Journal of Advanced Bioinformatics Applications and Research ISSN 0976-2604
Vol 3, Issue 1, 2012, pp 267-270 http://www.bipublication.com

IN SILICO GENE MAPPING AND FUNCTIONAL ANNOTATION IN SHOTGUN SEQUENCE OF LACTOBACILLUS HILGARDII ATCC 8290

Nikhat Sultana Syed Jeelani
Department of Bioinformatics, Shankarlal Khandelwal College Akola,
SGBA University Maharashtra, (India)
Corresponding email: niks.sultana@gmail.com

ABSTRACT:-
There are number of strains of genus lactobacillus, some of them are responsible for spoilage of wine, and *Lactobacillus hilgardii* is one of them, which survive under harsh wine condition and spoil the wine by producing the histamine. In our study focus was made to investigate the genome of *lactobacillus hilgardii* which was partially sequenced in the form of contigs, those are the fractioned part of the genome which were used for sequencing and when aligned together helps us to know the complete genome sequence. The genome sequence needs to be marked by the gene prediction program to locate gene, their positions, and total number of genes available in the genome drafts by sequencing. The obtained amino acid sequences were scanned, and predicted the available conserved domain in the protein sequence. The further study has been made computationally, such as the prediction of the molecular weight, iso-electric point, subcellular location, lipoprotein, tertiary structure, protease activity along with function prediction of the proteins coded in the contigs of *L. hilgardii,* these all were predicted using programs namely TMHMM, LipoP, Peptide Cutter, Compute Pi and Molecular Weight, PsortB, CDD BLAST, GeneMark.hmm (prokaryotic), PS^2 Protein structure prediction server.

Keywords*: Lactobacillus hilgardii,* contigs, gene marking, tertiary structure, colonization, conserved domain.

INTRODUCTION

Lactic acid bacteria are responsible for malolactic fermentation, an important step in winemaking [13]. However, some of them induce spoilage [10,11]. Therefore, specific detection and identification are useful for quality control. Precision and reliability of the phenotypical descriptions are not sufficient enough, and some misidentifications occur [5,6]. Molecular approaches offer new opportunities to characterize micro-organisms. Lonvaud-Funel e t al. (1990, 1991) describes the identification of lactic acid bacteria during vinification and wine storage by DNA-DNA hybridization[9]. Different strains of the genus Lactobacillus can be regularly isolated from must and wine samples. By various physiological activities, they can improve or reduce the wine quality. *Lactobacillus hilgardii* that is known to survive under harsh wine conditions is classified as a spoilage bacterium, e.g. due to the production of histamine. Many lactobacilli form an S-layer as the outermost cell wall component which has been found to facilitate the colonization of special ecological niches [2]. Water kefir is a homemade fermented beverage based on
a sucrose solution with fruit extracts. The inoculum of such fermentations consists of macroscopic granula containg lactic and acetic acid bacteria, and yeasts, which are embedded in an exopolysaccharide (EPS) matrix. A strain of *L.hilgardii* producing large amount of the granule-forming dextran were isolated. The glycosyltransferase (Gtf) commonly called glucan sucrose responsible for the production of this dextran was purified from *L.hilgardii[15].*In *Lactobacillus hilgardii*, together with the reciprocal antagonism between arginine deimination and histidine decarboxylation, offer clue keys to the understanding of the accumulation of lactate, amine, ammonia and athylcarbamate in wine,with consequent implications on different health risk controls[7].Putrescine, one of the main biogenic

amines associated to microbial food spoilage,can be formed by bacteria from arginine via ornithine decarboxylase(ODC),or from agmatine via agmatine deiminase(AgDI).putrescine production from agmatine could be linked to the aguA and ptcA genes in *Lactobacillus hilgardii* X1B, *Enterococcus faecalis* ATCC 11700, and *Bacillus cereus* ATCC 14579 [8]. Various poly phenols wine polyphenols shows the inhibitory effects on the growth of enological lactic acid bacteria such as *Lactobacillus hilgardii* and reported to inhibit the growth in its particular concentration/s[3].This study offerd a systematic steps to decipher the functionality in the draft genome of *L.hilgardii and the retrieved data for several contigs of L.hilgardii were managed in the form of database,and the aim to manage entire obtaind data using Microsoft access is to retrieve the data easily and assisted us in better retrieval,editing and management of the data of contigs in the L.hilgardii.*

MATERIALS AND METHODS

The whole genome of the *L.hilgardii* was studied by computationally by using online tools and databases. The sequences of *L.hilgardii* were obtained in the form of contigs from NCBI database . And the further study has been made by using online tools and databases. They are enlisted as follows: GeneMark.hmm (prokaryotic) is a program used to predict the sequences from contigs [12]. CDD BLAST program used to predict the conserved domain in the protein sequence .TMHMM is an online tool those predicted transmembrane helices in proteins and locates hydrophobic regions in several proteins.LipoP program predicted lipoproteins and discriminates between lipoprotein signal peptides, other signal peptides and n-terminal membrane helices [14]. Peptide Cutter predicted potential cleavage sites cleaved by proteases or chemicals in a given protein sequence .Pi and Molecular Weight calculated by tool provided by expasy server[4].PsortB predicted protein subcellular localization in an organism[16].PS^2 Protein structure prediction server used to build tertiary structures of proteins[1] .and the results were stored in Microsoft excel, and managed in Microsoft access as database **.**

RESULTS

The sequences of *L.hilgardii* ATCC8290 were retrieved in the form of whole genome shotgun sequencing project which contain sequences as contigs 3, 4, 5,6,7,8,9,11 & 12 by using NCBI database. Each contig marked with genes predicted by Gene marking (prokaryotic) program, as contig3 marked with41 genes, contig4 with 26, contig5 with 3, contig6 with 48, contig7 with 2, contig8 with 2, contig9 with 47, contig11 with 5, and contig12 with 15 genes, i.e. 189 genes were predicted. The available conserved domain of the predicted protein sequence was searched by CDD BLAST. Out of 189 proteins, 40 proteins categorized with transmembrane helices by TMHMM. Out of 189 proteins 32 proteins categorized as lipoprotein, 149 proteins as cytoplasmic protein by LipoP. Subcellular localization, Pi and Molecular Weight of each gene were predicted and results noted in Microsoft excel. Peptide Cutter predicted potential cleavage sites in each protein which has shown with the dotted line in the file with the probable cleavage site in Microsoft access as an attachment. The three dimensional structures of 103 proteins out of 189 proteins were build succsfully by using $(PS)^2$ server.

DISCUSSION

The computational study of shotgun sequence with reference to gene mapping and functional annotation of *L.hilgardii* provided us with the partial genome sequence in the form of contigs. The sequences need to be marked by the genes in order to understand the functional domain available in the organism. Our study highlighted several programs which could be implemented in the functional annotation of the uncharacterized regions of the genome. While, further purification and subcellular functionality could be linked based on the obtained data of each protein. Several proteins investigated remained uncharacterized and nominated as hypothetical proteins. Further the sequence homology approach assisted us to predict the probable function in the proteins governed by the *L.hilgardii*. This study involved the potent web

tools and Database creation tools to figure out the probable protein coding ability in the draft genome, and easy data analysis and retrieval respectively. Such an approach helped us to understand the distribution of gene sets on the genome and probable protein information coded by them to assist wet lab work in greater detail. This study helped us to know many other things about *Lactobacillus hilgardii.*

ACKNOWLEDGMENT

I thank full to Dilip G.Gore Sir, Director of Sai Bioinfosys institute to guide me for this work, and I thank full to Chandan.N.Dipke Sir, H.O.D of Bioinformatic and Dr. Jagdish Mangilalji Saboo Sir, principle of Shankarlal Khandelwal College Akola, to allow me to do this work under Dilip G.Gore Sir.

REFERENCES

[1] Chih-Chieh Chen, Jenn-Kang Hwang and Jinn-Moon Yang (2006). (PS)2: protein structure prediction server Nucleic Acids Res.

[2] Dohm N, Petri A, Schlander M, Schlott B, König H, Claus H. Molecular and biochemical properties of the S-layer protein from the wine bacterium Lactobacillus hilgardii B706. Arch Microbiol. 2011 Apr;193(4):251-61.

[3] García-Ruiz A, Moreno-Arribas MV, Martín-Álvarez PJ, Bartolomé B. Comparative study of the inhibitory effects of wine polyphenols on the growth of enological lactic acid bacteria. Int J Food Microbiol. 2011 Feb 28;145(2-3):426-31.

[4] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The ProteomicsProtocols Handbook, Humana Press (2005).

[5]Kandler, 0. & Weiss, N. (1986). Regular, non-sporing Grampositive rods. In Bergey's Manual of Systematic Bacteriology, pp. 1208-1260. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams & Wilkins.

[6]Lafon-Lafourcade, S. & Joyeux, A. (1979). Techniques simplifiees pour le denombrement et l'identification des microorganismes vivants dans les moiits et les vins. Connaiss Vigne Vin 13,295-310.

[7] Lamberti C, Purrotti M, Mazzoli R, Fattori P, Barello C, Coïsson JD, Giunta C,Pessione E. ADI pathway and histidine decarboxylation are reciprocally regulated in *Lactobacillus hilgardii* ISE 5211: proteomic evidence. Amino Acids. 2011 Jul;41(2):517-27.

[8] Landete JM, Arena ME, Pardo I, Manca de Nadra MC, Ferrer S. The role of two families of bacterial enzymes in putrescine synthesis from agmatine via agmatine deiminase. Int Microbiol. 2010 Dec;13(4):169-77. PubMed PMID: 21404211.

[9] Lonvaud-Funel, A., Fremaux, C., Biteau, N. & Joyeux, A. (1991). Speciation of lactic acid bacteria from wine by DNA hybridization with DNA probes. Food Microbiol8, 2 15-222.

[10] Lonvaud-Funel, A., Joyeux, A. & Rolland, C. (1990). Etude d'alteration des vins par les bacteries lactiques. In Actualit& Enologiques 1989, pp. 378-382. Edited by P. Ribereau-Gayon & A. Lonvaud. Paris: Bordas.

[11] Lonvaud-Funel, A. & Joyeux, A. (1982). Application de la bioluminescence au denombrement des microorganisms vivants dans les vins. *Connaiss Vigne Vin* 4, 241-256.

[12] Lukashin A. and Borodovsky M., GeneMark.hmm: new solutions for gene finding, NAR, 1998, Vol. 26, No. 4, pp. 1107-1115.

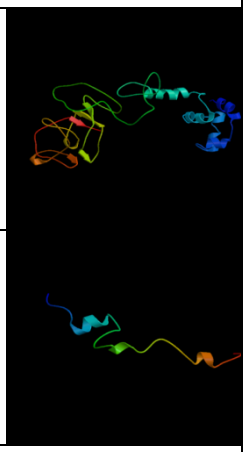
[13] Renault, P., Gaillardin, C. & Heslot, H. (1988). Role of malolactic fermentation in lactic acid bacteria. Biochimie 70, 375-379.

[14] Rahman O., Cummings S. P., Harrington D. J. and Sutcliffe I. C.. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteriaWorld Journal of Microbiology and Biotechnology 24(11):2377-2382 (2008).

[15] Waldherr FW, Doll VM, Meissner D, Vogel RF. Identification and characterization of a glucan-producing enzyme from Lactobacillus hilgardii TMW1.828 involved in granule formation of water kefir. Food Microbiol. 2010Aug;27(5):672-8.

[16] Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., Brinkman, F.S.L. (2010) PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics26(13):1608-1615.

TABLE: Functional proteomics of *L. hilgardii*

Contig	CDDBLAST	Structural homologs	PsortB	LipoP	TMHMM	pi	Mol.Wt	(PS)2:protein structure
Contig 3								
1	phage-related integrases	1z1bB	Cytoplasmic	CYT	no	9.7	44336.4	
2	Hypothetical protein	1h4pA	Cytoplasmic	CYT	no	5.4	12527.1	
Contig 4								
1	Hypothetical protein	2jesA	Cytoplasmic	CYT	no	9.8	28186.1	
2	Endo-N-acetylmuramidases (muramidases)	2j8gA	Extracellular	CYT	no	9.3	30592.8	
Contig5								
1	tryptophanyl-tRNA synthetase	1yi8A	Cytoplasmic	CYT	no	5.6	37807.9	
2	Ppx/GppA phosphatase	1t6cA	Cytoplasmic	CYT	no	9.9	34785.2	