Journal of Computer Science & Systems Biology - Open Access

JCSB/Vol.3 Issue 1

Research Article OPEN ACCESS Freely available online doi:10.4172/jcsb.1000051

A Phylogenetic Approach for Assigning Function of Hypothetical Proteins in *photorhabdus luminescens* Subsp. *laumondii* TT01 Genome

Razia M[1], Karthik Raja R[1], Padmanaban K[1], Sivaramakrishnan S[1] and Chellapandi P[2,*]

[1]Department of Biotechnology, Bharathidasan University, Tiruchirappalli-620024, Tamilnadu, India
[2]Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli-620024, Tamilnadu, India

Abstract

Polyketides are larger family of structurally diverse natural products with a broad range of biological activity. The occurrence of polyketide synthase gene family/cluster in bacteria is capable to synthesis the polyketides. In this study, sequence and structural similarities of some hypothetical proteins of *Photorhabdus luminescens* subsp. *laumondii* TT01 analyzed to assign the functional relationship with polyketide synthases (PKSs) using bioinformatics tools. Many hypothetical proteins of this organism have shown homologies to PKS family on which a significant homolog found to be located at genomic region 2247626-2249476 bp in chromosome 1 carrying identical function. On searching motif and domain, it showed a strong similarity to ketoacyl (KA) synthase I and then to acyl carrier protein. The ketosynthase-acyltransferasedidomain module 5 (2HG4) of *Saccharopolyspora erythraea* found as a good ortholog and the best template for modeling 3D structure from the sequences of hypothetical proteins. ProFunc and Castp servers used to annotate the structure-function relationship of protein models. The structural aspects at primary and secondary levels also showed a close resemblance to KA synthase. Phylogenetic analysis of this sequence and protein model ensured its function would be β-KA synthase showing the functional reliability like ketosynthase, and it has evolutionary relationship with soil bacteria. There was a horizontal gene transfer event to acquire this domain in *P.luminescence* genome. Consequently, an abundance of PKS gene in the genome of entomopathogenic bacteria will obviously helpful to protect its host nematode from other pathological pervasiveness.

Keywords: Entomopathogenic nematodes; *Heterorhabdus*; *Photorhabdus*; Ketoacyl synthase; Phylogeny; Polyketides; Functional assignment

Introduction

Entomopathogenic nematodes (EPNs) are unique model for the study of parasitism, pathogenicity, and symbiosis (Sandhu et al., 2006). *Photorhabdus* spp. is a gram-negative bacteria symbiotically associated with entomopathogenic nematodes of genera *Heterorhabdus* belongs to the family Heterorhabditidae. The bacterium is thought to release a wide variety of potential virulence factors, high-molecular-weight toxin complexes, lipopolysaccharide, proteases, lipases and range of different antibiotics, and bacteriocins that appear to inhibit other saprophytic microorganisms in the insect cadaver, and provide nutrients utilized by the nematodes (Ernst, 2000). Thus, the bacteria–nematode complex has been exploited as a potential biocontrol agent against several insect pests (Forst and Nelson, 1996; Ffrench-Constant et al., 2000; Sandhu et al., 2006).

The genome of *Photorhabdus luminescens* subsp.*laumondii,* symbiont of *Heterorhabdities bacteriophora*, has been fully sequenced. The 5.7 Mb bacterial genome contains 4839 predicted genes and a diverse array of potential virulence factor-encoded genes (*yops*, a yersinibactin-like siderophore and the low calcium-response stimulan), including genes for several classes of toxins. The potential for horizontal gene transfer raises the intriguing possibility that the virulence factors present in invertebrate pathogens may also be present in vertebrate pathogens (Ffrench-Constant et al., 2000; Ginolhac et al., 2005). *Photorhabdus* spp. produces a various metabolites having the properties like antibacterial, antifungal, cytotoxic and nematicidal, which are apparently of mixed peptide–polyketide origin (Hranueli et al., 2001). Polyketide synthases (PKS) polymerize simple fatty acids into a large variety of different products called polyketides by successive decarboxylating Claisen condensations. PKS can be divided into 2 groups, modular type I PKS consisting of one or more large multifunctional proteins (Scott et al., 2003) and iterative type II PKS, complexes of several monofunctional subunits (Hranueli et al., 2001; Brachmann et al., 2007). Dittmann et al., (1997) stated that PKS is responsible for nonribosomal synthesis of a diverse array of compounds involved in processes ranging from fatty acid synthesis to antibiotic production (including production of inhibitors of eukaryotic protein phosphatases). Some of the predominant classes of polyketide synthetase-like sequences, 31 are hit with a syringomycin synthetase from *Pseudomonas syringae* pv. *syringae* and the syringomycin synthetase gene cluster. Syringomycin itself has a wide range of antibacterial and antifungal properties (Gerth et al., 1994; Kealey et al., 1998; Hoffmann and Valancia, 2004).

Deployment of similar antibiotics produced by *P. luminescens*

***Corresponding author:** Chellapandi P, Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli-620024, Tamilnadu, India, Tel: +91-431-2407071; Fax: +91-431-2407045; Email: pchellapandi@gmail.com

Received January 19, 2010; **Accepted** February 08, 2010; **Published** February 08, 2010

Citation: Razia M, Raja KR, Padmanaban K, Sivaramakrishnan S, Chellapandi P (2010) A Phylogenetic Approach for Assigning Function of Hypothetical Proteins in *Photorhabdus luminescens* Subsp. *laumondii* TT01 Genome. J Comput Sci Syst Biol 3: 021-029. doi:10.4172/jcsb.1000051

W14 may, aid to maintain a bacterial monoculture in an insect cadaver (Ffrench-Constant et al., 2000). *P. luminescens* also contains a sequence that is similar to tolaasin (another lipodepsipeptide), which exploits for self-protection in *Pseudomonas tolaasii*, and colicin activity proteins, colicin transport proteins and pyocin immunity proteins. In addition to genes for specific mechanisms for antibiotic production and self-protection, the W14 genome contains numerous sequences that exhibit homology to genes for other antibiotic resistance mechanisms (Ffrench-Constant et al., 2000). These sequences include genes involved in resistance to penicillin (penicillinase and penicillin-binding protein), bicyclomycin, and a range of other antibiotics (tetracycline, rifampin, and kasugamycin) via a variety of different mechanisms (Metsa-Ketela et al., 2002; Jenke-Kodama et al., 2005). The large number of sequences exhibit homology to genes for different multiple-drug-like export systems, including Emr-like and Mdl-like systems that export drugs ranging from chloramphenicol to acriflavin. Stilben (3, 5-dihydroxy-4-isopropyl-stilben) produced in *P.luminescens* TTO1 due to the presence of key enzyme phenylalanine ammonia-lyase has antibacterial, anti-fungal and anti-nematode activity (Williams et al., 2005). Recently, a biosynthesis gene cluster responsible for the production of anthraquinones (AQs) from the *P.luminescens* has also been identified (Brachmann et al., 2007). These are potentially deploying broad-spectrum antibiotics to repel other organisms that might colonize the insect cadaver.

Inferring protein function is a challenging task as global protein sequence and structure similarities are often unreliable for function prediction and their evolution histories contain direct functional information (Liang et al., 1998). Thus, the present study was aimed to use sequence and structural similarities and identities of polyketide synthase to assign the function of hypothetical proteins in *P.luminescens* subsp. *laumondii* TT01. In addition, phylogenetic relationship of the protein sequences and modeled 3D structures was studied to ensure its functional fidelity and to understand the evolutionary mechanism of PKS family in entomopathogenic bacteria. Antagonist effects of this bacterium on other invading bacteria in nematodes and endosymbiotic mechanism have not yet been determined well. Perhaps, this work will describe how this gene was acquired and diverged in the genome of *P.lminescens* during evolution.

Materials and Methods

Sequence retrieval and similarity search

Sequences of polyketide synthase encoded genes were retrieved from iHOP server (Hoffmann and Valencia, 2004) and the deduced protein sequences were used to retrieve similar sequences from a completed genome of *P.luminescens* subsp. *laumondii* TT01 by using BLASTn, BLASTp and PSI-BLAST tools (Altschul et al., 1997) with default parameters. Sequences with low e-values and high identity scores were screened to obtain PKS homologs in which three sequences such as NCBI accessions NP_929573, NP_929153 and NP_929148 gave many hits to PKS sequences of other organisms. BLASTp and PSI-BLAST were carried out to individual coding regions of these sequences for searching similarity sequences to be retrieved from SwissProt and Protein Data Bank (PDB). Sequences with a reasonable similarity to query sequence were selected and then clustered using MEGA 4.0 software (Tamura et al., 2007) implemented with ClustalW algorithm. All the aligned sequences were inspected manually and edited to unreliable sequences to minimize the gaps.

Prediction of protein structure and function

The selected coding regions were subjected to domain and motif predictions using MyHits server (http://hits.isb-sib.ch/cgi-bin/PFSCAN) with motif databases. Conserved domain classification of query sequence was searched in NCBI conserved domain architecture (Marchler-Bauer et al. 2005). Primary and secondary structural features of proteins were computed with ProtParam (http://expasy.org/tools/protparam.html) and SOPMA server (Geourjon and Deleage, 1995), respectively. Swiss-Model is an automatic comparative protein modeling server (Schwede et al., 2003), where the query sequence was uploaded to build three dimensional structures based a template. A quality of modeled protein structure was further validated with SAVS (Structure Analysis and Verification) server (http://nihserver.mbi.ucla.edu/SAVS/) using Prove and ProCheck algorithms. Structural alignment of a protein model and a template was carryout by DaliLite server (Holm and Park, 2000) and then RMSD and Z-score were computed. Castp server (Binkowski et al., 2003) with weighted Delaunay triangulation and the alpha complex for shape measurements and Active Site Prediction server (Brylinski et al., 2006) with fussy-oil drop model was used to search functional regions in binding packets of a protein model. The function of each model was annotated by ProFunc server (Laskowski et al., 2005) uploading PDB file. Protein-coding and promoter regions of our query sequence were predicted by WebGene server (Milanesi et al., 1999) and Web Promoter Scan service (bimas.dcrt.nih.gov/molbio/proscan). Primers sequence and hybridization probes were designed by Primer3 server (Rozen and Skaletsky, 2000) with default settings.

Phylogenetic tree building and analysis

The orthologs and paralogs searches were carried out among all organisms by Kegg-SSDB search using Smith-Waterman (SW) score (Smith and Waterman, 1981). Both homologous sequences and structures were used for searching phylogenetic trees. The selected sequences were clustered with complete deletion of gaps using ClustalX 2.0 software (Thompson et al., 1997). Then after, Neighbor joining (Gascuel and Steel, 2006) and Minimum evolution (Kumar, 1996) trees were searched homogeneous patterns among all lineages using MEGA 4.0 software with 1000 bootstraps values (Bradley et al., 1996), JTT model (Jones et al., 1992) along 0.25 gamma distribution, at uniform rates among sites. Structural based phylogenetic analysis was done by uploading model protein (in PDB format) in ConSurf server (Landau et al., 2005) and structural alignment and created tree were viewed in Rasmol (http://www.umass.edu/microbio/rasmol) and MEGA 4.0 software, respectively.

Results and Discussion

Sequence similarity analysis

Above 50 PKS sequences obtained from iHOP database that were used to search similarity sequences in *P. luminescens* subsp. *laumondii* genome using NCBI and KEGG BLASTp and PSI-BLAST tools. It resulted that most PKS sequences from different organisms matched with three hypothetical proteins, NP_929153, NP_929573 and NP_929148 of this organism out of which a hit NP_929153 showed the significant e-value (3e-

102) with many organisms. Hence, the best hit (NP_929153) was chosen as a candidate of this study for further analyses of its function and structure. The PSI-BLAST search results obtained from PDB shown that the ketosynthase-acyl transferase sequence (2HG4) of *Saccharopolyspora erythraea* was highly matched with query sequence, and obtained e-value was 3e-78 (Table 1). Later on, query sequence showed its sequence similarity with β-ketoacyl - acyl carrier protein (ACP) synthase of *Homo sapiens* and of many enterobacterial families. The e-value was ranged from 3e-78 to 6e-13. Although, when it searched for similar sequences from SwissProt we obtained significant PKS hits with diverged functional activities from *Bacillus subtilis, Mycobacterium tuberculosis*, and *S. erythraea*. This preliminary data has supported to suggest the hypothetical protein having some functional properties like PKS family.

Motif and domain similarities analyses

Motif regions of this protein sequence analyzed with My Hits, followed by collective information retrieved from Prosite database. It showed that 5 regions related to β-ketoacyl (KA) synthases and 2 regions corresponded to 3-oxo (ACP) synthase on which β-KA synthases active site (201-217) predicted as a motif region. The motif region was G - {A} - {KGR} - x (2) - [LIVMFTAP] - {R} - x - [AGC] - C - [STA] (2) - [STAG] – x (2) - {LI} - [LIVMF] where C was the active site residue. While searching conserved domain of query sequence, it matched with PKS (1- 422 aa) and FabB (1-375 aa) domains with e-values 7e-139 and 6e-41, respectively. The analyses of conserved domain and motif similarities pointed out that there was a strong functional relationship between hypothetical protein and PKS family. The extraordinary high diversity of polyketide products, however, is achieved by an optional use of domains for the modification of keto groups and by the use of different substrates for chain initiation and extension. It has been calculated that a PKS system comprising six modules is theoretically able to produce over 100 000 possible structures (Gonzalez-Lergier et al., 2005). The majority of bacterial PKS I consist of multiple sets of domains, or modules, that normally correspond to the number of acyl units in the product (Staunton and Weissman, 2001). A minimal module is composed of a ketoacyl synthase (KS) domain, acyltransferase (AT) domain and acyl carrier protein (ACP) domain. Frequently ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains are embedded in the multifunctional mega synthases (Hranueli et al., 2001; Bode and Muller, 2005).

Structural similarity analysis

Unlike fatty acids, the structures of polyketides are far more diverse due to variations in the fatty acid synthesis theme and post-PKS modifications. This structural diversity is moreover reflected in diversity in their biological modes of action (Hopwood, 1990). Protein structure is fundamental to phenotype and yet little previous effort has been devoted to characterizing its impact on evolution (Nick et al., 1998). Secondary structural similarities and identities may also be impacted on the conservation of amino acids in a protein. Since, ProtParam program used to predict the primary structure of query protein sequence such as negatively (41) and positively charged amino acids residues (33), grand average of hydrophobicity (- 0.016), pI (5.39), and aliphatic index (89.5). These features were closely related with β-KA synthase and fatty acid synthase, suggesting the functional resemblance even at the primary structure level. Secondary structure of this protein has 33.81 % α-helix, 39.95 % random coil, 8.27 % β-turn and 17.95 % extended coil in which majority of these features, particularly random coils and α-helix matched with KA synthase of different organisms (Figure 1). It is also important to note that secondary structural elements are more conserved than the precise atomic structure (Mizuguchi

Function	PDB	Organism	e-value
Ketosynthase-Acyltransferase Didomain of Module 5 From Debs	2HG4	*Saccharopolyspora erythraea*	3e-78
Mitochontrial β-keto acyl ACP synthase	2IWY	*Homo sapiens*	5e-27
Mitochontrial β-keto acyl ACP synthase	2C9H	*Homo sapiens*	6e-26
3-Oxoacyl –(ACP)	1J3N	*Thermus Thermuphilus*	3e-23
Mitochontrial KAS	1WOI	*Arabidopsis thaliana*	4e-21
β-Ketoacyl (ACP) synthase Ii (Mtkasb)	2GP6	*Mycobacterium tuberculosis*	7e-20
β-Ketoacyl (ACP) Synthase Ii	1KAS	*Escherichia coli*	2e-19
Fab f (kasii)	2GFW	*Escherichia coli*	2e-19
Fabf (K335a) mutant with covalently linked Dodecanoic acid	2GFY	*Escherichia coli*	1e-18
Actinorhodin ketosynthase chain length factor	1TQY	*Streptomyces coelicolor*	2e-18
Fab f (Kasii) C163q mutant	2GFV	*Escherichia coli*	7e-18
β–Ketoacyl (ACP)	1DD8	*Escherichia coli*	4e-16
β–Ketoacyl(ACP) synthase Ii	1OXO	*Streptococcus pneumoniae*	4e-16
β–Ketoacyl(ACP) synthase I in complex with Thiolactomycin	1FJ4	*Escherichia coli*	6e-16
β–Ketoacyl(ACP) synthase I K328r	1H4F	*Escherichia coli*	1e-15
β–Ketoacyl (ACP) synthase Ii (Fab f)	2GQD	*Staphylococcus aureus*	3e-15
β–Ketoacyl(ACP) synthase Lys 328 ala mutant	2BYW	*Escherichia coli*	3e-15
β–Ketoacyl(ACP) synthase Ii	2ALM	*Streptococcus pneumoniae*	3e-15
Kas I H298q Mutant In complex with C12 fatty acid	2BYZ	*Escherichia coli*	5e-15
Kas I H298e Mutation	2BYY	*Escherichia coli*	6e-15
β-Acyl carrier protein Synthase Ii (Kasii)	1E5M	*Synechocystis sp. Pcc6803*	1e-14
β-Ketoacyl (ACP) synthase	1F91	*Escherichia coli*	6e-13
β-Keto (ACP) synthase I in complex with C10 Fatty Acid Substrate	1EK4	*Escherichia coli*	6e-13
Actinorhodin ketosynthase Chain length Factor	1TQY	*Streptomyces coelicolor*	3e-10

Table 1: Comprehensive information of protein structures used in the phylogenetic analysis.

```
TARGET   69   APFFGISPRV AAMMDPQQRM LLELTWQAIE DSGANPLGYS GSKTGVFIGS
2hg4D   100   AGFFGISPRE ALAXDPQQRT XLEISWEALE RAGHDPVSLR GSATGVFTGV

TARGET        hhh   hhh hh   hhhhh hhhhhhhhhh h     hh       ssssss
2hg4D         hhh   hhh h   hhhhhh hhhhhhhhhh h     hh       ssssss

TARGET   169   SSGLTALTQA VNSLRSGECQ QAIVGSVNLL SNTFNMAAYY RAGMLSKDGC
2hg4D    200   SSGLTALHLA XESLRRDECG LALAGGVTVX SSPGAFTEFR SQGGLAADGR

TARGET        hhhhhhhhhh hhhhhh     ssssssss      hhhhhhh
2hg4D         hhhhhhhhhh hhhhhh     ssssssss      hhhhhhh

TARGET    269  NSLTSPNPEQ QIALVKDCLL QAGISAEQIS YLEAHGTGTS LGDPIEFNAL
2hg4D     300  NGLTAPSGPA QQRVIRRALE NAGVRAGDVD YVEAHGTGTR LGDPIEVHAL

TARGET               hhh hhhhhhhhhh hh         ssss     h hhhhhhhhhh
2hg4D                hhh hhhhhhhhhh hh       s ssss     h hhhhhhhhhh

TARGET    369  PSAAFQRLNP EIDSVDSRLQ LATEENSWRV GAGQKRFAGL SSFGLGGSNS
2hg4D     398  RTLHFDEPSP QIEW-DLAVS VVSQARSWPA GERPRR-AGV SSFGISGTNA
TARGET               sss           s ss   sss          sss ssss  ssss
2hg4D                sss           s ss   sss       ss sss ssss  ssss
```

Figure 1: Comparison of secondary structure of predicted β-ketoacyl synthase I (target) and structural template (2GH4) (h- helix; s-bend region).

and Go, 1995) and that protein architecture depends on constraints related to bring key residues close in space. Thus, natural selection acts on both the secondary structure elements, because of architectural constraints, and on a few critical residues directly involved in catalysis.

Apart from these resemblances, 3D structure of protein is one of main concerns to emphasis our idea about structure-function relationships due to more conserved constraints. Thus, 3D structure of this protein predicted from the sequence by searching homologous structures with a low energy conformer from PDB (Figure 2). It was structurally similar to keto synthase-acyl transferase didomain of module from *S.erythraea* (2HG4) with 41% sequence identity. According to Ramachandran plot, this model has conformational stability and likelihood to template. Maximally 69.5% quality factor, 0.6 RMSD and 68.1 Z-score have been computed to the model and thereby it reflected the close similarity to structure of many β-KA synthase.

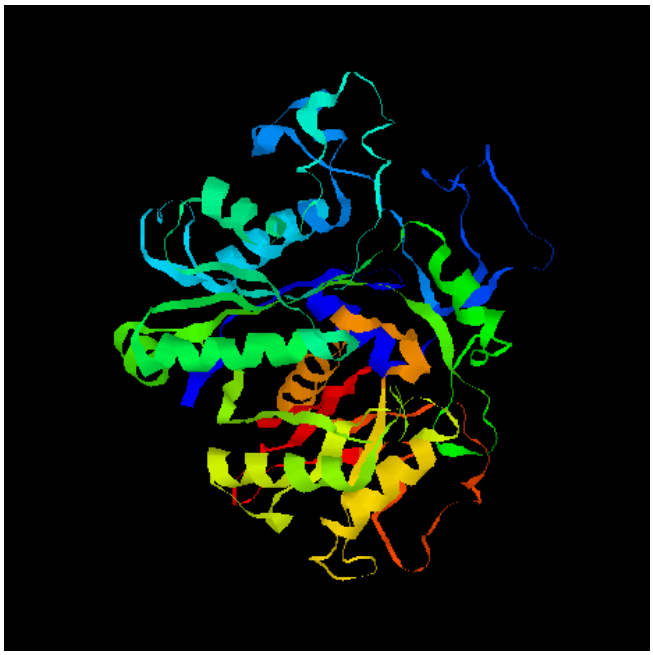

Figure 2: Modeled structure of β-ketoacyl synthase I based on the template 2HG4 using SWISS MODEL. This is a graphical view (Ribbon model) of RASMOL representing all the helixes and coils of the model.

Functional similarity analysis

The Castp, ProFunc and active site prediction tools used to predict the function of modeled protein. It shown that there were two functional regions, 43-464 and 43-288 aa, resembled to KA synthase on which second region was more significant due to less e-value (7.02e-74) and other proteins gave similarities to AMP binding domain as shown in Table 2. The predicted active site residues were Gly414, Gly415 and Ser416 (score 3.102) with average conservation score 0.769. It revealed the structure-function relationship of query protein from which its function can be assigned with more effort.

Homologous similarity analysis

Orthologous and paralogous characters are determinant factors for assigning the function of any protein. PKS of *Myxococcus xanthus* identified as the best orthologs with 1474 SW score, 0.418 identity and 622 overlap (Figure 3). Nevertheless, many orthologous hits belonged to β-KA synthase of *Burkholderia pseudomalleii*, suggested the possible ways how this gene was acquired in *P.luminescens* genome through evolutionary process. As far as we excavated data there were no significant paralogous hits from this microorganism on which many of them belonged to hypothetical proteins, however, a PKS family related protein, 3-oxoacyl-(ACP)-synthase I (404-416 aa), found with a low SW similarity score (275-451) and identity (0.254-0.292). In such case, surface similarity search based on scoring matrix constructed can lead to more sensitive and specific method for predicting protein function (Liang et al., 1998).

Sequence based phylogenetic inference

Functional predictions are of great value in guiding research and in sorting through huge data, and the increased use of phylogenetic methods can only serve to improve the accuracy of functional predictions (Eisen, 1998). Using sequence and structures of PKS from different organisms phylogenetic trees have been made to reveal how this hypothetical protein of *P.luminescens* conserving functional relationship with KA synthase. Minimum evolution tree formed 10 sub-clades include 3-oxoacyl-(ACP) synthases (I-III), trifunctional enzyme subunit beta, nodulation protein E, fatty acid synthase and actinorhodin, tetracenomycin C, granaticin and, oxytetracyclin β-KA synthases (I-II) (Figure

Accession#	Position	PDB hit	Score	E-value	ID	Motif	Nest	Quality factor*
NP_929153	43-464	2HG4 A	41%	1.63e-81	PS00606B	Ketoacyl synthase	12	63.92
	43-288	2HG4 A	41 %	7.02e-74	PS00 606B	Ketoacyl synthase	11	69.56
NP_929573	1219-1396	1AMU B	27%	2.46e-35	PS00154x2	AMP Binding	14	58.33
	3065-3466	1AMU A	31%	3.75e-56	PR00154x2	AMP Binding	9	70.99
	3560-3624	1AMU A	32%	4.32e-52	PS00455	AMP Binding	8	69.39
NP_929148	226-896	1AMU B	27%	2.42e-37	PF00501	AMP Binding	13	54.41
	484-903	1AMU B	27%	2.46e-35	PF00455	AMP Binding	14	58.33
	1090-1387	1AMU A	27%	3.08e-44	PR00154x2	AMPBinding	10	70.81
	1317-1957	1AMU A	27%	3.47e-40	PR00154x2	AMP Binding	10	70.96
	2046-2109	2HG4 A	29%	7.23e-27	PF00698	Acyl transferase	3	63.46

# denotes NCBI accession

* Quality factor of every protein models was computed by SAVS server.

Table 2: Comparative information for predicted function of hypothetical proteins in *P.luminescence* genome using ProFunc server.

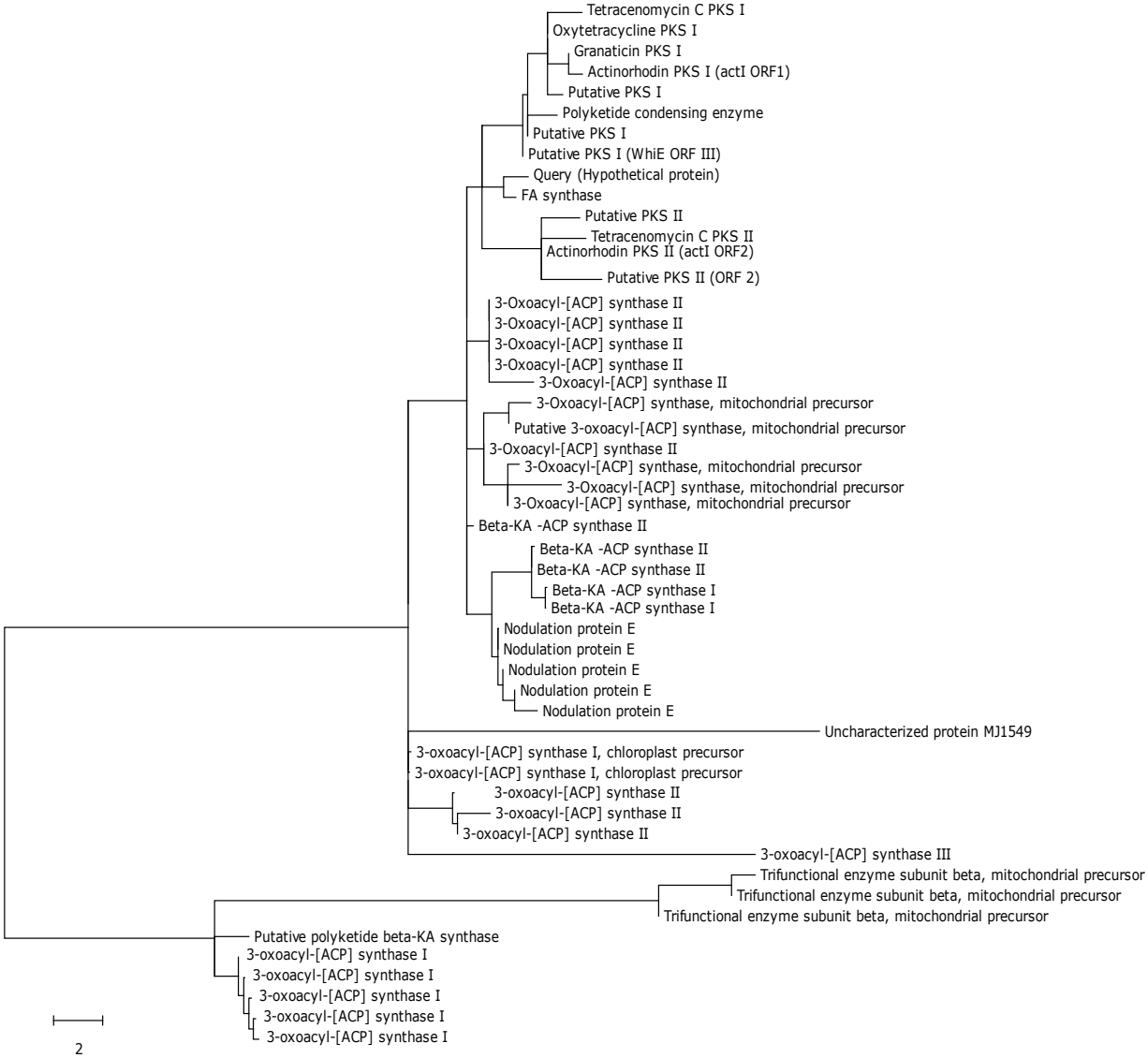

Figure 3: Phylogeny based on the sequences of PKS domains constructed by Minimum evolution algorithm. Numbers above branches indicate bootstrap support values using 1000 pseudo sequence replicates. Branch length indicates number of inferred amino acid changes per position. Tips of the tree give the names of the proteins (if annotated in the database). Major clades and subclades are indicated by vertical bars, each of which shares a common organization of domains (those in parentheses are variable in their presence or absence within that clade). Branch length indicates number of inferred amino acid changes.

3). These groups clustered separately in this tree according to their relatedness wherein our query protein formed a cluster with fatty acid synthase and then with β-KA synthases involved in biosynthesis of antimicrobial metabolites. The result indicated that β-KA synthase activity to query protein might be acquired from later one via fatty acid synthase domain. A combination of nodulation and the release of antibacterial peptides (Lavine and Strand, 2002) of *Photorhabdus* are reported to evade the insect immune system, suggesting the functional diversity of PKS family as revealed in the study. On basis of this protein sequence it was phylogenetically related with *B.subtilis, M.bovis, S.erythraea, E.coli, Streptomyces coelicolor, S.glaucescens* and *S.rimosus*. Among these organisms, many of the microorganisms belonged to soil bacteria, suggested the occurrence of KA

Accession#	Gene	Organism	Length	SW-score	Identity	Overlap#
Orthologs						
MXAN_3938	Polyketide synthase	*Myxococcus xanthus*	2088	1474	0.418	622
Psyr_4314	β-ketoacyl synthase	*Bacillus amyloliquefaciens*	3231	1443	0.396	626
RBAM_022000	DfnG	*Burkholderia pseudomallei* 1710b	5204	1422	0.385	637
BURPS1710b_A2618	OnnB	*Burkholderia pseudomallei* 668	4539	1391	0.393	595
BURPS668_A1478	Polyketide synthase, type I	*Burkholderia pseudomallei* 668	4614	1390	0.393	595
BURPS1106A_A1393	Polyketide synthase, type I	*Burkholderia pseudomallei* 1106a	4555	1390	0.393	595
BPSS1006	Polyketide synthase	*Burkholderia pseudomallei* K96243	4574	1390	0.393	595
Haur_3959	β-ketoacyl synthase	*Herpetosiphon aurantiacus*	2230	1384	0.411	586
BMA10247_A1129	Polyketide synthase	*Burkholderia mallei* NCTC 10247	1820	1384	0.392	595
BMA10299_0449	DszB	*Burkholderia mallei* NCTC 10229	1778	1384	0.392	595
Paralogs						
plu1880	Hypothetical protein	*Photorhabdus luminescens*	4160	1109	0.390	525
plu2321	Hypothetical protein	,,	3908	961	0.346	546
plu1212	Hypothetical protein	,,	1267	943	0.347	513
plu2831	3-Oxoacyl-(ACP) synthase I	,,	416	451	0.292	439
plu2217	Hypothetical protein	,,	421	305	0.246	443
plu3184	3-Oxoacyl-(ACP) synthase I	,,	404	275	0.254	343
plu4676	Hypothetical protein	,,	421	258	0.265	343
plu4191	Hypothetical protein	,,	428	230	0.220	413
plu4190	Hypothetical protein	,,	371	124	0.204	225
plu0171	Uxu operon transcriptional regulator	,,	247	102	0.280	193
plu3623	Pyruvate dehydrogenase subunit E1	,,	887	101	0.224	205

# denotes KEGG accession

Table 3: KEGG-SSDB homologs search results for query protein (NP_929153).

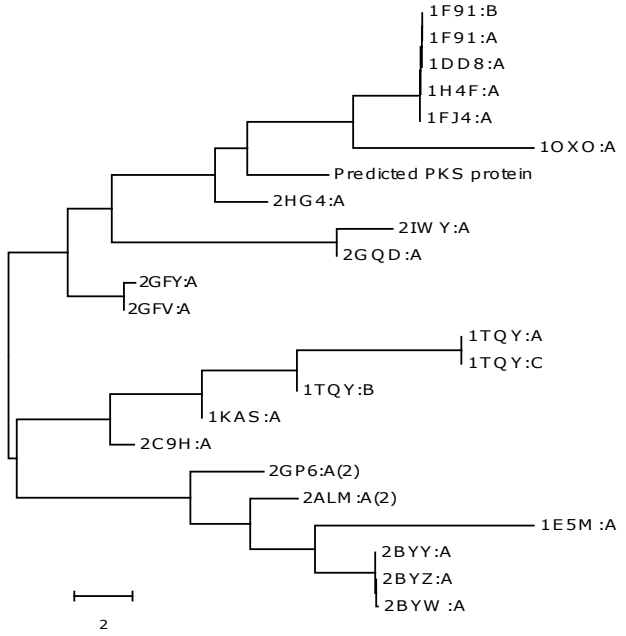

Figure 4: Phylogeny based on the structure of PKS domains constructed by Neighbor joining algorithm. Numbers above branches indicate bootstrap support values using 1000 pseudo sequence replicates. Branch length indicates number of inferred amino acid changes per position. Tips of the tree give the names of the proteins (if annotated in the database). Major clades and subclades are indicated by vertical bars, each of which shares a common organization of domains (those in parentheses are variable in their presence or absence within that clade). Branch length indicates number of inferred amino acid changes.

synthase in *P.luminescens* as a result of horizontal gene transfer mechanism (HGT) by the event of soil microbial interaction. Genetics and biochemistry of bacterial type I polyketide biosynthesis has been well investigated for the biosynthesis of the aglycone of erythromycin in *S.erythrea* (Donadio et al., 1991). A comparative sequence analysis is also revealed incongruity between the KS and 16S rRNA phylogenetic trees, which strongly suggested that there has been horizontal transfer of aromatic polyketide biosynthesis genes among *Streptomyces* species (Metsa-Ketela., 2002). In some event, the distribution of PKS among bacteria is due to duplication patterns, and the involvement of HGT processes (Ginolhac et al., 2005; Jenke-Kodama et al., 2005) and a recent horizontal acquisition followed by duplication (Stonesifer et al., 1986; Stinear et al., 2004) under selective pressure. One of the most important advances in the reconstruction of evolutionary trees has been the consideration of heterogeneity of evolutionary rates among sequence sites (Yang, 1996). Nick et al. (1998) reported that rate heterogeneity is strongly associated with structural environment. A higher rate of replacement for exposed sites is seen for each secondary structure type.

Structure based phylogenetic inference

Similarly, structure based phylogenetic analysis revealed that there was a closed proximity between our modeled protein and KA synthase of *Streptococcus pnuemoniae* and *S. erythraea*. Both structures were formed a separate cluster in neighbor joining tree (Figure 4) include β-KA I synthase, actinorhodin keto synthase , 3-oxoacyl- (ACP) synthase, fatty acid synthase f and β-

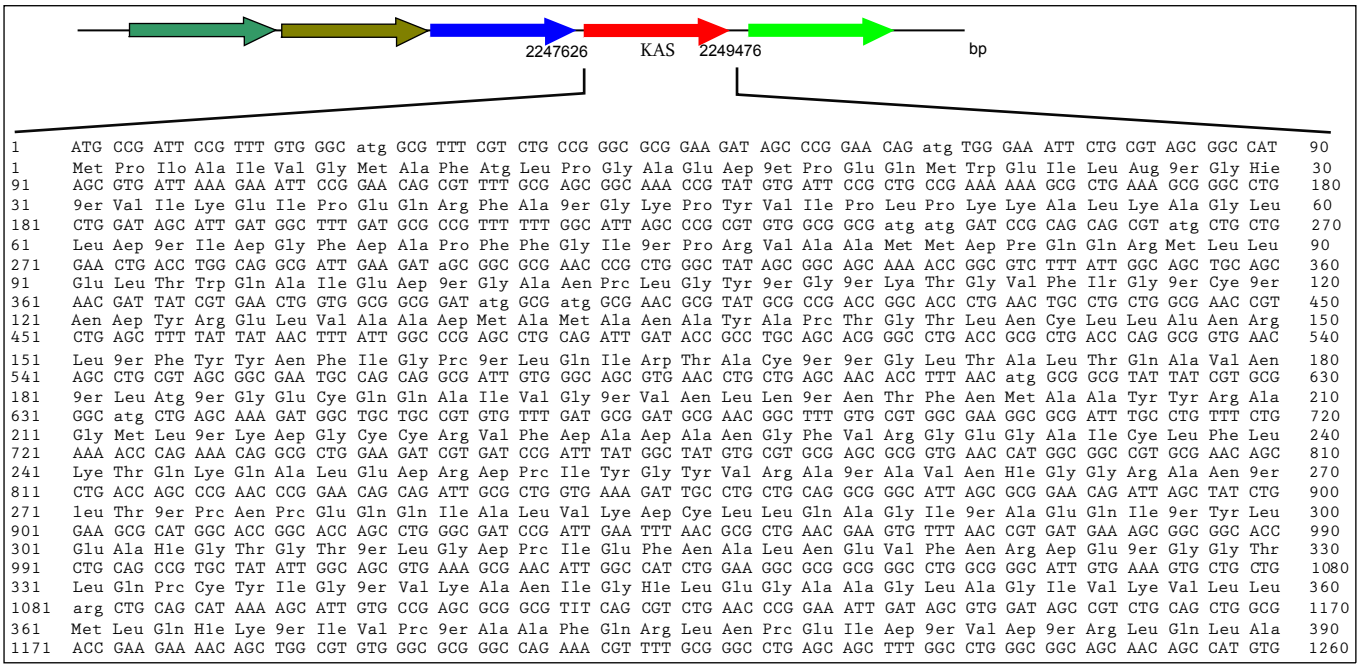

Figure 5: Structure of predicted β-ketoacyl synthase I gene and its codon usages.

ACP synthase. It may be attributed by structural diversity of PKS involved in polyketides metabolism that is likely to have evolved from fatty acid biosynthesis (Hopwood, 1990). Unlike sequence based phylogenetic tree, our query protein was more corresponded with enterobacteria, particularly *E.coli*, *M. tuberculosis* etc., but functionally showed well relatedness as similar to previous tree. A phylogram of a 60% majority rule consensus tree had a topology similar to that of the maximum likelihood tree. A bootstrap analysis with 1000 replications, performed to check the robustness of the most-parsimonious tree, yielded a tree with more phylogenetic relationship with enterobacteria and soil bacteria for sharing KA domain. There was a close correspondence between the taxonomic grouping of the bacteria and the taxonomic grouping of their nematode associates. Thus, these phylogenetic analyses strongly support us for existence of β-KA synthase as a functional form in this organism, which might be originated from PKS family of soil microbes through evolutionary mechanisms either HGT or recombination process.

Recombination frequency analysis

Recombination of common families among each other contributes more to the process of divergence at the level of domain combinations than common families combining with kingdom-specific families (Apic et al., 2001). Phylogenetic distribution of acetyl transferase and KS domain sequences revealed that multiple gene duplications, gene losses, as well as HGT have contributed to the evolution of PKS I in bacteria. The impact of these factors seems to vary considerably between the bacterial groups (Jenke-Kodama et al., 2005). Reconstruction of phylogenies from sequences with known structures involves less uncertainty and is therefore expected to be more accurate than reconstruction of phylogenies from sequences with unknown structures (Jeffrey et al., 1996).

Genomic location

According to these inferences, the function of this hypothetical protein was assigned to be β-KA synthase I that was located in chromosome I at the genomic position of 2249476-2249476 in this bacterium. It has 1269 bp in length (58.23% G+C and 41.77% A+T contents) encoded protein with 423 amino acid resides. Moreover, as there were no family related genes adjacent (neighboring) or flanking this gene, the existence of PKS related operon or gene cluster would be absent. This suggests that a possible way for occurrence of this gene in the genome of *P.luminescens* subsp. l*aumondii* TT01 as a consequence of HGT from soil symbionts. The codon usages and a proposed gene architecture are depicted Figure 5. ATG (methionine) served as a start codon for transcription of this gene, but there were no transcription regulatory elements such as promoter and transcription binding sites. Therefore, adopting promoter sites will hopefully helpful to express this gene in an expression vector.

Conclusion

Overall, β-ketoacyl synthase I encoded gene identified and annotated in this bacterium by bioinformatics database and tools. The function of this protein is predicted and ensured for biological reliability through structural data and phylogenetic approach. Furthermore, this is first report to identify ketoacyl synthase I function from this genome. This study provides a new vision to understand the mechanism involved in biosynthesis of antimicrobial and antinematicidal metabolites in this bacterium, and to enlighten the capability of producing ketoacyl synthase in nematode host that are used to inhibit consumption of the insect cadaver by competing organisms. However, experimental investigation of these systems has the potential of opening new vistas in the understanding of the crucial processes of such metabolites synthesis and regulation.

Acknowledgment

This study was supported by a grant from the DBT (Department of Biotechnology)-Young investigator, New Delhi, India.

Refernces

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al.(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search program. Nucleic Acids Res 25: 3389-3402. » CrossRef » PubMed » Google Scholar
2. Apic G, Gough J, Teichmann SA (2001) Domain combinations in archaeal,

eubacterial and eukaryotic proteomes. J Mol Biol 310: 311-325. » CrossRef » PubMed » Google Scholar

3. Binkowski TA, Naghibzadeg S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 31: 3352-3355. » CrossRef » PubMed » Google Scholar
4. Bode HB, Muller R (2005) The impact of bacterial genomics on natural product research. Angewandte Chemie 44: 6828-6846. » CrossRef » PubMed » Google Scholar
5. Brachmann AO, Joyce SA, Jenke-Kodama H, Schwar G, Clarke DJ (2007) A type II polyketide synthase is responsible for anthraquinone biosynthesis in *Photorhabdus luminescens*. ChemBioChem 8: 1721-1728. » CrossRef » PubMed » Google Scholar
6. Bradley E, Elizabeth H, Susan H (1996) Bootstrap confidence levels for phylogenetic trees. Proc Nat Acad Sci 93: 13429-13429. » CrossRef » PubMed » Google Scholar
7. Brylinski M, Konieczny L, Roterman I (2006) Ligation site in proteins recognized *in silico*. Bioinformatics 1: 127-129. » CrossRef » PubMed » Google Scholar
8. Dittmann E, Neilan BA, Erhard M, Von Dohren H, Borner T (1997) Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Mol Microbiol 26: 779-787. » CrossRef » PubMed » Google Scholar
9. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252: 675-679. » CrossRef » PubMed » Google Scholar
10. Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res 8: 163-167. » CrossRef » PubMed » Google Scholar
11. Ernst JD (2000) Bacterial inhibition of phagocytosis. Cell Microbiol 2: 379-386. » CrossRef » PubMed » Google Scholar
12. Ffrench-Constant RH, Waterfield N, Burland V, Perna N, Daborn PJ (2000) A genomic sample sequence of the entomopathogenic bacterium *Photorhabdus luminescens* W14: potential implications for virulence. Appl Environ Microbiol 66: 3310-3329. » CrossRef » PubMed » Google Scholar
13. Forst S, Nealson K (1996) Molecular biology of the symbiotic pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp. Microbiol Rev 60: 21-43. » CrossRef » PubMed » Google Scholar
14. Gascuel O, Steel M (2006) Neighbor-joining revealed. Mol Biol Evol 23: 1997-2000. » CrossRef » PubMed » Google Scholar
15. Geourjon C, Deleage G (1995) SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments. Comp Applications Biosci 11:681-684. » CrossRef » PubMed » Google Scholar
16. Gerth K, Bedorf N, Irschik H, Hofle G, Reichenbach H (1994) The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiotics 47: 23-31. » CrossRef » PubMed » Google Scholar
17. Ginolhac A, Jarrin C, Robe P, Perriere G, Vogel TM, et al. (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. Journal of Mol Evol 60: 716-725. » CrossRef » PubMed » Google Scholar
18. Hoffmann R, Valencia A (2004) A Gene Network for Navigating the Literature. Nature Genet 36: 664-664. » CrossRef » PubMed » Google Scholar
19. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16: 566-567. » CrossRef » PubMed » Google Scholar
20. Hopwood DA, Sherman DH (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annual Rev Genet 24: 37-66. » CrossRef » PubMed » Google Scholar
21. Hranueli D, Peric N, Borovicka B, Bogdan S, Cullum J, et al. (2001) Molecular biology of polyketide biosynthesis. Food Technol Biotechnol 39: 203-213. » CrossRef » PubMed » Google Scholar
22. Jeffrey LT, Nick G, David TJ (1996) Combining protein evolution and secondary structure. Mol Biol Evol 13: 666-673. » CrossRef » PubMed » Google Scholar
23. Jenke-Kodama H, Sandmann A, Muller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol 22: 2027-2039. » CrossRef » PubMed » Google Scholar
24. Jones DT, Taylar WR,Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comp Applications Biosci 8: 275-282. » CrossRef » PubMed » Google Scholar
25. Kealey J, Liu L, Santi DV, Betlach M, Barr PJ (1998) Production of a polyketide natural product in non-polyketide producing prokaryotic and eukaryotic hosts. Proc Nat Acad sci 95: 505-509. » CrossRef » PubMed » Google Scholar
26. Kumar S (1996) A stepwise algorithm for finding minimum evolution trees. Molecular Biol Evol 13: 584-593. » CrossRef » PubMed » Google Scholar
27. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, et al. (2005) ConSurf 2005: Presenting the evolutionary rate of amino acid sites on protein structures. Nucleic Acids Res 33: W299-W302. » CrossRef » PubMed » Google Scholar
28. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33: W89-W93. » CrossRef » PubMed » Google Scholar
29. Lavine MD, Strand MR (2002) Insect hemocytes and their role in immunity. Insect Biochem Mol Biol 32: 1295-1309. » CrossRef » PubMed » Google Scholar
30. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7: 1884-1897. » CrossRef » PubMed » Google Scholar
31. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, et al. (2005) CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res 33: 192-196. » CrossRef » PubMed » Google Scholar
32. Metsa-Ketela M, Halo L, Munukka E, Hakala J, Mantsala P, et al. (2002) Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. Appl Environ Microbiol 68: 4472-4479. » CrossRef » PubMed » Google Scholar
33. Milanesi L, D'Angelo D, Rogozin IB (1999) Gene Builder: interactive in silico prediction of genes structure. Bioinformatics 15: 612-621. » CrossRef » PubMed » Google Scholar
34. Mizuguchi K, Go N (1995) Comparison of spatial arrangements of the secondary structure elements in proteins. Protein Eng 8: 353-362. » CrossRef » PubMed » Google Scholar
35. Nick G, Jeffrey LT, David TJ (1998) Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genet 149: 445-458. » CrossRef » PubMed » Google Scholar
36. Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. Methods Mol Biol pp 365-386. » CrossRef » PubMed » Google Scholar
37. Sandhu SK, Jagdale GB, Hogenhout SA, Grewal PS (2006) Comparative analysis of the expressed genome of the infective juvenile entomopathogenic nematode, *Heterorhabditis bacteriophora*. Mol Biochem Parasitol 145: 239-244. » CrossRef » PubMed » Google Scholar
38. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31: 3381-3385. » CrossRef » PubMed » Google Scholar
39. Scott K, Louise GN, John WT, Yoder OC, Gillian TB (2003) Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Nat Acad sci 100:15670-15675. » CrossRef » PubMed » Google Scholar
40. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599. » CrossRef » PubMed » Google Scholar

41. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24: 4876-4882. » CrossRef » PubMed » Google Scholar

42. Williams JS, Thomas M, Clarke DJ (2005) The gene *stlA* encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in *Photorhabdus luminescens* TT01. Microbiol 151: 2543-2550. » CrossRef » PubMed » Google Scholar

43. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analysis. Tree 11: 367-372. » CrossRef » PubMed » Google Scholar