-B.flexus      TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTA 1304
                           ************************************************** 

gi-631252602-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1379
gi-298239721-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1359
gi-392514580-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1399
gi-538283896-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1356
gi-723266040-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1358
gi-669636729-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1385
gi-353260321-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1400
gi-722093272-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1382
gi-340026005-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1400
CRWB2                      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1352
gi-507481883-B.flexus      GTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA 1354
                           ************************************************** 

gi-631252602-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1429
gi-298239721-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1409
gi-392514580-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1449
gi-538283896-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1406
gi-723266040-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1408
gi-669636729-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1435
gi-353260321-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1450
gi-722093272-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1432
gi-340026005-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1450
CRWB2                      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1402
gi-507481883-B.flexus      CACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTA 1404
                           ************************************************** 
```

Unlike the pairwise alignments in BLAST that align only the best matching segments of the hit and query sequences, ClustalW finds the best alignment over the full length of each sequence submitted. The sequences formatted by the game are the full length sequences of the query and hits as they appear in GenBank or GenPept. The scores obtained as each sequence was aligned in a pairwise alignment with each other sequence. In a phylogenic tree, the alignment is built by first aligning the two most similar sequences and then adding the other sequences to the alignment in descending order of similarity. With aligned DNA sequences, the character * was only appeared and it specify that the same base is found at that position in all sequences in the alignment. It has been observed that there are a fair number (at least 35-50%) of similarities which spread over the full lengths of the sequences and they are likely to be related. The best sequence matches can be achieved through phylogenetic analysis based on dendrogram approach.

Table 1: Multiple Sequence Alignment search result

Number of sequences	10
Alignment Score	1404
Sequence format	Pearson
Sequence type	aa
Clustal W version	2.1
Guide tree file	DND

Phylogram

>CRW.B2

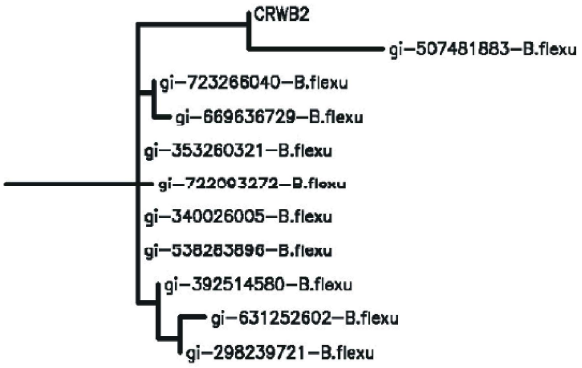

Fig. 5: Dendrogram of bacterial effective pattern based on similarity

The dendrogram based on bacterial effective pattern is represented in Fig. 5. The similarity and dissimilarity were grouped in different clusters. It has been shown that the three groups have the maximum linkage. The second cluster is grouped into two sub groups and the seventh cluster is grouped into two groups which shown the maximum linkages according to the similarity. In the present study, it was found that the gi-507481883 differed from other sequences by their major cluster grouping by having better result in all the aspect.

From the above dendrogram analysis it can be concluded that the gi-507481883 - B. flexus is the parent strain (species) of the genus as this has came in the root cluster with its unique identity. All other strain probably might have originated from parental B flexus in the process of genetic evolution. Thus dendrogram analysis revealed the possibility of this method as a taxonomic tool.

CONCLUSION

The culture “CRW.B2” shows significant similarity with Bacillus flexus based on nucleotidehomology and phylogenetic analysis.

REFERENCES

1. Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool, J. Mol. Biol., 215: 403-410.
2. Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang and Z. Zhang, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25: 3389-3402.
3. Pearson, W.R. and D.J. Lipman, 1988. Improved tools for biological sequence comparison, Proc Natl Acad Sci USA., 85: 2444-2448.
4. Carbonell, J. and J. Goldstein, 1998. The use of MMR, Diversity-based re ranking for reordering documents and producing summaries, In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp: 333-336.
5. Hopwood, D.A., F. Malpartida, H.M. Kieser, H. Ikeda, J. Duncan, I. Fujii, B.A.M. Rudd, H.G. Floss and S.O. Mura, 1985. Production of ‘hybrid’ antibiotics by genetic engineering, Nature. 314: 642-644.
6. Stratagene and La Jolla, 2007. USA.
7. Promega and Mannheim, 2003.Germany.
8. White, T.J., T. Bruns, S. Lee and J. Taylor, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: a guide to methods and applications. (Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds). Academic Press, New York, USA, pp: 315-322.

9. Lord, J.M., A.S. Markey and J. Marshall, 2002. Have frugivores influenced the evolution of fruit traits in New Zealand? In: Levey, D.J., Silva, W.R. and Galetti, M. (eds) Seed Dispersal and Frugivory: Ecology, Evolution and Conservation. CABI Publishing, Wallingford, UK, pp: 55-68.
10. Muyzer, G., E.C. De Waal and A.G. Uitterlinden, 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol., 59: 695-700.
11. Thompson J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res., 25: 4876-4882.