RESEARCH ARTICLE OPEN

In-Silico Studies of Halophilic Archaeon DL31 Plasmids for Gene Annotation and Structure Prediction

Archis Panday*, Azeem uddin Siddiqui**, Swapnil Sanmukh*** And Krishna Khairnar***

*Ashok and Rita Patel Institute of Integrated Study and Research in Biotechnology and Allied Sciences, Sardar Patel University New Vallabh Vidya Nagar-388121, Anand, Gujarat.
**Indian School of Mines (ISM), Dhanbad - 826004, Jharkhand, India.
***National Environmental Engineering Research Institute (NEERI), Nagpur-440020, Maharashtra (India)

ABSTRACT
The Insilico structure prediction, gene annotation and sub-cellular location were determined for the hypothetical proteins in two plasmids from Unclassified Halophilic archaeon DL31 plasmids. The probable structures were predicted for 92 hypothetical proteins by the use of PS^2 structure prediction server. The functional annotation for 68 proteins was possible by the use of Bioinformatics web tools like CDD-Blast, Interproscan and pfam by searching the presence of conserved domains amongst the different known proteins and their families through online databases. The Sub-cellular location predictions were done for all the 92 unclassified hypothetical proteins by using CELLO v 2.5 servers. Through comparative genomics approach this study revealed various functional hypothetical proteins in two plasmids of Unclassified Halophilic archaeon DL31.

Keywords - Insilico, gene annotation, sub-cellular location; Bioinformatics web tools, conserved domains, comparative genomics.

I. INTRODUCTION

The Halobacteriaceae form a monophyletic group within the phylum Euryarchaea of the domain Archaea. The Halobacteriaceae include 26 named genera each with at least one cultured species (http://www.the-icsp.org/taxa/halobacterlist.htm) (Oren, 2008; Sorensen et al., 2005). It has long been recognized that *Archaea* originating from a variety of diverse environments are able to N-glycosylate numerous proteins (Eichler and Adams, 2005). In comparison to other groups of extremophilic microorganisms such as the thermophiles and the alkaliphiles, the halophiles of all three domains have been relatively less exploited in biotechnological processes, with some exceptions like β-carotene (Oren, 2010), secondary metabolites (Han et al., 2009, Lu et al., 2008), bioremediation, biofuel, bioplastics, hydroxyectoine (Quillaguaman et al., 2010, Van-Thuocet al., 2010). The problem of petroleum contamination can be neutralized by halophiles like *Alcanivorax dieselolei*, which are capable to grow on crude oil, diesel and pure aliphatic hydrocarbons but unable to degrade aromatic compounds. The presence of Poly hydroxy alkanoate (PHA) granules in Haloarcula marismortui (Oren et al. 1990) was first reported in 1972 (Kirk and Ginzburg 1972) followed by Haloferax, Halobiforma, and Haloquadratum,

Pfam (http://pfam.sanger.ac.uk/) (Alex et al., 2004), which shows the ability to search the defined conserved domains in the sequences and assist in the which have been found to accumulate poly (3-hydroxybutyrate) or poly (3-hydroxybutyrate-co-hydroxyvalerate) (Fernandez-Castillo et al. 1986; Hezayen et al. 2002; Burns et al. 2000).

The present paper reports the comparative genomic studies for determining the probable functional properties of some unclassified hypothetical proteins in the two plasmids of Halophilic archaeon DL31, which may prove helpful for identifying novel enzymes and protein candidates with possible applications in the near future.

II. MATERIALS AND METHODS

Sequence Retrieval

The Complete sequence of chromosome of Halophilic archaeon DL31 was retrieved from the KEGG database (http://www.genome.jp/kegg/) (Lucas, et al., unpublished).

Functional Annotations

The functional annotation for the hypothetical protein genes for were carried out by using bioinformatics web tools like CDD-Blast (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997; Schaffer et al., 2001; Aron et al., 2006), Interproscan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (Zdobnov and Rolf., 2001; Quevillon et al., 2005) and classification of proteins in appropriate family (Sanmukh et al., 2011).

Sub-cellular Localization of the hypothetical proteins

The CELLO v 2.5 is an online sub-cellular localization predictor server (http://cello.life.nctu.edu.tw/) which identifies the sub cellular location of given protein sequences of Gram positive, Gram negative and Eukaryotic cells. The identification of sub-cellular localization of the proteins is helpful in better understanding of their functional properties that in turn are helpful in proper classification of unclassified proteins in their respective families (Yu et al., 2004; Yu et al., 2006).

Protein Structure Prediction

The three dimensional structures of hypothetical proteins were generated using online PS^2 protein structure prediction server (http://ps2.life.nctu.edu.tw/). The server accepts the protein sequence in FASTA format and uses the strategies of Pair-wise and multiple alignments to generate resultant proteins 3D structures. These structures are constructed using structural positioning information of atomic coordinates in PDB format using best scored alignment data. The selection of template is based on the conserved domain and must be available for modeling purpose.

III. RESULTS AND DISCUSSION

The comparative genomic studies for characterizing 92 genes of two plasmids from Unclassified Halophilic archaeon DL31 was carried out. The functional annotation and classifications of proteins was done by using sequence similarity search with close orthologous family members available in various protein databases using the web tools. The sub-cellular localization of the proteins was also predicted for assigning proper functional character/s. The online-automated PS^2 server was used for the prediction of 3-D structure of hypothetical proteins. These study revealed structure prediction for 92 hypothetical proteins from which functions were assigned to 68 and 24 of them showed unknown functional properties, which were unavailable in the online databases. The analysis of proteins by using web tools for classification of 92 proteins into particular protein family based on conserved domain available in the sequence are represented in respective Table 1 and Table 2. The $(PS)^2$ protein structure prediction server produce the three dimensional structures for 92 proteins satisfactorily using best scored orthologous template, which are represented in the order as Template ID, Identity, Score and E-value separated by hyphen in structure column of Table 1 and Table 2.

IV. CONCLUSION

These results obtained in the in-silico studies for characterization of unknown proteins in the Halophilic archaeon DL31 plasmids, showed the ability of computational tools for annoting unclassified hypothetical proteins successfully. We were able to categorize 68 hypothetical proteins by their functions and predict 92 protein structures within the unclassified Halophilic archaeon DL31 plasmids. The function prediction web tools have shown variable results depending upon the information available in databases, when searched for the conserved domains or functional sites in the submitted protein sequences under study. The data generated in this study revealed some novel functions in the hypothetical proteins which were still unidentified. These functional proteins may prove helpful for understanding the life cycle of the Halophilic archaeon DL31 bacterium.

Table 1. Predicted Structures, annotations and sub-cellular localization of hypothetical proteins in Halophilic archaeon DL31 through comparative genomic approach.

Halophilic archaeon DL31 (Plasmid 1)

Gene ID	CDD-Blast	Interproscan	Pfam	Cello	PS2 structure
11093414	NO	NO	IncA protein; Laminin Domain II; Secretion system effector C (SseC); Transcription factor/nuclear export subunit protein 2;RNA polymerase II transcription mediator complex subunit 9; Baculovirus polyhedron envelope protein, PEP, C terminus; Calcium binding and coiled-coil domain (CALCOCO1) like; Eukaryotic glutathione synthase; Tetrahydromethanopterin S-methyltransferase, subunit G	Nuclear 3.739 *	1m1jC- 18- 36- 0.007
11093420	No	No	No	Cytoplasmic 3.675 *	1TyfKA -21- 32- 0.003
11093424	This family of uncharacterized proteins contain a conserved HXXP motif. A similar motif is seen in protein families in the His-Me finger endonuclease superfamily which suggests this family of proteins may also act as endonucleases.; Uncharacterized conserved protein [Function unknown]	No	HNH endonuclease	Cytoplasmic 3.382 *	1cdoA- 22- 33- 0.001
11093425	Divergent AAA domain	No	Divergent AAA domain.C(AAA 4)	Cytoplasmic 4.591 *	2qgpB- 17- 32- 0.004
11093426	No	No	No	Cytoplasmic 2.700 *	2hihA- 10- 31- 0.008
11093449	No	No	No	Cytoplasmic 3.663 *	1zakA- 18- 45- 3e-05
11093453	No	No	No	Cytoplasmic 4.549 *	1mc3A- 37- 31- 0.008
11093464	helix_turn_helix, Arsenical Resistance Operon Repressor;	Winged helix-turn-helix DNA-binding domain	Helix-turn-helix domain	Cytoplasmic 4.859 *	1hw1A- 14- 36- 0.005
11093469	No	No	No	InnerMembrane 4.835 *	2b5uA- 16- 50- 1e-08
11093473	Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]	No	No	Cytoplasmic 3.275 *	1o4wA 19 43 3e-05

11093495	Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).	von Willebrand factor, type A	von Willebrand factor type A domain	Cytoplasmic 2.290 *	2odpA- 14- 39- 6e-04
11093517	No	No	snRNA-activating protein of 50kDa MW C terminal, Protein of unknown function (DUF3268), F420H2 dehydrogenase subunit FpoO	Cytoplasmic 2.651 *	1sqhA- 21- 37- 0.001
11093520	No	No	No	Cytoplasmic 4.072 *	2oqrA- 25- 40- 4e-04
11093523	McrBC 5-methylcytosine restriction system component; Members of this family of bacterial proteins modify the specificity of mcrB restriction by expanding the range of modified sequences restricted.	5-methylcytosine restriction system component	McrBC 5-methylcytosine restriction system component	Cytoplasmic 4.436 *	2v62B- 34- 33- 0.002
11093525	No	No	Domain of unknown function (DUF1934)	Cytoplasmic 4.161 *	2o18C- 25- 35- 0.005
11093527	No	No	FAD binding domain, Sugar-specific transcriptional regulator TrmB	Cytoplasmic 3.984 *	1i1gA- 18- 38- 0.001
11093535	Seryl-tRNA synthetase N-terminal domain	No	Plasmid replication region DNA-binding N-term, Protein of unknown function (DUF3584), UV radiation resistance protein and autophagy-related subunit 14, Tat binding protein 1(TBP-1)-interacting protein (TBPIP), bZIP transcription factor, RNA pol II promoter Fmp27 protein domain, Protein of unknown function (DUF465).	Cytoplasmic 3.361 *	2dfsA- 13- 43- 2e-04
11093542	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Domain of unknown function DUF234, DEXX-box ATPase C-terminal	Archaeal ATPase, Archaea bacterial proteins of unknown function, MarR family	Cytoplasmic 4.839 *	2qenA- 20- 130- 5e-31
11093547	No	No	No	InnerMembrane 4.239 *	2pziA- 9- 33- 0.002
11093548	Putative lysophospholipase; This domain is found in bacteria and eukaryotes and is approximately 110 amino acids in length.	unintegrated	Alpha/beta hydrolase family, TAT (twin-arginine translocation) pathway signal sequence .	Cytoplasmic 3.529 *	1l7aA- 18- 60- 2e-09

11093552	No	No	No	InnerMembrane 4.896 *	1ciiA- 23- 33- 0.002
11093562	It contains several conserved aspartates and histidines that could be metal ligands., HerA helicase [Replication, recombination, and repair]	Domain of unknown function DUF87	Domain of unknown function DUF87, HAS barrel domain	Cytoplasmic 4.062 *	1e9rB -12- 99- 2e-23
11093573	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Winged helix-turn-helix DNA-binding domain	Sugar-specific transcriptional regulator TrmB, Polyketide synthesis cyclase	Cytoplasmic 4.359 *	2jscB- 37- 34- 0.005
11093574	Plasmid stabilisation system protein; Members of this family are involved in plasmid stabilisation.	No	No	Cytoplasmic 3.267 *	1wmiA- 19- 39- 3e-04
11093586	putative transcriptional regulator	Transcription regulator LuxR, C-terminal	Bacterial regulatory proteins, luxR family	Cytoplasmic 4.167 *	1p4wA -28- 36- 0.007
11093602	No	Integrase-like, catalytic core	No	Cytoplasmic 4.854 *	1z1bB- 12 -52- 2e-07
11093621	DnaQ-like (or DEDD) 3'-5' exonuclease domain superfamily; The DnaQ-like exonuclease superfamily is a structurally conserved group of 3'-5' exonucleases, which catalyze the excision of nucleoside monophosphates at the DNA or RNA termini in the 3'-5' direction.	Ribonuclease H-like domain	RNase_H superfamily	Cytoplasmic 3.839 *	1wn7A -17- 104- 2e-23
11093647	Plasmid pRiA4b ORF-3-like protein; Members of this family are similar to the protein product of ORF-3 found on plasmid pRiA4 in the bacterium Agrobacterium rhizogenes.	Plasmid pRiA4b, Orf3	Plasmid pRiA4b ORF-3-like protein	Cytoplasmic 3.298 *	2i1sB- 13- 76- 8e-15
11093654	No	No	Bacterial protein of unknown function (DUF883)	InnerMembrane 4.393 *	2b5uA- 21- 37- 0.008
11093662	MarR family; The Mar proteins are involved in the multiple antibiotic resistance, a non-specific resistance system.	Winged helix-turn-helix DNA-binding domain	MarR family	Cytoplasmic 3.601 *	3boqB- 22- 39- 5e-04

11093663	PemK-like protein; PemK is a growth inhibitor in E. coli known to bind to the promoter region of the Pem operon, auto-regulating synthesis. This Pfam family consists of the PemK protein in addition to ChpA, ChpB and other PemK-like proteins.	mRNA interferase PemK-like protein	PemK-like protein	Cytoplasmic 2.714 *	1ne8A- 20- 58- 6e-10
11093665	No	No	Domain of unknown function (DUF1965)	Cytoplasmic 1.399 * InnerMembrane 1.166 *	1yy9A- 18- 39- 2e-04
11093670	helix_turn_helix, Arsenical Resistance Operon Repressor	Winged helix-turn-helix DNA-binding domain	Helix-turn-helix domain, Family of unknown function (DUF716)	Cytoplasmic 4.722 *	1hw1A- 14- 36- 0.007
11093717	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Winged helix-turn-helix DNA-binding domain	Helix-turn-helix domain	Cytoplasmic 4.563 *	1mkmB- 24- 35- 0.008
11093718	PemK-like protein; PemK is a growth inhibitor in E. coli known to bind to the promoter region of the Pem operon, auto-regulating synthesis.	Plasmid maintenance toxin/Cell growth inhibitor	PemK-like protein, Pyridoxine 5'-phosphate oxidase C-terminal dimerisation region	Cytoplasmic 3.411 *	1ne8A- 26- 55- 7e-09
11093719	MarR family; The Mar proteins are involved in the multiple antibiotic resistance, a non-specific resistance system.	Winged helix-turn-helix DNA-binding domain	MarR family, Domain of unknown function (DUF4423)	Cytoplasmic 3.892 *	2fbhA- 31- 34- 0.005
11093732	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Winged helix-turn-helix DNA-binding domain	Helix-turn-helix domain	Cytoplasmic 3.451 *	2fbhA- 25- 37- 0.001
11093734	No	No	Photosystem II Pbs27, Domain of unknown function (DUF1707)	Cytoplasmic 3.505 *	1u2wB- 18- 35- 0.006
11093737	No	No	No	Cytoplasmic 4.304 *	1q16B- 19- 36- 0.002
11093742	No	No	Putative peptidoglycan binding domain	Cytoplasmic 3.494 *	2qyuA- 23- 36- 0.005
11093756	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	No	Bacterial regulatory protein, arsR family	Cytoplasmic 4.816 *	2ia0A- 20 -37- 0.004
11093764	Piwi_piwi-like_ProArk: PIWI domain, Piwi-like subfamily found in Archaea and Bacteria.	Stem cell self-renewal protein Piwi	Piwi domain	Cytoplasmic 4.648 *	1w9hA 15 138 2e-33

11093778	Uncharacterized conserved protein [Function unknown]	Protein of unknown function DUF159	Uncharacterised ACR, COG2135	Cytoplasmic 4.198 *	2f20A- 30- 152- 3e-38
11093780	Predicted permease; This family of integral membrane proteins are predicted to be permeases of unknown specificity.	Protein of unknown function DUF318, transmembrane	Predicted permease, YHS domain, YHS domain	InnerMembrane 4.768 *	2inpA- 23- 54- 8e-08
11093791	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Winged helix-turn-helix DNA-binding domain	MarR family, Golgi phosphoprotein 3 (GPP34), MarR family, Domain of Unknown Function (DUF746)	Cytoplasmic 4.787 *	1p4xA- 11- 38- 0.003
11093792	Ferric uptake regulator(Fur) and related metalloregulatory proteins; typically iron-dependent, DNA-binding repressors and activators; Ferric uptake regulator (Fur) and related metalloregulatory proteins are iron-dependent, DNA-binding repressors and activators mainly involved in iron metabolism.	Transcription regulator PadR N-terminal	Transcriptional regulator PadR-like family, Serine-threonine protein kinase 19, Hypothetical protein (DUF2513), Coxiella burnetii protein of unknown function (DUF807), YjcQ protein, Ribulose bisphosphate carboxylase, small chain, Phage X family, Protein of unknown function DUF86	Cytoplasmic 4.423 *	2fe3B- 23- 43- 2e-05
11093808	No	No	No	Cytoplasmic 4.393 *	1q16B -18- 37- 0.003
11093810	No	No	No	Cytoplasmic 2.212 * Periplasmic 1.835 *	1em8A- 27- 38- 0.002
11093825	Ferric uptake regulator(Fur) and related metalloregulatory proteins; typically iron-dependent, DNA-binding repressors and activators; Ferric uptake regulator (Fur) and related metalloregulatory proteins are iron-dependent, DNA-binding repressors and activators mainly involved in iron metabolism.	Transcription regulator PadR N-terminal	Transcriptional regulator PadR-like family, Hypothetical protein (DUF2513), Serine-threonine protein kinase 19, YjcQ protein, Phage X family, Ribulose bisphosphate carboxylase, small chain, Protein of unknown function DUF86	Cytoplasmic 4.173 *	2fe3B- 21- 43- 2e-05
11093827	No	Zinc finger, SWIM-type	SWIM zinc finger	Cytoplasmic 3.958 *	1wgpA- 39- 40- 5e-04
11093845	Cupredoxin-like domain; The cupredoxin-like fold consists of a beta-sandwich with 7 strands in 2 beta-sheets, which is arranged in a Greek-key beta-barrel.	Twin-arginine translocation pathway, signal sequence	TAT (twin-arginine translocation) pathway signal sequence,Copper binding proteins, plastocyanin/azurin family	Periplasmic 2.045 * Extracellular 1.434 *	1pmyA- 18- 40- 7e-04

11093852	CRISPR/Cas system-associated protein Cas4; CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and associated Cas proteins comprise a system for heritable host defense by prokaryotic cells against phage and other foreign DNA; Cas4 is RecB-like nuclease with three-cysteine C-terminal cluster.	unintegrated	PD-(D/E)XK nuclease superfamily	Cytoplasmic 2.337 *	1avqA- 16- 37- 0.010
11093856	No	No	No	Cytoplasmic 4.348 *	2bk7A- 16- 30- 0.010
11093859	No	Quinoprotein amine dehydrogenase, beta chain-like	No	Cytoplasmic 3.743 *	1pbyB- 10- 44- 5e-05
11093860	No	No	Protein of unknown function (DUF1703)	Cytoplasmic 3.893 *	1yy9A- 23- 31- 0.005
11093861	No	unintegrated	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	Cytoplasmic 4.829 *	1uylA- 19- 52- 1e-07
11093872	Plasmid stabilisation system protein; Members of this family are involved in plasmid stabilisation.	unintegrated	Plasmid stabilisation system protein	Cytoplasmic 2.831 *	1wmiA- 12- 48- 8e-07
11093874	No	No	No	Cytoplasmic 4.656 *	2aw5B- 29- 38- 0.001
11093878	Homeodomain-like domain.	No	Homeodomain-like domain, Bacterial protein of unknown function (DUF883)	Cytoplasmic 3.784 *	2qm3A- 26- 37- 0.003
11093885	No	No	Domain of unknown function (DUF1998)	Cytoplasmic 3.140 *	1s3eA -14- 31- 0.007
11093886	Catalytic domain of phospholipase D superfamily proteins; Catalytic domain of phospholipase D (PLD) superfamily proteins.	Phospholipase D/Transphosphatidylase	PLD-like domain	Cytoplasmic 3.582 *	2ze9A -16- 55- 1e-08
11093890	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	Winged helix-turn-helix DNA-binding domain	Helix-turn-helix domain	Cytoplasmic 3.464 *	2d1hA- 20- 37- 0.004
11093918	No	unintegrated	No	Cytoplasmic 4.821 *	2ewrA- 11- 37- 0.007
11093919	Winged helix DNA-binding domain	MarR-type HTH domain	Winged helix DNA-binding domain, Helix-turn-helix domain	Cytoplasmic 4.712 *	2jscB- 19- 43- 6e-05

11093922	No	No	PD-(D/E)XK nuclease superfamily	Cytoplasmic 3.847 *	1q16B- 21- 38- 6e-04
11093930	Predicted nucleic acid-binding protein, contains PIN domain [General function prediction only]	No	No	Cytoplasmic 2.283 *	1o4wA- 20- 44- 1e-05
11093938	No	No	No	Cytoplasmic 3.558 *	1zakA- 18- 43- 9e-05
11093941	No	No	No	Extracellular 1.418 * Cytoplasmic 1.267 * OuterMembrane 1.208 *	2d73A -13 -36- 2e-04
11093945	Plasmid pRiA4b ORF-3-like protein; Members of this family are similar to the protein product of ORF-3 found on plasmid pRiA4 in the bacterium Agrobacterium rhizogenes.	Plasmid pRiA4b, Orf3	Plasmid pRiA4b ORF-3-like protein, Activator of Hsp90 ATPase homolog 1-like protein	Cytoplasmic 3.326 *	2i1sB- 16- 120- 1e-28
11093960	Ferritin-like superfamily of diiron-containing four-helix-bundle proteins; Ferritin-like, diiron-carboxylate proteins participate in a range of functions including iron regulation, mono-oxygenation, and reactive radical production.	Twin-arginine translocation pathway, signal sequence	Ferritin-like domain	Cytoplasmic 1.788 * Extracellular 1.689 *	2ib0A- 18- 99- 1e-21
11093963	DNA breaking-rejoining enzymes, intergrase/recombinases, C-terminal catalytic domain.	Integrase, catalytic	Phage integrase family	Cytoplasmic 2.718 *	1a0pA -14- 113- 5e-26
11093965	Plasmid pRiA4b ORF-3-like protein; Members of this family are similar to the protein product of ORF-3 found on plasmid pRiA4 in the bacterium Agrobacterium rhizogenes.	Plasmid pRiA4b, Orf3	Plasmid pRiA4b ORF-3-like protein	Cytoplasmic 3.170 *	2i1sB -18 -129- 2e-31
11093967	No	No	No	Cytoplasmic 3.545 *	2bj7B- 19- 38- 9e-04
11093972	Glyoxalase-like domain; This domain is related to the Glyoxalase domain pfam00903.	Glyoxalase-like domain	Glyoxalase-like domain	Cytoplasmic 4.619 *	1cjxA- 15- 39- 0.001
11093976	Sulfatase	Alkaline phosphatase-like, alpha/beta/alpha	Sulfatase	Cytoplasmic 4.681 *	2qzuA -15- 63- 6e-11
11093978	4-amino-4-deoxy-L-arabinose transferase and related	No	No	InnerMembrane 4.809 *	2ahxB -20- 35- 3e-04

	glycosyltransferases of PMT family [Cell envelope biogenesis, outer membrane]				
11093997	Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein von Willebrand factor (vWF).	von Willebrand factor, type A	von Willebrand factor type A domain	Cytoplasmic 2.290 *	2odpA- 14- 39- 6e-04
11094010	type 2 lantibiotic biosynthesis protein LanM; Members of this family are known generally as LanM, a multifunctional enzyme of lantibiotic biosynthesis.	Six-hairpin glycosidase-like	Pectic acid lyase	Cytoplasmic 3.993 *	2ahfA -11- 41- 0.001
11094014	Predicted membrane-associated, metal-dependent hydrolase [General function prediction only]	Alkaline phosphatase-like, alpha/beta/alpha	Domain of unknown function (DUF3512)	Cytoplasmic 4.655 *	2qzuA- 13 -44- 2e-05
11094019	Bacterial PH domain	Bacterial PH domain	Bacterial PH domain, Protein of unknown function (DUF1469)	InnerMembrane 4.365 *	1yy9A- 16- 39- 3e-05
11094034	DDE superfamily endonuclease	Homeodomain-like	Winged helix-turn helix, DDE superfamily endonuclease	Cytoplasmic 4.844 *	1pdnC- 16- 38- 0.003
11094047	No	No	Uncharacterised nucleotidyltransferase, Carboxypeptidase activation peptide	Cytoplasmic 4.348 *	2ewrA- 12- 58- 2e-09
11094051	PIN domain;	Regulator of G protein signalling superfamily	PIN domain	Cytoplasmic 4.381 *	2fe1A- 24- 40- 0.001
11094057	Tetratricopeptide repeat domain	Zinc finger, C2H2	Tetratricopeptide repeat, Drought induced 19 protein (Di19), zinc-binding, Protein involved in formate dehydrogenase formation	Cytoplasmic 4.359 *	1w3bB- 12- 42- 2e-04
11094058	Calcium binding; CcbP is a Ca(2+) binding protein which, in Anabaena, is thought to bind Ca(2+) by protein surface charge.	Uncharacterised protein family, calcium binding protein, CcbP	Calcium binding	Cytoplasmic 3.700 *	2p0pA- 41- 116- 2e-27
11094069	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors.	unintegrated	MarR family	Cytoplasmic 4.553 *	1hw1A- 16- 36- 0.007
11094100	No	No	No	Cytoplasmic 4.841 *	1rz4A- 27- 36 - 0.007

Halophilic archaeon DL31 (Plasmid 2)

Gene ID	CDD-Blast	Interproscan	Pfam	Cello	PS2 structure
11057176	No	No	Hydrogenase expression/synthesis hypA family, Double zinc ribbon, DnaJ central domain, NMD3 family, Cytochrome c7, Aspartate carbamoyltransferase regulatory chain, metal binding domain	Periplasmic 1.589 * Extracellular 1.533 * Cytoplasmic 1.318 *	2pvxB- 28- 34- 0.005
11057178	No	No	Winged helix-turn-helix DNA-binding	Cytoplasmic 3.535 *	1q1hA- 26- 37- 0.003
11057186	No	unintegrated	No	Cytoplasmic 3.056 *	1itkB- 30- 35- 0.006

ACKNOWLEDGEMENTS

Archis Panday (M.Sc. trainee) and Azeem Siddique (M.tech. trainee) wants to thanks Swapnil Sanmukh (Research fellow) and Krishna Khairnar (Scientist) for their assistance and help during the whole study period.

REFERENCES

[1] Alex, B., Lachlan, C., Richard, D., Robert, D. F., Volker, H., Sam, G.J., Ajay, K., Mhairi, M., Simon, M., Erik, L. L. S., David, J. S., Corin Y., Sean, R. E., (2004). The Pfam families' database. Nucleic Acids Research, Vol. 32, D138-D141.

[2] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., (1997). Gapped BLAST and PSI-BLAST: "a new generation of protein database search programs". Nucleic Acids Res. 25 (17), 3389-402.

[3] Aron, M. Bauer., John, B. A., Myra, K. D., Carol, D. S., Noreen, R. G., Marc, G., Luning, H., Siqian, H., David, I. H., John, D. J., Zhaoxi, K., Dmitri, K., Christopher, J. L.,Cynthia A. L., Chunlei, L., Fu, L., Shennan, L., Gabriele, H. M., Mikhail, M., James, S. S., Narmada, T., Roxanne, A. Y., Jodie, J. Y., Dachuan, Z., Stephen, H. B., (2006). CDD: "*a conserved domain* database for interactive domain family analysis. "Nucleic Acids Research, Vol. 35, D237–D240.

[4] Benlloch, S., Martínez-Murcia, A.J., and Rodríguez-Valera, F. (1995a) Sequencing of bacterial and archaeal 16S rRNA genes directly amplified from a hypersaline environment. Syst Appl Microbiol 18: 574–581.

[5] Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML (2000) Cultivation of Walsby's sqare haloarchaeon. FEMS Microbiol Lett 238:469–473

[6] Eichler, J., and M. W. Adams. 2005. Posttranslational protein modification in archaea. Microbiol. Mol. Biol. Rev. 69:393–425.

[7] Fernandez-Castillo R, Rodriguez-Valera F, Gonzalez-Ramos J, Ruiz- Berraquero F (1986) Accumulation of poly(-hydroxybutyrate) by halobacteria. Appl Environ Microbiol 51:214–216

[8] Graf, R., S. Anzali, J. Bu¨nger, F. Pflu¨cker, and H. Driller. 2008. The multifunctional role of ectoine as a natural cell protectant. Clinics Dermatol. 26:326–333.

[9] Han, J., Q. Lu, L. Zhou, H. Liu, and H. Xiang. 2009. Identification of the polyhydroxyalkanoate (PHA)-specific acetoacetyl coenzyme A reductase among multiple FabG paralogs in Haloarcula hispanica and reconstruction of the PHA biosynthetic pathway in Haloferax volcanii. Appl. Environ. Microbiol. 75:6168–6175.

[10] Hezayen FF, Tindall BJ, Steinbüchel A, Rehm BHA (2002a) Characterization of a novel Halophilic archaeon, Halobiforma haloterrestris gen. nov., sp. nov., and transfer of Natronobacterium nitratireducens to Halobiforma nitratireducens comb. nov. Int J Syst Evol Microbiol 52:2271–2280

[11] Kanz, C. et al. (2005) The EMBL Nucleotide Sequence Database. Nucleic Acids Res., 33, D29–D33.

[12] Kirk RG, Ginzburg M (1972) Ultrastructure of two species of halobacterium. J Ultrastruct Res 41:80–94

[13] Lu, Q., J. Han, L. Zhou, J. Zhou, and H. Xiang. 2008. Genetic and biochemical characterization of the poly (3-hydroxybutyrate-co-3-hydroxyvalerate) synthase in Haloferax mediterranei. J. Bacteriol. 190:4173–4180.

[14] Lucas, S., Han,J., Lapidus,A., Cheng, J.F., Goodwin,L., Pitluck,S., Peters,L., Mikhailova,N., Teshima,H., Detter,J.C., Han,C., Tapia,R., Land,M., Hauser,L., Kyrpides,N., Ivanova,N., Pagani,I., Dyall-Smith,M., Cavicchioli, R. and Woyke,T. Complete sequence of chromosome of Halophilic archaeon DL31 (Unpublished).

[15] Mascellani, N., X. Liu, S. Rossi, J. Marchesini, D. Valentini, D. Arcelli, C. Taccioli, M. H. Citterich, C.-G. Liu, R. Evangelisti, G. Russo, J. M. Santos, C. M. Croce, and S. Volinia. 2007. Compatible solutes from hyperthermophiles improve the quality of DNA microarrays. BMC Biotechnol. 7:82.

[16] Mescher, M. F., and J. L. Strominger. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. J. Biol. Chem. 251:2005–2014.

[17] Oren A (2008) Correct names of taxa within the family Halobacteriaceae – May 2008.

[18] Oren A, Ginzburg M, Ginzburg BZ, Hochstein LI, Volcani BE (1990) Haloarcula marismortui (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. Int J Syst Bacteriol 40:209–210

[19] Oren, A. 2010. Industrial and environmental applications of halophilic microorganisms. Environ. Technol. 31:825–834.

[20] Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997). CATH--a hierarchic classification of protein domain. Structures.15: 5(8):1093-108.

[21] Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. (2005) InterProScan: protein domains identifier. Nucleic Acids Res. 33 (Web Server issue):W116-W120

[22] Quillaguaman, J., H. Guzman, D. Van-Thuoc, and R. Hatti-Kaul. 2010. Synthesis and production of polyhydroxyalkanoate by halophiles: current potential and future prospects. Appl. Microbiol. Biotechnol. 85:1687–1696.

[23] Rodríguez-Valera, F. (1988) Characteristics and microbial ecology of hypersaline environments. In Halophilic Bacteria.

[24] Rodríguez-Valera, F., Ruíz-Berraquero, F., and Ramos-Cormenzana, A. (1981) Characteristics of the heterotrophic bacterial populations in hypersaline environments of different salt concentrations. Microb Ecol 7: 235–243.

[25] Rodríguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J.F. (1985) Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. Microb Ecol 11: 107–115.

[26] Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S. Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F., (2001*)." Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements".* Nucleic Acids Res. 29(14), 2994-3005.

[27] Schnoor, M., P. Voß, P. Cullen, T. Boking, H. J. Galla, E. A. Galinski, and S. Lorkowski. 2004. Characterization of the synthetic compatible solute homoectoine as a potent PCR enhancer. Biochem. Biophys. Res. Commun. 322:867–872.

[28] Sorensen KB, Canfield DE, Teske AP, Oren A (2005) Community composition of a hypersaline endoevaporitic microbial mat. Appl Environ Microbiol 71: 7352–7365.

[29] Swapnil G. Sanmukh, Waman N. Paunikar, Tarun K. Ghosh and Tapan Chakrabarti 2011. Structural & Functional Prediction of Hypothetical Proteins In Bacteriophages Against Halophilic Bacteria - An In Silico Approach. Int J Pharm. Bio. Sci. Vol 2 (2), B61-B70

[30] Van-Thuoc, D., H. Guzman, J. Quillaguama´n, and R. Hatti-Kaul. 2010. High productivity of ectoines by Halomonas boliviensis using a combined two-step fed-batch culture and milking process. J. Biotechnol. 147:46–51.

[31] Yu CS, Chen YC, Lu CH, Hwang JK: Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics 2006, 64:643-651.

[32] Yu CS, Lin CJ, Hwang JK: Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science 2004, 13:1402-1406.

[33] Zafer, A., Yucel, A., Mark, B. Protein secondary structure prediction for a single-sequence using hidden semi-Markov models, BMC Bioinformatics ,7, 178, 2006.

[34] Zdobnov, E. M., Rolf, A. Interproscan- an integration platform for the signatures recognition methods in InterPro. Bioinformatics 17,847-848, 2001.