

**Research Article**

**Genome sequence analysis of *solanum lycopersicum* showing the phylogenetic relationship based on multiple sequence alignment and conserved domain proteins.**

**Uma kumari<sup>1</sup> and Ashok kumar choudhary<sup>2</sup>**

<sup>1</sup>Department of Biotechnology, Jharkhand rai university,  
Ranchi, Jharkhand, India

<sup>2</sup>Department of Botany, Ranchi University,  
Ranchi, Jharkhand, India

**ABSTRACT**

Phylogenetics analysis has become essential in researching the evolutionary relationship between sequence alignment and conserved domain protein evolutionary relationship are identified from open reading frame rather than from complete sequences. A reading frame is a set of consecutive, nucleotide, non overlapping triplets of three consecutive nucleotide. The national center for biotechnology information NCBI provide many tools for comparing database-stored nucleotide or protein sequence, including the well-known BLAST algorithms. NCBI also provide database such as the genebank and SNP, in which biologist can seek homology or specific function. The EMBL (European molecular Biology laboratory), provide freely available molecular data and online bioinformatics tools of scientific community. The proposed platform correctly identifies the evolutionary relationship between the well studied references species, showing phylogenetics relationship based on multiple sequence alignment and conserved protein domain.

**Keywords:-**Protein evolution, Tomato genomic resources database, Phylogeny genome evolution, Orf reading frame, Diversity analysis.

**INTRODUCTION:-**

Several biological organization have implemented bioinformatics tools on website. The national center for biotechnology information (NCBI) Lindberg DA (2000) provide many tools for comparing database-stored nucleotide or protein sequence, including the well known blast algorithms understanding the evolutionary relationship between groups of organism has become increasingly reliant on phylogenetic analysis. Reconstructing ancestral sequence can reveal the timing and directionality of mutation. The orf indicate candidate protein coding region in a DNA sequence. cloud computing is the practice of using a network of remote servers

hosted on the internet to store, manage and process data, rather than a local server or a personal computer. cloud computing services categorise in three broad services like IaaS (infrastructure as a services), PaaS (platform as a services), SaaS (software as a services). Develop a high -availability, large scale Orf phylogenetic analysis. Cloud service based on virtualization technology. Diversity demonstrates species variance at a specific position in the protein sequence are very similar at that position. in contrast, a high diversity value denotes low similarity at the position. A frame with high variance also indicate that this frame mutate

easily. These Orfs are encoded within each of the six translation frame (3 in the forward direction and 3 in reverse direction), so that identify the translation frame yielding the longest protein coding sequence. Several biological organization have implemented bioinformatics tools on the website.

## 2. MATERIAL and Method

Orf phylogenetic analysis services combining framework, virtualization technology, phylogenetic tree tools, and diversity analysis. A fasta sequence alignment software package used to functional and evolutionary relationship between sequences.

### Solanum lycopersicum non-specific lipid-transfer protein 2 (LE16), mRNA

NCBI Reference Sequence: NM\_001247024.2

>NM\_001247024.2 Solanum lycopersicum non-specific lipid-transfer protein 2 (LE16), mRNA

```
TTGTACTACTATACTATAATCTTTACTCAATTTTTTT
ATAAGAATGGAAATGTTTGGAAAAATTGCATG
CTTTGTGGTTTTTTCATGGTGGTGGTTGCACCCCAT
GCAGAGTCACTGAGCTGCGGCGAGGTTACGTCT
GGCTTGGCTCCTTGTCTCCCTTATCTCGAGGGTTCGCG
GCCCTCTAGGAGGGTGTGTGGTGGTGTAAAGG
GTCTATTGGGCGCAGCCAAGACCCAGAAAGACCGG
AAGACAGCGTGCACCTGCCTTAAATCGGCAGCTAA
TTCTATTAAGGGCATTGATACAGGAAAAGCCGCTGG
GCTCCCTGGAGTTTGTGGAGTCAACATTCCTTAC
AAGATCAGCCCTTCCACTGATTGCTCAACGGTCCAG
TAAGGTTGATGAAAGCTAAATCTGGCTTGTGGAG
ATGAAGAATAAGATGGATTATCTATCGAGTTTGAT
CCATCATTATATATATATATATATATATATATAT
ATGTTGTCCTTTCTTTTGTAAATTTGGAGTACTTAT
ATATTGTCGAGTCTTGTAAATGAACATTGGTGGT
TGCTTATGTTACAACCAATCTTCATATAGTAATACA
TGTTATTTCTGCTTT
```

### 2.1 Database and Corresponding Web services

| Database name | Web services type:   |
|---------------|--|
| NCBI          | E-Utility web services ( <a href="http://www.ncbi.nlm.nih.gov">http://www.ncbi.nlm.nih.gov</a> ) |
| BLAST         | <a href="http://www.ebi.ac.uk/tools/sss/ncbiblast">www.ebi.ac.uk/tools/sss/ncbiblast</a>         |
| FASTA         | <a href="http://www.ebi.ac.uk/tools">www.ebi.ac.uk/tools</a>                                     |
| EMBL/EBI      | EMBL-EBI web services ( <a href="http://www.ebi.ac.uk/tools/">http://www.ebi.ac.uk/tools/</a> )  |

Uniprot KB Programmatic access services (<http://www.uniprot.org>)

ORF <https://www.ncbi.nlm.nih.gov/orffinder>

## 3. RESULTS AND DISCUSSION

The Ncbi house a series of databases relevant to bioinformatics tools and services. Epigenomic database of the ncbi (National center of biotechnology information) Sayers E. (2013) at NIH (National Institute of Health) means to collect the maps of epigenetics modification and the occurrence across the human genome. List of accession number may be provided in an interactive mode of a uploading a text file. Searching and browsing the database and generating curated dataset is an essential for processing analysis and interpretation. Ncbi, Fasta, Blast, Orf finder, this need for searching subset of sequence/data from new widely used database providing e-value in blast, occurring by chance with the observed the score/high score in E-value. Detecting open reading frame is the functional Orf are extracted from sequence. The Orf finder locates all open reading frame of a specified minimum size in a sequence. Diversity analysis among open reading frames the diversity usually depicts the number of different identities. All input Rna sequence are translated to protein sequences based on genetic code. The genetic code is the set of rules by which Rna sequence information is translated in to protein. Constructing phylogenetic tree of solanum lycopersicum based on open reading frame a evolutionary tree is a branching between (tree) diagram showing the evolutionary relationship between biological species. The phylogenetic tree is computed using clustalW. This algorithm builds two phylogenetic tree; one based on full sequence and other for orf only.

## CONCLUSION

Cloud computing online delivering of computer resources as software. Sequence database showing phylogenetics relationship based on multiple sequence alignment and conserved protein

domains. A three best matches in the sequence database getting together by using of smart blast analysis tools. E-value is increased from default value, larger list with more scoring hits can be reported based on quality of alignment (the score) and size of the database by applying the sequence alignment method and bioinformatics tools.

### ACKNOWLEDGEMENT

We extended our sincere thanks to Dr. Savita Sanger "vice chancellor" of Jharkhand rai university; Ranchi, India for kindly providing me the platform to carry out the research.

### REFERENCES

- [1]. A.M. Maxam, W. Gilbert A new method for sequencing DNA Proc. Natl. Acad. Sci. U. S. A., 74 (1977), pp. 560–564
- [2]. Canese K. Jentsch J. & Carol Myers J. (2002). The NCBI handbook 2nd edition, Chapter 2, PubMed: The Bibliographic Database.
- [3]. Deonier, Richard; Simon Tavaré; Michael Waterman (2005). Computational Genome Analysis: an introduction. Springer-Verlag. p. 25. ISBN 0-387-98785-1.
- [4]. Lipman, DJ; Pearson, WR (1985). "Rapid and sensitive protein similarity searches". Science. 227 (4693): 1435–41. doi:10.1126/science.2983426. PMID 2983426.
- [5]. L.M. Smith, et al. Fluorescence detection in automated DNA sequence analysis Nature, 321 (1986), pp. 674–679
- [6]. Lindberg DA (2000). "Internet access to the National Library of Medicine". Eff Clin Pract. 3: 256–60. PMID 11185333.
- [7]. M.L. Metzker Emerging technologies in DNA sequencing Genome Res., 15 (2005), pp. 1767–1776
- [8]. Mount, D. W. (2004). Bioinformatics: Sequence and Genome Analysis (2nd ed.). Cold Spring Harbor Press. ISBN 978-0-87969-712-9.
9. "Open reading frame". U.S. National Library of Medicine. 2015-10-19. Retrieved 2015-10-22.
- [10]. Pearson, WR; Lipman, DJ (1988). "Improved tools for biological sequence comparison". Proceedings of the National Academy of Sciences of the United States of America. 85 (8): 2444–8. doi:10.1073/pnas.85.8.2444. PMC 2800136. PMID 3162770.
- [11]. Sayers E. (2013). The NCBI handbook, 2nd edition, NCBI Protein Resources
- [12]. Slonczewski, Joan; John Watkins Foster (2009). Microbiology: An Evolving Science. New York: W.W. Norton & Co. ISBN 978-0-393-97857-5. OCLC 185042615
- [13]. Scott Jw, Harbaugh Bk:Micro Tom: A miniature dwarf tomato. Florid Agar Experiment 1989.
- [14]. Taylor Willie, Higgins Des 2000, Bioinformatics :Sequence structure and database practical approach —,1st Edition October 2000 ,Published by Oxford university press.
- [15]. Whitworth, W.A. (1901) Choice and Chance with One Thousand Exercises. Fifth edition. Deighton Bell, Cambridge. [Reprinted by Hafner Publishing Co., New York, 1959.]
- [16]. Zanet, J.; Benrabah, E.; Li, T.; Pelissier-Monier, A.; Chanut-Delalande, H.; Ronsin, B.; Bellen, H. J.; Payre, F.; Plaza, S. (2015). "Pri sORF peptides induce selective proteasome-mediated protein processing". Science. 349 (6254): 1356–1358. doi:10.1126/science.aac5677. ISSN 0036-8075.
- [18]. Zhao, K.; Chu, X. (2014). "G-BLASTN: accelerating nucleotide alignment by graphics processors". Bioinformatics 30 (10):138491. doi:10.1093/bioinformatics/btu047. PMID 24463183.