เมตาจีโนมจากดินป่าชายเลน: แหล่งของจีนชนิดใหม่ที่เกี่ยวข้องกับกระบวนการชีวสังเคราะห์ของสารออกฤทธิ์ทางชีวภาพ และความหลากหลายทางชีวภาพของจุลินทรีย์

นายอนิรุทธิ์ ลิ้มตระกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาจุลชีววิทยาทางการแพทย์ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

MANGROVE MUD METAGENOME: THE SOURCE OF NEW GENE INVOLVING IN BIOACTIVE COMPOUND BIOSYNTHESIS AND MICROBIAL DIVERSITY ACCESS

Mr.Anirut Limtrakul

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Medical Microbiology
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2011

Thesis Title	MANGROVE MUD METAGENOME: THE SOURCE OF NEW GENE INVOLVING IN BIOACTIVE COMPOUND BIOSYNTHESIS AND MICROBIAL DIVERSITY ACCESS
By	Mr. Anirut Limtrakul
Field of Study	Medical Microbiology
Thesis Advisor	Associate Professor Nongluksna Sriubolmas, Ph.D.
Thesis Co-advisor	Associate Professor Pintip Pongpech, Ph.D.
Thesis Co-advisor	Associate Professor Benjamas Wongsatayanon, Ph.D.
Thesis Co-advisor	Lily Eurwilaichitr, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

……....Dean of the Graduate School
(Associate Professor Pornpote Piumsomboon, Ph.D.)

THESIS COMMITTEE

……..Chairman
(Assistant Professor Anan Chongthaleong, M.D., MBA)

……..…….Thesis advisor
(Associate Professor Nongluksna Sriubolmas, Ph.D.)

……..Thesis Co-advisor
(Associate Professor Pintip Pongpech, Ph.D.)

……..Thesis Co-advisor
(Associate Professor Benjamas Wongsatayanon, Ph.D.)

……..Thesis Co-advisor
(Lily Eurwilaichitr, Ph.D.)

……..Examiner
(Tanittha Chatsuwan, Ph.D.)

……..External Examiner
(Assistant Professor Suthep Wiyakrutta, Ph.D)

อนิรุทธิ์ ลิ้มตระกูล: เมตาจีโนมจากดินป่าชายเลน: แหล่งของจีนชนิดใหม่ที่เกี่ยวข้องกับกระบวนการชีวสังเคราะห์ของสารออกฤทธิ์ทางชีวภาพ และความหลากหลายทางชีวภาพของจุลินทรีย์ MANGROVE MUD METAGENOME: THE SOURCE OF NEW GENE INVOLVING IN BIOACTIVE COMPOUND BIOSYNTHESIS AND MICROBIAL DIVERSITY ACCESS อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร. นงลักษณ์ ศรีอุบลมาศ, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร. พิณทิพย์ พงษ์เพ็ชร, รศ.ดร. เบญจมาส วงศ์สัตยานนท์, ดร.ลิลี่ เอื้อวิไลจิตร, 159 หน้า

ในการศึกษานี้ ได้ทำการสืบค้นยีนที่เกี่ยวข้องกับกระบวนการชีวสังเคราะห์ของสารโพลีคีไทด์ ซึ่งเป็นสารที่มีความหลากหลายทั้งด้านโครงสร้างและการออกฤทธิ์ ในเมตาจีโนมของชุมชนจุลินทรีย์จากดินป่าชายเลน (ป่าโกงกาง) จากตำบลคลองโคน จังหวัดสมุทรสงคราม การตรวจหายีนของ KS domain ของ type I PKS ในเมตาจีโนมจากดินป่าชายเลนด้วยการทำปฏิกริยาลูกโซ่พอลิเมอเรส (polymerase chain reaction, PCR) ซึ่งใช้ degenerate primers ที่จำเพาะกับ conserved motif VDPQQR และ histidine active site motif HGTGT พบยีนของ KS domain ที่หลากหลาย 12 ยีน ลำดับกรดอะมิโนที่ถอดรหัสของ domain ใหม่เหล่านี้ มีความเหมือนกัน 34-70% ส่วนใหญ่ มีความสัมพันธ์ทางวิวัฒนาการใกล้ชิดกับกลุ่มของเอนไซม์ cis-AT type I PKS ที่มาจากแบคทีเรียหลายชนิด ที่เหลือพบมีความสัมพันธ์ใกล้ชิดกับกลุ่มลูกผสมระหว่างเอนไซม์สังเคราะห์สารโพลีคีไทด์ และสารเพปไทด์โดยไม่อาศัยไรโบโซม (NRPS) ซึ่งมีสองกลุ่มใหม่ที่เสนอขึ้นจากการศึกษาครั้งนี้คือ hybrid PKS/NRPS I-II และ hybrid PKS/NRPS III การวิเคราะห์ลำดับเบสของยีน 16S rRNA ในโคลนที่สร้างจากผลิตผลของ PCR พบว่าแต่ละไรโบไทป์มีความแตกต่างกัน และมีความ สัมพันธ์ทางวิวัฒนาการที่ใกล้ชิดกับกลุ่มของแบคทีเรียที่ต่างกันไปของไฟลัม *Proteobacteria*, *Acidobacteria*, *Deferribacteres*, *Chloroflexi*, *Verrucomicrobia*, *Lentisphaerae* และมีหนึ่งไฟโลไทป์มีความสัมพันธ์ทางวิวัฒนาการใกล้ชิดกับแบคทีเรียที่ยังไม่ได้จัดจำแนกกลุ่ม ในการสืบหายีน PKS จากคลังเมตาจีโนมของดินป่าชายเลนที่ได้สร้างขึ้นโดยใช้ fosmid vector พบว่าโคลน KKFOS_KSI_pool37_088 ซึ่งมีชิ้น insert DNA ขนาดประมาณ 29 กิโลเบส มียีนของ PKS และ NRPS ชนิดใหม่อยู่ในบริเวณใกล้เคียงกัน แสดงให้เห็นว่าดินป่าชายเลนคลองโคนประกอบด้วยยีนของ PKS ชนิดใหม่และแบคทีเรียชนิดใหม่ที่มีความหลากหลาย ซึ่งสามารถเข้าถึงได้โดยการศึกษาเมตาจีโนม

สาขาวิชา จุลชีววิทยาทางการแพทย์ ลายมือชื่อนิสิต.....

ปีการศึกษา 2554 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.....

4789158020: MAJOR MEDICAL MICROBIOLOGY
KEYWORDS: MANGROVE SOIL/ METAGENOME/ METAGENOMIC LIBRARY/ POLYKETIDE SYNTHASE/ PKS/ KETOSYNTHASE/ KS DOMAIN/ POLYKETIDE SYNTHASE-NONRIBOSOMAL PEPTIDE SYNTHASE HYBRID/ HYBRID PKS/NRPS.

ANIRUT LIMTRAKUL: MANGROVE MUD METAGENOME: THE SOURCE OF NEW GENE INVOLVING IN BIOACTIVE COMPOUND BIOSYNTHESIS AND MICROBIAL DIVERSITY ACCESS. ADVISOR: ASSOC. PROF. NONGLUKSNA SRIUBOLMAS, Ph.D., CO-ADVISOR: ASSOC. PROF. PINTIP PONGPECH, Ph.D., ASSOC. PROF. BENJAMAS WONGSATAYANON, Ph.D., LILY EURWILAICHITR, Ph.D., 159 pp.

In this study, the polyketide synthases (PKS) genes involving in biosynthesis of polyketides, the groups of prolific and various structural and functional bioactive compounds were searched on metagenome of soil microbial community from Klongkone mangrove forest, Samut Songkhram Province, Thailand. Detection of ketosynthase (KS) domain gene of type I polyketide synthase (PKS) in the mangrove mud metagenome by polymerase chain reaction (PCR) using degenerate primers specific for the conserved motif VDPQQR and the histidine active site motif HGTGT resulted in 12 diverse KS domain genes. Deduced amino acid sequences of these novel domains were 34-70% homology. Most of the discovered KSs were phylogenetically related to cis-AT type I PKS, one was related to trans-AT and the others were related to NRPS preceding hybrid PKS/NRPS, which groups of hybrid PKS/NRPS I-II and III were newly proposed in this study. A variety of novel 16S rRNA gene phylotypes were also found by analysis of PCR-product clones. They were phylogenetically related to bacteria in phyla *Proteobacteria*, *Acidobacteria*, *Deferribacteres*, *Chloroflexi*, *Verrucomicrobia*, and *Lentisphaerae*. One phylotype was related to unclassified bacteria. In searching for PKS gene in the mangrove soil metagenomic library constructed with fosmid vector, novel PKS and NRPS modules were found in clone KKFOS_KSI_pool37_088 with the 29 kb DNA insert. These revealed that Klongkone mangrove soil harbored various novel diverse PKS genes and bacteria which could be accessed through metagenomics.

Field of Study: Medical Microbiology
Academic Year: 2011

Student's Signature .....
Advisor's Signature .....
Co-advisor's Signature .....
Co-advisor's Signature .....
Co-advisor's Signature .....

ACKNOWLEDGEMENTS

This study was financially supported by "The 90th Anniversary of Chulalongkorn University Ratchadaphiseksomphot Endowment Fund" and faculty of Pharmaceutical sciences, CU fund.

I am deeply grateful to my advisor, Associate Professor Dr. Nongluksna Sriubolmas for her valuable advices, kind assistance and support throughout my study and research project, and her guidance on my dissertation and manuscript.

I would like to express my deep appreciation to my thesis co-advisor, Associate Professor Dr. Pintip Pongpech for her guidance, suggestions and consultation, Associate Professor Dr. Benjamas Wongsatayanon for her support and persistent help throughout my study, and Dr. Lily Eurwilaichitr for her advice and guidance on constructing metagenomics library. I would like to thank chairman of thesis committee, Assistant Professor Anan Chongthaleong, M.D., Dr. Tanittha Chatsuwan, and Assistant Professor Dr. Suthep Wiyakrutta, an external examiner, for their valuable advices and valuable suggestions to write a dissertation

I gratefully acknowledge Professor Dr. Vithaya Meevootisom for his kind and generous supporting on next generation sequencing of the metagenomic clones, and.Associate Professor Dr. Nittharatana Paphavasit for her advice and kind guidance on suitable source for mangrove soil samples

Many thanks also go to the staffs and graduate students at Enzyme Technology Laboratory, Bioresources Technology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC) for providing facilities and technical assistance on metagenomic library construction, at B600 laboratory, Microbiology department, Faculty of Sciences, Mahidol University for providing facilities and technical assistance on metagenomic library analysis,

Most of all, I would like to thank my parents who love me and encourage me endlessly, without them the success in my life could not have come.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

A	=	Adenylation domain
ACP	=	Acyl carrier protein domain
AMP	=	Adenosine monophosphate
AT	=	Acyl transferase domain
ATCC	=	American Type Culture Collection
ATP	=	Adenosine triphosphate
BAC	=	Bacterial Artificial Chromosome
bp	=	Base pair
$^{\circ}C$	=	Degree Celsius
C	=	Condensation domain
CDD	=	Conserved Domain Database
Cm^R	=	Chloramphenicol resistant
CTAB	=	Cetyl trimethylammonium bromide
DH	=	Dehydratase domain
DMSO	=	Dimethyl sulfoxide
DNA	=	Deoxyribonucleic acid
cm	=	Centimeter
E	=	Epimerization domain
e.g.	=	For example
ER	=	Enoyl reductase domain
et al.	=	And other
EDTA	=	Ethylenediaminetetraacetic acid
g	=	Gravitational force
HEPES	=	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	=	Human immunodeficiency virus
i.e.	=	That is (id est)
IPTG	=	Isopropyl-beta-D-thiogalactopyranoside
kb	=	Kilo base pairs
KR	=	Ketoreductase domain
KS	=	Ketosynthase, Ketoacyl synthase

LB	=	Luria Bertani
LB_{Amp}	=	Luria Bertani with ampicillin supplement
LB_{AmpXGI}	=	Luria Bertani with ampicillin, IPTG and X-Gal supplement
LB_{Cm}	=	Luria Bertani with chloramphenicol supplement
LB_{CmAr}	=	Luria Bertani with chloramphenicol and arabinose supplement
LB_{CmXGI}	=	Luria Bertani with chloramphenicol, IPTG and X-Gal supplement
µl	=	Microliter
M	=	Methytransferase domain (MT)
ml	=	Milliliter
MH	=	Muller-Hinton
min	=	Minute
mm	=	Millimeter
mM	=	Millimolar
MT	=	Methytransferase domain (M)
ng	=	Nanogram
nm	=	Nanometer
NRPS	=	Nonribosomal peptide synthase, Nonribosomal peptide synthetase
OD	=	Optical density
ORF	=	Open reading frame
PCP	=	Peptidyl carrier protein (T)
PCR	=	Polymerase chain reaction
PFGE	=	Pulsed field gel electrophoresis
pH	=	Power of hydrogen
PKS	=	Polyketide synthase
rRNA	=	Ribosomal ribonucleic acid
rpm	=	Round per minute
SDA	=	Sabouraud's Dextrose Agar
SDS	=	Sodium dodecyl sulfate
sp.	=	Species
T	=	Thiolation domain (PCP)
%T	=	Transmittance (percentage)

TBE	=	Tris, Boric acid and EDTA
Te	=	Thioesterase domain
UV	=	Ultraviolet
X-Gal	=	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

CHAPTER I

INTRODUCTION

1.1 Introduction

Antibiotics have been discovered and put on used for more than 70 years, which was the beginning of the curable infectious disease era. Those antibiotics usually are or derive from low-molecular-weight bioactive compounds produced by microbes, plants, or animals (Newman *et al.*, 2000; Newman *et al.*, 2003; Butler, 2004; Butler, 2005; Koehn & Carter, 2005; Davies, 2006). About 60 to 80% of the drugs used for treating cancer and infectious diseases today are derived from natural products (Newman *et al.*, 2003; Gullo *et al.*, 2006). However, with the beginning introduction of antibiotics for therapy, drug resistant pathogens have also been evolutionally selected. With this selective pressure of antibiotic usages, the bacterial strains capable of resisting to the currently in use antibiotics will be selected soon in hospital or community setting (Stubbings & Labischinski, 2009; Maillard *et al.*, 2010). Many reviews and surveys have been reported that in the past two decades the discover and development of antibiotics by the major pharmaceutical companies are in decline and few new antibiotic drugs have reached the market, while the drug resistant and multiple drug resistant pathogens have been emerged with an increasing rate. These are the sign for urgent need of new antibiotic drugs (Barrett & Barrett, 2003; Boggs & Miller, 2004; Bush, 2004; Livermore, 2004; Clardy *et al.*, 2006; Jones *et al.*, 2008; Dent *et al.*, 2010; Ray *et al.*, 2010).

To be prepared for these drug-resistant pathogens, and those that will come, the urge of need for new antibiotics is agreed by many studies (Schito, 2006; Spizek *et al.*, 2010). Despite to the fact concerning the decline of antibiotic production in pharmaceutical industry for over 25 years, and the discovering of only two new developed antibiotic classes, natural products are still believed to be the main sources of lead compounds for drug development and production (Conly & Johnston, 2005; Newman & Cragg, 2007; Butler & Cooper, 2011)

For long, microorganisms have been recognized as major sources of the biologically active molecules (Lorenz & Eck, 2005). These microorganisms, especially from soil, have been cultivated and screened for natural products with antibiotic activity (Handelsman *et al.*, 1998). Bacteria in genus *Streptomyces* isolated from soil seem to be the largest antibiotic-producer in microbial world discovered so far (Watve *et al.*, 2001). Up until 2002, more than 7000 antibiotics are made by this genus alone, while the total of about 9000 bioactive compounds made by all actinomycetes (Demain, 2006). Besides antimicrobial antibiotics, soil microorganisms also produce a large number of bioactive compounds, such as, immunosuppressive agents (cyclosporins from *Trichoderma* and *Tolypocladium* sp., rapamycin from *Streptomyces* sp.), cholesterol-lowering agents (mevastatin from *Penicillium* sp., lovastatin from *Aspergillus* sp.), antihelmintics and antiparasitic drugs (ivermectins from *Streptomyces* sp.), and a potential antidiabetic agent from *Pseudomassaria* sp. The unique unclassified bioactive compound with antifungal activity has been isolated from *Actinomadura* sp. (Badji *et al.*, 2006). Among these bioactive compounds, polyketides also contributed to many groups and represented a large portion of them.

As most of the readily culturable soil microorganisms had been exhausted, the rate of discovering of new compounds decreased. Some researchers have turned to look for rare actinomycetes from different or exotic ecosystems (Lazzarini *et al.*, 2000), and many have turned to marine environment with over 15,000 structurally diverse natural products discovered (Wang, 2006).

However, most of microbial bioactive molecules (both terrestrial and marine origin) found today are mainly from microorganisms which could be cultivated in the laboratories. These cultivable microorganisms accounting for only a fraction of existing microorganisms in a given environmental habitats. By monitoring cellular carbon content of microorganism in a given habitat, the estimated amount of microbial cells are about 10^{29} and 10^{30} cells in terrestrial and oceanic habitat, respectively (Whitman *et al.*, 1998). While using the reassociation kinetics of soil bacterial DNA (Torsvik *et al.*, 1990; Torsvik *et al.*, 1998) and molecular genetics techniques (Torsvik *et al.*, 1998; Torsvik & Ovreas, 2002), soil and sediment were estimated to harbor a very high bacterial number (10^{10} bacterial cells per gram soil) and diversity (might be as high as 10,000 different bacterial species). These

estimations make the cultivable bacteria from these soil and sediment samples accounting for only 0.1-1% of the total bacterial numbers. For the other environments, only 0.001-0.1% of microorganisms in seawater, 0.25% in fresh water, and 0.25% of sediment microorganisms were found to be cultivable (Amann *et al.*, 1995). Conversely, over 99% of microbial diversity in any given habitat is not accessible by cultivation (Langer *et al.*, 2006). Therefore, natural bioactive compounds discovered from cultivable microbes mentioned above are accounted for those from lesser than 1% of the producers. Even though there are low rate and chances in discovery of a new bioactive compound from those 1% pool, the 99% non-cultivable microorganisms is the promising reservoir for finding a new compound. In addition, today, the technology used for accessing genetic and function of non-cultivable microorganisms is also available. This technology is the metagenomics.

The metagenomics is a technique or technology used for studying or analyzing genomes of the entire microbial community or population within environmental samples without the need of cultivation of the microorganisms of interest. These collective genomes are called metagenomic DNA or metagenome which is different from clonal genome obtained from pure-culture. Metagenomics also includes all applications of the entire genomics technology that suites for metagenomic DNA. The basic procedures of metagenomics imitate the genomic library construction of the clonal genome. In metagenomic analysis, the processes involve isolation of DNA from the environmental sample, inserting the isolated DNA into a suitable vector, transforming the recombinant DNA into a host bacterium, and screening for the resulting transformants. By isolating the DNA from the environmental sample makes metagenomics possible for capturing DNA or genes of the majority non-cultivable microorganisms within the environmental sample. Furthermore, transforming of the clone into cultivable host provides the accessibility for manipulation and analysis of the DNA or genes of uncultivable microbes in the environmental sample.

Mangrove forest mud or soil is the area with high organic clay which is submerged under the salt water during high tide and exposed to ambient atmospheric air and dryness during low tide. In addition, it is the place with the combination of terrestrial and oceanic environment; thus, it possesses a very specific and unique ecological niche which could also support highly unique life forms. As mention

above, both soil and sea water are environmental sites that are rich with microbial diversity and its bioactive compound. Then, the mangrove forest seems to be the promising environmental habitat with the specifically uniquely diverse microbial community (Ghosh *et al.*, 2010; Liang *et al.*, 2006). Consequently, this environment is one of the most probable places which the unique and potentially novel genetic distinction and the production of bioactive compound could be found.

1.2. Hypothesis

Mangrove mudflat is the reservoir of novel genes involving in bioactive compound biosynthesis, and the source of microbial diversity.

1.3. Objectives

1.3.1 To construct the mangrove mudflat derived metagenomic libraries.

1.3.2 To analyze and describe genes in mangrove mudflat derived metagenomic libraries involving in biosynthesis of bioactive compounds.

1.3.3 To sample microbial diversity of mangrove mudflat by analyzing its metagenome.

CHAPTER II

LITERATURE REVIEW

2.1 Metagenomics

Microorganisms have been widely known to present everywhere, but only fraction of them is known through the conventional cultivation processes. The studies using genomic DNA reassociation, and some other molecular techniques developed thereafter. Soil bacterial community was found to sustain a vast variety of bacterial diversity than it appeared by cultivating isolation methods (Table 2.1) (Torsvik *et al.*, 1990; Torsvik *et al.*, 1998; Torsvik & Ovreas, 2002). The majority of the existing bacterial population about 99% could not be cultivated using conventional cultivating technology (Table 2.2) (Amann *et al.*, 1995; Torsvik & Ovreas, 2002). Even more efficient cultivating techniques have been developed, those cultures still be the small portion comparing to the uncultured ones (Zengler *et al.*, 2002; Keller & Zengler, 2004; Stevenson *et al.*, 2004; Zengler *et al.*, 2005; Alain & Querellou, 2009; Kenters *et al.*, 2011). Consequently, any studies requiring clones of pure cultures, e.g. microbial metabolic capabilities or functionalities within their community, could not be fulfilled by those majorities of the uncultured microorganisms. Including, the conventionally tapping of the expected wealthy genomic information of the secondary metabolites and natural products or the actual secondary metabolites and natural products themselves from those uncultured would be limited (Handelsman *et al.*, 1998; Lefevre *et al.*, 2008). Isolation of the microbial natural product, conventionally, begins with the extraction of the natural compound from microorganisms themselves or from the culture broth supernatant thereof (Sarker *et al.*, 2006). This way the microorganisms have to be cultivated prior to undergo the extraction processes, which is not applicable for the majority of the microorganisms in the natural population. This is where the metagenomics comes into the view.

Table 2.1 Bacterial abundance observed by fluorescence microscopy, bacterial community complexity, and equivalent genomic numbers in relative to *Escherichia coli* genome (4.1×10^6 bp) (Torsvik & Ovreas, 2002)

DNA source	Abundance (cells cm^{-3})	Community genome complexity (bp)	Genome equivalents
Forest soil	4.8×10^9	2.5×10^{10}	6000
Forest soil, cultivated prokaryotes	1.4×10^7	1.4×10^{10}	35
Pasture soil	1.8×10^{10}	$(1.5\text{-}3.5)\times10^{10}$	3500-8800
Arable soil	2.1×10^{10}	$(5.7\text{-}14)\times10^9$	140-350
Pristine marine sediment	3.1×10^9	4.8×10^{10}	11,400
Marine fish-farm sediment	7.7×10^9	2.0×10^8	50
Salt-crytallizing pond, 22% salinity	6.0×10^9	2.9×10^7	7

Table 2.2 Culturability of bacteria in different habitats (Amann *et al.*, 1995; Torsvik & Ovreas, 2002).

Habitat	Culturability (%)
Seawater	0.001-0.1
Fresh water	0.25
Mesotrophic lake	0.1-1.0
Unpolluted estuarine water	0.1-3.0
Activated slugde	1.0-15.0
Sediments	0.25
Soil	0.3
Forest soil	0.3

In opposed to genome, the clonal genome, the term metagenome or metagenomic DNA, representing the collective of the genomic DNAs of the whole microbial community in a given natural environment was coiled by Handelsman and colleagues in 1998 (Handelsman *et al.*, 1998). Hence, the molecular technology involving in the investigation of the metagenome in a given interest environment by means of functional or sequencing analyses is called metagenomics (Handelsman, 2004a, 2004b). There are other less commonly used terms to describe the same technology such as environmental DNA libraries, zoolibraries, soil DNA libraries, eDNA libraries, community genome, and others (Riesenfeld *et al.*, 2004). As the metagenomics using the metagenome from the natural microbial community dwelling within the environmental sample, it is not limited by cultivability of the subjective microorganisms, but the wealthy genetic and its deduced functional information of those 99% uncultured microorganisms could be accessed. The technology originally and usually involves cloning and the genomic library construction which is not necessarily obligated since the advancement of the sequencing technology, e.g. next generation sequencing, and computational methods for metagenome analyses have developed (Hoff *et al.*, 2009). This additional direct sequencing of the metagenome has compensated the missing genomic information in the metagenomic library caused by toxic expression of the inserted DNA. The clones carrying the host toxic encoding insert would not survive after the transformation, and the genetic information held in the insert would be lost (Forns *et al.*, 1997).

The idea of cloning the metagenome for the molecular studies to gain the genomic information from the uncultured microorganisms was developed before the term was created (Schmidt *et al.*, 1991; Riesenfeld *et al.*, 2004). The field of metagenomics has expanded rapidly in this past decade; today it has covered the studying of many natural microbial communities, e.g. marine water (Venter *et al.*, 2004; Leitao *et al.*, 2009), marine sediment (Bhuiyan *et al*. 2011), acid mine drain (Tyson *et al.*, 2004), mangrove soil (Jiang *et al.*, 2011), terrestrial soil (Rondon *et al.*, 2000; Chung *et al.*, 2008), termite gut (Warnecke *et al.,* 2007), hot spring (Tirawongsaroj *et al*, 2008), and human oral flora (Willner *et al.,* 2011). Although, with the advancement of the current sequencing technology and bioinformatics tools, some studies would omit the cloning step or some would use the metagenome

sequencing to compensate the lost information by cloning biases (Hoff *et al.*, 2008; Hoff, 2009; Hoff *et al.*, 2009; Park *et al.*, 2011), its common processes would comprise environmental extraction for the metagenome, the metagenomic library construction, and the analysis of the metagenomic library clones.

Although metagenomics have been developed to circumvent the cultivation problems, they still have many limitations (Committee on Metagenomics, 2007; Dupre and O'Malley, 2007). First of all, high microbial diversity in the environmental samples, especially from soil, requires a large metagenomic library to cover the entire metagenomic DNA, even with BAC vector. Second, the current technologies, e.g. sequencing technology, bioinformatic tools and metagenomic databases, cannot efficiently support large amount of complex data derived from metagenomes. There are also many technical biases within the metagenomics processes and procedures, e.g. environmental sampling, cell disruption for DNA extraction, the extraction processes, and cloning and expression systems. Furthermore, most of the metagenomes are usually obtained from the genome of major population within the environmental community. The information on the minor members of communities with an important role is obscured, and required the development of new technology to sort out those that still be hidden. Another problem in metagenomic research is an inadequate of reference genomic data, as the genomic data generated by metagenomics is too large and accumulate too fast to be annotated. Some sequences in database do not have functional data which lead to inadequate number of reference genome to identify microbial functions. Finally, the expression system in functional screening mostly relies on *E. coli*. Those genes that cannot be expressed in *E.coli*, their genomic information will be lost and new gene-expression systems are needed.

2.1.1 Metagenome extraction

There are many methods for extracting the metagenome from the environmental samples which may largely be categorized into direct extraction and indirect extraction methods. These subcategories mainly are applied to soil and dirt samples; as, for those air, fresh, or sea water, the microbial cells are most readily separated from the environmental body where the microorganisms harbore. The indirect extraction means that the environmental sample is not directly subjected to

the DNA extraction procedure. Instead, the microbial cells dwelling within the sample are separated from their habitat prior to subject for DNA extraction. This method is preferred in the study aimed for the very high molecular weight DNA, and very large pieces of metagenome. Therefore, cell lyses and the following DNA extraction procedures are performed on the obtained cells embedded in agarose gel to preserve the microbial genome from any shearing forces. This procedure would allow high molecular weight DNA from phylogenetically diverse groups to be obtained, but the obtained DNA would rather low (Gabor *et al.,* 2003).

For the direct extraction methods, the environmental sample is directly subjected to the DNA extraction buffer; and, the microbial cells are directly lysed within the environmental sample. There are many lysis methods with different extent of shearing effects that apply to the sample. The chemical (e.g. detergents, and cell integrity disrupting enzymes) lyses would be gentler than mechanical shearing force of shaking, bead beating or ultrasonic vibrating cell disruption. The more forceful method the more likely higher diversity but smaller of DNA fragments are obtained. For example, in the extraction of mangrove soil from China, the chemical or mechanical lysis procedure alone could not extract the genome of all taxonomic groups of microorganisms within the mangrove soil. Only certain groups of them were obtained. Combinations of all chemical and mechanical methods together might extract the DNA from most various microbial life forms, the archaeal, eubacterial, and fungal cells (Jiang *et al.*, 2010). During the extraction processes the impurity from the environmental body prone to co-extracted with the obtained metagenome is another factor to be regarded. Moreover, different environmental samples coherently have different in impurity extent. The example of one with most profound effect on the metagenome purity and its downstream applications is humic acid in soil samples. Each type of soil has different level of humic acid contamination, and its concentration level at 0.8-51.7 μg/ml, and 0.2-0.48 μg/ml could inhibit restriction endonucleases, and Taq polymerases, respectively. Transformation efficiency is also affected by this humic acid in the concentration of 100 μg/ml (Tebbe & Vahjen, 1993). The extra steps to purify the co-extracted impurity, would normally add more shearing effect to the DNA. To choose the extraction methods, the balance between the size and the diversity of the extracted DNA must suit the research question and the

final analysis step. For instant, if the functional-based screening of the product encoded by a large gene (covering large portion of DNA), the gentle procedures are more likely preferable. After obtaining the metagenomic DNA, generally, it will be used in cloning step to construct the metagenomic library.

2.1.2 Metagenomic construction

In this cloning step, the metagenomic DNA fragment is ligated into the suitable vector and transformed into the host of choice. There are many vectors used in metagenomics, plasmid, cosmid, fosmid, or bacterial artificial chromosome (BAC). Plasmid has the smallest cloning capacity (i.e. <10 kb) while cosmid and fosmid has the cloning capacity ranging about 30-45 kb, and the BAC has the biggest cloning capacity of 50-300 kb (Watson *et al.*, 2007). In this step, the vector and the host chosen are indicated by the metagenome insert size which ultimately depends on the final analysis step. For penicillin amidase gene with average size of 2.7 kb, the metagenomic libraries constructed with plasmid pZErO-2 (insert size of 4-6 kb) resulted in the discovery of the novel penicillin amidase (Gabor *et al.*, 2004). An antifungal positive clone carrying 40 kb insert DNA was successfully cloned with fosmid vector (Chung *et al.*, 2008). BAC metagenomic library constructed from marine microbe with insert size over 160 kb was demonstrated (Beja *et al.*, 2000). For the full potential of BAC vector, the 316 kb insert was successfully cloned with this vector; however it was a tea tree plant library, not a metagenomic library (Lin *et al.*, 2011).

2.1.3 Metagenomic library screening

There are two main approaches, the functional-based and the sequence-based, are available for metagenomic analyses. The functional-based is the biological activity screening strategy that relies on the phenotypic expression of the traits of interest by the analyzed library clones. There is no need of prior knowledge about the gene to be discovered this way, only the functional trait to be monitored, which is the most advantage of this approach. Thus, genes or biological information discovered by this strategy have the real novelty potential, and have never been discovered before.

Many antibiotics, and 15 catalytic enzymes from agarolytic consortium cosmid (Voget *et al.*, 2003), were discovered by this functional based analysis. Limitation of this approach can be contributed by many factors that cause lacking of trait expression. Host-expression compatibility is of the most importance for this approach if the foreign genes come with their own promoter. As the metagenomic DNA fragments are most likely compose of ones from a wide variety of microorganisms. Many of them may be evolutionary distance from host organisms used for the library construction, mostly *E. coli*, the one proved with high heterologous expression capability (Daniel, 2004; de Lorenzo, 2005). The expression machinery, codon usages, protein folding and protein secretion system were different and the inherent trait carried within their DNA insert may not be able to express in *E. coli* (Streit & Schmitz, 2004). The size of the genes of interest and insert DNA fragments used to construct the library was another factor contributing to the challenge. If the size differences were too narrow (assuming the inserts were bigger than the gene with function of interest), most of the random shearing generated DNA fragments would rarely have the intact gene, with promoter on the same single insert fragment. Chance of being expressed and detected would be scarce. Therefore, the high-throughput screening would be needed to search for one positive clone in 20,000 to over 100,000 clones (Brady, 2007; Chung *et al.*, 2008) of this functional-based strategy (Parsley *et al.*, 2010).

For sequence-based, the analysis is relied on the similarity comparison of the metagenomic sequences to those sequences on public database. At least in part, the sequences of the gene or homologous gene of interest must be known and used for comparison, and be referred to by (de novo) sequences from metagenomic insert DNA. The analysis of metagenomic library for gene of interest could be done by gene specific degenerate primers, or hybridization probes, for specific gene function or groups. Sequentially, the positive clone with the insert DNA carrying gene of interest would be detected and analyzed. Alternatively, the whole insert fragment of the entire library are sequenced and traced back to the clone from which sequence of the interesting gene come. Although, this strategy would unlikely obtain the new gene with new function, new variant of the targeting trait have been discovered (Park *et al.*,

2008; Gontang *et al.*, 2010; Parsley *et al.*, 2010a; Parsley *et al.*, 2010b; Parsley *et al.*, 2011).

2.2 Metagenomic and natural product discovery

In general, natural products are organic compound obtained from living organisms, which are divided into two groups, the primary and secondary metabolites. The former groups are usually high molecular weight and fundamentally could be found in every organisms and mostly tend to be used within the producer, the later ones are the low molecular weight compounds (<3,000) found in specific groups of organism (Lefevre *et al.*, 2008; Dewick, 2009). As most of secondary metabolites usually exert some biological effects to others being around them, these bioactive compounds are targeted to be used in drug discovery and development. Bioactive compounds from cultured microorganisms have played the important role in modern drug discovery for many decades. (Handelsman *et al.*, 1998; Gillespie *et al.*, 2002; Bérdy, 2005; Bode & Muller, 2005; Newman & Cragg, 2007; Lefevre *et al.*, 2008). Despite their being prolific resources, they represent only about 1% of the existing microbial community in the nature. The wealthy reservoir of those bioactive compounds of the 99% uncultured microorganisms have been proved accessible by technology of metagenomics (Handelsman *et al.*, 1998; Rondon *et al.*, 2000; Gillespie *et al.*, 2002; Courtois *et al.*, 2003; Streit & Schmitz, 2004; van Elsas *et al.*, 2008a, 2008b). Although they cannot yet be cultivated, their genetic material and information involving in the biosyntheses of their bioactive compounds have been captured and stored in metagenomic libraries. New class of acyl-amino acid antibiotic was the early example of the power of the technology to search for novel bioactive compound biosynthetic gene from as-yet uncultured (Brady & Clardy, 2005, 2003; Brady *et al.*, 2004).

2.3 Polyketides

Polyketides are a class of natural products, which have also been placed as a member under the lipid classification system. Linear polyketides, polyenes, polyether, macrolides and lactone polyketides, aromatic polyketides, and flavonoids are

examples of their sub-categories, representing each structurally different groups of molecular carbon backbone of the polyketides (Fahy *et al.*, 2005). Up to 2006, they comprise about 7,000 known structures of all natural products discovered. They are very diverse in term of their function which could be antibiotics, immunosuppressants, antiparasitics, cholesterol-lowering, or antitumoral agents (Fig 2.1) (Weissman & Leadlay, 2005; Hertweck, 2009).

These highly complex substances and amazingly diverse in functionality are biosynthesized by polyketides synthases (PKSs). These enzymes are evolutionary closely related to fatty acid synthases (Jenke-Kodama *et al.,* 2005). Their biosynthetic pathways are very alike, both in substrates and mechanisms of chain assembly. They both synthesize their products from simple building blogs, e.g. acetate and propionate, by repeating the decarboxylative condensation reaction of the activated carboxylic acid starter or extender units in orderly repeated until reaching the defined molecular chain length of their unique core complex molecular products. However, the fatty acid synthases are more rigid in their biological pathway. Their intermediate products have to undergo all the reduction processes to become fatty acid. On the other hand, the PKSs are more creative and use all possible combinations of those reduction steps to construct varieties of structurally end products and so function (Hertweck, 2009). In part, this might be due to the fatty acid synthases are responsible for producing primary metabolites, while another are making secondary metabolites (Fig. 2.2). There are three main types of these enzymes, type I PKSs (modular types of multifunctional megapeptide enzyme), type II PKSs (complex of bacterial iteratively monofunctional enzymes), and type III PKSs (lacking of acyl carrier protein (ACP), chalcone synthase type).

For modular type I PKSs, these enzymes are multifunctional enzymes with series of active domains for each functional group covalently link together into sets or modules on a single peptide molecule. The typical minimal domains for a simple module comprise ketosynthase (KS), acyl transferase (AT), and ACP (Fig. 2.2, 2.3, 2.4a). In type I PKSs, the series of modules catalyze the incorporation of the acylated monomers into the polyketide backbone chain in order one to one, one module for one monomer incorporation in respect to their physical order on the PKS megapeptide. As

each module operates only one, and relay its product to the next neighboring module until it reaches the last module of the enzyme. This orderly repeated reaction may also be called assembly-line fashion (Fischbach & Walsh, 2006).

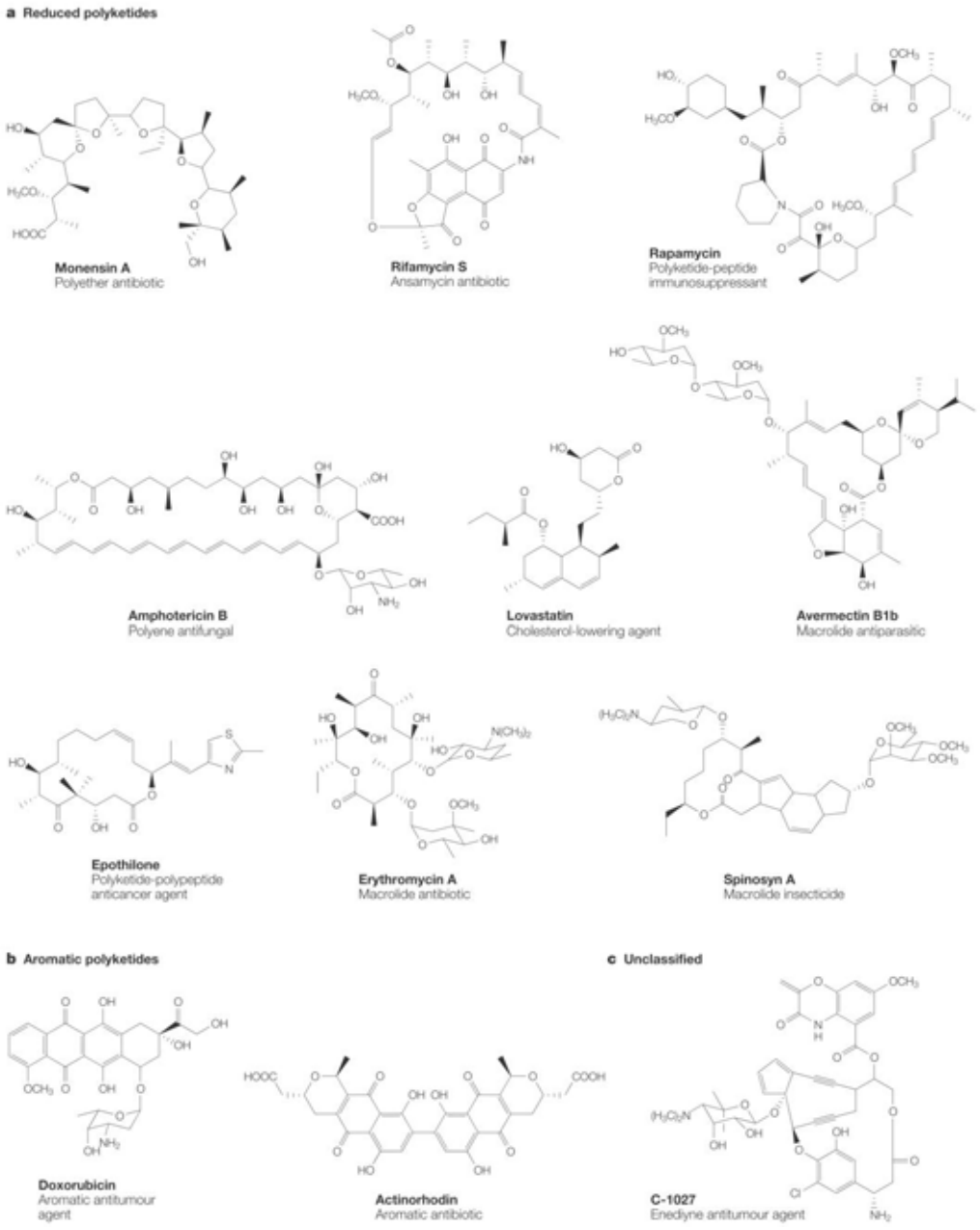


Figure 2.1 Samples of polyketides showing their highly structural and functional diversity. The varieties of polyketides are shown and grouped according to their mode of biosynthesis, (a) reduced polyketides, (b) aromatic polyketides, and (c) unclassified. (Weissman & Leadlay, 2005)

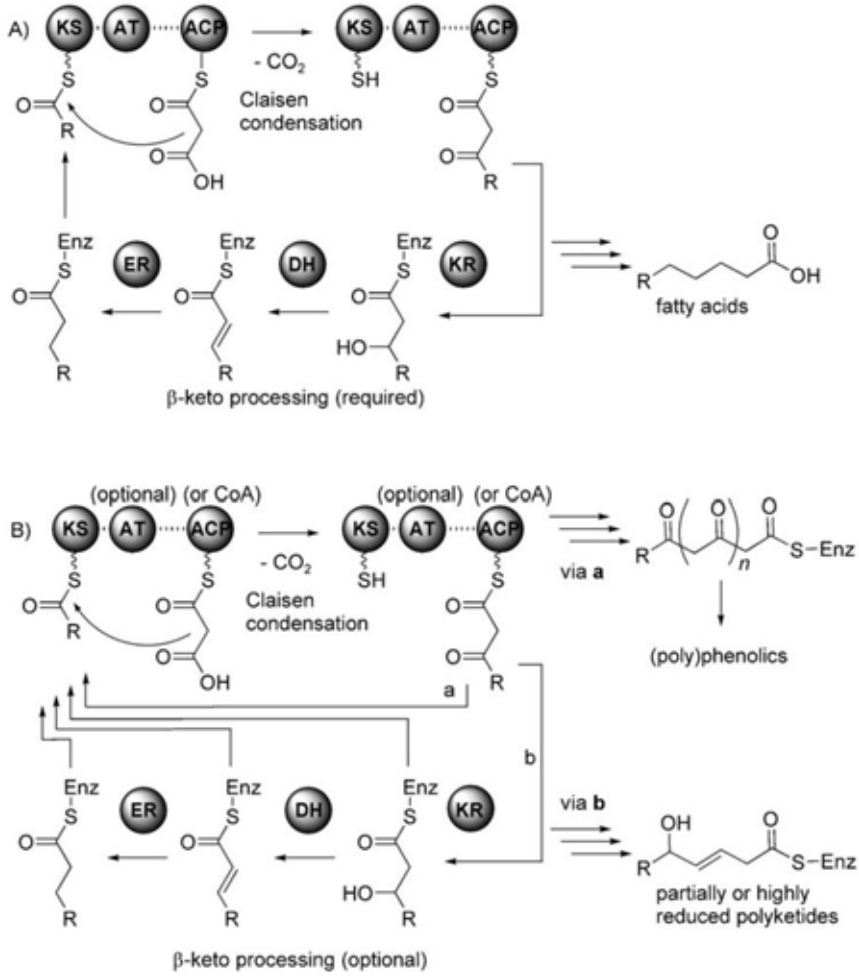


Figure 2.2 The diagrams depicting the basic mechanism comparison between fatty acid synthases (A) and PKSs. (B). For fatty acid synthases (a), there is only one loop for their intermediate core structural products to go through to become fatty acid. Polyketides have a lot more options (B); their intermediate products can avoid all reducing processes via route "a", or could be partially or completely reduced via route "b". Enzyme domains shown in the diagrams are ketosynthase (KS), acyl transferase (AT), acyl-carrier protein (ACP), ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER). The first three are considered minimal core enzyme for PKSs, and another three are may considered optional. (From Hertweck 2001)

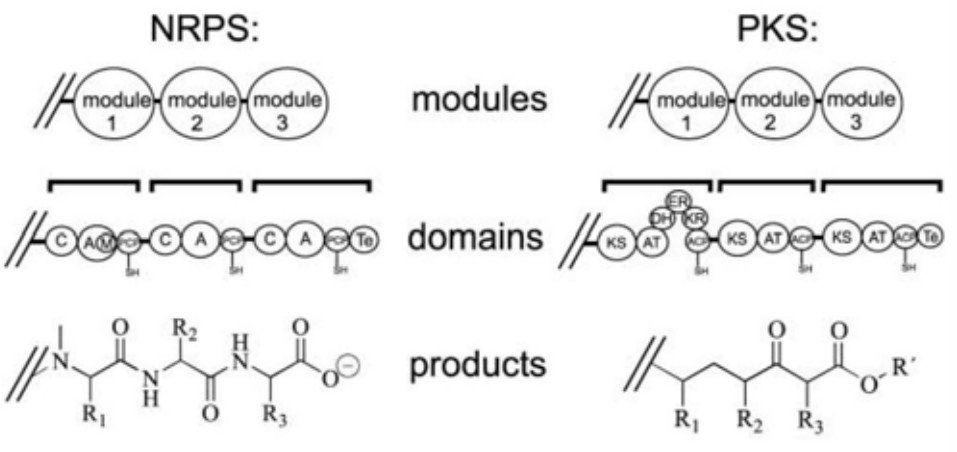


Figure 2.3 The diagrams depicting the structural organization of modules, domains, and their products show the similarity between type I PKS and NRPS. They are multimodular megapetide, within each module resides the functional domains orderly organized and operated in assembly-line manner to generate precisely predictable product. [NRPS, nonribosomal pepetide synthetase: (A) adenylation, (C) condensation, (M) methyltransferase, methylation, (PCP) peptidyl carrier protein; (PKS) polyketide synthase: (ACP) acyl carrier protein, (AT) acyl transferase, (KS) ketosynthase, (DH) dehydratase, (ER) enoyl reductase, (KR) dehydratase, and (Te) thioesterase] (From Schwarzer and Marahiel, 2001)

The additional modification may occur with the present of additional reducing domains like ketoreductase (KR), dehydratase (DH), or enoyl reductase (ER), in a module (Fig. 2.2B and 2.4b). This typical organization of the type I PKS functional groups may sometimes called archetypal erythromycin which may frequently find in *Actinobacteria*, *Myxobacteria*, *Pseudomonades*, and *Cyanobacteria* (Jenke-Kodama *et al.*, 2005). The variety combination organization of these functional groups fruits into the diverse structural polyketide products.

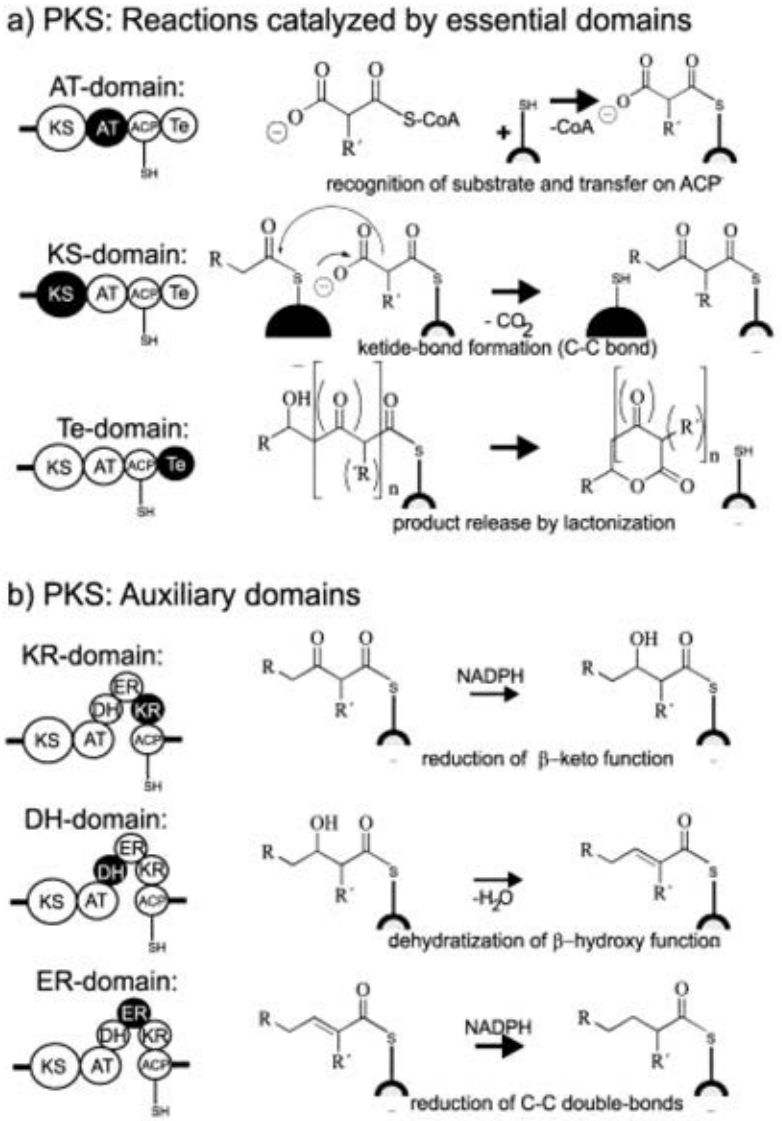


Figure 2.4 Essential PKS domains (a), PKS optional domains (b). (Schwarzer, and Marahiel, 2001)

The modular type I PKSs are divided into multimodular or iterative types. While the iterative type I PKSs use an enzyme module repeatedly to synthesize their products, the multimodular type I PKSs use each of their enzymatic modules once for their product assembly. In addition, sub-typing of the multimodular type I PKSs can be further grouped according to their domain organization. The cis- or trans- AT groups, are categorized based on the possessing of the integrated AT domains on their enzymatic modular proteins (cis-AT), or lacking of the AT domain on their enzymatic

modular proteins (trans-AT). The required AT activity is then provided by the free AT enzyme encoded from the different open reading frame from which the rest of the modular domains encoded. Although, PKS enzymes catalyze the construction of the extraordinarily structurally and functionally diverse compounds, the KS catalytic domains are recognizable as an enzyme superfamily found across the living domains, prokaryotes, fungi, and plants. All sub-types of both type I multi-modular and iterative PKSs mentioned above can be distinguished on the phylogenetic trees reconstructed based on their KS domains amino acid sequences, as each of them would clustered into each unique clade of their own groups. Additionally, for the cis-AT type I multimodular, the clustering among KS of the same taxonomic groups is frequently observed (Moffitt & Neilan, 2003). Another phylogenetic unique groups on the KS phylogenetic tree is the special KS domain of the nonribosomal peptide synthetases (NRPSs) associated PKSs, the hybrid PKS/NRPS, they responsible for the production of the compound with amino acid on the polyketides moiety or vice versa. This special KS domain locates in the PKS module lies adjacent downstreamed from NRPS module.

As the phylogenetic analyses of KS domain amino acid sequence could distinguish and group the subtypes of type I PKSs, they have been used for detection and identification the evolutionary related of the newly discovered KS domains (Jenke-Kodama *et al.*, 2005; Kim & Fuerst, 2006; Zhao *et al.*, 2008). More conveniently, despite their low statistical support, the truncated sequences of KS domains generated from KS specific gene polymerase chain reaction (PCR) are powerful enough to distinguish these type I PKS subgroups, especially between cis- and trans-AT (Piel *et al.*, 2004).

2.4 Nonribosomally synthesized peptides

Nonribosomally synthesized peptides are other secondary metabolites with highly bioactivity, and similar to polyketides in the aspects of their biosyntheses through the similar assembly-line fashion by large multifunctional domains megapeptide enzymes, nonribosomal peptide synthetases (NRPSs). As with the PKSs, NRPSs compose of functional domains organizing into set of modules, each module respond for each activated amino acid incorporated into the final core structural peptide. The functional domains in NRPSs are also named the similar way, and function in similar manner as with the minimal core domains in PKSs (Fig. 2.5).

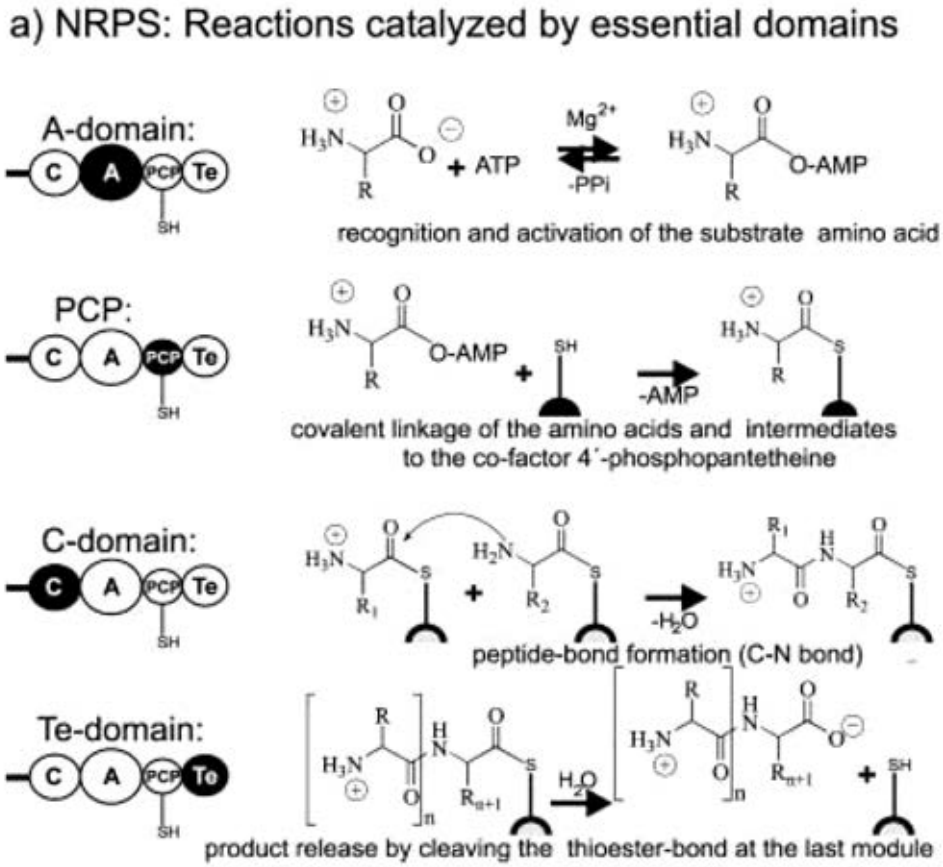


Figure 2.5 Reactions catalyzed by essential NRPS domains: A domain; PCP, T domain; C-domain; and, Te-domain. (From Schwarzer, and Marahiel 2001)

The adenylation (A) domain (analogous to AT) catalyzes amino acid adenylation; it selectively activates the carboxylate with ATP to form aminoacyl-AMP intermediate, and incorporates the aminoacyl group of the activated amino acid monomer onto thiolation (T) domain, (similar to ACP) which is responsible for phosphopentetheine thioester formation and acting as the attachment site for the growing peptide chain molecule, the peptidyl carrier protein (PCP) domain. For NRPSs, their substrate can be either proteinogenic or nonproteinogenic amino acid. The condensation (C) domain (analogous to KS) forms the peptide bond and binds substrates together. At the C-terminus of the NRPSs, there would be thioesterase (Te) domain to release the peptide chain from the NRPSs. A set of these domains form a minimal NRPS module in the form of A-T-$[C\text{-}A]_n$-Te (Schwarzer & Marahiel, 2001; Fischbach & Walsh, 2006).

There are three types of these NRPSs according to their domain architecture and assembly mechanism, the linear NRPS (type A), the iterative NRPS (type B), and the non-linear (type C). Type A, the classical arrangement, the modules and domains arrange corresponding to their peptide product (one module one amino acid), and the DNA sequence of the gene. Type B is also the modular type similar to type A, but the modules are reusable during peptide syntheses; thus there are more amino acid incorporated in the peptide than the module present on the enzyme. Type C is an unorthodox module arrangement; or, it could be monomodular, multiple duplicate of single module.

Besides those domains of the minimal set, NRPSs also have the optional domains as in PKSs. For example, the epimerization (E) domain is often found in NRPSs in gram-positive bacteria where D-amino acid is required in the peptide stricture, but is scarce or absent in cell of the producer. The E domain would epimerize L-amino acid to D-amino acid before the condensation. Therefore, the domain with an -E-C- organization is expected to incorporate the D-amino in the nonribosomal peptide. The C domain in the module function in that selecting and activating L-amino acid but condensing D-amino acid into the product are assigned as $^{D}C_{L}$ domain.

2.5 Hybrid PKS/NRPS

The hybrid PKS/NRPS is the enzymatic system responsible for biosynthesis of natural product with the mixed structural backbone of amino acid and short carboxylic acids. Many bioactive compounds are synthesized by this hybrid system. Some are shown with their domain organization in Fig 2.6f-j. Earlier, this system had been recognized from the bioactive compounds produced by myxobacteria. Their bioactive compound, myxovirescin, epothilone (Fig 2.6g), myxothiazols (Fig. 2.6f), and myxolamid showed the polyketides backbone with the amino acid incorporated (Silakowski *et al.*, 2001). Later, two hypothesized scenarios of hybrid were considered by Shen 2002 using bleomycin biosynthetic gene cluster as a model (Shen *et al.*, 2002). One was the PKS module positioning in front of the NRPS module, and

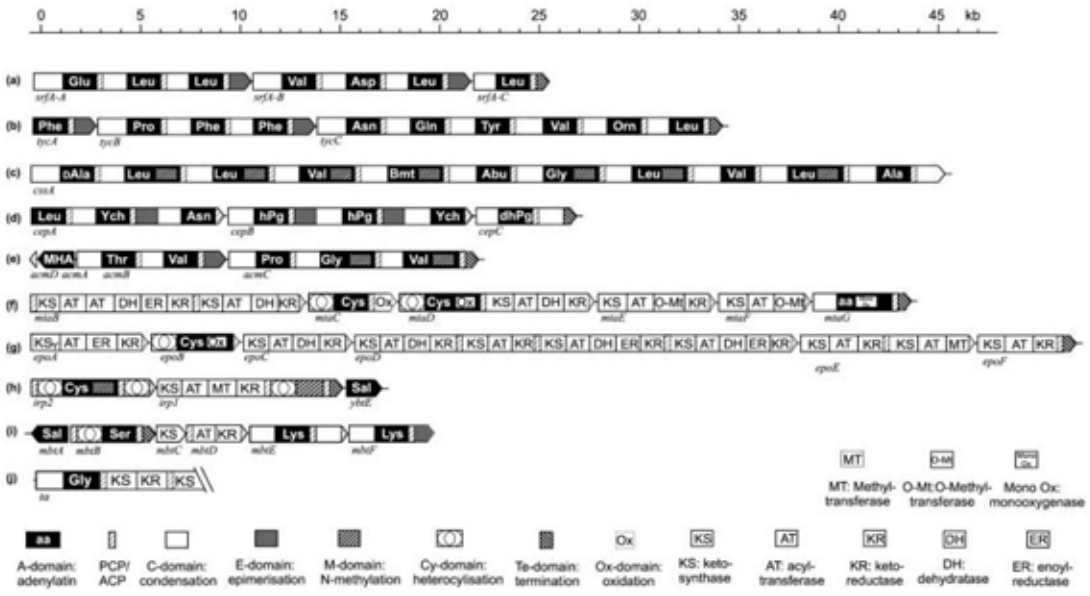


Figure 2.6 Examples for the organization of NRPS and PKS/NRPS hybrid gene clusters. The cluster for the: (a) surfactin from *Bacillus subtilis* ATCC21332, (b) tyrocidine from *Bacillus brevis* ATCC 8185, (c) cyclosporine from *Tolypocladium inflatum*, (d) chlorocremomycin from *Amycolatopsis orientalis*, (e) actinomycin from *Streptomyces chrysomallus*; PKS/NRPS hybrid (f) myxothiazol from *Stigmatella aurantiaca*, (g) epothilone from *Sorangrium cellulosum* So ce90, (h) yersiniabactin from *Yersinia pestis*, (i) mycobactin from *Mycobacterium tuberculosis*, (j) antibiotic TA from *Myxococcus xanthus.* (Schwarzer, and Marahiel, 2001) [larger at Fig. B17]

another was the vice versa. They found and hypothesized that the KS of the later combination, where the NRPS module positions in front of the PKS module, was unique because this KS domain had to act on and condense peptide onto the polyketide unit. This unique KS exhibited the phylogenetic well define and evolutionary apart from other KS (Fig. 2.7) by many studies. It carried the distinct characteristic conserved motifs (Fig. 2.8a) "N(D/E)KD" and "VQTACSTS" (distinct amino acid is shown with underline, the active site is shown with double underline) around the cysteine active site (Moffitt & Neilan, 2003; Ginolhac *et al.*, 2004; Ginolhac *et al.*, 2005; Zhao *et al.*, 2008). Variant of this hybrid signature sequence was found and sub-groups were proposed (Fig 2.8b) (Zhao *et al.*, 2008).

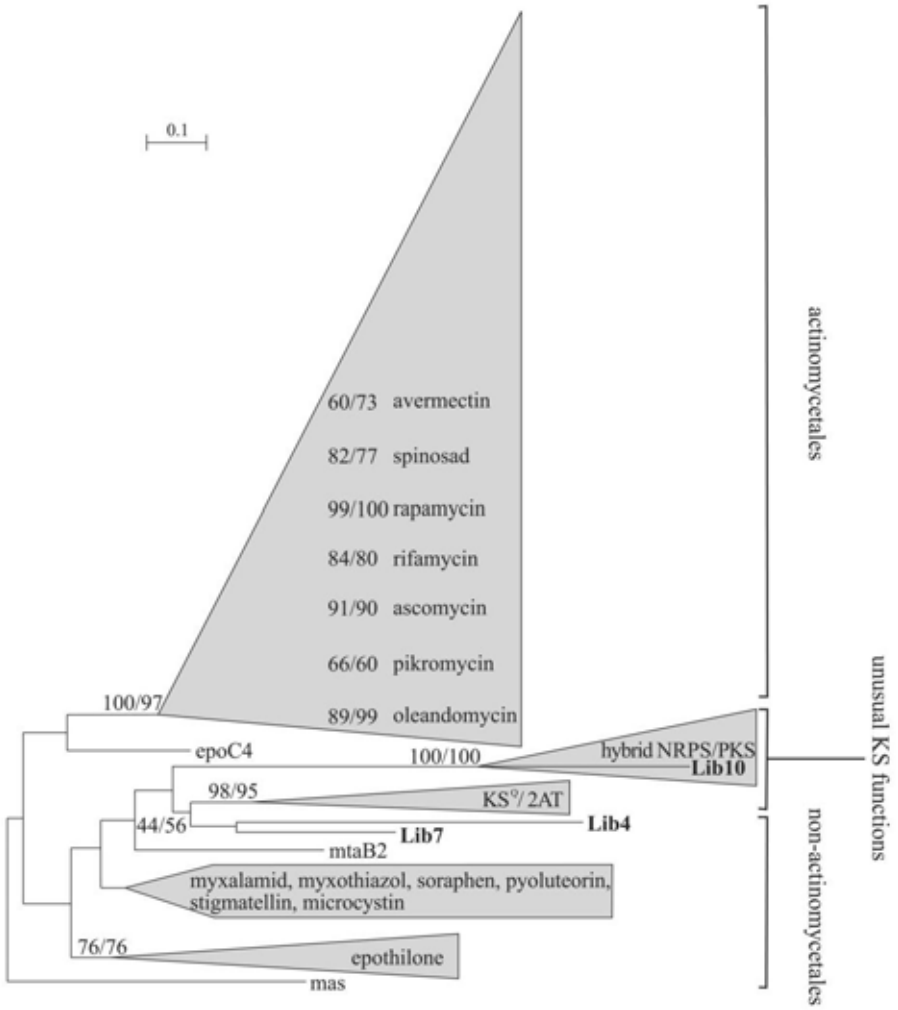


Figure 2.7 The PKS KS compressed tree showing the well defined and evolutionary distinctive from both actinomycetales and non-actinomycetales clade of hybrid NRPS/PKS (Ginolhac *et al*., 2004)

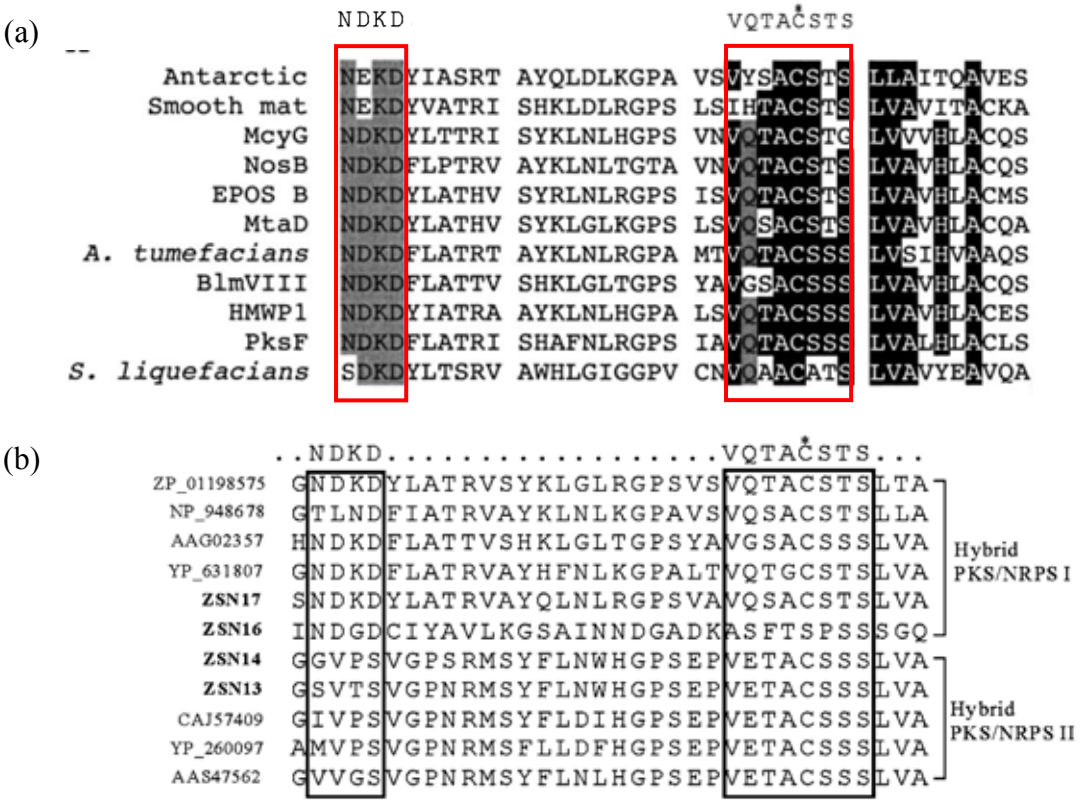


Figure 2.8 Sequence alignments showing the signature motif of PKS in PKS/NRPS hybrid (the NRPS module locating in front of PKS module). (a) Distinct characteristic conserved motifs "N(D/E)KD" locating at 22 amino acids upstream from the active site cysteine in "VQTACSTS" (distinct amino acid is shown with underline, the active site is shown with double underline) (Moffitt, and Neilan, 2003). (b) Hybrid PKS/NRPS I with the signature N(D/E)KD" and Hybrid PKS/NRPS II with variant at the signature sequence **(**Zhao *et al.*, 2008).

One of the factors contributing to the vast diverse structure of polyketides and nonribosomal peptides is believed to be the variety of possible combination and organization of their enzymatic functional domains and modules. The mixed of these two major microbial bioactive compound biosynthesis systems results in even greater diversity of the combination of bioactive compound, and opportunity for those who investigate the way to engineer the combinatorial biosynthesis.

2.6 Mangrove soil microbial community

Mangrove forest is an intertidal and unique ecosystem that could be found along 60-75 % of the tropical and subtropical coastlines around the world. It has been known for high productivity that nurtures and supports life beyond its own system (Nabeel *et al.*, 2010). Most information on those productivity were based on chemical based measuring, e.g. total carbon mass, gas production, or mass shifted by microbial growth (Barrera-Alba *et al.*, 2007; Kristensen *et al.*, 2008). Little studies have been observed the microbial community and diversity in mangrove system. Microorganisms in mangrove sediment have been found that they contribute greatly to the nutrient distribution along the food web within mangrove forest and off shore. The bacteria discovered by cultured based 70% was *Gammaproteobacteria*, 20% was *Firmicutes*, and the rest was *Actinobacteria*. (Nabeel *et al*., 2010). This was similar to other microbial investigation that most discovered by cultivation; only bacteria in phylum *Firmicutes* and *Proteobacteria* were found. Some of them found to function as nitrogen fixer and help promote plant (mangrove) growth (Kathiresan & Selvam, 2006); some expressed valuable lytic enzymes which could degrade the plastic waste (Kumar et al., 2007). In the study of Sudha (2009), the *Proteobacteria* capable of producing L-asparaginase, the enzyme used in lymphoblastic leukemia treatment, was isolated. Using culture-independent methods, the diversity of those cultured microbe were included. 16S rDNA clones from Futien mangrove forest was study and found that the majority (67%) of bacterial phylotypes were fall among five classes of phylum *Proteobacteria*. *Gammaproteobacteria* were the most abundant among them. Over ten phyla were also found in minor portion, including *Cytophaga*, *Flexibacter*, *Bacteroides*, *Actinobacteria*, *Chloroflexi*, *Firmicutes*, *Fusobacteria*, *Clamydiae*, *Verrucomicrobia*, *Fibrobacteres*, *Acidobacteria*, and *Planctomycetes*. Detected over ten bacteria phyla, the mangrove soil microbial community was considerably harboring high diversity of bacteria (Liang *et al.*, 2006). Using the same approach, mangrove sediment in India, was found to harbor at least 8 bacterial phyla. Four classes of *Proteobacteria*, *Flexibacter*, *Actinobacteria*, *Chloroflexi*, *Firmicutes*, *Acidobacteria*, *Gemmatimonadetes*, and *Planctomycetes* were found (Ghosh *et al.*, 2010). As human polluting all other environment, mangrove ecosystem is not excluded, it was found that pollution could shift the microbial community (Ghosh *et*

al., 2010; Taketani *et al.*, 2010), and in some cases could reduce the population diversity (Kristensen *et al*., 2008; Sjoling *et al*., 2005). For the Klongkone mangrove itself, Sakami (2008) found that the number of bacteria in the microbial community at mangrove forest might be lower than shrimp culture pond; however the population complexity were higher and quite stable all around the forest area to mudflat.

To search for new natural products, there are many approaches available (Demain and Sanchez, 2009). Screening from unexplored environment dwelling new microbial taxa by mining microbial genomes would be the most promising approach (Donadio *et al*., 2010). The mangrove forests are one of the richest environments promising for unique microorganisms.

CHAPTER III

MATERIALS AND METHODS

3.1 Materials

3.1.1 Culture media and antibiotics

Luria Bertani (LB) culture medium was used as general all purpose for bacterial culture medium. LB with 1.5% (w/v) agar (Difco) supplemented with ampicillin (T.P. Drug laboratories) (100 μg/ml) (LB_{Amp}) or chloramphenicol (Sigma) (12.5 μg /ml) (LB_{Cm}) were used for the cultivation of appropriate recombinant clones. LB agar with appropriate antibiotic supplement with the addition of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) (Sigma) (40 μg/ml) and isopropyl-beta-D-thiogalactopyranoside (IPTG) (Sigma) (40 μg/ml) (LB_{AmpXGI} or LB_{CmXGI}) was used in blue white selection process. LB agar containing chloramphenicol (12.5 μg /ml) and arabinose (Sigma) (0.01%w/v) (LB_{CmAr}) was used in fosmid high copy number inducing experiment. Sabouraud's dextrose agar (SDA) (Merck) or SDA supplemented with Ketoconazole (USP24; Karingo, Italy) was used for cultivation of *Candida albicans* ATCC 90028 in the antifungal assay. Muller-Hinton (MH) agar (Merck) was used for cultivation of test strains in the antibacterial assay. 7% sheep Blood agar was used for hemolytic functional screening.

3.1.2 Chemicals

Chemicals used for preparing buffers and solutions needed in this study were as listed: agarose (Vivantis), boric acid (BioScience Inc), cetyl trimethylammonium bromide (CTAB) (BioScience Inc), chloroform (Merck), absolute ethanol (Merck), ethidium bromide, ethylenediaminetetraacetic acid (EDTA) (BioScience Inc), glycerol (Fisher Scientific), glycogen (Thermo Fisher Scientific), glacial Acetic acid (Merck), hydrochloric acid (Merck), isopropanol (Merck), sodium chloride (Merck), sodium acetate (Merck), SYBR® Gold (invitrogen), and tris (hydroxymethyl) aminomethane (Tris) (BioScience Inc).

3.1.3 Disposable lab wares

Plastic wares used in this study were as the following: 50 ml and 15 ml centrifuge tubes (Corning), 30 ml centrifuge tubes (Nalgene®), microfuge tubes 0.2, 0.5 and 1.5 ml (Axygen), pipette tips for volume 20, 200, and 1000 ml (Axygen), disposable plastic Petri dish (Hycon). For DNA isolation from agarose gel by electroelution into dialysis tube, SnakeSkin® pleated dialysis tubing (Thermo Scientific) was used. Glass beads (undrilled, 3 mm; Ajax Finechem Pty Ltd) were used to spread bacteria suspension on agar plate.

3.1.4 Microorganisms, and growth condition

Escherichia coli DH5α was used as a host cell for PCR amplicons library construction. *E.coli* EPI300™-T1R (Epicentre® Biotechnologies), the host for fosmid library construction, was supplied as part of The CopyControl™ Fosmid Library Production Kits. pSuperBAC1 carrying *E. coli*, DH10B JW366 (obtained from Department of Plant Pathology, University of Wisconsin, Madison, USA.) was used for the construction of chloramphenicol-resistant *Bacillus subtilis*. *Candida albicans* ATCC 90028, *Serratia marcescens* ATCC 27117, and chloramphenicol-resistant *Bacillus subtilis* (*B. subtilis* Cm^R) were used as test strains. *B. subtilis* Cm^R was constructed earlier in the laboratory by electrotransformation of pSuperBAC1 into *B. subtilis* ATCC 6633 cell. *Streptomyces peucetius* var. *caesius* TISTR1253 purchased from Thailand Institute of Scientific and Technological Research (TISTR) Culture Collection was used as a source of type I PKS KS gene, a positive control for its gene specific PCR.

Otherwise indicated, all bacteria used in the study normally were grown at 37 ºC overnight, with 200 rpm shaking for broth culture. *Candida albicans* ATCC 90028 was normally cultivated at 30 ºC on SDA.

3.1.5 Vectors, primers, DNA marker, and molecular kits

pCC2FOS™ vector (Epicenter) provided within CopyControl™ Fosmid Library Production Kit was the vector used for fosmid library construction. pGEM®-T Easy vector (Promega) was a T-A cloning vector used for PCR amplicon

library construction. Primers used in this study were listed in Table 3.1. Lambda DNA *Hin*dIII digest marker (Fermentas), GeneRuler DNA marker mix (Fermentas), VC 1kb DNA Ladder (Vivantis), VC100bp Plus DNA Ladder (Vivantis), Low Range PFG Marker (New England BioLab) were DNA marker used in this study. The Taq enzyme and buffer for PCR reaction throughout the study were prepared using GoTaq® Colorless Master Mix (Promega).

High-Speed Plasmid Minikit (Geneaid) and Gel/PCR DNA Fragments Extraction Kit (Geneaid) were used for general plasmid extraction and PCR purification, respectively. FosmidMax® (Epicenter) was used for high yield fosmid extraction.

Table 3.1 Summary of primers used in this study

Name	Sequences	used for	References
MDPQQRf	5' RTRGAYCCNCAGCAICG 3'	KS specific PCR	Kim *et al.*, 2006
HGTGTr	5' VGTNCCNGTGCCRTG 3'	KS specific PCR	
27f	5' AGAGTTTGATCMTGGCTCAG 3'	16S rDNA specific PCR	Lane. 1991
1492r	5' TACGGYTACCTTGTTACGACTT 3'	16S rDNA specific PCR	
pCC1™/pEpiFOS™	5' GGATGTGCTGCAAGGCGATTAAGTTGG 3'	fosmid end sequencing	Epicenter
pCC1™/pEpiFOS™	5' CTCGTATGTTGTGTGGAATTGTGAGC 3'	fosmid end sequencing	

3.2 Instruments

For centrifugation of DNA, 15 ml and 50 ml centrifuge tubes were centrifuged in Heraeus Megafuge 1.0R with refrigeration (DJB Labcare, England) and 30 ml centrifuge tubes were used with refrigerated centrifuge (Sigma 2K 15, B. Braun Biotech International, Pennsylvania, USA). The agarose gel electrophoresis was performed using i-MyRun.NC Electrophoresis system (Cosmo Bio Co., LTD., Japan). UV visible spectrophotometer (UV-160A UV-Visible recording spectrophotometer, SHIMADZU, Kyuto, Japan) was used for optical density (OD) and % transmittance

measurement. Automated thermal cycler (Mastercycler gradient, Eppendorf, Hamburg, Germany) was used for PCR. MicroPulser™ (Bio-Rad, California, USA) was used for electroporation. CHEF-DR®III (Bio-Rad, California, USA) was used for pulsed field gel electrophoresis (PFGE).

3.3 Mangrove mud samples

Mangrove mud at the depth of 0-10 centimeters from surface was collected from Klongkone mangrove forest, Samut Songkhram Province, Thailand during July to August. The sample collection was planned ahead for the low tide period occurring in the day time as we did not equipped with any special tools for collecting the sample under the water. The tide forecast information from Tahjeen (Tachin) tide station, Thailand, was accessed through web site of "Mobile Geographic" (http://tides.mobile geogra phics.com/locations/6300.html). The website provided information regarding tide forecast from tide station around the world. For Klongkone mangrove forest, the Tahjeen tide station was the closest one. Date and month during rainy season was searched for ones with low tide during the daytime. Sequentially, the tide prediction diagram for that particular date was examined. For example, the tide diagram for July, 13th 2007 (Fig. 3.1), the diagram provided the information telling that the sun rise at about 6:00 a.m., while the low tide started sometime before midnight (12:00 p.m.) and would reach the lowest point below sea level at 9:42 am., the tide would rise and reach the highest 3 meters above sea level at sunset (around 6:48 p.m.) of that day. The tide stays below sea level from around 8:00a.m. to around 12:00 p.m. which mean four hours period for sample collection in that day.

Nine mangrove soil cores from within two square meters were collected using soil core sampler, a PVC pipe, with about 1 inch diameter, and kept on ice until it reached the laboratory. The DNA extraction of the soil sample was performed immediately upon the arrival as described below; the rest was kept at -20 ºC for long term storage. The pH was monitored at the sampling time and before DNA extraction, measured by pH test strips (Merck).

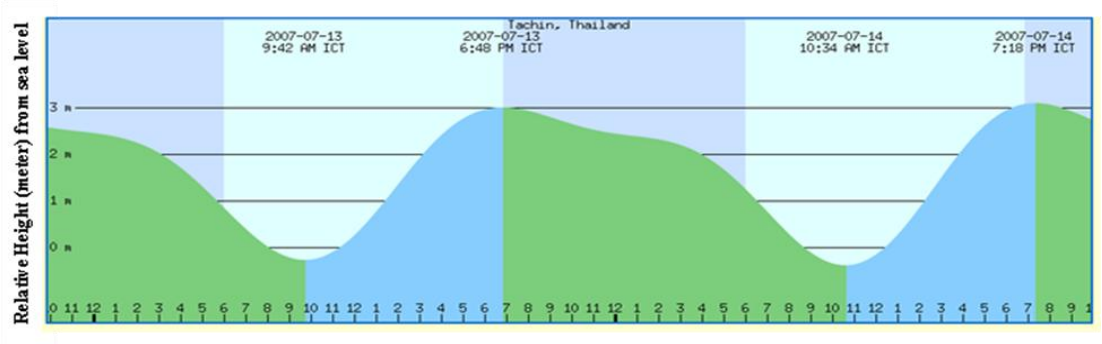


Figure 3.1 Tide prediction diagram at Tahjeen tide station obtained from "Mobile Geographic" (http://tides.mobilegeographics.com/locations/ 6300.html). The cartoon for tide level at specific time of a particular day is useful for planning the sampling time. The Y axis is relative height of water comparing to standard sea level. X axis is point of time in a day displayed in 12 hour format. The plotting areas in green with declining slope are the time for low tide. The blue plotting areas with inclining slope are representing high tide. The alternative dark and light shade above the potting area is represent night and day time, respectively

3.4 Klongkone mangrove metagenome preparation

3.4.1 Metagenomic DNA extraction

Upon arrival to the laboratory, 10 cm of the top layer of mangrove soil cores were pooled into one Klongkone mangrove soil sample. After thoroughly mixed, the sample was sieved through 4 mm and 2 mm meshes, serially, to remove animal or plant parts. The metagenome or the DNA from the soil sample was extracted directly using chemical and heat as microorganism cell disrupting means (Brady, 2007). A total of 12.5 grams of sieved soil sample was weighed into each 50 ml conical centrifuge tubes. Preheated (70 ºC) lysis buffer containing 2% (w/v) SDS was added into the soil containing tubes. The tubes were, then, gently inverted and well mixed prior to incubate for 2 hours in 70 ºC water bath. The tubes were inverted to evenly resuspend the bottom settled soil every 30 min during the incubation. Next, the soil lysates were left to cool down before they were centrifuged at 3500×g for 10

min at 4 ºC. The obtained supernatants were transferred by poured to new clean centrifuge tube and repeated the centrifugation process. In every supernatant transferring, the waxy white material should be avoided. Sequentially, the supernatants were centrifuged again for 20 min before equally aliquot were distributed into new clean centrifuge tubes for DNA precipitation with 0.7 volume of isopropanol. After the mixtures of crude DNA solution and isopropanol were gently well mixed and incubated at room temperature for 30 min, the DNA was pelleted by centrifugation at 3500×g for 30 min at 4 ºC. The DNA pellets were washed with 70% (v/v) ethanol, and left air dried at room temperature for few hours to get rid all the alcohol. The pellets were dissolved in sterile deionized (DI) water. At this point, a brownish crude mangrove soil DNA was obtained. The further purification and size selection of the DNA was done by agarose gel electrophoresis.

3.4.2 DNA purification

The compressed gel electrophoresis was used for DNA purification with CHEF-DR® III Pulsed Field Electrophoresis (Bio-Rad Laboratories). The crude DNA was loaded on combined well in 0.8% (W/V) agarose gel in 0.5X TBE, and was subjected to pulsed field gel electrophoresis (PFGE) with the 2 to 3 hours running condition 9.0 Volt/cm, 0.2 second initial switching time, 5 second final switching time, 120º angle, 14 ºC. The low range PFG ladder (New England Biolabs) was used as DNA molecular marker. The DNA in agarose gel portion that cleared of brown substance was stained with SYBR Gold (invitrogen), and observed on the in-house blue light box with amber filter. Agarose noodle embedded with high molecular weight DNA fragments about 48 kb and above was collected for isolation of the DNA by electroelution with dialysis tubing.

3.4.3 DNA isolation from agarose gel using real time electroelution

To perform electroelution using dialysis tubing (Sambrook, 1989), high molecular weight DNA embedded agarose noodle was placed in, pre-soaked and 0.5x TBE filled, one-end sealed dialysis tube. Then the open end was carefully and securely sealed to prevent any air bubble to be present inside the tube. After the tube

was immerged under 0.5x TBE in the transparent electrophoresis chamber, the gel noodle was positioned to the tube wall near the negative polar of the chamber. The complete migration of the DNA was observed real time on blue light box while electromotive force (EMF) was supplied to the DNA (Fig. 3.2). The EMF polarity was reversed for 10 second to loose the DNA from the dialysis tube wall, and emptied agarose noodle was discarded. The DNA solution in the dialysis tube was collected into 15 ml conical tube and centrifuge at 4000xg for 15 minutes to clear out the possible carried over small pieces of agarose. To be more cautious and avoid the invisible small agarose pieces, only 90-95% of the supernatant was carefully transferred to new tubes without disturbing the bottom of the tubes where the pellet might possibly located. The obtained DNA solution, then, was undergone the DNA precipitation with the aid of glycogen (Fermentas), according to the manufacturer's manual, to ensure the recovery of the DNA. Briefly, 1/10 volume of 3M sodium acetate, 1/400 volume of 20 mg/ml glycogen, and 2.5 volume of ethanol was serially added one by one to the DNA solution, respectively. Most importantly, the DNA solution must be gently mixed thoroughly after and before each addition of each reagent. The mix solution was aliquot into microfuge tubes and spun down at top speed for 10 minutes before the supernatant was discarded. The pellet was washed twice with 70% (v/v) ethanol before it was air dried and resuspended in sterile DI water. Quality and quantity of the purified mangrove soil DNA was assessed by PFGE and spectrophotometry at $OD_{\lambda 230nm}$, $OD_{\lambda 260nm}$, and $OD_{\lambda 280nm}$. The ratio of $OD_{\lambda 260nm}$/ $OD_{\lambda 230nm}$ was used for estimating impurity from polysaccharides, phenolate ion, thiocyanates, and other organic compounds including humic acid, while $OD_{\lambda 260nm}$/ $OD_{\lambda 280nm}$ was used for the evaluation of the impurity from proteins. The acceptable ratio for ODλ260nm/ ODλ230nm and ODλ260nm/ ODλ280nm for DNA was ≥ 1.5and ≥ 1.8, respectively (Wilfinger *et al.*, 1997; LaMontagne *et al.*, 2002).

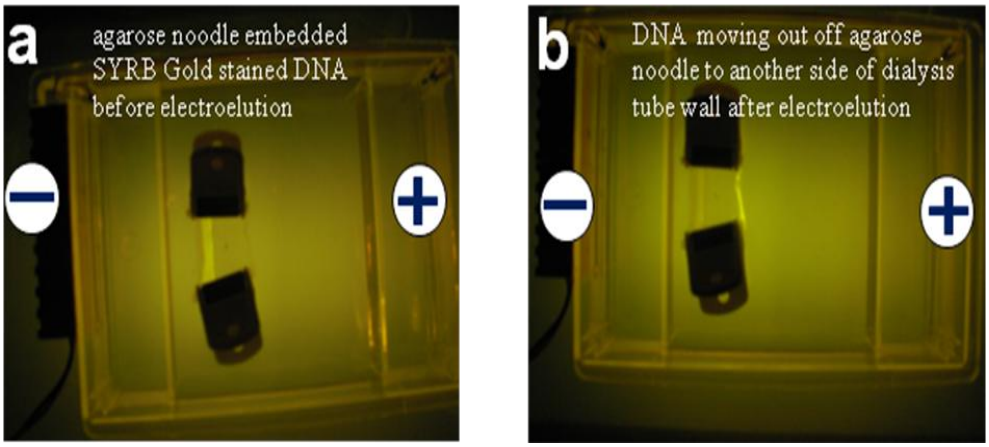


Figure 3.2 Real time electroelution in dialysis tubing. (a) The illuminated SYBR gold stained DNA embedding agarose gel inside a sealed dialysis tube was observed under blue light through amber filter before the electroelution. (b) After the electroelution was finished, the illuminated DNA was migrated to the opposite inner wall of the dialysis tube leaving the agarose gel noodle dimmed.

3.5 Assessing of potential novel PKS and bacterial diversity.

3.5.1 Type I-PKS KS gene specific PCR

The amplification of specific type I-PKS KS domain gene was performed using the gene specific degenerated primers, MDPQQRf/HGTGTr (Kim *et al*., 2006) designed for amplification of type I-PKS KS gene between MDPQQR conserved motif and HGTGT histidine active site motif (Fig. 3.3 and Table 3.1) from wide host range. The purified mangrove mud metagenome was used as template in down scaled 20 µl reaction of touchdown PCR with GoTaq® Colorless Master Mix (Promega) and 10 µM of each primer. The PCR cycles were started with the denaturation step at 96 ºC for 5 min, followed by 7 cycles of 96 ºC for 1 min, 65 ºC annealing, which would decreasing 1 ºC per cycle for 1 min. the polymerization step was carried out at 72 ºC for 1 min, followed by 40 cycle of 96 ºC for 1 min, 58 ºC for 1 min, and 72 ºC for 1 min, and finished up with the final extension at 72 ºC for 7 min. The PCR product was observed on agarose gel electrophoresis. The DNA in the

agarose gel with correct size amplicon was isolated using the Gel/PCR DNA Fragments Extraction Kit (Geneaid).

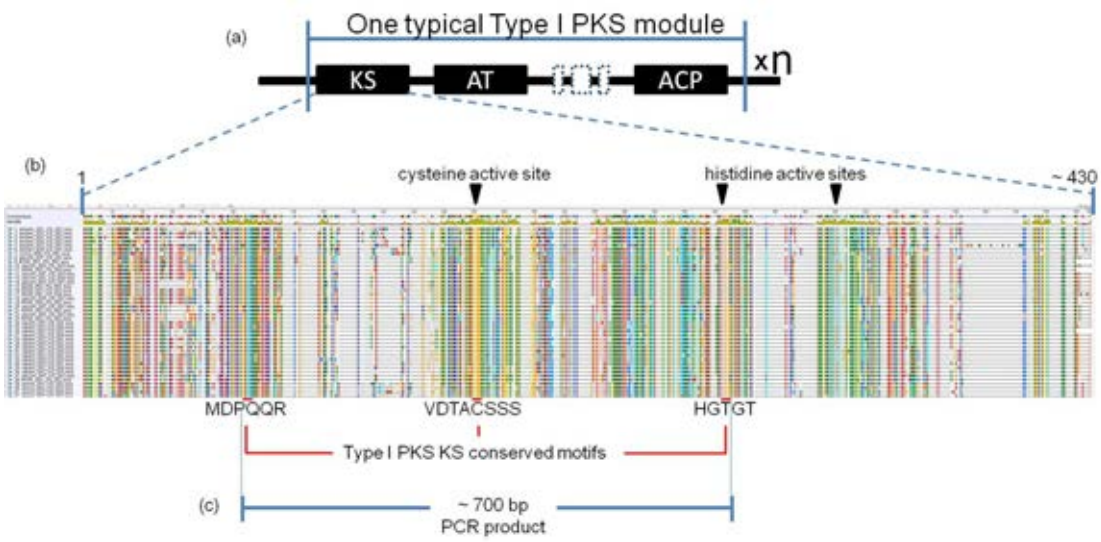


Figure 3.3 A typical module organization of type I-polyketide synthase (PKS) megapeptide. (a) The minimal set comprising ketosynthase domain (KS), acyl transferase domain (AT), and acyl carrier protein (ACP). The addition functional domain may present in many modules of PKS, and are shown in dashed boxes. (b) The alignment of type I-PKS KS domain. The active sites are marked with arrow heads and the conserved motifs are underlined and labeled with their conserved amino acid sequences. The motifs "MDPQQR" and "HGTGT" are used for PCR primer design (Ginolhac *et al.*, 2004). (c) The 700 bp PCR product.

3.5.2 Amplification of 16S rDNA from Klongkone mangrove mud metagenome

The almost complete 16S rRNA gene was amplified using 16S rRNA universal primers, 27f and 1492r, as described by Lane (1991). The PCR reaction was set up using GoTaq® Colorless Master Mix (Promega) and 10 µM of each primer. The purified Klongkone mangrove mud metagenome no more than 100 ng was used as template. The amplification condition after the initial heating at 95 °C for 5 min was 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min followed by a final

extension step of 5 min at 72°C. PCR products were monitored by agarose gel electrophoresis. The DNA in the agarose gel with correct size amplicon was isolated using the Gel/PCR DNA Fragments Extraction Kit (Geneaid).

3.5.3 Preparation of electrocompetent *E. coli* DH5α

Electrocompetent *E. coli* DH5α cells were prepared by glycerol washing adapted from the instruction in Biorad's *E. coli* Pulser™ operating manual. The overnight culture of *E. coli* DH5α in a volume of 1 ml was used to inoculate into 200 ml of LB broth in 1,000 ml Erlenmeyer flask, which was incubated at 37 °C with 220 rpm shaking, until the freshly grown *E. coli* DH5α reached the cell density of approximately 0.3 but not exceeded the 0.6 $OD_{\lambda 600}$. Every step onward, every solutions and lab-wares had been pre-chilled and kept cold throughout the procedure. The cells were transferred into pre-chilled centrifuge bottle and spun down at 4000×g and 4 °C for 10 min. Supernatant was discarded. Cell pellet was gently resuspended to the original volume with ice-cold sterile DI water. Cell suspension was spun down to collect the bacterial cells. The washing process was repeated three more times using 10% (v/v) glycerol with the volume stepped down to 0.5, 0.1 and 0.05 volume of the original, respectively. Then, the competent cells were finally suspended in 0.001× of the original volume by 10% glycerol and divided into 50 µl aliquot for each electroporation reaction. The competent cells were immediately used and the left over aliquots were kept at -80 °C.

The electrotransformation was operated using MicroPulser™ (Bio-Rad) with pre-programmed Ec2 operating mode automatically setting. *E. coli* cells in 0.2 cm gap electroporation cuvette were electroporated with voltage of 2.5 kV and field strength (E) of 12.5 kV/cm.

3.5.4 Amplicon libraries construction

As the amplicon products, both partial type I-PKS KS and 16S rRNA gene generated from the metagenomic DNA were mixed products. The generated amplicons were ligated into T-A type cloning vector, i.e. pGEM®-T Easy Vector (Promega), and electrotransformed into electrocompetent *E. coli* DH5α, as described

in section 3.5.3, to sort out the individual PCR products. Circular pGEM®-T Easy vector (extracted from blue colonies from prior experiments) was used as positive control. One hundred white transformant colonies from 18-24 hours at 37 ºC incubation on 100 μg/ml ampicillin, 40 μg/ml IPTG, and 40 μg/ml X-Gal supplemented LB (Luria Bertani) agar plate (blue/white selection) were randomly collected.

3.5.5 Clones selection for sequencing

The streak plate for single colony isolation of those randomly selected white clones were done on LB agar plate with 100 μg/ml ampicillin, 40 μg/ml IPTG, and 40 μg/ml X-Gal supplementation to test for late and low expression of β-galactosidase activity by keeping the plate at 4 ºC for another 24 hours after 18 hours of 37 ºC incubation. The remained white clones, then, were used for colony PCR and determination of the expected size of recombinant plasmid. The colony PCR for the partial type I-PKS KS amplicon was performed as described in section 3.5.1, but using only toothpick tip picked part of colony as source of DNA template. For 16S rRNA gene transformant clones, the colony PCR was performed using the vector specific primers for the insert flanking region, M13f and M13r to avoid the problem which may cause by host 16S rDNA. The results were visualized by agarose gel electrophoresis.

To observe the size of the recombinant plasmid, for each of the selected clones, the used toothpick for picking the recombinant clone for colony PCR was used for inoculating the LB broth supplemented with 100 μg/ml ampicillin. After overnight incubation, cells from 1 ml of each clones were collected, washed with sterile DI, and extracted for its recombinant plasmid using High-Speed Plasmid Minikit (Geneaid). Each extracted plasmid was digested with *Eco*RI and observed on agarose gel electrophoresis. The sizes of the pGEM®-T Easy vector, the partial type I-PKS KS insert and the 16S rDNA were about 3 kb, 700 bp and 1.5 kb, respectively. The positive recombinant plasmids, from the clones having both correct colony PCR product and correct size of the recombinant insert, were sequenced at Biodesign Co., Ltd. (Thailand) and 1st BASE Pte Ltd (Singapore) sequencing services.

3.5.6 Bioinformatic analyses

Bioinformatic tool packages available online, the NCBI Entrez, blastn, blastp, and pblastx (Altschul *et al.,* 1990; Altschul & Lipman, 1990; Zhang & Madden, 1997; Camacho *et al.,* 2009) were used for searching and downloading homologous sequences; CDD searching service (Marchler-Bauer *et al.,* 2002; Marchler-Bauer *et al.,* 2003; Marchler-Bauer *et al.,* 2005; Marchler-Bauer *et al.,* 2007; Marchler-Bauer *et al.,* 2009; Marchler-Bauer *et al.,* 2011) was used for identification guiding for conserved domains and domains organization with in deduced amino acid sequences (http://www.ncbi.nlm.nih.gov/); and, ClustalW2 (Larkin *et al.,* 2007) and MAFFT multiple alignment services at EBI (http://www.ebi.ac.uk/Tools/msa/) were used in multiple alignment and % homology analysis. PKS and NRPS domains and module detection and organization were done at PKS/NRPS Analysis Web-site (Bachmann & Ravel, 2009), and at Structure Based Sequence Analysis of Polyketide Synthases (SBSPKS) web services (Yadav *et al.,* 2003a, 2003b; Ansari *et al.,* 2004; Yadav *et al.,* 2009; Anand *et al.,* 2010). For analyses performing on local PC, the bioinformatic tools and package software, the freeware ,BioEdit (V.7.0.9) (http://www.mbio.ncsu.edu/bioedit/page2.html), and the commercial one, Geneious (V. 5.4.6) (Drummond *et al.,* 2010) purchased from Biomatters Ltd, were also used in this study. For DNA sequences which were not encoding for protein such as 16S rRNA gene sequences, blastn was the main tool for searching and collecting the homologous sequences. For protein encoding sequences, for example, type I PKS KS domain gene, the correct open reading frame needed to be assessed and the expected deduced amino acid sequences were generated with BioEdit and Geneious software. blastx and blastp were used for the homology assessment and for retrieving those homologous sequences on the public databases, and also for collecting them for later study and analyses. Online ClustalW2 from EBI was frequently used for the multiple alignment and the similarity assessment of the homologous sequences. For local PC, multiple alignments and phylogeny of the obtained homologous sequences were calculate and analyzed by Bio Edit and Geneious.

3.6 Fosmid library construction.

Purified Klongkone mangrove mud metagenome was used as a source of genomic DNA for fosmid library construction in pCC2FOS^{TM} vector by CopyControl^{TM} Fosmid Library Production Kit (Epicentre technologies). The construction was made as recommended by manufacturer's manual protocol with some modification. As the purified DNA situated with the size readily for fosmid cloning system, the shearing step was omitted. After the end-repairing processes, the metagenome fragments were ligated into pCC2FOS^{TM} vectors. The ligated DNAs then were packed into phage packaging proteins (MaxPlax Lambda Packaging Extracts); subsequently, they were adsorbed and transferred into EPI300-T1R cloning host cells. The fosmid clones were plated on LB agar containing chloramphenicol (12.5 μg/ml). After overnight incubation at 37°C, each plate of transformants were washed by 20% glycerol and stored in 1.5-ml microfuge tube. Pooled transformants were labeled and kept at -80°C for further studies.

3.7 Screening of the metagenomic library

Metagenomic library was screened for clone(s) carrying gene involving in biosynthesis of bioactive compounds.

3.7.1 Functional-based screening

For functional-based approaches, pigment production, hemolytic activity, and antimicrobial activity were observed. Many pigments were reported to be compound with biological activities, e.g. red pigment-antibiotic from Egyptian soil sample (Hussein & El-Gammal, 1976), yellow pigment with antibiotic activity (Selvameenal *et al.,* 2009), or the red aminoglycoside antibiotic "violamycin" (Fleck *et al.*, 1974), while the hemolytic activity may indicate the present of eukaryotic cell disrupting agents. The activities against microorganisms were tested on *C. albicans* ATCC 90028, *Serratia marcescens* ATCC 27117, and *B. subtilis* Cm^{R}. These test strains were the representative of fungal, gram negative, and positive bacterial cells, respectively. All test strains were chloramphenicol resistant strain.

3.7.1.1 Pigment production and hemolytic activity

The fosmid library clones at about 10^3 CFU/ml were plated on LB_{Cm} and LB_{CmAr} for observation of pigment production every day up to 6 days after incubation at 37 ºC. In the same way, the fosmid library clones at about 10^3 CFU/ml were plated on 7% sheep blood agar plate supplemented with 12.5 µg/ml chloramphenicol and 0.01% (w/v) arabinose to be observed for hemolytic zone every day up to 6 days after incubation at 37 ºC.

3.7.1.2 Preparation of test microorganisms

B. subtilis Cm^R was cultivated on LB agar plate containing 12.5 µg/ml of chloramphenicol at 37°C overnight. Then, one colony was inoculated into tryptic soy broth and incubated at 37°C and shaking at 200 rpm for 2-3 h. In the same way, *S. marcescens* fresh cell suspension was prepared from an overnight single colony inoculum incubated at 37°C and shaking at 200 rpm for 2-3 h. *C. albicans* ATCC 90028 was grown on SDA at 30 ^oC for 24 h and suspended in 0.85% sodium chloride solution.

The turbidity of microbial cell suspension was adjusted to 50% T at 580 nm. Sequentially, the soft agar seed media of each microbial inocula were prepared. Each bacterial inoculum was inoculated in a final concentration of 1% (v/v) into molten Mueller Hinton agar (0.5% agar). Yeast inoculum was added into two flasks of molten SDA (0.5% agar) containing subinhibitory concentration of ketoconazole (0.125 µg/ml) and equivalent amount of dimethylsulphoxide used to dissolve ketoconazole. These seed media were used for antimicrobial activity screening. Seed medium containing ketoconazole was used to screen for compound which had synergistic activity with ketoconazole against *C. albicans*.

3.7.1.3 Antibacterial and antifungal activity Screening

Each pool of metagenomic library was serially diluted in 0.85% sodium chloride solution to about 10^3 CFU/ml. A 100 µl of diluted pool was spread on LB_{CmAr} agar plate using sterile glass beads. This was done in 20 replicas per pool. After incubation for 5 days at 25°C, the fosmid clones were tested for

antibacterial and antifungal activities, 5 replicas of each test strains. Five milliliters of prepared seed media (as described in section 3.7.1.2) were gently spread on each library plate. The overlaid plates were incubated at 37°C for overnight. Inhibition zone was observed under colony counter magnifying glass. Clones with inhibition zone were selected and streak for isolation on LB_{Cm}. Finally, they were cross tested with all of test organisms for the confirmation of antimicrobial activities (Gillespie *et al.*, 2002).

3.7.2 Sequence-based analysis

Sequence-based analysis is the analysis of metagenomic library clones mostly based on the similarity of their DNA sequences captured in the library to the sequences on public database. This type of analysis, the gene of interest usually must be, at least in part, submitted on the database to be used as reference for DNA probe or primer design.

To locate the type I-PKS KS gene fragment from metagenomic library using PCR, degenerate primers for KS gene fragments were used to screen them from the library pools. The recombinant fosmids from each fosmid Klongkone metagenomic pools were extracted using FosmidMax® according to the manufacturer's instruction. The extracted recombinant fosmid was then used as DNA template for type I-PKS KS gene PCR amplification using MDPQQRf/HGTGTr primers as described in section 3.5.1. Type I-PKS KS positive pool with the PCR product size about 700 bp were serially diluted to about 10^3 CFU/ml and plated on LB_{Cm} plates. After overnight incubation, at least 500 library clones per pool were replicated on LB_{Cm}, and LB_{CmAr} plates, 50 clones per sub-pool plate (in 5×10 matrix). After overnight incubation, sub-pool replicas on LB_{Cm} plates were kept refigerated, and those on LB_{CmAr} plates were used for colony PCR. Colonies on a sub-pool plate were collected and resuspended together in sterile water, and then they were boiled for 5 min. Their cell lysate was used as PCR template. After the type I-PKS KS gene specific PCR, the refrigerated sub-pool plate corresponding to the sub-pool with positive PCR reaction was further replicated on one LB_{Cm}, and two LB_{CmAr} plates. Next morning, sub-pool on the LB_{Cm} plate was kept as template in the refrigerator,

and ones on LB_{CmAr} plates were used for another round PCR screening. On one LB_{CmAr} replica plate, the clones were collected by their 10 columns sub-pools. Another LB_{CmAr} replica plate, the clones were collected by their 5 rows sub-pools. Each sub-pool was resuspended in sterile water and boiled for 5 min. Supernatants of each cell lysate were screened for type I-PKS KS gene by PCR again. The matrix of the PCR results from row and column sub-pools was drawn to pick out the positive clone carrying type I-PKS KS gene. The PCR of the positive recombinant fosmid clone was repeated to confirm the result. The extracted recombinant fosmid was sequenced using the MDPQQRf and HGTGTr as sequencing primers.

3.8 Sequencing and further analysis of the recombinant clones with putative bioactive compound genes

The recombinant clones which had been found to carry metagenomic fragments and give at least one positive result from the tests above, are subjected to the complete sequencing of the insert fragment.

3.8.1 Primer walking

As the conventional capillary base sequencing, at best, it could generate about 700 good quality base pair reads at a run. To sequence the whole insert DNA fragment of fosmid cloning system capable of holding up to about 40 kb, many round of sequencing would be needed. The sequencing result of the recombinant clone using type I-PKS KS gene specific as primers was used as core sequence for designed the outward primers, i.e. 5’ out primer and 3’ out primer, for sequencing further DNA sequences beyond the PCR targeting sequence. Each of the returned sequencing results were used for designing other new outward sequencing primers. In the same time, the end-sequencing from both end of the insert site started with the vector primers, pCC1™/pEpiFOS™f and pCC1™/pEpiFOS™r, was parallel performed. The new primers designed from these sequencing results were named 5’ in or 3’ in, as they were designed for sequencing from the vector ends inward the insert DNA sequence. Supposedly, these processes were repeated until the whole insert DNA was completely sequenced. Sequentially, the insert sequence was analyzed and annotated using the bioinformatics tools described above.

3.8.2 Next generation sequencing

As mentioned above the limitation of conventional capillary base sequencing, at best, it could generate about 700 good quality base pair reads at a run. Many round of sequencing were needed to sequence the whole insert DNA fragment of fosmid cloning system capable of holding up to about 40 kb. These processes were laborious and time consuming without the guaranteed results. Next generation was one of the best alternative routes. There are many platforms for next generation sequencing; one that available at hand was GS Junior 454 sequencing from Roche. The technologies were based on emulsion PCR and pyrosequencing based reaction. The sequencing procedure was done as instructed by manufacturer guideline. In brief, the DNA sample, the recombinant fosmid, was shear into about 500-1000 bp pieces by high speed nitrogen shearing force in the nebulization step. The sheared DNA fragments were end-repaired and ligated with binding and sequencing adapters. After each DNA fragments was bound to sequencing beads in one to one ratio, the water in oil emulsion PCR was proceeded. The beads filled with DNA fragment PCR products, the template, were loaded into pico-titer plate one bead per well. After the pyrosequencing enzyme mixes was filled into the wells, the sequencing was performed and the results were recorded on the GS-Junior sequencing system. The obtained sequencing raw data was assembled into a contiguous sequence using Geneious (v 5.4.6) (Drummond *et al.,* 2010). Consequencely, the consensus sequence, was searched for type I-PKS KS conserved motifs. As the results, the boundary of the ORF containing all of the conserved motifs was searched and annotated. The ORFs within the insert DNA was detected using bioinformatic software Geneious (v.5.4.6). The predicted ORFs were used as query sequences for blastx. Each ORF was putatively annotated for their gene function as obtained from its top hit blast results. For those that were putatively annotated as genes related to PKS were further investigated for their novelty.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Mangrove mud samples and metagenomic DNA preparation

The environmental sample using as the source of metagenomic DNA in this study was mangrove forest mud at the depth of 0-10 centimeters from surface collected from Klongkone mangrove forest, Samut Songkhram, Thailand. This mangrove forest was part of the "mangrove reforestation scheme", and was well studied and documented under the "Ecological Assessment of Mangrove Reforestation/Afforestation in Samut Songkhram Province on Zooplankton and Benthic communities" research project during 1994-2002. As indicated in the report, the reforestation successfully reintroduced the stable mangrove ecosystem and brought up its highest food web complexity and diversity richness. In addition, it also reported that the most of soil organic level was found on the top surface and was higher during the rainy season when, also, the highest primary production was observed (Paphavasit *et al*., 2002). These were the reasons for collecting the sample during the rainy season, besides the convenient access to the mud/soil during the day-time and low-tide with the simplest means and tools required.

After the sample was collected, the collected mangrove mud/soil cores were observed and shown to be a neutral pH by pH test strips (Merck), both at the sampling site and at the laboratory. The 10 centimeters portion from surface-end of the cores (Fig. 4.1 a) were gathered and mixed before sieved to remove any parts (or whole) of animal or plant that might be in the soil, and used for the DNA extraction as described in Material and Methods.

In this study, high molecular weight DNA was obligated; however, the most diverse DNA types were also desired. Thus, the direct DNA extraction and electroelution purification according to Brady (2007) was adopted with some modification as described in Material and Methods. As the soil sample was rich with soil organic matters, the obtained crude DNA was in dark brown color (Fig. 4.1 b-d). After the PFGE and electroelution, purified DNA was obtained, however, with the unexpected smaller size than what had been anticipated. The obtained purify DNA

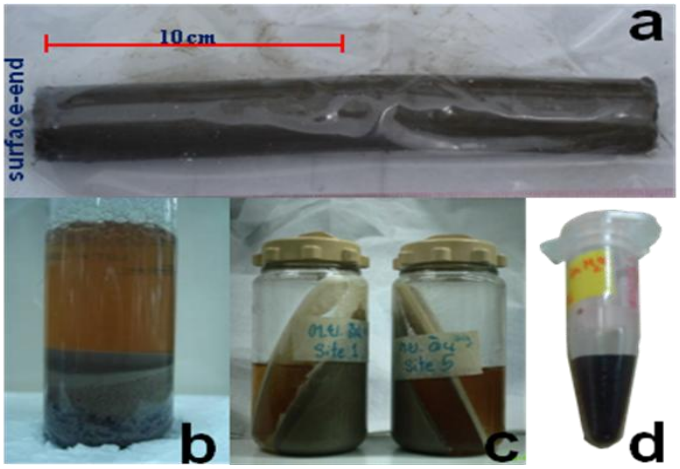


Figure 4.1 A mangrove mud core and its extracted metagenomic DNA solution. (a) Mangrove mud/soil core collected from Klongkone mangrove forest, Samut Songkhram, Thailand. The length-bar displayed the depth of the core to be used for DNA extraction. (b and c) The brownish supernatant containing the DNA obtained after extraction. (d) A dark brown crude DNA obtained after a series of alcohol precipitation and concentration.

from 36 to 50 kb DNA embedding gel noodle (Fig. 4.2 b dashed box with **) showed only the size between 10 kb up to 23 kb (Fig. 4.3 a). In the same way, majority of 30 to 50 kb purified DNA (Fig. 4.3 b) was obtained from agarose gel noodle embedded DNA of the size from 50 to over 95 kb (Fig 4.2b dashed box with **). This size reducing of the DNA could have been due to the shearing from isolation of DNA from gel processes. Pipetting transfer of DNA solution is well known cause for high molecular weight DNA shearing (Sambrook, 2001). Another possibility, the crude DNA might be bound or intercalated by the cationic soil organic substances causing the reduction of the effective charge and slow migration through agarose gel by electromotive force of the electrophoresis (Sigmon & Larcom, 1996; Webb & Ebeler, 2004; Wanunu *et al.*, 2009). The purified DNA, on the other hand, supposedly was free of impurity; thus, this DNA migrated through the gel on its own charge and mass, which apparently shifted to the right position for the correct size.

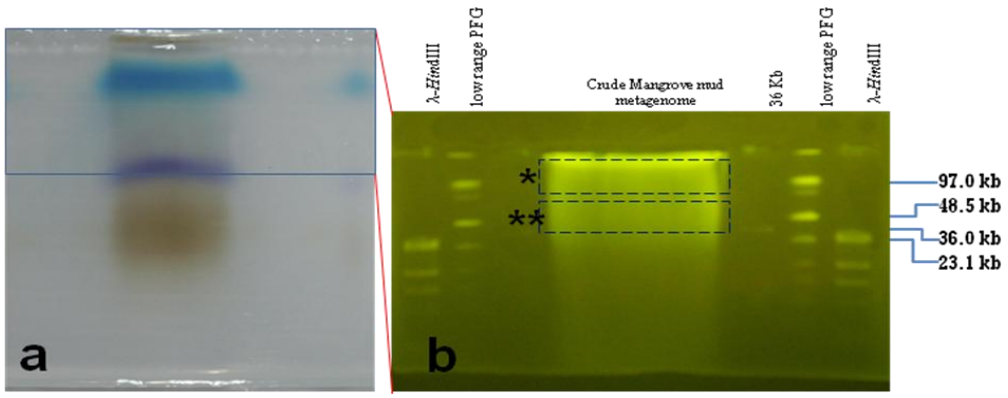


Figure 4.2 Mangrove mud metagenomic DNA separated in 0.8% agarose gel using Pulse field Gel Electrophoresis. (a) The brown substance separated from the high molecular weight DNA above the blue stain band of xylene cyanol dye. (b) The gel portion without the brown substance stained with SYBR Gold and observed under blue-light through amber filter. Gel embedded with DNA in the size of 50 kb to over 97 kb, and the one with DNA about 36-50 kb (dash box with *, and **, respectively) were excised from the gel; and, DNA was extracted from the gel by electroelution.

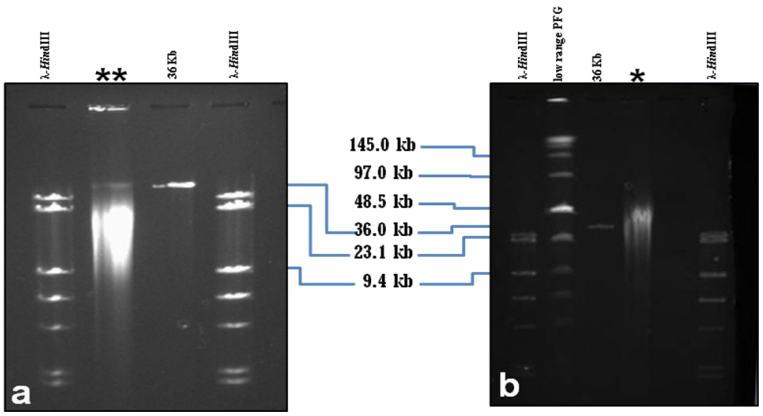


Figure 4.3 Agarose electrophoresis gel showing purified DNA from real-time electroelution. (a) Purified DNA** with the majority sized range between 9 to 23 kb was obtained from agarose gel noodle embedded DNA in the expected size range of 36 to 50 kb. (b) Purified DNA* with the majority sized range between 30 and 50 kb was obtained from agarose gel noodle embedded DNA in the expected size range of over 50 kb.

Nevertheless, the obtained 30-50 kb DNA was the perfect size for fosmid library construction without the need for the shearing step (Epicentre® Biotechnologies). This purified Klongkone mangrove mud metagenomic DNA was dissolved in 50 µl sterile DI water, with concentration about 400 ng/ µl. According to UV spectrophotometry measurement, the $OD_{\lambda 260nm}/OD_{\lambda 280nm}$ of the purified DNA was 1.89 which was acceptable purity with low protein contamination. This purified DNA also exhibited an acceptable $OD_{\lambda 260nm}/OD_{\lambda 230nm}$ ratio, which was 1.48. Although, it was lower than 1.5 that had been reported by LaMontagne (2002), this purified DNA could be successfully used for PCR reaction (in section 4.2), indicating less contamination of humic acid that could impair the downstream processes for fosmid library construction (Wilfinger *et al.*, 1997; LaMontagne *et al.*, 2002). Moreover, the recent study had shown that DNA template with $OD_{\lambda 260nm}/OD_{\lambda 230nm}$ as low as 1.24 could still be used for successful PCR reaction (Fatima *et al.*, 2011). The purified mangrove mud metagenome was kept at 4 ºC until further uses.

4.2 Type I-PKS KS gene diversity and Bacterial diversity held within Klongkone mangrove metagenome

To observed whether Klongkone mangrove mud metagenome was potentially bearing novel gene involving in bioactive compound biosynthesis, and a variety of bacterial taxa, the PCR investigations on the representative genes, type I-PKS KS gene and 16S rRNA gene, were performed.

4.2.1 PCR amplification of type I-PKS KS domain from Klongkone mangrove mud metagenome

Mangrove mud metagenome was sought for type I-PKS KS domain gene with their gene specific primers, MDPQQRf/HGTGTr (Kim and Fuerst, 2006). The forward primer MDPQQRf priming at the conserved motif about 270 bp (90 amino acids) upstream of its cysteine active site, and the reverse primer HGTGTr priming the histidine active site about 430 bp (140 amino acids) downstream from cysteine active site. Genomic DNA of *Streptomyces peucetius* var. *caesius*

TISTR1253 was used as positive control. As shown in Fig. 4.4, the PCR products were 700 bp as expected.

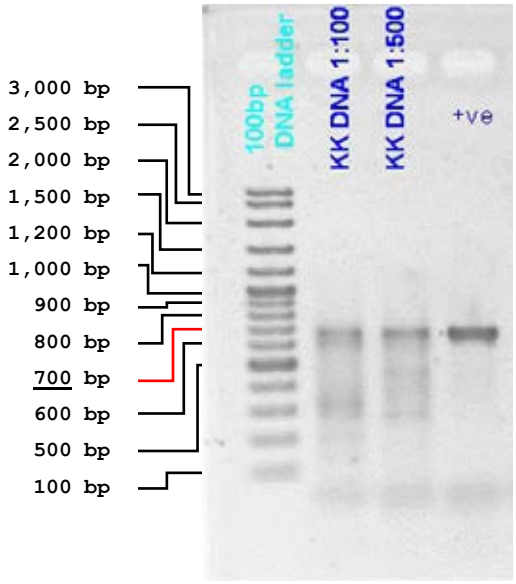


Figure 4.4 Agarose electrophoresis gel showing PCR products obtained from amplification of the partial type I-PKS KS domain gene in mangrove mud metagenome.

The purified mixed amplicons of corresponding size (700 bp) were ligated into pGEM®-T Easy Vector (Promega), and transformed into *E. coli* DH5α, as described in section 3.5.4. One hundred white colonies grown for 18-24 hours on LB_{AmpXGI} agar plate were randomly picked. Slow and uneven color development of blue/ white selection was found, partly due to its instability, light sensitive, and water insoluble of X-Gal, made it difficult to distinguish at the early colony formation (Heuermann and Cosgrove, 2001). After re-streaking the picked colony on LB_{AmpXGI} and keeping them cold after incubation, 45 clones remained white. Only twenty clones were found to yield positive PCR amplification products (700 bp) of partial type I-PKS KS domain, as shown in Fig. 4.5.

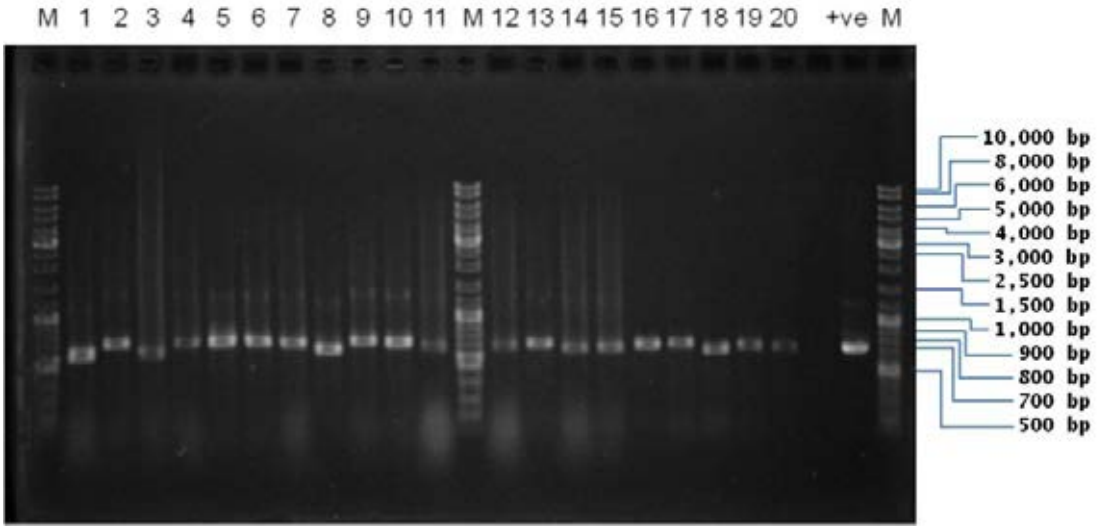


Figure 4.5 PCR products (about 700 bp) obtained from amplification of the partial type I-PKS KS domain gene in the 20 remained white recombinant clones. M: 1 kb ladder; +ve: positive control (*S. peucetius* var. *caesius* TISTR1253).

4.2.2 Sequence analyses type I-PKS KS domain from Klongkone mangrove mud metagenome

The AB1 sequencing files from both directions (forward and reverse) of each clone were contig and their vector portions were trimmed. The insert DNA sequences were then detected for their open reading frames and deduced amino acid sequences. As all these amplicons were partial Type I PKS KS domain protein sequences; thus, at least, one open reading frame of their sequences must be read through in a single frame without any stop codon to encode for the correct partial protein. Additionally, the deduced amino acid sequences were created using the standard genetic code (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi). Thirteen clones, KK_KSI_002, 003, 010, 013, 017, 019, 021, 023, 026, 027, 029, 033, and 035, were obtained with the insert sequences exhibiting, at least one, intact open reading frame without stop codon for the entire DNA insert length. Their deduced amino acid sequences were monitored for the characterizing amino acid conserved motifs for KSs, MDPQQRL, VDTACSSS, and HGTGT. As mentioned in previous chapter, MDPQQR is conserved motif used as the forward primer site; the VDTACSSS is the conserved motif around the cysteine active site (underlined), and

HGTGT, the site for reverse primer, is the amino conserved motif at histidine active site (underlined). These conserved motif features of the type I-PKS KS gene were all present in every deduced amino acid sequences of all inserts of the thirteen clones, some, with few variations (Fig. 4.6 to4.19).

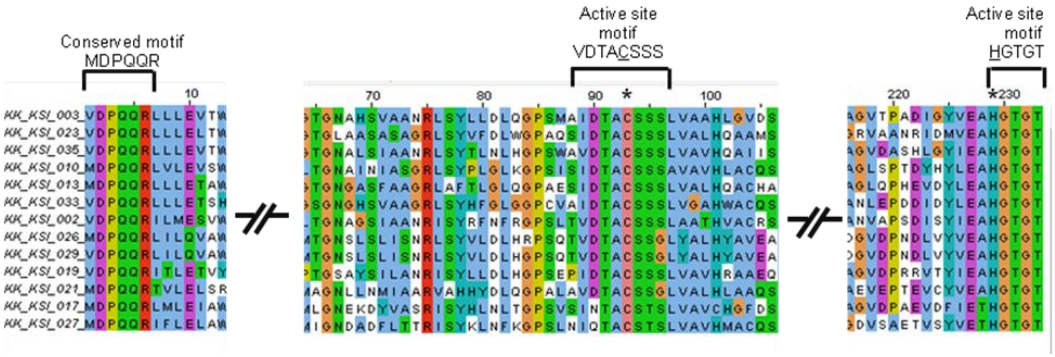


Figure 4.6 The clustalW2 alignment of partial sequence Type I-PKS KSs from mangrove mud metagenome showing the important characteristics of the conserved motifs for Type I-PKS KS proteins, the cysteine (*) active site motif, VDTACSSS. (The enlarged figure can be seen on Fig. B18)

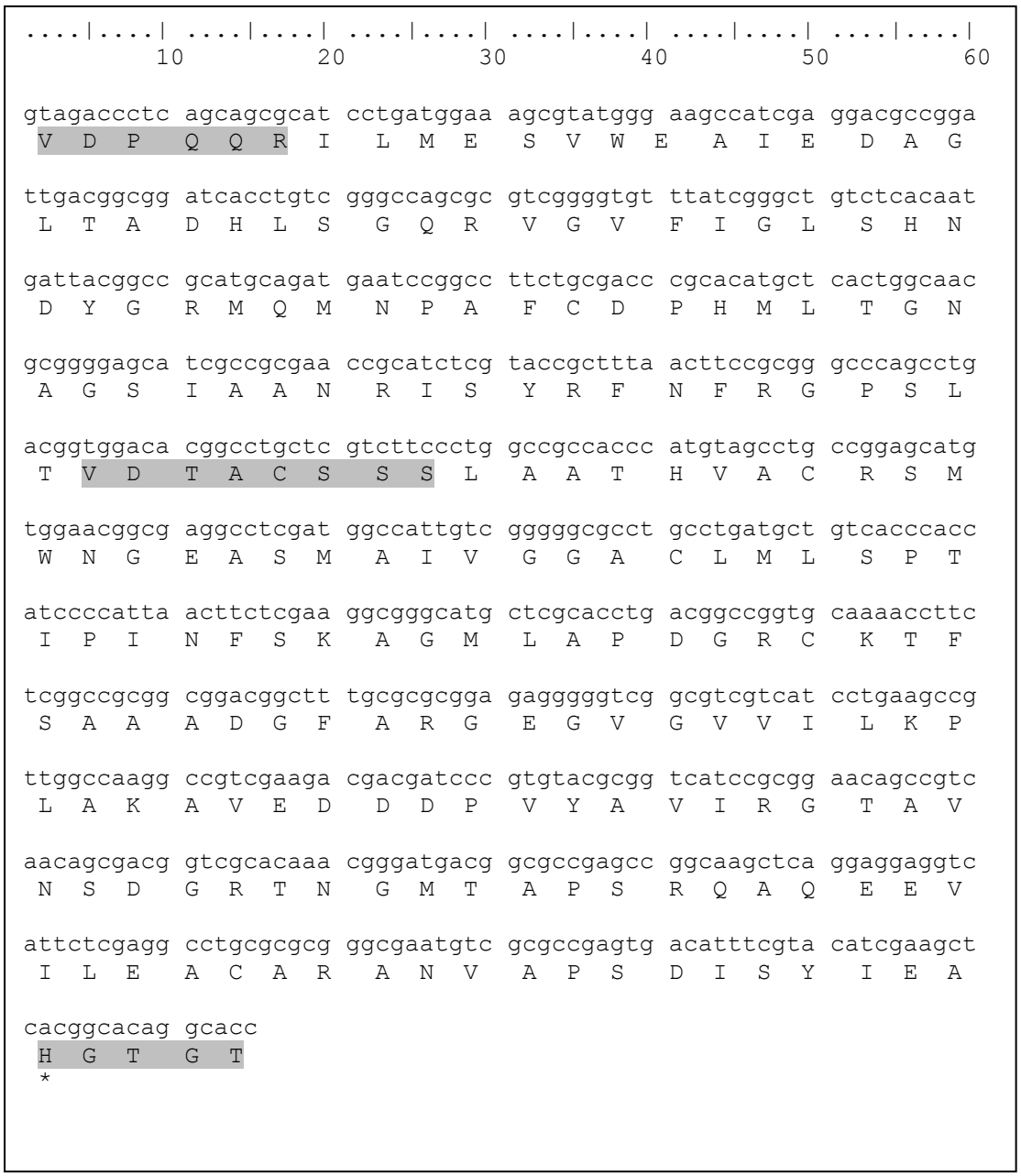


Figure 4.7 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_002. This sequence contained 675 base pairs with 124 a, 219 c, 219 g, and 113 t. The type I-PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtggatccac agcagcgact gcttttggag gtaacctggg aagcgctgga acgatcggga
 V  D  P   Q  Q  R  L   L  L  E   V  T  W   E  A  L  E   R  S  G

cagaacccgt ctgccttggc gggcagcgcc accggcgtct tcatcggcat cagcagttcc
 Q  N  P   S  A  L  A   G  S  A   T  G  V   F  I  G  I   S  S  S

gattattcgc ggctgcagtt cagacatccc ggcttgattg atgcttatgc cgggacagga
 D  Y  S   R  L  Q  F   R  H  P   G  L  I   D  A  Y  A   G  T  G

aacgcgcaca gtgtggccgc caaccggctt tcatacctgc ttgatctaca agggcccagc
 N  A  H   S  V  A  A   N  R  L   S  Y  L   L  D  L  Q   G  P  S

atggccatcg acaccgcctg ctcgtcctcg ctggtggctg cacacctggg cgtggacagc
 M  A  I   D  T  A  C   S  S  S   L  V  A   A  H  L  G   V  D  S

ttgcgcagtg gtgaggtcga tctggcactg gtcggcggcg tgaacctgct cctggcacct
 L  R  S   G  E  V  D   L  A  L   V  G  G   V  N  L  L   L  A  P

gatttaacca tcacttttc ccaggcacag atgatggccc ctgacggccg ctgcaaaact
 D  L  T   I  T  F  S   Q  A  Q   M  M  A   P  D  G  R   C  K  T

ttcgacgcgc gggctgacgg ttatgtccgc ggcgaaggct gtggcgtgat tgttttgaag
 F  D  A   R  A  D  G   Y  V  R   G  E  G   C  G  V  I   V  L  K

cgcctctccg atgcgcagcg cgacggcgac ccgattctgg ccttgctgcg aggttccgct
 R  L  S   D  A  Q  R   D  G  D   P  I  L   A  L  L  R   G  S  A

gtgaaccaga acggccgttc caatggtctc accgccccca acggtctggc ccagcaggca
 V  N  Q   N  G  R  S   N  G  L   T  A  P   N  G  L  A   Q  Q  A

gtcattcgcc gtgccctggc taatgccggg gtaacccctg ccgatatcgg ctacgtcgag
 V  I  R   R  A  L  A   N  A  G   V  T  P   A  D  I  G   Y  V  E

gcccatggca ccggtacc
 A  H  G   T  G  T
```

Figure 4.8 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_003. This sequence contained 678 base pairs with 117 a, 215 c, 207 g, and 139 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggatcctc agcagcgtct cgtgctcgaa gtcagctggg aggcgctcga gcgggcaggg
 M  D  P   Q  Q  R  L   V  L  E   V  S  W   E  A  L  E   R  A  G

atctccgccc ggaccctcga tgggagccct acaggtgtct acctcggcat cacctacacc
 I  S  A   R  T  L  D   G  S  P   T  G  V   Y  L  G  I   T  Y  T

gactaccaga agctgccgcg gttcggggag gatctccgcc gcatcacccc ccacaccttg
 D  Y  Q   K  L  P  R   F  G  E   D  L  R   R  I  T  P   H  T  L

acgggtaacg cgataaacat cgcttcgggg cggctatcct acccccctcgg gctcaaggga
 T  G  N   A  I  N  I   A  S  G   R  L  S   Y  P  L  G   L  K  G

ccgagcatct cgatcgacac cgcctgctcg tcatcggcgg tcgcagtcca cctggcttgt
 P  S  I   S  I  D  T   A  C  S   S  S  A   V  A  V  H   L  A  C

cagagcctac gggcagacga gtccgaactt gccctcgccg ggggtgtgaa cctgatcctc
 Q  S  L   R  A  D  E   S  E  L   A  L  A   G  G  V  N   L  I  L

tcgcccgagg tgatgatctg cctgagcagc acccaggccc tctctccggg agatcgttgc
 S  P  E   V  M  I  C   L  S  S   T  Q  A   L  S  P  G   D  R  C

cgggttttcg acgcgggagc agacgggttc gtgcgcggtg aaggctgcgg cgtactggtg
 R  V  F   D  A  G  A   D  G  F   V  R  G   E  G  C  G   V  L  V

ctcaagcgct tcgccgacgc ggtccgtgat ggcgatcgtg tcctcgcttt gattcgctca
 L  K  R   F  A  D  A   V  R  D   G  D  R   V  L  A  L   I  R  S

acggcggtca accaagacgg caggagcagc cggctgaccg cccccaacgg gcgtgcccag
 T  A  V   N  Q  D  G   R  S  S   R  L  T   A  P  N  G   R  A  Q

gaggcggtga tagcgcgggc gctcgaggac gccggcctga gcccaacgga ctatcactac
 E  A  V   I  A  R  A   L  E  D   A  G  L   S  P  T  D   Y  H  Y

ctggaggcgc acggcaccgg gact
 L  E  A   H  G  T  G   T
```

Figure 4.9 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_010. This sequence contained 684 base pairs with 116 a, 231 c, 222 g, and 115 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggatcctc agcagcgact actgttggag accgcttggg aggccctaga gtcagctgcc
 M  D  P   Q  Q  R  L   L  L  E   T  A  W   E  A  L  E   S  A  A

ttatccccgc aagggttgca cggttcgcgg accggcgtct ttgtggggat taccaaaacg
 L  S  P   Q  G  L  H   G  S  R   T  G  V   F  V  G  I   T  K  T

gattacgccc agttggcgat gttggactac gagcagatcg acgtctacgc tggaaccggc
 D  Y  A   Q  L  A  M   L  D  Y   E  Q  I   D  V  Y  A   G  T  G

aatggcgcgt cgttcgctgc tgggaggctg gctttcaccc tcggactgca agggccggcc
 N  G  A   S  F  A  A   G  R  L   A  F  T   L  G  L  Q   G  P  A

gagtcgattg acacggcttg tagctcctcg ctggtggctc tgcaccaggc ctgccacgct
 E  S  I   D  T  A  C   S  S  S   L  V  A   L  H  Q  A   C  H  A

gtccgcgacg gccagtgcga gacggccttg gtcggcggcg ttcatctgat gttgacgccg
 V  R  D   G  Q  C  E   T  A  L   V  G  G   V  H  L  M   L  T  P

gccagcacca tatttctttc tcgtgcgaag gcccttgcgc ccgatggacg ctgcaagaca
 A  S  T   I  F  L  S   R  A  K   A  L  A   P  D  G  R   C  K  T

ttcgacgcca gcgcggacgg gttcgcgcgg gcggaaggtt gtggcgtctt ggtcttgaaa
 F  D  A   S  A  D  G   F  A  R   A  E  G   C  G  V  L   V  L  K

aaactggctg acgccgagcg tcacggcgat cggattctgg ctgtgattcg cgggacggcc
 K  L  A   D  A  E  R   H  G  D   R  I  L   A  V  I  R   G  T  A

gtcaaccacg atgggccgag cagcggcctt accgtaccca acggtccggc gcagcaacaa
 V  N  H   D  G  P  S   S  G  L   T  V  P   N  G  P  A   Q  Q  Q

gtgattcggg acgcgttggc tcgagcgggc ctccagcctc acgaagtcga ctacttagaa
 V  I  R   D  A  L  A   R  A  G   L  Q  P   H  E  V  D   Y  L  E

gcccacggca ccggtacc
 A  H  G   T  G  T
```

Figure 4.10 Nucleotide sequence and its deduced amino acid sequence of partial type I PKS KS gene from KK_KSI_013. This sequence contained 678 base pairs with 118 a, 205 c, 223 g, and 132 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggacccac agcagcggct catgcttgag atcgcctggg aggcgctcga agatgcggga
 M  D  P   Q  Q  R  L   M  L  E   I  A  W   E  A  L  E   D  A  G

tacccgcccg acgcgatcga ggggctggtc ggggtcttcg ccgggatggg caacaacgcc
 Y  P  P   D  A  I  E   G  L  V   G  V  F   A  G  M  G   N  N  A

tacttccccg cgaacctgcg gagccgcccc gatgtcgttc gctctgccgg cgaactgcag
 Y  F  P   A  N  L  R   S  R  P   D  V  V   R  S  A  G   E  L  Q

acgatgttgg ggaacgagaa ggactatgtg gcgtcccgca tctcctatca cctgaacctc
 T  M  L   G  N  E  K   D  Y  V   A  S  R   I  S  Y  H   L  N  L

acgggcccca gcgtcagcat caacactgcg tgctcgacgt cattggtcgc ggtatgccac
 T  G  P   S  V  S  I   N  T  A   C  S  T   S  L  V  A   V  C  H

ggcttcgaca gtcttctcaa ctaccagtgc gaccttgcgt tggcaggagg gctcacggtt
 G  F  D   S  L  L  N   Y  Q  C   D  L  A   L  A  G  G   L  T  V

caccttcctc agaggtcggg ctacctctgg cgggaggtca tgatcttctc gaaggacggt
 H  L  P   Q  R  S  G   Y  L  W   R  E  V   M  I  F  S   K  D  G

cactgccggc cgttcgatgc ggcggcgagt ggaaccgtcg ccagcaacgg cggtggcctc
 H  C  R   P  F  D  A   A  A  S   G  T  V   A  S  N  G   G  G  L

gttgtcctga agcgcctcga ggacgcggtt gccgacggcg accgcatcta tggggtgatt
 V  V  L   K  R  L  E   D  A  V   A  D  G   D  R  I  Y   G  V  I

cgcggaacgg gcgtcaacaa cgacggttcc aacaagatga gcttcatggc ccccagcgtg
 R  G  T   G  V  N  N   D  G  S   N  K  M   S  F  M  A   P  S  V

gagggtcaga cggccgccat cgccacggcg ctggccgaag cgggagtcga tcccgcagag
 E  G  Q   T  A  A  I   A  T  A   L  A  E   A  G  V  D   P  A  E

gtcgatttca tcgagaccca cggcaccggc acc
 V  D  F   I  E  T  H   G  T  G   T
```

Figure 4.11 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_017. This sequence contained 678 base pairs with 125 a, 221 c, 227 g, and 120 t. The type I PKS KS important amino acid conserved motifs are highlighted. Signature motifs for hybrid PKS/NRPS were highlighted and underlined.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtggatcctc agcagcgcat caccctcgag accgtttatc atgctctgga agatgctgct
 V  D  P   Q  Q  R  I   T  L  E   T  V  Y   H  A  L  E   D  A  A

ctcgcgccgc agtccatcaa aggcagcgac accggcatct tcataggtgt gtccacctgg
 L  A  P   Q  S  I  K   G  S  D   T  G  I   F  I  G  V   S  T  W

gactatcgta atctggttca tctgcacccg gaacaaaaca gccaggccca ggttcccacc
 D  Y  R   N  L  V  H   L  H  P   E  Q  N   S  Q  A  Q   V  P  T

ggttcggcct attccatttt ggcgaaccgc atttcttacc tgctggacct ccacggaccc
 G  S  A   Y  S  I  L   A  N  R   I  S  Y   L  L  D  L   H  G  P

agtgagccca tcgatacggc ctgctcttct tccctggtcg ccgtccaccg cgccgcggaa
 S  E  P   I  D  T  A   C  S  S   S  L  V   A  V  H  R   A  A  E

cagatccggc tcggcaattg ttcgatggcg atcgttggcg gggtcaatgc cctgcttacc
 Q  I  R   L  G  N  C   S  M  A   I  V  G   G  V  N  A   L  L  T

cccgaactga gtcgctcctt ccaccaggca ggcatgttga gtgaggacgg ccgctgcaaa
 P  E  L   S  R  S  F   H  Q  A   G  M  L   S  E  D  G   R  C  K

acctttgatc aaagcgccaa cggttacgtg cgtggtgaag gtgtaggcat ccttatcctc
 T  F  D   Q  S  A  N   G  Y  V   R  G  E   G  V  G  I   L  I  L

aagccgctca gccgggccga agccgacggc gaccgaatct acggactcat tctgggtacg
 K  P  L   S  R  A  E   A  D  G   D  R  I   Y  G  L  I   L  G  T

gctgaaaacc acggcggccg ggccaatacc ctgacctcgc ccaacccgca ggcacaaaaa
 A  E  N   H  G  G  R   A  N  T   L  T  S   P  N  P  Q   A  Q  K

gaactgctcc tcaaagccta ccgccaggcg ggtgtagatc cccgccgggt cacctacatc
 E  L  L   L  K  A  Y   R  Q  A   G  V  D   P  R  R  V   T  Y  I

gaagcccatg gcaccggcac c
 E  A  H   G  T  G  T
```

Figure 4.12 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_019. This sequence contained 681 base pairs with 140 a, 232 c, 179 g, and 130 t. The type I PKS KS important amino acid conserved motifs are highlighted. Proline residue at the cysteine active site motif was in red and underlined. This residue appeared only on trans-AT hybrid PKS/NRPS clade (Fig.4.22).

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggatcctc agcagcgaac ggtgctggaa ctgtcgcgcg atctgctcga gcccctggcc
 M  D  P   Q  Q  R  T   V  L  E   L  S  R   D  L  L  E   P  L  A

ggcgagcgcg acatcggcgt gttcgtgggc gccggaaacc acgcctacag cgaggcgatc
 G  E  R   D  I  G  V   F  V  G   A  G  N   H  A  Y  S   E  A  I

tccgcgcacc tcggtgagcc gctgcatccc aacgccatgg cgggcaacct gctcaacatg
 S  A  H   L  G  E  P   L  H  P   N  A  M   A  G  N  L   L  N  M

atcgcggcgc gggtggcgca ccactacgac ctgcagggcc ccgcgttggc cgtggacacc
 I  A  A   R  V  A  H   H  Y  D   L  Q  G   P  A  L  A   V  D  T

gcgtgcagct cgggcctggt cgccctgcac ctcgcggccc agagcctggc caccggcgag
 A  C  S   S  G  L  V   A  L  H   L  A  A   Q  S  L  A   T  G  E

tgccgctacg ccatcgccgg cggcgtgcac ctcaacctca cgcccgccca acaccagctg
 C  R  Y   A  I  A  G   G  V  H   L  N  L   T  P  A  Q   H  Q  L

ttcgacaacg caggagccct gtcgccgacc ggccagtgcc gcccgttcca tcccgatgcc
 F  D  N   A  G  A  L   S  P  T   G  Q  C   R  P  F  H   P  D  A

gacggcatgg tccccggtga aggcaccgtg ctgttcctgc tccagccagc agacgccgcc
 D  G  M   V  P  G  E   G  T  V   L  F  L   L  Q  P  A   D  A  A

cgggccgagg gaagagcccc catcggcatc ctccgggcgt cggccatcaa caacgacggc
 R  A  E   G  R  A  P   I  G  I   L  R  A   S  A  I  N   N  D  G

accagcctcg gggtgatggc gcccaacccc gcggggcagg aggcggtcat ccgccgcgcc
 T  S  L   G  V  M  A   P  N  P   A  G  Q   E  A  V  I   R  R  A

ctgcgccagg cggaggtgga gcccaccgag gtctgctacg tcgaggccca cggcaccgga
 L  R  Q   A  E  V  E   P  T  E   V  C  Y   V  E  A  H   G  T  G

acc
 T
```

Figure 4.13 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_021. This sequence contained 663 base pairs with 105 a, 259 c, 217 g, and 82 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtggatcctc agcagcgcct cttgctggag ctcacctggg acgccctggg tcaggcagga
 V  D  P   Q  Q  R  L   L  L  E   L  T  W   D  A  L  G   Q  A  G

attgtaccta cgtcgctcgg gggcagtccg actggcgtct acatcggcat ggcctcgggc
 I  V  P   T  S  L  G   G  S  P   T  G  V   Y  I  G  M   A  S  G

gactatggca aactggcgtc ggcgtcagcg cctgctaacg cctatacagg gacgggcctg
 D  Y  G   K  L  A  S   A  S  A   P  A  N   A  Y  T  G   T  G  L

gcggcgagcg ccagcgccgg tcggttgtcg tacgtgtttg acctatgggg cccggctcag
 A  A  S   A  S  A  G   R  L  S   Y  V  F   D  L  W  G   P  A  Q

tcgattgaca cggcttgctc gtcgtccttg gtggctctgc atcaggcggc catgagttta
 S  I  D   T  A  C  S   S  S  L   V  A  L   H  Q  A  A   M  S  L

cgctacggcg agaccgatct cgcgctcgtc gccggtgtga acgcgatgct gttagccgac
 R  Y  G   E  T  D  L   A  L  V   A  G  V   N  A  M  L   L  A  D

acaaccgtgg cgttttcgca ggctcgcatg ctgagcgggg atggttgctg caagaccttt
 T  T  V   A  F  S  Q   A  R  M   L  S  G   D  G  C  C   K  T  F

gacgcacggg ctgatggcta tgtgcgaagc gagggttgcg gcgtcatggt cttgcagcgt
 D  A  R   A  D  G  Y   V  R  S   E  G  C   G  V  M  V   L  Q  R

ggcagggatg cgcggcgcga tggcaaccgg ccgttagccc tcgtcgtagg aacggcggtc
 G  R  D   A  R  R  D   G  N  R   P  L  A   L  V  V  G   T  A  V

aatcaggacg gtcgaagcca agggctgacg gcgcccaacg ggctgagcca gcaggcggtc
 N  Q  D   G  R  S  Q   G  L  T   A  P  N   G  L  S  Q   Q  A  V

gtggggcaag cattggcgaa cgggagggta gcggcgaacc gtatcgatat ggtcgaggcg
 V  G  Q   A  L  A  N   G  R  V   A  A  N   R  I  D  M   V  E  A

cacggcaccg gtacc
 H  G  T   G  T
```

Figure 4.14 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_023. This sequence contained 675 base pairs with 110 a, 192 c, 247 g, and 126 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggaccctc agcagcgcct aatccttcaa gtggcgtggg aggcgttggg gcatgcgggc
 M  D  P   Q  Q  R  L   I  L  Q   V  A  W   E  A  L  G   H  A  G

ctgccgcctt cttcgctgcc gaaggatcgc accggcgtct atatcggcgc ctccagctcc
 L  P  P   S  S  L  P   K  D  R   T  G  V   Y  I  G  A   S  S  S

gattactcga accgtttcta tctcgacccg gcgtccatcg acagccagtt catgaccggc
 D  Y  S   N  R  F  Y   L  D  P   A  S  I   D  S  Q  F   M  T  G

aattccctga gtcttatctc caaccgactt tcctacgtgc ttgatctgca taggccgagc
 N  S  L   S  L  I  S   N  R  L   S  Y  V   L  D  L  H   R  P  S

cagacggttg acacagcatg ctcctccggc ctttacgccc tccactacgc ggtggaggcg
 Q  T  V   D  T  A  C   S  S  G   L  Y  A   L  H  Y  A   V  E  A

ttgaagtccg gacggatcga caccgccata gtcggcgcgg tcaacatgct gctctcgccc
 L  K  S   G  R  I  D   T  A  I   V  G  A   V  N  M  L   L  S  P

tttcctttcg tcggtttttc ccgcgcctcg atgctgtcga agaagggcct ctgccgcgct
 F  P  F   V  G  F  S   R  A  S   M  L  S   K  K  G  L   C  R  A

ttcgacgccg acggcgacgg gtatgtgcgc tcggagggcg ccgccgtctt cgtgcttcgc
 F  D  A   D  G  D  G   Y  V  R   S  E  G   A  A  V  F   V  L  R

gccgagcatg tggcgatcgc cgagggcgac cggatccgcg gctacgtggc ggcgaccggc
 A  E  H   V  A  I  A   E  G  D   R  I  R   G  Y  V  A   A  T  G

gtcaacaccg acggccgcac gcccggggtc tcgcagccga gcgccgaccg gcaggcggcg
 V  N  T   D  G  R  T   P  G  V   S  Q  P   S  A  D  R   Q  A  A

ctgctgcgct cgatctaccg cgaggacggc gtcgatccga acgacctcgt ctacgtcgag
 L  L  R   S  I  Y  R   E  D  G   V  D  P   N  D  L  V   Y  V  E

gcccacggca ccggaact
 A  H  G   T  G  T
```

Figure 4.15 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_026. This sequence contained 678 base pairs with 103 a, 238 c, 207 g, and 130 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
atggatcctc agcagcgtat ctttcttgaa cttgcctggg aagcacttga gtcagccgga
 M  D  P   Q  Q  R  I   F  L  E   L  A  W   E  A  L  E   S  A  G

tacaaccctg atacgtatga tggcctcatt ggggtttttg caggtaccag tggcaatgat
 Y  N  P   D  T  Y  D   G  L  I   G  V  F   A  G  T  S   G  N  D

tatcgtaaaa actttgccgc aaatcagtta agcatctcat caggaatgga atcatttgag
 Y  R  K   N  F  A  A   N  Q  L   S  I  S   S  G  M  E   S  F  E

atgatgatag gcaatgatgc agattttctg acaacgcgca tttcatacaa gttaaacttt
 M  M  I   G  N  D  A   D  F  L   T  T  R   I  S  Y  K   L  N  F

aaaggaccaa gccttaatat acagactgca tgttccactt cacttgtggc agtacacatg
 K  G  P   S  L  N  I   Q  T  A   C  S  T   S  L  V  A   V  H  M

gcgtgtcaga gcctgctgac gtatcaaagc gacatggcca tggccggagg tatatgtatc
 A  C  Q   S  L  L  T   Y  Q  S   D  M  A   M  A  G  G   I  C  I

aggtttccgc agggccacgg ttacatgtac caggaaggca tgatctggtc gccggatgga
 R  F  P   Q  G  H  G   Y  M  Y   Q  E  G   M  I  W  S   P  D  G

cattgccggc cgtttgatgc aaaggcacag ggaacgctcc tggggcaggg tggcgggatc
 H  C  R   P  F  D  A   K  A  Q   G  T  L   L  G  Q  G   G  G  I

gttgttctca agagacttgc tgatgcattg caagatggtg acacagttct tgcaatcata
 V  V  L   K  R  L  A   D  A  L   Q  D  G   D  T  V  L   A  I  I

aaagggtcag caataaacaa tgacggctca atgaaagtag gctttacagc tccaagtgtt
 K  G  S   A  I  N  N   D  G  S   M  K  V   G  F  T  A   P  S  V

gacggccagt ctgaagccat ttccatggcc ctggcactgg gcgatgtttc agctgaaaca
 D  G  Q   S  E  A  I   S  M  A   L  A  L   G  D  V  S   A  E  T

gtcagctatg ttgaaaccca cggcacaggc acg
 V  S  Y   V  E  T  H   G  T  G   T
```

Figure 4.16 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_027. This sequence contained 693 base pairs with 185 a, 155 c, 183 g, and 170 t. The type I PKS KS important amino acid conserved motifs are highlighted. Signature motifs for hybrid PKS/NRPS were highlighted and underlined.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtagaccctc agcagcgcct aatccttcaa gtggcgtggg aggcgttgga gcatgcgggc
 V  D  P   Q  Q  R  L   I  L  Q   V  A  W   E  A  L  E   H  A  G

ctgccgcctt cttcgctgcc gaaggatcgc accggcgtct atatcggcgc ctccagctcc
 L  P  P   S  S  L  P   K  D  R   T  G  V   Y  I  G  A   S  S  S

gattactcga accgtttcta tctcgacccg gcgtccatcg acagccagtt catgaccggc
 D  Y  S   N  R  F  Y   L  D  P   A  S  I   D  S  Q  F   M  T  G

aattccctga gtcttatctc caaccgactt tcctacgtgc ttgatctgca tgggccgagc
 N  S  L   S  L  I  S   N  R  L   S  Y  V   L  D  L  H   G  P  S

cagacggttg acacagcatg ctcctccggc ctttacgccc tccactacgc ggtggaggcg
 Q  T  V   D  T  A  C   S  S  G   L  Y  A   L  H  Y  A   V  E  A

ttgaagtccg gccggatcga caccgccata gtcggcgcgg tcaacatgct gctctcgccc
 L  K  S   G  R  I  D   T  A  I   V  G  A   V  N  M  L   L  S  P

tttcctttcg tcggtttttc ccgcgcctcg atgctgtcga agaagggcct ctgccgcgct
 F  P  F   V  G  F  S   R  A  S   M  L  S   K  K  G  L   C  R  A

ttcgacgccg acggcgacgg gtatgtgcgc tcggagggcg ccgccgtctt cgtgcttcgc
 F  D  A   D  G  D  G   Y  V  R   S  E  G   A  A  V  F   V  L  R

gccgagcatg tggcgatcgc cgagggcgac cggatccgcg gctacgtggc ggcgaccggc
 A  E  H   V  A  I  A   E  G  D   R  I  R   G  Y  V  A   A  T  G

gtcaacaccg acggccgcac gcccggggtc tcgcagccga gcgccgaccg gcaggcggcg
 V  N  T   D  G  R  T   P  G  V   S  Q  P   S  A  D  R   Q  A  A

ctgctgcgct cgatctaccg cgaggacggc gtcgatccga acgacctcgt ctacgtcgag
 L  L  R   S  I  Y  R   E  D  G   V  D  P   N  D  L  V   Y  V  E

gcccacggca ccggaact
 A  H  G   T  G  T
```

Figure 4.17 Nucleotide sequence and its deduced amino acid sequence of partial type I PKS KS gene from KK_KSI_029. This sequence contained 678 base pairs with 102 a, 239 c, 207 g, and 130 t. The type I PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtggacccgc agcagcggct gctgttggag accagccacg aggccctgga agacgccggc
 V  D  P   Q  Q  R  L   L  L  E   T  S  H   E  A  L  E   D  A  G

atggcgcccg accgcgtttc cggccggccg ggcggtgtgt tcgtggggat ctgcgggatc
 M  A  P   D  R  V  S   G  R  P   G  G  V   F  V  G  I   C  G  I

gactactcga agcggatcac ccggcgcgac ccgcgcttga tcgacgccta catcggctcg
 D  Y  S   K  R  I  T   R  R  D   P  R  L   I  D  A  Y   I  G  S

ggcaacggac acagcgtagc ggccgggcgg ctctcgtatc acttcgggct cggcgggccg
 S  L  R   A  G  E  C   D  F  A   L  A  G   G  V  N  L   L  L  D

tgtgtggcca tcgacaccgc ctgttcgtcc tcgctggtcg gcgcgcactg ggcctgccag
 C  V  A   I  D  T  A   C  S  S   S  L  V   G  A  H  W   A  C  Q

tcgcttcggg ccggcgagtg cgacttcgcc ctggccggcg gcgtgaacct cctgctcgac
 S  L  R   A  G  E  C   D  F  A   L  A  G   G  V  N  L   L  L  D

cccgaactga gcatcaactt ctcgaaggcc aacatgctgg ctcccgacgg gcgctgcaag
 P  E  L   S  I  N  F   S  K  A   N  M  L   A  P  D  G   R  C  K

acgttcgaca cccgggccga tggctacgtc cgcggcgagg gtgcgggcat ggtcgtgctc
 T  F  D   T  R  A  D   G  Y  V   R  G  E   G  A  G  M   V  V  L

aagcggctct ccgacgcccg ggccgacggc gatcgcatcc tggcggtgat tcgcggctcg
 K  R  L   S  D  A  R   A  D  G   D  R  I   L  A  V  I   R  G  S

gccgtcaacc aggacggccc cagcagcgga ctaaccgtgc ccaacggtcc ggcccagcag
 A  V  N   Q  D  G  P   S  S  G   L  T  V   P  N  G  P   A  Q  Q

gacgtgatcc gccgggccct cagcgcagcc aacctggaac ccgacgacat cgactacctc
 D  V  I   R  R  A  L   S  A  A   N  L  E   P  D  D  I   D  Y  L

gaagcccacg gcactggcac t
 E  A  H   G  T  G  T
```

Figure 4.18 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_033. This sequence contained 681 base pairs with 103 a, 246 c, 232 g, and 100 t. The type I-PKS KS important amino acid conserved motifs are highlighted.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
gtggatcctc agcagcggct gcttttagag gtgacctggg aggcgctgga atatgccggc
 V  D  P   Q  Q  R  L   L  L  E   V  T  W   E  A  L  E   Y  A  G

ataccgccgg acagcctgtc cggcagccag accggagttt ttattgggat tagcagccac
 I  P  P   D  S  L  S   G  S  Q   T  G  V   F  I  G  I   S  S  H

gattatattc aattacagac aacacccacc ccatttagcg gcacgggtaa cgccttgagt
 D  Y  I   Q  L  Q  T   T  P  T   P  F  S   G  T  G  N   A  L  S

attgctgcca atcgactgtc ttataccctc aacctgcacg gccccagttg ggccgtcgat
 I  A  A   N  R  L  S   Y  T  L   N  L  H   G  P  S  W   A  V  D

acggcctgtt cctcgtcact ggtagctgtt catcaggcca taataagctt gcagcagggg
 T  A  C   S  S  S  L   V  A  V   H  Q  A   I  I  S  L   Q  Q  G

gagtgccatc tggccctggc gggcggggtc aatctcattc taacgccgga gctgaccgac
 E  C  H   L  A  L  A   G  G  V   N  L  I   L  T  P  E   L  T  D

atctttttc aggccggcat gttagcgccc gacggccggt gtaaaacatt tgatgccaaa
 I  F  F   Q  A  G  M   L  A  P   D  G  R   C  K  T  F   D  A  K

gctgatggtt atgtgcgggg tgaaggggcc ggtattgttg ttctcaaacg cctggcggat
 A  D  G   Y  V  R  G   E  G  A   G  I  V   V  L  K  R   L  A  D

gcccggcaag ccggagataa cattctggcg gtcattcgcg gctcggcggt aaaccaggat
 A  R  Q   A  G  D  N   I  L  A   V  I  R   G  S  A  V   N  Q  D

ggccggagca acggcctgac cgcgcctaat gggcctgctc aacaggcggt aatcagacag
 G  R  S   N  G  L  T   A  P  N   G  P  A   Q  Q  A  V   I  R  Q

gcgctacaaa aagcgggagt cgatgccagt caccttggct acattgaagc gcatggcacc
 A  L  Q   K  A  G  V   D  A  S   H  L  G   Y  I  E  A   H  G  T

ggcact
 G  T
```

Figure 4.19 Nucleotide sequence and its deduced amino acid sequence of partial type I-PKS KS gene from KK_KSI_035. This sequence contained 666 base pairs with 141 a, 185 c, 199 g, and 141 t. The type I PKS KS important amino acid conserved motifs are highlighted.

The similarities among Type I-PKS KSs from mangrove mud metagenome were assessed using clustalW2 to calculate the identity percentage among each of their deduced amino acid pairs. Besides KK_KSI_026 and KK_KSI_029 with 99 % identity to each others, all of the rest showed no more than 70% identity (ranging from 34% to 70%) among themselves (Table 4.1). These suggested that most of the obtained ketosynthase sequences were diverse. Although, from their deduced amino acid identity percentage alone, they were insufficient to be concluded. They were highly possible from different PKS clusters or organisms of origin.

Table 4.1 Identity percentage among deduced amino acid Type I-PKS KS partial sequences from Klongkone mangrove metagenomic DNA calculated by ClustalW2.

KK_KSI_002													
KK_KSI_003	55												
KK_KSI_010	49	60											
KK_KSI_013	51	60	59										
KK_KSI_017	40	45	43	43									
KK_KSI_019	50	53	44	50	39								
KK_KSI_021	37	43	41	42	39	39							
KK_KSI_023	45	63	52	54	39	47	40						
KK_KSI_026	41	49	42	43	38	45	39	46					
KK_KSI_027	43	42	43	41	54	37	37	39	34				
KK_KSI_029	43	50	43	43	38	46	39	46	99	34			
KK_KSI_033	52	68	58	62	44	49	44	55	42	41	43		
KK_KSI_035	58	70	59	60	46	55	44	59	47	46	49	65	
Clones	**KK_KSI_002**	**KK_KSI_003**	**KK_KSI_010**	**KK_KSI_013**	**KK_KSI_017**	**KK_KSI_019**	**KK_KSI_021**	**KK_KSI_023**	**KK_KSI_026**	**KK_KSI_027**	**KK_KSI_029**	**KK_KSI_033**	**KK_KSI_035**

(Excluded KK_KSI_026 and 029, the highest % identity was highlighted in dark shade; the lowest was highlighted in light shade.)

The blastx, and blastp were operated using the insert DNA sequences, and their deduced amino acid sequences as query sequences, respectively, in attempt to identify the homology of the obtained sequences to the proteins on the public database. The BLAST results were the same for the two methods. The conserved domain database (CDD) searching service was also used. It searched and displayed the most likely homologous protein or protein family instantly along with the blast searching (Fig. 4.20).

Confirmed by BLAST and CDD searching, all Klongkone mangrove mud metagenomic DNA inserts from the thirteen (twelve diverse) clones were significantly homology to type I-PKS KS domain genes, as shown in Fig. 4.20. Their maximum % identities were in the range of 53.3% to 78.3% identity (11 out of 12, were ranging at 60s% identity or below) to their closest match on public Database. It very likely suggested the evolutionary distance between the type I-PKS KS from Klongkone mangrove soil and ones on the public database. The top blast hit results were summarized in Table 4.2. It was found that most of the obtained deduced amino acid sequences exhibited homology below 70% to those on the public database. These suggested the diversity and novelty tendency of the KSs, ultimately their linked functional PKS pathway, within the Klongkone mangrove soil metagenome. Their sequences were deposited on GenBank, and their accession numbers are shown in Table 4.3. Further phylogeny of these sequences was studied to determine their homology and novelty.

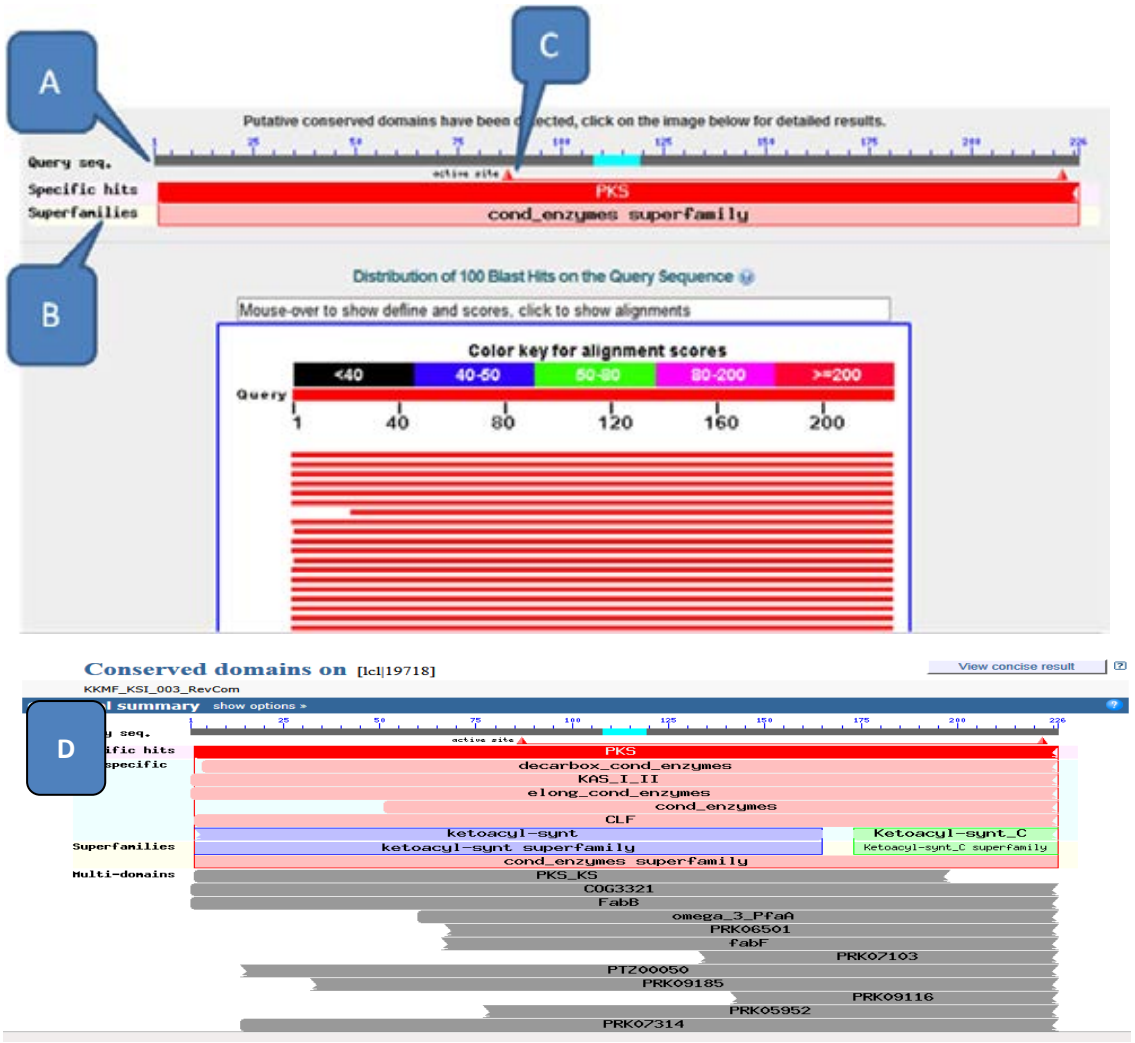


Figure 4.20 The conserved domain database (CDD) searching services provided alongside with new blast results. The diagram (A) showed the query sequence length and position. The summaries of conserved domain hit results were showed in diagram (B). The annotation for the position (C) of the active site for functional enzyme proteins were also displayed. Selecting on the displayed area brought up the full CDD service with detailed results (D).

Table 4.2 Summary of Blast results, using partial type I PKS KS sequences from Klongkone mangrove mud metagenome as queries, showing the sequences from GenBank returned as the highest % identity.

Clones KK_KSI	Top Blast Hit		Known species with maximum %identity *		
	% identity	Description (accession number)	% identity	Description (accession number)	protein
002	63.7	uncultured bacterium KS (ABH08749)	63.3	*Mycobacterium* sp. JDM601 (*Actinobacteria*) (YP_004521734)	-
003	78.3	uncultured bacterium KS (ACC99565)	73.9	*Anabaena variabilis* ATCC 29413 PKS KS (*Cyanobacteria*) (AAX44126)	-
010	59.2	uncultured bacterium PKS KS (ACC99568)	64.0	*Chondromyces crocatus* PKS KS (*Deltaproteobacteria*) (CAQ18829)	AjuB
013**	64.9	*Mycobacterium vanbaalenii* PYR-1 hybrid PKS/NRPS (*Actinobacteria*) (YP_951844)			
017	61.0	*Hyphomicrobium denitrificans* KS (*Alphaproteobacteria*) (YP_003755474)			-
019	65.2	uncultured bacterium KS (ADE08320)	60.8	*Sorangium cellulosum* PKS KS (*Proteobacteria*) (AAY32964)	DszA (DisA)
021	53.3	*Streptomyces griseus* NRPS/PKS (*Actinobacteria*) (YP_001821790)			-
023	60.6	*Planktothrix rubescens* NIVA-CYA 98 PKS (*Cyanobacteria*) (CAQ48259)			MicA
026	64.2	uncultured bacterium KS (AAX86994)	63.7	*Maribacter* sp. W-9 PKS KS (*Bacteroidetes*) (ACR55656)	
027	64.9	*Leptolyngbya* sp. PCC 73110 PKS KS (*Cyanobacteria*) (AAX44111)			
029	64.6	uncultured bacterium KS (AAX86994)	64.1	*Maribacter* sp. W-9 PKS KS (*Bacteroidetes*) (ACR55656)	
033***	67.0	*Lyngbya majuscula* PKS KS (*Cyanobacteria*) (AAS98782)			JamK
035	73.9	*Anabaena variabilis* ATCC 29413 PKS KS (*Cyanobacteria*) (AAX44126)			

* For ones with the top blast hit were uncultured, the next hit sequence with known species of origin was determined.

Hit with highest % identity to *Chondromyces crocatus* PKS KS (*Deltaproteobacteria*) (CAQ18829); with (65.9%) **, (69.6%) ***

Table 4.3 Summary data of partial Type I PKS KS amplicon insert DNA from Klongkone mangrove mud metagenome

Clones	Partial Type I-PKS KS amplicon insert DNA (carried within the clone)		
	Accession number	Length of nucleotide sequence (bp)	Sequence details (Figure)
KK_KSI_002	JN224477	675	4.7
KK_KSI_003	JN224478	678	4.8
KK_KSI_010	JN224479	684	4.9
KK_KSI_013	JN2244780	678	4.10
KK_KSI_017	JN2244781	693	4.11
KK_KSI_019	JN2244782	681	4.12
KK_KSI_021	JN2244783	663	4.13
KK_KSI_023	JN2244784	675	4.14
KK_KSI_026	JN2244785	678	4.15
KK_KSI_027	JN2244786	693	4.16
KK_KSI_029	JN2244787	678	4.17
KK_KSI_033	JN2244788	681	4.18
KK_KSI_035	JN2244789	666	4.19

4.2.3 Klongkone type I-PKS KS phylogenetic analysis

To determine the diversity and novelty of the obtained type I-PKS KS sequences, the rooted neighbor-joining phylogenetic trees were constructed using top five blast hits of each insert sequences and used eubacterial FabFs (KASIIs) as out groups with 1000 replication bootstraps. The constructed phylogenetic tree revealed the diversity of the type I-PKS KS acquired from Klongkone mangrove mud metagenome. They were distributed into four different type I PKS sub-classification groups, hybrid PKS/NRPS type I, type II (trans-AT), type III, and cis-AT PKS (Figs. 4.21 and 4.22). Type I-PKS KS from clones KK_KSI_017 and 027 were definely

clustering together with 100% bootstrap supported score. The phylogenetic cluster of this clade was previously assigned as PKS/NRPS hybrid type I group, and the PKS KS members of this clade were described as NRPS preceded PKS system (Moffitt & Neilan, 2003; Ginolhac *et al.*, 2004; Ginolhac *et al.*, 2005; Zhao *et al.*, 2008). In this study, however, had recognized that this PKS/NRPS hybrid type I was separated into two sub clusters. One including KK_KSI_027, showed all the signature motifs [N(DE)K] and VQTACSTS for PKS/NRPS hybrid type I described by Zhao (2008), hence, the cluster was labeled hybrid PKS/NRPS I-I. On the other hand, its sister clade, had the variable signature motif at the active site motif. Instead of having VQTACSTS, it had V*TACSTS, where * could be asparagines (N), serine (S), tyrosine (Y), histidine (H), or valine (V). As the members of this clade possessed variant of hybrid PKS/NRPS I signature, this clade was assigned as hybrid PKS/NRPS I-II; and, KK_KSI_017 was in this group.

Another group assigned as hybrid PKS/NRPS II by Zhao (2008) with the VETACSSS signature motif, but without 22 upstream amino acid [N(DE)K] pattern was closely clustered and shared the common ancestor with significant support by bootstrap value (100%) to KK_KSI_019 and domain KS_4 of DszA/ DisA (AAY32964), part of the disorazole synthase gene cluster. Although, KK_KSI_019 and KS_4 of DszA/ DisA (AAY32964) did not show the VETACSSS signature motif, their sequences together with all others in this clade exhibited proline amino acid residue at the upstream position next to the cysteine active site signature motif (Fig. 4.22). Proline at this position did not present on any other KS sequences on the tree, but only the sequences on this clade. Moreover, the analysis of the intact sequences (where it possible) from which the KS blast results were obtained. The domain organization of these sequences revealed their lacking of AT domain within the PKS module; thus, they were trans-AT modular PKS (Table A1 and A2). As all of the clade members with the intact protein information could be identified as trans-AT PKS, this clade was putatively assigned “trans-AT hybrid PKS/NRPS”. The proline residue of the sequences in this clade might be related to the trans-AT functionality. Therefore, the significance of this proline residue, present in all clade members, in the relative to the function of the PKS would be very interesting subject to be pursued.

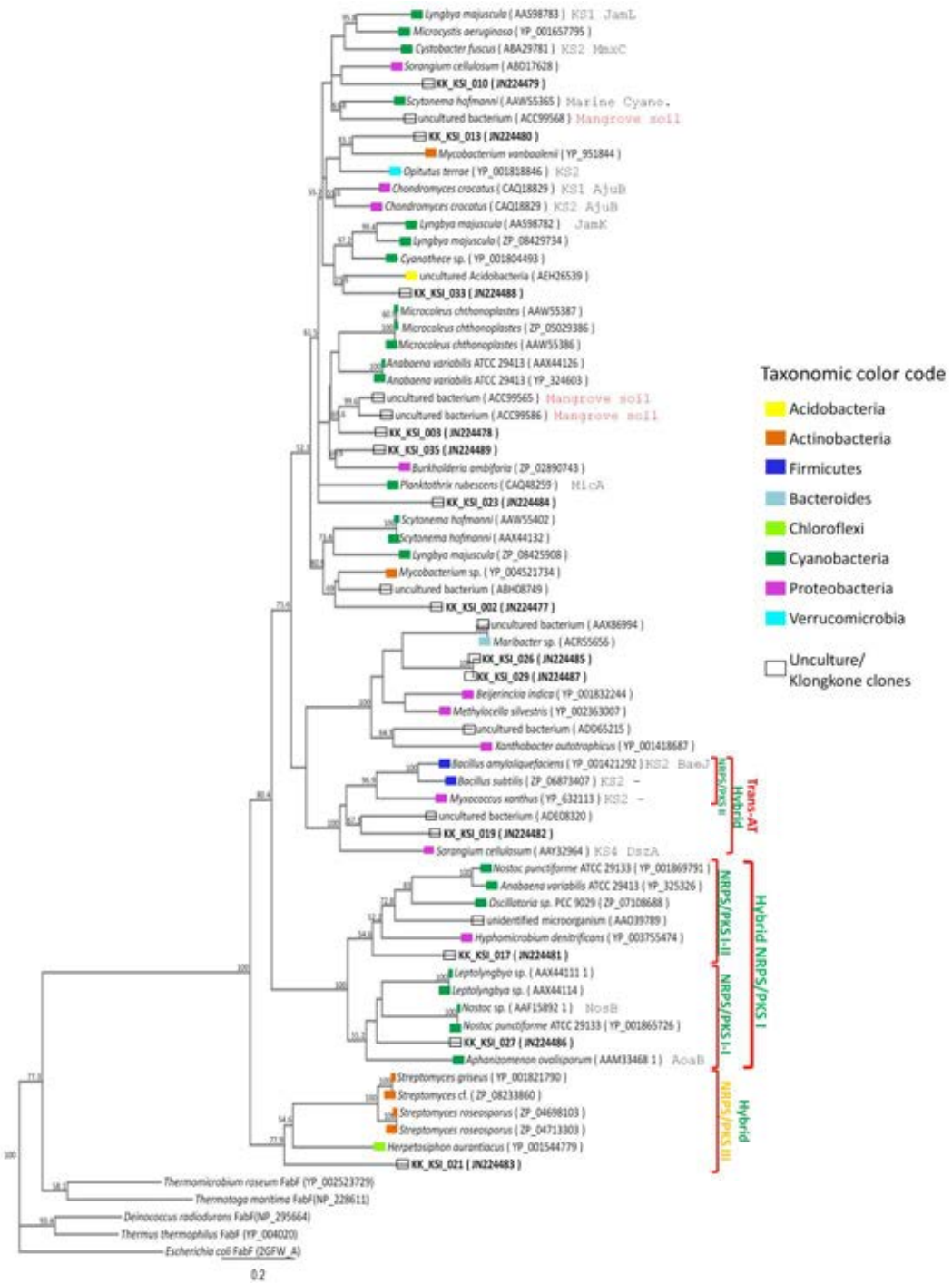


Figure 4.21 Phylogenetic analysis of type I-PKS KS domains from Klongkone mangrove forest soil, constructed by neighbor-joining under Geneious version 5.4.6. Bootstrap values of >50%, calculated from 1,000 bootstrap trees are indicated. Sequences from this study are in boldface. The accession numbers of those sequences retrieved from Genbank are in brackets. The scale bar indicates 0.2 amino acid substitution per position.

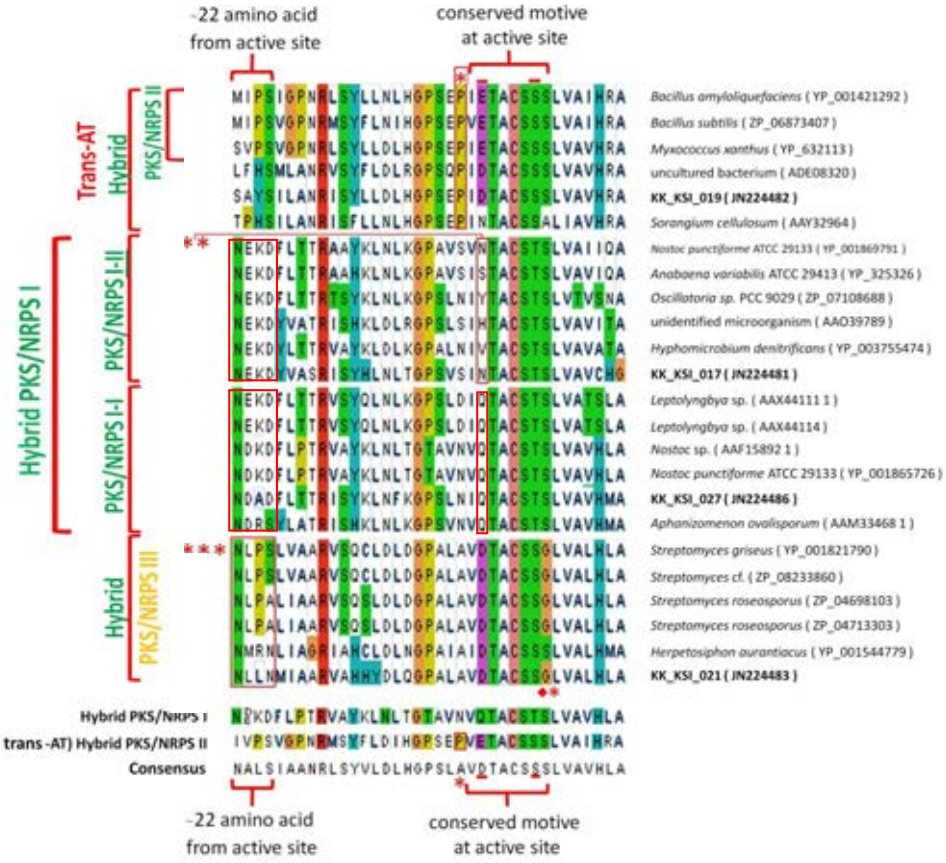


Figure 4.22 Amino acid alignment showing signature sequences around cysteine active site motif and about 22 residues from the active site toward N-terminal pattern of ketosynthase domains for hybrid polyketides synthase/non-ribosomal peptide synthetase. The typical consensus of the KS active site motif (VDTACSSS), the previous assigned hybrid PKS/NRPS I [N(D/E) KD and VQTACSTS], and hybrid PKS/NRPS II (VETACSSS) signatures are present at the top of the alignment. Sequences from this study are in boldface. The remarks on the alignment for hybrid PKS/NRPS of each distinct clades are indicated. "*" indicates the proline residue present only on trans-AT hybrid clade, "**" indicate residue in the signature active site motif of assigned hybrid PKS/NRPS I-II that are varied (polar: asparagine, N; serine, S; histidine, H, or hydrophobic: tyrosine, Y; valine, V) amino acids. "***" indicates the possible other patterns at about 22 residues from the active site toward N-terminal (NLXX) which seem to vary according to the organisms of origin in assigned hybrid PKS/NRPS III. Marked with a diamond and an asterisk is glycine third residue downstream from active site dominantly found in this clade.

For the most distinct clade on the tree, it was the clade, where the KK_KSI_021 was placed on with 77.9% bootstrap support. This clade was distinctively separated from the rest of the tree. The domain organization from the sequences on the clade that could be analyzed showed that the module the KS belong to was placed between NRPS modules (Table A1 and A2). In addition, the last serine (S) residue on the cysteine active site motif mostly was replaced by glycine (G). As the KSs on this distinctive clade were representing those KS domain from PKS module sandwiched by NRPS module on the same megapeptide; and, it did not belong to the hybrid PKS/NRPS I. thus, this clade was hypothetically proposed as hybrid PKS/NRPS III.

The last and biggest was the cis-AT clade. This clade comprised many clusters and many of the Klongkone mangrove metagenomic type I-PKS KSs were placed here, including KK_KSI_002, 003, 010, 013, 023, 026, 029, 033, and 035. No significant distinguish on the conserved motifs or their domain organization of their clade members could be used to sub-categorized them apart. All of the minor clades seemed to be the clusters of KSs from mixed groups of organisms of their origins. For example, KK_KSI_026 and KK_KSI_029 the similar sequences pair were placed on the clade of mixed *Proteobacteria* and *Bacteroidetes*. KK_KSI_010 placed closely related to KS from *Sorangium cellulosum* (ABD17628). They were clustered together with KS from *Cyanobacteria* and *Proteobacteria*, while, KK_KSI_002 which was closely related to KS from Actinobacteria, was placed among KS from *Cyanobacteria*. It was highly unlikely to be able to conclude anything deeper than to conclude that all of these type I-PKS KS domains from Klongkone mangrove metagenome were highly possible cis-AT PKS. Another interesting point on this clade was the KS sequences from the uncultured bacteria (ACC99586 and ACC99565) placed closely related to KK_KSI_003 on the tree. They were also obtained from mangrove soil, but, from China. If these KSs from China mangrove soil and KK_KSI_003, the Thailand mangrove soil, were related or come from the same PKS gene cluster, it might suggest that this PKS gene cluster was common or might be important in response to environmental stresses (niche specific) in mangrove ecosystem (Elsaied *et al*. 2007; Elsaied *et al.* 2011), harboring the same gene even geographically far apart.

It was clear that metagenome from Klongkone mangrove soil, contained variety of type I-PKS KS domains. There were both cis-AT and trans-AT together with many types of PKS/NRP hybrid systems, some of which were very likely the new systems. In fact, two of PKS/NRP hybrid systems were newly proposed in this study. In addition, as can be seen from the table. 4.2 and phylogenetic tree (Fig. 4.21), most of the homologous PKS KS domains from public database to the Klongkone mangrove metagenome's deduced amino acid sequences were from *Proteobacteria* and *Cyanobacteria*, few of them were from *Firmicutes*, *Acidobacteria*, *Verrucomicrobia*, *Chloroflexi*, and *Actinobacteria*. Most of these organisms of origin of the PKS KS, except those from *Cyanobacteria* and *Actinobacteria*, were in agreement with the 16S rDNA diversity study below. These may, in part, be caused by, the limitation of the metagenome extraction method used in this study. Using only SDS and heat might not be able to extract the DNA of *Actinobacteria* from mangrove sediment. The more vigorous extraction combination of both chemical and mechanical may be needed to reduce the DNA extraction bias (Jiang *et al.*, 2010). Therefore, those hit blast result sequences from *Cyanobacteria* and *Actinobacteria* would most likely represent the origin of PKS KS domain resulting from horizontal gene transfers which have been recognized and documented (Ginolhac *et al.*, 2005). Even, with the biases of bacterial variation of this Klongkone mangrove mud metagenome, it certainly was the good source for searching for new PKS gene.

4.2.4 PCR amplification of 16S rRNA from Klongkone mangrove mud metagenome.

Not only the diversity of the PKS genes was monitored, but the variety of bacteria within the Klongkone mangrove mud metagenome was also observed. The universal 27f and 1492r primers (Table 3.1) targeting bacterial 16S rRNA gene were used. The length of the amplicons generate by this pair of primers was about 1500 bp (Fig. 4.23). As the amplicons were generated using Taq DNA polymerase, they were able to ligate into T-A type cloning vector, i.e. pGEM®-T Easy Vector (Promega), and transformed into *E. coli* DH5α. One hundred white colonies grown for 18-24 hours on 100 μg/ml ampicillin supplemented LB agar plate were randomly picked. The

confirmation of successful transformation was assessed by plasmid extraction and colony PCR with pUC/M13 forward and pUC/M13 reverse primer pair. The pGEM®-T Easy_16S rRNA amplicon recombinant plasmids from 15 clones were sequenced. Clone KK_16S_035 was suspected to be the result of chimeric amplicon and was removed during sequence submission to GenBank. The accession numbers for other 14 sequnces are JN802261 to JN802274 as shown on in the table 4.5.

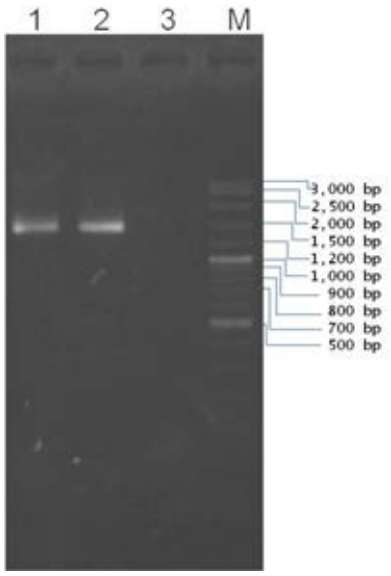


Figure 4.23 Agarose electrophoresis gel showing PCR products obtained from amplification of the 16S rRNA gene in mangrove mud metagenome. *E. coli* genomic DNA was used as positive control (lane 1). Lane 2 and 3 were mangrove mud metagenome and water, respectively. Lane M was 100 bp DNA ladder.

4.2.5 Sequence analyses of 16S rRNA gene from Klongkone mangrove mud metagenome

Multiple alignment and sequence homology of 16S rDNA obtained from Klongkone mangrove metagenome were performed and calculated using ClustalW2. It was found that the range of 16S rDNA sequence identity percentage among themselves were between 72% and 85% (Table 4.4). As the similarity levels of

the 16S rDNA sequences were below 97%, these indicated that all of the 16S rDNA sequences obtained from Klongkone mangrove metagenome were come from different bacterial species (Janda 2007; Stackebrandt 1994; Fox 1992).

Table 4.4 Identity percentage among 16S rDNA sequences from Klongkone mangrove metagenomic DNA calculated by ClustalW2. Total number of nucleotides in each sequence (1452-1526 bp) was shown in Fig. B1 to B14.

018														
019	79													
023	81	80												
034	77	80	78											
036	78	84	80	81										
037	79	83	80	79	85									
038	79	80	81	79	81	82								
039	80	81	81	78	80	81	84							
040	85	80	82	78	78	80	80	80						
041	77	77	77	76	76	76	77	75	75					
042	74	78	75	75	76	75	75	76	76	74				
043	72	74	73	75	74	76	74	72	73	72	75			
044	78	83	82	79	83	81	81	82	79	77	78	75		
045	77	78	77	76	77	76	78	76	75	79	75	73	77	
Clones KK_16S	**018**	**019**	**023**	**034**	**036**	**037**	**038**	**039**	**040**	**041**	**042**	**043**	**044**	**045**

(The values with highlighted in dark or light shades were the highest or lowest % identity, respectively.)

To determine the novelty of the bacterial species inhabited within Klongkone mangrove soil, the obtained 16S rDNA sequences were used as query sequences to blast against GenBank database. All of the top blast results for each query sequences (Table 4.5) were uncultured bacteria, and most of them were from marine environmental samples. The top blast results for KK_16S_018, KK_16S_034, KK_16S_037, KK_16S_040, KK_16S_043, KK_16S_044, and KK_16S_045 were returned with 16S rDNA sequences with the homology level below 97%. These

suggested that half of the randomly selected 16S rDNA sequences represented novel bacterial species. For those that had been classified (KK_16S_018, KK_16S_034, and KK_16S_037), they were distributed among bacterial phyla, *Acidobacteria* and *Proteobacteria*. Another four sequences (KK_16S_040, KK_16S_043, KK_16S_044, and KK_16S_045) had not been classified.

On the other hand, the top hit blast results of KK_16S_019, KK_16S_023, KK_16S_036, KK_16S_038, KK_16S_039, KK_16S_041, and KK_16S_042 were returned with 16S rDNA sequences showing the homology equal or over 97% but not 99.5%. These implied the species novelty represented by these 16S rDNA sequences comparison was uncertainty, since; the identity score of the 16S rDNA alone is not enough to clarify the classification rank from the species level and beyond (Fox 1992). To identify as reliable same species, the genomic reassociation over 70% identity is required, which is not applicable for this study. Additionally most of them were still unclassified, for those classified [alphaproteobacterium (DQ811848), the top hit for KK_16S_02 and deltaproteobacterium (JF344660), the top hit for KK_16S_038], they were belong to phylum *Proteobacteria*. However, with these great 16S rDNA homology, it could not be denied the closely relationship of the Klongkone mangrove phylotypes to those already discovered elsewhere. Especially, when their counter hit blast results were also from marine environmental samples. In particular, EF125450, DQ811848, and EF125455 the top hit blast results of KK_16S_019, KK_16S_023, and KK_16S_039, respectively, were all come from mangrove soil samples. Even though it could not be concluded with confidence clearly, these Klongkone mangrove mud 16S rDNA sequences and their counter blast results were very closely related, and highly possible the same species. Although, these would mean that many of them were already discovered and not novel (which more than half that were), it confirmed that the obtained 16S rDNA sequences from Klongkone mangrove mud metagenome were come from and part of the real mangrove bacterial community.

Table 4.5 Summary of Blast results, using 16S rDNA sequences from Klongkone mangrove mud metagenome as queries, showing the hit results with the highest % identity.

Clones (accession number)	Top Blast Hit			
	% identity	Description	Accession number	Classification group /environmental site *
KK_16S_018 (JN802261)	93 1414/1509	Uncultured sediment bacterium clone JSS S04 317	HQ191049	*Gammaproteobacteria*/ Janssand intertidal sediment (Germany)
KK_16S_019 (JN802262)	98 1486/1516	Uncultured bacterium clone MSB-2F3	EF125450	Unclassified/ Haikou mangrove soil (China)
KK_16S_023 (JN802263)	97 1426/1456	Uncultured alpha proteobacterium clone MSB-3A6	DQ811848	*Alphaproteobacteria*/ Haikou mangrove soil (China)
KK_16S_034 (JN802264)	94%**	Uncultured bacteria clone ODP1230B34.16	AB177173	Unclassified /subseafloor sediment (Peru)
KK_16S_036 (JN802265)	98% 1485/1522	Uncultured bacterium clone GoC_Bac_29_D1_C0_M0 ()	FJ813582	Unclassified/ volcano mud, the Gulf of Cadiz (Belgium)
KK_16S_037 (JN802266)	93% 1399/1500	Uncultured Acidobacteria bacterium clone HAHS13.83	HQ396922	*Acidobacteria*/ haloalkaline soil (salt paddy) (India)
KK_16S_038 (JN802267)	99% 500/1520	Uncultured delta proteobacterium clone ANOX-098	JF344660	*Deltaproteobacterium*/ Galicia marine oil-polluted sediments (Spain)
KK_16S_039 (JN802268)	97% 1485/1527	Uncultured bacterium clone MSB-2G1	EF125455	Unclassified/ Haikou mangrove soil (China)
KK_16S_040 (JN802269)	92% 1403/1531	Uncultured bacterium clone BP100	HQ190558	Unclassified/ Zhonguan oil field (China)
KK_16S_041 (JN802270)	97% 1468/1518	Uncultured bacterium clone BS009	GU145394	Unclassified/ Black sea suboxic zone
KK_16S_042 (JN802271)	98% 1435/1466	Uncultured bacterium clone SCS_HX36_194	HM598232	Unclassified/ Deep sea sediment South China sea (China)
KK_16S_043 (JN802272)	90% 1375/1526	Uncultured bacterium clone HWB2224-2-58	HM243858	Unclassified/ Honghu lake sediment (China)
KK_16S_044 (JN802273)	95% 1451/1530	Uncultured bacterium clone 034E59	EU925843	Unclassified/ Yellow Sea sediment (China)
KK_16S_045 (JN802274)	89% 1332/1504	Uncultured bacterium clone C13S-66	EU617882	Unclassified/ northern Bering Sea sediment (China)

*The information on the sequences was from GenBank
** Uncultured bacterium (AB177173): 94% identity of 93% query coverage

4.2.6 Klongkone 16S phylogenetic analysis

Phylogenetic tree was constructed using the 16S rDNA sequences of Klongkone mangrove metagenomic 16S rDNA, their top hit blastn search, and their

related 16S rDNA from ribosomal database projects (RDP) Release 10 (Update 27 with total 1,921,179 16 rDNA sequences, with 962,279 good quality over 1200 bp of bacterial sequences). The sequences from RDP were obtained using seqmatch searching tool provided at RDP website. Three related 16S rDNA sequences from non type strains and three from type strain hit matches of each Klongkone mangrove metagenomic 16S rDNA sequences were made up the related sequences from RDP. Also from RDP database, a type strain from non related bacterial phyla which did not present in the seqmatch search results was included for the phylogenetic tree construction. The phylogenetic tree constructed with these 16S rDNA, visually displayed the distribution of Klongkone mangrove metagenomic 16S rDNA among 16S rDNA sequences from many bacterial phyla. From the phylogenetic tree, the 16S rDNA sequences from Klongkone mangrove mud metagenome were placed on clades from 6 bacterial phyla, the *Proteobacteria, Acidobacteria, Chloroflexi, Verrucomicrobia, Lentisphaerae*, and *Deferribacteres*. 16S rDNA sequences from 6 clones were placed on the proteobacterial clade. 16S rDNA sequences from clones KK_16_018 and 040 were on *Gammaproteobacteria* sub-clade, while KK_16S_023 were clustered within *Alphaproteobacteria*, and KK_16S_044, 039, and 038 were placed on *Deltaproteobacteria*. Those of KK_16S_037, 036, and 019 were clustered on *Acidobacteria* clade (Fig. 4.24_A.). Each of the 16S rDNA sequences from KK_16S_041, 042 and 045 was distinctively placed clostly to phylogenetic clade of *Verrucomicrobia, Chloroflexi, and Lentisphaerae*, respectively (Fig. 4.24_B).

The 16S rDNA sequences from clones KK_16S_034 was found to be placed among phylotypes and bacterium classified into *Deferribacteres* phylum which were the bacterial group found in deep sea ocean floor (Inagaki *et al.,* 2006). The 16S rRNA gene of uncultured bacterium clone ODP1230B34.16 was the one most related to the Klongkone mangrove mud 16S rDNA KK_16S_034. Although together with ODP1230B21.12's sequence, they were deposited on public database as unclassified. However, phylogenetic tree, in this study, (Fig. 4.24_B) clustered them with *Caldithrix abbyssi* (T) LF13T (AJ430587), the bacterium in phylum *Deferribacteres* (Mori *et al.,* 2008). This was agreed with phylogenetic tree from the study of Inagaki (2006) that those uncultured bacterium clone ODP1230B34.16 and ODP1230B21.12

were clustered on *Deferribacteres* clade as well. Thus, it was most likely the insert of KK_16S_034 would belong to the bacterium in this phylum.

For KK_16S_043, each of them was phylogenetically placed on its own distinct clade that was evolutionary closely related to those environmental unclassified 16S rDNA. In addition to its most homologous sequence from blast results was an unclassified bacterium with only 90% identity, and the phylogenetic tree suggesting the evolutionary distinct apart from all other existing bacterial phyla. It was possible this Klongkone mangrove metagenomic 16S rDNA belonged novel unclassified bacterial group.

As can be seen, no bacterium from *Actinobacteria*, the main sources of natural product in soil, was identified in this study. The limited number of 16S rDNA studied may be one of the reasons. However, the most possible reason should be the limitation of the DNA extraction method used in this study. The direct extraction with chemical (SDS) and heat was used for extraction of the Klongkone mangrove metagenome. This most gently extraction procedure was used because of the promising of high molecular weight DNA result. However, the diversity of genomic DNAs was compromised, and the ones from actinobacteria could not be extracted from mangrove soil (Jiang *et al.*, 2010). Despite the suspected limitation of bacterial phyla from the extraction bias of metagenomic DNA, the most abundant bacterial group found in this study was agreed with the study of Sakami (2008). From four of their DGGE major bands, three of them were representing proteobacterial groups (Sakami *et al*., 2008). However, as they did not identify the 16S rRNA gene of those variety minor bands, the diversity of bacterial phylotypes present in klongkone mangrove forest had not been concluded in their study.

This study had demonstrated that Klongkone mangrove metagenome contained bacterial genomic information of diverse groups of bacteria, at least from 6 phyla and one unclassified, and half of them were very likely new bacterial phylotypes (at the species level) which could be accessed through this metagenomic DNA.

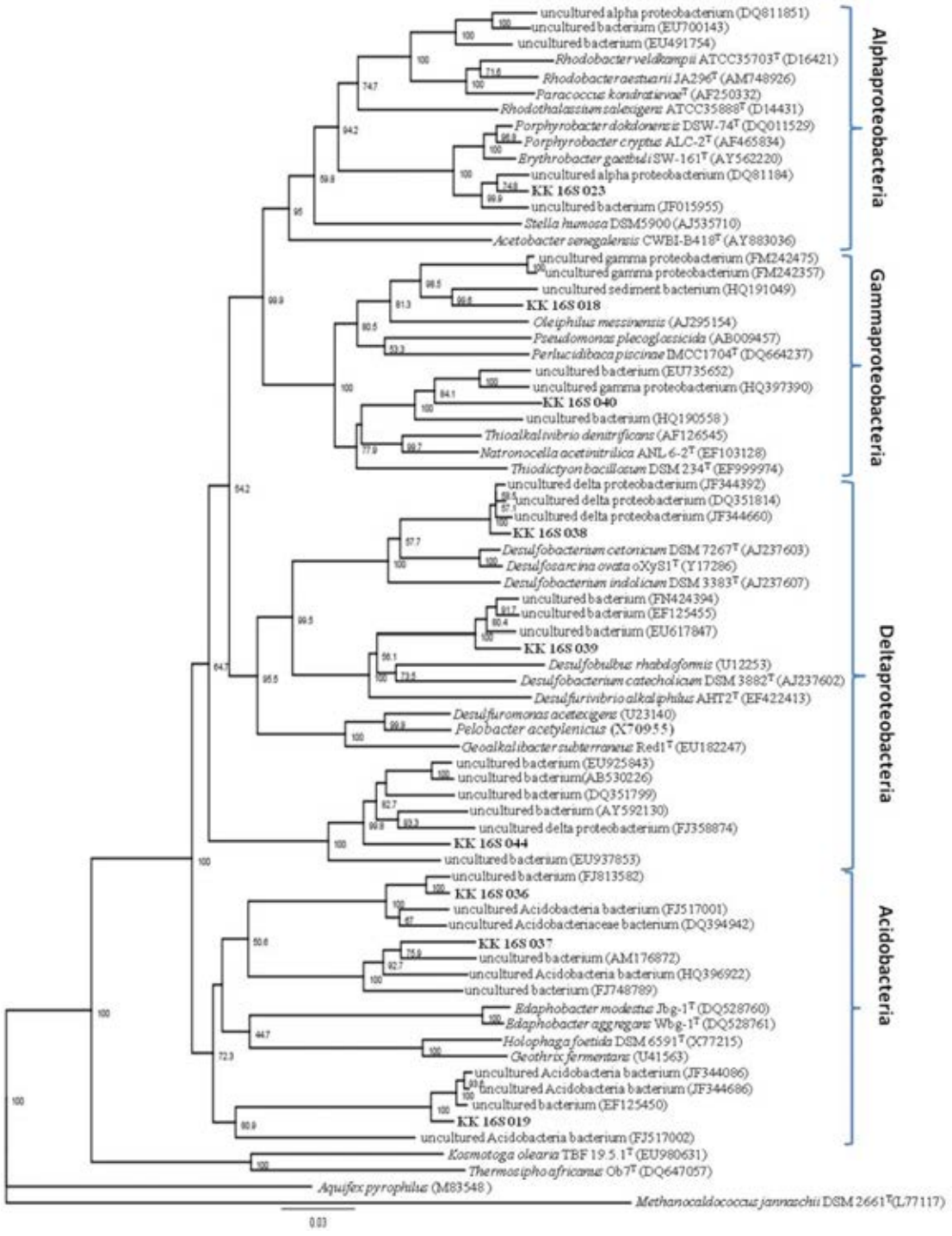


Figure 4.24_A Phylogenetic tree based on 16S rRNA gene, inferred by Neighbor-joining, showing the distribution of klongkone metagenomic derived 16S rRNA genes among bacterial phyla. Bootstrap values over 50% from 1,000 resampling were shown at each node. The scale bar represents 0.05 nucleotide substitutions per site.

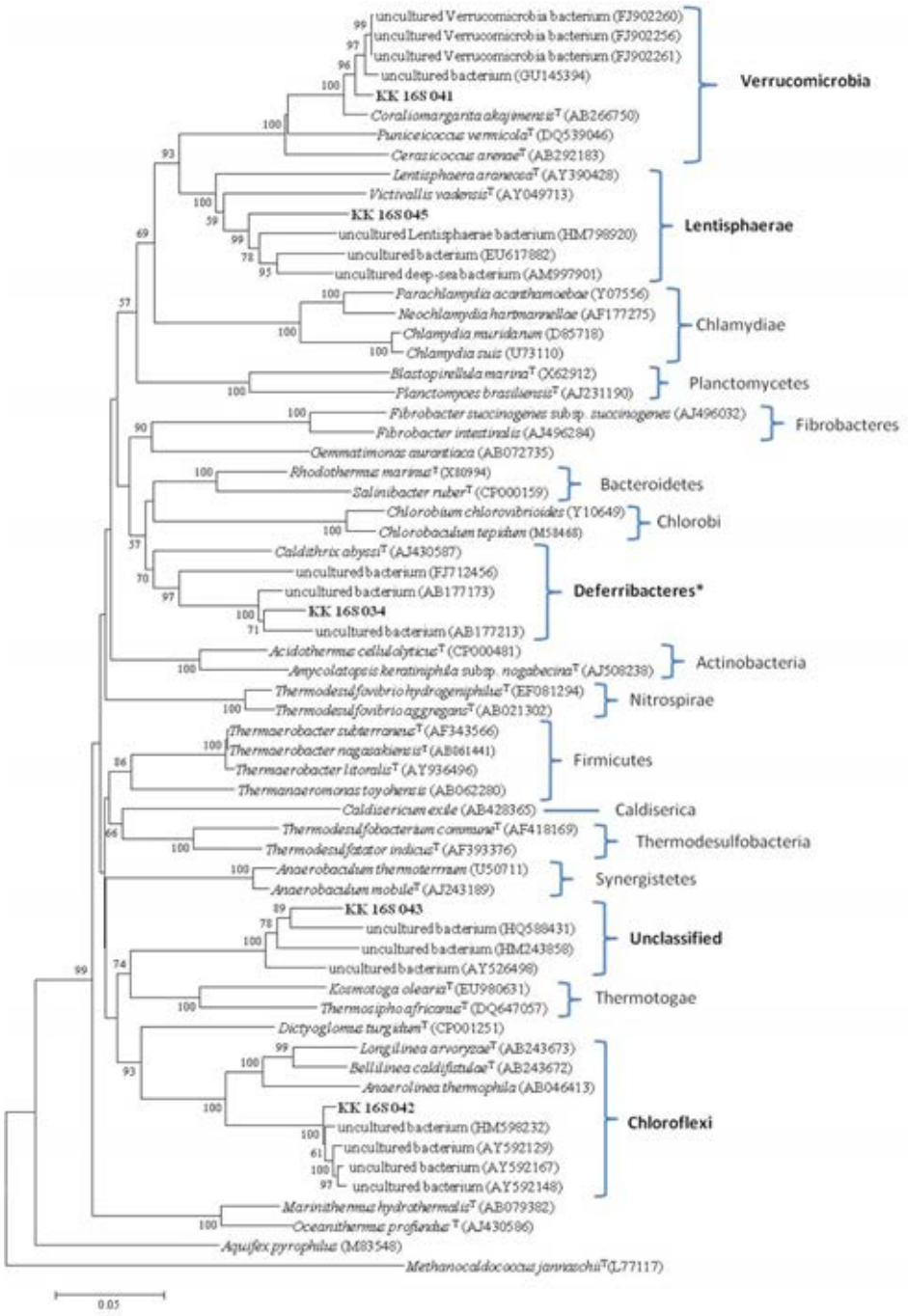


Figure 4.24_B Phylogenetic tree based on 16S rRNA gene, inferred by Neighbor-joining, showing the distribution of klongkone metagenomic derived 16S rRNA genes among bacterial phyla. Bootstrap values over 50% from 1,000 resampling were shown at each node. The scale bar represents 0.03 nucleotide substitutions per site. (*) based on phylotypes from Inagaki (2006). Clades with 16S rRNA gene from Klongkone metagenome were displayed in bold.

4.3 Mangrove mud metagenomic fosmid library construction

From the type I-PKS KS gene diversity, it had been shown that numbers of type I PKSs could be detected in the Klongkone mangrove mud metagenome. Most of them potentially belonged to new type I PKS. However to be more practical and putting the PKS gene into work more than partial type I PKS KS domain gene was needed, and at least the whole KS domain gene or an intact PKS modular gene should be obtained for using it in the future expression or combinatorial biosynthesis study. To capture a large piece of metagenomic DNA that could hold at least an intact minimal PKS module, fosmid library of Klongkone mangrove mud metagenome was constructed using CopyControl^{TM} Fosmid Library Production Kit. The constructed library contained about 14,000 fosmid clones. The library was kept as pooled of clones as suggested by the manufacturer's instruction manual. The 95 sub-pools contained about 150 clones per pool were kept at -80 ˚C until used.

4.4 Screening of metagenomic fosmid library

4.4.1 Functional based screening

The clones of metagenomic library were observed for their ability to express the desirable traits, the pigment production, hemolytic activity, and antimicrobial activities. All pooled clones were grown on LBA_{Cm} or LB_{CmAr} and the pigment formation of the clone was observed for up to six days. No clone with pigment or color was developed. The similar result was received from the hemolytic activity test on 7% (v/v) sheep blood agar. There was no hemolytic zone could be observed from 39 tested pools (1-30, 41-44, 91-95), which was corresponding to approximately 5800 clones. Unsuccessful screening for functional properties on the antimicrobial activity was repeated as the same pools as used in hemolytic activity screening were used. From all three testing strains, any clones that might exhibit a hollow or cleared zone around their colony at any sizes were collected, 479 suspected clones were obtained. The suspected clones were undergone the confirmational test against all test strains (*C. albicans*, *B. subtilis, and S. marcescens*) (Table 4.6). The results were inconsistent and not reproducible which led to the conclusion of no antimicrobial active clone was found.

It was not surprised that no positive clone from any tests was found. There are many factors contributing to the success rate of the functional based screening. Environmental sample choosing was one of the important factors. The sample that supported or enriched with the traits of interest would enhance the chance success rate, such as the discovery of novel agarase genes from agarolytic consortium (Voget *et al.*, 2003). The enrichment stretagy was inapplicable in this study because. the uniqueness of mangrove forest was required. Incomplete gene, gene cluster, or operon would lead to no expression and function. Incompatible host, lack of the required expression machineries, or unequipped with the right transporting systems, all could also lead to no expression and function. The example studying of Schmitz and colleagues in 2008, the screening for genes with hemolytic activity from phage genomic DNA, and the hit rate was as high as one in 5,000 screened clones (Schmitz *et al.*, 2008). The phage genomic DNA, which could readily be expressed in bacterial host, was speculated as one of the reason for such high hit rate. The expression of toxic elements was another factor contributed to low rate of discovering the positive results and added more challenges to the procedure. A chance of finding one positive clone in 20,000 clones to less than 1 in 100,000 clones was considered high success screening rates for the antimicrobial activity and pigment production to be found (Brady, 2007). With the modest size of the fosmid library in this study (about 15,000 clones), the functional-based screening strategy currently used was too much time consuming and the chance of finding positive clones was slim. Therefore, it was detained, and the screening strategy was steered to mainly focus on sequence-based screening until the better and more effective functional-based screening method would be found.

Table 4.6 Number of clone with suspicious inhibition zone against specific test organisms

Pool	*C. albicans*	*C. albicans* + ketoconazole	*B. subtilis**	*S. marcescens**	Total
1	0	4	10	0	14
2	0	3	16	0	19
3	0	0	20	0	20
4	0	3	11	0	14
5	0	1	72	0	73
9	0	0	19	0	19
10	0	0	12	0	12
31	0	0	30	0	30
35	0	0	22	0	22
36	0	0	10	0	10
42	22	8	75	0	105
43	0	18	69	0	87
44	0	2	52	0	54
Total	22	39	356	0	479

*Chloramphenicol resistant strain.

4.4.2 Sequence based screening

This was searching processes for genes of interest captured within the metagenomic library based on the prior knowledge of the nucleotide sequences of that specific gene. Type I PKSs, one of the most important genes involving in biosynthesis of bioactive natural products, the wealthy sources for drug development and production, were the targets in this screening.

4.4.2.1 Identification of type I PKS KS domain gene from Klongkone mangrove mud metagenomic library using PCR

The sub-pools of Klongkone mangrove mud metagenomic library clones were screened for type I PKS KS domain genes. From 24 screened sub-pools (21-44), sub-pool number 37 was found to produce about 700 bp amplicons from the type I-PKS KS domain gene specific PCR. This successful hit rate of one from approximately 3600 might seem low comparing to 7 hit from 1000 clones. However, that high hit rate (7/1000) was from the screening of PKS within the enriched source of polyketides producing marine sponge and its associated microorganisms (Schirmer *et al.*, 2005). Therefore, the hit rate in this study might be considered modest. The clone carrying the DNA template for this partial KS domain gene amplicon was sought out as described in section 3.7.2. The results were shown in Fig. 4.25. Three isolates (KKFOS_KSI_pool37_050, 088, and 106) were found to be responsible for the positive PCR reaction for the target gene. However the type I-PKS KS sequences of their recombinant fosmid vectors were identical, they were believed to be the same clone. These clones were also tested for the functional activities. There was no functional trait of interest discovered from them. The pKKFOS_KSI_pool37_088 recombinant fosmid vector extracted from KKFOSKSI_pool_37_088 was used for further sequencing of the entire insert fragment DNA within its.

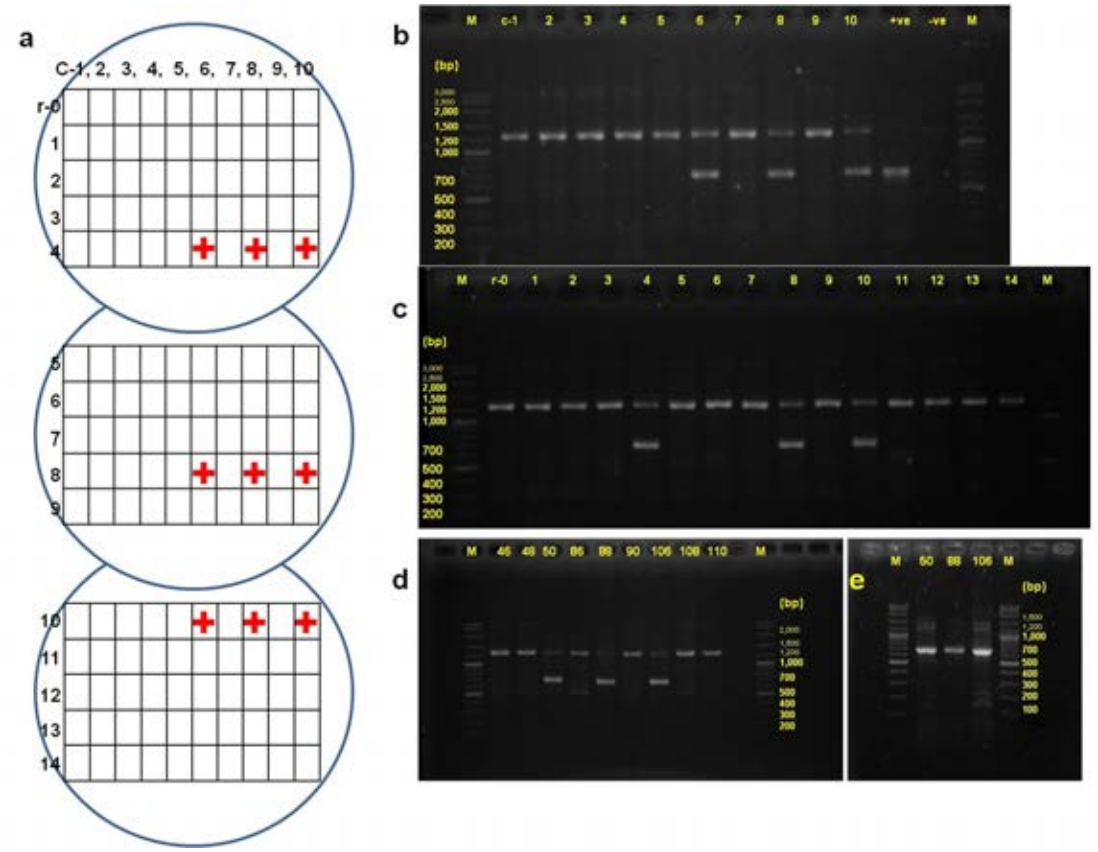


Figure 4.25 Type I PKS KS gene specific colony PCR pool matrix diagram. (a) The location of individual clone. Triplicate of these subpool matrix were prepared, two sets (on LB_{CmAr}) for pooling in column and row manner, another (LB_{Cm}) for later individual selection. Colony PCR result of pool of the matrix columns c-1 to c-10 and row r-0 to r-14 visualized on agarose gel electrophoresis (b) and (c), respectively. Nine possible positive clones [red crosses marked in (a)] were tested for type I PKS KS gene individually, and the results was visualized on agarose gel electrophoresis (d). The positive clones from (d) were confirmed for positive type I-PKS KS gene carrying (e).

4.4.2.2 Identification of novel PKS module by primer walking sequencing

The pKKFOS_KSI_pool37_088 fosmid was extracted and purified using FOSMID MAX (Epicentre), then it was sent for it DNA insert to be sequenced at 1st BASE (Singapore). The first outward primers were the reverse complement of MDPQQRf and HTGTr primers. All other designed primers and their properties were listed in the table 4.7. All primers were designed using Primer3 and were confirmed with NetPrimer and OligoAnalyzer to ensure their properties, e.g. Tm, and potential hairpin loop and dimer formation, were fitted for sequencing. Nucleotide sequencing of the inserted DNA by primer walking resulted in 4,468 bp long with about 68.6 % GC content. The entire sequence could be read through into a deduced amino sequence on a single open reading frame of translation, which also contained all three Type I-PKS KS gene characteristic motifs. The conserved motif VDPQQR (5'-GTCGATCCTCAGCAGCGT-3') was found at base position to 2,039-2,056, the cysteine active site motif VDTACSSS (5'-GTGGATACGGCCTGCTCCTCGTCG-3') was found at nucleotide position 2,294 to 2,317 , and the histidine active site motif HGTGT (5'-CATGGCACGGGGACG-3') located at the nucleotide number 2,711 to 2,725 (Fig. 4.26).

Table 4.7 Primers used in primer walking sequencing of pKKFOS_KSI_pool37_088 insert DNA.

Walking primers					
Name	**Sequences**	**Length (Bases)**	**Tm (˚C)**	**Hairpin**	**Dimer**
Out5'_01	GCC GAG ACA AGC CGA AGA GCG A	22	61.4	2	0
Out5'_02	GGG AAG ACT CGA GGA GCC AGG GTC GAT C	28	65.0	6	4
Out5'_03	CTC GCG CAG GAG ATC CCA GTA CCG ACC GAC	30	67.9	4	1
Out5'_04	GAA GGA GAG GTC CCG GTG CTT GGG GTA GTT	30	65.7	4	0
Out3'_01	CGG TCA AGA CGA ACA TCG GAC ATA C	25	58.0	6	0
Out3'_02	GAG CTC GCG TCT CTC CCG CCG CTT C	25	68.0	6	0
Out3'_03	GTG CTA TTC GCT TTC GAG TAT GCC CTC	27	60.0	4	0
Out3'_04	TCT CTT CAC CGG CCA GGG CTC GCA GTT C	28	67.5	4	0
Out3_05	CCT CGA GAG ATG CGA CGA GC	20	57.1	6	2

```
....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50
CCACGTCGAG CGAGCGCTCA ACGTCAGGGC CAGTGGTTTT CTCCATCGCC 50
GCTATTTCGA CGGCGTACGC GACGTCGTCG CTTCGATCTT CGACACGAAG 100
CCTTTCGAAG CTCAGCCCCG TTACGTCGTG GATACCGGCT GCGGGGACGG 150
CACCTTCCTG CGGTCGATCT ACGAGACGGT CCGAGCGCGC ACGGCTCGGG 200
GCAAGGTTCT CGATCAGTAC CCTCTCGTCC TCGTCGGGGT CGACCTCAAC 250
GATGCGGCGC TCGAGGAGGC GAAGGGCACG CTCGACGATC TTCCGCACGT 300
TCTCTTGAAG GGCGACGTCG CCGACCCTAA CGGGATCGAG GCCGAGCTCG 350
AGCGTCGAGG AATCGCGGCC GACGACGTCC TGCACGTGCG CTCGTTTCTC 400
GATCACGACT CGCACCTCCC GATGGTCACG GACCGCCTCG CGGCGCAGGC 450
GCGCGCGGAC GTCCCGTACC GCGGCGTGTC GGTGGGCCCG GCGGGTGAGC 500
TCGTTCCGCA GTCGGTCGTG ATGCAGGGGC TCGTCGAGCA TCTCACCCGA 550
TGGGCGAGTG TGCTGACGCG ACACGGGATG ATCCTGCTCG AGGTTCACTC 600
GGTCGACCCC GTCACCGCGG CTCGCTACCG CGAGTCGACG GAGAGCCTCA 650
GCTTCGACGC TTTTCAGGCG TTCTCGCTCC AGTACCTCGT GGAGGCGTCC 700
GACTTCCTGC TCGCGGTGGC CGAGGCGGGG CTCTTCGCGC GGCCGGCCTT 750
CGCGCGTCGC TACCCGACCG CCCTTCCTTT TACACGAATC TCGCTCAACT 800
GGCTCGAAAG GCGCCCCTAC CGGGTCCGCC TGGCACGCGC GTCCGACATC 850
CCCGCGATGA TGGAGATCGA GGCGGCGGCC TGGCCCGAGC CCCTCTCGGC 900
GTCTCGCGAG GAGATCGAGC GCCGTCTCGT CACCGATCCA CGTGGTCAGA 950
TCGTCGTGGA ATCGGAAGGG CGTCTCGTCG CCGTCCTCTA TTCCCAGCGC 1000
ATCGGGTCCG TCGACCATCT CGAGGGGACG ACCCACCGGG ACGTCGGCTC 1050
GATCGCGCGG GAGGACGGTC CGGTCCTTCA GCTTCTCGGG ATCTTCGTGC 1100
ACCCCTCGGT GCAACACCTG GGGATCGGCG ACCAGCTCCT CGAGCTCGCC 1150
CTTCAGGTCG CCGAGGTCAC CGCGGAGATC CAGCTCGTTG CCGGCGTCAC 1200
GCGATGCAGG AACTACCCCA AGCACCGGGA CCTCTCCTTC GAGGAGTACG 1250
TAGAGAGCAT GGGGAGCTCG AGCGTCCCGG TCGACCCCAT CCCTCTGATG 1300
CACACGAGCC ACGGCGCCCG GATCGTCAAA GTGCTCGCGG GGCATCGACC 1350
CGAGGACGTC GAGAACGACG GGGCGGGCGT TCTCGTGACC TACGACGTTC 1400
GCGGGGGCGG TGGGGGCAGA GCCCCTTCGA CCGAGCTCGG GAAGCGAGGT 1450
CCGGGGATCG ACGTCGCCGA CGAGGGGTCG CTCGCGGCGA AGATCGAGAA 1500
AGCGGTCCGA AAGGTGCTTC GCCCCGAGCG CGAGCGCGTC TTCTCGTCGA 1550
GCCGCGCTCT GATGGACATG GGCTTCGACT CGGCCGACCT CCTCGAGCTT 1600
CGGGTCCTGC TGGGTGATCT CTTTGCTCTG GAGCTGGATC CGACGCTCTT 1650
CTTCCGATAC CCCACCGCCG ACGCGCTCGC GAGCTACCTG GCGGGACGAG 1700
GGAAGGTCGA GCCCACGCCT TCGCCGACGC CCTGGCTCGC GCCGGTCGAG 1750
AGGACGCCGA GCTTTGCGCG GTCGGAGCCG CACCGGCGCG ACGTACAACC 1800
GCGAGCGACG GGCTCGAGCC GGATCGCGAT CGTGGGAATC GGCTGCCGGT 1850
TTCCCGGCGC CTCGAACGTC GGTCGGTACT GGGATCTCCT GCGCGAGGGA 1900
CGCGATGCCG TCGCCGAAGC GCCCGGGGAT CGACCCTGGC TCCTCGAGTC 1950
TTCCCACGCT CGCTTCGGCG GTTTTCTCGA AGCGGTGGAC CGGTTCGACG 2000
CGTCGCTCTT CGGCTTGTCT CGGCGCGAGG CGAACCAGGT CGATCCTCAG 2050
CAGCGTCTCC TTCTCGAGAC CTGCTGGACG GCCCTCGAGA ACGCGGGCAT 2100
CGCTCCCCCT TCACTGGAAG GCACGAGGAC GGGCGTCTTT CTGGGGATCA 2150
TGTCGCACGA CTACGAGCTC TTGCAGGTTC GGCGCGGCCA GGTCGGCGCG 2200
AAGGCCGACC CCTACTTCGC CTCCGGCAAC TCTCTCGCCG TCGCCGCAGG 2250
```

Figure 4.26 Insert nucleotide sequence obtained from primer walking sequencing. The conserved characteristics motif for type I PKS KS domain gene were highlighted in yellow, green and blue for MDPQQR conserved, cysteine active site , and histidine active site motif, respectively.

```
CCGCCTCGCC TACGTGTTCG GCTTTCGGGG GCCGGCCATC TCGGTGGATA 2300
CGGCCTGCTC CTCGTCGCTC GTCGCGGTTC ACCTCGCCGC CGAGAGCCTT 2350
TTTCGCGGTG AGTGCGAGGT CGCGATCGCG GCCGGCGTTC AGCTTCTGCT 2400
CGCTCCCGAG CTCACGGCCT CCTACGCCAA GGCGGGTATG TTGTCGCCGG 2450
ACGGAAGGTG CAAGACCTTC GACGCCTCCG CGAACGGCTA CGTGCGGAGC 2500
GAGGGGGTGG GAGCGGTCGT GCTCAAGCGG CTCGAGGACG CGCTTCGCGA 2550
CGGCGACGAC GTGCTGGCGG TCGTTCGCGG CTCGGCGCTC AACCAGGACG 2600
GGAGCAGCAA CGGGCTCACG GCGCCGAGCG CCGCCGCCCA GGAAGAGGTC 2650
ATCCGGGAAG CGCTCCGAAG GGCCGAGCTG AGCCCCTCCG AGATCTCTTA 2700
CGTGGAAGCT CATGGCACGG GGACGCCGCT CGGAGATCCG ATCGAGTTCG 2750
ACGCGCTCGC CGCCGTCTAC GGCGAGTCCC GGGACGAAGA GAACCCGCTG 2800
CACCTGGGCT CGGTCAAGAC GAACATCGGA CATACCGAGG GCGCGGCCGG 2850
CATCGCGGGC CTCATCAAAG TCGCGCTCGC CTTGCGTCAC GCCACCATCC 2900
CCGCGCACCT CCACTTCGAG CGGCCGAACC CCTCGATCGA TCTGACGAGG 2950
ATCCCGGCAC GCATTCCCCG CGAGAGCCTG GGCTGGTCTC TCGGGAAAGG 3000
CGCGACCAGC CGCCTGGCCG CGGTCAGCGC CTTCGGCTTC AGCGGCACCA 3050
ACGCCCACGT CGTGCTCGAG GAGTGGCTCG CTCCGCCGCC GCCGGTGCCG 3100
AGCGGTCGTC GACCCCTGCA CCTGCTCACG GTCTCGGCCG CGAGCGAGGC 3150
CTCGCTTCAC GAGCTCGCCT CGTCGTACGC CGAGCGCCTT CGACGCGACG 3200
AGACGGACTT CGGCGACTTC GCCTTCAGCG TCAACACCGG GCGCGCCCAA 3250
TTCGAAGAGC GCTCCGTGAT CCTCGCCGGC TCGAGCGCCG AAGCGGCGGG 3300
AAAGATCGAC GGCCTGTCGA AGGCGGGGGC GGTCGAGCCG GGTGTGCACC 3350
GAGCTCGCGT CTCTCCCGCC GCTTCGCCCA CCATCGCGTT TCTCTTCACC 3400
GGCCAGGGCT CGCAGTTCGC GGAGATGGGA AGACAGCTCT TCGCGAGCGA 3450
GCCCGTGTTC CGCCGCACCC TCGAGAGATG CGACGAGCTC CTGACGGACC 3500
GGCTCGCGGG CGGACTTCTC GCCGCCCTGT TCCCTTCGAT GGAGGGCGGT 3550
GTGCCCGCCG CCCCCATCGA CGAGACCGCC TATACGCAGC CGGTGCTATT 3600
CGCTTTCGAG TATGCCCTCG CCGAGCTCTG GCGCTCGTGG GGCGTCGAGC 3650
CCGCCTTCGT TCTCGGACAC AGCGTGGGAG AGTACGTCGC CGCTTGCGTC 3700
GCGGGCGTTC TCGACCTGGA GGGCGCGCTC GATCTCGTGG CCGAGCGCGG 3750
GCGGCTCATG CAGGAGCTGC CCGAGAAAGG CGCCATGGCC GCCGTCGCGG 3800
CGACGGAGTC CGAGGTCGCG GCCTGGATCG GAAGCTTCGG AGACGATCTC 3850
TCGGTAGCCG CGGTCAACGC GCCGTCGAGC GTCGTGGTCT CCGGCCGGTG 3900
CGAAGCTTTG AACGAGCTCC AGAAAGATCT CGAATCGCGC GGCGTTCGGG 3950
TGAGACGCCT CAGGGTCTCG AACGCCTTCC ACTCCGCCTT GATGGAGCCG 4000
ATGCTCGGCG CGCTCGAGGC CGAGGCCGCG AAGCTCGAGC TGCGAGCTTC 4050
GTCCATCCGG CTCGTCTCGA ACCTCGACGG CCGGTTCGTG GAGCCCGAGT 4100
CCCTGACCCC CGACTATTGG GCCCGGCACG CACGCCGCGC GGTGCGCTTC 4150
TCTGACGGGA TCCGTACGCT CGTCGCGAAC GGGTGCGACA CCTTCCTGGA 4200
GATCGGTCCG GGAACGACGC TCCTGACCCT CGGCCGGATG TCGGCGGGTG 4250
AGGTGCCGGC GCTTTGGCTT CCGAGCCTCC GTCCCGGGGG CACGGATTGG 4300
GAGTCGATGC TCAGCTCGCT CGCCGAGCTC CATCTCCGGG GGGTGAACAT 4350
CGACTGGCTC CGCTACGACG CGCCTCACGC GCGTCGGCGT ATCGAGGTTC 4400
CCAACTACCC GTTCTCGCGG GAGCGCTACT GGTTCGACGA GGGGTCGGAG 4450
GGATCTTCCG AGGAGAGC 4468
```

Figure 4.26 Insert nucleotide sequence obtained from primer walking sequencing. The conserved characteristics motif for type I PKS KS domain gene were highlighted in yellow, green and blue for MDPQQR conserved, cysteine active site , and histidine active site motif, respectively. (Continue)

Blastp using deduced amino acid of an entire sequence as query sequence returned with the polyketide synthase module of CurA from *Lyngbya majuscula* 3L (AEE88289) and beta-ketoacyl synthase of *Burkholderia ambifaria* AMMD (ABI91470) as the most homologous protein on the public database with 45.7 and 41.4% identity. The result from blastx using the nucleotide sequence as a query sequence, those two most homologous proteins from blastp were returned as top 5 most homology, and the top blast was PKS from *Myxococcus xanthus* DK 1622, with 41.0% identity of very low query coverage. There were other blastx results with higher % homology, but the query coverage percentage was much lower, implying the smaller portion of query was used in comparison during the search for homologous sequences. These low query coverage results might not be the reliable representative of the whole protein molecule. Using the blast results and the characteristic motifs of entire KS proteins (N-terminal motif, EPIAIVGACD; motif between histidine active site and C-terminal motif, KTNIGHL; and the terminal motif, GTNAHVIEE (Pawlik *et al.*, 2007), the whole KS protein was located with quite distinct N-terminal motif, SRIAIVGIGC (1817-1845, 5'-AGCCGGATCGCGATCGTGGGAATCGGCTGC-3') and C-terminal motif of GTNAHVVLE (3044-3070, 5'-GGCACCAACGCCCACG TCGTGCTCGAG-3'). Using only the whole KS portion and the core conserved part of KS as queries sequences, blastx was performed along side with blastp using their deduced amino acid sequences. The results from the core conserved sequence of KS domain were the same for both blast programs; the KS domain within the insert of pKKFOS_KSI_pool37_088 was homologous to type I-PKS KS domain of uncultured bacteria with 65.5% identity. For the homology search of the whole KS domain with blastp and blastx, the top 10 hits returned were similar. They were PKS KS domain from mainly *Proteobacteria*, *Cyanobacteria*, and few from firmicutes, with the highest homology at 58.9% identity [beta-ketoacyl synthase from *Burkholderia ubonensis* Bu, (ZP_02381208)]. The homology of type I-PKS KS domain with those sequences in the public database suggested that the KS domain in pKKFOS_KSI_pool37_088 fosmid potentially belonged to new PKS.

Moreover, from the blast results of the obtained 4,468 bp sequence, the flanking regions of KS domain might also belong to other domain of PKS. These portions of the sequence were investigated using blastp. The N-terminal characteristic

motif of AT domain (AFLFTGQGAQR) was found. The sequences AFLFTGQGSQF located at nucleotide position 3,386-3,418, but the C-terminal motif GVAVDWXXA could not be located. The deduced amino acid sequence of the insert, then, was used as query to search against conserved domains database (CDD) (Marchler-Bauer *et al.*, 2005; Marchler-Bauer *et al.*, 2007; Marchler-Bauer *et al.*, 2009; Marchler-Bauer *et al.*, 2011). It was also submitted to PKS/NRPS Analysis Web-site (Bachmann & Ravel, 2009), and at Structure Based Sequence Analysis of Polyketide Synthases (SBSPKS) web services (Yadav *et al.*, 2003a, 2003b; Ansari *et al.*, 2004; Yadav *et al.*, 2009; Anand *et al.*, 2010) to be analyzed for the PKS/NRPS domains and domain organization within the sequence. The results were all agreed that the domains on the sequence, from N-terminal, were NRPS T domain, PKS KS domain, and PKS AT domain (Fig. 4.27). The blast results for each of these individual domains, confirm the novelty of the sequence and its deduced amino acid sequence, as the highest similarity to the sequences on GenBank for each domain was lower than 60 % identity. Despite being multifunctional domain protein with three intact obtained domains, this sequence has not contained a complete single gene and its deduced amino acid was not an intact protein. Additional sequencing was needed to obtain the intact coding region of this gene or sequence. However, as informed by the sequencing operator, the sequencing step had advanced into the higher G+C rich region, a lot of trial and

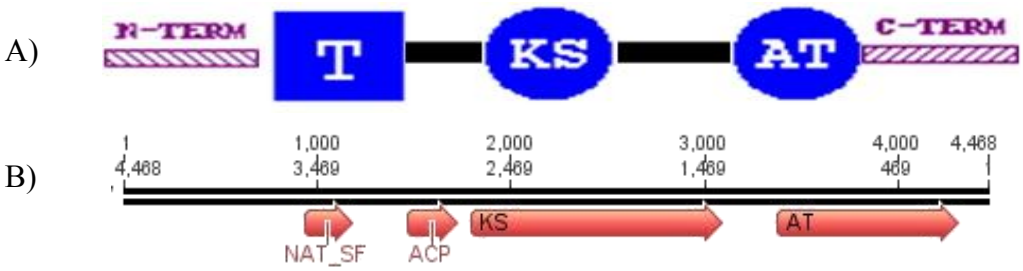


Figure 4.27 The diagram showing the the enzymatic domains predicted by (A) SBSPKS, and (B) CDD within the pKKFOS_KSI_pool37_088 insert DNA sequence obtained by primer walking. (NAT_SF: N-acetyl transferase superfamily; ACP: acyl carrier protein (predicted as T by SBSPKS); KS: ketosynthase; and AT: acyl transferase)

error for primer design and sequencing, implying that the unsuccessful sequencing often occurred and must be acceptable. Besides highly expending of precious DNA template, a lot more time was spent for little sequencing progression. Alternative sequencing methods was considered.

4.4.2.3 Identification of novel gene involving in bioactive compound biosynthesis within the recombinant fosmid insert by Next Generation Sequencing

With the support from Professor Vithaya Meevootisom, Ph.D., Mahidol University, Thailand, the entire inserted DNA sequence, almost 30 kb long, of clone KKFOS_pool_37_088 was generated from a single run using next generation sequencing, the mass sequencing technology allowing many megabases of DNA to be sequenced at once. The procedure for 454 GS Junior, the small scale, system (Roche) capable of at least 35 and possibly up to 70 megabases sequencing per run was briefly described in section 3.8.2. The obtained sequences with average 68.1% G+C nucleotide were assembled into the single contiguous fragment of DNA from 3,981 read fragments with the average dept of 50 folds for each concluded nucleotide (Fig. 4.28). The insert DNA was analyzed for ORFs with the size over 800 bp that begin with start codon in all frame. The 24 of them were found with the size range from 840 to 9,921 bp and used as query sequences for blastx. Eight ORFs were putatively annotated according to its top hit blast results (Fig. 4.29). Two largest open reading frame, the ORF 5 and 7 were putatively identified as homologous sequence to NRPS and PKS genes, respectively. Their domain organizations were determined by web based tools for analysis of conserved domain [CDD], and PKS/NRPS domain organization ["DOMAIN SEARCH PROGRAM for NRPS and PKS" from SBSPKS (Yadav *et al.,* 2003; Ansari *et al.,* 2004; Yadav *et al.,* 2009; Anand *et al.,* 2010), and "PKS/NRPS Analysis" (Bachmann & Ravel, 2009)], the functional domains of ORF 5 and ORF 7 were deduced and illustrated (Fig. 4.30, 4.31).

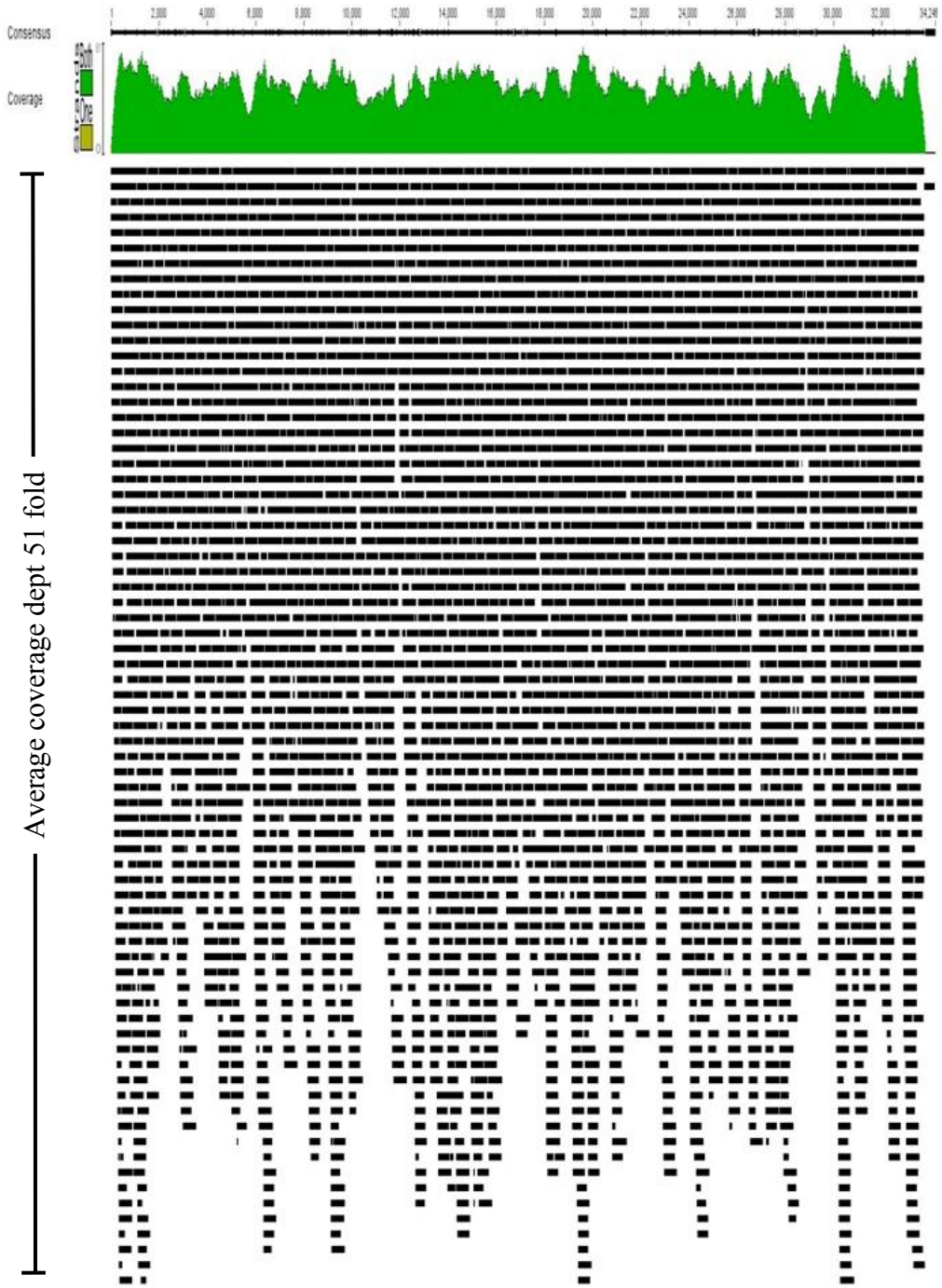
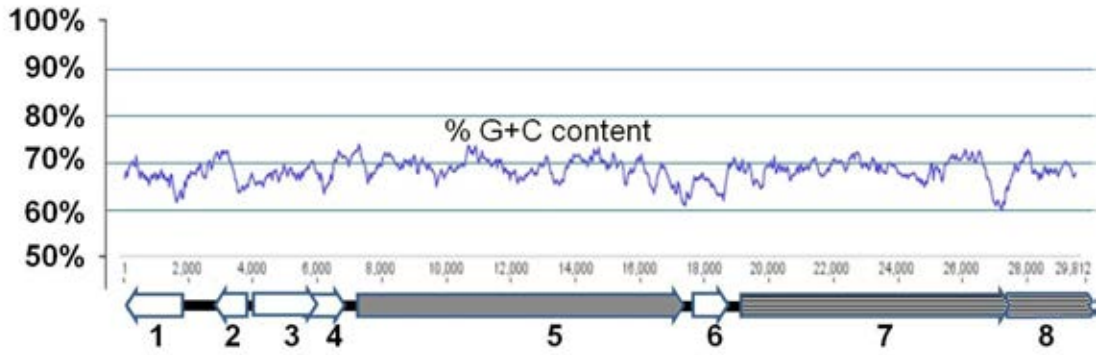


Figure 4.28 A contiguous consensus fragment of the insert DNA from fosmid clone. The 3,892 DNA fragments generated from GS junior 454 sequencing system were assembled into a single contig with the size of 34,248 base long by Geneious software version 5.4.6 (Biomatters). The average dept of the assembly was 51 bases.



ORF	Size (bp)	Top Blast hit	Closest taxon	E-value	Identity (%)
1	1587	integral membrane protein MviN (YP_003843702)	*Clostridium cellulovorans* 743B	2.7e-40	26.4
2	852	hypothetical protein Vapar_3187 (YP_002945071)	*Variovorax paradoxus* S110	2.9e-33	49.7
3	1887	putative ABC transporter ATP-binding/permease protein (YP_002763257)	*Gemmatimonas aurantiaca* T-27	0	50.0
4	807	Microcystin synthetase associated thioesterase (ZP_07114033)	*Oscillatoria* sp. PCC 6506	2.9e-63	43.0
5	9921	non ribosomal peptide synthetase (Chondramide synthetase; CmdD) (Q0VZ70)	*Chondromyces crocatus*	0	38.6
6	879	taurine catabolism dioxygenase TauD/TfdA (YP_001866800)	*Nostoc punctiforme* PCC 73102	1.6e-18	25.2
7	8055	polyketide synthase module (CurL) (ZP_08432360)	*Lyngbya majuscula* 3L	0	39.8
8*	2594	JamK (AAS98782)	*Lyngbya majuscula*	0	51.0

Figure 4.29 Predicted ORFs and their putative function identified according to their Blast results. Striped arrows represent those proteins related to polyketides biosynthetic pathway. Gray arrow was ORF likely to encode for NRPS. For others, the ORFs were in white.

* the incomplete ORF 8.

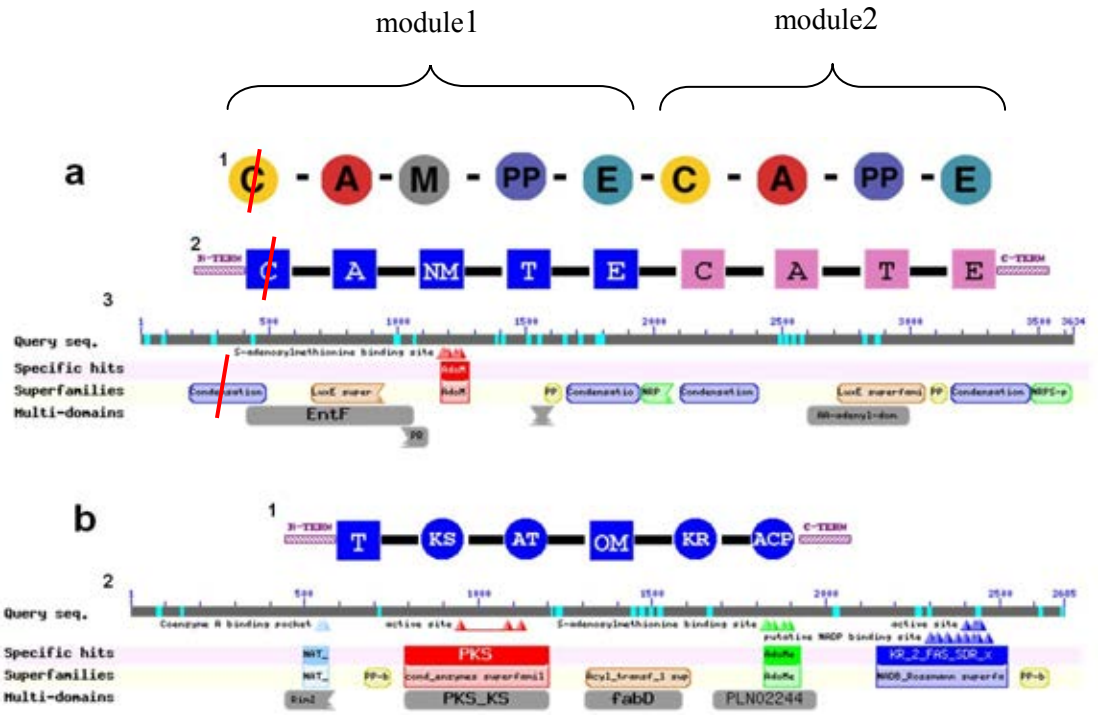


Figure 4.30 Domain organization of predicted ORF5 (a) and ORF7 (b). Using deduced amino acid from ORF5 and ORF7, their domain organizations were analyzed and predicted using web based analysis tool from (a1) PKS/NRPS Analysis Web-site (Bachmann & Ravel, 2009), (a2 and b1) Structure Based Sequence Analysis of Polyketide Synthases (SBSPKS) web services (Yadav *et al.*, 2003a, 2003b; Ansari *et al.*, 2004; Yadav *et al.*, 2009; Anand *et al.*, 2010), and (a3 and b2) CDD search service at NCBI (Marchler-Bauer *et al.*, 2011). There were two modules of NRPS containing ORF with incomplete N-terminal condensation (C) domain (a) as depicted in Fig. 4.31, and one PKS module with NRPS T domain at N-terminal (b).

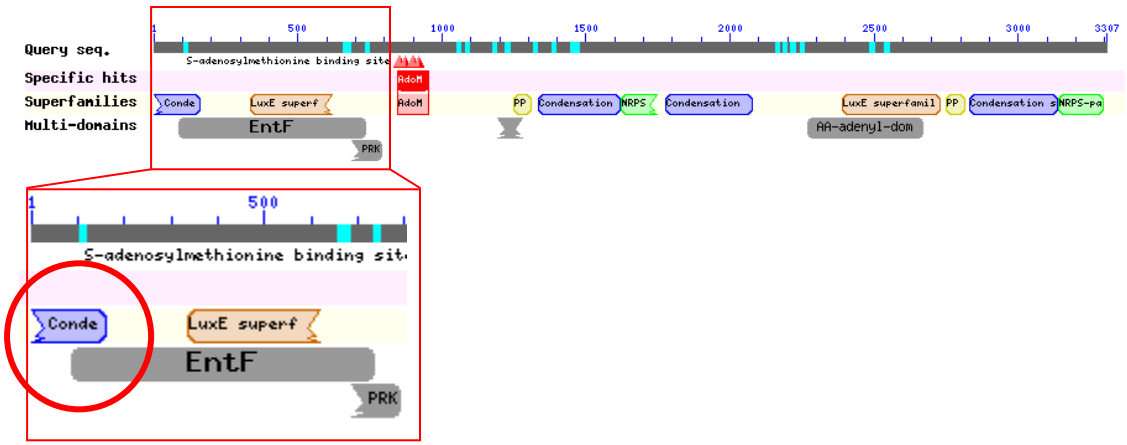


Figure 4.31 The putative functional domains of ORF 5 predicted by conserved domain database (CDD) searching results. It revealed the truncated of N-terminal condensation (C) domain (circled) of the putative ORF 5.

The ORF 5 was predicted to locate between nucleotide number 7,230 and 17,151 of the 29,812 bp insert DNA. Its top hit blastp result with the significant homology (e-value 0), ORF 5 was only about 39% identical to CmdD protein, the part of enzyme involving in potent anticancer agent, chondramide, biosynthesis. Revealed by CDD (Fig. 4.31), this predicted ORF contained the truncated N-terminal condensation (C_1) domain. This ORF was predicted to contain two NRPS modules orderly organized from N-terminal as incomplete condensation1 (C_1), adenylation1 (A_1), methyltransferase1 (M_1, NM_1, or MT_1), peptidyl carrier protein1 (PCP_1, PP_1, or T_1), epimerization1 (E_1), condensation2 (C_2), adenylation2 (A_2), peptidyl carrier protein2 (PCP_2), and epimerization2 (E_2). According to blastp, C2 was similar to the condensation domain within the amino acid adenylation domain protein (ZP_08494730) from *Microcoleus vaginatus* FGP-2 (ZP_08495279) with 48% identity (e-value of 4e-121). Its active site residues "HHVILDG" were similar to "HHIILDG" active site residues from McyB of microcystin biosynthesis enzyme, which catalyzes the condensation between D-alanine and leucine. Similarly, the

selected and activated putative amino acid to be attached to the thiolation (T) domain of NRPS could be predicted by amino acid residues at core motif involving in the catalytic pocket of the adenylation domains (A) (Challis, Ravel, & Townsend, 2000; Ansari, 2004; Rausch, 2005; Anand, 2010; Rottig, 2011). It was found that A_1 with the core motif of "D A F W L G G T" was predicted to select and activate alanine or D-valine by SBSPKS, or valine by "PKS/NRPS Analysis". For the E_1 (epimerization) domain with the active site residue of "HHLVVDV", it was predicted to epimerize alanine. A_2 with the core motif of "D A W F L G H V" was predicted to select and activate leucine by all three bioinformatics softwares. These suggested that the A_1 domain of the deduced NRPS from KKFOS_pool37_088 insert should tentatively activate alanine which would be epimerized by E_1 domain. The D-alanine was then condensed with A_2-selected leucine by C_2 domain. This concerted function might be tested by subcloning of the ORF 5 into expression plasmid, and transform into suitable host. The heterologous host as *E. coli* could be used for convenience. Subsequently the expressed enzyme could be tested for amino acid-dependent ATP-pyrophosphate exchange reactions of radiolabeled [^{32}P] pyrophosphate in the selection and activation of the amino acids (Edwards *et al.,* 2004; Binz *et al.,* 2010). For E_2 domain, however, its active site residue "HHLVVDA" was predicted to be closer to E domain catalyzing the epimerization of alanine than the one of leucine. This also might be tested by TLC of the epimerized targeted amino acid (Stachelhaus & Walsh, 2000).

To confirm and evaluate the novelty of the predicted functional domains, their deduced amino acid sequences were used as query sequences for blastp operation. The result of C_2 domains was mentioned above. For A_1 and A_2 domains, their most homologous counterparts were amino acid adenylation domain-containing proteins from *Nostoc punctiforme* PCC 73102 (YP_001869792) with 51% identity (e-value 3e-158), and gramicidin synthethase subunit C (Q70LM5) from *Brevibacillus parabrevis* with 49% identity (e-value 8e-137), respectively. For methylation (M) domain, the N-methylation (NM) is catalyzed by methyltransferase (MT), these interchangeable names were used in different research groups (Ansari *et al.,* 2004; Fischbach & Walsh, 2006; Bachmann & Ravel, 2009; Anand *et al.,* 2010). Protein from public database with the highest homology to M_1 domain, according to blastp, was S-

adenosylmethionine-dependent methyltransferases of amino acid adenylation domain protein from *Pedosphaera parvula* Ellin514 (ZP_03627826) with 44% identity (e-value 3e-91). While T_1 domain was similar to phosphopantetheine attachment site from *Anabaena variabilis* ATCC 29413 non-ribosomal peptide synthethase (YP_324595) with 67% identity (e-value 3e-21), T_2 domain was similar to one from *Streptomyces violaceusniger* Tu 4113 (YP_004815010) with 69% identity (e-value 4e-19). The most homologous E domain from public database to E_1 and E_2 were the ones of amino acid adenylation domain-containing proteins from *Nostoc punctiforme* PCC 73102 (YP_001866471) with 38% identity (e-value 6e-94), and YP_001866468 with 43% identity (e-value 2e-120), respectively. Besides T_1 and T_2 domains which exhibited about 67-69 % identity, all other domains showed only about 50% identity to their closest relative proteins.

In comparison between ORF 5 and its top hit blastp result, the differences could be distinguished at the C-terminal. While CmdD (Fig. 4.32) was ended with thioesterase (TE) indicating the end of the assembly-line (Rachid *et al.,* 2006), ORF 5 (Fig. 4.30) ended with epimerization2 (E_2) indicating further step of the assembly-line, which was most likely beginning with $^{D}C_L$ (the condensation of D-amino acid epimerized from L-amino acid) (Fischbach & Walsh, 2006). In addition, it had been experimentally confirmed that A_6, the first A domain of CmdD with core motif "D G V Q M A G V" selected and activated L-tryptophan, and A_7, the second A domain with core motif "D G S T I T A V", involved in tyrosine selection and activation (Rachid *et al.,* 2006). These emphasized the potential functional novelty of the NRPS modules in ORF 5. Their currently most homologous NRPS modules available in the public database selected and activated the substrates to be incorporated into the intermediate carbon backbone molecule differently from the predicted ones for NRPS modules in ORF 5. Despite their similar domain organization, their function and products were highly possibly different.

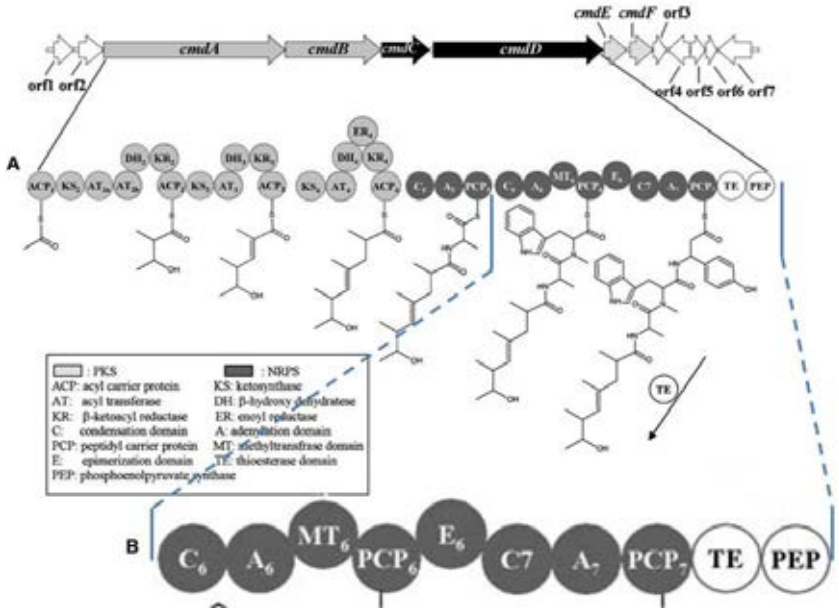


Figure 4.32 Gene cluster responsible for chondramide biosynthesis. (A) Modules and domain organization for condramide biosynthetic gene cluster. (B) Domain organization of CmdD module (modified from Rachid *et al.,* 2006).

For ORF 7 located at nucleotide number 19,162 to 27,216 of the pKKFOS_KSI_pool37_088 insert DNA, it was predicted as a single PKS module comprised thiolation (T), ketosynthase (KS), acyltransferase (AT), methyltransferase (M), ketoreductase (KR), and acyl carrier protein (ACP) organized in order from N-terminal, as shown in Fig. 4.30b,. From its predicted KR domain, it was expected one step reductive activity of beta-keto group by this PKS module. Indicated by blastp and psi-blast, this PKS module was homologous to CurL, the module before the termination module of enzyme involving in another anticancer agent, curacin A (Fig. 4.33), with about 40% identity. By using the "DOMAIN SEARCH PROGRAM for NRPS and PKS" from SBSPKS, the predicted domain organization of ORF 7 (Fig. 4.30) and CurL (Fig. 4.33) have been shown to be very similar, except for the N-terminal of ORF 7. It was predicted as thiolation domain (T) (which was the same to PCP, or PP). This indicated that the substrate for enzyme encoded from ORF 7 would most likely be aminoacyl- or peptidyl- molecule, while it was most likely an acylated

short carboxylic molecule for CurL. This T domain on ORF 7 exhibited the most

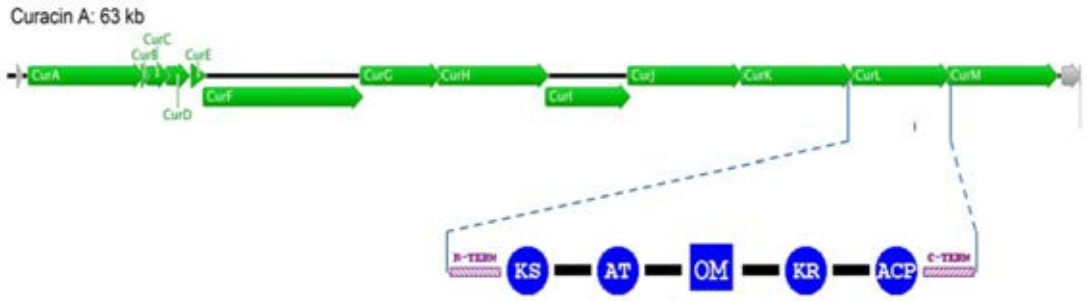


Figure 4.33 Curacin A biosynthetic gene cluster, showing domain organization of CurL module. (modified from Jones *et al.*, 2011).

homology to the phosphopantetheine attachment site of putative type I polyketide synthase from symbiont bacterium of *Paederus fuscipes* (ABM63529) with 47% identity (e-value 4e-10). KS and S-adenosylmethionine-dependent methyltransferases (OM) domain were found to be similar with ones from *Lyngbya majuscula* 3L but from different module. KS was similar to one from CurA with 58% identity (e-value 5e-159), while methyltransferase domain was similar to one from CurL 48% identity (e-value 7e-78). The putative AT domain in ORF 7 was homologous to the acyltransferase domain of polyketide synthase from *Oscillatoria* sp. PCC 6506 with 53% identity (e-value 1e-95). Top hit of blast result for ketoreductase (KR) domain of ORF 7 was returned with beta-ketoacyl reductase (KR) domain of polyketide synthase from *Nostoc* sp. PCC 7120 with 47% identity (e-value 1e-49). The final domain, acyl carrier protein (ACP) domain, of ORF 7 was identified as low complexity and was inappropriate to be used as a query in blast searching because its results are not produced from the sequences sharing homology. Therefore, its significant similarity could not be found.

However, for those domains that could produce blast results with significant homology, the blast results confirmed their belonging functional domain to be as predicted. Although, this ORF 7 was significantly homologous to CurL, they likely catalyzed the different product biosynthesis. The CurL upstreamed module, the PKS

module (CurK) (Chang *et al.,* 2004), has been reported to deliver polyketide intermediate for CurL to be condense onto it selected ketide molecule. On the contrary, the T domain at the N-terminal of ORF 7 was predicted to bring the aminoacyl or peptidyl molecule to be condensed on ketide molecule. These two homologous modules seemed to function differently.

These comparison of domain organization and putative functionality of the predicted ORFs within the insert DNA of KKFOS_pool37_088 to their top hit blast results, they seemed most likely different, and suggested the novelty of PKS and NRPS modules from Klongkone mangrove mud metagenome.

CHAPTER V

CONCLUSION

Cultural-based have proved microorganisms are the excellent source for bioactive compound needed in drug development (Newman, and Cragg, 2007). However, there are about 99% of them that have not been exploited due to unsupported current cultivation technology (Torsvik *et al.,* 2002). Metagenomics, the culture independent technology have been developed to study the collective genomes of the whole microbial community in the environmental sample (Handelsman *et al.*, 1998). This technology could reach into those microorganisms that have not been cultured, so their genes involving in bioactive compound biosynthesis could be accessible.

In this study, metagenomics was the tool of choice, and the environmental source was mangrove forest soil from Klongkone, Samut Songkhram Province, Thailand. The metagenome was extracted directly using detergent and heat as cell disrupting agent. The obtained about 30-40 kb in length of purified DNA was used in finding type I PKS KS gene, observing the bacterial source of the metagenome, and constructing the fosmid library. From the metagenome, 12 partial KS PCR amplicons were homology to type I-PKS KS gene on the public database with % identity ranging from 53.3 to 78.3. Phylogenetic tree also suggested that they were all type I PKS KS domain genes, and most of them were possible hybrid PKS/NRPS. In addition, KSs from KK_KSI_017 and 027 were phylogenetically grouped as preceding NRPS hybrid, and each of them was placed on the different sister clades. Hybrid PKS/NRPS I-I and hybrid PKS/NRPS I-II was proposedly assigned to each clade for KK_KSI_027 and KK_KSI_017, respectively, according to the variation on their cysteine active site motif. Moreover the hybrid PKS/NRPS III was also proposed based on the appearing of "NRPS domains-PKS-NRPS domains" formation on the PKS module which definely placed on the distinctive phylogenetic clade of KS gene.

For 16S rDNA obtained from Klongkone mangrove mud metagenome, a total of 7 out of 14 sequences exhibited homology below 97% to their top hit Blast results, suggesting highly possible new species. Phylogenetic analysis showed that there were

at least 6 different phyla in the metagenome. The actinobacteria were not included, which is possible due to the extraction bias (Jiang *et al.,* 2011). These had been implied that the Klongkone metagenomic DNA came from about 6 different, mainly *Proteobacteria*, phyla; so, did the type I PKS KS genes.

In library construction the library had about 14,000 clones covering about 500 Mb of genomic DNA. About one third was undergone functional base screening without success. Other strategy, like antibiotic susceptibility test might be more convenient and high throughput. As those who produce antibiotics should have the mechanisms to protect themselves.

Using type-I PKS KS gene specific PCR, one positive clone in the library was detected by this sequence based technique. The insert DNA was sequenced and revealed two complete open reading frames being modular multifunctional domain genes. One was a module of PKS. Another contained two modules of NRPS. This insert is highly possible to be novel and a part of hybrid PKS/NRPS gene cluster.

This Klongkone metagenome could be used as a source for finding new gene involving in bioactive compound biosynthesis via sequence based screening, and it could be used for accessing bacterial diversity as well.

REFERENCES

Alain, K., and Querellou, J. (2009) Cultivating the uncultured: limits, advances and future challenges. *Extremophiles* 13: 583-594.

Altschul, S.F. & Lipman, D.J. (1990) Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* 87: 5509-5513.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.

Anand, S., Prasad, M.V., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., and Mohanty, D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res* 38:W487–496.

Ansari, M.Z., Yadav, G., Gokhale, R.S., and Mohanty, D. (2004) NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res* 32: W405-413.

Bachmann, B.O., and Ravel, J. (2009) Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol* 458: 181-217.

Badji, B., Zitouni, A., Mathieu, F., Lebrihi, A., and Sabaou, N. (2006) Antimicrobial compounds produced by *Actinomadura* sp. AC104 isolated from an Algerian Saharan soil. *Can J Microbiol* 52: 373-382.

Barrera-Alba, J.J., Gianesella, S.M.F., Moser, G.A.O., and Saldanha-Corrêa, F.M.P. (2007) Bacterial and phytoplankton dynamics in a sub-tropical estuary. *Hydrobiologia* 598: 229-246.

Barrett, C.T., and Barrett, J.F. (2003) Antibacterials: are the new entries enough to deal with the emerging resistance problems? *Curr Opin Biotechnol* 14: 621-626.

Beja, O., Suzuki, M.T., Koonin, E.V., et al. (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2: 516-529.

Bérdy, J. (2005) Bioactive Microbial Metabolites. *J. Antibiot*. 58: 1-26.

Bhuiyan, F.A., Nagata, S., and Ohnishi, K. (2011) Novel chitinase genes from metagenomic DNA prepared from marine sediments in southwest Japan. *Pak J Biol Sci* 14: 204-211.

Binz, T. M., Maffioli, S.I., Sosio, M., Donadio, S., and Müller, R. (2010) Insights into an unusual nonribosomal peptide synthetase biosynthesis: identification and characterization of the GE81112 biosynthetic gene cluster. *J Biol Chem* 285: 32710-32719.

Bode, H.B., and Muller, R. (2005) The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl* 44: 6828-6846.

Boggs, A.F., and Miller, G.H. (2004) Antibacterial drug discovery: is small pharma the solution? *Clin Microbiol Infect* 10 Suppl 4: 32-36.

Brady, S.F. (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protoc* 2: 1297-1305.

Brady, S.F., and Clardy, J. (2005) N-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. *Org Lett* 7: 3613-3616.

Brady, S.F., and Clardy, J. (2003) Synthesis of long-chain fatty acid enol esters isolated from an environmental DNA clone. *Org Lett* 5: 121-124.

Brady, S.F., Chao, C.J., and Clardy, J. (2004) Long-chain N-acyltyrosine synthases from environmental DNA. *Appl Environ Microbiol* 70: 6865-6870.

Bush, K. (2004) Antibacterial drug discovery in the 21st century. *Clin Microbiol Infect* 10 Suppl 4: 10-17.

Butler, M.S. (2005) Natural products to drugs: natural product derived compounds in clinical trials. *Nat Prod Rep* 22: 162-195.

Butler, M.S. (2004) The role of natural product chemistry in drug discovery. *J Nat Prod* 67: 2141-2153.

Butler, M.S., and Cooper, M.A. (2011) Antibiotics in the clinical pipeline in 2011. *J Antibiot* (Tokyo) 64: 413-425.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Challis, G. L., Ravel, J., and Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* 7: 211-224.

Chang, Z., Sitachitta, N., Rossi, J.V., et al. (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *J Nat Prod* 67: 1356-1367.

Chung, E.J., Lim, H.K., Kim, J.C., et al. (2008) Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Appl Environ Microbiol* 74: 723-730.

Clardy, J., Fischbach, M.A., and Walsh, C.T. (2006) New antibiotics from bacterial natural products. *Nat Biotechnol* 24: 1541-1550.

Conly, J., and Johnston, B. (2005) Where are all the new antibiotics? The new antibiotic paradox. *Can J Infect Dis Med Microbiol* 16: 159-160.

Committee on Metagenomics: Challenges and functional applications, National Research Council. (2007) *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet.* The National Academics Press, Washington.

Courtois, S., Cappellano, C.M., Ball, M., et al. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* 69: 49-55.

Daniel, R. (2004) The soil metagenome-a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15: 199-204.

Davies, J. (2006) Are antibiotics naturally antibiotics? *J Ind Microbiol Biotechnol* 33: 496-499.

de Lorenzo, V. (2005) Problems with metagenomic screening. *Nat Biotechnol* 23: 1045; author reply 1045-1046.

Demain, A.L. (2006) From natural products discovery to commercialization: a success story. *J Ind Microbiol Biotechnol* 33: 486-495.

Demain, A.L., and Sanchez, S. (2009) Microbial drug discovery: 80 years of progress. *J Antibiot* (Tokyo) 62: 5-16.

Dent, L.L., Marshall, D.R., Pratap, S., and Hulette, R.B. (2010) Multidrug resistant *Acinetobacter baumannii*: a descriptive study in a city hospital. *BMC Infect Dis* 10: 196.

Dewick, P.M. (2009) *Medicinal Natural Products : A Biosynthetic Approach*. John Wiley., and Sons, Chichester, U.K.

Donadio, S., Maffioli, S., Monciardini, P., Sosio, M., and Jabes, D. (2010) Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J Antibiot* (Tokyo) 63: 423-430.

Drummond, A., Ashton, B., Buxton, S., et al. (2010) Geneious v5.3, Available from http://www.geneious.com.

Dupre, J., and O'Malley, M.A. (2007) Metagenomics and biological ontology. *Stud. Hist. Phil. Biol. Biomed. Sci.* 38(4): 834-846.

Edwards, D. J., Marquez, B. L., Nogle, L.M., et al. (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chem Biol* 11: 817-833.

Elsaied, H., Stokes, H.W., Kitamura, K., Kurusu, Y., Kamagata, Y., and Maruyama, A. (2011) Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J.* 5: 1162–1177.

Elsaied, H., Stokes, H.W., Nakamura, T., Kitamura, K., Fuse, H., and Maruyama, A. (2007) Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ Microbiol*. 9: 2298-312.

Fahy, E., Subramaniam, S., Brown, H.A., et al. (2005) A comprehensive classification system for lipids. *J Lipid Res* 46: 839-861.

Fatima, F., Chaudhary, I., Ali, J., Rastogi, S., and Pathak, N. (2011) Microbial DNA extraction from soil by different methods and its PCR amplification. *Biochem. Cell. Arch.* 11: 85-90

Fischbach, M.A., and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 106: 3468-3496.

Fleck, W., Strauss, D., Koch, W., and Prauser, H. (1974) Violamycin, a new red-pigment antibiotic. *Zeitschrift für allgemeine Mikrobiologie* 14: 551–558. doi: 10.1002/jobm.19740140702

Forns, X., Bukh, J., Purcell, R.H., and Emerson, S.U. (1997) How Escherichia coli can bias the results of molecular cloning: Preferential selection of defective genomes of hepatitis C virus during the cloning procedure. *Proc Natl Acad Sci USA* 94: 13909-13914.

Fox, G.E., Wisotzkey, J.D., and Jurtshuk, P., Jr. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* 42: 166-170.

Fried, M., and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 9: 6505-6525.

Gabor, E.M., de Vries, E.J., and Janssen, D.B. (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* 6: 948-958.

Gabor, E. M., de Vries, E. J., and Janssen, D. B. (2003). Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *FEMS Microbiol. Ecol.* 44: 153-163.

Ghosh, A., Dey, N., Bera, A., Tiwari, A., Sathyaniranjan, K., Chakrabarti, K., and Chattopadhyay, D. (2010) Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline Systems* 6: 1.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., et al. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* 68: 4301-4306.

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

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

Gontang, E.A., Gaudencio, S.P., Fenical, W., and Jensen, P.R. (2010) Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria. *Appl Environ Microbiol* 76: 2487-2499.

Gullo, V.P., McAlpine, J., Lam, K.S., Baker, D., and Petersen, F. (2006) Drug discovery from natural products. *J Ind Microbiol Biotechnol* 33: 523-531.

Haefner, B. (2003) Drugs from the deep: marine natural products as drug candidates. *Drug Discov Today* 8: 536-544.

Handelsman, J. (2004a) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669-685.

Handelsman, J. (2004b) Soil-The metagenomics Approach. *Microbial Diversity and Bioprospecting* (Bull AT, ed.), pp. 109-119. ASM Press, Washington, D.C.

Handelsman, J., Liles, M., Mann, D., Riesenfeld, C., and Goodman, R.M. (2002) Cloning the Metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods in Microbiology*, Vol. 33 (Wren B., and Dorrell N, eds.), p. 241. Academic Press, New York.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-249.

Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* 48: 4688-4716.

Heuermann, K., and Cosgrove, J. (2001) S-Gal: an autoclavable dye for color selection of cloned DNA inserts. *Biotechniques* 30: 1142-1147.

Hoff, K.J. (2009) The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics* 10: 520.

Hoff, K.J., Lingner, T., Meinicke, P., and Tech, M. (2009) Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res* 37: W101-105.

Hoff, K.J., Tech, M., Lingner, T., Daniel, R., Morgenstern, B., and Meinicke, P. (2008) Gene prediction in metagenomic fragments: a large scale machine learning approach. *BMC Bioinformatics* 9: 217.

Hussein, A.M., and El-Gammal, A. (1976) Characterization of a red pigment-antibiotic produced by *Streptomyces* sp. strain NRC-C7. *Z Allg Mikrobiol*. 16: 27-32.

Inagaki, F., Nunoura, T., Nakagawa, S., et al. (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc Natl Acad Sci U S A.* 103: 2815-20.

Janda, J.M., and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 45: 2761-2764.

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

Jiang, Y.X., Wu, J.G., Yu, K.Q., Ai, C.X., Zou, F., and Zhou, H.W. (2011) Integrated lysis procedures reduce extraction biases of microbial DNA from mangrove sediment. *J Biosci Bioeng* 111: 153-157.

Jones, A.C., Monroe, E.A., Podell, S., et al. (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci U S A.* 108: 8815-20

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., and Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature* 451: 990-993.

Kathiresan, K., and Selvam, M.M. (2006) Evaluation of bebeficial bacteria from mangrove soil. *Bot Mar* 49: 86-88

Keller, M., and Zengler, K. (2004) Tapping into microbial diversity. *Nat Rev Microbiol* 2: 141-150.

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., and Janssen, P.H. (2011) Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *J Microbiol Methods* 84: 52-60.

Kim, T.K., and Fuerst, J.A. (2006) Diversity of polyketide synthase genes from bacteria associated with the marine sponge *Pseudoceratina clavata*: culture-dependent and culture-independent approaches. *Environ Microbiol* 8: 1460-1470.

Koehn, F.E., and Carter, G.T. (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* 4: 206-220.

Kristensen, E., Bouillon, S., Dittmar, T., and Marchand, C. (2008) Organic carbon dynamics in mangrove ecosystems: a review. *Aquatic Botany* 89: 201-219.

Kumar, S., Hatha, A.A.M., and Christi, K.S. (2007) Diversity and effectiveness of tropical mangrove soil microflora on the degradation of polythene carry bags. *Rev Biol Trop* 55: 777-786

LaMontagne, M.G., Michel, F.C. Jr., Holden, P.A., and Reddy, C.A. (2002) Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *J Microbiol Methods* 49: 255-264.

Lane, D.J. (1991) 16S/23S rRNA Sequencing. *Nucleic Acid Techniques in Bacterial Systematic*, (Stackebrandt E., and Goodfellow M, eds.), pp. 115-175. John Wiley, and Sons, Chichester, UK.

Langer, M., Gabor, E.M., Liebeton, K., et al. (2006) Metagenomics: an inexhaustible access to nature's diversity. *Biotechnol J* 1: 815-821.

Larkin, M., Blackshields, G., Brown, N., et al. (2007) ClustalW and ClustalX version 2 *Bioinformatics* 23: 2.

Lazzarini, A., Cavaletti, L., Toppo, G., and Marinelli, F. (2000) Rare genera of actinomycetes as potential producers of new antibiotics. *Antonie Van Leeuwenhoek* 78: 399-405.

Lefevre, F., Robe, P., Jarrin, C., et al. (2008) Drugs from hidden bugs: their discovery via untapped resources. *Res Microbiol* 159: 153-161.

Leitao, E., Moradas-Ferreira, P., and De Marco, P. (2009) Evidence of methanesulfonate utilizers in the Sargasso Sea metagenome. *J Basic Microbiol* 49 Suppl 1: S24-30.

Liang, J-B., Chen, Y-Q., Lan, C-Y., Tam, N.F.Y., Zan, Q-J., and Huang, L-N. (2006) Recovery of novel bacterial diversity from mangrove sediment. *Mar Biol* 150: 739-747.

Lin, J., Kudrna, D., and Wing, R.A. (2011) Construction, characterization, and preliminary BAC-end sequence analysis of a bacterial artificial chromosome library of the tea plant (Camellia sinensis). *J Biomed Biotechnol* 2011: 476723.

Livermore, D.M. (2004) The need for new antibiotics. *Clin Microbiol Infect* 10 Suppl 4: 1-9.

Lorenz, P., and Eck, J. (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* 3: 510-516.

Maillard, O., Corvec, S., Dantal, J., Reynaud, A., Lucet, J.C., Bemer, P., and Lepelletier, D. (2010) Emergence of high ampicillin-resistant *Enterococcus faecium* isolates in a kidney transplant ward: role of antibiotic pressure and cross transmission. *Microb Drug Resist* 16: 123-128.

Marchler-Bauer, A., Panchenko, A.R., Shoemaker BA, Thiessen PA, Geer, L.Y., and Bryant, S.H. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* 30: 281-283.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., et al. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 33: D192-196.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., et al. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35: D237-240.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., et al. (2003) CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 31: 383-387.

Marchler-Bauer, A., Lu, S., Anderson, J.B., et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39: D225-229.

Marchler-Bauer, A., Anderson, JB., Chitsaz, F., et al. (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* 37: D205-210.

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

Mori, K., Sunamura, M., Yanagawa, K., et al. (2008) First cultivation and ecological investigation of a bacterium affiliated with the candidate phylum OP5 from hot springs. *Appl Environ Microbiol* 74: 6223–6229.

Nabeel, M.A., Kathiresan, K., Rajendran, N., Ohnishi, H., Hamaoka, H., and Omori, K. (2010) Contribution by microbes to the foodweb of a mangrove biotope: the approach of carbon and nitrogen stable isotopes. *Afr J Mar Sci* 32: 65-70.

Newman, D.J., and Cragg, G.M. (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70: 461-477.

Newman, D.J., Cragg, G.M., and Snader, K.M. (2003) Natural products as sources of new drugs over the period 1981-2002. *J Nat Prod* 66: 1022-1037.

Newman, D.J., Cragg, G.M., and Snader, K.M. (2000) The influence of natural products upon drug discovery. *Nat Prod Rep* 17: 215-234.

Newman, D.J., and Hill, R.T. (2006) New drugs from marine microbes: the tide is turning. *J Ind Microbiol Biotechnol* 33: 539-544.

Paphavasit, N., Piumsomboon, A., Aumnuch, E., et al. (2002) *Ecological Assessment of Mangrove Reforestation/ Afforestation in Samut Songkharm Province on Zooplankton and Benthic Communities*, 1st ed. National Research Council of Thailand, Bangkok.

Park, E.J., Kim, K.H., Abell, G.C., Kim, M.S., Roh, S.W., and Bae, J.W. (2011) Metagenomic analysis of the viral communities in fermented foods. *Appl Environ Microbiol* 77: 1284-1291.

Park, S.J., Kang, C.H., Chae, J.C., and Rhee, S.K. (2008) Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS Microbiol Lett* 284: 28-34.

Parsley, L.C., Consuegra, E.J., Kakirde, K.S., Land, A.M., Harper, W.F. Jr., and Liles, M.R. (2010a) Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. *Appl Environ Microbiol* 76: 3753-3757.

Parsley, L.C., Consuegra, E.J., Thomas, S.J., et al. (2010b) Census of the viral metagenome within an activated sludge microbial assemblage. *Appl Environ Microbiol* 76: 2673-2677.

Parsley, L.C., Linneman, J., Goode, A.M., et al. (2011) Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiol Ecol* 78: 176-187.

Pawlik, K., Kotowska, M., Chater, K.F., Kuczek, K., and Takano, E. (2007) A cryptic type I polyketide synthase (cpk) gene cluster in Streptomyces coelicolor A3(2). *Arch Microbiol* 187: 87-99.

Piel, J., Hui, D., Fusetani, N., and Matsunaga, S. (2004) Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol* 6: 921-927.

Rachid, S., Krug, D., Kunze, B., et al. (2006) Molecular and biochemical studies of chondramide formation-highly cytotoxic natural products from Chondromyces crocatus Cm c5 *Chem Biol* 13: 667-81.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* 33: 5799-5808.

Ray, D., Sinha, S., Saha, S., et al. (2010) A preliminary sentinel surveillance report on antibiotics resistance trend of *Streptococcus pyogenes* in Kolkata Region, India. *Al Ameen J Med Sci* 3: 6.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38: 525-552.

Rondon, M.R., August, P.R., Bettermann, A.D., et al. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66: 2541-2547.

Rottig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2 - a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* 39 (Web Server issue): W362-367.

Sakami, T., Fujioka, Y., and Shimoda, T. (2008) Comparison of microbial community structures in intensive and extensive shrimp culture ponds and a mangrove area in Thailand. *Fisheries Science* 74: 889-898.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Painview, N.Y.

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harber Laboratory Press, New York.

Sarker, S.D., Latif, Z., and Gray, A.I. (2006) *Natural Product Isolation*. Humana Press, Totowa, NJ.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F., and Hutchinson, C.R. (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl Environ Microbiol* 71: 4840-4849.

Schito, G.C. (2006) The importance of the development of antibiotic resistance in Staphylococcus aureus. *Clin Microbiol Infect* 12 Suppl 1: 3-8.

Schmidt, T.M., DeLong, E.F., and Pace, N.R. (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 173: 4371-4378.

Schmitz, J.E., Daniel, A., Collin, M., Schuch, R., and Fischetti, V.A. (2008) Rapid DNA library construction for functional genomic and metagenomic screening. *Appl Environ Microbiol* 74: 1649-1652.

Schwarzer, D., and Marahiel, M.A. (2001) Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* 88: 93-101.

Selvameenal, L., Radhakrishnan, M., and Balagurunathan, R. (2009) Antibiotic pigment from desert soil actinomycetes; biological activity, purification and chemical screening. *Indian J Pharm Sci.* 71: 499–504.

Shen, B., Du, L., Sanchez, C., Edwards, D.J., Chen, M., and Murrell, J.M. (2002) Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. *J Nat Prod* 65: 422-431.

Sigmon, J., and Larcom, L.L. (1996) The effect of ethidium bromide on mobility of DNA fragments in agarose gel electrophoresis. *Electrophoresis.* 17: 1524-7.

Silakowski, B., Kunze, B., and Muller, R. (2001) Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. *Gene* 275: 233-240.

Sjöling, S., Mohammed, S.M., Lyimo, T.J., and Kyaruzi, J.J. (2005) Benthic microbial diversity and nutrient processes in mangroves; impact of deforestation. *Estuar Coast Shelf Sci* 63: 10.

Spizek, J., Novotna, J., Rezanka, T., and Demain, A.L. (2010) Do we need new antibiotics? The search for new targets and new compounds. *J Ind Microbiol Biotechnol* 37: 1241-1248.

Stachelhaus, T. and Walsh, C. T. (2000) "Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase." *Biochemistry* 39: 5775-5787.

Stackebrandt, E., Ludwig, W., Weizenegger, M., et al. (1987) Comparative 16S rRNA oligonucleotide analyses and murein types of round-spore-forming bacilli and non-spore-forming relatives. *J Gen Microbiol* 133: 2523-2529.

Stevenson, B.S., Eichorst, S.A., Wertz, J.T., Schmidt, T.M., and Breznak, J.A. (2004) New strategies for cultivation and detection of previously uncultured microbes. *Appl Environ Microbiol* 70: 4748-4755.

Streit, W.R., and Schmitz, R.A. (2004) Metagenomics-the key to the uncultured microbes. *Curr Opin Microbiol* 7: 492-498.

Stubbings, W., and Labischinski, H. (2009) New antibiotics for antibiotic-resistant bacteria. *F1000 Biol Rep* 1:40.

Sudha, S.S. (2009) Marine Environment: A potential Source for L-Asparaginase Producing Microorganisms. *IUP J Biotechnol* 3: 57-62.

Taketani, R.G., Franco, N.O., Rosado, A.S., and van Elsas, J.D. (2010) Microbial community response to a simulated hydrocarbon spill in mangrove sediments. *J Microbiol* 48: 7-15.

Tebbe, C.C., and Vahjen, W. (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol* 59: 2657-2665.

Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., et al. (2008) Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* 133: 42-49.

Torsvik, V., and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* 5: 240-245.

Torsvik, V., Daae, F.L., Sandaa, R.A., and Ovreas, L. (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* 64: 53-62.

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

Treusch, A.H., Kletzin, A., Raddatz, G., et al. (2004) Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ Microbiol* 6: 970-980.

Tyson, G.W., and Banfield, J.F. (2005) Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol* 13: 411-415.

Tyson, G.W., Chapman, J., Hugenholtz, P., et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37-43.

van Elsas, J.D., Costa, R., Jansson, J., et al. (2008a) The metagenomics of disease-suppressive soils - experiences from the METACONTROL project. *Trends Biotechnol* 26: 591-601.

van Elsas, J.D., Speksnijder, A.J., and van Overbeek, L.S. (2008b) A procedure for the metagenomics exploration of disease-suppressive soils. *J Microbiol Methods* 75: 515-522.

Venter, J.C., Remington, K., Heidelberg, J.F., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74.

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E., and Streit, W.R. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol* 69: 6235-6242.

Wang, G. (2006) Diversity and biotechnological potential of the sponge-associated microbial consortia. *J Ind Microbiol Biotechnol* 33: 545-551.

Wanunu, M., Sutin, J., and Meller, A. (2009) DNA profiling using solid-state nanopores: detection of DNA-binding molecules. *Nano Lett*. 9: 3498-3502.

Warnecke, F., Luginbuhl, P., Ivanova, N., et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450: 560-565.

Watson, J.D., Caudy, A.A., Myers, R.M., and Witkowski, J.A. (2007) *Recombinant DNA : Genes and Genomes : A Short Course*, 3rd ed. W.H. Freeman: Cold Spring Harbor Laboratory Press, New York.

Watve, M.G., Tickoo, R., Jog, M.M., and Bhole, B.D. (2001) How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol* 176: 386-390.

Webb, M.R., and Ebeler, S.E. (2004) Comparative analysis of topoisomerase IB inhibition and DNA intercalation by flavonoids and similar compounds: structural determinates of activity. *Biochem J*. 384: 527–541

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

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95: 6578-6583.

Wilfinger, W.W., Mackey, K., and Chomczynski, P. (1997) Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques* 22: 474-476, 478-481.

Willner, D., Furlan, M., Schmieder, R., et al. (2011) Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. *Proc Natl Acad Sci U S A* 108 Suppl 1: 4547-4553.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2003a) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J Mol Biol* 328: 335-363.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2003b) SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res* 31: 3654-3658.

Yadav, G., Gokhale, R.S., and Mohanty, D. (2009) Towards prediction of metabolic products of polyketide synthases: an in silico analysis. *PLoS Comput Biol* 5: e1000351.

Zengler, K., Walcher, M., Clark, G., et al. (2005) High-throughput cultivation of microorganisms using microcapsules. *Methods Enzymol* 397: 124-130.

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., and Keller, M. (2002) Cultivating the uncultured. *Proc Natl Acad Sci U S A* 99: 15681-15686.

Zhang, J., and Madden, T.L. (1997) PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res* 7: 649-656.

Zhao, J., Yang, N., and Zeng R (2008) Phylogenetic analysis of type I polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediment. *Extremophiles* 12: 97-105.

APPENDICES

APPENDIX A

Table A1 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig. 4.21) (for those that available), sorted by accession number.

Order On Tree	Accession	Name	Domain Organization
58	AAF15892	*Nostoc* sp.	KS-AT-ACP
61	AAM33468	*Aphanizomenon ovalisporum*	A-A-T-KS-AT-DH
53	AAO39789	unidentified microorganism	KS
13	AAS98782	*Lyngbya majuscula*	KS-AT-KR-ACP (NRPS)
1	AAS98783	*Lyngbya majuscula*	KS-AT-DH-ER-KR-ACP-C-A-T
6	AAW55365	*Scytonema hofmanni*	KS
20	AAW55386	*Microcoleus chthonoplastes*	KS
18	AAW55387	*Microcoleus chthonoplastes*	KS
30	AAW55402.1	*Scytonema hofmanni*	KS
56	AAX44111	*Leptolyngbya* sp.	KS
57	AAX44114	*Leptolyngbya* sp	KS
21	AAX44126	*Nostoc* sp	KS
31	AAX44132.1	*Scytonema hofmanni*	KS
36	AAX86994	*uncultured bacterium*	KS
49	AAY32964	*Sorangium cellulosum*	KS-DH…(trans-AT)
3	ABA29781	*Cystobacter fuscus*	KS-AT-KR-ACP-KS-AT-DH-KR-ACP (NRPS)
4	ABD17628	*Sorangium*	KS
34	ABH08749.1	uncultured bacterium	KS
23	ACC99565	uncultured bacterium	KS
7	ACC99568	uncultured bacterium	KS
24	ACC99586	uncultured bacterium	KS
37	ACR55656	*Maribacter* sp	KS

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree; (trans-AT) indicates lacking AT.

Table A1 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig 4.21) (for those that available), sorted by accession number. (Continue)

Order On Tree	Accession	Name	Domain Organization
42	ADD65215	uncultured bacterium	KS
47	ADE08320	uncultured bacterium	KS
16	AEH26539	uncultured Acidobacteria	KS-AT–CM-KR-ACP-KS-AT-KR-ACP
12	CAQ18829	*Chondromyces crocatus*	(module1)KS-AT-DH-KR-ACP (NRPS)
11	CAQ18829	*Chondromyces crocatus*	(module2)KS-AT-KR-ACP (NRPS)
28	CAQ48259	*Planktothrix rubescens*	KS-AT-ACP-C-A-T
60	JN22446	KKDKSI_027	KS
35	JN224477	KKDKSI_002	KS
5	JN2244779	KKmKSI_010	KS
25	JN224478	KKmKSI_003	KS
8	JN224480	KKmKSI_013	KS
55	JN224481	KKmKSI_017	KS
48	JN224482	KKmKSI_019	KS
67	JN224483	KKmKSI_021	KS
29	JN224484	KKDKSI_023	KS
38	JN224485	KKDKSI_026	KS
39	JN224487	KKDKSI_029	KS
17	JN224488	KKDKSI_033	KS
26	JN224489	KKDKSI_035	KS
43	YP_001418687	*Xanthobacter autotrophicus*	KS-AT-DH-ER-KR-ACP
44	YP_001421292	*Bacillus amyloliquefaciens*	A-T-C-A-T-KS-KR-ACP-ACP-KS-KR-ACP-KS (trans-AT)
66	YP_001544779	*Herpetosiphon aurantiacus*	A-T-KS-AT-KR-ACP-C-A-T-C

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree; (trans-AT) indicates lacking AT.

Table A1 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig 4.21) (for those that available), sorted by accession number. (Continue)

Order On Tree	Accession	Name	Domain Organization
2	YP_001657795	*Microcystis aeruginosa*	KS-AT (NRPS)
15	YP_001804493	*Cyanothece sp.*	KS-AT-DH-CM-ER-KR-ACP
10	YP_001818846	*Opitutus terrae* PB90-1	KS-CM-KR-ACP-KS-AT-KR-ACP
62	YP_001821790	*Streptomyces griseus*	A-T-KS-AT-ACP-KR-ACP-C-A-T-C
40	YP_001832244	*Beijerinckia indica* ATCC 9039	KS-AT-ER-KR-ACP
59	YP_001865726	*Nostoc punctiforme* ATCC 29133	KS-AT-ACP
50	YP_001869791	*Nostoc punctiforme* ATCC 29133	KS-AT-ACP .(AAT_I)
41	YP_002363007	*Methylocella silvestris*	KS-AT-CM-ER-KR-ACP
54	YP_003755474	*Hyphomicrobium denitrificans*	KS-AT-ACP-.(AAT_I)
33	YP_004521734	*Mycobacterium* sp. JDM601	A-T-KS-AT-KR-ACP
22	YP_324603	*Anabaena variabilis* ATCC 29413	A-T-KS-AT-
51	YP_325326	*Anabaena variabilis* ATCC 29413	KS-AT-ACP- (AAT_I)
46	YP_632113	*Myxococcus xanthus*	ER-KS-DH-KR-ACP -KS-DH-ACP-KS-DH-KR-ACP-KS
9	YP_951844	*Mycobacterium vanbaalenii*	KS-AT ACP-C-A-T-C
27	ZP_02890743	*Burkholderia ambifaria*	A-T-KS-DH (trans-AT)
64	ZP_04698103	*Streptomyces roseosporus*	A-T-KS-
65	ZP_04713303	*Streptomyces roseosporus*	A-T-KS-AT-KR-ACP-C-A-T-C-
19	ZP_05029386	*Microcoleus chthonoplastes*	A-T-KS-AT
45	ZP_06873407	*Bacillus subtilis* subsp. *spizizenii*	A-T-C-A-T-KS-KR-ACP-ACP-KS-KR-ACP-KS-KR-ACP-KS (trans-AT)
52	ZP_07108688	*Oscillatoria* sp. PCC 9029	KS-AT-ACP- (AAT_I)
63	ZP_08233860	*Streptomyces* cf.	A-T-KS-AT-KR-ACP-C-A-T-C
32	ZP_08425908.1	*Lyngbya majuscula*	KS-AT-KR-ACP
14	ZP_08429734	*Lyngbya majuscula*	KS-AT-DH-CM

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree; (trans-AT) indicates lacking AT.

Table A2 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig 4.21) (for those that available), sorted by position on the tree.

Order On Tree	Accession	Name	Domain Organization
1	AAS98783	*Lyngbya majuscula*	KS-AT-DH-ER-KR-ACP-C-A-T
2	YP_001657795	*Microcystis aeruginosa*	KS-AT (NRPS)
3	ABA29781	*Cystobacter fuscus*	KS-AT-KR-ACP-KS-AT-DH-KR-ACP (NRPS)
4	ABD17628	*Sorangium*	KS
5	JN2244779	KKmKSI_010	KS
6	AAW55365	*Scytonema hofmanni*	KS
7	ACC99568	uncultured bacterium	KS
8	JN224480	KKmKSI_013	KS
9	YP_951844	*Mycobacterium vanbaalenii*	KS-AT ACP-C-A-T-C
10	YP_001818846	*Opitutus terrae* PB90-1	KS-CM-KR-ACP-KS-AT-KR-ACP
11	CAQ18829	*Chondromyces crocatus*	(module2)KS-AT-KR-ACP (NRPS)
12	CAQ18829	*Chondromyces crocatus*	(module1)KS-AT-DH-KR-ACP (NRPS)
13	AAS98782	*Lyngbya majuscula*	KS-AT-KR-ACP (NRPS)
14	ZP_08429734	*Lyngbya majuscula*	KS-AT-DH-CM
15	YP_001804493	*Cyanothece sp.*	KS-AT-DH-CM-ER-KR-ACP
16	AEH26539	uncultured Acidobacteria	KS-AT–CM-KR-ACP-KS-AT-KR-ACP
17	JN224488	KKDKSI_033	KS

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree.

Table A2 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig 4.21) (for those that available), sorted by position on the tree. (continue)

18	AAW55387	*Microcoleus chthonoplastes*	KS
19	ZP_05029386	*Microcoleus chthonoplastes*	A-T-KS-AT
20	AAW55386	*Microcoleus chthonoplastes*	KS
21	AAX44126	*Nostoc* sp	KS
22	YP_324603	*Anabaena variabilis* ATCC 29413	A-T-KS-AT-
23	ACC99565	uncultured bacterium	KS
24	ACC99586	uncultured bacterium	KS
25	JN224478	KKmKSI_003	KS
26	JN224489	KKDKSI_035	KS
27	ZP_02890743	*Burkholderia ambifaria*	A-T-KS-DH (trans-AT)
28	CAQ48259	*Planktothrix rubescens*	KS-AT-ACP-C-A-T
29	JN224484	KKDKSI_023	KS
30	AAW55402.1	*Scytonema hofmanni*	KS
31	AAX44132.1	*Scytonema hofmanni*	KS
32	ZP_08425908.1	*Lyngbya majuscula*	KS-AT-KR-ACP
33	YP_004521734	*Mycobacterium* sp. JDM601	A-T-KS-AT-KR-ACP
34	ABH08749.1	uncultured bacterium	KS
35	JN224477	KKDKSI_002	KS
36	AAX86994	*uncultured bacterium*	KS
37	ACR55656	*Maribacter* sp	KS
38	JN224485	KKDKSI_026	KS
39	JN224487	KKDKSI_029	KS
40	YP_001832244	*Beijerinckia indica* ATCC 9039	KS-AT-ER-KR-ACP
41	YP_002363007	*Methylocella silvestris*	KS-AT-CM-ER-KR-ACP
42	ADD65215	uncultured bacterium	KS
43	YP_001418687	*Xanthobacter autotrophicus*	KS-AT-DH-ER-KR-ACP

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree; (trans-AT) indicates lacking AT.

Table A2 Accession number and Domain organization of sequences included in KS phylogenetic tree (Fig 4.21) (for those that available), sorted by position on the tree. (continue)

44	YP_001421292	*Bacillus amyloliquefaciens*	A-T-C-A-T-KS-KR-ACP-ACP-KS-KR-ACP-KS (trans-AT)
45	ZP_06873407	*Bacillus subtilis* subsp. *spizizenii*	A-T-C-A-T-KS-KR-ACP-ACP-KS-KR-ACP-KS-KR-ACP-KS (trans-AT)
46	YP_632113	*Myxococcus xanthus*	ER-KS-DH-KR-ACP -KS-DH-ACP-KS-DH-KR-ACP-KS (trans-AT)
47	ADE08320	uncultured bacterium	KS
48	JN224482	KKmKSI_019	KS
49	AAY32964	*Sorangium cellulosum*	KS-DH...(trans-AT) (NRPS)
50	YP_001869791	*Nostoc punctiforme* ATCC 29133	KS-AT-ACP .(AAT_I)
51	YP_325326	*Anabaena variabilis* ATCC 29413	KS-AT-ACP- (AAT_I)
52	ZP_07108688	*Oscillatoria* sp. PCC 9029	KS-AT-ACP- (AAT_I)
53	AAO39789	unidentified microorganism	KS
54	YP_003755474	*Hyphomicrobium denitrificans*	KS-AT-ACP-.(AAT_I)
55	JN224481	KKmKSI_017	KS
56	AAX44111	*Leptolyngbya* sp.	KS
57	AAX44114	*Leptolyngbya* sp	KS
58	AAF15892	*Nostoc* sp.	KS-AT-ACP
59	YP_001865726	*Nostoc punctiforme* ATCC 29133	KS-AT-ACP
60	JN22446	KKDKSI_027	KS
61	AAM33468	*Aphanizomenon ovalisporum*	A-A-T-KS-AT-DH
62	YP_001821790	*Streptomyces griseus*	A-T-KS-AT-ACP-KR-ACP-C-A-T-C
63	ZP_08233860	*Streptomyces* cf.	A-T-KS-AT-KR-ACP-C-A-T-C
64	ZP_04698103	*Streptomyces roseosporus*	A-T-KS-
65	ZP_04713303	*Streptomyces roseosporus*	A-T-KS-AT-KR-ACP-C-A-T-C-
66	YP_001544779	*Herpetosiphon aurantiacus*	A-T-KS-AT-KR-ACP-C-A-T-C
67	JN224483	KKmKSI_021	KS

shaded in grey is NRPS domains; (NRPS) indicates the present of NRPS module on the same gene cluster; If there are more than one module on the PKS enzyme protein; underline indicates the module of the KS used on the tree; (trans-AT) indicates lacking AT.

APPENDIX B

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
    1 agagtttgat catggctcag attgaacgct ggcggcaggc ctaacacatg caagtcgagc
   61 ggaaacgaag aggagcttgc tcctttggcg tcgagcggcg gacgggtgag taacgcgtgg
  121 gaatgtgccc agtagcgggg gatagcccgg ggaaacccgg attaataccg cataagcccc
  181 tcgggggaaa gcaggggatc ttcggacctt gtgctactgg atcagcccgc gtcagattag
  241 ctagttggtg gggtaaaggc ctaccaaggc gacgatctgt agctggtctg agaggatgat
  301 cagccacact ggaactgaga cacggtccag actcctacgg gaggcagcag tggggaatat
  361 tgggcaatgg gcgcaagcct gacccagcca tgccgcgtgt gtgaagaagg ccctagggtt
  421 gtaaagcact ttaagcaggg aggaaggcta taaggttaat acccttgtag attgacgtta
  481 cctgcagaat aagcaccggc taaatccgtg ccagcagccg cggtaatacg gatggtgcaa
  541 gcgttaatcg gaattactgg gcgtaaagcg cgcgtaggtg gttcgttaag ttggatgtga
  601 aagccccggg cttaacctgg gaactgcatc cgatactggc gaactagagt ataggagagg
  661 gaggtagaat ttccggtgta gcggtgaaat gcgtagagat cggaaggaat accagtggcg
  721 aaggcggcct cctggcctaa tactgacact gaggtgcgaa agcgtgggga gcaaacagga
  781 ttagataccc tggtagtcca cgccgtaaac gatgtctgct agccgttgga gtccttagag
  841 gctttagtgg cgtagttaac gcgttaagca gaccgcctgg ggagtacggc cgcaaggtta
  901 aaactcaaat gaattgacgg gggcccgcac aagcggtgga gcatgtggtt taattcgatg
  961 caacgcgaag aaccttacca gggcttgaca tccagcgaac tttccagaga tggattggtg
 1021 ccttcgggaa cgctgtgaca ggtgctgcat ggctgtcgtc agctcgtgtc gtgagatgtt
 1081 gggttaagtc ccgcaacgag cgcaaccctt gtccttagtt gccagcacat tatggtggga
 1141 actctaagga gactgccggt gacaaaccgg aggaaggtgg ggacgacgtc aagtcatcat
 1201 ggcccttacg tcctgggcta cacacgtgct acaatggtcg gtacagaggg ctgcgaaccc
 1261 gcgagggtaa gccaatctct taaagccgat cgtagtccgg attggagtct gcaactcgac
 1321 tccatgaagt cggaatcgct agtaatcgcg aatcagaatg tcgcggtgaa tacgttcccg
 1381 ggccttgtac acaccgcccg tcacaccatg ggagttgact gcaccagaag taggtagcct
 1441 aaccgcaagg agggcgctta ccacggtgtg ctcaatgact ggggtgaagt cgtaacaagg
 1501 tagccgta
```

Figure B1 Nucleotide sequence of 16S rDNA from KK_16S_018. This sequence contained 1,508 base pairs with 374 a, 347 c, 485 g, and 302 t

```
                10         20         30         40         50         60
        ....|....| ....|....| ....|....| ....|....|....|....|....|....|
      1 agagtttgat cctggctcag aatgaacgct ggcggcgtgc ttaacacatg caagtcgagc
     61 gagaacgtcg acttcggtcg actagtaaag cggcgaacgg gtgagtaaca cgtgggcaac
    121 ctgccccgga gtgggggata acctggggaa actcgggcta ataccgcata cgatctcagc
    181 actgcggtgc tgcgatgaaa gctggccaat tcatgaaagc tggcgctcct ggaggggccc
    241 gcgcccgatt agcttgttgg tgaggtaacg gctcaccaag gcgacgatcg gtagccggcc
    301 tgagagggtg atcggccaca ctgggactga gacacggccc agactcctac gggaggcagc
    361 agtggggaat attggacaat gggcgcaagc ctgatccagc aacgccgcgt ggaggacgaa
    421 ggccttcggg tcgtaaactc ctgtcagttg ggacgaaaag tcgtcgatca atagtcggcg
    481 attatgacgg taccagcaga ggaagccccg gctaactccg tgccagcagc cgcggtaata
    541 cggaggggggc tagcgttatt cggaattatt gggcgtaaag ggcgcgtagg cggcttggta
    601 ggtcaaaggt gaaatccctc agctcaactg aggaactgcc tttgaaacca cctcgcttga
    661 ggctgggagg gggtagcgga attcccggtg tagcggtgaa atgcgtagat atcgggagga
    721 acaccggtgg cgaaggcggc tacctggacc agttctgacg ctgaggcgcg aaagtgtggg
    781 gagcaaacag gattagatac cctggtagtc cacactgtaa acgatgggca ctcggtgccg
    841 tgggtgttga cccctgcggt gccttagcta acgcgttaag tgccccgcct ggggagtacg
    901 gtcgcaaggc tgaaactcaa aggaattgac gggggcccgc acaagcggtg gagcatgtgg
    961 tttaattcga tgcaacgcga agaaccttac ctgggcttga actgtggagg accgttctgg
   1021 aaacaggacc ttctcttcgg agacccccat agaggtgctg catggctgtc gtcagctcgt
   1081 gtcgtgagat gttgggttaa gtcccgcaac gagcgcaacc cctatcccta gttgccagcg
   1141 agtgatgtcg ggaactctag ggagactgcc ccggaaatcg gggaggaagg tggggatgac
   1201 gtcaagtcct catggccttt atgtccaggg ctacacacgt gctacaatgg gcggcacaga
   1261 gggccgcgac agcgcgagct cgagctaatc ccaaaaaacc gtcctcagtt cggattgcag
   1321 tctgcaactc gactgcatga agttggaatc gctagtaatc ccggatcagc atgccggggt
   1381 gaatacgttc ccgggccttg tacacaccgc ccgtcacatc acgaaagtcg gctgtaccgg
   1441 aagtcggtgg gctaaccttc gggaggcaac tgcccatggt atggtcggcg attggggtga
   1501 agtcgtaaca aggtagccgt a
```

Figure B2 Nucleotide sequence of 16S rDNA from KK_16S_019 bp. This sequence contained 1,521 base pairs with 350 a, 380 c, 503 g, and 288 t

```
                10         20         30         40         50         60
        ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
      1 agagtttgat catggctcag aacgaacgct ggcggcatgc ctaacacatg caagtcgaac
     61 gaacccttcg gggtgagtgg cgcacgggtg cgtaacgcgt gggaacctgc ccttaggttt
    121 ggaataactc agagaaattt gagctaatac caaataatgt cttcggacca aagatttatc
    181 gcctttggat gggcccgcgt aggattagct tgttggtgag gtaaaggctc accaaggcga
    241 cgatccttag ctggtcttag aggatgatca gccacactgg gactgagaca cggcccagac
    301 tcctacggga ggcagcagtg gggaatattg gacaatgggc gaaagcctga tccagcaatg
    361 ccgcgtgagt gatgaaggcc ttagggttgt aaagctcttt tacccgggat gataatgaca
    421 gtaccgggag aataagcccc ggctaactcc gtgccagcag ccgcggtaat acggaggggg
    481 ctagcgttgt tcggaaatac tgggcgtaaa gcgcacgtag gcggcgccgt aagtcagggg
    541 tgaaatcccg gggctcaacc ccggaactgc ccttgaaact gcagtgctag aatcttggag
    601 aggcgagtgg aattccgagt gtagaggtga aattcgtaga tattcggaag aacaccagtg
    661 gcgaaggcga ctcgctggac aagtattgac gctgaggtgc gaaagcgtgg ggagcaaaca
    721 ggattagata ccctggtagt ccacgccgta aacgatgata actagctgtc cgggttcaca
    781 gaacttgggt ggcgcagcta acgcattaag ttatccgcct ggggagtacg gtcgcaagat
    841 taaaactcaa aggaattgac gggggcctgc acaagcggtg gagcatgtgg tttaattcga
    901 agcaacgcgc agaaccttac cagcgtttga catcctgatc gcgaatagca gagatgcttt
    961 tcttcagttc ggctggatca gtgacaggtg ctgcatggct gtcgtcagct cgtgtcgtga
   1021 gatgttgggt taagtcccgc aacgagcgca accctcgtcc ttagttgcca tcatttagtt
   1081 gggaactcta aggaaaccgc cggtgataag ccggaggaag gtggggatga cgtcaagtcc
   1141 tcatggccct tacacgctgg gctacacacg tgctacaatg gcaactacag tgggcagcga
   1201 actcgcgagg gtgagctaat ctccaaaagt tgtctcagtt cggattgttc tctgcaactc
   1261 gagagcatga aggcggaatc gctagtaatc gcggatcagc atgccgcggt gaatacgttc
   1321 ccaggccttg tacacaccgc ccgtcacgcc atgggagttg gtttcacccg aagatggtgc
   1381 gctaaccttt taggaggcag ccagccacgg tgggatcagc gactggggtg aagtcgtaac
   1441 aaggtaaccg ta
```

Figure B3 Nucleotide sequence of 16S rDNA from KK_16S_023 bp. This sequence contained 1,452 base pairs with 364 a, 337 c, 447 g, and 304 t

```
                  10         20         30         40         50         60
          ....|....| ....|....| ....|....| ....|....|....|....|....|....|
         1 agagtttgat catggctcag gacgaacgct agcggcgcgc ttaatacatg caagtcgaac
        61 gccaacggtt gacttcggtt ggcctagtag cgtggcgaac gggtgagtaa cgcgtaggca
       121 acctgccctt gagcggggaa taagcccggg aaaccgggta caataccgca tgttgttccg
       181 agatcgcatg gttttggaat taaagccttc gggcgctcag ggatgggcct gcgttctatt
       241 agcttgttgg tgaggtaacg gctcaccaag gcaacgatgg atagttggtc tgagaggacg
       301 atcagccaca ctgggactga gatacggccc agactccttc gggaggcagc agtaaggaat
       361 attgcgcaat ggacgaaagt ctgacgcagc gacgccgcgt gtgcgactaa gcccttcggg
       421 gtgtaaagca ctgtccagtg ggaagaatgc ccattacggc gggagagacg gtaccactgt
       481 agaaagctcc ggctaactcc gtgccagcag ccgcggtgat acggggggag caagcgttgt
       541 ccggatttac tgggcgtaaa gggcgtgtag gcgggacggt cagtcgtggg tgaaatcttc
       601 aggcttaacc tggaaattgc ccccgatact gccgttcttg agtgcaggag agggtagtgg
       661 aattcccggt gtagcggtgg aatgcgcaga tatcgggagg aacaccagtg gcgaaggcgg
       721 ctgcctggcc tgacactgac gctaaggcgc gaaagcgtgg ggagcaaaca ggattagata
       781 ccctggtagt ctacgctgta aacgatgggt actaggtgtt ggaggaatcg acccctccgg
       841 tgccgcagtt aacgcattaa gtaccccgcc tggggagtac gatcgcaagg ttgaaactca
       901 aaggaattga cggggccccg cacaagcggt ggagcatgtt gtttaattcg atgcaacgcg
       961 aagaacctta cctggcctgg aagcacaact gctcatccgg tgaaagccgg actccttcga
      1021 gggtgttgtg gaggtgctgc atggctgtcg tcagctcgtg tcgtgagatg ttgggttaag
      1081 tcccgcaacg agcgcaaccc ctatcgttag ttgccatcag atcttcggat gctgggaact
      1141 ctaacgagac tgcccgggtt aaccgggagg aaggtgggga tgatgtcaag tcctcatggc
      1201 ccttacggcc agggctacaa acgtgctaca atgggtggta cagagggaag cgataccgcg
      1261 aggtggagcc aatcccaaaa aaccatcccc agttcggatt gcagtctgca actcgactgc
      1321 atgaagttgg aatcgctagt aatcgctgat cagcaggcag cggtgaatac gttcccgggg
      1381 cttgtacaca ccgcccgtca caccatggga gccggtagca cccgaagtcg ctgtgctaac
      1441 cgcaaggagg caggcgccga aggtgagatc ggtgactggg gtgaagtcgt aacaaggtaa
      1501 ccgta
//
```

Figure B4 Nucleotide sequence of 16S rDNA from KK_16S_034 bp. This sequence contained 1,505 base pairs with 356 a, 361 c, 492 g, and 296 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
    1 agagtttgat catggctcag aatcaacgct ggcggcgtgc ctaacacatg caagtcgagc
   61 gagaaagggg gcttcggccc ttgagtacag cggcggacgg gtgagtaacg cgtgggtaat
  121 ctgcccttga gtggggaata acactgggaa actggtgcta ataccgcatg acgtcttcgg
  181 atcctcgggt tcggagatca aagctgggga ccgtaaggcc tagcgctcgg ggaggagccc
  241 gcgtccgatt agcttgttgg tgaggtaatg gctcaccaag gctccgatcg gtagccggcc
  301 tgagagggcg gacggccaca ctgggactga gacacggccc agactcctac gggaggcagc
  361 agtggggaac tttgcgcaat gggggaaacc ctgacgcagc aacgccgcgt ggaggatgaa
  421 ggccttcggg tcgtaaactc ctgtcaatcg ggacgaaagc gctccgacct aatacgtcgg
  481 gacgttgact gtaccggtgg aggaagctcc ggctaactct gtgccagcag ccgcggtaat
  541 acagagggag caagcgttgt tcggaattac tgggcgtaaa gggcgcgcag gcggcctggt
  601 cagtctcgtg tgaaatccct cggctcaact gaggaattgc acgggaaact gcctggcttg
  661 agttcgggag agggaagcgg aattccgggt gtagcggtga aatgcgcaga tatccggagg
  721 aacaccagtg gcgaaggcgg cttcctggac cgtgactgac gctgaggcgc gaaagctagg
  781 ggagcaaacg ggattagata ccccggtagt cctagctgta aacgttgagt gctgggtgta
  841 gggggtattg acccccccctg tgccgaagct aacgcattaa gcactccgcc tggggagtac
  901 ggtcgcaagg ctgaaactca aaggaattga cgggggcccg cacaagcggt ggagcatgtg
  961 gttcaattcg acgcaacgcg aagaacctta ccggggtttg aactgtacgg gacagctgca
 1021 gagaggcagt tttccttcgg gacccgtata gaggtgctgc atggctgtcg tcagctcgtg
 1081 tcgtgagatg ttgggttaag tcccgcaacg agcgcaaccc ttgcctcctg ttgccagcag
 1141 gtaatgctgg gcactctgga gagactgccg gtgataaacc ggaggaaggt ggggatgacg
 1201 tcaagtcctc atggccttta tgccccgggc tacacacgtg ctacaatggc tggaacaaag
 1261 ggttgcaaaa ccgtgaggtg gagctaatcc caaaaaacca gtctcagttc ggattggagt
 1321 ctgcaactcg actccatgaa gctggaatcg ctagtaatcg cggatcagca tgccgcggtg
 1381 aatacgttcc cgggccttgt acacaccgcc cgtcacatca cgaaagctgg ctgtactaga
 1441 agtcgctgag ctgacccttc gggggggcag gcgcccacgg tatggttagt gattggggtg
 1501 aagtcgtaac aaggtagccg ta
```

Figure B5 Nucleotide sequence of 16S rDNA from KK_16S_036 bp. This sequence contained 1,522 base pairs with 347 a, 374 c, 511 g, and 290 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
     1 agagtttgat catggctcag aatcaacgct ggcggcgtgc ctaacacatg caagtcgaac
    61 gggaaaagtc cttcgggact gagtagagtg gcgaacgggt gagtaacgcg tgggtgatct
   121 accctaaaga gggggataac ccgccgaaag gcgggctaat accgcatgag cctgtggggt
   181 cggacctaca gggaaaggcc ttcgggtcgt tttaggagga gcccgcgtcg gattagctag
   241 taggtgaggt aagggctcac ctaggcgacg atccgtagcc ggtctgagag gacggacggc
   301 cacactggga ctgagacacg gcccagactc ctacgggagg cagcagtggg gaattttggg
   361 caatgggcgc aagcctgacc cagcaacgcc gcgtggagga agaagttttt cggaacgtaa
   421 actcctgtcc tgagggacga agccagtgac ggtacctcgg gaggaagccc cggctaactc
   481 cgtgccagca gccgcggtaa tacgggggcg gcaagcgttg ttcggaatta ctgggcgtaa
   541 agggcgcgta ggcggcatgg gaagtcatgg gtgaaagccc cgggctcaac tcgggaatgg
   601 cctgtgaaac cactgtgctg gagtgctgga gagggaagcg gaattcccag tgtagcggtg
   661 aaatgcgtag atattgggag gaacatcggt ggcgaaggcg gcttcctgga cagacactga
   721 cgctgaggcg cgaaagccag gggagcaaac gggattagat accccggtag tcctggctgt
   781 aaacgatgaa cacttggtgt ggggggtgtt gaaagtccct ccgtgccgaa gaaaactcat
   841 taagtgttcc gcctggggag tacggccgca aggctgaaac tcaaaggaat tgacgggggc
   901 ccgcacaagc ggtggagcat gtggtttaat tcgacgcaac gcgaagaacc ttaccggggt
   961 ttgaactgtc cgagccgcct ctagagatag ggggttcctt tcggggactc gggcagaggt
  1021 gctgcatggc tgtcgtcagc tcgtgtcgtg agatgttggg ttaagtcccg caacgagcgc
  1081 aacccttgtc tgtagttgcc accaggtgaa gctgggcact ctacggagac tgtcggcgaa
  1141 aagccggagg aaggtgggga tgacgtcaag tccccatggc ctttatgtcc cgggctacac
  1201 acgtgctaca atgggcggaa cagagggcag cgacctcgca agagcaagcg aatcccagaa
  1261 atccgctctc agttcggatt gtagtctgca actcgactac atgaaggtgg aatcgctagt
  1321 aatcgcggat cagaacgccg cggtgaatac gttcccgggc cttgtacaca ccgcccgtca
  1381 catcacgaaa gtgggttgta ctagaagtcg ccaggctaac ccttcgggga ggcaggtgcc
  1441 gacggtatga tccatgattg gggtgaagtc gtaacaaggt aaccgta
```

Figure B6 Nucleotide sequence of 16S rDNA from K_16S_037 bp. This sequence contained 1,487 base pairs with 352 a, 354 c, 509 g, and 272 t

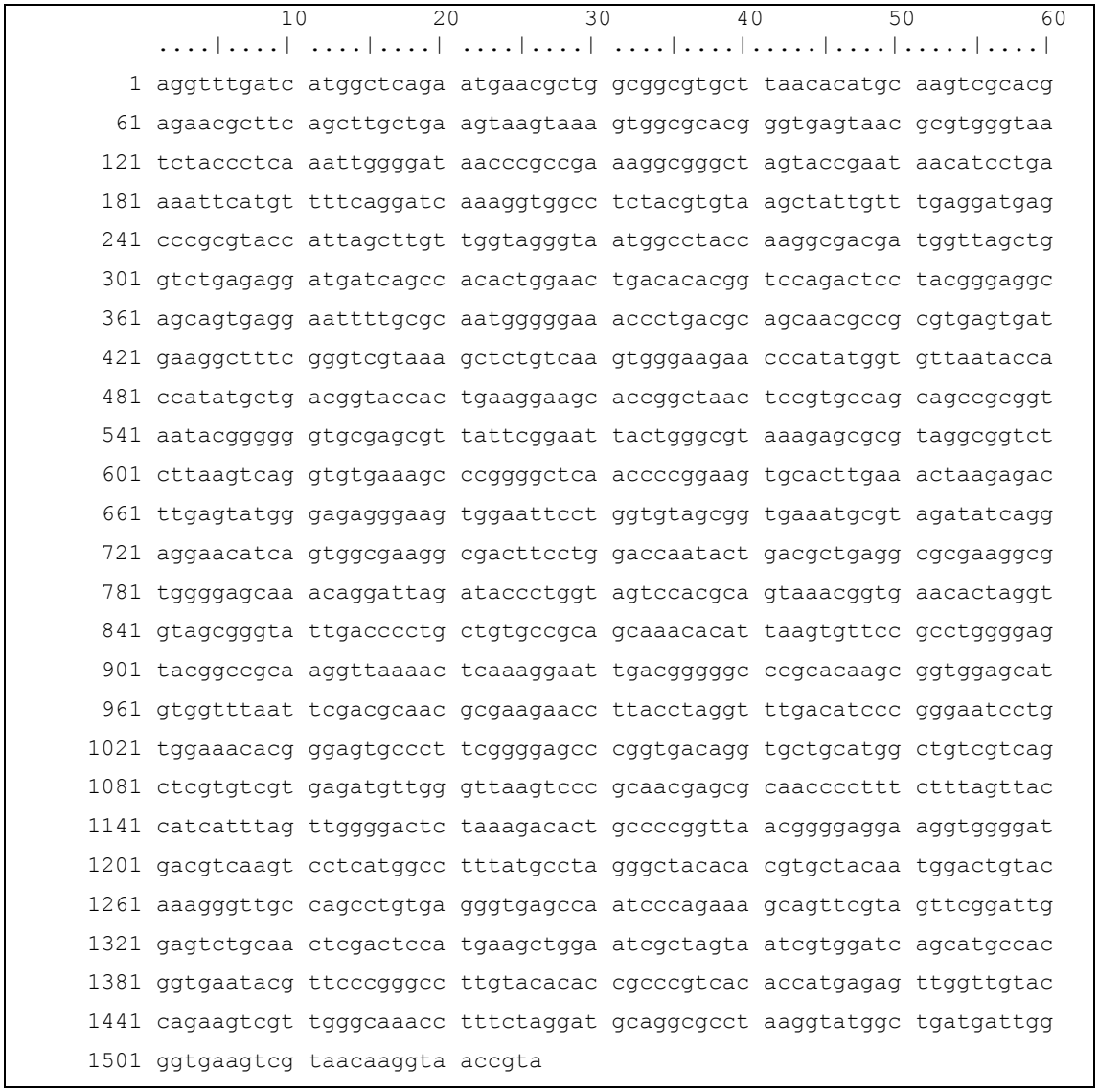


Figure B7 Nucleotide sequence of 16S rDNA from KK_16S_038 bp. This sequence contained 1,526 base pairs with 384 a, 346 c, 472 g, and 324 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
     1 agagtttgat cctggctcag aacgaacgct ggcggcgtgc ttaacacatg caagtcgaac
    61 gagaaagttt ccttcgggaa acgagtagag tggcgcacgg gcgagtaacg cgtagataat
   121 ctacccttat atctgggata acattgggaa actggtgcta ataccggata cacttcttat
   181 ctgcggggat aaggaggaaa ggtggcctct tcttgaaagc taccgtatag ggatgagtct
   241 gcgtaccatt agctagtagg tggggtaacg gcccacctag gcgacgatgg ttagcgggtc
   301 tgagaggatg atccgccaca ctggaactgg aacacggacc agactcctac gggaggcagc
   361 agtgaggaat attgcgcaat gggggaaacc ctgacgcagc gacgccgcgt gagtgaggaa
   421 ggtcttcgga tcgtaaagct ctgtcagaag ggaagaagtg ttatatagct aatacctatg
   481 taatttgacg gtaccttcag aggaagcacc ggctaactcc gtgccagcag ccgcggtaat
   541 acggagggtg cgagcgttgt tcggaattac tgggcgtaaa gggcgcgtag gcggccggat
   601 atgtcagatg tgaaagtcca cggctcaacc gtggaagtgc atttgaaact gtccggcttg
   661 agtatcggag gggagtgtgg aattcccggt gtagaggtga aattcgtaga gatcgggagg
   721 aacaccggtg gcgaaggcga cactctggac gaatactgac gctgaggcgc gaaagcgtgg
   781 ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatgtga actagaggta
   841 ggtggtgttg atcccatctg tctcgcagct aacgcattaa gttcaccgcc tggggagtac
   901 ggtcgcaaga ttaaaactca aaggaattga cgggggcccg cacaagcggt ggagtatgtg
   961 gtttaattcg acgcaacgcg aagaacctta cctgggcttg acatcccgag aatttcctgg
  1021 aaacaggaaa gtgcgtcatt agatgaactc ggtgacaggt gctgcatggc tgtcgtcagc
  1081 tcgtgtcgtg agatgttggg ttaagtcccg caacgagcgc aacccctatc tttagttgcc
  1141 agcagttcgg ctgggcactc tagagagact gccggtgtca aaccggagga aggtggggat
  1201 gacgtcaagt cctcatggcc tttatgtcca gggctacaca cgtactacaa tggtcggtac
  1261 aaagggctgc aagcacgcga gtgtaagcca atcccagaaa gccggtctca gtccggattg
  1321 gagtctgcaa ctcgactcca tgaagttgga atcgctagta atcgcggatc agcatgccgc
  1381 ggtgaatacg ttcccgggcc ttgtacacac cgcccgtcac accacgaaag tcggttgttc
  1441 cagaagtagt tgagctaacc ttcgggaggc aggctaccaa ggaatgatcg gtaattgggg
  1501 tgaagtcgta acaaggtaac cgta
```

Figure B8 Nucleotide sequence of 16S rDNA from KK_16S_039 bp. This sequence contained 1,524 base pairs with 391 a, 340 c, 476 g, and 317 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
     1 agagtttgat catggctcag attgaacgct ggcggcatgc ttaacacatg caagtcgaac
    61 ggtaacaggc cttcgggcgc tgacgagtgg cggacgggtg agtaacgcgt gggaatctgc
   121 ccttgagtgg gggacaacat tcggagacga atgctaatac cgcatgatgt ctacggacca
   181 aagttgcctt cgggtgacgc ttgaggatga gcccgcgttg gattagctag taggtggggt
   241 aaaggctcac ctaggcgacg atccatagct ggtctgagag gatgatcagc cacactggga
   301 ctgagacacg gcccagactc ctacgggagg cagcagtggg gaatattgga caatgggcgc
   361 aagcctgatc cagcaatgcc gcgtgtgtga agaaggcctg cgggttgtaa agcacttttc
   421 tgagagaaga aagcagccag gttaagagct aggctgtgtg acggtatctc aggaataagc
   481 accggctaac tccgtgccag cagccgcggt aatacggagg gtgcaagcgt taatcggaat
   541 tactgggcgt aaagggtacg taggcggcta tttaagtcgg atgtgaaagc cccgggctta
   601 acctgggaat gtcgttcgat actggatagc tagagtctgg cagaggctgg tggaattccc
   661 ggtgtagcgg tgaaatgcgt agatatcggg aggaacatta gtggcgaagg cggccagctg
   721 ggtcaagact gacgctgagg tacgaaagcg tggggagcaa acaggattag ataccctggt
   781 agtccacgcc ctaaacgatg aacactagac gttggttcca tttaagggat tagtgtcgaa
   841 gcaaacgcgt taagtgttcc gcctggggag tacggccgca aggttgaaac tcaaaggaat
   901 tgacgggggc ccgcacaagc ggtggagcat gtggtttaat tcgatgcaac gcgaagaacc
   961 ttaccagccc ttgacatcct cggaatcttt cagagatgaa agagtgcctt cgggaatcga
  1021 gtgacaggtg ctgcatggct gtcgtcagct cgtgtcgtga gatgtcgggt taagtcccgc
  1081 aacgagcgca acccttgtcc ctagttgcca gcacttcggg tgggaactct agggagactg
  1141 ccggtgataa accggaggaa ggtggggatg acgtcaagtc atcatggccc ttatgggctg
  1201 ggctacacac gtgctacaat ggtcggtaca gagggccgca aagccgcgag gtggagcaaa
  1261 tcccagaaag ccgatcttag tccggatcgc agtctgcaac tcgactgcgt gaagtcggaa
  1321 tcgctagtaa tcgcgaatca gcaatgtcgc ggtgaatacg ttcccgggcc ttgtacacac
  1381 cgcccgtcac accatgggag ttggctgcac cagaagccgg tagctcaaca tcgagcgctg
  1441 tccgcggtgt ggtcaatgac tggggtgaag tcgtaacaag gtagccgta
```

Figure B9 Nucleotide sequence of 16S rDNA from KK_16S_040 bp. This sequence contained 1,489 base pairs with 367 a, 338 c, 485 g, and 299 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
    1 agagtttgat cctggctcag agtgaacgct ggcggcgtgg ttaagacatg caagtcgaac
   61 gggattgtta aagttgacgc ttcggttgat tctttaacat gagagtggca aacgggtgcg
  121 taacacgtga gcaacctgcc ctaaagcctg ggatagctcg gagaaattcg aattaatacc
  181 ggatgtggca cgaaaacaca tgtttttagt gctaaagctt gtaacggcac tttaggaggg
  241 gctcgcggcc tatcagcttg ttggtgaggt aacggctcac caaggctaag acgggtagct
  301 ggtctgagag gatgatcagc cacactggaa ctgagacacg gtccagacac ctacgggtgg
  361 cagcagtttc gaatcattca caatgggggа aaccctgatg gtgcaacgcc gcgtgaggga
  421 tgaaggcctt cgggtcgtaa acctctgtca ccaaggagca acaagcaggt tcatagcctg
  481 ccctgagtta acttggagag gaagcagtgg ctaactccgt gccagcagcc gcggtaatac
  541 ggagactgca agcgttactc ggattcactg ggcgtaaagg gtgcgtaggc cgctaagcgt
  601 gtcaggtgtg aaatctcggg gcttaacctc gaaactgcgc ctgaaactgt ttagcttgag
  661 tattggagag gtaagcggaa tttctggtgt agcggtgaaa tgcgtagata tcagaaggaa
  721 caccaatggc gaaggcagct tactggacaa taactgacgc tgaggcacga aagcgtgggt
  781 agcgaaaggg attagatacc cctgtagtcc acgccgtaaa cgttgtacac taggtcttgg
  841 gggtttcgac cccttcagga ccccagctaa cgcgataagt gtaccgcctg aggactacgg
  901 ccgcaaggct agaactcaaa ggaattgacg ggggcccgca caagcggtgg agcatgtggt
  961 ttaattcgat gcaacgcgaa gaaccttacc taggcttgac atgtatcgga cggtttccag
 1021 agatgggttc tttccttcgg gactgataca caggtgctgc atggccgtcg tcagctcgtg
 1081 tcgtgagatg tttggttaag tccagcaacg agcgcaaccc tcgtccttag ttgccagcac
 1141 gttatggtgg ggactctaag gagacaaact tctttcagaa gtgggaaggt ggggatgacg
 1201 tcaggtcagt atggccctta cgcctagggc tacacacgtg ctacaatgcc cggtacagtg
 1261 ggacgcaata ccgcgaggtg gagcaaatcc tcaaaaccgg gcccagttcg gattggagtc
 1321 tgcaactcga ctccatgaag tcggaatcgc tagtaatgac gtatcagcta tgacgtcgtg
 1381 aatacgttcc cgggccttgt acacaccgcc cgtcacatca tgaaagccgg ttttgcccga
 1441 agtacgtgcg ctaacttcgg aagcagcgtc ctaaggcagg gccggtgatt gggatgaagt
 1501 cgtaacaagg taaccgta
```

Figure B10 Nucleotide sequence of 16S rDNA from KK_16S_041 bp. This sequence contained 1,518 base pairs with 383 a, 350 c, 458 g, 327 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
    1 agagtttgat cctggctcag gatgaacgct agcggcgtgc ctaatgcatg caagtcgaac
   61 gggaaccttt ggaattattc tgaaggggag agtggcggac gggtgagtaa cacgttggtg
  121 acctgcccca gagaggggga taaccattgg aaacgatggc taataccccg gatgtcatta
  181 tggttagaga gtaatgacta aagctccggc gctctgggag gggcctgcgg cccatcagct
  241 tgttggtagg gtaacggcct accaaggcga agacgggtag ggggcgtgag agcgtgaccc
  301 cccacactgg tactgaaaca cggaccagat acctacgggt agcagcagca aggaatattg
  361 cccaatggac gaaagtctga ggcagcaacg ccgcgtggag gatgaaggcc ttcgggttgt
  421 aaactccttt tctgggggaa gagaaaggac ggtaccccag gaataagtct cggctaacta
  481 cgtgccagca gccgcggtaa aacgtaggag gcgagcgtta tccggattta ctgggcgtaa
  541 agcgcgtgca ggcggttttc taagtcggac gtgaaagctc ccggctcaac tgggagaggt
  601 cgttcgaaac taggaggctt gagggtggta gaggagagtg gaattcccgg tgtagtggtg
  661 aaatgcgtag atatcgggag gaacaccagt ggcgaaggcg gctctctggg ccattcctga
  721 cgctaagacg cgaaagctag gggagagaac gggattagaa accccggtat tcctagccgt
  781 aaacgatgtc aactaggtgt gggggggtatc caaccccccct gtgctgcagc aaacgcgata
  841 agttgaccgc ctggggacta cggccgcaag gctaaaactc aaaggaattg acggggaccc
  901 gcacaagcag cggagcgtgt ggtttaattc gaggctacgc gaagaacctt accagggttt
  961 gacatgacgg tggtaggaag gcgaaagccg accgacccctt cggggagccg tcacaggtgc
 1021 tgcatggctg tcgtcagctc gtgccgtgag gtgttaggtt aagtcctgta acgagcgcaa
 1081 ccctcatcgt tagttatacg tgtctaacga gactgccggt gataaaccgg aggaaggtga
 1141 ggatgacgtc aagtcagcat ggcctttata tcctgggcta cacacacgct acaatggccg
 1201 gtacaatggg tagcgaaacc gcgaggtgaa gccaatcctc caaagccggt ctcagttcgg
 1261 attgcaggct gcaactcgcc tgcatgaagt cggagttgct agtaaccgcg cgtcagcaac
 1321 agtgcggtga atacgttccc gggtcttgta cacaccgccc gtcacgtcat ggaagttggc
 1381 aacgcctgaa gtcagtgagc taaccttcac gggaggcaac tgcccaaggt ggggtcggta
 1441 actgggacga agtcgtaaca aggtagccgt a
```

Figure B11 Nucleotide sequence of 16S rDNA from KK_16S_042 bp. This sequence contained 1,471 base pairs with 367 a, 344 c, 480 g, and 280 t

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|.....|....|.....|....|
     1 agagtttgat cctggctcag gactagcgct ggcggtgtgt cttagacatg caagtcgagc
    61 gggattcact ccggtagcaa taccggggtg cgcctagcgg cgaacgggtg agtaatgcaa
   121 gggtaaccta ccctgaggtg aggcacaacc cgtcgaaagg cgggctaatt cctcatgtgt
   181 tgcagcctcg gcatcctggc tgcagcaaag acgccgcaag gtgtcgcctc ttgaggggct
   241 cttgttctat cagcttgatg gcggggtaac ggcccaccat ggctgcgacg gatagtcggc
   301 gtgagagcgt gaccggccac aggggaactg agatacggtc cccactccta cgggaggcag
   361 cagtctagaa atttgggcaa tgggcgaaag cctgacccag cgacaccgcg tggaggatga
   421 agtctttcgg gatgtaaact cctgtcaggg tgaacgatgc cgttccggag taactgccgg
   481 ggcggtgacg gtactcccaa aggaagctcc ggctaactac gtgccagcag ccgcggtgat
   541 acgtaggggg caagcgctgt ccggattcac tgggcgtaaa gggtgtgtag gcggactggt
   601 gggtcgatgg tgaaagcttc gggcttaacc cgggaattgc tgtcgaaact accagtctag
   661 agagcaggag agggaactgg aacttccggt gtagcggtag aatgcgtaga tatcggaagg
   721 aacgctaata gcgaaggcag gttcctggaa tgcatctgac gctgagacac gaaagctagg
   781 ggagcgaaca ggattagata ccctggtagt cctagctgta aacgatgtac attaggcgtg
   841 ggggcctagt ctccgtgccg cagctaacgc gataaatgta ccgcctgggg actacggccg
   901 caaggctgaa actcaaagga attgacgggg acccgcacaa gcggtggagg atgtggttta
   961 attcgatgat acgcgaagaa ccttacctgg gtttgacatg caagtggtag ggacccgaaa
  1021 ggggatcgac cggggtttat ctctggagct tgcacaggtg ctgcatggct gtcgtcagct
  1081 cgtgccgtga ggtgtatggt taagtcccgc aacgagcgca acccctgccc ttagttgcca
  1141 accggtgacg gtgcactcta aggggactgc ctccgtcaag gaggaggaag gtggggacga
  1201 tgtcaagtca tcatggcctt tatgcccagg gctacacacg tcctacagtg gccgctacaa
  1261 tgggcagcga catcgcgagg tggagcgaat ccccaaaggt ggccatggtt cggatagcag
  1321 gctgcaactc gcctgcttga agacggaatc gctagtaatc gctgatcaga acgcagcggt
  1381 gaatacgttc ccgggtcttg tacacaccgc ccgtcacgcc atgggagttg gcaacgcccg
  1441 aagtccctcc tttgcggggg cccaaggcgg ggccgatgac tggggcgaag tcgtaacaag
  1501 gtaaccgta
```

Figure B12 Nucleotide sequence of 16S rDNA from KK_16S_043 bp. This sequence contained 1509 base pairs with 344 a, 373 c, 503 g, and 289 t

```
                10         20         30         40         50         60
        ....|....| ....|....| ....|....| ....|....|....|....|....|....|
      1 agagtttgat catggctcag aacgaacgct ggcggcgtgc ctaacacatg caagtcaaac
     61 gggaaagtcc tcttcggggg gcgagtacag tggcgaacgg gtgagtaaca cgtgggcaac
    121 ctgccctcag gattgggata acctcgcgaa agtggggcta ataccggata agaccacgct
    181 gtctacgggc agtggggtaa aaggtggcct ctccatggaa gctatcacct gaggatgggc
    241 ccgcggccta ttagcttgtt ggtgaggtaa cggctcacca aggcgacgat gggtagccgg
    301 cttgagaggg tgtacggcca cactggcact gagacacggg ccagactcct acgggaggca
    361 gcagtgggga attttgcgca atgggggaaa ccctgacgca gcaacgccgc gtgagtgatg
    421 aaggccttcg ggttgtaaag ctcttttagt ggggaagaaa caccctggga ctaataattc
    481 cagggcttga cggtacccac agaaaaagtt ccggctaact acgtgccagc agccgcggta
    541 atacgtaggg agctagcgtt gttcggaatc attgggcgta aagagcgcgt aggcggcgtg
    601 gcaagtccgt tgtgaaatcc caaagcttaa ctttggaact gcatcggata ctgccatgct
    661 agagttcggg agaggagagt ggaattccca gtgtagaggt gaaattcgta gatattggga
    721 ggaacaccgg tggcgaaggc ggctctctgg accgatactg acgctgaggc gcgaaagcta
    781 ggggagcaaa cgggattaga taccccggta gtcctagccg taaacgatgg gcactaggtg
    841 ttgggggtat cgaccccttc agtgccgcag ctaacgcatt aagtgccccg cctggggagt
    901 acggtcgcaa gactaaaact caaaggaatt gacgggggcc cgcacaagcg gtggagcatg
    961 tggtttaatt cgacgcaacg cgaagaacct tacctgggtt tgaactgcag tggaccggtg
   1021 cagagatgta cctttccttc gggactgctg cagaggtgct gcatggctgt cgtcagctcg
   1081 tgtcgtgaga tgttgggtta agtcccgcaa cgagcgcaac ccatgccctt agttaccagc
   1141 ggttcggccg gggactctaa gggaactgcc ggtgataaac cggaggaagg tgtggatgac
   1201 gtcaagttct catggccttt atgcccaggg ctacacacgt gctacaatgg tcggtacaaa
   1261 gggttgcaat accgtgaggt ggagccaatc ccaaaaagcc ggtctcagtt cggattgtag
   1321 gctgcaactc gcctgcatga aggtggaatc gctagtaatc ccggatcagc atgccggggt
   1381 gaatacgttc ccgggccttg tacacaccgc ccgtcacacc acgaaagccg gttgtaccag
   1441 aagtcgctga gccaacccgc aagggaggca ggtgccgaag gtatggccgg tgattggggt
   1501 gaagtcgtaa caaggtaacc gta
```

Figure B13 Nucleotide sequence of 16S rDNA from KK_16S_044 bp. This sequence contained 1523 base pairs with 368 a, 362 c, 496 g, and 297 t

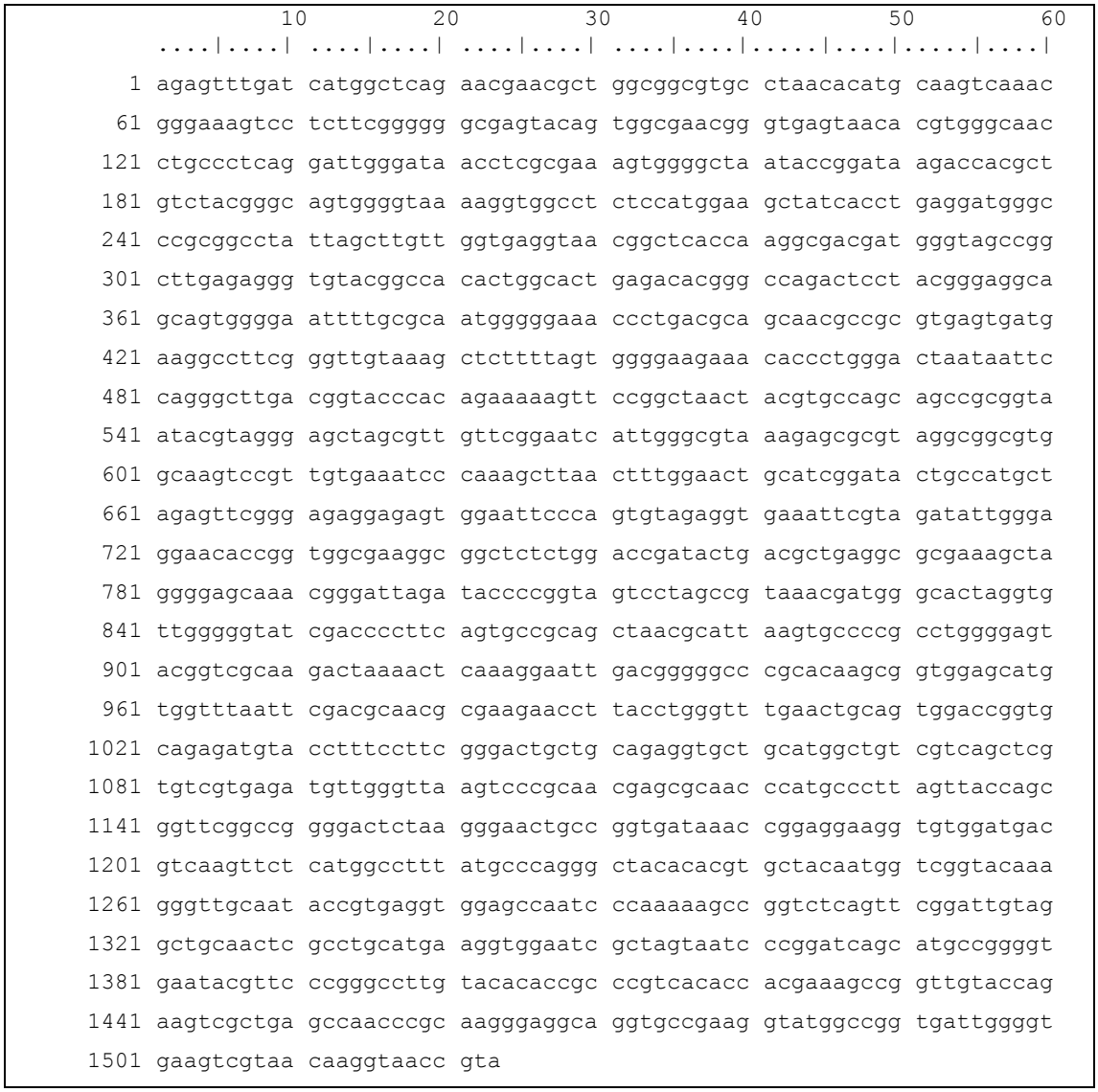


Figure B14 Nucleotide sequence of 16S rDNA from KK_16S_045 bp. This sequence contained 1,483 base pairs with 366 a, 340 c, 473 g, and 304 t.

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
     1 ggatcatttg gtctcctcgg cgctcggcca tgccatggag gacgccggcg agcgccgcga
    61 acgtcgagcc aggctgtgga agacggcccg gtattgatcg agctcctcgg ggcacatgag
   121 gcggaggagc agaacgtagc tcgcgaggaa aacgagcccg acaagcacga gctcgacgat
   181 ggcgatctcg ggcagtcggt ggctcgtgag ccggtccgtg agaagcgagc ccccgagcgc
   241 gagaagcgcc gccgcgacga ccttcgcgct gtctccaagg accctcgccg tgctcgccgt
   301 gcccaggatc gtgcgcaagt agacgaagag gatgccgagc ttgatatagg cggacacggt
   361 caccgccacg gcgatgccct ggtggccgag gctgggagag aggagcgcgc agagggcgag
   421 gttcagaacg acccgcgcga gccccgaacg gaggggcacc tggggctcct gacggctgtg
   481 gaacaccgag acgaggagct cgttgagccc gagcgcgagg atgaccggtg cgaggatgcg
   541 gagcaccgcg accgtgagcg cggtcgagcc cgcgtcgaac gagccgcgct cgaagagagc
   601 ccgaacgagg ggagtcgccg cgacgtagac gaaccccgcg agcggcgcga agaagaccac
   661 cgtcgtgacg agggctcttc tcacggactc gcgcgcctgg gcgctctctt tcgccaggaa
   721 ctgttcggaa aggaagggga acagagcgcc ggtcatcgat tttcccagga tccgaatggg
   781 cgactcgacc agcaaggtgg cgtattggag cgaggagacc gccccgacgc ccaggagcga
   841 ggcgaaggcc cggccgacga cgacgttgag cttccccgtc gccgtgccca cgtagatcgg
   901 gagcgacagg cggatcgctt tgaccaccgt ggcgtcccgg agatcgaagt cgggccggat
   961 ctggtgtcgg cggttccgga gcgcgatccc gatgatgagg agctggacgg ctccgcccac
  1021 gatctgaccc cagacgaggg cgaggacgcc ccaggagtgg aagatcagga aataagtggt
  1081 ggagacgatc tggaatacgg cctcggagct cgcaccggcg aggaaacgtt tctcggcata
  1141 gagcaccatc gcgagcagac gcgccgcgcc caccggggcc gccatccaca ggaggagccg
  1201 ggtcatggtc acggtggctt cgcgcccttc cgctccgaga ccgggtccga cgacgcgcag
  1261 cagaggctcg gtggcaatcg acgcgatcgc gatcgccgag acgaggagca gcagccaata
  1321 ggtgagcatc ttccccgcgg agcgccaggc gcctctctcg gagacgcttc gctcggaggc
  1381 gaagagcggg atgaagttga aagcacccgt ccccgcgaga tgcatggtga ggtccggaat
  1441 cacgagggcg aggaggtagg tgtccatggc gcccgacgtg ccgaagaagt aggccgcgag
  1501 cacgttcttc aggaagccga cgagcgcgac ggtcacggag gacgcgccga gcacggcgaa
  1561 ggcgcggagg atgtcgacgg cgaggcccat gtcagcgtcg cttttccagt tgggcgaaac
  1621 gggccaagtg agggctcacg cgctcgcggg aagcgaatcg atcgaacgac gtcctgcccc
  1681 cgcgtctcgt tccattgtca tcccaagctc tcgagctcga accatggcgc tccaccggca
  1741 tcgggttcgc gggcctgggg ggcacccacg ggaagtccga gtctccgtcc ggctggcaga
  1801 atacgccgtc cctagggttg ggattagggg aaagcgtaac ttttatggga atttcgattc
  1861 cctgtcaagg tcgagagagc gcgtccactc gcaaaccgtc gacaagaggt cgccttgtga
  1921 gcttcggtgg gtgaagatgt gatccgtccc gcgaatcaag tgcgtcgtca ccgcgcccgt
  1981 tcggcgaagc ggcccgaccg ggcgcttcac gccggtccgg tagaagtagt gggacgggct
  2041 gccttgcgag tagacgagaa gggtacgggt gccgcgccgt cccagcgctt cgaggctcga
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
2101 gagaatctcg ctcttcgacg gcacgaggga cagggcgtct tcctcgggtc ttgcgtcggc
2161 tccgttcgct tgcgtcttct tgccgtccct ctcggcggcc gagcgtttca acatccgaat
2221 cacgtcctgc cagatctcgc tccggcccga gagcagtcgc agccagcttg ccggccggag
2281 gagcttcttc cgatacagat ggcgcaggta ggcggtcgtc ggtacggcgt ggccctcgat
2341 gagcaccgcg cccaccacgc ggggatctcg agcggcggcg cgcagcgcgt tgtcggctcc
2401 ggaacagagc ccgacgagca cgaagcgctc gatccccctc gacgaagcga ggtaatcgag
2461 cgcctccttc acctcctcga gcgagctctg ggtaaagggc cgcccgtcgc gtcgaggctc
2521 gctgtcgccg attcccgaga agtcgaaacg catggtcacg aagcccgcct cggcgagctc
2581 gcgcgcgagc ttgacgtgca aacggttcgg cccgacgcga tggacgatcc ccgcgttgag
2641 aatgacgacg ccgccgcgcg gccggccgcc gttcgaagcg gccgggtcgg tgacgattcc
2701 caccagggtc ttccactcac cgaagaggag cgctttctcc gtcaagtcgc tctccccgcc
2761 agccaagacg cgatctcgga tacggttggc ttcgggacga cggcctggtc gattccctct
2821 tctcgaagcc agaaccgggc gctcggaatc cgggagcggc tcacctccac accccgcgcc
2881 cggagggcct ccgcagcggc ttcgacgtcc tcgtcgccgt cgccctccgc gccggtcacg
2941 atgagggccc tctgcgccgg cgctctatcg aggcgcgtca gatcgatgga cgcgattgct
3001 tccctcaagc cgcgcggcaa ggggaagccg agggcttcgt cctcctcgcc ctcgccttcg
3061 ccttcgcctt cgccttcgcc gccccccggg tgagctttcg cccaggctcg tctcagcgca
3121 tcgagctcgt cgaggtacct ggcgccgcgg acgacgggct cccagagcac gaggttcgag
3181 aagcgcaggc tttcgcgctc ggcggcgagc acggacaagg cggcgccgag cctcagcccg
3241 acggcggaga tatccgtcgc cgatgagccc tcggcgagct cctcgtgggc gagcgcgacg
3301 tcccgaaccc actggtcgag gctgccctcc tcgccgctcc ccgcggagtc gccgcagccg
3361 aagtagtcga accgaaggct ctcccgcccc tcgttcgaga gcgcttccgc gagtgcccgc
3421 agggctcggt gcgcccgaag gtactcgcgc ccgaaggggt tgcacatcag gacggctcgc
3481 ttggctcctc ccgacgcggc gaggtggtgg acaccgaaga gcgctctccc cgaggagccg
3541 aagaagaagg gtctcatggc gggtcgggat ccgcgtcctg gcgttcgctt catcgtcaca
3601 atggggcggg gggagccgat actagcacag tcgcgtgggc cgcgattcga gccggcgcaa
3661 cgtcaccggg accgagcgtc gcgacgttct cggtaaggcc acctggttgc agcgaagcgg
3721 cactgtggta gtgtagaccc gggtcgggac ccaactcctc ggctgaaaat cgaacacgac
3781 gcgacttcgg ggaaactagc cgtctcttcg tcggcgagag ctgaggcaaa tggcacgaga
3841 catcaagctc agggatgctt ggcgcgagac caagcgcatt ctctgggagc accgtcgccg
3901 gctcaccctc gccttcctcc tgacgctctt cagccggatg gccggcctcg ccgtccccgc
3961 gtcgaccaag tttctgatcg atgacgtcat cggcggcggg cggggtgagc tgctcgagac
4021 ggtcgcgatg ctcctcgtct tcgccatcgt cgctcaggcc gcgagcgcct tcgtcatgtc
4081 gcagctcatc ggcgtcgccg cgcagaggac catcaacgat cttcgcaccc ggttacagag
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
4141 ccacgtgacg cggcttccgg tgaagttctt ccagtccgag cacaccggaa gcctcatcgc
4201 ccggatcatg accgaccccg aggcggtgcg gaacctcgtg ggcacgggcc tgatccagct
4261 cgtgggggggc tccttgaccg ccgtgctcgc tctcggcgtg ctgctctatc tgaactggcg
4321 tctgacgttc ctgaccgttc tcgccctgac gctcttcgcc ggggtgatca ccttcgcgtt
4381 cgtctggatg cgtcccttgt atcgcgagcg gcgcgagatc tacgcggggg tgagcgggcg
4441 gttgtccgag tccctgagcg gcatccagac cgtgaagtgt tatcgcgccg agaagcgaga
4501 agacgaggtc ttcgttcgag gcgtggacga cctcttccgt aacgtggcga ggacgatcac
4561 cgggggcgcg gccgtcggct cggtctcgat catcatcgtg ggcggcgtgg tcgtcgcgat
4621 ttttctcgtg gggggacggg cgatcctcgc cggcaccatg accctcggcg acctcgcgat
4681 gtacgtggcc ttcaccgcgg tgatggtcgc gcccctgacg cagatcgccg gggtggccac
4741 ccaactgagc gacgccttct cggggctcga tcgcatccgc gagctcctcg acctcgagac
4801 cgagagcgac gaggacgagg gaagagcacc gctcctgtcg ctccagggcg acatcgtgtt
4861 cgagaacgtg gcgttcgagt acgagccgga aactcccgtg ctgaagggga tcgacttcga
4921 agcgcgtgcc gcgaccacca ccgcgctcgt cgggccgagc ggagccggga agagcaccgt
4981 catcgggctg gtgatggggt ttcaccgtcc caccgtgggg cgggttctcg tcgacggccg
5041 cgatctgtcg acgattcgcc tcgccgatta tcgctcctac ctcggcgtgg tgctccagga
5101 gaactttctc ttcgacggca ccgtggccga caacattcgc tacgcccgcc ccgccgcgac
5161 ggacgaggag gtccgccgcg ccgccgagat cgcgcattgc gacgagttcg tctcgcgctt
5221 caaggagggc tacgacacgg tggtgggcga gcgcggggtc aagctctcgg gcggcgagcg
5281 ccagcgggtc agcatcgccc gcgcgattct cgcgaaccct cggatactga ttctggacga
5341 agcgacctcg agcctcgaca gcgagaacga ggcgctgatt caggatgggc tgcgttcgct
5401 gcgggcgggg aggacgacgt tcgtcatcgc ccaccggctc tcgacgatcc ggtcggcgga
5461 tcagatcctc gtcgtggacg ggggccggat cgtggagcgg ggtgaccacc acacgctcct
5521 cgagaagaga ggacggtacc ggcagctcta cgaccgccag tactccgcgg agtcgaaccg
5581 ctacgtcaac ccgggcgaag agttctcggt ccggagcgag aaggccgccg aggccgagga
5641 cggcacccat gaggagcttc cgatgacgga ggcgatgccc aacgtcggcg ccgcgcttcg
5701 gtacgacgac aagtagtagg aaggtctacg agatgagcac cagcgcgaac ccgaaccgca
5761 cgacgcctct caatcgcacc ggcgggagcc cgtggatcgt acggccccag gcgagggacg
5821 gcgcgcgttt gcgcctcttg tgctttccct acgcgggtgg cggtccctcc gcctttcgga
5881 cctggcgcaa ccgcatgccc ggggacgtcg agtgctgcct cgtgcaaccg ccgggcaagg
5941 agagccgggt cctcgaagcc ccggtgcgga gcatcgacga gatcgtcgcc tccgtctacg
6001 aagccctgcg cccgctgctc gacctgcccg tcgtcttctt cgggcacagc atgggaggca
6061 aggtctcgtt cgagctcgcg cgcctcctga gacggagggg agagcggatt cccagcgcgc
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
6121 tcgtgatctc cgcgacgcgc gcgcccacgg tgcccgaccc cgacccgccc gtgagacatt
6181 tgcccgacaa ggagttcgtg gcggagatcc agagccgcta cgacgccatc ccgagtggcg
6241 tgatggagaa cgaagagctt ctggagctca tcatgccggg ccttcgcgcc gacttcgagg
6301 cgatggagac ccacgttcac caggaggagg agccgctcgc ctgtcccatc atctgcatgg
6361 gcggtgaaga ggacgaccgg gtcggcgaag cggggctgcg tgcctgggaa gcggagacca
6421 cgagcgactt cgaggccacg atttttcccg gcgaccattt ctacatccaa tcgcaagagg
6481 atcgcgtcgt gccgtttctc gccgagcggc tccagaagat tctcgagcgc tgcccttagg
6541 cgccatgccg gagcgcgcga gcccgggata ggcgagaagg atgggggagt cgtggagctg
6601 acggatcttc tcgctctcgt ccgggagcgc gacatcctgc tctgggcgga gggtgagaag
6661 ctccgctacg acgcgcccaa aggcgcgctg acccccgatc tccgcgagcg gctggccgag
6721 catcgcgccg aggtcctgga gttcctgcga gaggcgcgaa ccttcgcgga cgagcccgcg
6781 acgagcgggc ccgtggcgat cgtcgagcgc gatggaccgc tccttccgag cttcgcgcag
6841 gagcggctcc tgtttctgcg ggagctcgct cccgagagcc cgttctacaa cgtacccgtg
6901 gcgctccggc tgcgaggccg gctcgacgtc gcggcgctcg aggcgagcct cggtcggctc
6961 gtggagcggc acgaaacact ccggacgacg ttccggaacg ggaggcaggt cgtcggcgct
7021 ccacgaccgg tcgttctccc tcacgtcgac ctgctcgggc ttccggttgc ggatcgtgag
7081 cgcgaggcgc gacgtctcgc gaatgcggag gcggcgcggc tgttcgacct cgcgcgcggg
7141 ccccttcttc gcgcgcggct gtttcagctc gcggaggacg agcatttgct ctcgctcacc
7201 ctccaccacg tgatcgcgga cggttggtcc atgagtgttc tgctcagaga catggcggag
7261 ctctacgacg ccgccgcgtc gcggcggccg cccgcgctcg ccgagctcga ggcgcagtac
7321 gcggacttcg cggcctggca gcgagagacg gccaggggcg agaccctcga ccgccacgtg
7381 gcgttctggc ggcggaagct cgagaacctc gagcccctcg accttcccac cgaccggccg
7441 cggcccgcgt cgccgcgcta cgacggcgcc cgagaatggc gcgcgctccc cccggatctc
7501 gtggacgagc tccgtgcgct cgcacgctcc gaggacgtca cgctcttcac gctccttctc
7561 gccgccttct tcgtgctcct gcaccgctac accggccagg acgatctcac cgtgggctcg
7621 ccattcgcgg gtcggacccg ccccgagctc gaggagctca tcggcttttt cgtgaacagc
7681 ctggtgatgc gcgtcgacgt gtcctccgat ccctccttcc gcgcgcttct cggccgggtg
7741 gggtcgatga cgaaggacgc cctcgagcac tcggacgcgc cgttcgagaa gctcgtcgag
7801 acgatcgagc cggaccgaga gctcggtcga aacccccttct ttcaggtcgt cttcgccctt
7861 cagagcacgc catctaccgc gctcgcgatg ggcgatctcc gcctggagtg gggctgggcg
7921 gagacgggtg cgacgcgctt cgacctcgag gcgcacgtct gggacgacgg tgcgagcttg
7981 accctcgcct tcgtctacgg cagctcgatc ttcgaggccg ccacgatccg tcgcctgcag
8041 gcgcactacg agcacctgct ccggtcgatc gtccggtccc cgggagcgcg tctctccgag
8101 ctcgcctatc tcgacgacgc ggagcgagcc tcgctcctcg cctgggcggg agcctcgagc
8161 ccttatcgga acgcttcgac gattcccgag ctcttccagg agcaagcgcg gctgcgaccg
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
 8221 gaagccgtcg cgctctccca ccgctcgggc tcgatgagct acgccgagct cgaccaccgc
 8281 tcgagtcggc tggcgtgtgt gcttcgccgc cacggcgtcg gcacggagac gcgcgtcgcg
 8341 gtcgcctgcg agcgaagccc ggcgctcgtc gtcgccctcc tggggatcct caaggcgggc
 8401 ggggcctacg tacccgtcga cccggacctt cccgagggaa ggcgcctctt cgttctcgat
 8461 gacaccgagt cgcgcgtgct gctgacccgc gcggagctgg cgcctcggtt tgggggttac
 8521 gcgggcgcgc tcctcgtctt cgaagccatc gaagccgaga tcgaggcggg aagtcccgac
 8581 cttgattggg agcccatccc gggcgtggtg ggagccgaga acctcgccta cgtcacctat
 8641 acgtcgggct cgaccggccg ccccaaaggc gtcgcgatcg cccaccgcgg cgtggtccgt
 8701 ctcgtcaagg acaccgacta cgtcgcgctc ggccccgagg acgtcgtgct ccagctcgcg
 8761 cccgtctcgt tcgacgcctc gacgttcgag atctggggag gcctgctgaa cggatcgagg
 8821 ctcgtcctct atccgggcga gacccccacg ctcgaggagc tcgcccgggt tcttcgcgac
 8881 gaaggcgtca cgacgctctg gctcacgggg tcgctcttcc accagatggt cgaccgcgag
 8941 cgcgagcgag agactctcgc ctcggtcgag acgctcctcg ccggcggcga agcgctctcg
 9001 ccccctcacg tggcgaggat gctcgaggcg ctcgcgcccg gctgccgcct cgtgaacggc
 9061 tacggtccga ccgagggcac gaccttcacg tgctgccacc ccatgacgcg aaacacggag
 9121 ccgagcgagc tcttgccttc cgttcccatc ggccggccca tcgcgaacac ccgggtccat
 9181 ctgctcgatc gtcatttgga gctcgtgccc gtcggcgtcc cgggagagct ctatctcggt
 9241 ggaggcgggc tcgcgcgcgg gtacctcgcc cgccccgagc tcaccgcgga gcgcttcgtc
 9301 cccgaccccct tcgcgaccga aaggggaggc aggctctacg cgacgggaga cttctgtcgc
 9361 tatcggccgg acgggacgat cgagtttctc ggccgaaagg acgatcaagt gaaggtccgc
 9421 ggcttccgcg tggagctcgc ggagatcgag gcccgtctgc gcgaagcgcc cgggatcgcg
 9481 gacgccgtcg tgctcgcgcg gcgagacgcg accgcagaga cacgtctcgt ggcttacgtc
 9541 accgccgccg acggcgcgga gaacgacgac gacgagcatg cccacgtcga ggcctgggca
 9601 gacctcttcg accagactta cggaggcgcg ggctccgccg agccgaccgc tcccgacttc
 9661 acgggctgga acagcagcta caccggtgag cccattccgc gcaaagagat ggaagcgtgg
 9721 ctcgacgcca tcgtcgcgcg cgtcctttcg ctccgtcccg aaaaggtgct cgagatcggc
 9781 tgcggcacgg gactcctcgt ggaacgaatc gcgccgagct gcgagcgcta cctgggcacc
 9841 gacttctccg cgtccgcgat cgcgaagctc ggggccaggc tcgcgactcg aaagagcctt
 9901 tcgcacgtcg ccctcctcga gcgcgaggcc gcggacttcg aaggcatcga gagcgaatcg
 9961 ttcgatacgg tcatcctcaa ctccgtcgtg cagtactttc cctccgcgga ctacctgctc
10021 cgggtgatcg agggggcgtc gacctgtctc tcctcgggcg gccggatcgt cctgggcgac
10081 gtccgcagcc ttcccctcct cgaggcgttt cacgcgtcgg tgttgctcgc gcgttcggcg
10141 ggctccgacg cgagagaggc gctccgtcaa ggcgtcgcgg aagccgtcga ggacgaaagc
10201 gagctcgtgc tggcgccctc ctttttttctc gcgctacccg ggcgtctccc cgctgtcacg
10261 gacgtcgagg tcctgctcaa agagggacgg cacgagaacg agctctcgaa gttccgctac
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
        ....|....| ....|....| ....|....| ....|....|....|....|....|....|
10321 gacgtcatcc tgcacatcgg acgggtggag cgagcgccgg acgtcggtcg ctggctcgac
10381 cgggagcgcg gtgcgctcac gctcggtgat ctcgatgcgg cgctcgcatc gggggagagc
10441 ttcggggtgc gcgggattcc cgacccgaga ctcacgaccg aagaagcgct cctcgcctgg
10501 ctcgcgcggg aagatggaac gcacaccgtc gacgggtttc gcgccgagct ctccggtcga
10561 gacgcgaagc ctctggagct atccgcgatc gcccgcctgg ctcgcgaccg tggattccac
10621 gccgagctca cctattcttc cgaaggcgag gaccgcgtgg atgcgctctt ccggccgcgc
10681 gcagagcggg agcgcttcac ggtcttctgg tcgatggccc gcgccgggag ggacgagaga
10741 gaagcggcga accatccgct ccgcgctcga cgcgcgcgag cactcgccgc tcgcctccgc
10801 cagcacctga aagagcgcct cccggactac atgatcccgg cgagcttcct gctcctcgag
10861 agcttccccg tcggtccgac gggaaagctc gaccggcgcg cgctcccgag cccgagccgc
10921 gcccgggcgg accgcgcggg cgagtaccgg gctccgcgct cgcgcacgga gaagctcctc
10981 gtccgcatct gggccgatgt gctgcggctc gaccgggtcg ggattcacga caacttcttc
11041 gagctcggtg gagactccat cctcgggatc caggtcgcgg cccgggctcg ggacgaaggg
11101 ctcgccgtcg cggtgcgcca gctcttcgag caccagaccg tggcggagct cgccgcttcc
11161 gtcgcgcctt cgaaagcgcg cccggacgag cctccgcccg aggagggtcc ggttccgctg
11221 acgccgatcc agcgctggtt cttcgagctt cgtctctccg agcctcacca cttcaaccag
11281 gcgctggtgc tcgcgccccg tgttcccctc gaagcggatc gactgcggcg ggccgtgggg
11341 cgcctcgtcg agcaccacga cgcgctccgc ctgcggttcc accgagaagg acgacgagcg
11401 ggagggaggc gtcaggagct cgtcgccttc gacggggagg ctcccttcga agcggtggag
11461 ctcggcgagc ggagccttct cgaggtcgcg agcgaggtgc agcatcgcct cgacctggag
11521 agcggccctt tgctgcgggc cgtgctcgta cggggtgacg acgacgagca cgcacgcctc
11581 gttctcgtcg cgcaccacct cgtcgtggac gtcgtctcgt ggcgcgttct tctcgaggat
11641 ctggagcggc tcgttcgaga ccccgaggcg aggctctcga ggacgacggc tttcgccacc
11701 tggtcgcgag agctcgagcg cttcgcgcgc gggcgcgact tcgccgaaga gctcccgttc
11761 tggctcgccc aaggggaggc ctcggtcccg ccgctcccga gagacgtcgt gagcggagcc
11821 ccatcggcgg cgagcctgga tcgcgtcgtc gagaccttcg acgaagcgga aacggagtcg
11881 ctgcttcgaa gagcgccggg ccgccccgac gccgagctcg ccgagcacgt cctcacggcg
11941 ctgagcctcg ctttttcttc ctggacgaga cagaggaagc tccgggtgtc tctcgaaggg
12001 cacggccgtg aagagctctt cgagggtgca gatctctcgc ggacggtagg ctggttcacg
12061 tccctctacc cgatcgtgct ggaggcggag cgcgacgccg acccgagggc cgcgctcgaa
12121 ggcgtgaaga aggtcctgag agcggtgccc tcgccgcgcg tcggcttcgg cctcttgcgc
12181 tatctgcgag aagagccgag gctcgcgagc gtccccgctc ccgaggtgag cttcagctat
12241 ctcggtcagc tcgacgcgac gcttccgtcg agcgcgatct tcgagaccac cgaggacgat
12301 gtcgggccgc tccagagccc gaagggaacg agagcgcacc tcatcgacgt gaccgctcgc
12361 atccggcgag ggcgcctcac ggttcacttc tacttcagcc ggtcggtgca tcgtcgcgag
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
12421 acgatcgaag cgctggcggc ggctttcgcc tcgaaccttc ggaagcttct cggcgccttc
12481 ttcctggccg agtatccgga ggcccgtctc gagccttcac agtgggagga gctcgacagc
12541 cggtacggag ccgaggcgtt cgaggacgtc taccggctca gccccatgca gcgggggatg
12601 ctcttcgaga gcttctacca gcccgacgcg ggcgtctacg tcgagcagta cgcgctcgag
12661 ctctccgagc tggatccgag cgctttcctc cgatcctggc agcgggtcgt ggatcggcac
12721 cccgcgctcc gcacctcgat cgtgggcacg gcatgggaag agcctcacca gatcgtgtgg
12781 aaacgagcgg cgctcccggt cgaggatcac gacttgcgag gccggaagcc cgacgcgcag
12841 cgagcgtggc tcgagtgtta ccgtagccgt gaccgggagc gaggcatcga tctcgagtcg
12901 ccccccctca tgcgcctcgc gctcttcgcg ctgggcgctg ggcggtacga atgcctctgg
12961 acgttccacc acgtcatcct ggacgggtgg agcgttccgc tcgtgctcgg cgaggtcttt
13021 cagatctacc agagctcacg ccgcggtgag gatctcgtgc ttcccgcgcc gccgatcttt
13081 cggagcttca tcggctggct cggccgtcgg agcaccgcgg acgcggagcg ctactggcgg
13141 cagaggctcc ggggcatcac gtccgcgacg ccgctcccgc tcgcctcggc ctcgcgcgag
13201 gcgcgctcga gcacgaagga gtacggcgaa gagcgtatcg tgttgtctcg gaccgtcacg
13261 gagcaggtgc gcgcggccgc ggccggctgt cgcgtcaccg tcaatatctg cgtgctcggc
13321 gcctggtcga ttctcctgag ccgctatgcc ggcgaggaag acatcgtgct cggcacgacg
13381 gtctcggggc gacccgccga gctcgagggc gtcgagtcga tcgtgggcct cttcatcaac
13441 accctgcccc tccgggtacg cgtggacccc gccgtgaacg tggagcaatt cctgcgagag
13501 gtgcagtccc gtcagatcga gcagcgggag tacgagcaca gccctctgtg gcaggtgcag
13561 caatggagcg acgtgcccgc gggggcgagc ctgttcgaga ccttttcgt cttcgagaat
13621 taccccgccg aggccatgcg tctcatcgaa gcctcggatc tgaggggccg gtccaagccg
13681 agcgtggagc gctcgggtta tcccctcgcg ctcgcggcca gccttcggga ggagctttcg
13741 cttcggctct tctacgatcg cgaccgcttc gagccggaag cggttcgagc tctcgcggct
13801 cgtctcgagc gagcgctcga agccatttcc tccggggcct cgaaccgctc ccgaacgctg
13861 agcgggctga accttctggg agagaccgag cgccgacttc tcctggggcc ctggagcgag
13921 ggaccgatgc ccgtctcgat cgaaagcgcg acgagcgaga cggtcgtgga gctcgtcgag
13981 cgagtcgccc ggcgtgagcc ctcgagggtg gcgctcgtcg acgccgaggt caccctccgc
14041 tatggcgagc tgaacgctcg cgcgaaccag atcgcgcgcc gtctcctcga tggcggcgtg
14101 ggaccggagg ccgtcgtcgc cgtccacctc gagcgcgggt gggagctcgt cgtagccaag
14161 ctcgcgattc tgaaggcagg aggcgcttac cttccgctcg atccggaagc gccccggctc
14221 catcgggagc tcgtgctcga ggacgcgcgg ccccgggcga tcgtcacgcg cggggatcag
14281 gattggctcc tggagaggac ggtccccgtc atccgtctcg acgaggcgct cgtcccgggg
14341 gaggggggacc gggaccgtga agaggccttc gcgccggcga cgccgcgcga tctcgatagc
14401 ctcgcgtacg tcatctacac ctcgggctcc accgggcggc cgaagggagt cgaggtgacc
14461 caccgcggcc tcgcgaacct cgttcgctgg cacaacgaag cttacggaat cacccacgac
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
14521 gatcgcgcca gccaggtggc cgcgtcgacg ttcgacgcct gcgcgtggga gatctggccg
14581 cacttggccg ccggggcgag cctctacgtg atcgacgacg agacccgttc ggaccccgtg
14641 aagctctggg ccgccctcgc acggttcggc gtcaccatct cctttctccc cacaccgctc
14701 gcggaggcgg cgctcgcgga gcgcgagcgt cccgccgggc tgtcgttgcg cctgctcttg
14761 accggcggcg accagctcgg ccgtgtcgcc aaggacgagc gcttcgggtt cgtcaaccac
14821 tacggcttga ccgagaccgc ggtggtcgcg acggcgggcg aagcgggtac ggtcgcgccg
14881 ccgccgattg gccagcccat ctccggcgcc cgcgtctatc tcgtcgacgc gtcgatcgag
14941 ctcgttcctc ccggagcgaa aggggagatc gtggtggggg gcctcggcgt cgcgcgggga
15001 tacctcgatc gcccggatct caccgccgag cgcttcgtcc cggatccttt ctcttcgtca
15061 ccgggcgctc gcctctaccg aacgggggat ctcgggcgtc acttgccgga cggaagcatc
15121 gcgttcctcg gccgcgggga tcgccagatc cagctccggg gattccgggt cgagctcgcc
15181 gagatcgagc acttgctgtc ggcgcatccc ggcgtcgcgc aggcggccgt cgtcgttcgc
15241 gaggaagggg agcgcaagag aatcgtcgct tacgttcgta ccgggaatcc gggcccatcc
15301 gagagcgggc tccgcgctta tctgagcgac cggctgcctc cgcagatggt gcccgcggct
15361 gtggtcgcgc tcgacgcctt tccctcgacg ccgagcggga agatcgatcg cgagcgcctc
15421 ccggcccccg aaccgacgag cgagcgggag cgcgtggcgc cgcggagccc caccgagatc
15481 cggctcgccg cgatgtgggc ggaagttctg cgagtcccga ctcccggcgt ccacgacaac
15541 tttttcgaga tcggaggcga ctcgatcctg tcgattcagc tcgtcgcccg ttcgcggagc
15601 cagggcctct cgctcaccgc aaaggacgta ttccagcacc cgaccatcgc ggatctcgcc
15661 gcgctcgtcg atcgatccat cgaggcgccc gcctcgcccg actcgcccga cgagctcgag
15721 gccgtgggac ccataccgct cacgcccatc cagcgctggt tcttcgagca ggagctcccc
15781 gagccgcacc acttcaacca accgctgctt ctcactttga acgaacccgt ggaccccggc
15841 aggctcgagc acgcccttta cttcgtcgtc gaccatcacg aggcgtttcg tttccgctat
15901 cggcgggtcg gcggcgaatg gcgccaggag cgggaccccg caggcccgag ggcgactttc
15961 gtcgtgatcg actcacggga gctcgcaccc gaggcgcgac gagccgtcat gacccgccag
16021 gcgagcgcgc ttcaagcgag cctccgcctc gacgagggtc cgctcttccg cgccgcgctc
16081 ttcgacttcg gagcggacga gcccgcgagg ctcctcctca tcgcccatca cctcgtggtc
16141 gacgcggtct cgtggcgcat cctgcacgag gatctggagc gggccctgtc ttcgagggac
16201 gcggcgaaac gcgagccggt gagcttcccc gagacgacgc cgttctccgc gtgggcgaaa
16261 gcccttctcg atcacgccgc gtccggcgat ctctcggccg cggtcgcgtt ctatcgggcg
16321 ctcgatccgg cgcccagcgc ccgtttgccg cgcgacaacc ccaccgtcga gaacgacgtc
16381 ggcggcctgg agcagctgag agtgtccatg ggagaggagg agacgaaggc tcttttgcat
16441 cgagcacccg cggcgtacaa cgcgcgggtg aacgaaatgc tcgtcaccgc actcgtcgaa
16501 tgcatcgcgg gctacacggg ggatcgccga atgtggctca cgctcgaggg acacggcaga
16561 gaagagttgt tcccgggggt gaacgtctct cgtaccgttg gatggttcac gacgatgttt
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
16621 cccgtcctgc tcgagctgag cggcacgcgc ggaccgggtg acgctctgaa gcacgtgaag
16681 gagcaactgc gggccgttcc cgagcgcggc ttggcctatg gactcatccg gcacgcgggc
16741 gaggaccttg aagcgtcgag cctcctcgcg aggctgccgg cgcccgaggt ttccttcaac
16801 tacctcggcc ggctcgaggc ggtcggtgtc tcgggctccc tcctgtcccg ctccgacgag
16861 cccgtcggtc cgacgcgaag ccctctcggt cgaaggctgc acctcctgga catcacggcc
16921 gcggtcgtcg cgagcgagct gcagatccac tgggccttcc atcgtgggat ccaccggaac
16981 gagacggtcc agggaatggc ggagcgcttc gtggcgtcgc tgcgatcgct gatcgcgcat
17041 tgtctcgagc cgggcgcggg aggctatacc ccctccgact tctcgctctc cggcttgacc
17101 aaggcccagc tcgatcagct gatgaagcag aagtcccgaa ggccgagctg aaggctaacg
17161 tgcggtcgga atcacgaccg accgtccgat ccacgctgag gcgtcctcgt acgcgttggt
17221 gaacgcgtcc tcgagctccg gtgggcagct cgcgagatcg ttctcattca ggctggagat
17281 cgcgagcccg aggatggcgg gcaaaaacac gctcgccgcc gcgaaccagg agctctcgcg
17341 cgagcgcccg gcagatcgag taggtggaga ggagcaacaa gagaaagaaa ggaatctgcc
17401 cgaagttgag ccagtggtcg ctccccggga cgagggccgt ccacacgtag acgagcgcga
17461 cgtgccccgt cgagctctcg ccgcgggaca tcaacttctc cgatcctccg gggtcatcgg
17521 actcttcctt ttctctcttt ctggggcgtt ggcgttccag agcaaagaaa gtatcatggt
17581 cgacctcgga ggaatgccgg ctcacctccg gaaatggggg gatttccatg gaaggaatgc
17641 cactcagtcc cgagctcggc ctcgttttga cggggtgcgc ttccgacgag cctccgtcac
17701 ccgagaagac cgtcgagctc ctcgctcgcc acggcgtctt gctggtgcgg gggacttccg
17761 ccgaccagga gaagttcctc gaatggaccc ggagcgtcgc tccgggattg ccgcgcggcg
17821 tcgtcgacga caacgtgggc ctcgacttcc acggcgaggg ttattacctc ccttaccgga
17881 tcgacgtgct ctggttctac tgcatcgtgc cccccggccg gggcggtgag acgaagttcg
17941 tcgatggggt ttccgtgtgg cgcgagctcg gctcgagcac gcgggagttc ttcgagagcc
18001 acccgctgag ctacgagctg acccttccgc agaacctctg gacgcccctg ctgaagttcc
18061 tcgacgtcgg ccgccgcgac ccgggagacc ctttctatct tcccgccagt ccggaagccg
18121 ccgccggcgc gttcggtgag ctcggtcttc cctgcgtcgt ctccgcgagc ggctcgctct
18181 tcggtcgcta ccagcaccga gcaacgatcg agacgcggtt cggaggcgag gacgccttcg
18241 tgaacaccct ccttcacgcc gtcgatcccg tcttcatccc gaggcagaac taccggttga
18301 cgaccgtgat accgggtgac gtcctcgccg aagtcaaacg aacgaccgcg cggctcacga
18361 tgctcgtccg gtggcaagag cgcgacatgg tcgcggtcga caacacgcgc gtgatgcatg
18421 gtcggctccc gttcgaaggg ccgagacgga tccttgcggt gaacgggcgc ttcgtcacgg
18481 aaccgactct ctccgccgcg tcttgatcgt accgggggcg agcgccctcg tcgccaccgg
18541 tcattccgtc cagtcgtaat cattcgcccg gagaagcggg ccttcctgaa aagcgggcgc
18601 gtcaatgtgt tgcaagcgtc acgctgccct tgacacctgg gaccatccga tagaaactta
18661 cgcgaccttt ctcccaacca tcggtcggag tttcgttgct caccgcgtgg gtacgcgggt
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
18721 caggaggaga cacccctgag ctcgccgccc ccacgtcggt tgcatcgtat gcgggtatta
18781 gcgacttaga cagcctggcg ggggcctcgc tgcggaccaa gaacgcggcg agtcggtgga
18841 ggagaaacct ttggggctcg aaacaaacag gggattcgac gccccagttt ctcgcacgcc
18901 ggcgcgcggc tcggcttcgt ggctcgtgct cgtgctcgcg ctcgcgctcg cgaagcccgg
18961 ctcccggcgc ccgggacggg aaggtcggtc ttgatcgagc tgctgaaccg ctacgcccag
19021 gggctcgtgg ccgtcccggt ggtcgtggcg tgccggcgcc gagggctgtt cgacctgctg
19081 gagcctcacg aggagctatc actcgagacg atcggacgga agctctcggc caactcgggc
19141 catctcgccg ccgccgttcg gatgcttctc ggcctcggct ggttggagcg gagggatggc
19201 ggatacgtgc tcggaccgtc ggcgtccgag ctcgactcgg tgccggacga tctcgtctcg
19261 cgcttggcgt ttcccggcga agcttatctg cgaggcgagg cttccctgtc gctcgccccc
19321 tggctcgacc tctcggcatc ccgatggggc tcgtccgacc gcgtcctcgc ggaccagctg
19381 gacggcgtgt tgctggcccg cgttcttctc gcgctcggag gcatcggcga cgaagcggca
19441 tccgccgcgc tctcggagtc caggccgttg ttcgggaaga cgccggagcc ggctcgtcat
19501 gagctcagga gagtctttcg agatcgcgag tgggcttcgg gagaaggtcg ggagctccac
19561 ctgaacgaca cggggcgcta cgtgctcgag catgccgtcg tgctcgcgct cgccgcctcc
19621 tacgcgccga tgctcggccg gatcgacgag ctcttgttcg gcgacgcgac ggcgccgttc
19681 tcgcaggcgg gtgccgaaag ccacgtcgag cgagcgctca acgtcagggc cagtggtttt
19741 ctccatcgcc gctatttcga cggcgtacgc gacgtcgtcg cttcgatctt cgacacgaag
19801 cctttcgaag ctcagccccg ttacgtcgtg gataccggct gcggggacgg caccttcctg
19861 cggtcgatct acgagacggt ccgagcgcgc acggctcggg gcaaggttct cgatcagtac
19921 cctctcgtcc tcgtcggggt cgacctcaac gatgcggcgc tcgaggaggc gaagggcacg
19981 ctcgacgatc ttccgcacgt tctcttgaag ggcgacgtcg ccgaccctaa cgggatcgag
20041 gccgagctcg agcgtcgagg aatcgcggcc gacgacgtcc tgcacgtgcg ctcgtttctc
20101 gatcacgact cgcacctccc gatggtcacg gaccgcctcg cggcgcaggc gcgcgcggac
20161 gtcccgtacc gcggcgtgtc ggtgggcccg gcgggtgagc tcgttccgca gtcggtcgtg
20221 atgcaggggc tcgtcgagca tctcacccga tgggcgagtg tgctgacgcg acacgggatg
20281 atcctgctcg aggttcactc ggtcgacccc gtcaccgcgg ctcgctaccg cgagtcgacg
20341 gagagcctca gcttcgacgc ttttcaggcg ttctcgctcc agtacctcgt ggaggcgtcc
20401 gacttcctgc tcgcggtggc cgaggcgggg ctcttcgcgc ggccggcctt cgcgcgtcgc
20461 tacccgaccg cccttccttt tacacgaatc tcgctcaact ggctcgaaag gcgcccctac
20521 cgggtccgcc tggcacgcgc gtccgacatc cccgcgatga tggagatcga ggcggcggcc
20581 tggcccgagc ccctctcggc gtctcgcgag gagatcgagc gccgtctcgt caccgatcca
20641 cgtggtcaga tcgtcgtgga atcggaaggg cgtctcgtcg ccgtcctcta ttcccagcgc
20701 atcgggtccg tcgaccatct cgaggggacg acccaccggg acgtcggctc gatcgcgcgg
20761 gaggacggtc cggtccttca gcttctcggg atcttcgtgc acccctcggt gcaacacctg
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
20821 gggatcggcg accagctcct cgagctcgcc cttcaggtcg ccgaggtcac cgcggagatc
20881 cagctcgttg ccggcgtcac gcgatgcagg aactacccca agcaccggga cctctccttc
20941 gaggagtacg tagagagcat ggggagctcg agcgtcccgg tcgaccccat ccctctgatg
21001 cacacgagcc acggcgcccg gatcgtcaaa gtgctcgcgg ggcatcgacc cgaggacgtc
21061 gagaacgacg gggcgggcgt tctcgtgacc tacgacgttc gcgggggcgg tgggggcaga
21121 gccccttcga ccgagctcgg gaagcgaggt ccggggatcg acgtcgccga cgaggggtcg
21181 ctcgcggcga agatcgagaa agcggtccga aaggtgcttc gccccgagcg cgagcgcgtc
21241 ttctcgtcga gccgcgctct gatggacatg ggcttcgact cggccgacct cctcgagctt
21301 cgggtcctgc tgggtgatct ctttgctctg gagctggatc cgacgctctt cttccgatac
21361 cccaccgccg acgcgctcgc gagctacctg gcgggacgag ggaaggtcga gcccacgcct
21421 tcgccgacgc cctggctcgc gccggtcgag aggacgccga gctttgcgcg gtcggagccg
21481 caccggcgcg acgtacaacc gcgagcgacg ggctcgagcc ggatcgcgat cgtgggaatc
21541 ggctgccggt ttcccggcgc ctcgaacgtc ggtcggtact gggatctcct gcgcgaggga
21601 cgcgatgccg tcgccgaagc gcccggggat cgaccctggc tcctcgagtc ttcccacgct
21661 cgcttcggcg gttttctcga agcggtggac cggttcgacg cgtcgctctt cggcttgtct
21721 cggcgcgagg cgaaccaggt cgatcctcag cagcgtctcc ttctcgagac ctgctggacg
21781 gccctcgaga acgcgggcat cgctcccccc tcactggaag gcacgaggac gggcgtcttt
21841 ctggggatca tgtcgcacga ctacgagctc ttgcaggttc ggcgcggcca ggtcggcgcg
21901 aaggccgacc cctacttcgc ctccggcaac tctctcgccg tcgccgcagg ccgcctcgcc
21961 tacgtgttcg gctttcgggg gccggccatc tcggtggata cggcctgctc ctcgtcgctc
22021 gtcgcggttc acctcgccgc cgagagcctt tttcgcggtg agtgcgaggt cgcgatcgcg
22081 gccggcgttc agcttctgct cgctcccgag ctcacggcct cctacgccaa ggcgggtatg
22141 ttgtcgccgg acggaaggtg caagaccttc gacgcctccg cgaacggcta cgtgcggagc
22201 gaggggggtgg gagcggtcgt gctcaagcgg ctcgaggacg cgcttcgcga cggcgacgac
22261 gtgctggcgg tcgttcgcgg ctcggcgctc aaccaggacg ggagcagcaa cgggctcacg
22321 gcgccgagcg ccgccgccca ggaagaggtc atccgggaag cgctccgaag ggccgagctg
22381 agcccctccg agatctctta cgtggaagct catggcacgg ggacgccgct cggagatccg
22441 atcgagttcg acgcgctcgc cgccgtctac ggcgagtccc gggacgaaga gaacccgctg
22501 cacctgggct cggtcaagac gaacatcgga cataccgagg ccgcggccgg catcgcgggc
22561 ctcatcaaag tcgcgctcgc cttgcgtcac gccaccatcc ccgcgcacct ccacttcgag
22621 cggccgaacc cctcgatcga tctgacgagg atcccggcac gcattccccg cgagagcctg
22681 ggctggtctc tcgggaaagg cgcgaccagc cgcctggccg cggtcagcgc cttcggcttc
22741 agcggcacca acgcccacgt cgtgctcgag gagtggctcg ctccgccgcc gccggtgccg
22801 agcggtcgtc gacccctgca cctgctcacg gtctcggccg cgagcgaggc ctcgcttcac
22861 gagctcgcct cgtcgtacgc cgagcgcctt cgacgcgacg agacggactt cggcgacttc
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
22921 gccttcagcg tcaacaccgg gcgcgcccaa ttcgaagagc gctccgtgat cctcgccggc
22981 tcgagcgccg aagcggcggg aaagatcgac ggcctgtcga aggcgggggc ggtcgagccg
23041 ggtgtgcacc gagctcgcgt ctctcccgcc gcttcgccca ccatcgcgtt tctcttcacc
23101 ggccagggct cgcagttcgc ggagatggga agacagctct tcgcgagcga gcccgtgttc
23161 cgccgcaccc tcgagagatg cgacgagctc ctgacggacc ggctcgcggg cggacttctc
23221 gccgccctgt tcccttcgat ggagggcggt gtgcccgccg cccccatcga cgagaccgcc
23281 tatacgcagc cggtgctatt cgctttcgag tatgccctcg ccgagctctg gcgctcgtgg
23341 ggcgtcgagc ccgccttcgt tctcggacac agcgtgggag agtacgtcgc cgcttgcgtc
23401 gcgggcgttc tcgacctgga gggcgcgctc gatctcgtgg ccgagcgcgg gcggctcatg
23461 caggagctgc ccgagaaagg cgccatggcc gccgtcgcgg cgacggagtc cgaggtcgcg
23521 gcctggatcg gaagcttcgg agacgatctc tcggtagccg cggtcaacgc gccgtcgagc
23581 gtcgtggtct ccggccggtg cgaagctttg aacgagctcc agaaagatct cgaatcgcgc
23641 ggcgttcggg tgagacgcct cagggtctcg aacgccttcc actccgcctt gatggagccg
23701 atgctcggcg cgctcgaggc cgaggccgcg aagctcgagc tgcgagcttc gtccatccgg
23761 ctcgtctcga acctcgacgg ccggttcgtg gagcccgagt ccctgacccc cgactattgg
23821 gcccggcacg cacgccgcgc ggtgcgcttc tctgacggga tccgtacgct cgtcgcgaac
23881 gggtgcgaca ccttcctgga gatcggtccg ggaacgacgc tcctgaccct cggccggatg
23941 tcggcgggtg aggtgccggc gctttggctt ccgagcctcc gtcccggggg cacggattgg
24001 gagtcgatgc tcagctcgct cgccgagctc catctccggg gggtgaacat cgactggctc
24061 cgctacgacg cgcctcacgc gcgtcggcgt atcgaggttc ccaactaccc gttctcgcgg
24121 gagcgctact ggttcgacga ggggtcggag ggatcttccg aggagagccg cgggcgcggc
24181 tccgatgacg ccgttcccat cgagcctctg ctggccgaag tggaaagtct ctccgagccg
24241 gacctccacc gagcgcttca gcccgacggt aacggcggtg gaccatccga tctcgaacgc
24301 cgcctgacgg gcctctcgcc cgagcggctg aagctcttcc ttcgcaaact ggcgtcgcgg
24361 ccttcgatgg gcgcctcgga gcggagcgtc gttcacgatt actacgactc gctgagcggt
24421 gtctggcgtg acgagaaggc gcgcgagtac gaagcccacg agcgcttcct caccttcggc
24481 gtcttccccg cggtcgtgcc cggcttttcc tggtcgcgca ccctcaccga ccccgagcac
24541 gatcccgagc atttccgcat cgcgctccag gcccagaagg agctgcgatc gctccttttc
24601 cggaaggtcg acctcgaggg ccggcgagtg ctcgacatcg gctgcggcta cggctccgat
24661 ctcgtctccc tcgcccgggc taatcccgac ctcaccgccg tcggctacac cttgtcggac
24721 cgacaggccg agatcgcgag gaagaaagtc gcggagctcg gcctcgagac gcgcgtgagc
24781 atcgagaagg gggacagctc ccgggatccg ttccccggcg agttcgacgt caccatcggc
24841 ttcgaggtcg cgcaccacgt acgggacaaa gccgcgctct tcaccaacgt tgccgagcac
24901 ctcgtggagg gaggcaccgt tcttctggcc gatttcgtct ccaacgccgc cttcgccatc
24961 gagcatcgcg agacctcgtc gttcttcgcg cgaaaggacg agtgggtcga gctgctcgcg
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
25021 ccgcgcggcc tcgtcgtcga ggagtccgtc gacgtgagcc gggagatcgc gaattttctc
25081 gaggacgcgg agtttcagac cacgctgcgg gagatcgcgg caaagggcgg cgaggacgtc
25141 cgcaaggcgt tccgctccta cgaccagctg ggacggctcc tgcggaaaga gctcgccagc
25201 tacgtcctcc tcaccctccg gaagcgtccc gacatggcgg ccgacgatcg cgcgagggcg
25261 agccgtgagg cgctcgagca cccgcgacgc tacgcagacg tctcccccccg gggcgccgtg
25321 tacggcgtcg tatggaacga ggtcgatacg acgcgcgagc gaacgagagg aaactttgcg
25381 tcgagcacgg atctgataat cgacgacacc cgacgcctcg gctccgggct cgcccgggag
25441 ctcgagaagg aaggcgtcgc cgtttcgctc gcggcccccg gggaccggtt cgagcggctt
25501 tcgcggaatc gttggcgaat ccgactggac cgcccggaag atctgaaacg cttcgtcgcg
25561 gaggcgacga ttccgggggc atcgggcttt cgggtcgcct atctgccgtc gctcgacgct
25621 tcgacggagg acggaacgga cgagcgcgcg caagactggg aaaaggagct cgaccgcggt
25681 cccgtcggtc tcctccacct ggttcaggcc ctcgccgggg ccgacgcggg cgaagtccgc
25741 ctcctggtcg taacccgtgg ggtgcacggc gctgtcggca ccgagcccgt cgccgtcacc
25801 gagtctccgc tctccggcgt cgccgcgagc cttcgtcgcg agtaccccgc ctggagcacc
25861 atccatctcg acctcgatcc cgagcgggat cccgatgtgg agcagctgcg cgagcaattc
25921 ctgtcgggtt cggagccggc gctcgcgctc cggggcggcg cgcgactacg accgaggctg
25981 gagcgtcgaa tcgtgctccc cgtcgaaagc gactttcggt tcggcgatga gaccacgacg
26041 ctcatcacgg gcgggatggg cggcatcggt ctgcgaatcg cggcctgggc gagctcgcgc
26101 ggcgcgcggc atctcctctt gtgcggccgg agcgatccgt caccgcacgc gcgtgaggtc
26161 atcgcggagc tccgcgacgc cggcgtcgag gtcgtcgtcg agaaagcgga cgtgtcccgc
26221 gaggaagacg tcgagcgcct gatgagctcg atgacgcgcc tcgggggat gcccccgctc
26281 cgggccgtct tccacgccgc ggggaccgcc gaggacgcca cgatcggaag cctcacccgg
26341 gagaggcttc tcgacacctg cgccgcgaag atccggggcg cctggaacct gcatcgcgcc
26401 acgttcgaca tcccgctcga cgtcttcgtc ctctgctcct cggtggtcgc ggtcctcggc
26461 aacgctggcc aggccgccta tggggccgcc aacgcttttc tcgatgcgct cgccgagcgg
26521 cgaaggcaac aggggcggag cgccttgagc gtcgcctggg ggccgtgggc gtcggtcggc
26581 atggccgcaa ggctcggcga caaggatgct cgccgatggg aggagatggg cttcaccccg
26641 atggatcccg ccgtcgcgct gagaggtctc gagcagctcc tgtctctttc ccccgccgtc
26701 ccgcgcgcgc tcgtcgcggt cgtggactgg gatcgctacc ggtcgagggc gggcgagagc
26761 ccgctcatcg aggcgctcgt ctcccggccg aaggacggcg acggcacggc cgcgagcttc
26821 agcgacaagc tcgtgcgctc gaccgcgaga aagcgacacc agctgctggt cgagcacctg
26881 cgtctcgaag tcgccaaggt cctcggctcg tactccgcgc gcagcatcga tccggagaaa
26941 gggttctccg cgctcgggat ggactcgctg atgtcggtcg agctgagaaa ccgcctccaa
27001 tcgagcctcg gcctctcgct ttcctcgacc ctgaccatgg attacccaaa cgtaacggcc
27061 ctggcgactc atctcgacgg ccgcctgaca ccgtcgacga ccgaagacgg agcctcgagc
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
      ....|....| ....|....| ....|....| ....|....|....|....|....|....|
27121 gcctcgcgct cctcgctgga ggagacggac ctcgcgggac tcagtgagag cgaagcacaa
27181 ctactcctgg aaacgctcaa ggagttggag cactaatcat ggaacgagag agtaccgacg
27241 agaagagaac gcccgcgctc gaaaaagcgc tcagagcctt gacgctgatg aaggagaagc
27301 tcgaggccct ggagtatcgg aagcgcgagc ccatcgccgt tgtcggcatg tcctgccggt
27361 ttcccggcga tgccgacacg ccggagaagt tctgggagaa tctctcgaac ggcgtcgacg
27421 cgatcaccga gatccccaag gatcgtttcg acatcgatcg ctactacgac cccaggcccg
27481 ccgttcctgg aaagacgaac actcgatgcg ggggcttcgt tcgggggatc gaggagttcg
27541 accccacgtt tctcggagtc gctcccaagt acgcgtacgg tgtcgatcct cagttccgct
27601 tgttcggggc ggtgtgctgg gaggcgctcg agcgcgccgg ctacgcgccc ttgagcctcg
27661 acgggagcct gaccggggtc tacgtcggca tctggagcat cgactactgg caccggctcg
27721 cttcgcgacg gcccgacctc atcgacgcgg gcatggtggg aggcaacact catagcgtcg
27781 ccgccggggg catctcctac gtcctcggtc tcaaggggcc gagcatggcg gtggataccg
27841 cctgctcctc ctctctcgtg acggtcgacc tcgcgtgcca ggcgcttcgc gccggcacct
27901 gcgacatggc gctcgcggga ggcgtgaacg ccattctctc gcacgagaac ttcgtgtccc
27961 tctccagcat gcaggtgctc gcgcccgacg gacgctgcaa gaccttcgac gcctccgcgg
28021 acggtttcag ccgcggcgag ggcgcgggcg tggtcgtgct caagcggctc tccgacgcgc
28081 tcgccgacgg cgatccggtg ctggccgtga ttcggggcac cgccgtctat caggacggca
28141 agacgagcgg gatcacggtg cccaacgggc cttcgcaaca ggagaccgcg cgccgggcgc
28201 tcgcgcaggc ggggctcacc gccgccgacg tgagctacgt cgaagcccac gggacgggca
28261 cctcgctcgg ggatcccatc gaagtgcagt cgctcgccga cgtgtaccgg ccggggagct
28321 ccgagacgct cctcatcggc tcggtcaagt ccaacatcgg ccacctggag accgcggcgg
28381 ggatcgcggg cctcatcaaa gtgattctgt cgctcgagaa cgaggccttg ccgccgaacc
28441 ttcatttcaa gaacccgaac ccgcatatcc cctgggatcg tctcccggtg aaggtggtgg
28501 cggagaagac cccctggccg cgcggcgcca agcccaggcg agccggggtc agctccttcg
28561 gggcgagcgg cacgaacgcg cacgtgatcg tggaggaggc cccgatcgtt ccccggaagc
28621 cggcggaaaa gcccgagcgg gacgtccacc tgttcacctt ctcctccaag aacggcaagc
28681 gcttgccgga agtggccgct cgggtcgggg aattgctccc cgacgccgac atgcagatcg
28741 ccgacgcttg ctacaccgct agcgtcgcgc gctgtcactt cagcgagcgc gtcgccatca
28801 tggggcgaag ccgtgaggag atcagcgaca ggctcgcggc tttcgccaac gggaagagtg
28861 ggccgggggt tttccgcggg cggctgaaag ggagcgctcg gccttcgctc gccttcctct
28921 tcaccgggca gggcgcccag cacgtcaaga tgggccagag gctctacgag accgagcccc
28981 ggttccgcga ggtcatcgac cgctgcgacg agacgctgtc cgcgcttctt cccctgcgtc
29041 tgctcgacgt cctgtacccc ggagacgggg ggtcgagccc gattcacgag acgacctaca
29101 cccagcccgc tttgttcgcg atcgagctcg gtctcgccag gctctggcag cactgggggg
29161 tggagccgac gcacgtcatg ggtcacagcg tcggcgagta cgtcgccgct tgtatcgccg
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

```
                10         20         30         40         50         60
        ....|....| ....|....| ....|....| ....|....|....|....|....|....|
29221 aggtcttcag cgaagaagac ggtttgcggc tgatcgcgga gcgagcgcgg ctgatgggaa
29281 gcctgcctcg cggcggcgtc atggccgcgg tgctcgcccc cgagtcccgg gtgctcgaga
29341 tgctcgaggt gctcgcgaag ggtcgtgagc gctcgggggc cgtcgcggtc gccgccgtca
29401 acggtcccga gaacgtcgtg ctctccggca acgaggaggc cgtcgcgtca gtgctggccg
29461 agctcgaggc ttccggcatc gaggccaaga cgcttcaagt ctcccacgcc ttccactcgc
29521 cgctcatgga gcccatgctc gcttcgttcg aggcggtggc gcgggaggtc gagttccgtc
29581 agcctcagat tcctttcgtc tccaacctca ccggccagct cgcgagtgac gaggtcaccg
29641 tccccgagta ctgggtcagc cacgttcgcc acgcggtgcg attcgccgac ggcatcaaga
29701 cgctcggcca gctcggctgc cggctctttc tggagatcgg accgaagccc gtgctcgtgg
29761 gcatggggcg cagctccctt cccggcggcg ccacctggct gccgagcctg cg
```

Figure B15 Nucleotide sequence of KKFOS_KSI_pool37_088 insert DNA. This sequence contained 29,812 base pairs. (Continue)

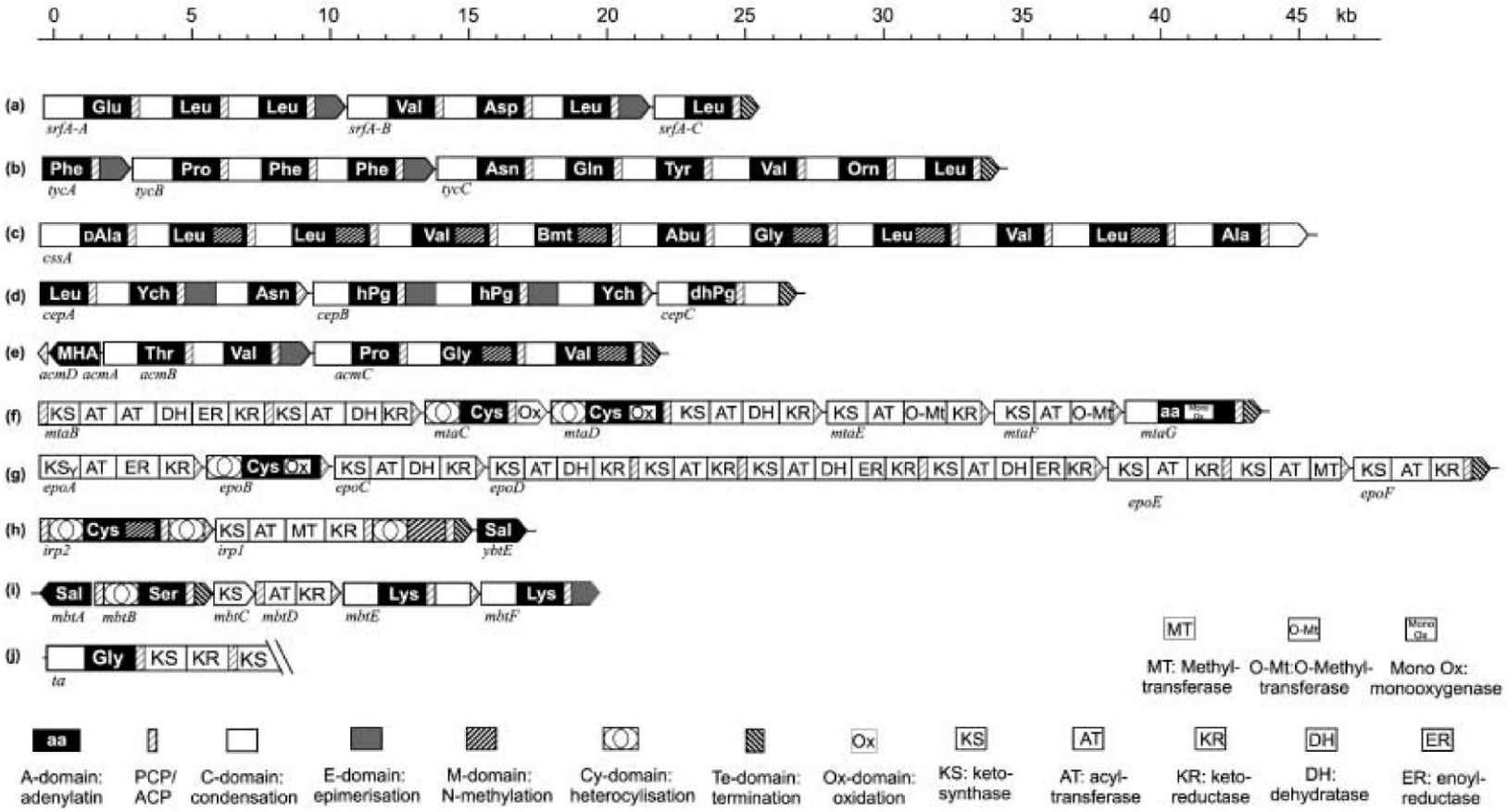


Figure B16 Examples for the organization of NRPS and PKS/NRPS hybrid gene clusters. The cluster for the: (a) surfactin from *Bacillus subtilis* ATCC21332, (b) tyrocidine from *Bacillus brevis* ATCC 8185, (c) cyclosporine from *Tolypocladium inflatum*, (d) chlorocremomycin from *Amycolatopsis orientalis*, (e) actinomycin from *Streptomyces chrysomallus*; PKS/NRPS hybrid (f) myxothiazol from *Stigmatella aurantiaca*, (g) epothilone from *Sorangrium cellulosum* So ce90, (h) yersiniabactin from *Yersinia pestis*, (i) mycobactin from *Mycobacterium tuberculosis*, (j) antibiotic TA from *Myxococcus xanthus*. (Schwarzer and Marahiel, 2001) (Enlargement of figure 2.6)

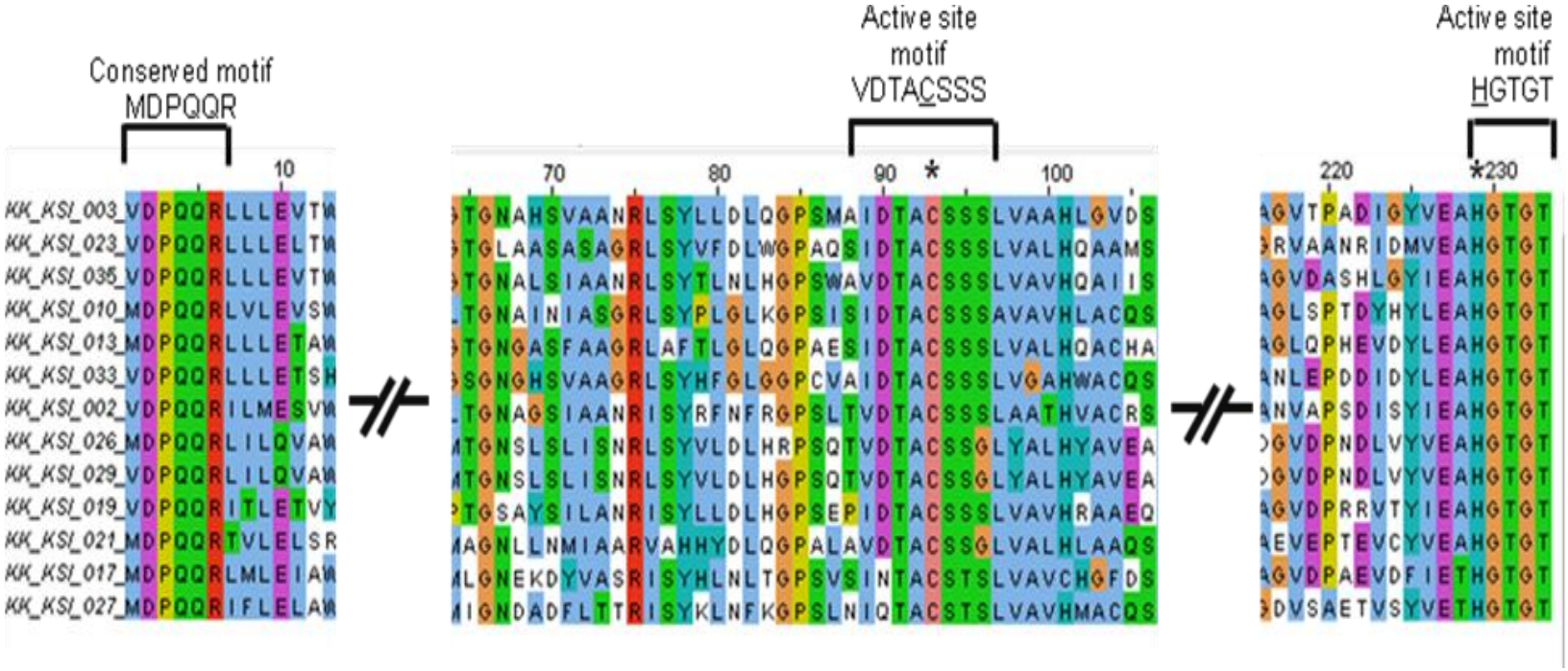


Figure B17 The clustalW2 alignment of partial sequence Type I-PKS KSs from mangrove mud metagenome showing the important characteristics of the conserved motifs for Type I-PKS KS proteins, the cysteine(*) active site motif, VDTACSSS. (Enlargement of figure 4.6)

APPENDIX C

REAGENTS AND BUFFERS

1. EDTA 0.5M (pH 8.0)

Dissolve 186.1 g disodium EDTA dihydrate in 700 ml water.
Adjust pH to 8.0 with 10 M NaOH (~50 ml; add slowly).
Add water to 1 liter and filter sterilize.

2. Glycerol (10% v/v)

Glycerol	1	Volume
H_2O	9	Volume

3. Lysis buffer

Tris-HCl	100	mM
EDTA	100	mM
NaCl	1.5	mM
CTAB	1%	(w/v)
SDS	2%	(w/v)

Adjust pH to8.0

4. NaCl, 5 M

NaCl	292	g
H_2O to 1 liter		

5. SDS, 20% (w/v)

Dissolve 20 g SDS (sodium dodecyl sulfate or sodium lauryl sulfate) in H_2O to 100 ml total volume with stirring. Filter sterilize using a 0.45-µm filter.

6. Sodium acetate, 3 M

Dissolve 408 g sodium acetate trihydrate ($NaC_2H_3O_2 \cdot 3H_2O$) in 800 ml H_2O.

Adjust pH to 4.8, 5.0, or 5.2 (as desired) with 3 M acetic acid

Add H_2O to 1 liter, Filter sterilize

7. TBE (Tris/borate/EDTA) 5×

Tris base (890 mM)	54	g
boric acid (890 mM)	27.5	g
H_2O	960	ml
0.5 M EDTA, pH 8.0,(20 mM final)	20	ml

8. TE (Tris/EDTA) buffer

Tris·Cl, pH 7.4, 7.5, or 8.0	10	mM
EDTA, pH 8.0	1	mM

9. Tris·Cl, 1 M

Dissolve 121 g Tris base in 800 ml H 2 O

Adjust to desired pH with concentrated HCl

Adjust volume to 1 liter with H_2O

Filter sterilize if necessary

Store up to 6 months at 4°C or room temperature

Approximately 70 ml HCl is needed to achieve a pH 7.4 solution, and ~42 ml for a solution that is pH 8.0.

BIOGRAPHY

Mr. Anirut Limtrakul has been graduated with a Bachelor of Science (Microbiology) from the Faculty of Sciences, Prince of Songkla University, Thailand in 1995. Since 1996, he has been working at Srinakharinwirot University. He has entered the degree of Doctor of Philosophy Program in Interdisciplinary Program in Medical Microbiology, Graduate School, Chulalongkorn University since 2004.

During the Ph.D. study, He has obtained the scholarship from the 90th Anniversary of Chulalongkorn University Ratchadaphiseksomphot Endowment Fund.