Full Length Research Paper

Directly accessing the diversity of bacterial type I polyketide synthase gene in Chinese soil and seawater

Jing-Wen Song, Xiao-Yi Dong, Bing-Hua Jiao* and Liang-Hua Wang*

Department of Biochemistry and Molecular Biology, Faculty of Basic Medical Science, Second Military Medical University, Shanghai 200433, P. R. China.

Accepted 30 April, 2013

Many bacteria are known to produce a wide range of biomedically important secondary metabolites synthesized by type I polyketide synthases (PKSs). These enzymes own a modular structure that could be used for combinatorial biosynthesis to yield novel drug candidates. To directly access PKS gene diversity of soil and seawater in China, a set of degenerate oligonucleotide primers were designed to amplify the ketoacyl synthase (KS) fragments belonging to the PKS I genes from the soil and the seawater samples. Twenty-three (23) new KS fragments were obtained. Their predicted amino acid sequences showed 45 to 85% identities to the known KS domains in the GenBank. Phylogenetic analysis indicated that 14 of them belonged to the “normal” KS groups that catalyzed the condensation of the acyl groups and the other nine KS fragments belonged to the hybrid PKS/NRPS (non-ribosomal peptide synthase) groups which used an amino acid moiety as a starter unit. All the four KS fragments isolated from the seawater belonged to the former group. No significant difference was found between the soil KS fragments and the seawater KS fragments. Several KS fragments were endowed with some distinct characters that will be useful as probes for future studies.

Key words: Uncultured microorganisms, polyketide synthase, ketoacyl synthase, soil, seawater.

INTRODUCTION

The biosphere is dominated by microorganisms, which produce numerous secondary metabolites with various biological activities. Efforts to discover new bioactive molecules from microbes have lasted one century. However, the number of novel compounds discovered in recent years has not increased in proportion to the progress of culture-based screening methods (Lu et al., 2008; Firn and Jones, 2000), because only a small fraction of all microbes can be cultured by traditional methods. The knowledge about the “underexplored majority” is still poor (Sogin et al., 2006). To avoid culture limit, modern biological technology approaches have provided direct access to explore genes or gene clusters that are responsible for the synthesis of microbial secondary metabolites (Cowan et al., 2005; Carlson et al.,2010). Polyketide synthase (PKS) gene clusters are suitable to be the screening target. PKS synthesize polyketides, a large family of secondary metabolites that include many clinically important drugs such as erythromycin (antibacterial), epothilone (antitumor), soraphen (antifungal), rapamycin (immuno-suppressant) and lovastatin (antihypercholesterolemic) (Staunton and Weissman, 2001; Fisch et al., 2009). Among the three types of PKS that have been reported to date, only PKS I shows a modular organization, allowing the comparative study of modules within or between
enzymes. A minimal module is composed of a ketoacyl synthase (KS) domain; an acyltransferase (AT) domain and an acyl carrier protein (ACP), each module catalyzes the condensation of the acyl groups to the growing acyl chain. Frequently ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains are also embedded in the module to modify the growing acyl chain (Trindade-Silva et al., 2013; Shen et al., 2003). Using standard genetic engineering, enzymes were deleted or added, or their substrate specificity was turned in a particular manner, the numbers of novel polyketides potentially accessible by this route could be enormous (Mondol et al., 2011). Therefore recognition of the availability and diversity of PKS domains in the environment is very important for future drug discovery and combinatorial biosynthesis efforts (Woo et al., 2010). This study focus on the conserved KS modules belonging to the PKS I gene clusters.

KSs represent a superfamily of complex biosynthetic pathway-associated enzymes found in prokaryotes, fungi, and plants (Moffitt et al., 2003). Previous phylogenetic studies of KS domains in PKS I have revealed several distinct groups: "normal" KS domains, KS\(^2\) domains that provide a decarboxylative activity, KS domains in the modules following hybrid non-ribosomal peptide synthase (NRPS) modules and KS domains in the trans-AT modules (Cheng et al., 2003; Jenke-Kodama et al., 2005; Moffitt et al., 2003). KS-specific PCR has been used to screen different environmental samples and various laboratory bacterial or fungal strains. The obtained KS domains have been successfully used to detect the PKS gene clusters from unculturable bacteria which were participating in known or unknown polyketide biosynthesis pathway in recombinant clones from many metagenomic libraries include soil (Courtois et al., 2003; Ginolhac et al., 2004), beetles (Piel, 2002), sponges (Piel et al., 2004; Schirmer et al., 2005) and bryozoans (Lim and Haygood, 2004; Lopanik et al., 2006). On the other hand, KS-specific PCR methods have succeeded in the discovery of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction (Kim et al., 2006).

In this study, a set of degenerate oligonucleotide primers, designed for amplification of KS domains, had been employed to identify KS gene fragments from soil and seawater DNA samples. Our purpose was to develop a culture-independent method to directly access PKS I gene diversity in Chinese soil and seawater, because the related knowledge was incomplete. This would open up the possibilities of using the results for DNA fingerprints of secondary metabolites and as the basis for a search for attractive antibiotic biosynthesis genes that could be used in the heterogeneous expression and combinatorial biosynthesis. Furthermore, the amplified KS gene fragments can also be used as homologous hybridization probes to detect the clones harbored PKS gene clusters in the recombinant metagenomic libraries that would be constructed in the following researches.

**MATERIALS AND METHODS**

**Sampling sites**

The soil samples were obtained from a depth range of 5 to 20 cm in an area located 1 km southeast of the Yangshan Harbor (Shanghai, China) in November 26, 2005. Nearly 10 L seawater was obtained from a depth range of 1 to 3 m below the sea level near the coast of the Yangshan Harbor (Shanghai, China) in November 26, 2005. The seawater was first filtered with 0.8 μm cellulose acetate membrane, and the filtrate was put through another 0.15 μm cellulose acetate membrane, both the membranes were washed with sterilized water three times, and then the mixture on the membranes was collected in Eppendorf tubes and stored at -20°C. Isolations of DNA from these samples were performed within two weeks after collection.

**DNA extractions**

DNA extractions were carried out by using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Inc.). Compared with an adaptation of the procedure described by Zhou et al. (1996), the MO BIO kit produced DNA with a higher level of purity that can be used for the following PCR directly. DNA extractions of the soil and seawater samples were performed following the manufacturers' instructions. The DNA yields were calculated from the A260. The DNA was loaded on a 1.0% agarose gels and ethidium bromide staining to determine size and concentration. All extracted DNAs were stored at -20°C until use.

**PCRs**

With the help of The CODEHOP designer (http://blocks.fhcrc.org/codehop.html) and manually correction, a set of degenerate PCR primers were designed from conserved regions of KS domains of bacterial PKS I genes (Rose et al., 2003). The forward primers KSF (5'-CGC TTC ATG GAY CCS CAR CA-3') were based on the conserved motif SDPQQR. The reverse primers KSR (5'-GTC CGG GTG CCR TGS SHY TCS A-3') were based on the conserved motif HGTGT. The specificity of the primer set had been confirmed by testing with a collection of polyketide-producing strain (Streptomyces rimosus 8229, Streptomyces coelicolor ATCC101478, Streptomyces avermitilis, ATCC 31271) prior to the following PCR reactions with the environmental samples. The specific fragments amplified with KSF-KSR were about 700 bp in length. The 25 μl PCR mixture consisted of 2.5 μl 10×Ex Taq Buffer(Mg\(^{2+}\) Plus), 2 μl dNTP Mixture (2.5 mmol/L each), 0.125 μl TaKaRa Ex Taq(5U/μl), 1.5 μl DMSO(dimethyl sulfoxide), 0.25 μl each primer (20μmol/L), 1 μl template DNA (0.03 g/L), and 17.375 μl ddH2O. Thermal cycling was performed in a GeneAmp PCR System 2400 Thermocycle (Perkin Elmer). The initial denaturation step at 94°C for 5 min was followed by 35 cycles of DNA denaturation at 94°C for 1 min, primer annealing at 65°C for 1 min, and DNA strand extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. PCR products were analyzed by electrophoresis in 1% agarose gels and ethidium bromide staining.

**Cloning and Sequencing**

The result PCR products about 700bp were purified on agarose gels (mini-DNA rapid purification kit, BioDev-Tech) and then cloned
RESULTS AND DISCUSSION

DNA extraction

There were mainly two different extraction techniques that have been developed for actual research demand. The first methods (direct methods) were based on the in situ lysis of microbial cells in the samples prior to DNA recovery and purification. The second strategy (indirect methods) included a bacterial cells enrichment step prior to cell lysis and DNA extraction. The latter procedure could produce longer DNA that suit the construction of metagenomic libraries but may reduce the yield (Cowan et al., 2005). Considering the PCR requirements of this study, both a commercial Kit and a protocol first described by Zhou et al. (1996) that belong to the direct methods were compared in the DNA extractions, the MO BIO method includes lyses by bead beating resulted in more clear and higher purity which were supported by the following PCR reactions. All the three environmental DNA fragments extracted by the MO BIO Power Soil kit were larger than 21 kb and similar to the others (Figure 1). DNA extracted from the soil had the highest DNA yield (16.8 µg/g). DNA extracted from seawater had a very lower yield (0.4 µgDNA/liter seawater to 0.8 µm membrane sample and 0.1 µgDNA/liter seawater to 0.15 µm membrane sample).

PCRs

Though there were quite some differences in the DNA yield of the samples, the degenerate oligonucleotide primers set KSF/ KSR based on the DPQQQRLL and HGTGT motifs worked efficiently for these environmental samples. To improve the probability and efficiency of successfully amplifying, some of the forward primers and the reverse primers were combined one to one (the primers used in this study are shown in Table 1).

About twenty 700 bp fragments amplified from the soil sample and three from the seawater sample were purified from agarose gels and then cloned into a pMD18-T Vector. Approximately sixty positive recombinants were selected for the following sequencing. After excluding the identical clones, 23 unique nucleotide fragments (ranged from 630 bp to 690 bp after cutting off the primer and vector sequences) were obtained. Among them 19 clones were amplified from soil (DQ640993, DQ640997, DQ641926, DQ641927, and DQ673151) and 4 from the seawater (01529, 0848-1, 0848-2 and DQ673151). The PCR results are shown in Figure 2 and 3.

Interestingly, it had been observed that these isolated clones are different sequences. For example, both clones dz-ks-DG-1 and clone dz-ks-DG-3 were isolated from the PCR product of the primer pair KSF/DKSR. Similar instances happened frequently. In the case of the primer pair KSF/KSRH, which we called “super primer pair” (in soil DNA samples these primers amplified 4 different KS domains), five clones (dz-ks-A1, dz-ks-P1, dz-ks-DH-1, dz-ks-DH-2 and dz-ks-DH-3) using this primer pair were obtained from soil sample and three clones (01529-2, 0848-1, 0848-2) from seawater sample. This special primer
Table 1. Sequence of primers used in this study.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>sequences (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>KSFA</td>
<td>cgCTccATggAcccccAACa</td>
</tr>
<tr>
<td>KSFB</td>
<td>cgCTccATggAcccccAgcA</td>
</tr>
<tr>
<td>KSFC</td>
<td>cgCTccATggAccgcAccA</td>
</tr>
<tr>
<td>KSFD</td>
<td>cgCTccATggAccgcAccA</td>
</tr>
<tr>
<td>KSFE</td>
<td>cgCTccATggATcccccA</td>
</tr>
<tr>
<td>KSFF</td>
<td>cgCTccATggATcccccAgcA</td>
</tr>
<tr>
<td>KSFG</td>
<td>cgCTccATggATccgcAccA</td>
</tr>
<tr>
<td>KSFH</td>
<td>cgCTccATggATccgcAccA</td>
</tr>
<tr>
<td>KSRA</td>
<td>gTccggTccgTgcgccTccA</td>
</tr>
<tr>
<td>KSRB</td>
<td>gTccggTccgTgcgccTcga</td>
</tr>
<tr>
<td>KSRB</td>
<td>gTccggTccgTgcgcTTccA</td>
</tr>
<tr>
<td>KSRD</td>
<td>gTccggTccgTgcgcTTcga</td>
</tr>
<tr>
<td>KSRD</td>
<td>gTccggTccgTgcgcTTcga</td>
</tr>
<tr>
<td>KSRF</td>
<td>gTccggTccgTgcgccATcga</td>
</tr>
<tr>
<td>KSRG</td>
<td>gTccggTccgTgcgcAcccTccA</td>
</tr>
<tr>
<td>KSRH</td>
<td>gTccggTccgTgcgccTcga</td>
</tr>
<tr>
<td>KSRJ</td>
<td>gTccggTccgTgcgccTcga</td>
</tr>
<tr>
<td>KSRK</td>
<td>gTccggTccgTgcgccTcga</td>
</tr>
<tr>
<td>KSRM</td>
<td>gTccggTccgTgcgccATcga</td>
</tr>
<tr>
<td>KSRN</td>
<td>gTccggTccgTgcgccAcccTccA</td>
</tr>
</tbody>
</table>

pair even successfully amplified several novel KS fragments from some bacterial strains in our laboratory (data not shown).

It could be concluded that our simple and rapid PCR strategy used in this study worked effectively. On the other hand, the results also confirmed that there were plenty of unknown PKS genes belonging to the underexplored microbe had not been detected, which could predict the existence of many novel microbial secondary metabolites.

Sequence analysis

The alignment of the predicted protein sequences revealed that all the nucleotide sequences encoded the highly conserved region corresponding to the active site of the beta-ketoacyl synthetase consensus region of PKS I genes. The results are summarized in Figure 4.

As described by previous studies, the active-site sequence motif of KS domains was analyzed in relation to their own function (Moffitt et al., 2003). For instance, in the KS domains of niddamycin synthase and tylosin synthase (Bisang et al., 1999), the active-site cysteine were replaced by a glutamine residue (called KSQ domains) which provided a decarboxylative activity. Analysis of the KS domains belong to the NRPS/PKS complexes which used an amino acid moiety as a starter

![Figure 2. Agarose gel electrophoresis of 700 bp KS fragments amplification products from DNA of soil using different forward primer (KSFA-H) and the same reverse primer (KSR-A-N). Lanes1-8 represented the PCR products using primer KSFA-KSFH, row A-N represented the PCR products using primer KSRA-KSRN.](image)

![Figure 3. Agarose gel electrophoresis of 700 bp KS fragments amplification products from DNA of sea water. Lane 1: Tiangen DNA marker III; lanes 2-4: blank control; lane 5: KSFD/KSRH, 0.15 μm membrane sample; lane 6: KSFB/KSRI, 0.15 μm membrane sample; lane7: KSFD/KSRH, 0.8 μm membrane sample.](image)
Figure 4. Clustal V alignment of predicted amino acid sequences of soil and seawater PKS I gene fragments. (*, identity; :, strong similarity; ., weak similarity). Soil KS genes obtained in present study and their identities in the GenBank were as follows: dz-ks-1, 63% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (215-amino-acid [aa] alignment); dz-ks-2, 53% identity to Nostoc sp. PCC 7120 PKS gene (223aa alignment); dz-ks-A1, 64% identity to Nostoc punctiforme PCC 73102 PKS gene (218aa alignment); dz-ks-P1, 66% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (215aa alignment); dz-ks-8103, 61% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (226aa alignment); dz-ks-HH-1, 61% identity to Crocosphaera watsonii WH 8501 beta-ketosynthase gene (216aa alignment); dz-ks-HH-2, 85% identity to uncultured bacterium beta-ketosynthase gene (226aa alignment); dz-ks-DG-1, 61% identity to Myxococcus xanthus DK 1622 PKS/NRPS synthase gene (221aa alignment); dz-ks-DG-3, 66% identity to Myxococcus xanthus DK 1622 PKS/NRPS synthase gene (221aa alignment); dz-ks-DH-1, 63% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (216aa alignment); dz-ks-DH-3, 86% identity to Crocosphaera watsonii WH 8501 beta-ketosynthase gene (193aa alignment); dz-ks-DH-2, 86% identity to Mycobacterium sp. JLS beta-ketoacyl synthase gene (217aa alignment); dz-ks-DH-3, 66% identity to Crocosphaera watsonii WH 8501 beta-ketosynthase gene (193aa alignment); dz-ks-DH-3, 58% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (220aa alignment); dz-ks-DH-2, 67% identity to Anabaena sp. WH 0404 KS gene (215aa alignment); dz-ks-BK-1, 58% identity to Nostoc punctiforme PCC 73102 PKS gene (220aa alignment). Seawater KS genes obtained in present study and their identities in the GenBank were as follows: 015-2, 75% identity to Nitrosonomas europaea ATCC 19718 WcbB gene (216aa alignment); 01529-2, 64% identity to Nostoc sp. PCC 7120 all1648 gene (216aa alignment); 0848-1, 45% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (210aa alignment); 0848-3, 76% identity to Mycobacterium vanbaalenii PYR-1 beta-ketosynthase gene (217aa alignment).
Among the 23 KS sequences obtained in this study, 9 sequences displayed a unique pattern N(D/E)KD, 22 amino acids upstream from the cysteine active site in the KS domain and the conserved pattern VDTACSSS was replaced by VQTACSTS (these amino acid residues were underlined in Figure 4). These two patterns were shown to identify KS domains belonging to hybrid NRPS/PKS systems, the remaining 15 KS sequences (the 4 KS sequences derived from the sea water included) showed typical conserved patterns of KS domains. No KS-like fragment was found from all the 23 KS fragments.

Using BLAST programs provided by the National Center Biotechnology Information (NCBI), these KS gene fragments showed 45 to 85% identities to the known PKS I amino acid sequences in the GenBank database. DQ673147 was the most homologous with the identities 85% and if something is known about the source organisms and the PK-pathways they are participating in. Then, based on a well calculated phylogeny - this information could then be taken to interpret from which species groups these KS-domains were amplified and possibly which PK-biosynthesis pathways would be active in these isolated samples. If other researchers would have also amplified some molecular markers that would allow identifying the microbes present in their samples (Figure 4).

Many of the closest matched genes in the GenBank were related to useful antibiotic biosynthesis. For example, dz-ks-811-1 has a 61% identity to Lyngbya majuscula JamP gene that belongs to a 58 kb jamaicamide A biosynthetic gene cluster (jam). The jam was a remarkable example of a co-linear pathway for the assembly of a complex lipopeptide (Edwards and Gerwick, 2004). The assembly of jamaicamide A involved approximately thirty separate biochemical steps, and were rich in biochemical transformations novel to PKS or NRPS biosynthetic systems. This indicated that the product of dz-ks-811-1 may play a role in the biosynthesis of an expectable antibiotic. Similar results were found for the other KS fragments such as dz-ks-HN-2, dz-ks-HK-3, 015-2 and dz-ks-HN-1. Their characters suggested that these KS gene fragments could be used as homologous hybridization probes to detect positive clones harbored PKS gene clusters in the recombinant metagenomic libraries which would be constructed in the following researches.

**Phylogenetic analysis**

cyanobacteria, proteobacteria and actinobacteria was studied for their capacity of producing various secondary metabolites, including polyketides (Ehrenreich et al., 2005; Staunton and Weissman, 2001). A plenty of PKS gene clusters came from these bacteria had been identified and well studied. In the present study, 22 KS amino acid sequences obtained in the present study and their closest 1 or 2 matches sequences derived from the NCBI were placed in a phylogenetic tree (Figure 5). The dz-ks-48-3 fragments were not included because of its incomplete sequence. This tree showed the high diversity of KS domains from the soil and seawater may derive from cyanobacteria, proteobacteria and actinobacteria. In the bacterial type I PKS gene group, 6 KS domains (DQ640993, DQ641926, DQ641927, DQ673146, DQ673150 and 01529-2) were most homologous to *cyanobacteria* KS genes, these sequences might be actinobacteria-derived. All the 4 KS sequences from the sea water (sequences been underlined in this paragraph) belonged to the bacterial type I PKS gene group. The other 9 KS gene fragments were clustered with the sequences from hybrid NRPS/PKS clusters which had been predicted by their active-site analyses with NCBI. No obviously different characters were observed between the sequences from the soil and those from the seawater. Though some branches showed low bootstrap value, the topologies of the phylogenetic trees based on the NJ, Minimum Evolution and UPGMA computed by the MEGA3.1 revealed the similar results.

Although modern biotechnologies will increase the cultivated bacterial number greatly in the near future, the majority of the microbe will remain uncultivated. Thus the culture-independent method will allow the direct access to the unknown majority.

In this study, a set of degenerate oligonucleotide primers were designed to amplify the KS fragments belong to the PKS I genes from the soil and the seawater samples in China. These primers worked effectively, 23 DNA fragments (19 from soil and 4 from sea water) had been isolated, their predicated amino acid sequences shared 45 to 85% identifies to the known KS genes in the GenBank.

Phylogenetic analyses and the active-site based prediction showed the cloned KS fragments can be divided into two distinct functional groups, since they were all relatively far in evolutionary distance from each other. Fourteen (14) of them belonged to the normal PKS I domains that catalyzed the condensation of the acyl groups, and the other 9 fragments belonged to the PKS/NRPS groups which used an amino acid moiety as a starter unit. All the 4 KS fragments isolated from the seawater belonged to the former group. No significant difference was found between the soil KS isolated from the seawater KS fragments. Due to the diversity of the PCR products, it can be concluded that the degenerate oligonucleotide primer set worked effectively in the PCR amplifying reaction of the different environmental samples. Furthermore, Several KS fragments showed characters that could be used in the following studies.

Data presented in this study showed that the PCR method using degenerate primer to isolate the secondary metabolites biosynthesis gene fragments from the environmental samples were practically effective. This
Figure 5. Phylogenetic analysis of the KS fragments amplified by PCR from soil and seawater DNA and the closest sequences derived from the NCBI. The reconstruction was computed for all 64 KS amino acid sequences by the distance method (NJ, Poisson correction distance model) with interior branch length supports 1000 replicates. E. coli FabF was used as outgroup. Only bootstrap value ≥50% are shown (▲ KS fragments that belong to Hybrid or Mixed PKS/NRPS enzymes complexes; ● KS fragments that belong to the Type I ketosynthase).
study will lay the foundation for the biosynthesis and heterogeneous expression of polyketides and contribute to the exploitation of microbial genetic resources belong to the “underexplored majority”.

ACKNOWLEDGEMENTS

This work was supported by Grants 30670043 of the Chinese National Natural Foundation and 2006AA09Z434 and 2013AA092904 of the High Technology Research and Development Program of China.

REFERENCES


