ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 7 (4): 660-668, 2008
ISSN 1682-296X
© 2008 Asian Network for Scientific Information

Phylogenetic Analysis of Type I and Type II Polyketide Synthase from Tropical Forest Soil

Mei-Fong Pang, Geok-Yuan Annie Tan, Noorlidah Abdullah,
Choon-Weng Lee and Ching-Ching Ng
Institute of Biological Sciences, Faculty of Science, University of Malaya,
50603 Kuala Lumpur, Malaysia

Abstract: Culture-independent approach was employed to retrieve diverse type I and type II polyketide synthase (PKS) ketosynthase (KS) domains from community DNA extracted from forest topsoil. Type I KS domains detected were from four phyla which comprised Cyanobacteria, Proteobacteria, Actinobacteria, Chloroflexi and uncultured bacteria. The type II KS_{α} domains were derived from four suborders of actinobacteria in addition to uncultured bacteria. Approximately 25% of the KS domains were recovered from uncultured bacteria and many sequences of type I KS were derived from Myxobacteria. BLASTP results showed that the type I and type II KS domain amino acid sequences were between 52 to 93% identical to the comparative sequences in the GenBank database. Phylogenetic analysis for novelty prediction of KS domains showed 14 and 9 novel clades of type I KS and type II PKS KS_{α} domains, respectively. These phylogenetically distinct novel clades might represent a new subclass of KS domains. Present data suggests the possibility of further discovery of novel genes encoding bioactive compounds which may have medical and pharmaceutical value.

Key words: Culture-independent, tropical soil, polyketide synthase (PKS), ketosynthase (KS), phylogenetic analysis

INTRODUCTION

It has been estimated that there are approximately 2.6×10^{29} prokaryotes residing in soil (Whitman *et al*., 1998). Yet, only a small fraction of them (0.1 to 1%) are cultivable with current techniques (Torsvik *et al*., 1990). Most of the microbial diversity has not been revealed due to the limitation of cultivation (Hugenholtz and Pace, 1996). Since most prokaryotic microbes are known to produce pharmaceutically important secondary metabolites through polyketide synthase (PKS) pathways (Hertweck *et al*., 2007; Staunton and Weissman, 2001) this massive uncultured community is a large genetic reservoir that contains numerous promising source of novel polyketide chemical structures (Pettit, 2004). Diversity of novel polyketide chemical structure could be utilized to curb the emerging resistance to existing antibiotics among infectious pathogens.

Polyketides are secondary metabolites which possess pharmacologically important activities such as antimicrobial, antifungal, antiparasitic, antitumor and agrochemical properties (Metsa-Ketela *et al*., 1999; Staunton and Weissman, 2001). To date, three different classes of PKS genes have been discovered (Shen, 2003). Type I PKS consists of large, multi-domain and highly modular proteins that produce polyketides by successive condensation of simple carboxylic acid units. Type I PKS encodes for macrolides such as erythromycin (antibiotic) and rapamycin (immunosuppressant) which have an indispensable role in medical treatment (Staunton and Weissman, 2001). Type II PKS comprises of aggregates of mono-functional proteins which catalyze formation of compounds through aromatization and cyclization. Doxorubicin (anticancer) and tetracyclines (antibiotic) are among the products synthesized by the type II PKS biosynthesis pathways (Hertweck *et al*., 2007). Type III PKSs is chalcone synthase-like PKS that is involved in synthesis of chalcones (CHS) and stilbenes in plants and polyhydroxy phenols in bacteria (Hutchinson, 1998). Chalcone and stilbenes are plant-specific PKS which

Corresponding Author: Dr. Ching-Ching Ng, Institute of Biological Sciences, Faculty of Science, University of Malaysia, 50603 Kuala Lumpur, Malaysia
Tel: +60-(0) 3-79675872 Fax: +60-(0) 3-79675908

perform complex reactions to produce an enzyme-bound tetraketide intermediate (Hopwood and Sherman, 1990). These tetraketide intermediates will fold to form a new aromatic ring system (Hopwood and Sherman, 1990). Type I PKS is modular. It has been shown that genetic engineering of novel polyketides can be achieved by combinatory biosynthesis (Weissman and Leadlay, 2005). By manipulating PKS catalytic domains within PKS I modules, hybrid PKS system can be generated (Kittendorf and Sherman, 2006). These hybrid PKS systems, with successful manipulation, will produce hybrid polyketides which could potentially become drug candidates (Gokhale *et al*., 1999; McDaniel *et al*., 1999). Therefore, accessing the PKS diversity in an environment is important for drug discovery and combinatorial biosynthesis of novel PKS (Moffitt and Neilan, 2003).

Numerous screening efforts have been carried out over the years in order to accelerate the rate of discovering novel pharmaceutically important polyketides. However, these efforts have been hindered by the limitations of cultivation because not all bacteria can be grown in standard laboratory conditions (Hugenholtz and Pace, 1996). If the diversity of chemistry produced by cultured bacteria is an indicator of the chemical capacity of these uncultured bacteria, then many biocatalysts and perhaps pharmaceutically important drugs, remain to be discovered from soil microorganisms (Gillespie *et al*., 2002).

The purpose of this study was to investigate and identify novel type I and type II PKS KS domains of tropical forest topsoil using a culture-independent method. Phylogenetic analysis of the sequences can be used to predict the novelty of the polyketide from the identification of the KS domains (Ginolhac *et al*., 2004; Wawrik *et al*., 2005). These novelty predictions are crucial in detecting novel polyketides and can provide invaluable information on combinatorial biosynthesis to design novel drugs (Moffitt and Neilan, 2003).

MATERIALS AND METHODS

DNA extraction from soil: Forest topsoil was collected from Rimba Ilmu, University of Malaya, Malaysia. Rimba Ilmu is one of the most important biological conservatories in Malaysia. It is a secondary tropical rain forest and contains living collections of over 1600 species. Soil samples were sieved through a 2 mm mesh to remove roots and large particles. A soil sample (20 g) was suspended in 50 mL of DNA extraction buffer (Yeates *et al*., 1998). Sample was further incubated at 65°C with sodium dodecyl sulfate (2 mL, 20%) and proteinase K treatment for 2 h with occasional gentle inversion. After removing the soil residues, supernatant was precipitated by using half-volume of polyethylene glycol (30%)/sodium chloride (1.6 M) and further incubated at room temperature for 2 h. The extracted DNA was resuspended in 20 mL of Tris-EDTA (10 mM Tris-1 mM EDTA pH 8.0) buffer before being further purified by equal volume of phenol/chloroform/isoamyl alcohol (25:24:1). Precipitated DNA was washed using 70% alcohol and dissolved in 1 mL TE buffer.

PCR amplification of 16S rRNA genes; conserved region of type I and type II PKS KS domain: To gain direct access to the type I and type II PKS gene diversity from forest soil, conserved region of KS domain of both PKS genes were amplified from the forest soil metagenome and cloned into plasmid vector. PCR reactions (total volume 50 μL) contained 1 μL of DNA template (5 ng), 5 μL of 10X PCR buffer, 1 μL of 10 mM dNTPs mix, 1 Unit of DyNAzyme™ 2 DNA Polymerase (Finnzymes Oy, Finland) and 5% of Dimethyl Sulfoxide (Sigma-Aldrich Inc. USA). PCR amplification was performed with a GeneAmp PCR System 2400 (Perkin-Elmer Corporation, USA) and the PCR parameters as shown in Table 1.

Generation of PCR clones libraries of type I and type II PKS KS domain: PCR products were gel purified by using GeneClean kit (Qbiogene, USA) and

Table 1: PCR primers and PCR conditions used in this study

Target gene	PCR primers	PCR conditions	PCR product (bp)	Reference
Type I PKS KS domain	KSLF: 5'-CCSCAGSAGCGCSTSYTSCTSGA-3' KSLR: 5'-GTSCCSGTSCCGTGSGYSTCSA -3'	1 cycle of 95°C 5 for min 40 cycles of 95°C for 1 min 58°C for 1 min 72°C for 1 min 1 cycle 72°C for 15 min	700	Ginolhac *et al*. (2004)
Actinomycetes type II PKS KSα domain	540F: 5'-GGITGCACSTCIGGIMTSGAC-3' 1100R: 5'-CCGATSGCICCSAGIGAGTG-3'	1 cycle of 95°C for 5 min 40 cycles of 95°C for 1 min 60°C for 1 min 72°C for 1 min 1 cycle of 72°C for 15 min	550	Wawrik *et al*. (2005)

cloned into pGEM-T Easy Vector (Promega, USA). Positive clones with inserts were identified with blue-white selection on X-Gal (5-bromo-4-chloro-3-indolylb-D-galactopyranoside)/IPTG (isopropyl-b-D-thiogalactopyranoside) (LB)/ampicilin Luria-Bertani (LB) plate. The forest soil type I and type II PKS KS domain PCR clones libraries were named LibKS1 and LibKS2. Clones were subjected to colony PCR by using M13 primers, M13F (5'-GGTTTTCCCAGTCACGAC-3') and M13R (5'-AGCGGATAACAATTTCACAC-3'). The colony PCR conditions were as follows: initial denaturation for 10 min at 94°C followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 45 sec, extension at 72°C for 1.5 min and final extension at 72°C for 5 min. The resulting PCR products were analyzed by Restriction Fragment Length Polymorphism (RFLP) with *Alu*I and *Hae*III (NEB, US). Clones with distinct RFLP patterns were subjected to DNA sequencing (Solgene, Korea).

Phylogenetic analysis: The vector sequences from the DNA sequencing results were removed. Type I and type II PKS KS conserved domain sequences were first translated into amino acid sequences before being compared to the NCBI protein database using BLASTP. The sequences were aligned using ClustalW packaged in MEGA (Molecular Evolutionary Genetics Analysis) version 4.0 (Tamura *et al*., 2007). Phylogenetic analysis was performed using the MEGA 4.0. Phylogenetic reconstruction for type I and type II PKS KS domains were performed by using Neighbor-Joining (NJ) and the PAM matrix. Bootstrap test of 1000 replicates were performed to determine confidence level.

Nucleotide accession numbers: DNA sequences obtained in this study were submitted in Gen Bank under the accession numbers EU445114 to EU445191.

RESULTS

Phylogenetic analysis of type I PKS KS domain: Amino acid sequence analysis revealed that the type I PKS KS domains of 38 clones with different RFLP patterns from LibKS1 were affiliated with members of phyla Cyanobacteria, Proteobacteria, Actinobacteria, Chloroflexi and uncultured bacteria. BLASTP of the LibKS1 KS domain showed low homology to the closest match in GenBank database. The amino acid sequences of KS domain were 51 to 73% similar to comparative sequences in the database.

Phylogenetic analysis demonstrated that the KS domains of clusters A1, A2, A3 and A4 from LibKS1 were clustered with KS domains of Myxobacteria *Polyangium cellulosum* and Actinobacteria *Streptomyces* sp., with high bootstrap values of 99 and 93%, respectively (Fig. 1). Two KS domains sequences, clones K307 and K151(2) were closely related to the KS domain of non-ribosomal peptide synthetase (NRPS) of *Myxococcus xanthus*. Another two KS domain sequences, clones K182 and K11 which were more related with KS domain of type I PKS of *M. xanthus* were placed in another clade. KS domains in cluster A7 were found to be grouped together within the cyanobacterial KS cluster. Clone K44(2) was found placed within a clade with probable peptide synthetase protein of *Ralstonia solanacearum* and the KS domain of an uncultured bacterial symbiont but this topology was not supported by a high bootstrap value. Phylogenetic analysis also showed KS domains in clusters A9, A10, A11 and A12 were clustered with KS domain sequences of uncultured bacteria (>83% bootstrap value) while clones K60(2) and K165 showed affiliation with Actinobacteria and Myxobacteria. The KS domain sequences in clusters A13 and A14 were grouped in a distinct clade with KS domains from members of Chloroflexi and Cyanobacteria with a high bootstrap value of 100%.

The KS domains of LibKS1 were found to be structurally distinct and displayed three types of conserved patterns of the cysteine active site. The first type of conserved pattern belonged to hybrids between non-ribosomal peptide synthetase (NPRS)/PKS I (Moffitt and Neilan, 2003). This conserved pattern was exhibited in 9 clones: K323, K161, K59(2), K39(2), K186, K12(2), K48(2), K7 and K8 from clusters A13 and A14. These KS domains contained consensus sequence N(D/E)KD at 22 amino acid upstream from the cysteine active site in the KS domain. Their conserved cysteine active site V**D**TACS**S**S was replaced by (V/I/L)**Q**TACS**T**S (boldface type indicated amino acid changes). The second type of conserved pattern, V**D**TACS**A**S was found in clone K13(2) in which the conserved serine was substituted by alanine. The last and most abundant type of conserved pattern (M/I/L/V)**D**TACS**S**S, was found in the remaining 28 clones in LibKS1.

Phylogenetic analysis of type II PKS KS domain: Sequence analysis revealed that the actinomycetes type II PKS KS_{α} domains of LibKS2 were derived from members of the suborders *Streptomycineae*, *Pseudonocardineae*, *Micromonosporineae* and *Frankineae*. According to

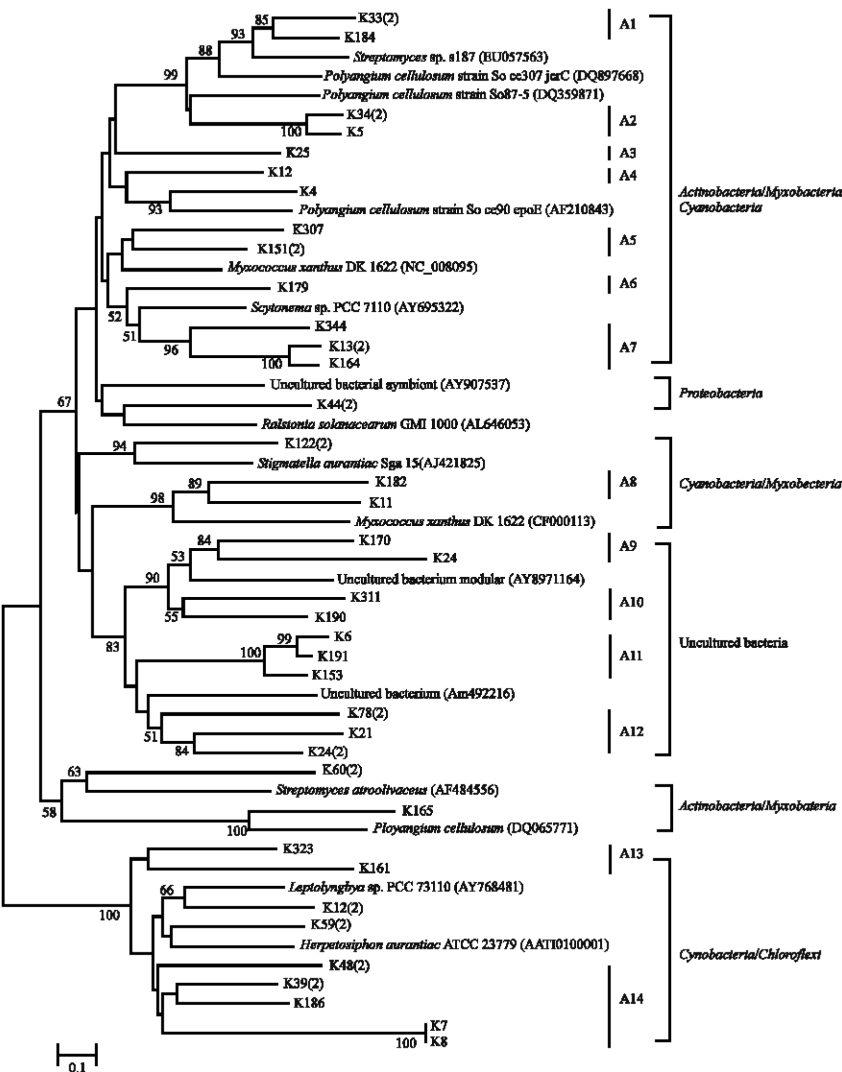

Fig. 1: Phylogenetic analysis of type I PKS KS domain from selected clones in LibKS1. Only bootstraps value more than 50% bootstrap value are shown

BLASTP results, the KS domain amino acid sequences of LibKS2 were 71 to 93% identical to known sequences in the GenBank database.

Most of the LibKS2 KS_{α} domain resembled the KS domain of members of *Streptomycineae*. KS_{α} domain in cluster B1 formed a distinct clade (as shown in Fig. 2) in the phylogenetic analysis but they were linked with *Streptomyces* sp. AM-7161. KS domains of cluster B2 from LibKS2 were also placed within the same clade. Both clades were supported by a high bootstrap value of 100%. One group of KS_{α} in cluster B3 from LibKS2 was clustered with the KS_{α} domain of *Kibdelosporangium aridum*. Phylogenetic analysis revealed that clones P137(2), P127(2), P130, P139, P28 and P111(2) were affiliated with beta-ketoacyl synthase of *Streptomyces cyaneus*. The relationship was supported by bootstrap value of 69%. KS_{α} domains in cluster B4 were placed within the clade of KS_{α} domains of uncultured bacteria but were related to *K. aridum* (>96% bootstrap value) while the KS_{α} domain sequence of clone P7 was shown to be related to that of *Streptomyces* sp. Phylogenetic analysis also demonstrated that the KS_{α} domains from the suborder of *Micromonosporineae* and *Frankineae* were clustered in the same clade. KS_{α} domain sequence of clone P193 was related with *Salinispora arenicola* whereas those in clusters B8 and B9 were related with *Frankia* sp.

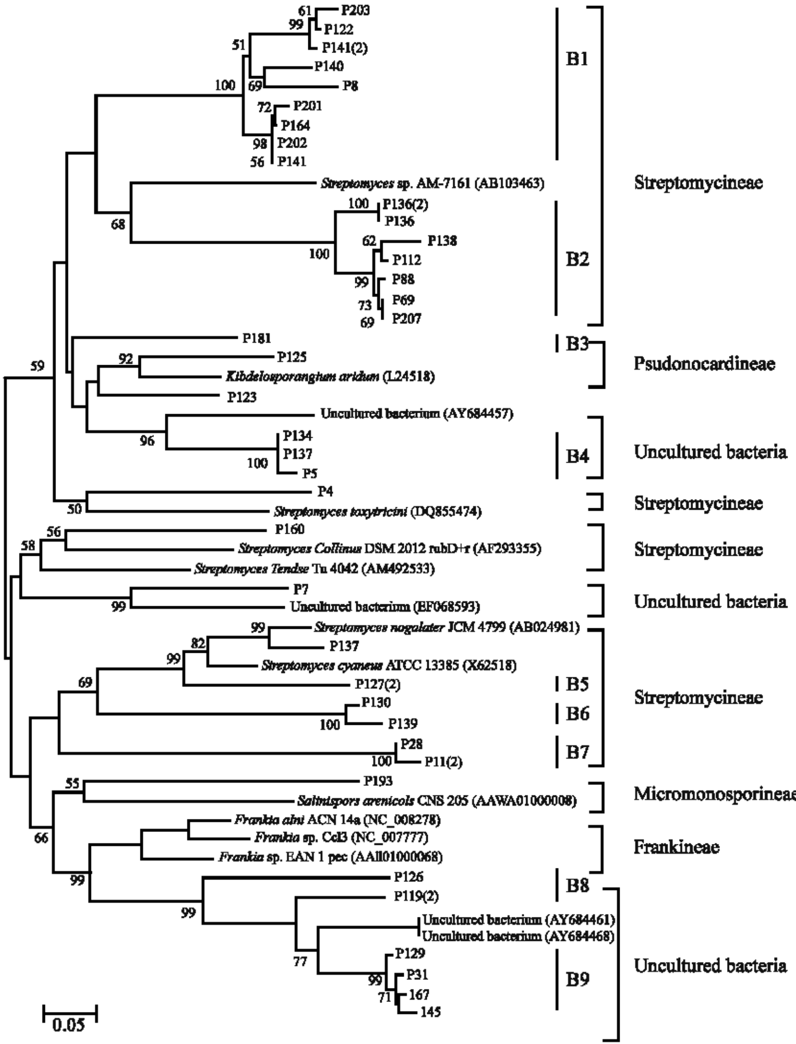

Fig. 2: Phylogenetic analysis of type II PKS KS $_{\alpha}$ domain from selected clones in LibKS2. Only bootstraps value more than 50% bootstrap value are shown. *B represents cluster

DISCUSSION

Phylogenetic analysis of KS domains in this study revealed that sequences were derived from the members of Actinobacteria, Proteobacteria, Cyanobacteria, Chloroflexi and uncultured bacteria. We discovered approximately 25% of KS domains were derived from uncultured bacteria and a large portion of KS of PKS I were derived from Myxobacteria. Myxobacteria belongs to members of subphylum Deltaproteobacteria and is known to produce structurally diverse secondary metabolites (Janssen, 2006).

From the phylogenetic analysis, KS domains of LibKS1 were divided into two types. The first type of the LibKS1 KS domains consisted of type I KS domain (Fig. 3). The second type of KS domains found in LibKS1 are KS domains specific for amino acid starter units in hybrid NRPS/PKS systems (Fig. 4). KS domains in clusters A1, A2, A3, A4, A5, A6, A7 and cluster A8 from LibKS1 were related with the KS domains of the Myxobacteria *Polyangium* KS of type I PKS. *Polyangium* strains are prolific producers of secondary metabolites (Gerth *et al*., 2003). These KS domains also showed relationship with the KS domains of Actinobacteria, Myxobacteria and Cyanobacteria but evolved independently. Clustering of cyanobacterial and myxobacterial KS domains is already well-known (Jenke-Kodama *et al*., 2005; Moffitt and Neilan, 2003). It was interesting to note that there was mixed clustering of Myxobacterial, Actinobacterial and Cyanobacterial KS

K323	--NKD---IQTACSTSLVAV
N. aurantiacus (NZ ATTI0100001)	--NKD---VQSACSTSLVAT
K39 (2)	--NKD---VQTACSTSLVAV
K-186	--NKD---IQTACSTSLVAV
L. *sp.* PCC 73110 (AY76841)	--NKD---IQTACSTSLVAT
K48 (2)	--NKD---VQTACSTSLVAV
K12 (2)	--NKD---VQTACSTSLVAV
K59 (2)	--NKD---VQSACSTSLVAT
K161	--NKD---VQSACSTSLAAV
K7	--NKD---LNTACSTSLTNI
K8	--NKD---LNTACSTSLTNI

Fig. 3: Alignment of the active sites of type I KS domains with distinct functions. Conserved cysteine (C) site is shown

K33	IDTACSSSLSAL
K183	VDTACSSSLTAL
Streptomycse sp. 3187 (EU057563)	VDTACSSSLVAL
S. cellusum So ce307 JerC (DQ897668)	VDTACSSSLVSL
K34 (2)	VDTGCSSSLVSL
K5	VETGCSSSLVSL
S. cellusum So87-5 (DQ359871)	VDTACSSSLVAV
K44 (2)	IDTACSSSLVAM
K12	VDTACSSSLVAV
K4	LDTACSSSLAAL
S. cellusum So ce90 epoE (AF210843)	VDTACSSSLVAI
K25	IDTACSSSLYAL
Uncultured bacterial smybiont (AY907537)	VDTACSSSLVAV
K307	IDTACSSSLVAI
K151 (2)	IDTACSSSLVAV
Nyxococcus xanthus DK1622 (NC_008095)	VDTACSSSLVSL
K179	IDTACSSSLVSV
K344	IDTACSSSLAAV
Scytinema sp. (AY695322)	IDTACSSSLVSV
K13 (2)	VDTACSASLVSV
K164	VDTACSSSLVSV
K182	VDTACSSSLVAT
K11	VDTACSSSLVAT
Nyxoccus xanthus (CP000113)	LDTACSSSMVAV
K311	IDTACSSALVSV
K190	VDTACSSSLVAV
Uncultured bacterium (AY897164)	VDTACSSSLVAL
K170	VDTACSSSLVAL
K24	MDTACSSSLVAL
K6	MDTACSSGLYAV
K191	MDTACSSALTAV
K153	MDTACSSSLTAV
Ralstonia solanacearum (AL 646053)	IDTACSSSLVAV
K78 (2)	LDTACSSSLVAV
K21	VDTACSSALVAV
K24 (2)	VDTACSSALVAV
Uncultured bacterium (AM492216)	VDTACSSSLVAV
K122 (2)	LDTACSSSLVAI
Stigmatella aurantiaca (AJ421825)	VNTACSSSLVAV
K60 (2)	IDTACSSGLVAL
Streptomyce atroovaceus (AF484556)	VDTACSSSLVAI
K165	VDTACSSSLTAI
Polyangium cellium (DQ065771)	MDTACSAMTVAI

Fig. 4: Alignment of type I KS domains in mixed or hybrid systems

domains. This finding is in concordance of the findings of Li *et al*. (2007) which also revealed that type I KS domains of *Sorangium* clustered together with myxobacterial, actinobacterial and cyanobacterial KS domains. Two KS

domains, clones K182 and K11 (A8) (GenBank accession no. EU445178 and EU445179) were clustered with the KS domain of *M. xanthus*. These KS domains also showed close relationship with the KS domains of uncultured bacteria suggesting that these KS sequences (clusters A9, A10, A11, A12) might encode for novel macrolide-like antibiotics. Although macrolide-like antibiotics are known to be produced by members of the genus *Streptomyces* (Xue *et al*., 1998), studies have shown that the nucleotide and organizational level of the PKS biosynthesis genes of the Actinobacteria and Myxobacteria share striking similarity (Varon *et al*., 1992). Both of the Actinobacteria and Myxobacteria have high G+C content in their genome (>65%), suggesting that genetic exchange e.g., horizontal gene transfer (HGT) could have occurred between Actinobacteria and Myxobacteria (Ginolhac *et al*., 2005; Schupp *et al*., 1995).

KS domains specific for amino acid starter units in hybrid NRPS/PKS systems in LibKS1 formed a single cluster. It was interesting to note that these KS domains of uncultured bacteria were clustered with the KS domains of Cyanobacteria *Leptolyngbya* sp. and Chloroflexi *Herpetosiphon* sp. with bootstrap value of 100%. Siphonazole is a new structural class and unusual secondary metabolites produced by *Herpetosiphon* sp. through the hybrid NRPS/PKS systems (Nett *et al*., 2006). The relationship between clustering of these KS domains suggested to us that the uncultured bacteria KS, clones K7 and K8 might contain genes that encoded siphonazole class polyketides. With between 56 to 63% of homology to known KS in database, it is tempting to suggest KS domains from A14 could belong to a novel group of KS domains and could possibly produce novel siphonazole polyketides.

The consensus sequence of the cysteine active sites of type I KS domain was represented by V**D**TACS**S**S (boldface type indicated conserved residues) (Moffitt and Neilan, 2003). Residue D (aspartic acid) and S (serine) were highly conserved in type I KS. Analysis of the KS domain of LibKS1 revealed that most of them displayed type I KS. However, the valine (5) residue in the consensus KS was often replaced by residue methionine (M), isoleucine (I) or leucine (L) in the type I KS although the conserved residues D and S remained unchanged. KS of K13(2) (GenBank accession No. EU445166) exhibited unique amino acid sequence in the cysteine active site (VDTACSAS) in which the serine was substituted by alanine. This conserved pattern of cysteine active site was also exhibited by PKS of *Polyangium cellulosum*. As for the hybrid NRPS/PKS I KS domains in LibKS1, all of the KS domains displayed **Q**TACS**T**S in the cysteine active site. Conserved residue D and S were replaced by glutamine (Q) and threonine (T). The valine (V) residue in the consensus KS could be replaced by residue I or L. All of the KS domains analysed did not show cysteine active site mutation specific in which the active cysteine was substituted by glutamine (Q) (KS^Q). KS^Q will lead to loss of condensation activity of KS domain but still have the decarboxylative activity of ACP-bound dicarbocylic acid, giving rise to the starter moiety (Weinig *et al*., 2003). This indicated all of the KS domains analysed in this study retained the condensation activity of the extender unit to the growing polyketide chain (Moffitt and Neilan, 2003).

Both KS_α of B1 and B2 showed affiliation with medermycin-producing *Streptomyces* sp. AM-7161. Medermycin is a class of aromatic antibiotics known as benzoisochromanequinones isolated from streptomycetes and is active against Gram-positive bacteria (Ichinose *et al*., 2003). The KS_α sequences from B1 and B2 which showed similarity ranging from 72-79% to known KS_α sequences in the database indicate that these KS domains are yet-to-be isolated. The KS domain sequences of clones P125 and P123 were closely related with that of *K. aridum* (supported by bootstrap of 92%). KS domain of P181 showed similarity with *K. aridum* while KS domains of uncultured bacteria from B4 (85% identical) were associated with KS_α of *K. aridum* with 96% of bootstrap value. These KS_α domains were phylogenetically distinct from the KS_α of *K. aridum*, suggesting that molecules of B4 could be chemically unique and structurally different.

KS_α of P137(2) was most similar to KS_α of biosynthetic pathway of the anthracycline antibiotic nogalamycin of *Streptomyces nogalater*. Nogalamycin is a potent cytostatic drugs that inhibits DNA-dependent RNA synthesis by binding to DNA. Nogalamycin possess cytostatic activity against tumor cells and have been used as antitumor drug (Bhuyan and Reusser, 1970). Sequences from B6 were also affiliated with KS_α of *S. nogalater*. Based on the phylogenetic analysis, KS_α domains of B6 would have further evolved independently. Sequences of B8 and B9 were placed in the same clade with the KS_α of uncultured bacteria. These KS_α domains were clustered together with the KS_α of *Frankia* sp. indicating that these KS_α shared similarity with the KS_α of *Frankia* sp. Members of *Frankia* are nitrogen fixing, symbiotic actinomycetes but produce antimicrobial compounds against soil microbes in order to facilitate their own survival (Sarma *et al*., 2003). Majority of the KS_α domains recovered from the tropical forest topsoil were derived from the suborder Streptomycineae and this suggests that members of Streptomycineae are dominant actinomycetes in LibKS2. Phylogenetic analysis revealed KS_α domains of Streptomycineae were divided into four

distinct genotypic groups which perhaps indicate their multi-ancestries or rapid evolutionary rates. It is thought that inherent genetic strategies coupled with high frequencies of HGT contributed to the significant diversity of Streptomycineae KS_{α}(Li *et al*., 2007; Lorenz and Wackernagel, 1994).

Present results showed that diverse set of type I and type II KS domains were recovered from the tropical forest topsoil. This suggested the culture-independent approach is a practical method for the discovery of novel and unique KS domains from the environment. It is also important to note that the diversity of the KS domains from this study does not necessarily represent the entire PKS domains in the tropical forest topsoil due to the bias in DNA extraction, PCR, cloning and number of clones studied.

Molecular phylogenetic analysis revealed 14 novel clades (A1-A14) of type I PKS KS domains and 9 novel clades (B1-B9) of type II PKS Ks_{α} domains. The phylogenetically distinct novel clades might represent a new subclass of KSs. Previous studies had shown that replacement of a domain within the PKS modular can lead to a functional complex which will synthesize a novel polyketide (Gokhale *et al*., 1999; McDaniel *et al*., 1999). Discovery of novel KS domains could facilitate combinatorial synthesis of novel PKS which will lead to the development of novel drugs. In addition to this, phylogenetic analysis of KS domains can also be used as marker or indicator for preliminary screening to estimate the possibilities of detecting novel PKS genes clusters from the soil metagenome. This is particularly important since the soil metagenome will become a platform for novel gene discovery.

ACKNOWLEDGMENTS

This study is funded by IRPA 09-02-02-008-BTK/ER-36 (Metagenomik) from the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. We thank MOSTI National Science Fellowship (NSF) for giving Mei-Fong, Pang a scholarship. We would also like to thank Dr. Chong-Lek Koh.

REFERENCES

Bhuyan, B.K. and F. Reusser, 1970. Comparative biological activity of nogalamycin and its analogs. Cancer Res., 30: 984-989.

Gerth, K., S. Pradella, O. Perlova, S. Beyer and R. Muller, 2003. Myxobacteria: Proficient producers of novel natural products with various biological activities-past and future biotechnological aspects with the focus on the genus Sorangium. J. Biotechnol., 106: 233-253.

Gillespie, D.E., S.F. Brady, A.D. Bettermann, N.P. Cianciotto and M.R. Liles *et al*., 2002. Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Applied Environ. Microbiol., 68: 4301-4306.

Ginolhac, A., C. Jarrin, B. Gillet, P. Robe and P. Pujic *et al*., 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Applied Environ. Microbiol., 70: 5522-5527.

Ginolhac, A., C. Jarrin, P. Robe, G. Perriere, T.M. Vogel, P. Simonet and R. Nalin, 2005. Type 1 polyketide synthases may have evolved through horizontal gene transfer. J. Mol. Evol., 60: 716-725.

Gokhale, R.S., S.Y. Tsuji, D.E. Cane and C. Khosla, 1999. Dissecting and exploiting intermodular communication in polyketide synthases. Science, 284: 482-485.

Hertweck, C., A. Luzhetskyy, Y. Rebets and A. Bechthold, 2007. Type 2 polyketide synthases: Gaining a deeper insight into enzymatic teamwork. Nat. Prod. Rep., 24: 162-190.

Hopwood, D.A. and D.H. Sherman, 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet., 24: 37-66.

Hugenholtz, P. and N.R. Pace, 1996. Identifying microbial diversity in the natural environment: A molecular phylogenetic approach. Trends Biotechnol., 14: 190-197.

Hutchinson, C.R., 1998. Combinatorial biosynthesis for new drug discovery. Curr. Opin. Microbiol., 1: 319-329.

Ichinose, K., M. Ozawa, K. Itou, K. Kunieda and Y. Ebizuka, 2003. Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: Towards comparative analysis of the benzoisochromanequinone gene clusters. Microbiology, 149: 1633-1645.

Janssen, P.H., 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Applied Environ. Microbiol., 72: 1719-1728.

Jenke-Kodama, H., A. Sandmann, R. Muller and E. Dittmann, 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol., 22: 2027-2039.

Kittendorf, J.D. and D.H. Sherman, 2006. Developing tools for engineering hybrid polyketide synthetic pathways. Curr. Opin. Biotechnol., 17: 597-605.

Li, Z.F., J.Y. Zhao, Z.J. Xia, J. Shi and H. Liu *et al*., 2007. Evolutionary diversity of ketoacyl synthases in cellulolytic myxobacterium Sorangium. Syst. Applied Microbiol., 30: 189-196.

Lorenz, M.G. and W. Wackernagel, 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev., 58: 563-602.

McDaniel, R., A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach and G. Ashley, 1999. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products. Proc. Natl. Acad. Sci. USA., 96: 1846-1851.

Metsa-Ketela, M., V. Salo, L. Halo, A. Hautala and J. Hakala *et al*., 1999. An efficient approach for screening minimal PKS genes from Streptomyces. FEMS Microbiol. Lett., 180: 1-6.

Moffitt, M.C. and B.A. Neilan, 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. J. Mol. Evol., 56: 446-457.

Nett, M., O. Erol, S. Kehraus, M. Kock and A. Krick *et al*., 2006. Siphonazole, an unusual metabolite from *Herpetosiphon* sp. Angew Chem. Int. Ed. Engl., 45: 3863-3867.

Pettit, R.K., 2004. Soil DNA libraries for anticancer drug discovery. Cancer Chemother. Pharmacol., 54: 1-6.

Sarma, H.K., B.K. Sharma and S.C. Tiwari, 2003. A novel calcimycin antibiotic from Gram-positive actinomycete Frankia microsymbiont. Curr. Sci., 85: 1401-1403.

Schupp, T., C. Toupet, B. Cluzel, S. Neff, S. Hill, J.J. Beck and J.M. Ligon, 1995. A *Sorangium cellulosum* (Myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: Cloning, characterization and homology to polyketide synthase genes from actinomycetes. J. Bacteriol., 177: 3673-3679.

Shen, B., 2003. Polyketide biosynthesis beyond the type 1, 2 and 3 polyketide synthase paradigms. Curr. Opin. Chem. Biol., 7: 285-295.

Staunton, J. and K.J. Weissman, 2001. Polyketide biosynthesis: A millennium review. Nat. Prod. Rep., 18: 380-416.

Tamura, K., J. Dudley, M. Nei and S. Kumar, 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol., 24: 1596-1599.

Torsvik, V., J. Goksoyr and F.L. Daae, 1990. High diversity in DNA of soil bacteria. Applied Environ. Microbiol., 56: 782-787.

Varon, M., N. Fuchs, M. Monosov, S. Tolchinsky and E. Rosenberg, 1992. Mutation and mapping of genes involved in production of the antibiotic TA in *Myxococcus xanthus.* Antimicrob. Agents Chemother., 36: 2316-2321.

Wawrik, B., L. Kerkhof, G.J. Zylstra and J.J. Kukor, 2005. Identification of unique type 2 polyketide synthase genes in soil. Applied Environ. Microbiol., 71: 2232-2238.

Weinig, S., H.J. Hecht, T. Mahmud and R. Muller, 2003. Melithiazol biosynthesis: Further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. Chem. Biol., 10: 939-952.

Weissman, K.J. and P.F. Leadlay, 2005. Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol., 3: 925-936.

Whitman, W.B., D.C. Coleman and W.J. Wiebe, 1998. Prokaryotes: The unseen majority. Proc. Nat. Acad Sci. USA., 95: 6578-6583.

Xue, Y., L. Zhao, H.W. Liu and D.H. Sherman, 1998. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity. Proc. Natl. Acad Sci. USA., 95: 12111-12116.

Yeates, C., M.R. Gillings, A.D. Davison, N. Altavilla and D.A. Veal, 1998. Methods for microbial DNA extraction from soil for PCR amplification. Biol. Proceed. Online, 1: 40-47.