

Genome sequencing of oryza sativa represent a large range of intraspecific variability

Uma Kumari

Assistant Professor, Department of Biotechnology, Marwari College, Ranchi, Ranchi University, India

ABSTRACT

With the advancement of technology next generation sequencing, oryza sativa reference genome and genomic variation information has become available. Genomic information accumulating in oryza sativa can serve as excellent genotype. Phenotype information that analyzing rice functional and structural large range of intraspecific variability. SmartBLAST accept the FASTA sequence to identify the matches in the sequence search and the SmartBLAST display available information about the query. By applying global and local sequence alignment align the entire sequence (end to end alignment). It usually done in Global alignment by comparing homologous gene like comparing two genes with same function. Tools smart BLAST accepts either a FASTA sequence or a protein accession as input. COBALT is a constant based alignment tool that implements a framework for multiple alignment of protein sequence. In genome sequencing the multiple sequence compare all six sequence to each other and produce optimal alignment between all sequence with ideal for presentation like a phylogenetic tree, that show the relationship among a set of sequence.

KEYWORDS:-Oryza, SmartBLAST, COBALT, Ribosomal protein sequence, Multiple sequence alignment.

Date of Submission: 20-12-2020

Date of Acceptance: 03-01-2021

I. Introduction

Chloroplast genome sequence of oryza sativa using reference gene performing BLAST search gene was responsible for the yellow, green phenotype chlorophyll is one of the most abundant biological molecules on earth and play an essential role in photosynthesis. The yellow green leaf phenotype in the whole plant growth duration with significantly reduce the level of chlorophyll. Rice is the most important crops in the world of the staple food for over half of population. In phylogenetic study of oryza by applying software (Smart BLAST) in chloroplast genome sequence region. The wild species contain valuable genetic diversity that has continued immediately to rice crop improvement. NCBI create public database, develop software tools for analyzing genome database, containing million of nucleotide sequence where records the molecular modeling database (MMDB) consisting of three dimensional structure as well as tools for their visualization and comparative analysis of 3D structure by applying bioinformatics software. BLAST has found similar sequence database to the query in database it is helpful to give idea whether the alignment is good and BLAST statistical theory to produce a bit score and E value of each alignment pair. Using bioinformatics tool finding the Similarity of gene expression provides the important clues for understanding the biological functions of genes, biological processes and metabolic pathways related to genes (Chen et al. 2013, Huang *et al.* 2013). In sequence alignment where the bit score gives an indication about how good the alignment, the higher the score indicate better alignment. In sequence alignment that has an E-value of 0.05 means that this similarity has 5 in 100. Multiple sequence alignment (MSA) play an important role in evolutionary analysis of biological sequence of oryza. In analysis almost all the pair in the input sequence can be aligned locally and globally. Intraspecific variability of sequencing of oryza, Chloroplast genome variation has been detected.

II. Material And Methods

Chloroplast genome sequencing annotation was performed by (Hang and Cronk 2005) using Oryza sativa reference sequence from genebank library of NCBI. A chloroplast genome map was drawn using genome software (Conant and walfe 2008). The sequence ribosomal protein S15 (Chloroplast). Rice research has been access to the high quality reference genome sequence generated in 2005 by the International Rice Genome sequencing Project (IRGSP). SmartBLAST uses combination of BLAST searches for the multiple sequence Alignment to produce its results. Improvement the annotation of sequence through the combination of computational advancement. Smart Blast searches a protein query sequences.

2.1 Database and Corresponding web services

Database name	Web services type: URL
NCBI	http://www.ncbi.nlm.nih.gov
FASTA	www.ebi.ac.uk/tools
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/
Smart BLAST	blast.ncbi.nlm.nih.gov/

2.2 Sequence using in Smart BLAST

Sequence 1(Smart BLAST)

ribosomal protein S15 (chloroplast) [Oryza sativa]

MKKKGGGRKIFGFMVKEEKEENWGSVEFQVFSFTNKIRRLASHLELHKKDFSSERGLRLLGKRQRLLA
YLAKKNRVRYKKLISQLDIRER

Sequence 2 (MULTIPLE SEQUENCE ALIGNMENT)

ribosomal protein S7 (chloroplast) [Oryza sativa]

>YP_009305373.1 ribosomal protein S7 (chloroplast) [Oryza sativa]
MSRRGTAEKRTAKSDPIFRNRLVNMVVNRIMKDGKKSLEYQILYRAVKKIQKTETNPLLVLRQAIIR
VTPNIGVKTRRNKKGSTRKVPKIEIGSKQGRALAIRWLLLEASQKRPGRNMAFKLSSELVDAAKGGGGAIR
KKEATHRMAEANRALAHFR

III. Result And Discussion

Smart BLAST finding sequence in a given database, second sequence run in the BLAST the graphical result segment of the coverage 100%, mismatches 0 and gap 0. The sequence bar is colored by the sources sequence and to generate that base in genome sequencing. Protein sequence database using the optimized version of BLAST. Second ,SmartBLAST perform a multiple sequence alignment on six sequences using the COBALT multiple sequence alignment program identifies subject sequence similar to the query and SmartBLAST uses the multiple sequence Alignment for the phylogenetic tree and the Graphical overview.

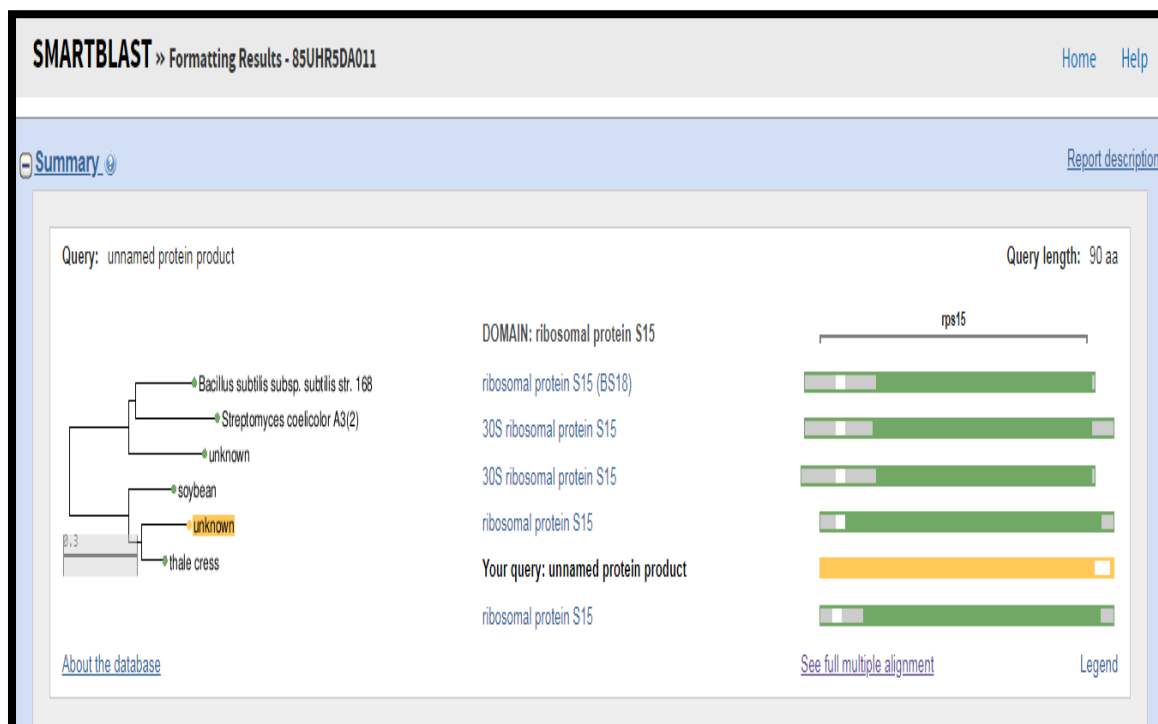


Figure 1 Shows the phylogenetic relationships based on multiple sequence alignment .The query sequence of this search highlight in yellow, Green indicate reference database.

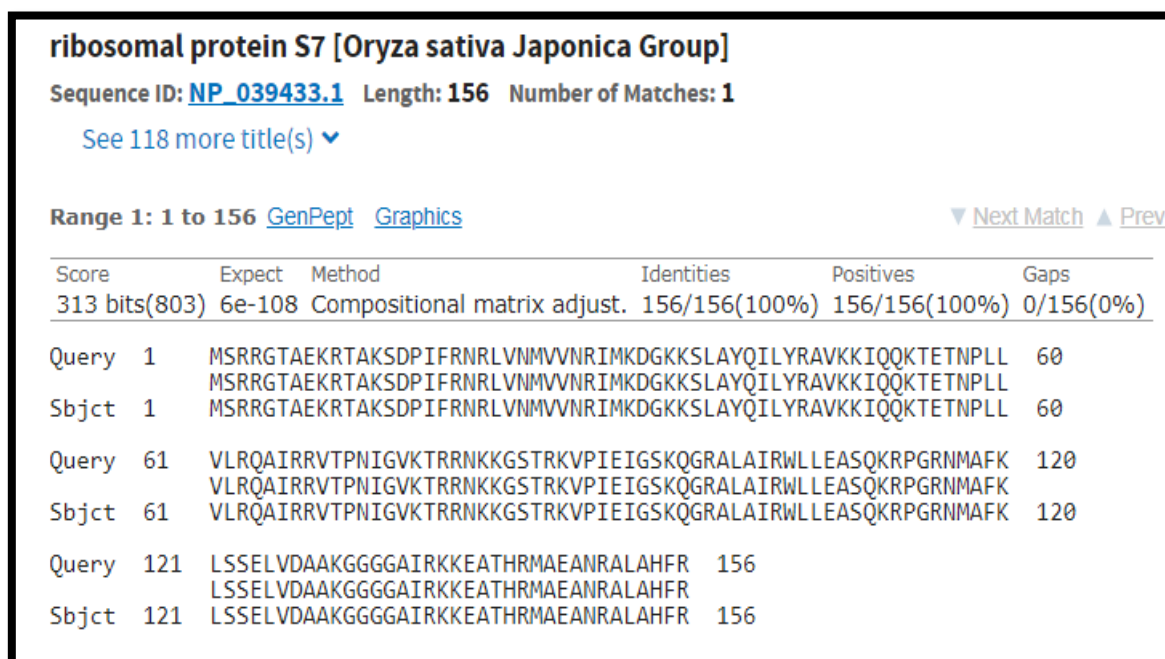


Figure 2 Smart BLAST Alignments

IV. Conclusion

In the BLAST the graphical result segment based phylogenetics analysis and result as a sorted list of matches between query and database sequence. Where the align sequence shows the zero gap the sequence aligned globally and locally the fast progress of sequence technology it is increasing tractable to generate whole genome sequencing data. As the function and associated traits of more gene in the rice genome are determined the global rice functional genomics. A nucleotide sequence database search tools that shows a computational result to the speed improvement compared to mega BLAST.

Reference

- [1]. Altschul, S.F., Madden T.L., Schaffer, A.A., Zhang J., Zhang Z., Miller W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402
- [2]. Daniels N.M., Gallant A., Peng J., Cowen L.J., Baym M., Berger B. Compressive genomics for protein databases. *Bioinformatics.* 2013; 29:i283–i290.
- [3]. Rost B., Yachdav G., Liu J. The predict protein server. *Nucleic Acids Res.* 2004; 32:W321–W326.
- [4]. Loewenstein Y., Raimondo D., Redfern O.C., Watson J., Frishman D., Linial M., Orengo C., Thornton J., Tramontano A. Protein function annotation by homology-based inference. *Genome Biol.* 2009; 10:207.
- [5]. Tatusov R.L., Galperin M.Y., Natale D.A., Koonin E.V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000; 28:33–36.
- [6]. Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J., Willey D.L., et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001; 409:928–933.
- [7]. Berkman P.J., Lai K., Lorenc M.T., Edwards D. (2012) Next-generation sequencing applications for wheat crop improvement. *Amer. J. Bot.* 99: 365–371.
- [8]. Chen J., Huang Q., Gao D., Wang J., Lang Y., Liu T., et al. (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4: 159
- [9]. Fujita M., Horiuchi Y., Ueda Y., Mizuta Y., Kubo T., Yano K., et al. (2010) Rice expression atlas in reproductive development. *Plant Cell Physiol.* 51: 2060–2081.
- [10]. Hamada K., Hongo K., Suwabe K., Shimizu A., Nagayama T., Abe R., et al. (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52: 220–229.
- [11]. Helmy M., Tomita M., Ishihama Y. (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. *BMC Plant Biol.* 11: 63.
- [12]. Huang X., Kurata N., Wei X., Wang Z.X., Wang A., Zhao Q., et al. (2012) a map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501
- [13]. Huang X., Wei X., Sang T., Zhao Q., Feng Q., Zhao Y., et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42: 961–967.
- [14]. International Rice Genome Sequencing Project. (2005) the map-based sequence of the rice genome. *Nature* 436: 793–800.
- [15]. Itoh T., Tanaka T., Barrero R.A., Yamasaki C., Fujii Y., Hilton P.B., et al. (2007) Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.* 17: 175–183.
- [16]. Kawahara Y., de la Bastide M., Hamilton J.P., Kanamori H., McCombie W.R., Ouyang S., et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* 6:4
- [17]. Komatsu S. (2005) Rice proteome database: a step toward functional analysis of the rice genome. *Plant Mol. Biol.* 59: 179–190.
- [18]. Kurata N., Yamazaki Y. (2006) Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.* 140: 12–17
- [19]. Li H., Durbin R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- [20]. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079
- [21]. Miyao A., Iwasaki Y., Kitano H., Itoh J., Maekawa M., Murata K., et al. (2007) A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol. Biol.* 63: 625–635
- [22]. Nagamura Y., Antonio B.A., Sato Y., Miyao A., Namiki N., Yonemaru J., et al. (2011) Rice TOGO Browser: a platform to retrieve integrated information on rice functional and applied genomics. *Plant Cell Physiol.* 52: 230–237
- [23]. Ohyanagi H., Takano T., Terashima S., Kobayashi M., Kanno M., Morimoto K., et al. (2015) Plant Omics Data Center: an integrated web repository for interspecies gene expression networks with NLP-based curation. *Plant Cell Physiol.* 56: e9.
- [24]. Ohyanagi H., Tanaka T., Sakai H., Shigemoto Y., Yamaguchi K., Habara T., et al. (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. japonica genome information. *Nucleic Acids Res.* 34: D741–D744.
- [25]. Uyang S., Zhu W., Hamilton J., Lin H., Campbell M., Childs K., et al. (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35: D883–D887.
- [26]. [Sakai H., Kanamori H., Arai-Kichise Y., Shibata-Hatta M., Ebana K., Oono Y., et al. (2014) Construction of pseudomolecule sequences of the aus rice cultivar Kasalath for comparative genomics of Asian cultivated rice. *DNA Res.* 21: 397–405.
- [27]. Sakai H., Lee S.S., Tanaka T., Numa H., Kim J., Kawahara Y., et al. (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54: e6.
- [28]. Sato Y., Takehisa H., Kamatsuki K., Minami H., Namiki N., Ikawa H., et al. (2013) RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res.* 41: D1206–D1213.
- [29]. Shenton M.R., Ohyanagi H., Wang Z.X., Toyoda A., Fujiyama A., Nagata T., et al. (2015) Rapid turnover of antimicrobial-type cysteine-rich protein genes in closely related *Oryza* genomes. *Mol. Genet. Genomics* 290: 1753–1770.
- [30]. Tanaka T., Antonio B.A., Kikuchi S., Matsumoto T., Nagamura Y., Numa H., et al. (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* 36: D1028–D1033.
- [31]. Kumari Uma 2019, "Molecular Analysis of *Oryza Sativa* Using A MEGABLAST" (TJPRC) June 2020, Vol 10, Issue 1, Pp 79-86
- [32]. Wang M., Yu Y., Haberer G., Marri P.R., Fan C., Goicoechea J.L., et al. (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46: 982–988
- [33]. Zhang Q.J., Zhu T., Xia E.H., Shi C., Liu Y.L., Zhang Y., et al. (2014) Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc. Natl. Acad. Sci. USA* 111: E4954–E4962.