

COMPUTATIONAL ANALYSIS OF *SOLANUM LYCOPERSICUM* (TOMATO) *LYCOPERSICONES CULENTUM* REPRESENTING A LARGE RANGE OF INTRASPECIFIC VARIABILITY

UMA KUMARI¹ & ASHOK KUMAR CHOUDHARY²

¹ Lecturer in Microbiology Department, Department of Biotechnology, JRU, Ranchi, Jharkhand, India,

²Department of Botany, Ranchi University, Ranchi, Jharkhand, India

ABSTRACT

Genome analysis and annotation using genome from the libraries. All of these crops are members of the Solanaceae or nightshade family, one of the world's most important vegetable plant families in terms of both economic value and production volume. The most widely programs for global multiple sequence alignment are from the Clustal series of programs. The tomato genomic resources database is an online and interactive relational database, developed using open sources software. If two sequences is an alignment share a common ancestors, mismatches can be interpreted as a point mutation. Local alignments are more useful for dissimilar sequence motif within their larger sequence. BLAST has a bio-informatics algorithm to align sequence as if they were found in the database search. When an expectation value is increase of default value. The (TGRD), the tomato genomic resources database is an online and interactive relational database, developed using open sources software. Tomato is the model for a number of biological studies important to agriculture, such as fruit development and ripening, disease resistance and biochemical pathway of important nutrient. The coding sequence is the actual region of DNA that is translated to form proteins. While the ORF may contain introns as well, the CDS refers to those nucleotides (exons) that can be divided into cordons. In sequence, alignment is a way of arranging the sequence of DNA, RNA, or protein to identify the functional, structural or evolutionary relationship between the sequence. If two sequences in an alignment share a common ancestors, mismatches can be interpreted as a point mutation. FASTA format is a text based format for representing either nucleotide sequence or peptide sequence, in which nucleotide or amino acid is represented using single letter code. Sequence homology is a general term that indicates evolutionary relatedness among sequence. Sequence similarity is a substitution with similar chemical properties. The Clustal Colored alignment also has the colour option in the output results. In addition to maintaining the gene bank nucleic acid, sequence database, NCBI provides data retrieval system and computational resources for the analysis of gene bank data and variety of other biological data made available through NCBI.

KEYWORDS: Bioinformatics, Genome Database, TGRD (Tomato Genomic Resources Database), Sequence Analysis, Data Compiled, Sequence Alignment & Solanum Lycopersicum

Received: Dec 19, 2017; **Accepted:** Jan 09, 2017; **Published:** Jan 12, 2018; **Paper Id.:** IJBTRFEB20181

INTRODUCTION

Tomato has been used extensively for genetic studies, because of several regions such as, its diploid genome, short generation time, availability of homozygous in breed lines, efficient transformation technology. An alignment will display by default, the following symbol denoting the degree of conservation observed in each column. Each bar drawn below the map represents a protein (or protein fragment) that matches the query sequence

the most similar hits are shown at the top in red, pink, green, blue and black bars follow, representing protein in decreasing order of similarity. The aim of the tomato genome sequencing is to reveal and explore the genetic variation availability in tomato. The tomato has been selected as a target crop, because it is economically one of the most important crop species. The program can run online from the EBI web server. The sources code, executable for Windows, Linux are available from the EBI. The coastal series of programs are widely used in molecular biology for the multiple alignment of both nucleic acid and protein sequence and for preparing the phylogenetic trees. Taylor Willie, Higgins Des 2000. Bioinformatics new features include NEXUS and FASTA format output, printing range numbers and fater tree calculations. Clustal was originally developed to run on local computers; numerous web servers have been setup, notably at the EBI(European Bioinformatics Institute). Tomato has been used extensively for genetic studies, because of several reason such as its diploid genome, short generation time, efficient transformation technology. The source code, executable for Windows, Linux are available from the EBI. The celestial series program is widely used in molecular biology for multiple alignment of both nucleic acid and protein sequence and for preparing the data can be submitted and accessed via the World Wide Web (Mount. David 2004). The tomato genome resources database is a interactive relational database developed using open sources bioinformatics software. Sequence analysis created a huge impact *solanaceae* research. Using pairwise alignment to find the best matching in query sequences. FASTA format is a text based format for representing either nucleotide sequence or protein sequence (Higgins, D. G.; Sharp, P. M. (1989). The format originates from the FASTA software package. For DNA and Protein, it is represented in one letter IUPAC nucleotide code and amino acid code. It finds the local similarity between the sequence and calculates the statistical significance of matches. Mismatch would be connected to a space. ClustalW is a widely used multiple sequence alignment tool in computer program (Higgins, D. G.; Bleasby, A. J.; Fuchs, R. (1992).. An alignment will display by default, the following symbols denoting the degree of conservation observed in each column. FASTA produces local alignment score and the comparison of the query sequence to every sequence in the database. Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. (1997). Sequence alignment or Sequence comparisons lie at the heart of the Bioinformatics, which describes the way of arrangement DNA and RNA to identify the regions of similarity among them. Several biological organizations have implemented bioinformatics tools on the website. NCBI computational biology branch focus on theoretical, analytical and applied computational approaches to a broad range of fundamental problem in molecular biology.

MATERIALS AND METHODS

The National Center of Biotechnology Information (NCBI) is a multidisciplinary research group that serves as resource for molecular biology information developing new method to deal with the volume and complexity of data searching and methods that can analyze the structure and function of macromolecules creating computerized systems for storing and analyzing data. A FASTA sequence alignment software package is used for functional and evolutionary relationship between sequences. The primary database retrieval system at NCBI, which links together several databases including gene bank. FASTA is available as a part of a package of program that constructs a local and global sequence alignment.

Database and Corresponding Web services

Database name	Web services type: URL
NCBI	E—Utility web services (http://www.ncbi.nlm.nih.gov)
FASTA	www.ebi.ac.uk/tools
BLAST	https://BLAST.NCBI.nlm.nih.gov/BLAST.cgi
Clustal omega	http://www.ebi.ac.uk/Tools/msa/ClustalW2/
EMBL/EBI	EMBL-EBI web services (http://www.ebi.ac.uk/tools/)
Uniprot KB	Programmatic access services (http://www.uniprot.org)
EBI/ftp site:	ftp://ftp.ebi.ac.uk/pub/software/ClustalW2/

RESULTS AND DISCUSSIONS

Alignment can be done by two methods slow/accurate, fast/appropriate. The FASTA file format is now largely used by other sequence database search tools, which take input as nucleotide or protein sequence program. (ClustalW) clustalW is a widely used multiple sequence alignment that manipulates existing alignment, profile analysis and create phylogenetic tree. Alignment can be done by two methods slow/accurate, fast/appropriate. Clustal Omega is a new multiple sequence alignment program that high profile technique to generate alignment between two or more sequences. Local sequence alignment program report alignment scores for the alignment constructed, and related(homologous) sequences will have higher alignment scores. The statistical significance of an alignment score is more widely accepted as a metric to comment on the relatedness of the two sequences being aligned. The ClustalW and clustalx multiple sequence alignment program have been completely rewritten in C++ (Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003). This will facilitate the further development of the alignment algorithms in the future and has proper portion of the program to the latest version of Linux, window operating system.(Availability-the program can be run online from the EBI web server. <http://www.ebi.ac.uk/tools/clustalW2>.Theclustal series of program are widely used in molecular biology for the multiple alignment of both nucleic acid and protein sequence and preparing phylogenetic trees clustal was originally developed to run on the local computer, numerous web server has been setup, notably at the EBI (European bioinformatics institute).

Pair wise Statistical Significant Estimation

Consider the pairwise statistical significance described in obtainable by the following function: where Sequence1 and Sequence 2, and SC is the scoring scheme (substitution matrix, gap penalties), and N is the number of shuffles.

>gi|1050193310|ref|NM_001329952.1| *Solanum lycopersicum* plastid-specific 50S ribosomal protein 5-like (LOC101252938), mRNA

```
TCCAGAAAAACCAAACTCAAACTGGAGAGATGGCTCTCCTTATCTTCACTGCAACAACTCCCTCTGTTC
TCCTCTCATCTCAATCTTCTCTACTTCACAAGCTTCTGCATTCCCTGCTTCGTTATCCTCCAGGTTCTG
CAACAATCACTTTACCCTGACACCTAAGTCTTATGCCAATGGTTATATTCAGCACCTTTTATCTTCCAA
AGGAGAGGTGCATTGATTGCTACAGCGGCTGCAGACATTGATAGTGTGCGGTTTCAGATAATCCTGAGCCTT
CACCAGAAAAAGGAGGAAAGTGTGCTGTGAGAATCTCCCTCTGGAGTCTAAGCTTCAAGAGAAGCT
TGAACAGAGATGAAGATGAAATTGGCAAAAAGCTTAGACTACGGAGGAAGAGACTCGTTAGGAAGCGC
CACCTAAGGAAGAAAGGACGATGGCCACCTTCAAAGATGAAGAAGAACAAGAATGTCTAACTTAACCTG
AAATGCCTTGCAAGTGTCTCGTTTTTCTCGTAGTCTTTATAATATCGAAATACTGTAATCTCTGAGATC
ATTTCTTCAACCTGTACCTGATACCTTATGAAATTGATTAGATTTTTTCCCGAAAAA
```

```
└─ sequence1TCCAGAAAAACCAAACTCAAA
└─ sequence2TCCAATTTCTCCACCTATAG
```

CONCLUSIONS

Analysis of the sequence has been developed with the objective of providing a single platform for customizable data from some of the major biological databases. An application of pairwise statistical significance to empirically determine the effective gap opening penalties and to generate alignments between a nucleotide and protein sequence referred to as query describe the overall quality of an alignment. An alignment will display by default the following symbol denoting the degree of observation can be seen through clado branch or phylo branch.

REFERENCES

1. Altschuhlen; Gish, Warren; Miller, Webb; Myers, Eugene; Lipman, David (1990). "Basic local alignment search tool". *Jol, StepuRNAI of Molecular Biology* 215 (3): 403–410
2. Andreas D. Baxevanis, B. F. FracisQuellette, "A practical guide to the analysis of Gene and protein. 3RD Edition october 2004. Published by Wiley, John and Sons.
3. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. *GenBank: Update. Nucleic Acids Research*, 2004, vol 32, Database Issue: D23-D26.
4. "ClustalW/ ClustalX: Multiple Sequence Alignment". Retrieved 1 October 2013.
5. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003). "Multiple sequence alignment with the Clustal series of programs". *Nucleic Acids Res* 31 (13): 3497–3500. doi:10.1093/nar/gkg500. PMC 168907. PMID 12824352.
6. Giovannoni, J. 2001, *Molecular biology of fruit maturation and ripening*, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 52, 725-749.
7. Higgins, D. G.; Sharp, P. M. (1989). "Fast and sensitive multiple sequence alignments on a microcomputer". *Computer Applications in the Biosciences (CABIOS)* 5 (2): 151–153. Doi: 10.1093/bioinformatics/5.2.151. PMID 2720464.
8. Higgins, D. G.; Bleasby, A. J.; Fuchs, R. (1992). "CLUSTAL V: Improved software for multiple sequence alignment". *Computer Applications in the Biosciences (CABIOS)* 8 (2): 189–191. Doi:10.1093/bioinformatics/8.2.189. PMID 1591615.
9. Higgins DG, Thompson JD, Gibson TJ. (1996). *Using CLUSTAL for multiple sequence alignments. Methods Enzymol.*, 266, 383-402.
10. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. (1998). *Multiple sequence alignment with Clustal X. Trends Biochem Sci.*, 23, 403-405.

11. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). *Clustal W and Clustal X version 2.0*. *Bioinformatics*, 23, 2947-2948.
12. Madden T. (2002). *The NCBI handbook*, 2nd edition, Chapter 16, *The BLAST Sequence Analysis Too*.
13. Mount. David 2004, *Bio-informatics:-sequence & Genome Analysis*”, published by Cold spring Harbour laboratory press.
14. NCBI Resource Coordinators (2012). "Database resources of the National Center for Biotechnology Information". *Nucleic Acids Research* 41 (Database issue): D8–D20.
15. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. (1994). "CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". *Nucleic Acids Research* 22(22): 4673–4680. doi:10.1093/nar/22.22.4673. PMC 308517. PMID 7984417.
16. Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. (1997). "The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools". *Nucleic Acids Research* 25 (24): 4876–4882. Doi: 10.1093/nar/25.24.4876. PMC 147148. PMID 9396791.
17. Taylor Willie, Higgins Des 2000, *Bio-informatics: Sequence structure and database practical approach “*, 1st Edition October 2000, Published by Oxford university press.
18. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.
19. Uma Kumari and Ashok Kumar Choudhary,” (May-June 2016) *Computational Analysis of sequences to determine expectation value commonly used in Bioinformatics Database.(IOSR) Vlo-18, Issue 3, Ver. 1, pp20-22*
20. Uma Kumari and Ashok Kumar Choudhary, **June 2016**, *Genome sequence analysis of Solanumlycopersicumby applying sequence alignment method to determine the statistical significance of an alignment(IJBTR)*”vol-6, Issue 2016, p9-12
21. Uma Kumari and Ashok Kumar Choudhary, **Dec 2016** *Genome sequence analysis of Solanumlycopersicum showing the phylogenetic relationship based on multiple sequence alignment and conserved domain proteins.(IJBR)*, Vol-7, Issue-4, 2016, pp2012-2014
22. Uma Kumari and Ashok Kumar Choudhary,(Nov-dec2016) “*Genome sequence analysis to evaluate the performance of pairwise statistical significance of Solanumlycopersicum (IOSR)volume 2, Issue 7, pp46-48, IOSR JouRNAl.*
23. Uma Kumari and Ashok Kumar Choudhary,”(13thnov2017) *Genome Sequence Analysis to Evaluate the Query Cover and E-Value Providing Genetic Statistical Significance in LycopersiconEsculentum (IJSET)*, Volume 5, Issue 5.
24. Uma Kumari and Ashok Kumar Choudhary (**June2017**) “*Genome Sequence Analysis of Solanumlycopersicum Showing the site of Restriction Enzymes of RNA Sequence by Applying NEB Cutter As Bioinformatics Tool*” (IJBTR); (VOL - 7, Issue – 3, pp-37-40

