การสร้างและการคัดกรองคลังเมตาจีโนมิกจากดินป่าชายเลนเพื่อหายีนของเอนไซม์สังเคราะห์เพปไทด์โดยไม่อาศัยไรโบโซม

นายณัฐวุฒิ ลีลากนก

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาเภสัชศาสตรมหาบัณฑิต
สาขาวิชาจุลชีววิทยา ภาควิชาชีวเคมีและจุลชีววิทยา
คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553

CONSTRUCTION AND SCREENING ON MANGROVE SOIL METAGENOMIC LIBRARY FOR NONRIBOSOMAL PEPTIDE SYNTHETASE GENE

Mr. Nattawut Leelakanok

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Pharmacy Program in Microbiology

Department of Biochemistry and Microbiology

Faculty of Pharmaceutical Sciences

Chulalongkorn University

Academic Year 2010

Thesis Title	CONSTRUCTION AND SCREENING ON MANGROVE SOIL METAGENOMIC LIBRARY FOR NONRIBOSOMAL PEPTIDE SYNTHETASE GENE
By	Mr. Nattawut Leelakanok
Field of Study	Microbiology
Thesis Advisor	Associate Professor Nongluksna Sriubolmas, Ph.D.
Thesis Co-Advisor	Associate Professor Pintip Pongpech, Ph.D.

Accepted by the Faculty of Pharmaceutical Sciences, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

…….. Dean of the Faculty of Pharmaceutical Sciences

(Associate Professor Pintip Pongpech, Ph.D.)

THESIS COMMITTEE

…….. Chairman

(Associate Professor Vimolmas Lipipun, Ph.D.)

…….……. Thesis Advisor

(Associate Professor Nongluksna Sriubolmas, Ph.D.)

…….. Thesis Co-Advisor

(Associate Professor Pintip Pongpech, Ph.D.)

…….. Examiner

(Assistant Professor Mali Wirotesangthong, Ph.D.)

…….. External Examiner

(Assistant Professor Suthep Wiyakrutta, Ph.D

ณัฐวุฒิ ลีลากนก: การสร้างและการคัดกรองคลังเมตาจีโนมิกจากดินป่าชายเลนเพื่อหายีนของเอนไซม์สังเคราะห์เพปไทด์โดยไม่อาศัยไรโบโซม อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. นงลักษณ์ ศรีอุบลมาศ, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร. พิณทิพย์ พงษ์เพ็ชร, 181 หน้า

การใช้เมตาจีโนมิกเพื่อศึกษาจุลชีพที่เพาะเลี้ยงไม่ได้ในดินตัวอย่างจากป่าชายเลนอาจช่วยให้ค้นพบเอนไซม์สังเคราะห์เพปไทด์โดยไม่อาศัยไรโบโซม (NRPS) ซึ่งสังเคราะห์เพปไทด์ที่มีความสำคัญทางคลินิก ดินตัวอย่างที่ใช้ในงานวิจัยเก็บจากป่าชายเลนคลองโคน จ. สมุทรสงคราม การสกัดดีเอ็นเอจากดินใช้วิธีทำให้เซลล์แตกโดยตรงและทำให้บริสุทธิ์โดย gel electrophoresis และ dialysis ทำการคัดกรองดีเอ็นเอจากดินเพื่อหาส่วน A domain ของยีน *nrps* ด้วยการทำ PCR โดยใช้ MTF2/MTR primer พบว่าผลิตภัณฑ์จาก PCR มีขนาด 1 กิโลเบส การโคลนผลิตภัณฑ์จาก PCR ที่ได้ ได้โคลนจำนวน 5 โคลน จากการหาลำดับเบสและ sequence alignment ของดีเอ็นเอจากโคลนพบว่ายีนที่เพิ่มจำนวนได้เป็นยีนใหม่ การวิเคราะห์ความสัมพันธ์ทางวิวัฒนาการของยีนเหล่านี้พบว่ามีความสัมพันธ์กับ NRPS ของแบคทีเรียกลุ่ม cyanobacteria, actinobacteria และ proteobacteria นอกจากนี้ พบว่าบางยีนไม่สามารถทำนายกรดอะมิโนที่จะไปกระตุ้นได้เนื่องจากมีลำดับเบสต่างจากฐานข้อมูล ผลดังกล่าวแสดงให้เห็นว่าดีเอ็นเอจากดินป่าชายเลนมียีน *nrps* ที่ใหม่และหลากหลาย เหมาะสมต่อการนำไปสร้างคลังเมตาจีโนม การสร้างคลังเมตาจีโนมิกจากดินป่าชายเลนได้คลังเมตาจีโนมิกขนาด 14,000 โคลน (95 pool) จากการคัดกรองคลังเมตาจีโนมทั้งหมด 31 pool ไม่พบโคลนที่มีฤทธิ์ยับยั้ง *Candida albicans* ATCC 90028 และ *Bacillus subtilis* ที่ดื้อต่อยา chloramphenicol ผลการคัดกรอง pool 30-39 ด้วย PCR โดยใช้ MTF2/MTR primer พบว่าโคลนจาก pool 36 และ37 มี A domain ของ *nrps* ชนิดใหม่ แสดงว่าในคลังเมตาจีโนมิกจากดินป่าชายเลนมียีน *nrps* ชนิดใหม่และหลากหลายซึ่งอาจมีศักยภาพที่จะนำไปสู่การค้นพบยาใหม่ได้

ภาควิชา ชีวเคมีและจุลชีววิทยา ลายมือชื่อนิสิต……..……..……..…..

สาขาวิชา ....จุลชีววิทยา.... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก…….....

ปีการศึกษา ....2553.... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม…..….....

5076609233: MAJOR MICROBIOLOGY

KEYWORDS: MANGROVE SOIL/ METAGENOME/ METAGENOMIC LIBRARY/ NONRIBOSOMAL PEPTIDE SYNTHASE/ *nrps* GENE.

NATTAWUT LEELAKANOK: CONSTRCUTION AND SCREENING ON MANGROVE SOIL METAGENOMIC LIBRARY FOR NONRIBOSOMAL PEPTIDE SYNTHETASE GENE. THESIS ADVISOR: ASSOCIATE PROFESSOR NONGLUKSNA SRIUBOLMAS, Ph.D., THESIS CO-ADVISOR: ASSOCIATE PROFESSOR PINTIP PONGPECH, Ph.D., 181 pp.

Metagenomic study of unculturable bacteria may discover enzymes involved in synthesis of novel natural products, e.g. nonribosomal peptide synthetase (NRPS). Soil sample collected from Klongkone mangrove, Thailand was used for soil direct DNA extraction. Soil metagenomes was then purified and screened for *nrps* by PCR using A domain specific primer, MTF2/MTR. One-kb amplicons were cloned and sequenced for further analysis. Sequence alignment and phylogenetic analysis of deduced amino acid sequences of amplified A domain revealed that they were evolutionary related to NRPSs of cyanobacteria, actinobacteria and proteobacteria. These implied the novelty and diversity of *nrps* from mangrove soil metagenome. These metagenomes were used for metagenomic library construction in *E. coli* resulted in 14,000 library clones divided into 95 pools. A total of 31 pools were screened for anti-*C. albicans* ATCC 90028 and anti-*B. subtilis* (chloramphenicol resistant) activities. PCR screening of pool 30-39 of the library found A domain of *nrps* in pool 36 and 37. Sequence alignment and phylogenetic analysis revealed the novelty of these *nrps.* This suggested that mangrove soil metagenomic library harbored the novel and diverse *nrps* genes which might potentially lead to new drug discovery.

Department: Biochemistry and Microbiology	Student's Signature……..….
Field of Study: Microbiology	Advisor's Signature……..…
Academic Year: 2010	Co-Advisor's Signature……

ACKNOWLEDGEMENTS

I would like to acknowledge the financial support from *Chulalongkorn University Graduate Scholarship to Commemorate the 72nd Anniversary of His Majesty King Bhumibol Adulyadej* which provided a scholarship during Master's Degree program at the Department of Biochemistry and Microbiology, Faculty of Pharmaceutical Sciences, Chulalongkorn University and The 90th Anniversary of Chulalongkorn University Fund for financial support throughout the research study.

Most of credits in this thesis should be justifiably passed to Associate Professor Dr. Nongluksna Sriubolmas for her valuable advice, kind help on my thesis and manuscript.

I would like to express my deep appreciation to my thesis co-advisor, Associate Professor Dr. Pintip Pongpech for her guidance, suggestion and consultation.

I would like to thank chairman of thesis committee, Associate Professor Dr. Vimolmas Lipipun for her comments and support for getting the Graduate Scholarship and Assistant Professor Dr. Suthep Wiyakrutta, an external examiner, for his valuable advice. I also would like to thank Assistant Professor Dr. Mali Wirotesangthong for being the thesis committee.

I would like to thank the faculty staff in section of Microbiology. I gratefully acknowledge the generous help of Assistant Professor Penpun Naenna for being my best consultant.

Special thanks for Peerapat Roongsattham and Ken Chuaphalakit for proofreading my thesis and manuscript and for every kinds of support. I am indebted to Anirut Limtrakul for a kind help, practical advice and technique teaching. Thank Woratouch, Chutima and Ssvalee for a valuable help. Thank you all of my friends for every kind of help that all of you have provided.

Finally, I am thankful to my family members who always have faith and believe in me. Thank you for the supports and strength given to me. Without them, I could not come this far. Thank you very much

CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

AMP	=	Adenosine monophosphate
ATCC	=	American Type Culture Collection
ATP	=	Adenosine triphosphate
BAC	=	Bacterial Artificial Chromosome
bp	=	Base pair
^{0}C	=	Degree Celsius
Cm^{R}	=	Chloramphenicol resistant
CTAB	=	Cetyl trimethylammonium bromide
DMSO	=	Dimethyl sulfoxide
DNA	=	Deoxyribonucleic acid
cm	=	Centimeter
e.g.	=	For example
et al.	=	And other
EDTA	=	Ethylenediaminetetraacetic acid
g	=	Gravitational force
HEPES	=	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	=	Human immunodeficiency virus
IPTG	=	Isopropyl-beta-D-thiogalactopyranoside
Kb	=	Kilo base pairs
LB	=	Luria Bertani
µl	=	Microliter
ml	=	Milliliter
MH	=	Muller-Hinton
min	=	Minute

mRNA	=	Messenger ribonucleic acid
mm	=	Millimeter
mM	=	Millimolar
ng	=	Nanogram
nm	=	Nanometer
NRPS	=	Nonribosomal peptide synthase, Nonribosomal peptide synthetase
OD	=	Optical density
PCP	=	Peptidyl carrier protein
PCR	=	Polymerase chain reaction
PVP	=	Polyvinyl pyrrolidone
pH	=	Power of hydrogen
PKS	=	Polyketide synthase
rRNA	=	Ribosomal ribonucleic acid
rpm	=	Round per minute
SDA	=	Sabouraud's Dextrose Agar
SDS	=	Sodium dodecyl sulfate
T	=	Transmittance
TBE	=	Tris, Boric acid and EDTA
USFDA	=	The Food and Drug Administration
UV	=	Ultraviolet
X-Gal	=	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

CHAPTER I

INTRODUCTION

1.1 Introduction

The enormous antibiotics market size with significant continuing growth rate (Chrisoffersen, 2006) makes new antibiotics discovery the most research field of interest. Additional reason for support the importance of novel antibiotics discovery is antibiotic resistance problems (World Health Organization, 2001) which have increased and become a major problem in global public health (World Health Organization, 2005). Drug resistance microorganisms have increased risk of morbidity, mortality and overall health care expenditures (Nicolau, 2009; Maragakis, 2008; Sipahi, 2008; Howard *et al.*, 2001). Many infectious diseases are untreatable by currently available antibiotics. Furthermore, development of new antibiotics in the pharmaceutical industry is inadequate for the demand to deal with drug resistance problem (Davies, 2006). As a consequence, new antibiotics discovery is always the need for treatment of infectious diseases.

Antibiotics being used nowadays were discovered from bacteria, actinomycetes and fungi, isolated from natural habitats e.g. soil, water, plants and animals. Soil seems to be the potential source of antibiotic producers because soil microorganisms have high diversity (Torsvik *et al*., 1990; Gans *et al*., 2005) and quantity (Pettit, 2004; Gans *et al*., 2005), especially in the first ten centimeters from the surface (Takahashi and Omura, 2003). The screening on soil microorganisms discovered many clinically useful anticancers e.g. bleomycin, daunorubicin, adriamycin, pentostatin, streptozocin, mithracin, mitomycin C, and actinomycin D. (Pettit, 2004), and antibacterials, e.g. daptomycin (Baltz, 2008), erythromycin (Malmborg, 1986), gramicidin (Gall and Konashev, 2001), novobiocin (Smith, 1956), rifamycin (Alvarez *et al*., 1990), vancomycin (McCormick *et al*., 1955-1956) and streptomycin (William, 2004). At present,

screening on soil microorganisms has a tendency to rediscover of known compounds. In order to reduce a redundancy of strains and compounds, the uncommon microorganisms from the unique or extreme habitats, such as mangroves (Kathiresan and Bingham, 2001), sea (Li and Qin 2005) and hot springs (Rainey and Oren 2006) should be screened (Faber, 2006). From these habitats, many novel enzymes and biomolecules were found (Schiraldi and Rosa, 2002; Den Burg, 2003; Ferrer *et al*., 2007).

The unculturable microorganisms become the attractive source of natural products because they are seldom studied. Since less than one percent of microbe population in soil sample could be cultured (Handelsman *et al*., 1998; Hallam *et al*., 2003), the unculturable population was proposed to dominate in habitats (Staley *et.al*, 1985; Amann *et al*., 1995; Stein *et.al*, 1996). Many procedures were developed for access to the unculturable population. The culture-dependent methods such as using new culture media (Balestra and Misaghi, 1997) or changing the culture condition (Mitsui *et al*., 1997) have many constraints for application in the study of unculturable population (Handelsman *et al*., 1998). Another approach is culture independent technique, metagenomics. The term metagenomics was first used to call the collective genomes of soil microflora and the process of metagenomics cloning and clones screening (Handelsman *et al*., 1998). For more updated version, they refer to “the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and cultivation of individual species" (Chen and Pachter, 2005).

By the metagenomics method, many known and novel natural products such as enzymes, antibiotics, anticancers and multi-enzyme complexes, e.g. polyketide synthases and nonribosomal peptide synthetases (NRPSs), have been found (Daneal, 2003; Kennedy *et al*., 2007). NRPS is one of the most interested enzyme complexs since they involves in the production of many clinically important antibacterial, antifungal, antiviral, immunosuppressant, and anticancer drugs, for example penicillins, cephalosporins, glycopeptides, cyclosporins and bleomycins. (Felnagle *et al*., 2007), and their gene, *nrps*, has an ability to be genetically engineered which aiming for

peptides structure modification. NRPS is organized in a repeated modular structure. Each module consists of structurally independent domains with a specific function. The minimal module consist of at least three core domains, adenylation (A) domain, thiolation (T) domain (also known as peptidyl carrier protein; PCP) and condensation (C) domain. The A domain recognizes and activates amino acid substrate to aminoacyl adenylate which binds to 4'-phosphopantetheinyl co-factor of the PCP domain. Amino acids bound to PCP domain are then elongated by C domains into peptidyl chain. Alternative domains such as epimerization domain which is responsible for the change of the Cα stereochemistry may be found in some module. In the last step of nonribosomal peptide biosynthesis pathway, release of the peptide from the NRPS is responsible by thioesterase (Te) domain. The number and order of the modules correspond to the amount of amino acids and the sequence of the peptide being synthesized, respectively (Lautru and Challis, 2004). They also affect the size of *nrps* which usually range from 5.9-48.5 kb (Kratzschmar *et al*., 1989; Stachelhaus and Marahiel, 1995; Stachelhaus *et al*., 1996; Marahiel *et al*., 1997).

Due to the potential of metagenomics in novel natural products discovery, using metagenomics approach to capture *nrps* from soil with unique-characteristics, for example mangrove soil, may increase the opportunity to discover the novel *nrps* that produces clinically important peptides which could bring this research area closer to the light at the other end of the tunnel.

1.2 Objectives

The objectives of this study were to discover *nrps* gene from mangrove soil metagenome. Specifically:

1. Screening on mangrove soil metagenome for new *nrps* gene by PCR
2. Construction of soil metagenomics library
3. Screening the soil metagenomics library for *nrps* gene

CHAPTER II

REVIEW OF LITERATURE

2.1 Producers of natural products

Natural products and their derivatives are the single most important source of novel drugs. Among the capable sources of natural products including microbes, plants, animals and minerals, microorganisms are the most productive ones (Newman *et al*., 2003; Ganesan, 2008; Harvey, 2008). Not all of them but some specific groups, e.g. actinomycetes, marine cyanobacteria and endophytes are active natural products producer. For details of natural products from marine actinomycetes and cyanobacteria, see Table A1 in Appendix A.

Actinomycetes refer to gram positive mycelia forming bacteria in order *Actinomycetales*. Actinomycetes, particularly in *Streptomyces* and *Micromonospora* genera, are the most pharmaceutically important because approximately half of the discovered bioactive metabolites including many clinically important antibacterial antibiotics, e.g. erythromycin, streptomycin, vancomycin and tetracycline, antitumor and immunosuppressant are also produced by them (Jensen *et al*., 2005; Lam, 2006). Actinomycetes usually have been isolated from terrestrial habitats but can be discovered from marine environment, as well. Some of marine actinomycetes may originate from terrestrial actinomycetes which were occasionally washed into marine water and been able to adapt to live in that environment, but others are true marine actinomycetes which can be found only in sea water (Bredholt *et al*., 2008). *Rhodococcus marinonascens* is the first marine actinomycetes that were taxonomically described. After that, *Salinispora* which has an obligate requirement of sodium for growth was discovered and classified (Fenical and Jensen, 2006). Marine actinomycetes adapt themselves to the extreme environment in the ocean, for

example high pressure, anaerobic conditions, extremely low or high temperature, high acidic condition, according to their habitats. It is believed that marine and terrestrial actinomycetes are different in characteristics since marine environmental conditions are significantly different from terrestrial ones. Therefore, they might produce different types of bioactive compounds. Many novel metabolites produced by marine actinomycetes usually have unusual structures and properties.

The cyanobacteria are one of the most productive groups of microalgae. They proliferate in marine and freshwater habitats, resulting in the formation of water red tide blooms. Cyanobacteria are suitable objects for natural product discovery since they usually produce toxins (Neilan *et al*., 1999). Many of their secondary metabolites are produced via non-ribosomal peptide synthetase (NRPS) or mixed polyketide-NRPS pathway (Tan, 2007). Filamentous and heterocystous cyanobacteria are the most likely sources of novel natural products within the phylum since they contain diverse and novel *nrps* and *pks* genes (Ehrenreich *et al*., 2005). Interestingly, cyanobacteria, which were found with symbiosis in bioactive compounds producing marine invertebrate, are usually the primary producer of the secondary metabolites, for example patellamides which produced from tunicate *Lissoclinum patella* are synthesized by *Prochloron didemni*, a unicellular cyanobacterial symbiont (Schmidt *et al*., 2005). One example of the important natural products from cyanobacteria is cyanovirin, a potent HIV fusion inhibitor. At present, it has been placed on an accelerated track for clinical development (Dunlap *et al*., 2007).

Other major natural products producers are endophytic microorganisms, microbes that host in the internal tissue of living higher plants without causing any immediate and apparent negative effects (Saikkonen *et al*., 1998; Saikkonen *et al*., 2004). Each individual plant can host one or more endophytes. Even though they live together in symbiotic or mutualistic relationships, endophytes can become the aggressive saprophytes or opportunistic pathogens in some conditions. The most common endophytes appear to be fungi and bacteria. The most frequently isolated endophytes are fungi which are usually fungi imperfecti or deuteromycetes. The estimate

number of endophytic fungi may be at least 1 million species. Endophytic fungi usually produce a specific phytochemicals which is a unique characteristic of the host because gene producing secondary metabolite of the endophyte might genetically recombine with the host gene during evolution. Endophytes studies not only allow the discovery of new metabolites but also facilitate the production of rare natural products. The collection and fermentation of endophytes are bypass for the production of many valuable bioactive compounds instead of harvest slow-growing and uncommon plants. Rational selection of plants for endophytes studies include those with unusual biology, novel strategies for survival, ethnobotanical history and are endemic (Strobel *et al*., 2004). Endophytes from mangrove species are significant sources of useful metabolites since mangroves usually produce unusual secondary metabolites, some of which are endemic and used in traditional medicine (Ananda and Sridhar, 2002).

2.2 Soil as a source of microorganisms

Soil in traditional meaning is “the natural medium for the growth of land plants” (Soil survey staff, 2006). In technical term, it refers to “a natural body comprised of solids (minerals and organic matters), liquid, and gases that occurs on the land surface, occupies space, and is characterized by one or both of the following: horizons, or layers, that are distinguishable from the initial material as a result of additions, losses, transfers, and transformations of energy and matter or the ability to support rooted plants in a natural environment” (Soil survey staff, 2006). From both of the above meaning, mangrove sediment is considered as soil (Ferreira *et al*., 2007). The upper limit of soil is end at air, water, plants or non-decomposition plant materials. The lower boundary of soil is thin cemented horizons that are impermeable to roots, but not below 200 centimeters (cm) from surface. Although approximately all soil consists of other composition other than mineral and organic components, most soils are dominantly divided into two categories, mineral soil and organic soil (histosol). Mineral soils have less than 20 to 35 percent

organic matter by weight. Organic soil refers to soil at the upper 80 cm that consists of organic matter more than a half (Soil survey staff, 2006). Soil organic matter, a derivative of biological substances including thermally altered materials contained within the soil matrix or on the soil surface, composes of living and nonliving component. Living components are organic materials associated with the living cells existed in soil (plants, animals or microorganisms). Non living components consist of dissolved organic matter, insoluble particulate organic matter, humus and inert organic matter. Humus is a mixture of amorphous organic materials mainly biomolecules (lipids, polysaccharides and proteins) and non-identifiable molecules, e.g. humic substances. Inert organic matter usually refers to carbonaceous organic material, for example charcoal, charred plant residues, graphite, and coal.

Mineral components in soil form matrix which absorbs organic materials. Soil matrix contributes to soil architecture which refers to the arrangement of pores in soil and soil particles. Since soil particle affects water, oxygen and decomposer organisms availability (via the entrapment and isolation of decomposers from organic materials), it influences the biological stability of organic materials (Baldock and Skjemstad, 2000). Soil mineral particles are regularly bound together into aggregates, the larger secondary particles. Among them are soil spaces which are the connection of various sizes of pores. Soil pores are ranging in size from diameter less than 0.1 millimeters (mm) (micropores) to more than 20 mm (macropores) (Baldock and Skjemstad, 2000). Individual micropores usually contain only one type of microorganisms. Categorizing the mineral particles according to their size is called soil separate. The three main separates - sand, slit and clay - have diameter in mm range from 2.0-0.05, 0.05-0.002 and less than 0.002, respectively. The sand separate is subdivided into very coarse sand (2.0-1.0 mm in diameter), coarse sand (1.0-0.5 mm), medium sand (0.5-0.25 mm), fine sand (0.25-0.10 mm) and very fine sand (0.10-0.05 mm) (National employee development staff, 1987). The relative percentage of sand, slit and clay classify soil into 12 major soil textural classes. Soil texture has an impact on diversity and quantity of microorganisms (Foster, 1988; Girvan *et al*., 2003) due to soil

pedogenesis (Ulrich and Becker, 2006), and also affects the attachment of microorganisms to soil particle (Bakken, 1985). Small organisms (0.3 mm in diameter) are found as a single cell in dense fabrics of clay or humified organic matter while larger bacteria form small colonies in the larger micropores or associated with substantial deposits of organic matter, e.g. fecal pellets and cell-wall debris. Soil microfauna and fungi mainly occupy the larger voids (Foster, 1988).

Soil microorganisms play an important role in soil biogeochemistry. They appear to have an influence in mineral and organic compound cycle, pedogenesis, soil structure, soil fertility, soil quality and above ground ecosystems (Alexander, 1964; Schloter *et al*., 2003; Kirk *et al*., 2004). On the contrary, soil properties directly affect microbial diversity and quantity (Aislabie *et al*., 2008). Microorganisms in soil are high in biodiversity (Torsvik *et al*., 1990; Kirk *et al*., 2004; Gans *et al*., 2005). Soil has a larger amount of microorganisms when compared to other environments (Pettit, 2004; Gans *et al*., 2005). Most of bacteria and fungi in soil, and other environment, cannot be cultured (Staley *et.al*, 1985; Amann *et al*., 1995; Stein *et.al*, 1996; Kirk *et al*., 2004). The culturable microorganisms in soil are approximately only 0.1-1 percent from all microbial population (Handelsman *et.al*, 1998; Hallam *et.al*, 2003). Some studies revealed that bacteria and fungi are found densely in soil from surface to one meter depth (Lavahun *et al*., 1996). For actinomycetes, they overcrowd at the first 10 cm of soil surface (Takahashi *et al*., 2003).

Soil microbes are responsible for the production of several clinically useful lead compounds. Anticancers, for example bleomycin, daunorubicin, adriamycin, pentostatin, streptozocin, mithracin, mitomycin C, and actinomycin D (Pettit, 2004) and antibacterials, e.g. daptomycin (Baltz, 2008), erythromycin (Malmborg, 1986), gramicidin (Gall and Konashev, 2001), novobiocin (Smith, 1956), rifamycin (Alvarez *et al*., 1990), vancomycin (McCormick *et al*., 1995-1996) and streptomycin (William, 2004) were found from soil screening. Other examples of drugs developed from lead discovered from soil microbes are from soil actinomycetes, e.g. acarbose, aztreonam, cephalosporins, ivermectin, pentostatin, orlistat, penems

and tacrolimus. Some are from soil bacteria, for example mupirocin and gusperimus. The rest are produce by soil fungi, for example caspofungin, lovastatin, ciclosporin (Ganesan, 2008).

2.3 Natural products form mangroves (Kathiresan and Bingham, 2001; Food and Agriculture Organization of the United Nations, 2007)

2.3.1 Overview

The term “mangrove” refers to both plants and forest ecosystem. In general, mangroves are salt-tolerant forests found along the intertidal zone in the tropics. They consist of certain plant families, e.g. Rhizophoraceae, Avicenniaceae and Combretaceae which have developed physiological, structural and morphological adaptations to the mangrove habitat, for example they have an aerial roots system, desalination mechanism or viviparous reproduction. Mangroves grow mainly on soft soil (Ferreira *et al*., 2007) and may be found as isolated patches of dwarf trees or as prolific forests under suitable environmental conditions. The exact number of mangrove species is ranged from 50 to 70 according to different classifications. Mangrove areas in Southeast Asia are the largest of any regions and are outstanding for their high biodiversity. More than 50 mangrove species grow along the coasts, some of them (*Aegiceras floridum, Camptostemon philippinensis, Heritiera globosa*) are endemic to the region. Some of the relatively common species are considered rare in the region as a whole, e.g. *Ceriops decandra, Osbornia octodonta, Scyphiphora hydrophyllacea, Sonneratia ovata*. Mangrove forests in Southeast Asian countries are well-constructed because of high rainfall, riverine inputs and the edaphic and coastal features of this region. In Thailand, trees may grow to a height of 20–30 meters along these coasts.

Mangroves are outstanding in uniqueness and diversity which make them a suitable area for novel drug discovery. Mangroves may be considered as an extreme environment since they

grow in brackish water which is varied in salinity and fluctuating tidal level. Mangrove soil is muddy, anoxic and contains low nitrogen and phosphorus content. In addition, estuaries, the place where mangroves found, often act as efficient reservoirs of pollutants either from river or ocean (Lugo, 1998; Kehriga *et al*., 2003). Therefore, mangrove soil in many areas is contaminated by contaminants, e.g. oils, organic solvents, heavy metal and toxic chemicals (Canestri and Ruiz, 1973; Volkman *et al*., 1994; Kehriga *et al*., 2003; Vane *et al*., 2009). Organisms in this area must morphologically and physiologically adapt and develop to this habitat and become unique. Plants are the best example. Environmental conditions within mangrove forests, especially salinity, make it extremely difficult for non-halophytic and non-wetland plants to grow and reproduce (Lugo, 1998). Only 34 species in nine genera and five families are considered as major species, true mangroves which occur exclusively in mangroves and have the ability to form pure stands. Furthermore, microorganisms in mangroves are also uncommon. They are mixed population of terrestrial and marine microbes. Terrestrial microorganisms usually inhabit at the top of mangroves where the salt water never reaches while marine microbes reside the lower part of plants (Nambiar and Raveendran, 2009). Soil microorganisms are halophiles and usually anaerobes. Some of them adapt to the contaminated environment and develop the ability to tolerate the specific toxicants found in mangroves, e.g. organic solvents tolerance bacteria (Sardessa and Bhosle, 2002) and metal tolerance fungi (Zhihong and Yang, 2009). Mangroves are also a unique habitat for a certain group of fungi called manglicolous fungi (Nambiar and Raveendran, 2009) which are an important decomposer in the mangrove ecosystem and usually found on ground mangrove materials. Beside the uniqueness of mangroves, they are also a diverse ecosystem. Mangrove ecosystems are among the world's most productive ecosystem since they are able to store large amounts of organic carbon derived from suspended material from the river and ocean (Ellison, 1998; Komiyama *et al*., 2008) and atmospheric carbon dioxide (Lal, 2005). Mostly, the organic carbon accumulates in the upper 1.5 meters of the mangrove soil (Kristensen *et al*., 2008). Organic matter is an important component of soil and related to quantity and quality

of above ground plant and microbial biomass in soil (Jia *et al*., 2005). Although mangrove forest is the most species-poor forest ecosystems in the tropics (Lugo, 1998), the mangrove microorganisms are very high in diversity. A variety groups of bacteria including sulfate-reducing bacteria, methanogenic bacteria, nitrogen fixing bacteria and photosynthetic bacteria including purple non-sulfur bacteria and green sulfur bacteria are found epiphytic or in soil. Some of mangrove bacteria are symbionts or parasites. Besides, surprisingly diverse fungal communities are seen in mangroves. Furthermore, mangroves are also the area where most of pharmaceutical important microbes are found, for example marine actinomycetes (Fenical and Jensen, 2006; Mitra *et al*., 2008.), marine cyanobacteria (Tan, 2007; Jones *et al*., 2009), and endophytic fungi (Tomita, 2003).

Mangrove forests are used as a source of wood, food, fodder and other forest products, e.g. tannin, alcohol, sugar, honey and medicine. In ecological viewpoint, mangrove ecosystem is a nursery habitat for juveniles of fish and can be used for aquaculture. Moreover, microorganisms isolated from mangroves also help in waste disposal, for example bioremediation of crude oil (Odokuma and Dickson, 2003), waste water (Ye *et al*., 2001; Wu *et al*., 2008) and plastic degradation (Kathiresan, 2003; Kumar *et al*., 2007) Mangrove forests also help to protect nearby marine ecosystem by entrapping upland runoff sediments from upland rivers. They reduce, protect, and prevent coastal erosion caused by the effects of wind, waves and water currents. In addition, mangroves are major sources of drug discovery since they have unique biochemical pathways which produce unique metabolites. From the historical period, human in tropical and subtropical region used mangroves in traditional medicine. Some of these folk medicines are still being used and researched in ethonopharmacology field. In addition, scientists found that not only mangroves itself, other members of the ecosystem, e.g. bacteria, fungi and animals also produce bioactive compounds.

Despite mangroves are highly beneficial to humans, they have often been undervalued and destroyed. Mangroves area in Thailand has been reduced annually. However, the promotion

in mangroves reforestation resulted in the decrease of deforestation rate from 1.1% per year in 1980-1990 to the annual rates of 0.3% in 2000-2005. The main causes of loss of mangroves area in Southeast Asia, including Thailand, are overexploitation and the development of commercial shrimp farms which has been promoted because of its high economic value. This activity causes the loss of ecosystem diversity. In particular, conversion of forest to agricultural ecosystems affects several soil properties especially soil organic carbon (Lal, 2005). Bacterial communities in mangroves are also affected by shrimp farming by the use of chemicals, antibiotics and exotic species (Sousa, 2006). Other factors affecting mangroves biodiversity are temperature, pH of water (Oliveira and Pampulha, 2006) and pollutants. Hydrocarbons and heavy metals have impact upon mangroves, microbes and marine organisms (Edwards *et al*., 2001; Agoramoorthy *et al*., 2007; Labud *et al*., 2007; Zhou *et al*., 2009). Hydrocarbons, e.g. gasoline (Labud *et al*., 2007) and polycyclic aromatic hydrocarbon (Zhou *et al*., 2009) inhibit microbial growth and also affect nutrient cycling in soil.

2.3.2 Bioactive compounds from mangroves

Mangrove plants commonly produce tannin, phenolic compounds and organic osmolytes for specific purpose (Kathiresan and Bingham, 2001). Tannins in mangroves might prevent the plant from insect herbivores and inhibit the growth of bacteria, fungi and phytoplankton. Tannins from mangroves are used for leather work and for the curing and dyeing of fishing nets. The example of medical use of tannin is the bark of *Rhizophora mangle*, the red mangrove. Red mangle bark has been used traditionally in folk medicine of Caribbean countries due to its antiseptic, astringent, haemostatic and antifungal properties. It contains tannin as major constituent and others, e.g. epicathechin, cathechin, gallic acid, ellagic acid, chlorogenic acid, fatty acids and carbohydrates as minor constituents. Tannin component in Rhizophora has been experimentally proven to possess antibacterial, heal wounds and have antiulcerogenic effects

(Berenguer *et al*., 2006). The *in vivo* studies showed that *Rhizophora mangle* has gastroprotective effect which appears through an antioxidant and prostaglandin-dependent way (Perera *et al*., 2001; Berenguer *et al*., 2006). Mangroves produce phenolics and peroxidases from oxygen in order to waste excess light energy since they grow in high sunlight tropical environments. In addition, phenolic compounds in mangroves, especially in root, act as root growth hormone. Phenolic compounds, as opposed to tannins, stimulate phytoplankton growth. As medicine, polyphenols from red mangle bark exhibited cyclooxygenase-2 inhibitory activity and secretory phospholipase A(2) inhibitory activity. These components contribute to anti-inflammatory activity of the aqueous extract from Rhizophora bark (Marrero *et al*., 2006). Organic osmolytes serve for salt regulating purpose maintaining osmotic balance. Example of organic osmolites from mangroves species are mannitol, proline, glycine betaine, asparagines, stachyose and purine nucleotides.

A large amount of bioactive compounds from mangroves have been studied. They possess various activities range from antiviral to insecticide. A few mangroves species, particularly those belonging to the family Rhizophoraceae, show particularly strong antiviral activity. Some have activity against clinically important pathogenic viruses including human immuno-deficiency virus (HIV), *Vaccinia* virus, encephalomyocarditis virus, new castle disease virus and hepatitis-B viruses (Nakashima *et al*., 1996; Premanathan *et al*., 1999; Li *et al*., 2006). Active components which contribute to potent anti-HIV activity may be acid polysaccharides (galactose, galactosamine, glucose and arabinose). True mangrove floras also contain other bioactive metabolites, for example diterpenoids, naphthoquinone and procyanidins which show antibacterial activities against many bacterial species including *Bacillus subtilis, Pseudomonas aeruginosa, Staphylococcus aureus* and mycobacteria (Rojas and Coto, 1987; Han *et al.*, 2005a; 2005b; Han *et al*., 2007; Wangensteen *et al*., 2009). Some of mangrove plants also exhibit antifungal activity (Rojas and Coto, 1987). Mangroves in genus Rhizophora exhibits antioxidant activity (Vijayavel *et al*., 2006; Suganthy *et al*., 2009) while those in family Verbenaceae show

strong cytotoxic (Xu *et al*., 2004; Jones *et al*., 2005; Han *et al*., 2007). In addition to chemotherapeutic agents, mangroves compounds are well known for mosquitoes repellant. Mangrove extracts kill larvae of the mosquitoes *Anopheles stephensi, Culex tritaeniorhynchus*, *Culex quinquefasciatus,* and *Aedes aegypti* either in smoke form or solution. Others compounds discovered from mangrove plants may also have potential to be developed as drugs for diabetes mellitus, (Tamrakar *et al*., 2008) atherosclerosis (Owen *et al*., 2007) and Alzheimer's disease (Suganthy *et al*., 2008). For details of natural products from mangroves, see Table A2 in Appendix A.

Another major source of natural product from mangroves is microorganisms which live either endophytic or in soil. Natural products from plants may be produced from endophytic microorganisms (Cheplick and Clay, 1988; Tomita, 2003; Lin *et al.,* 2005). Most of bioactive compounds from mangrove endophytes are produced from endophytic fungi and usually have antibacterial or cytotoxic activity (Huang, H., *et al*. 2007; Lin *et al*., 2008a; 2008b; Kjer *et al*., 2009; Xu *et al*., 2009). Mangrove soil contains an enormous amount of microorganisms which have specificity and diversity (Kathiresan and Bingham, 2001; Marchand, 2003). Moreover, mangrove soil harbors marine actinomycetes and cyanobacteria which are robust sources of novel natural products (Fenical and Jensen, 2006; Tan, 2007; Jones *et al*., 2009). In mangrove forests, actinomycetes are likely to be found in rhizosphere soil rather than in plant tissue (Hong *et al*., 2009).

2.4 Nonribosomal peptide synthetase

Polypeptides are mostly synthesized by ribosome, some are not. The first nonribosomal peptide discovered in 1963 was tyrocidine, a cyclic decapeptide produced by *Bacillus brevis*. Study in gramicidin S (Hori *et al*., 1989; Turgay, Krause and Marahiel 1992; Saito *et al*., 1994), tyrocidine (Mootz and Marahiel, 1997) and polymixin B synthesis introduced the existence of

ribosome-independent peptide synthesis pathway to public (Schwarzer *et al*., 2003). Biochemistry of nonribosomal peptide synthetase (NRPS) was hypothesized as modular enzymatic mechanism. The analysis of *nrps* gene found that the number of repeating encoded sequences in *nrps* gene is equivalent to the number of amino acids activated by NRPS (Schwarzer *et al*., 2003). Extensive studies of the structures and the functions of NRPSs nowadays combining with the advanced genetic engineering technology may lead to a possible potential NRPS drug modification which is important to new drug discovery since many nonribosomal peptide drugs are clinically important and widespread used for medically important conditions (Doekel *et al*., 2008; Velkov and Lawen, 2009).

nrps genes are found in a wide range of organisms, including bacteria and fungi, but are not known in plants and animals. NRPS systems in eukaryotes are always single polypeptides which have complete function while NRPS module in prokaryotes is often a multiple polypeptide which has to assemble in order to form a functional NRPS (Velkov and Lawen, 2009). In bacteria, *nrps* gene distribute vastly in phylum proteobaceria, firmicutes, actinobacteria, cyanobacteria and planctomycetes. Bacteria in phylum proteobaceria usually produce siderophores which act as iron chelating compounds while mycobactin-related siderophores are produced by actinobacteria. Cyanobacteria produce toxic secondary metabolites, e.g. microcystins and nodularins, which are cyclic peptides that cause acute hepatotoxicity and often lead to bloom formation in marine and fresh water (Donadio *et al*., 2007). In fungi population, NRPSs are much more abundant in euascomycetes than in basidiomycetes but rarely found in chytridiomycota, zygomycota, schizosaccharomycota, and hemiascomycota. Most of *nrps* genes discovered by genomic sequencing of fungi were from *Aspergillus* and *Cochliobolus* species (Jirakkakul *et al*., 2008). *nrps* genes in fungi generally evolve rapidly. This resulted in discontinuous distribution of the genes, and difficulties in identifying whether the genes have common ancestors. However, some of *nrps* genes, e.g. *nps2*, *nps4*, *nps6* and *nps10* in *Cochliobolus heterostrophus* are relatively conserved. The conservation pattern among *nrps* gene might relate to their function. *nrps* genes

involving in growth and development show less variation in copy number and more conserved in domain pattern in comparison with *nrps* genes involved in more niche-specific functions (Bushley and Turgeon, 2010). Examples of medically significant *nrps* genes from fungi are *acvA* gene and *simA* gene. The *acvA* gene which is produced from *Aspergillus nidulans* and *Penicillium chrysogenum* controls the production of the precursor of β-lactam antibiotics, for example penicillin. The *simA* gene from *Tolypocladium inflatum* controls the production of cyclosporin A.

Genes coding for NRPS are organized in operons or in clusters (Caboche *et al*., 2008) as shown in Figure 1.

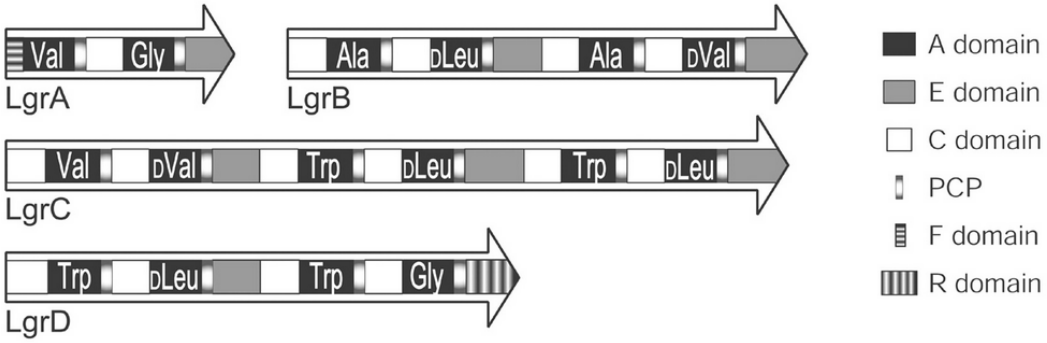

Figure 1: Organization of *nrps* gene (Linear tyrocidine synthethase gene) (Kessler *et al*., 2004)

nrps cluster usually vary from 5.9-48.5 kb in length (Kratzschmar *et al*., 1989; Stachelhaus and Marahiel, 1995; Stachelhaus *et al*., 1996; Marahiel *et al*., 1997). *nrps* genes of prokaryotes do not have intron, while those of eukaryotes usually have several introns and exons (Velkov and Lawen, 2009). *nrps* genes, either in bacteria or fungi, frequently undergo horizontal gene transfer, so their corresponding metabolites are not conserved across the kingdom (Brushley *et al*., 2008; Khaldi *et al*., 2008; Rounge *et al*., 2009). Considering domain conservativeness, condensation domain is less conserved than adenylation domain and thiolation domain. (Stein and Vater, 1996) However, various modules of *nrps* show several highly conserved motifs, e.g. highly conserved signature sequence, A1-A10, (as shown in Table 1) which are important for ligand

binding in the adenylation domain; and signature sequence for cofactor binding site in thiolation domain.

Table 1 Highly conserved core motifs of the adenylation domains of nonribosomal peptide synthases; M_a = Medium chain aliphatic amino acid (A, V, L, I and M); A_r = Aromatic amino acid (F, Y, H and W)

Core	**Consensus sequence (Marahiel *et al*., 1997)**	**Consensus sequence (Gulick *et al*., 2009)**
A1	L(TS)YxEL	M_a(ST)A_rx(EQ)M_a
A2	LKAGxAYL(VL)P(LI)D	(RKF)M_aGM_a
A3	LAYxxYTSG(ST)TGxPKG	$M_a M_a$X(ST)(STG)G(ST)TGxP
A4	FDxS	A_r
A5	NxYGPTE	A_r(GW)x(AT)E
A6	GELxIxGxG(VL)ARGYL	GEx (n = 10-14) GY
A7	Y(RK)TGDL	(ST)GD
A8	GRxDxQVKIRGxRIELGEIE	Rx(DK)x (n = 6) G
A9	LPxYM(IV)P	-
A10	NGK(VL)DR	PxxxxGKM_ax(RK)

Other signature sequences are also found in condensation domain (C1-C7), thioesterase domain, epimerization domain (E1-E7) and *N*-methylation domain (N1-N3), as shown in Figure 2 (Marahiel *et al*., 1997; Gulick *et al*., 2009). These amino acid sequences can be used for designing degenerative PCR primers (Tapi *et al*., 2010), e.g. A2f/A3r, A3F/A7R and MTF/MTR primers were specific for core A2/A3, A3/A7 and A2/A8 of adenylation domain, respectively. (Neilan *et al*., 1999; Martens *et al*., 2000; Sacido and Genilloud, 2004)

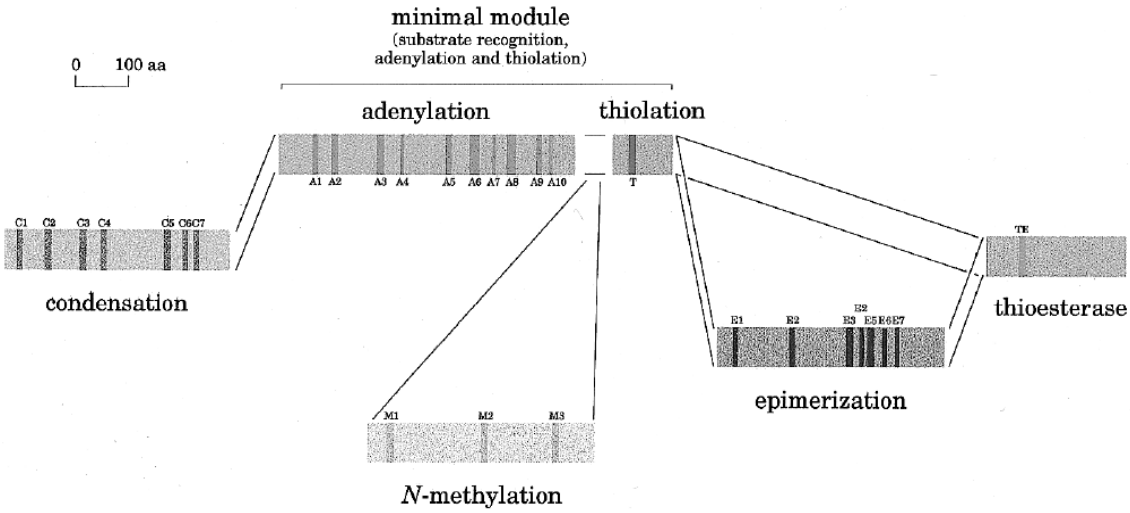

Figure 2: Core conserve region of NRPS (Marahiel *et al*., 1997)

NRPS consists of modules which are subdivided into domains as basic units. Each module is responsible for the incorporation of a specific monomer. Generally, each module composes of three core catalytic domains: adenylation (A) domains, thiolation (T) domains (peptidyl carrier protein: PCP) and condensation (C) domains. At C-terminus of NRPS enzyme, a thioesterase (Te) which catalyzes the release of the peptide from the NRPS is often found (Figure 3).

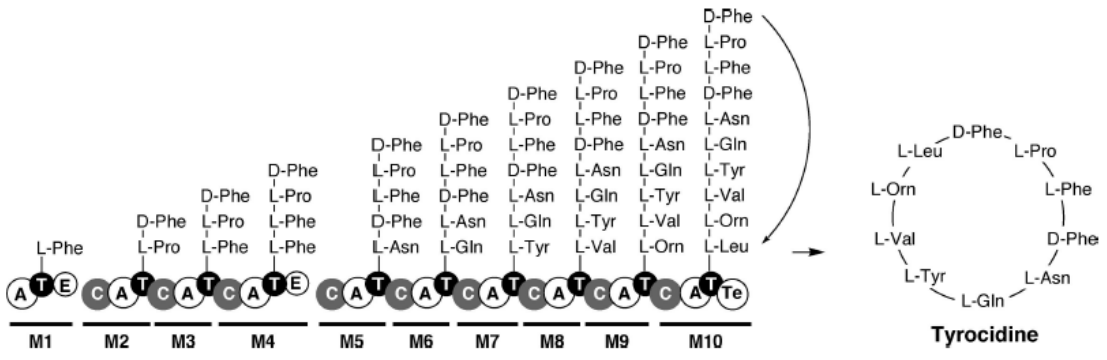

Figure 3: Modular structure of Tyrocidine synthsthase NRPS (Felnagele *et al*., 2007)

These domains form a minimal NRPS module which has classical A-T-[C-A-T]*n*-Te assembly (Schwarzer *et al*., 2003). Three types of NRPSs are classified based on domain architecture and assembly mechanisms (Mootz *et al*., 2002). Type A or linear NRPS has a classical arrangement of domain and module which leads to a parallel structure between the module and primary sequences of the peptide products. Type B or iterative NRPS also has a classical arrangement of module and domain similar to type A, but their modules and domains are re-used for synthesizing product with repeating units. Nonlinear or type C NRPS do not have classical arrangement, for example the yersiniabactin-synthetase consists of a Cy-Mt-T unit. NRPS may consist of a single A-T-C module (monomodular) or repeated A-T-C modules (multimodular). Some NRPS genes may encode an incomplete NRPS module but fused to a polyketide synthase (PKS) unit (Felnagele *et al*., 2007; Bushley and Turgeon, 2010).

The A domains belong to the adenylate-forming enzyme superfamily. They select, activate and incorporate the substrate into T domains of NRPS. Their substrate can be proteinogenic amino acids or nonproteinogenic amino acids as well as carboxy and hydroxyl acids, e.g. aryl acid (Jirakkakul *et al*., 2008). A domains have a high substrate specificity which is also seen in C and Te domains in the lesser extent. A domains activate the carboxyl group of the amino acid via ATP-dependent reaction to form aminoacyl-AMP intermediate. Aminoacyl intermediate is then transferred to 4'-phosphopantetheine of the thiolation domain and then covalently bound to thiol group of the 4'-phosphopantethein, which results in aminoacyl thioesters. Bounded aminoacyl thioesters then form peptide bond with aminoacyl thioesters or peptidyl thioesters which are bound with T domain of the adjacent module. Peptide bond formation is catalyzed by C-domain (Weber and Marahiel, 2001; Challis and Naismith, 2004).

Condensation (C) domain is in the acetyl coenzyme A dependent acetyl transferase superfamily. They contain two structural similar subdomains. Each subdomain contains distinct substrate-binding sites, donor site and acceptor site. While donor site binds nuclephilic aminoacyl thioesters from preceding module, acceptor site binds electrophilic aminoacyl thioester of the

corresponding T domain of the module. The acceptor site is more selective for the correct substrate than the donor ones. In several NRPSs, cyclization domain (Cy domain) may be found instead of C domain. The Cy domain has additional function that, after condensation, some amino acids, e.g. cysteine, serine or threoine are cyclized resulting in thiazoline or oxazaline ring. Besides, both C and Cy domains have high sequence similarity (Weber and Marahiel, 2001; Challis and Naismith, 2004; Roongsawang *et al*., 2005; Rausch *et al*., 2007).

The last module of NRPS, at the C-terminus, is typically a Te domain. Thioesterases belong to α,β-hydrolase family and are categorized into two types. Type I Thioesterases which are founded at the last domain of NRPS are responsible for the release of the peptide from enzyme complex by catalyzing the hydrolysis of the peptide bonds and intramolecular cyclization formed amide or ester bonds formation of nonribosomal peptide. The release of the peptide from enzyme, either by hydrolysis or cyclization, is a result of ester bond formation between terminal hydroxyl group of the nonribosomal peptides and serine residue of the Te domain, resulting in ester-linked intermediate. Substrate specificity of Te domains that responsible for hydrolysis and cyclization is different. Specificity of those which catalyze the cyclization is flexible while specificity of Te domains which hydrolyze linear peptides from an NRPS is unknown. On the other hand, type II thioesterases which often stand alone but are encoded within *nrps* gene clusters may play an important role in removing the inactivated acetyl groups of incorporated amino acid from the 4'-phosphopantetheine thiol of thiolation domain (Weber and Marahiel, 2001; Challis and Naismith, 2004).

After the release of nonribosomal peptide from Te domain or during the elongation step, nonribosomal peptide structure can be modified by tailoring enzymes. Glycosyltransferases add aglycones with particular deoxysugars after peptide is released while acyl carrier protein grows the acyl chain during chain elongation step. Tailoring enzymes which modify peptide during synthesis may act as separate subunits (in trans) or in relevant module (in cis). Examples of modifying domains which catalyze substrate during NRPS synthesis are C-methyltransferase (C-

MT domain), epimerization domain (responsible for transformation of L-amino acid into D-amino acid), cyclization domain (responsible for heterocyclization of cysteine and serine/threonine residues to thiazoline and oxazoline, respectively) and cytochrome P450s (responsible for oxidative cyclization of phenolic sidechains). Additional non-NRPS tailoring enzyme may modify either substrate or final peptide product by aminoacyl β-hydroxylation, halogenations and glycosylation (adding of sugar, glycone, to aglycone moiety of the molecule) (Walsh *et al*., 2001; Cadel-Six *et al*., 2008; Bushley and Turgeon, 2010).

Nonribosomal peptides differ from ribosomally synthesized peptides in several features. First of all, they can be linear like ribosomal peptides, but also branched or cyclic (Caboche *et al*., 2008). Second, NRPSs are able to incorporate non-proteinogenic amino acids, e.g. ornithine, hydroxyphenylor dihydroxyphenyl-glycine and (4*R*)-4-[(*E*)-2-butenyl]-4-methyl-L-threonine (Bmt) into nonribosomal peptide while ribosome are not. Besides, bond that connects amino acids in nonribosomal peptides can be pepide bond or ester bond. Next, nonribosomal peptides are usually heterocyclic or macrocyclic, branch or unbranch, and represent dimers or trimers of identical structural elements. In addition, these peptides may obtain side chain via *N*-methylations, *N*-formylations and glycosylations. Nonribosomal peptide may contain acetate or propionate units which are inserted by polyketide synthase and sometimes fatty acids inserted by fatty acid synthase (Schwarzer *et al*., 2003; Challis and Naismith, 2004). Finally, structure of nonribosomal peptide is diverse because of the differences in the selection of amino acids activated by A domains, modifying domains, for example E (epimerization) domains, cyclization and modifying enzymes, e.g. glycosyltransferases, carbamoyltransferases, and oxidases, in different NRPS complexs.

Various types of clinically significant pharmaceutical products are derived from nonribosomal peptides. (Felnagele *et al*., 2007) The most important example is β-lactam antibiotics, e.g. penicillins and cephalosporins, which are produced by a variety genera of fungi (*Penicillium, Cephalosporium* and *Aspergillus*) and bacteria (*Streptomyces, Nocardia,*

Flavobacterium and *Lysobacter*). Their mechanism of action is inactivation of the transpeptidation reaction in cell wall synthesis. β-lactam antibiotics biosynthesis pathway begin with the formation of tripeptide ACV from L-α-aminoadipate, L-Cysteine and D-Valine which is catalyzed by an NRPS designated as ACV synthetase. ACV is then modified by isopenicillin N synthetase into isopenicillin N which contains β-lactam and thiazolidine rings, the shared structure of β-lactam antibiotics. The next example is glycopeptides, e.g. balhimycin, bleomycins, chloroeremomycin, vancomycin and teichoplanin. Glycopeptides is heptapeptides synthesized by actinomycete from various species including *Actinoplanes* and *Streptomyces*. Vancomycin and teichoplanin are produced by *Amycolatopsis orientalis* and *Actinoplanes teichomyceticus*, respectively. Their mechanism of action is inhibition of cell wall synthesis of gram-positive bacteria by hydrogen bonding with D-Alanyl-D-Alanine of peptidoglycan precursor. This inhibits the transglycosylation of peptidoglycan for further elongation and cross-linking. Thus, peptidoglycan of the cell wall is weakened and cell wall is more susceptible to lysis. Vancomycin and teichoplanin is considered as clinically useful pharmaceuticals as the USFDA approved of vancomycin for the treatment of patients with infections caused by staphylococcal and streptococcal species. However while teichoplanin is not approved by the USFDA, it is used in Europe for the treatment of gram-positive infection. Cyclosporins, a group of 11 amino acid cyclic peptide produced by *Tolypocladium inflatum*, are also the best representative of nonribosomal peptide drugs. Cyclosporin A is the most medically useful because it has immunosuppression (due to T-cell suppression) and anti-inflammation activities. Nevertheless, it has high toxicity. Cyclosporin A is used clinically as an immunosuppressant in prevention of organ rejection after allogenic organ transplantation and in treatment of autoimmune disease. The other examples of nonribosomal peptide natural products are bacitracin, capreomycin, daptomycin lipopeptides, polymycin and quinoxalines (Kleinkauf, and Dohren, 1990; Felnagele *et al*., 2007).

2.5 Metagenomics

Jo Handelsman and colleagues first used the term metagenomics in 1998 (Handelsman *et al*., 1998) to term the collective genomes of soil microflora. Metagenomics refers to a set of research techniques that consist of metagenomics cloning and clones screening. The definition now has been updated and is “the science of discovering, modeling, understanding and ultimately managing at the molecular level the dynamic relationships between the molecules that define living communities and the biosphere” (Committee on Metagenomics, 2007). For a less complicated version, metagenomics means “the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species" (Chen and Pachter, 2005).

In the past, microbiological researches rely on cultivation method. Microorganisms including bacteria, archaea, eukarya and viruses in the environmental samples, e.g. soil, water and plants are isolated into a single colony. This approach cannot study the unculturable population which is estimated to be 99% of microbial population in the environment (Staley *et al*., 1985; Amann *et al*., 1995; Stein *et al.*, 1996; Handelsman *et al.*, 1998; Hallam *et al.*, 2003). This unculturable population is also largely diverse and unrelated to the cultured ones (Riesenfeld *et al*., 2004). Most of microbes in environmental sample cannot grow in selected culture media because of many reasons including inappropriate growth conditions, e.g. inoculums size, temperature, pressure, atmosphere, surface area, media type and incubation period, and microorganism factors, e.g. growth rate, symbionts and toxic products (Simu *et al*., 2004; Davis *et al*., 2005). These obstacles retard the development of cultivation method and raise the culture-independent approach as an alternative. As metagenomics is the culture bypassing method which microbial genome can be expressed, it may be the solution to this problem.

Metagenomics comprises four major processes. First, Metagenomes (collective genomes from environment) are extracted from sample of interest and are purified. Second, Metagenomic

DNA is cloned into a suitable cloning vector and transformed into host strains, resulting in metagenomic library. Third, metagenomic library is screened by either sequencing approach or functional approach for required phenotype. Finally, selected clones are collected for additional study (Streit and Schmitz, 2004). Types of samples collected depend on the objectives of the study. Suitable samples for novel bioactive compound screening are soils (Handelsman *et al*., 1998). It has been reported that marine water, mine drainage or animal samples are efficient samples for microbial communities and diversity studies, symbiosis researches and natural products research (Handelsman, 2004). Metagenomic DNA is extracted from a sample by direct or indirect method. Indirect method or cell extraction method is the method that isolates active microbial cells from sample and then extracted for metagenomic DNA. Although this method gives higher purity and diversity of genomic DNA, lower DNA yields limits its use (Steffan *et al*. 1988; Gabor *et al*., 2003). Direct DNA extraction or cell lysis method (Steffan *et al*. 1988; Picard *et al*., 1992; Ogram *et al*., 1997) is performed by breaking microbial cell directly and extracting DNA from the sample. Cell disruption can be done in various lytic treatment including mechanical forces, for example bead-mill homogenization, bead-beating, sonication, heating or thermal shock, and chemicals, e.g. cetyl trimethylammonium bromide (CTAB), proteinase K and sodium dodecyl sulfate (SDS) (Xia *et al*., 2006). Chemical or enzymatic lysis is relatively gentler than mechanical disruption methods (Rajendhran and Gunasekaran, 2008). This method yields much higher DNA quantity compare to indirect method. Though this procedure gives less DNA purity, the contaminants do not cause problems since they can be eliminated in the additional step of purification. Moreover, eukaryotic genome, which is co-extracted, is generally not expressed in bacterial host organisms. Purification processes are needed for both extraction methods but more extensive for direct extraction. At least four types of purification methods are commonly used: cesium chloride density gradient centrifugation, chromatography, electrophoresis and dialysis, and filtration. To remove all contaminants, several purification methods should be combined depending on the type of an environmental sample (Roose-Amsaleg *et al*., 2001). Extracted DNA

can be quantified and qualified by UV visible spectrophotometer using optical density (OD) at specific wavelength. OD at 230 nanometer (nm) indicates the amount of salt, solvent and humic acid contaminant. OD at 260 and 280 nm indicate the amount of DNA and protein impurity, respectively. Double strands DNA with concentration of 50 microgram per milliliter have an OD_{260} equivalent to one. DNAs which are suitable for downstream process should have OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios in the ranges of 1.8-1.9 and 1.4-1.9, respectively (Wilfinger *et al*., 1997; Roose-Amsaleg *et al*., 2001).

Metagenomic library construction is performed after the extraction and purification steps. Purified metagenomic DNA is cut by restriction enzyme and then inserted into vector. Recombinant vector is then transformed into competent host cell. Type of vector used in ligation reaction depends on metagenomic DNA size. A small sized insert DNA can be easily transformed without the need of very high DNA purity. However, it cannot contain large gene clusters or operons. Furthermore, a large number of clones are needed for library coverage. On the contrary, large DNA is suitable for a study of a large gene cluster or microbial genomes, but it is difficult to obtain and process (Streit and Schmitz, 2004; Daniel, 2005). Each vector type delivers DNA-insert which has different sizes. Small sized DNA [less than 10 kilo base pairs (kb)] is generally delivered by plasmid (Henne *et al*., 1999) while larger sized insert is usually delivered by cosmid, fosmid or bacterial artificial chromosome (BAC). Cosmid delivers insert which has length ranging from 25 to 35 kb (Entcheva *et al*., 2001). DNA with size approximately of 40 kb is usually delivered by fosmid (Beja *et al*., 2002). BAC is usually used with larger sized insert (200 kb) (Beja *et al*., 2000; Rondon *et al*., 2000). Plasmids have higher copy numbers than the other vectors. This property yields advantages for the detection of weakly-expressed foreign gene. Many transformation methods including chemical or mechanical transformation are used. Electroporation is the most popular method due to its ease, rapidity, high efficacy and reproducibility (Sambrook and Russell, 2001). For host cells used in metagenomic library construction, *Escherichia coli* are preferred because they are commonly employed in downstream

process and industrial fermentations (Daniel, 2004; de Lorenzo, 2005). Other hosts used include *Aspergillus* (Lubertozzi and Keasling, 2008), *Pseudomonas putida* (Li and Qin, 2005) and *Streptomyces lividans.* (Hopwood *et al*., 1985). *Streptomyces* host should be encouraged for use in drug discovery purpose because it is the well known antibiotics producer (Rajendhran and Gunasekaran, 2008).

Metagenomic libraries are screened by sequence–based analysis or functional screening (Schloss and Handelsman, 2003; Daniel, 2004). These two approaches have different advantages and disadvantages. Sequence-based screening is the screening of libraries for clones that contain required sequences by hybridization probes or PCR primers designed from conserved DNA sequences of already known genes or protein families, for example 16S rRNA gene. This approach relies on known conserved DNA sequences. It is also an expression-independent approach which allows the detection of unexpressed gene or incomplete gene fragment. In other word, it can be said that sequence based screening is not suitable for a full length gene. On the contrary, functional screening is expression-dependent. This approach begins with identification of clones that express phenotype of interest followed by characterization of the active clones by sequence-based method or biochemical analysis. Functional screening allows the detection of functional genes which express the functional gene products. It also has the potential to detect novel genes encoding new types or classes of bioactive compounds since it does not depend on known conserve sequence. However, expression problems of foreign gene in selected host cell may limit the detection ability of this method. Different codon usage, transcriptional co-factor, protein folding and protein secretion could be a reason for poor protein expression and low activities of the transformant (Streit and Schmitz, 2004).

Metagenomics contribute to various potential applications in many fields which microorganism involved (Committee on Metagenomics, 2007). In earth sciences, genome-based microbial ecosystem models may describe and predict global environment process. Metagenomics also facilitate in community-based microbial biology, ecology and evolution.

Microbial diversity in selected habitat can be accessed more easily via sequence-based analysis by using 16S rRNA gene. Metagenomics also assist in understanding of microorganisms' activities in energy production or waste remediation. The role of microorganisms in health of plants and animals may involve in agriculture. The better understanding about symbiosis and pathogenesis of microbials in domestics lead to more productions and loss of less animals and plants. Novel natural products, for example enzymes and bioactive compounds which may generate benefits in foods, cosmetics and pharmaceutical industry, are expected to be discovered from metagenomics. Metagenomics obviously have pharmaceutical application in novel drugs discovery. Functional screens of metagenomic libraries have identified both novel and previously described natural products, for example biotin, enzymes, antibiotics and biosynthetic pathways (Daneal, 2003; Schloss and Handelsman, 2003; Kennedy *et al*., 2007). Enzymes, e.g. agarase, alcohol oxidoreductase, amidase, amylase, β-galactosidase (Wang *et al*., 2010), chitinase, DNase, glycerol/diol dehydratase, 4-hydroxybutyrate dehydrogenase, lipase (Liaw, *et al*., 2010) protease and xylanase (Mo *et al*., 2010) were discovered via metagenomics. Antibiotics, e.g. ascidiacyclamide, discodermolide (Dunlap *et al*., 2007), indirubin (Osburne *et al*., 2000), N-acyltyrosine, N-acylaromatic long chain amino acid antibiotis, onnamide (Haygood and Davidson, 1997; Piel *et al*., 2004), terragines (Wang *et al*., 2000), turbomycins (Gillespie *et al*., 2002), and violacein (Brady *et al*., 2001), and anticancers, for example bryostatins (Haygood and Davidson, 1997; Davidson *et al*., 2001; Hildebrand *et al*., 2004; Dunlap *et al*., 2007) and patellamides (Bergmann and Feeney, 1950; Bergmann and Burke**,** 1995; De Rosa *et al*., 1995; Schmidt *et al*., 2005) which have potential in drug development were also found. Besides, in metagenomic libraries screening, biosynthetics pathway, e.g. biotin synthetic pathway, polyketide synthase and nonribosomal peptide synthetase gene cluster were discovered.

Although metagenomics may solve the cultivation problems, they still have many limitations (Committee on Metagenomics, 2007; Dupr'e and O'Malley, 2007). First of all, metagenomics DNA from some types of sample, especially from soil, require a large number of

clones in metagenomic library to cover the entire metagenomic DNA. Second, technology limitations are considered to be important. Current technologies, e.g. sequencing technology, bioinformatics and metagenomic database, cannot efficiently deal with large amount of complex data derived from metagenomes. The sophisticated study, for example comparison of bacterial communities or microbial system biology, needs the use of high technology. Metagenomic technique has various in-process technical biases, e.g. sampling, lysing cells for DNA extraction, cloning and expression systems. Furthermore, metagenomes usually contain DNA of major population. The information about minor members of communities with important role is still scarce. The techniques used to study rare community members, for example cell sorting or community DNA normalizations are still in the development process. Another problem in metagenomic research is an inadequate of genomic data. Some sequences in database do not have functional data which lead to inadequate number of reference genome to identify microbial functions. Some data, for example physical conditions is difficult to associate with functional data. Finally, the use of expression host cell in functional screening is usually restricted to *E. coli*. This limits the discovery of genes that cannot be expressed in *E.coli* and raises the need of novel gene-expression system.

2.6 **Soil metagenomics** (Daniel, 2005; Rajendhran and Gunasekaran, 2008)

Soil is one of the best microbial niches for metagenomic research due to the quantity and diversity of uncultured microbial population. Soil has many significantly different characters from other samples; therefore procedure of soil metagenomics differs from the others. Storage of soil samples in any stages: on site collection, shipping, and long term storage is critical because it has a strong influence on experimental results. Although different storage methods affect microbial properties of different soil types in different levels, freezing is proved to be the best storage method comparing with air-drying for all soil types (Wallenius *et al*., 2010). For some instances,

soil microbial enrichment can be done in order to increase the discovering frequency of desired population. The most critical process in soil metagenomic research is DNA extraction because it affects both the species abundance and composition of the bacterial community (Laurent-Martin *et al*., 2001). Soil DNA extraction, either by direct or indirect process, is difficult because soil microbe is firmly trapped by soil matrix. Clay component in soil attaches well to microbial cell wall. Soil with higher amount of clay is harder to be extracted but it appears to have less impurity. In other words, clay content of the soils is positively correlated with the purity of soil DNA but is negatively correlated with DNA yielded from extraction (Bakken, 1985). Yield and diversity of soil DNA is also affected by the extraction method used because different soil microorganisms have different susceptibilities to different cell lysis procedure. Chemical or enzymatic lysis is selective to particular cell types and penetrate soil matrix partially while mechanical disruption homogeneously disperses soil or sediment samples in lysis buffer. Therefore, mechanical treatment is more effective and less selective than chemical lysis. However, DNA shearing is the major disadvantage of mechanical treatment (Rajendhran and Gunasekaran, 2008). Different extraction procedures result in different yields of DNA. The amounts of DNA isolated from different soil types per a gram of soil range from less than one microgram to approximately 500 micrograms (Daniel, 2005). Direct lysis is preferred because it is suitable for various soil textures, less time consumption and better DNA yield (Roose-Amsaleg *et al*., 2001; Daniel, 2004). Another major problem of soil extraction is humic substances which are always co-extracted because they have physicochemical properties similar to that of nucleic acids. The impurities from humic compounds can interfere enzyme kinetics, e.g. restriction-enzyme and *Taq* polymerase. They also affect other downstream processes, e.g. transformation, cloning and DNA hybridization. To remove humic substances, purification steps often require PVP (polyvinylpolypyrrolidone) (Roose-Amsaleg *et al*., 2001). DNA from soil extraction should be precipitated using 5% polyethylene glycol instead of absolute ethanol or isopropanol in order to remove the humic impurity. Soil DNA measurement should be done with densitometric analysis

of ethidium bromide stained agarose gel instead of spectrophotometric because OD_{260} indicates the levels of humic substances rather than the DNA (Arbeli and Fuentes, 2007). In general, purity of soil genomic DNA can be evaluated by absorbance ratio at 260/230 nanometers (DNA/humic acid) and 260/280 nanometers (DNA/protein) (Roose-Amsaleg *et al*., 2001). Absorbance at higher wavelengths (320 or 340 nanometers) have been reported that they can be used to measure the level of humic acid and give the optical density that is independent from the OD of DNA and protein contents (Rajendhran and Gunasekaran, 2008). Soil DNA which is still contaminated with humic or matrix substances or sheared during purification step might be used to construct plasmid libraries but better to be discard. Another critical step is library construction. Soil metagenomic library needs to have a large size of inserts and high number of clones to cover the enormous amount and diversity of soil microbe. Approximately, more than 10^7 plasmid clones (5 kb inserts) or 10^6 BAC clones (100 kb inserts) are required to represent the genomes of all the different prokaryotic species presented in one gram of soil (Handelsman *et al*., 1998). Many soil metagenomic researches for novel drug discovery use fosmid, cosmid or BAC as a vector since they allow the detection of a large gene cluster, e.g. polyketide synthases. For the screening of soil metagenomics, high-throughput and sensitive screening methods are required. PCR is mostly used for sequence-based screening of soil-based library and often applied for identification of 16S rRNA gene in phylogenetic study. Genes encoding enzymes which contain highly conserved domains, e.g. polyketide synthases, gluconic acid reductases and nitrile hydratases are compatible with PCR based method. In most functional screening, *E. coli* has successfully been used as the expression host (Daniel, 2005).

Soil metagenomics have various applications, e.g. identification of functional genes, investigation of the microbial diversity and community dynamics, and assembly of complete genome of an uncultured organism. Using functional screening, many novel compounds with potential application in biotechnological and pharmaceutical industry were found. Mangrove soil metagenomics reveal the existence of new lipase subfamily (Couto *et al*., 2010). From soil

metagenomics, enzymes, for example 1,4-α-glucan branching enzyme agarase, 4-hydroxybutyrate dehydrogenase, alcohol oxidoreductase, agarase (which usually found in marine microbe), amidase, amylase, cellulases, glycerol/diol dehydratase, lipase, pectate lyases and protease, were discovered. Moreover, biosynthesis pathway of vitamin or its precursor, e.g. ascorbic acid, biotin and nitrile hydratases (nicotinamide precursor) was found. Clones with antibiotics activites, for example indigo blue, indirubin, N-acylaromatic long chain amino acid, terragine, turbomycin and violacein were also detected from this approach (Voget *et al*., 2003; Daniel, 2004; Lim *et al*., 2005; Schmeisser *et al*., 2007). Almost all of antibiotics were found form *E. coli*-constructed metagenomic library except terragine, which were in streptomyces host (Wang *et al*., 2000).

CHAPTER III

MATERIALS AND METHODS

3.1 Materials

3.1.1 Culture media and antibiotics

Luria Bertani (LB) agar (Difco) supplemented with ampicillin (T.P. Drug laboratories) (100 μg/ml) was used for the cultivation of *Escherichia coli* DH5α carrying pGEM®-T Easy vector. For blue-white selection of transformants that contain pGEM®-T Easy vector, LB agar containing ampicillin, 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) was used. Culture media for the cultivation of *E.coli* DH5α which was used for metagenomic library construction (section 3.6) was LB agar supplemented with chloramphenicol (Sigma) (12.5 μg/ml). For an induction of fosmid production, *E.coli* DH5α was cultured on LB agar containing chloramphenicol and arabinose (Sigma) (10 mg/ml). Sabouraud's dextrose agar (SDA) (Merck) was used for cultivation of *Candida albicans* ATCC 90028 in the antifungal assay. Ketoconazole (USP24; Karingo, Italy) was used in antifungal test. Muller-Hinton (MH) agar (Merck) was used for cultivation of *Bacillus subtilis* in the antibacterial assay.

3.1.2 Chemicals

Chemicals used in this study were as the followed: for soil DNA preparation, lysis buffer which composes of tris (hydroxymethyl) aminomethane (Tris) (BioScience Inc),

ethylenediaminetetraacetic acid (EDTA) (BioScience Inc), cetyl trimethylammonium bromide (CTAB) (BioScience Inc) and sodium chloride (Merck); sodium acetate (Merck), absolute ethanol (Merck), isopropanol (Merck), was used. Chemicals for agarose gel electrophoresis were TBE buffer which consists of Tris, boric acid (BioScience Inc) and EDTA; agarose (Vivantis), polyvinyl pyrrolidone (PVP) (Applichem). SYBR® (invitrogen) was used instead of ethidium bromide for DNA staining in purification process. For polymerase chain reaction (PCR), GoTaq® Colorless Master Mix (Promega) was used throughout the study. All of PCR primers in this research were synthesized by 1st Base, Singapore. Glycerol (Fisher Scientific) was used as cryoprotectant for library storage and used in electrocompetent cell preparation. HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (Applichem) was used in preparation of electrocompetent *B. subtilis*.

3.1.3 Plastic wares and extraction kits

Plastic wares used in this study were as the following: 50 ml centrifuge tubes (Corning), centrifuge tubes 30 ml (Nalgene®), microtubes 1.5 ml (Axygen), disposable plastic petri dish (Hycon). For DNA purification by dialysis, SnakeSkin® pleated dialysis tubing (Thermo Scientific) was used. For DNA extractions, High-Speed Plasmid Minikit (Geneaid) were used for plasmid extraction and FosmidMax® (Epicenter) was used for fosmid extraction. Glass beads (undrilled, 3 mm; Ajax Finechem Pty Ltd) were used to spread bacteria suspension on agar plate.

3.1.4 Microorganisms, host cells and cloning vectors

For screening of *nrps* gene on mangrove soil, *E. coli* DH5α was used as a host cell and pGEM®-T Easy vector (Promega) was used as a cloning vector. For construction of soil

metagenomic library, host cell and cloning vector was acquired from CopyControl™ Fosmid Library Production Kit with pCC2FOS™ vector (Epicenter). For sequence-based library screening, *Micromonospora chalcea* ATCC 12452 was used as a positive control in PCR which was used for *nrps* gene amplification. pSuperBAC in *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*.

3.1.5 Tested strains

For functional-based screening, *Candida albicans* ATCC 90028 was used in the antifungal assay. The microorganism was maintained on SDA and stored at 4°C. For the screening of antibacterial activity, chloramphenicol-resistant *Bacillus subtilis* was used. *B. subtilis* Cm^R was maintained on LB agar containing chloramphenicol (12.5 μg/ml) and stored at 4°C.

3.1.6 DNA marker

Different types of DNA markers were used depending on the length of DNA loaded. Lambda DNA *Hin*dIII digest marker (Fermentas), GeneRuler DNA marker mix (Fermentas) and VC100bp Plus DNA Ladder (ready to use) (Fermentas) were used in the experiment.

3.2 Instruments

For centrifugation of DNA, 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).

POWER PAK 300 (Bio-Rad, California, USA) was used for agarose gel electrophoresis. For optical density (OD) and % transmittance measurement, UV visible spectrophotometer (UV-160A UV-Visible recording spectrophotometer, SHIMADZU, Kyuto, Japan) was used. Automated thermal cycler (Mastercycler gradient, Eppendorf, Hamburg, Germany) was used for PCR. MicroPulser™ (Bio-Rad, California, USA) was used for electroporation. Colony replication was performed by replica plating tool (Scienceware, Pequannock, NJ, USA).

3.3 Sample collection

Soil sample was collected from Klongkone mangroves, Samutsongkhram province in September 2008. The soil sample were collected from surface to 10 cm depth and stored on ice during sample collection and transportation. For long term storage, soil samples were kept at -20°C.

3.4 Soil DNA preparation

DNAs of soil microorganisms were extracted by direct extraction method (Brady, 2007). After the extraction, soils DNAs were purified and size-selected by gel electrophoresis and dialysis (Brady, 2007). Purified soil DNAs were quantified and qualified by OD measurement at 230, 260 and 280 nm (Wilfinger *et al*, 1997). Ten reactions of direct soil extraction were performed in 50 ml sterile tubes. Twelve and a half grams of soil (total 125 grams of soil for ten reactions) and 15 ml of lysis buffer were added into each bottle which was then incubated at 70°C in water bath for 2 hours. Then, soil suspension was centrifuged at 5,000 rpm at 25°C in order to separate the supernatant and soil precipitate. Supernatant was put into 30 ml centrifuge tube. Next, 0.7 volume of cold isopropanol was added into soil supernatant to precipitate DNAs from the supernatant. DNA pellet was collected by centrifugation at 3,500 g for 30 minutes at 25°C.

Pellet was then washed by cold 70% ethanol. Washed DNA was then air-dried and dissolved in sterile distilled water.

Extracted DNA was purified by agarose gel electrophoresis. Agarose gel was prepared at 0.6% concentration in TBE buffer. PVP in a 2% final concentration was added into the gel (Young *et al*., 1993). Gel electrophoresis was performed in air-conditioning room using 30 voltage of electromotive force until visible brown band of humic substances move out of the agarose gel. Agarose gel was stained with SYBR gold and then excised under blue light, leaving only the required size (in this experiment, the DNAs longer than 23 kb were required). Excised agarose gel was placed into TBE buffer in dialysis bag. Agarose gel electrophoresis was performed in dialysis bag in order to extract size-selected DNA from agarose gel into TBE buffer. DNAs in TBE buffer was then precipitated by adding 0.1 volume of 3 molar sodium acetate and 0.7 volume of cold isopropanol. DNA pellet was washed by cold 70% ethanol, air-dried and dissolved in sterile distilled water. Purified soil DNAs was kept in the refrigerator for further process of the experiment.

Optical density of DNA at 230, 260 and 280 nm was measured by UV visible spectrophotometer in order to determine the quantity and purity of purified the DNA. OD_{230} indicates the amount of salt, solvent and protein contaminant in DNA solution while OD_{280} indicates the amount of protein contaminant. OD_{260} indicates an amount of DNA. Double strands DNA with concentration of 50 microgram per milliliter have an OD_{260} equivalent to one. Soil DNA which is suitable for further downstream process should have OD_{260}/OD_{280} and OD_{260}/OD_{230} ratios of 1.8-1.9 and 1.4-1.9, respectively.

3.5 Screening on mangrove soil metagenomes for *nrps* gene

3.5.1 PCR amplification of *nrps* gene from mangrove soil

Purified soil DNAs from section 3.4 were used as a DNA template for PCR amplification. *nrps* gene encoding A domain was amplified by using three primer pairs which are A2f (AAG GCN GGC GSB GCS TAY STG CC)/A3r (TTG GGB IKB CCG GTS GIN CCS GAG GTG) primers (Martens *et al.*, 2007); A3F (GCS TAC SYS ATS TAC ACS TCS GG)/ A7R (SAS GTC VCC SGT SCG GTA S) primers (Sacido and Genilloud, 2004) and MTF2 (GCN GGY GGY GCN TAY GTN CC)/MTR (CCN CGD ATY TTN ACY TG) primers (Neilan *et al*., 1999). PCR amplification was performed in a 20 µl reaction mixture consisting of GoTaq® Colorless Master Mix. The thermocycling program was performed using an automated thermal cycler and run as follows; For A2f/A3r primers, 95°C for 5 min; 40 cycles of 95°C for 1 min, 70°C for 1 min and 72°C for 2 min; with a final extension period of 7 min at 72°C; For A3F/A7R primers, 95°C for 5 min; 35 cycles of 95°C for 30 sec, 59°C for 2 min and 72°C for 4 min with a final extension period of 10 min at 72°C ; For MTF2/ MTR primers, 94°C for 5 min; 35 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 2 min; with a final extension period of 7 min at 72°C. DNA from *M. chalcea* ATCC 12452 was used as positive control for *nrps* gene amplification since its genomic DNA contains *nrps* gene cluster (Sacido and Genilloud, 2004).

3.5.2 Preparation of electrocompetent *E. coli* DH5α

Electrocompetent *E. coli* DH5α cells were prepared by chemical method using calcium chloride (Sambrook and Russell, 2001) with some modifications. Briefly, *E. coli* was inoculated into 5 ml of LB broth and incubated at 37°C with shaking at 200 rpm for overnight. Overnight culture in a volume of 300 µl was then inoculated into 30 ml of LB broth and

incubated at 37°C with shaking at 200 rpm to an OD_{600} of approximately 0.5. Cells were transferred into centrifuge bottle and spun down at 4°C for 10 min. Supernatant was discarded. Cell pellet was gently resuspended in 30 ml of ice-cold 10% glycerol. Cell suspension was spun down for the collection of bacterial cells. This process was repeated three times by changing the volume of 10% glycerol into 15, 3 and 2 ml, respectively. Cells were finally suspended in 2 ml of 10% glycerol and divided into 400 μl aliquot. Each aliquot was used for each electroporation reaction.

3.5.3 Sequencing of PCR products

PCR products with 200-300-bp, 700-bp and 1-kb in length which supposed to be the PCR product of *nrps* genes from A2f/A3r, A3F/A7R and MTF2/MTR, respectively, were ligated into pGEM®-T Easy vector according to the manufacturer's instruction and transformed into competent cells (*E. coli* DH5α). Transformation was performed by electroporation using MicroPulser™ according to the user instruction. Circular pGEM®-T Easy vector (extracted from blue colonies from prior experiments) was used as positive control for electroporation. Transformants were plated on LB agar containing ampicillin (100 μg/ml), X-Gal (40 μg/ml) and IPTG (40 μg/ml) for blue-white selection. After the overnight incubation at 37°C, white colonies were selected and screened for *nrps* gene by PCR amplification using A2f/A3r, A3F/A7R or MTF2/MTR primers as the condition described above. Recombinant plasmid of the PCR-positive clone was extracted using High-Speed Plasmid Minikit and directly subjected to sequencing by using T7 (TAATACGACTCACTATAGGG) and SP6 (TATTTAGGTGACACTATAG) primers at 1st BASE (Singapore).

3.5.4 Phylogenetic analysis of PCR products from mangrove soil

Nucleotide sequences of DNA insert in plasmid were trimmed off vector sequence and degenerate primer binding sites were identified and removed. Nucleotide sequences were translated into amino acid sequences using BioEdit Sequence Alignment Editor 7.0.5.3. (Hall, 1999). All of deduced amino acid sequences from mangrove soil were aligned by ClustalW2 (Larkin *et al*., 2007) serviced online by EBI (http://www.ebi.ac.uk/Tools/clustalw2/index.html) in order to generate the alignment score for sequence identity analysis.

With CLC Mainworkbench 5.6 (CLC bio A/S, Aarhus, Denmark), the obtained DNA sequence of partial *nrps* gene was used as the query sequence to search for similar sequences in non-redundant protein sequences database, reference proteins database and swissprot protein sequences database of NCBI using blastx program (Stephen *et al*., 1997). The redundant sequences from the same species were sorted out. Only the sequences with the highest score were selected and used for generating multiple sequences alignments by Muscle sequence alignment (Edgar, 2004). Aligned sequences were further adjusted manually for good alignment of core sequences. The unrooted phylogenetic tree was inferred using neighbor-joining algorithm and a bootstrap analysis of 1,000 replications.

3.5.5 Prediction of amino acid activated by A domain of *nrps* gene from mangrove soil

Amino acid substrate which is recognized by A domain of NRPS from mangrove soil was predicted by PKS/NRPS Analysis Web-site at http://nrps.igs.umaryland.edu/nrps/. (Bachmann and Ravel, 2009). The NRPS prediction BLAST server first extracted eight amino

acids from input A-domain sequences derived from mangrove soil. After that, Blast program searched the eight critical residues of the enquiry against a database of eight amino acids lining the active pocket of adenylation domains of assigned function.

3.6 Metagenomic library construction

Prepared soil DNA from section 3.4 was used for metagenomic library construction. CopyControl™ Fosmid Library Production Kit with pCC2FOS™ vector (Epicenter) was used for construction of the library according to user manual with some modifications. Briefly, purified DNAs from section 3.5 were end-repaired and then ligated into pCC2FOS™ vector. The process of "shearing of DNA" and "size selection of the End-repair DNA" were skipped because purified DNAs were already sheared and size-selected from the purification process. Recombinant fosmids were packed with MaxPlax Lambda Packaging Extracts. Packed fosmids were diluted with Phage Dilution Buffer in serial 10 fold dilution before adding to EPI300-T1® host cells in order to determine the titre of the packaged phage particles. Phages at the selected titer were mixed with EPI300-T1® cells for transfection. Transfected bacteria 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-μl microtube. Pooled transformants were labeled and kept in 20% glycerol. Metagenomic library was stored at -80°C for further analysis.

3.7 Metagenomic library screening

3.7.1 Functional-based screening

For functional-based approaches, antifungal (*C. albicans* ATCC 90028) and antibacterial (chloramphenicol-resistant *B. subtilis*) activities were tested.

3.7.1.1 Construction of chloramphenicol-resistant *B. subtilis*

Bacillus subtilis with chloramphenicol-resistant phenotype was constructed by introduction of pSuperBAC which contains chloramphenicol resistance gene (Williamson *et al*., 2005) into *B. subtilis* ATCC 6633 by electroporation technique. Preparation of *B. subtilis* electrocompetent cells was modified from the method described by Matsuno, Y., *et al*. (1992). Briefly, *B. subtilis* ATCC 6633 was inoculated in LB broth and cultured at 37°C with shaking at 200 rpm for overnight. Fresh overnight culture in a volume of 0.5 ml was inoculated into 10 ml of LB broth and incubated at 37°C with shaking at 200 rpm for 3 hours. Cells were harvested by centrifugation at 4°C. Cell pellet was washed twice in 1 mM HEPES and twice in 10% glycerol. Cells were finally suspended in 2 ml of 10% glycerol and divided into 400 µl aliquot. Each aliquot was used for each electroporation reaction. Transformation was performed by electroporation using MicroPulser™ according to the user instruction. LB broth in a volume of 1 ml was added into the electroporated bacterial suspension. The bacterial cells were incubated at 37°C for 1 h before spreaded on LB agar containing 12.5 µg/ml of choramphenicol. After incubation at 37°C for 24 h, the transformant was subcultured and maintained on LB agar containing chloramphenicol 12.5 µg/ml. This strain was designated as *B. subtilis* Cm^R and stored at 4°C.

3.7.1.2 Preparation of test microorganisms

B. subtilis Cm^R was cultured on LB agar containing 12.5 μg/ml of choramphenicol at 37°C overnight. Then, one colony was inoculated into tryptic soy broth and incubated at 37°C for 2-3 h.

C. albicans was grown on SDA at 37°C for 24 h and suspended in 0.85% sodium chloride solution.

The turbidity of microbial suspension was adjusted to 50% T at 580 nm. Bacterial inoculum was added in a final concentration of 1% to 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 Screening for antibacterial and antifungal activities

Each pool of metagenomic library was serially diluted in 0.85% sodium chloride solution to 10^{-6} dilution. A 100-μl diluted sample was spread on LB agar containing chloramphenicol (12.5 μg/ml) and arabinose (10 mg/ml) using sterile glass beads. This was done in 15 replicas. After incubation for 5 days at 25°C, a total of 5 and 10 plates were tested for antibacterial and antifungal activities, respectively. 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 magnifying glass and lamp. Clones with inhibition zone were observed again under stereomicroscope. The selected clones

were streak for isolation on LB agar containing chloramphenicol (12.5 μg/ml) and then tested with all of tested organisms for the confirmation of antimicrobial activities.

3.7.2 Sequence-based screening

PCR using MTF2/MTR primers were used for screening the *nrps* genes in metagenomic library. Each pool of metagenomic library was extracted for recombinant fosmid DNA using FosmidMax® according to the manufacturer's instruction. Extracted fosmid was then used as DNA template for PCR amplification using MTF2/MTR primers as described in section 3.5.1. Pool that gave a 1-kb PCR product which supposed to contain *nrps* gene cluster was selected for further analysis. PCR-positive pool was diluted into 10^{-6} dilution and spread on LB agar containing chloramphenicol (12.5 μg/ml) which, from now on, was called as a starter plates. The starter plates were incubated at 37°C overnight. Each of them was replicated once into one plate using replica plating tool on the next day. Replicated plates were incubated at 37°C overnight and then collected by washing with sterile distilled water. Wash water was boiled for 5 minutes. Then, boiled supernatant was collected and used as DNA template for colony PCR using MTF2/MTR primer pair with the condition as mentioned in section 3.5.1. Replicated plates that gave the PCR-positive result referred to the *nrps* gene-containing starter plates. Clones in starter plate that had the PCR-positive replicated plate were collected. Each colony in each starter plate was picked and streaked onto a single grid of fifty-grid LB plate containing chloramphenicol (12.5 μg/ml) and then incubated at 37°C overnight. Fifty-grid plates were replicated once into one plate using replica plating tool. The fifty-grid replicated plates were incubated in the condition as above. After the incubation, replicated plates were washed and boiled. Washed water of the replicated plates was used for colony PCR using MTF2/MTR primer. 50-grid LB plates that gave 1-kb PCR product may contain clones with *nrps* gene. Those plates were then replicated tree times onto one LB plate containing chloramphenicol (12.5 μg/ml) (used as a stock culture) and

two LB plates containing chloramphenicol (12.5 μg/ml) and arabinose (10 mg/ml) (plate A and B). Clones from each row of plate A and column of plate B were swept, resuspended in sterile water and pooled into microtube. Cell suspension from each row and column was boiled separately for 5 minutes. Cell supernate was then collected and used as DNA template for PCR using MTF2/MTR primers. Data of rows and columns with PCR-positive result were analyzed and used for identification of clones containing *nrps* gene. PCR was repeated to confirm the selected PCR-positive clones. PCR products from the positive clone were ligated into pGEM® -T Easy vector as mentioned in section 3.5.3. PCR-positive clones were extracted for plasmid DNAs which were directly subjected to sequencing by using MTF2/MTR primers at 1 st BASE (Singapore). Sequencing results were analyzed as mentioned in section 3.5.4 and 3.5.5. Clones that contained *nrps* gene were also functionally screened for biological activities as described in functional-based screening section.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Sample collection and soil DNA preparation

Mangrove soil sample used in the present research was collected from Klongkone mangrove, Samutsongkhram province. This mangrove forest is a suitable place for collecting the sample due to its abundance. In addition, there is no industrial plant or open aquaculture pond nearby (Paphavasit *et al*., 2002). Soil sample was collected from the surface to 10 centimeter depth. The sample was stored at 4°C during sample collection and transportation as described in the methodology, although it has been reported that phylogenetic diversity of soil bacterial community was not affected by temperature or length of storage (Lauber *et al*., 2010).

Direct extraction of mangrove soil was performed shortly after the collection since it has been reported that the enzymatic activities of soil greatly changed during the storage especially at the initial period (12 weeks) (Dadenko *et al*., 2009). It was noticeable that, in this research, soil with long term storage yielded low amount of DNA. Direct extraction of 125 grams of mangrove soil yielded enough crude soil DNA for the purification process (Figure 4A). Purification of extracted DNAs by agarose gel electrophoresis and dialysis with the use of PVP could yield good quality DNA with an OD_{260}/OD_{280} ratio of 1.83 and an OD_{260}/OD_{230} ratio of 2 which were higher than the acceptable ratio of 1.8 and 1.4, respectively (Wilfinger *et al*., 1997). This indicated that the purified DNAs were pure enough for downstream processes. Agarose gel electrophoresis of the purified DNAs demonstrated that they were approximately 23-kb and higher in length, as shown in Figure 4B, which were large enough to contain *nrps* gene cluster since several bacterial operons for the biosynthesis of the peptide antibiotics were only 18-45 kb

(Marahiel *et al*., 1997). This range of DNA length was compatible with CopyControl™ Fosmid Library Production Kit with pCC2FOS™ vector (Epicenter) without the need of further DNA shearing process as indicated in section 3.6 in methodology.

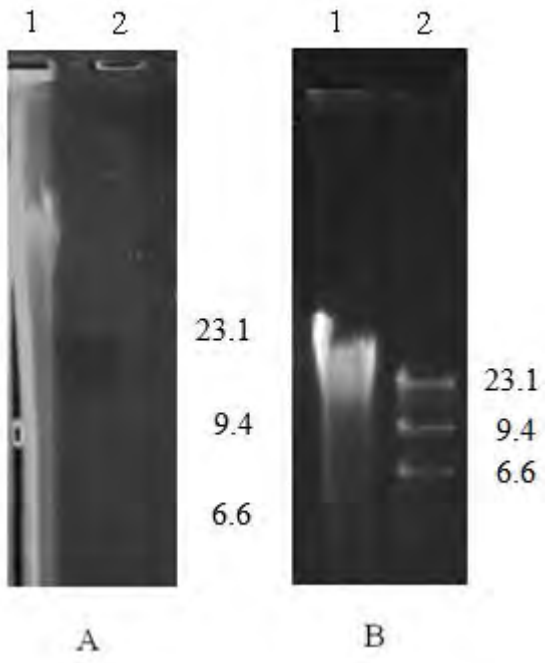

Figure 4. Agarose gel (0.6 %) electrophoresis of (A) crude DNA and (B) purified soil DNA. (A) Lanes: 1, Crude soil DNA.; 2, Lambda DNA *Hin*dIII digest marker in Kb length.; (B) Lanes: 1, Purified soil DNA; 2, Lambda DNA *Hin*dIII digest marker.

4.2 Screening on mangrove soil metagenome for *nrps* gene

4.2.1 PCR-amplified *nrps* gene from mangrove soil

Soil DNAs were screened for *nrps* gene by PCR amplification with three A-domain specific PCR primer pairs. A3F/A7R is the primer pair that is specific for *nrps* of soil actinomycetes (Sacido and Genilloud, 2004). A2f/A3r primer pair is specific for *nrps* of *Roseobacter* clade in proteobacteria (Martens *et al.*, 2007). MTF2/MTR is a specific primer pair of *nrps* of aquatic cyanobacteria (Neilan *et al*., 1999). Although A3F/A7R primer pair was successfully used in many literatures (Gonzalez *et al*., 2005; Gao and Huang, 2009; Qin *et al*., 2009; Chronakova *et al*., 2010), in this study, it could not amplify both soil DNA and DNA extracted from *M. chalcea* ATCC 12452 (as positive control) by the condition of PCR cycles and reagents indicated by Sacido and Genilloud, 2004. Problems with the use of A3F/A7R primer pair might arise from humic substances contaminated in DNA template. Humic substances inhibit DNA polymerase by chelating with Mg^{2+} in PCR. Inhibition concentration of humic substances could as low as 8 ng/µl. This inhibition problem could be overcome by increasing the Mg^{2+} concentration or further diluted the crude DNA template to 100-fold (Roose-Amsaleg *et al*., 2001). Although DMSO can enhance the PCR by interfering the secondary structure of the DNA (Chakrabarti and Schutt, 2001), it can also inhibit half of Taq polymerase activity (Hung *et al*., 1999). Therefore, final concentrations of $MgCl_2$ (1.5, 2.5, 3 and 4 mM), DNA template (1, 2 and 20 ng/µl) and DMSO (0 and 10% final concentration) were varied in order to optimize the PCR condition. However, these optimizations could not yield any PCR product. The use of GoTaq® Colorless Master Mix instead of indicated reagents still could not amplify either DNA samples or positive control. Gradient PCR was also performed in order to determine the optimum annealing temperature for the use of A3F/A7R with GoTaq® Colorless Master Mix. This attempt still could not yield any amplicons from this primer pair. Therefore, this primer pair was neglected.

A2f/A3r was the next primer pair that was used in this research. This primer pair could amplify PCR products with 200-300-bp in length from mangrove soil and genomic DNA of *M. chalcea* ATCC 12452 (Figure 5). To confirm *nrps* sequence in PCR products, they were ligated into pGEM® T Easy vector and transformed into *E. coli* DH5α as described in section 3.5.1 of the methodology section. This cloning process was unsuccessful. Only few white colonies grew on the selection media. Most of them gave 500-bp PCR product when performed colony PCR. One PCR-positive clone with approximately 300-bp PCR product was collected and designated as MA2_12. This clone was selected for further analysis of nucleotide sequence of DNA insert as described in section 3.5.3. Nucleotide sequence of DNA insert in plasmid of clone MA2_12 (Figure 6) was used as the query sequence to search for similar sequences in non-redundant protein sequences database of NCBI using blastx program. It was found that the

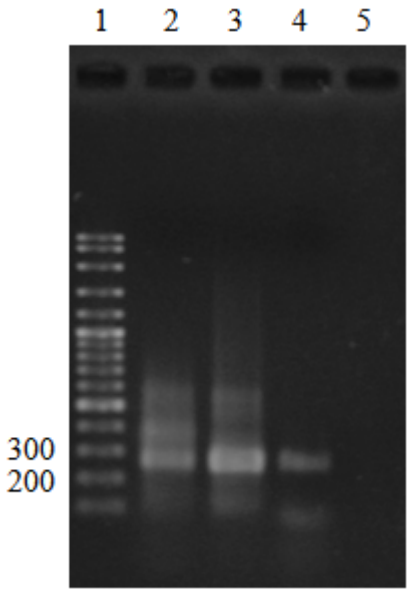

Figure 5. Agarose gel (1 %) electrophoresis of PCR products amplified with A2f/A3r primer pair, PCR products in range of 200-300 bp in length are shown. Lanes 1, VC100bp Plus DNA Ladder in bp length ; 2, PCR product from mangrove soil DNA; 3, PCR product from MA2_12; 4, PCR product of *Micromonospora chalcea* ATCC 12452 (as positive control); 5, water (as negative control).

DNA insert was not related to *nrps* gene. The highest sequence identity (95%) hits were protein of unknown function DUF255 of *Micromonospora* sp. L5 (ZP_06399683) and N-acylglucosamine 2-epimerase of *M. aurantiaca* ATCC 27029 (YP_003834076). With the conditions used in this study, A2f/A3r primers pair could not yield *nrps* related PCR product. Therefore, it was also neglected.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
GCGGGCGGTG CTTACGTTCC GGTGCTGATC TCGGTCGGTT ACGCGGCCTG CCACTGGTGT
CATGTCATGG CTCACGAGTC GTTCGAGAAC GAGGCAGTGG CCCGGCTGAT GAACGACGAC
TTCGTCTGCG TGAAGGTCGA CCGCGAGGAG CGCCCCGACG TTGACGCGGT CTACATGACC
GCCGCCCAGG CGATGACCGG GCAGGGCGGC TGGCCGATGA CCGTCTTCGC GACGCCGGAC
GGCACCCCGT TCTTCTGCGG CACATAAGCA CCACCTGCA
```

Figure 6. Nucleotide sequence of DNA insert from clone MA2_12

The last primers pair, MTF2/MTR, was used in many research for amplification of adenylation domain of cyanobacterial *nrps*, e.g. Ehrenreich *et al.*, 2005; Vizcaino *et al*., 2005; Tooming-Klunderud *et al*., 2007; Barrios-Llerena, *et al*., 2007; Pearson and Neilan, 2008; Zhao *et al*., 2008; Castle and Rodgers, 2009; Sipari *et al*., 2010. It was also used successfully for PCR in this study. MTF2/MTR primer pair was used for mangrove soil DNA amplification as described in section 3.5.1. MTF2/MTR primer pair could amplify PCR products with one-kb in length (Figure 7A) corresponding to the size of *nrps* gene encoding A domain (Neilan *et al*., 1999). To confirm that the PCR products were really amplified from *nrps* gene, they were subsequently ligated into pGEM®-T Easy vector and were transformed into competent cells. White colony transformants were collected and screened by PCR amplification which aimed for clones containing one-kb insert. Eight PCR-positive clones were found, as shown in Figure 7B. They

were designated as SM2_2, SM3, SM20, SM23, SM27, SM48, SM50 and SM51. Plasmids in these positive clones were extracted and nucleotide sequences of DNA inserts were analyzed as described in section 3.5.3. Six from eight clones were found to contain DNA insert which were related to *nrps* gene. The nucleotide sequences of DNA inserts in clone 27 and clone 50 were found to have 100% identity. The nucleotide and deduced amino acid sequences of the partial peptide synthetase gene of these clones, as shown in Figures 8-12, were submitted to Genbank. Their accession numbers were

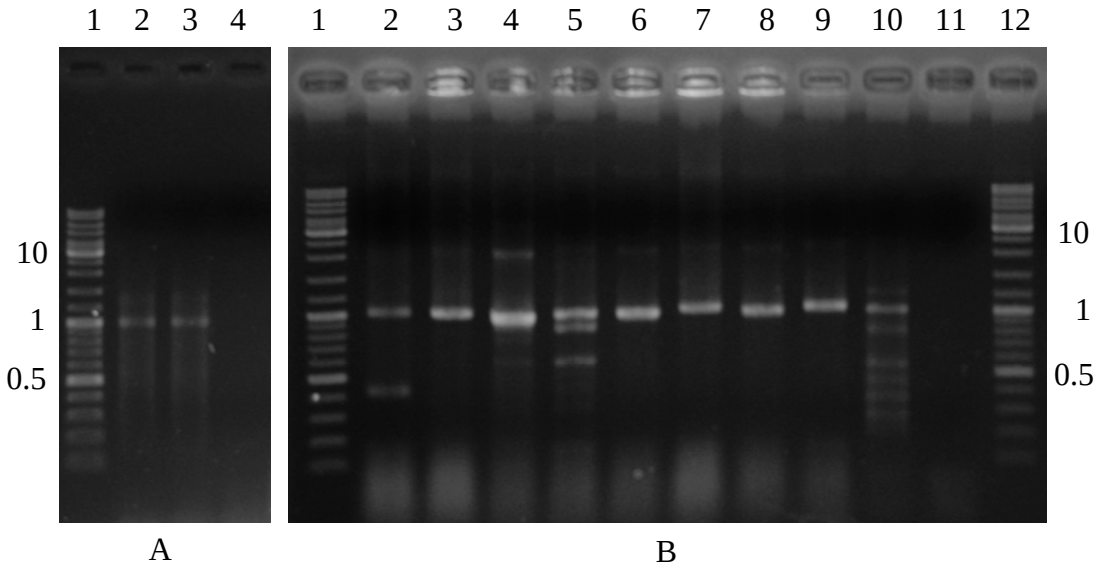

Figure 7. Agarose gel (1%) electrophoresis of PCR product from mangrove soil amplified with MTF2/MTR primer pair, PCR products with 1-kb in length are shown. (**A)** Lanes 1, GeneRuler DNA marker mix in Kb length; 2, PCR product from mangrove soil DNA; 3, PCR product of *Micromonospora chalcea* ATCC 12452 (as positive control); 4, water (as negative control). (**B)** Lanes 1 and 12, GeneRuler DNA marker mix in Kb length; 2, PCR product from SM2_2; 3, PCR product from SM3; 4, PCR product from SM20; 5, PCR product from SM23; 6, PCR product from SM27; 7, PCR product from SM48; 8, PCR product from SM50; 9, PCR product from SM51; 10, PCR product of *Micromonospora chalcea* ATCC 12452 (as positive control); 11, water (as negative control)

shown in Table 2. They consisted of partial A2 (P(LI)D), complete A3-A7 (A3, LAYxxYTSG(ST)TGxPKG; A4, FDxS; A5, NxYGPTE; A6, GELxIxGxG(VL)ARGYL; A7, Y(RK)TGDL) and partial A8 (GRxDx) conserved core sequences of adenylation domain of NRPS as described by Marahiel *et al*, (1997).

Table 2 Summary data of clones containing 1-kb PCR products amplified from mangrove soil

Clone	**Length of nucleotide sequence (bp)**	**Accession number**	**Figure**
SM2_2	964	HM592295	8
SM3	961	HQ290323	9
SM23	961	HQ290324	10
SM27	958	HQ286565	11
SM48	1000	HQ286566	12

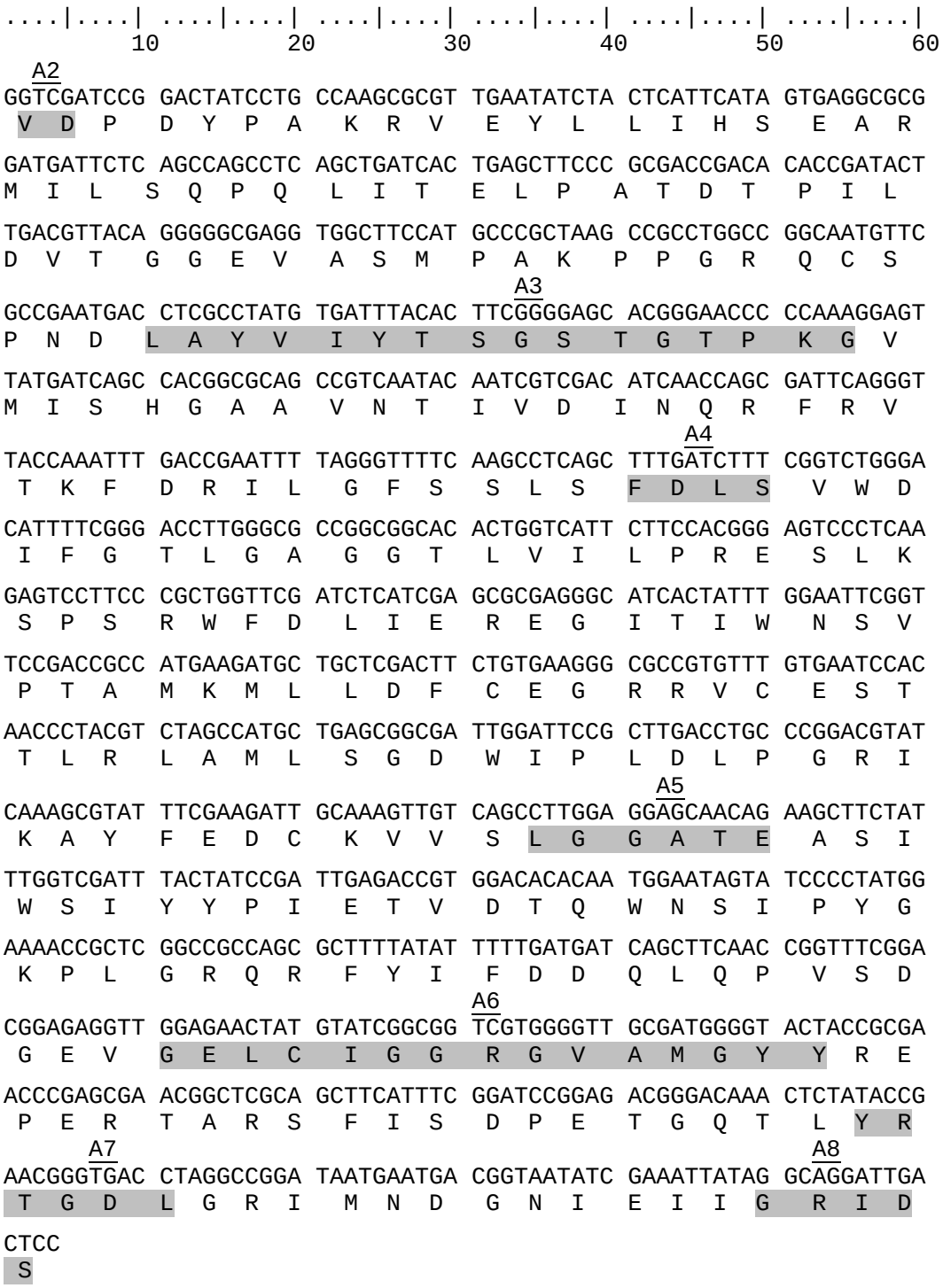

Figure 8 Nucleotide and deduced amino acid sequences of partial A domain of SM2_2 NRPS - conserved core motifs of A domain of NRPS are highlighted in gray.

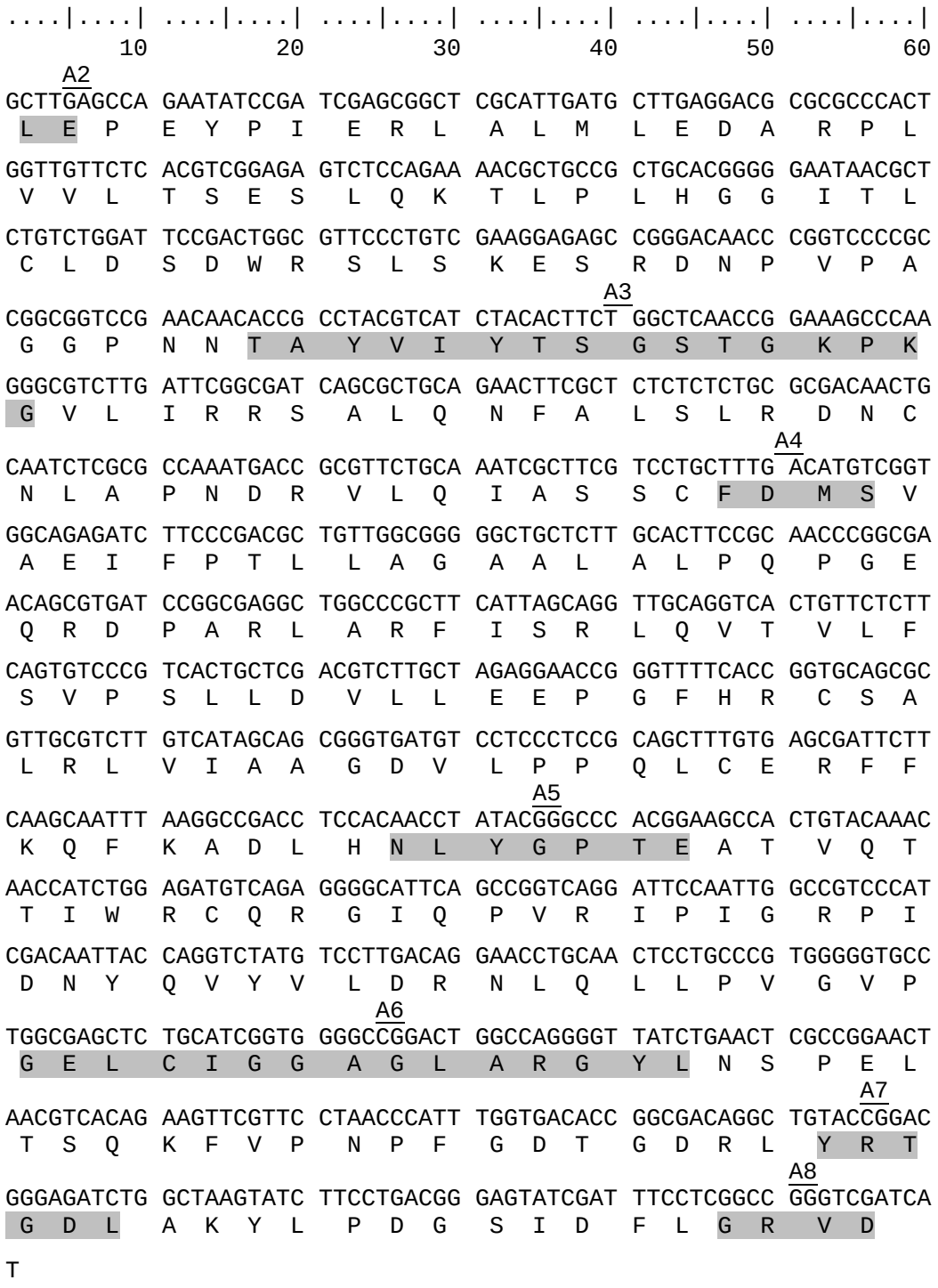

Figure 9 Nucleotide and deduced amino acid sequences of partial A domain of SM3 NRPS - conserved core motifs of A domain of NRPS are highlighted in gray.

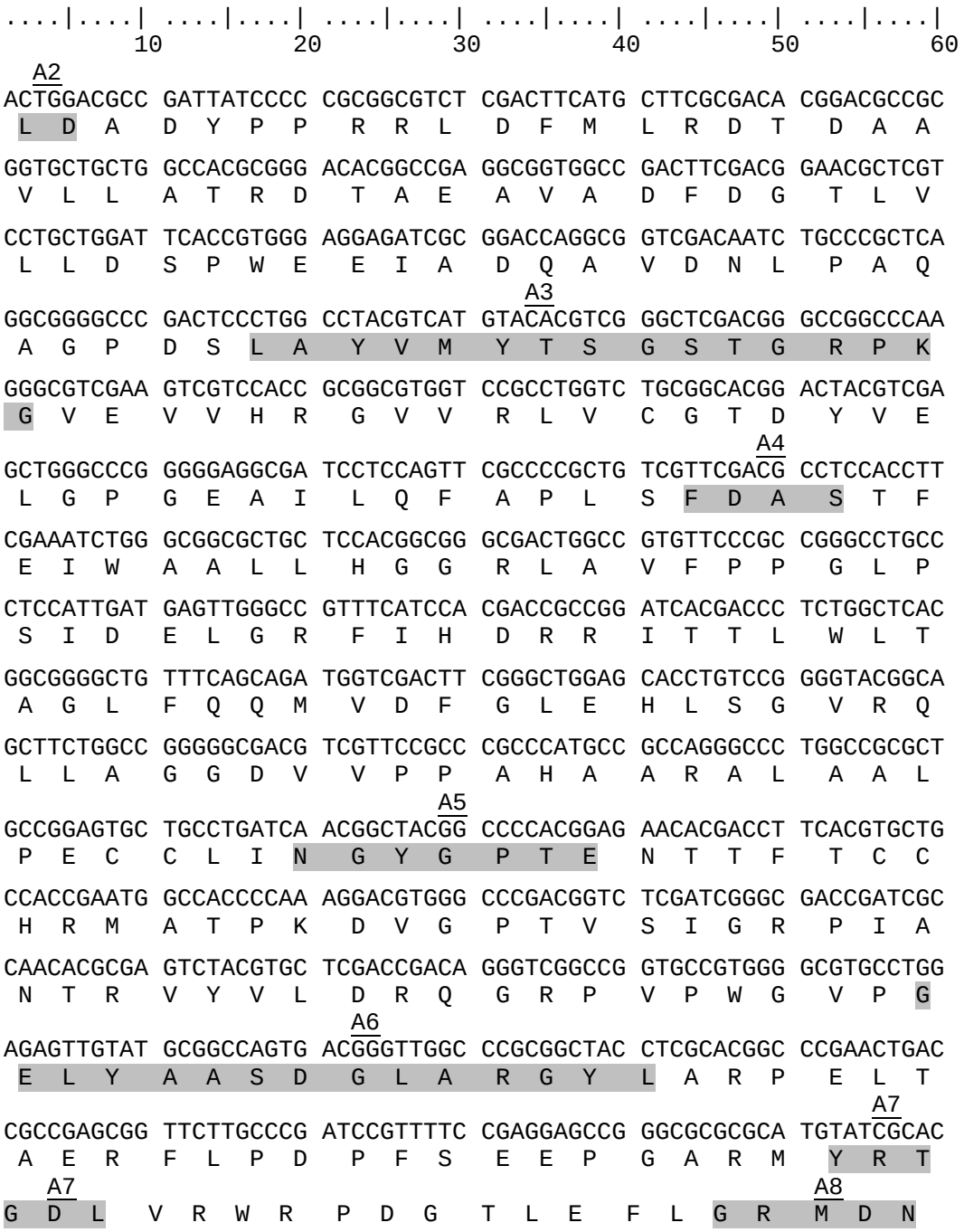

Figure 10 Nucleotide and deduced amino acid sequences of partial A domain of SM23 NRPS - conserved core motifs of A domain of NRPS are highlighted in gray.

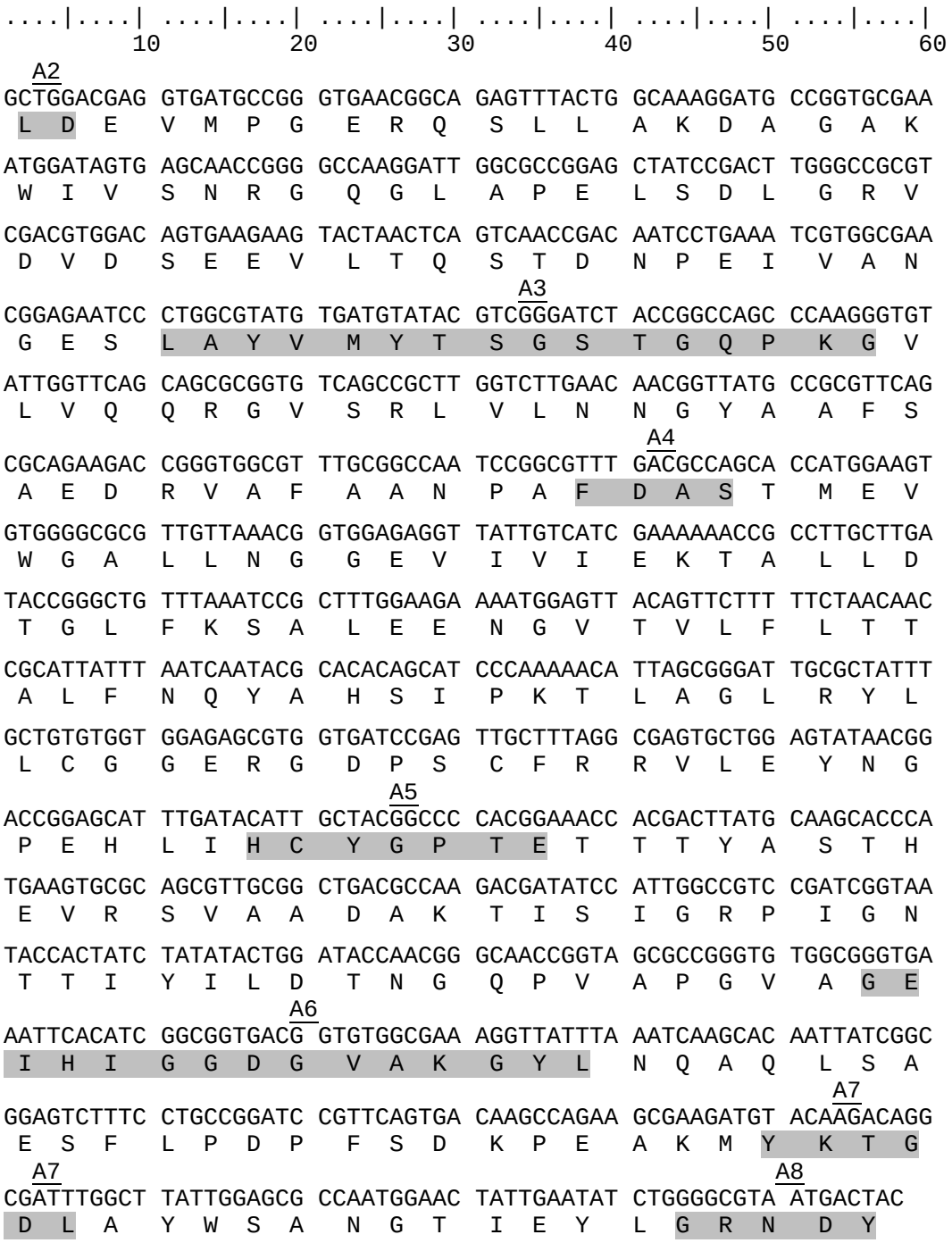

Figure 11 Nucleotide and deduced amino acid sequences of partial A domain of SM27 NRPS - conserved core motifs of A domain of NRPS are highlighted in gray.

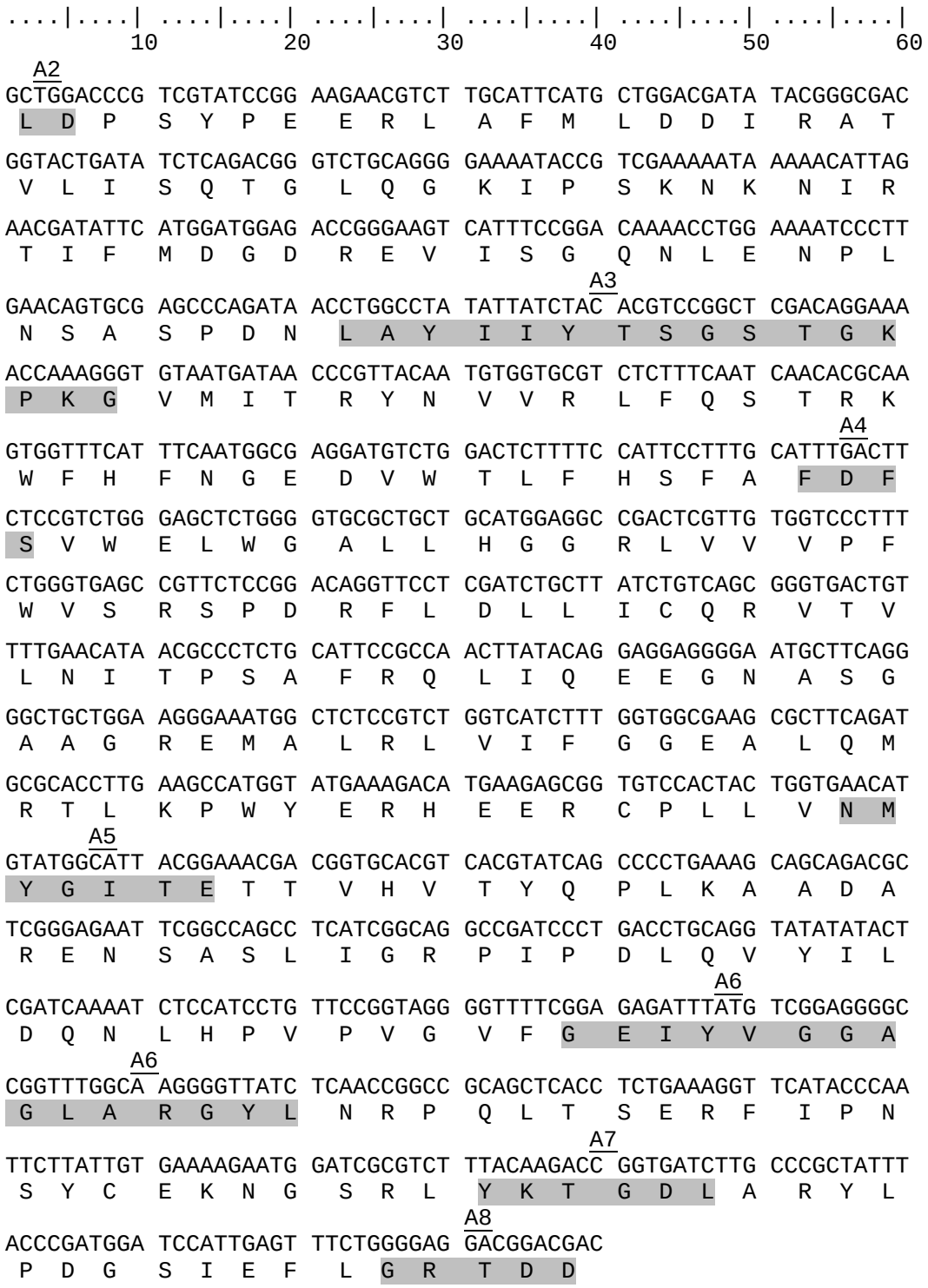

Figure 12 Nucleotide and deduced amino acid sequences of partial A domain of SM48 NRPS - conserved core motifs of A domain of NRPS are highlighted in gray.

All of the nucleotide sequences amplified directly from mangrove soil were translated into deduced amino acid sequences using BioEdit Sequence Alignment Editor 7.0.5.3. The amino acid sequences were aligned with ClustalW2. The alignment score of deduced amino acid sequence of partial A domain of SM2_2, SM3, SM23, SM27 and SM48 NRPS showed that they have low identity in ranged from 31% - 44% (Table 3). This suggested that five deduced amino acid sequences from mangrove soil were different from one another. The highly conserved core motifs in these sequences are in the catalytic A domain of peptide synthetase surrounding the active site where the amino acid substrates bind (Marahiel *et al*, 1997). It implied that the obtained A domains might not share cognate substrates.

Table 3 Percent identity of deduced amino acid sequences of partial A domain of NRPSs from Klongkone mangrove soil processed by ClustalW2. Total number of amino acid in each sequence (318-320) was shown in Figure 1B in Appendix B.

SM2_2	100				
SM3	34	100			
SM23	31	39	100		
SM27	34	32	42	100	
SM48	34	44	39	36	100
Clone	SM2_2	SM3	SM23	SM27	SM48

Nucleotide sequences of DNA insert in these clones was used as the query sequences to search for similar sequences in non-redundant protein sequences database, reference proteins database and swissprot protein sequences database of NCBI using blastx program. The maximum % identity and % positive of amino acid sequences were in the ranges of 52 – 62 and 69 – 78, as shown in Table 4. These results suggested the novelty of *nrps* gene obtained from mangrove soil

metagenome. Deduced amino acid sequence of each clone and A domains of known species hit (which showed the similar conserved core sequences with each of A domain) were used for phylogenetic analysis. Phylogenetic trees were constructed using Neighbor joining method. The reliability of an inferred tree was tested by bootstrapping with 1000 replication.

Table 4 Summary data of species hit with maximum score retrieved from NCBI by blast search with *nrps* sequences from Klongkone mangrove soil as the query sequences

Clone	**Known species with maximum score**	**% identity**	**% positive**	**Protein**
SM2_2	Proteobacteria (*Sorangium cellulosum*)	52 (171/236)	69 (226/236)	EpoB*
SM3	Cyanobacteria (*Microcystis* sp.)	52 (168/321)	66 (213/321)	McnA*
SM23	Proteobacteria (*Myxococcus xanthus*)	55 (174/319)	65 (206/319)	NRPS
SM27	Proteobacteria (*Pseudomonas entomophila*)	57 (185/320)	72 (232/320)	NRPS
SM48	Cyanobacteria (*Microcoleus chthonoplastes*)	62 (207/333)	78 (263/333)	NRPS

EpoB: epothilone synthetase B; McnA: *Microcystis* cyanopeptolin synthesis enzyme

Phylogenetic analysis of deduced amino acid sequence of SM2_2 with known species hit placed it in the clade of Cyanobacteria, as shown in Figure 13. This suggested that the novel *nrps* gene encoding A domain in clone SM2_2 was evolutionary related to NRPS from *Nostoc punctiforme* PCC 73102 (YP_001869919) with boostrap value of 62.7%.

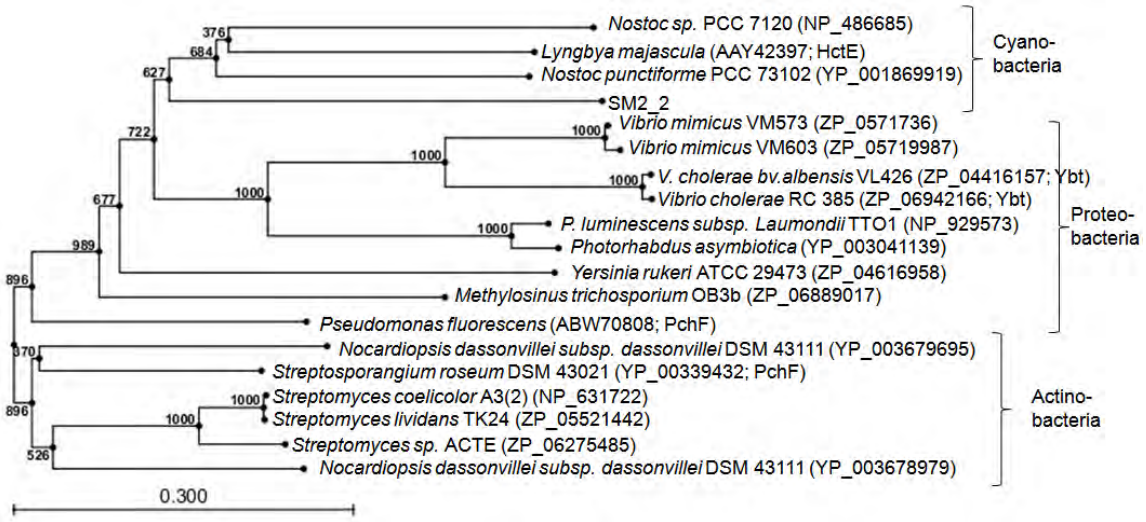

Figure 13 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of SM2_2 and 18 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.30 amino acid substitution per position. Abbreviations, HctE: hectochlorin synthetase E; Ybt: yersiniabactin synthetase ; PchF: pyochelin synthetase subunit F.

Phylogenetic analysis of SM3 A domain placed it as sister clade of bacteria in Phylum Cyanobacteria with boostrap value of 57.5%, as shown in Figure 14. This suggested that the novel *nrps* gene encoding A domain in clone SM3 might come from Cyanobacteria.

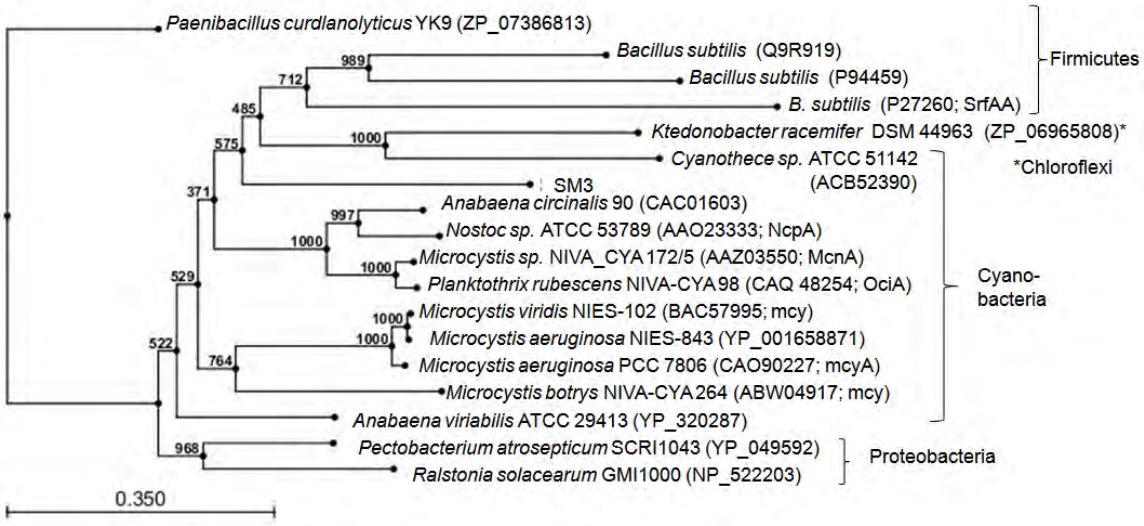

Figure 14 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of SM3 and 16 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.35 amino acid substitution per position. Abbreviations, SrfAA: surfactin synthase subunit 1; NcpA: nostocyclopeptide synthetase; McnA: *Microcystis* cyanopeptolin synthesis enzyme; OciA: *Planktothrix* cyanopeptolin synthesis enzyme subunit A; Mcy: microcystin synthetase

Phylogenetic analysis of A domain of SM23 placed it in the clade of Actinobacteria, as shown in Figure 15. This suggested that novel *nrps* gene encoding A domain in clone SM23 was evolutionary related to NRPS of *Streptomyces clavuligerus* ATCC 27064 with boostrap support of 85.6%.

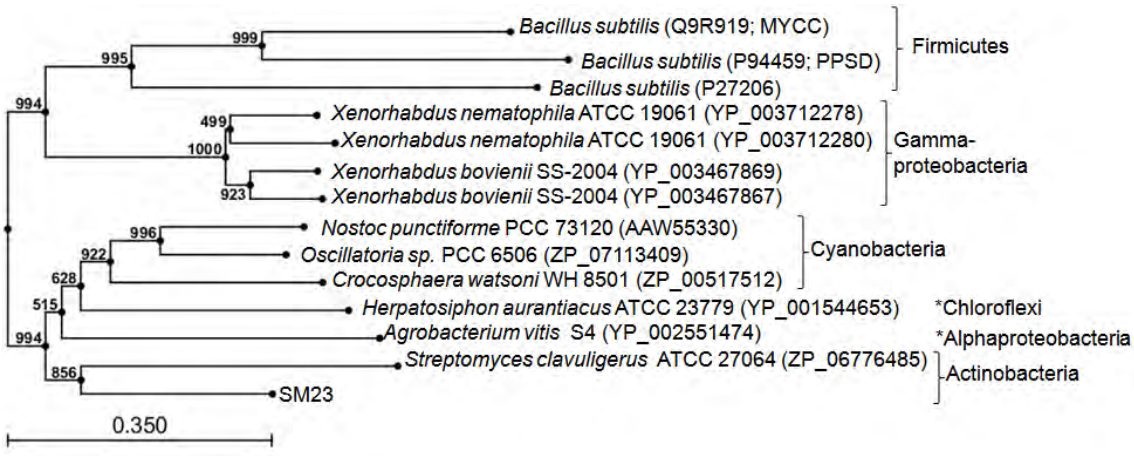

Figure 15 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of SM23 and 13 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.35 amino acid substitution per position. Abbreviations, MYCC: mycosubtilin synthase subunit C; PPSD: plipastatin synthase subunit D; YP_003712280 and YP_003467867: Phenylalanine racemase; YP_003467869: Ornithine racemase

Phylogenetic analysis of A domain of SM27 placed it in the clade of Gammaproteobacteria, as shown in Figure 16. , This suggested that novel *nrps* gene encoding A domain in clone SM27 was evolutionary related to massetolide B synthetic enzyme of *Pseudomonas fluorescence* (ABH06368) and NRPS of *P. fluorescence* SBW25 (YP_002872142) with boostrap support of 83.2%.

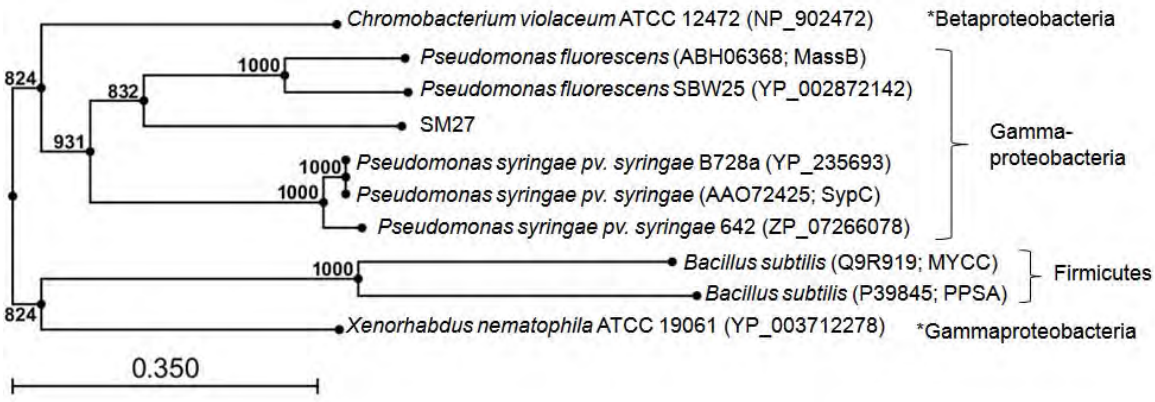

Figure 16 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of SM27 and 9 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.35 amino acid substitution per position. Abbreviations, MassB: massetolide B synthethase; SypC: syringopeptin synthetase subunit C; MYCC: Mycosubtilin synthase subunit C; PPSA: plipastatin synthase subunit A

Phylogenetic analysis of A domain of SM48 placed it in the clade of Cyanobacteria, as shown in Figure 17. This suggested that novel *nrps* gene encoding A domain in clone SM48 was evolutionary related to *Nostoc azollae* 0708 with boostrap support of 43.4%.

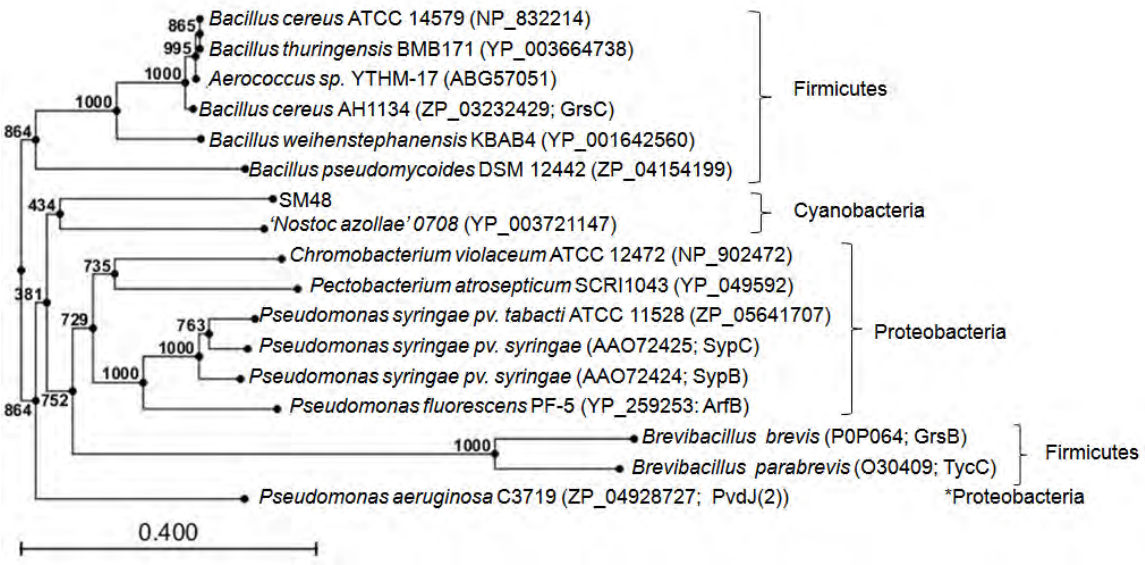

Figure 17 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of SM48 and 16 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.40 amino acid substitution per position. Abbreviations, GrsC: gramicidin synthetase subunit C; SypC: syringopeptin synthetase subunit C; SypB: syringopeptin synthetase unit B; ArfB: arthrofactin synthetase subunit B; PvdJ(2): pyoverdine synthase; GrsB: gramicidin S synthase subunit B; TycC: tyrocidine synthase subunit C

Table 5 summarizes phyla of bacteria having NRPSs evolutionary related to sequences from Klongkone mangrove soil metagenome. It was noticeable that NRPSs having evolutionary related to those of clones SM2_2 and SM23 and NRPSs having maximum % identity were from bacteria in different phyla. Although the primer pair MTF2/MTR is specific for A domain of *nrps* from aquatic cyanobacteria, data from phylogenetic analysis of five deduced amino acid sequences revealed that this primer pair could amplify the partial adenylation domain from actinobacteria and proteobacteria. This correlated with the results from phylogenetic analysis of 16s rRNA gene in the Sundarban mangrove sediment which revealed the major divisions of detected bacterial phyla including Proteobacteria and Actinobacteria (Ghosh *et al*., 2010).

Based on the identity of deduced amino acid sequences of A domains of NRPSs obtained from Klongkone mangrove soil (Table 3), the identity to known protein hits (Table 4), and phylogenetic analysis (Figures 13-17), it might indicate the diversity of *nrps* gene from mangrove soil metagenome.

Table 5 Summary data of phylogenetic analysis of A domain of NRPS amplified from mangrove soil.

Clone	**Maximum % identity to known species**			**Related phylum (% Boostrap value)**	**Figure of phylogenetic tree**
	Nonredundant proteins	**Reference proteins**	**Swissprot proteins**		
SM2_2	52	52	47	Cyanobacteria (62.7)	Figure 13
SM3	52	50	45	Cyanobacteria (57.5)	Figure 14
SM23	57	57	54	Actinobacteria (85.6)	Figure 15
SM27	57	58	42	Proteobacteria (83.2)	Figure 16
SM48	62	62	55	Cyanobacteria (43.4)	Figure 17

4.2.2 Predicted amino acid which is activated by A domain of NRPS from mangrove soil

Analysis of the crystal structure of adenylation domain of gramicidine synthetase NRPS (GrsA) that activates phenylalanine revealed eight (Challis *et al*., 2000) or ten (Rausch *et al*., 2005) residue positions which correlated with substrate specificity. These eight or ten decisive residues of GrsA could be extracted from the multiple sequence alignment of A domains. This example became the basis for prediction of specificity of A domain with unknown function, e.g. from PCR amplification or genome sequencing. Prediction of amino acid activated by A domain of *nrps* gene could facilitate site directed mutagenesis for the alteration of A domain specificity (Challis *et al*., 2000). In order to perform the prediction of amino acid activated by partial A domain of NRPS which was PCR amplified from Klongkone soil mangrove, five deduced amino acid sequences were used as inputs for the analysis by PKS/NRPS Analysis Web-site (Bachmann and Ravel, 2009). The processes of prediction were followed as indicated in the tutorial. The results from analysis were summarized in Table 6. Five A domains of NRPSs from mangrove soil recognized and activated different amino acid substrates. This finding correlated with the data from the analysis of the seven core motifs of these NRPS (Figure 8-12). Since core motifs were substrate binding sites, different in core motifs resulting in different binding sites. NRPS of clones SM2_2, SM3 and SM27 had residues in the binding pocket different from the sequences in database and could not be predicted for the activated amino acids. The differences of residues in the binding pocket of these three NRPS might represent the clones' novelty.

Table 6 Summary data of the prediction of amino acid activated by A domain of NRPS which was PCR amplified from Klongkone mangrove soil metagenome.

NRPS	Residues in the Binding pocket	Predicted amino acid*
SM2_2	DLWNRALT	No amino acid HIT
SM3	DMAFIALV	No amino acid HIT
SM23	DAFWLGXX	TycC-M4-Val
SM27	DAMFLGCT	No amino acid HIT
SM48	DFWNIGMV	CchH-M2-Thr

*TycC: Tyrocidine synthase subunit C; CchH: coelichelin synthase; Val: Valine; Thr: Threonine

Metagenomes extracted from soil mangrove were screened by PCR in order to search for novel *nrps*. The data from sequence alignment and prediction of activated amino acids of five sequences derived from the PCR products supported the novelty of five partial A domains of *nrps* from mangrove soil. Moreover, phylogenetic analysis data could reveal the diversity of *nrps* amplified from soil. Three of five tentative novel *nrps* genes were evolutionary related to those from Cyanobacteria. The rest of them were related to those from Proteobacteria and Actinobacteria. These findings agree with previous research (Zhao *et al*., 2008) in that sequence alignment of A domain amplified from soil could be used in the novel *nrps* detection. Data from phylogenetic analysis demonstrated that the majority of amplified genes belonged to Cyanobacteria phyla (Zhao *et al*., 2008). This is because the MTF2/MTR used in the study was designed by sequence alignment of A domain in known gene cluster of bacteria in the clade of Cyanobacteria, Actinobacteria, and Firmicutes (Neilan *et al*., 1999). The novelty and diversity of five deduced amino acid sequences suggested that Klongkone mangrove soil metagenomes might have a potential to be a source of novel biosynthetic genes which is valuable for the discovery of

new peptide antibiotics. Therefore, the metagenomic DNA from mangrove soil sample was used for metagenomic library construction in order to access the novel *nrps* genes.

4.3 Construction of metagenomic library

Sequencing results from section 4.2 confirmed that Klongkone mangrove soil contained novel *nrps* genes. Therefore, this soil sample was a suitable source of novel peptides. Purified soil DNA was used for metagenomic library construction with CopyControl™ Fosmid Library Production Kit with pCC2FOS™ vector. Transfection of EPI300-T1^{R} cells with recombinant fosmids were cultured on petri dish. This resulted in approximately 14,000 clones of metagenomic library (approximately 150 clones per plate). Transformants from each plate were washed with 20% glycerol and stored in 1.5-ml microtubes at -80°C. This process resulted in a total number of 95 pools of metagenomic library. There was a hypothesis that clones with strange morphology or pigment producing ability may produce uncommon secondary metabolites with antibiotic activities. These phenotypes were screened from library. Nevertheless, no clone with strange morphology or pigment-producing ability was found.

4.4 Screening of metagenomic library

4.4.1 Functional-based screening

Metagenomic library was screened for antibacterial and antifungal activities in order to search for *nrps* which may encode for the active peptides. *Candida albicans* ATCC 90028 and chloramphenicol resistance *Bacillus subtilis* constructed from *B. subtilis* ATCC 6633 were used as tested microorganisms. These microorganisms were selected for metagenomic library screening because they have an ability to grow on culture media containing

chloramphenicol. From 95 pools of the library, 29 pools (approximately 4,300 clones) were screened (1-20, 41-44, 91-95). A total of 417 clones were collected as they might have a small inhibition zone noticeable under magnifying glass (Table 7). Confirmation test against all test organisms resulted in no active clone. The result is contrast to a previous research in functional screening of mangrove soil metagenomic that could discover novel functional enzymes, e.g. lipase (from 2,400 library clones with approximately 25-kb insert (Couto *et al*., 2010)) and novel multicopper oxidase with laccase activity (from 8,000 library clones with approximately 5-kb insert (Ye *et al*., 2010)). It was not surprised because genes studied in those previous researches had a smaller size than *nrps* (lipase, 1-2 kb; multicopper oxidase with laccase activity, 1.5 kb; *nprs*, 5.9-48.5 kb). Small genes could be easily contained in large insert fragment, so those genes should have full length and could express. Moreover, exotic *nrps* gene could not express for a functional NRPS in *E. coli* for various reasons as discussed in Chapter II. DNA shearing, differences in codon usage, protein folding mechanisms, chaperone used in *E. coli* host cell and the lack of extracellular transportation mechanism could disrupt the *nrps* expression, resulting in no or inactive NRPS (Streit and Schmitz, 2004). Functional-based screening of the rest of library was recommended because this may lead to the discovery of active clones. Other screening which compatible with the use of chloramphenicol should be tried.

Table 7 Number of clone with suspicious inhibition zone against specific test organisms

Pool	*C. albicans*	*C. albicans* + ketoconazole	*B. subtilis*	Total
1	0	4	10	14
2	0	3	16	19
3	0	0	20	20
4	0	3	11	14
5	0	1	72	73
9	0	0	19	19
10	0	0	12	12
42	22	8	75	105
43	0	18	69	87
44	0	2	52	54
Total	22	39	356	417

4.4.2 Sequence-based library screening

4.4.2.1 *nrps* gene from metagenomic library

Metagenomic library was screened for *nrps* gene. At least 1,500 clones from ten pools (30-39) were screened by PCR using MTF2/MTR primer pair. Three pools (33, 36 and 37) were PCR positive and may contain 1-kb insert (Figure18, Lane 2-4). Pool 36 and 37 were selected for further study. Selected pool was 10^{-6} fold diluted with normal saline solution (0.9%) and spread on 10 plates of LB agar with chloramphenicol. After an overnight incubation, each plate was replicated into LB agar with chloramphenicol and arabinose. Replicated plates were collected for bacterial cells and extracted plasmid DNA by boiling and collecting for supernatant which was used as a DNA template for PCR screening. Replicated plates that

demonstrated positive PCR reaction referred to the starter plates that contain PCR-positive colonies. Theoretically, each colony in those plates should be collected as a single colony. From pool 36, two colonies were positive for PCR screening and were designated as 3671310 and 3671314. Because of susceptibility to contamination, those two clones were streak for isolation. Isolated clones were confirmed for *nrps* by PCR. Four clones from pool 36 were selected and designated as 3671310-03, 3671310-08, 3671310-11 and 3671314-25 (Figure18, Lane 5-8). For pool 37, two clones named as 3710 and 3746 was collected from metagenomic library (Figure18, Lane 9-10). It was noticeable that MTF2/MTR primer amplified the nonspecific 900-bp PCR products from clone of pool 36 and nonspecific 1,100-bp PCR products from clone of pool 37. The nonspecific PCR products cannot be eliminated by agarose gel electrophoresis and would interfere further DNA sequencing process. Therefore, PCR products from these six PCR-positive clones must be cloned into pGEM® -T Easy vector. White colonies containing 1-kb PCR product were screened by PCR using MTF2/MTR primers pair as mentioned in section 3.5.1 in methodology. Name and sequencing details of those PCR-positive transformants were shown in Table 8. The nucleotide sequences of the partial A domain of *nrps* amplified from from pool 36 and 37 clones from metagenomic library were submitted to GenBank. Their accession numbers were listed in Table 8. All of the nucleotide sequences from library were translated into deduced amino acid sequences using BioEdit Sequence Alignment Editor 7.0.5.3.

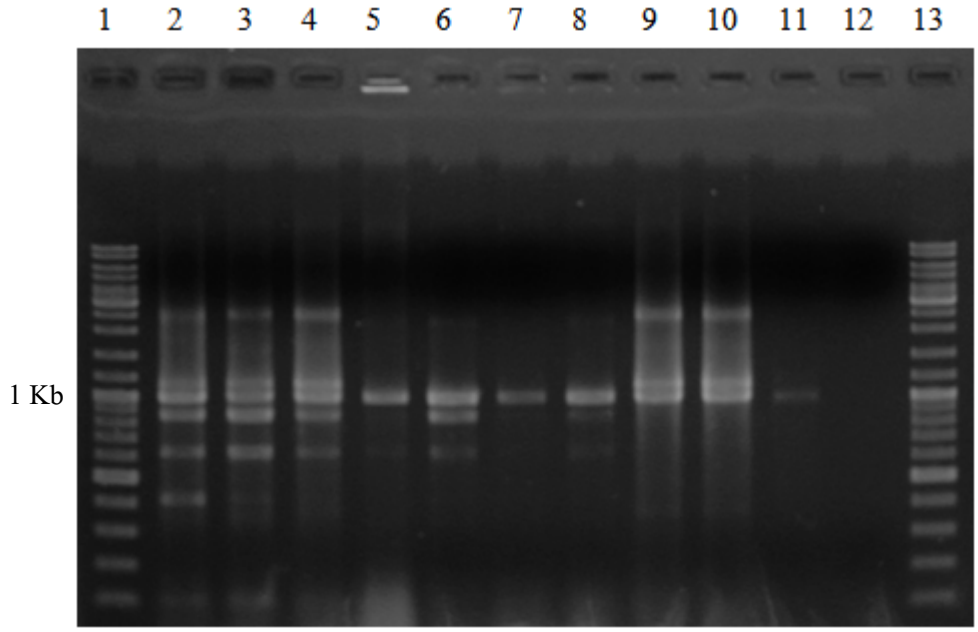

Figure 18 Agarose gel (1%) electrophoresis of PCR product from metagenomic library amplified with MTF2/MTR primer pair. PCR products with 1-kb in length are shown. Lanes 1 and 13, GeneRuler DNA marker mix in Kb length; 2, PCR product from metagenomic library pool 33; 3, PCR product from pool 36; 4, PCR product from pool 37; 5, PCR product from clone 3671310-03; 6, PCR product from clone 3671310-08; 7, PCR product from clone 3671310-11; 8, PCR product from 3671314-25; 9, PCR product from clone 3710; 10, PCR product from clone 3746; 11, *Micromonospora chalcea* ATCC 12452 (as positive control); 12, water (as negative control).

Table 8 Summary data of transformants derived from pool 36 and 37 of soil metagenomic library.

Transformant	NRPS containing clone from metagenomic library	Length of nucleotide sequence (bp)	Accession number of nucleotide sequences	Figure of nucleotide and deduced amino acid sequences
360305	3671310-03	970	HQ286558	Figure 19
360310	3671310-03	970	HQ286559	Figure 20
360312	3671310-03	970	HQ286560	Figure 21
360802	3671310-08	970	HQ286561	Figure 22
360804	3671310-08	970	HQ286562	Figure 23
361101	3671310-11	970	HQ286563	Figure 24
361102	3671310-11	970	HQ286564	Figure 25
371001	3710	988	HQ286567	Figure 26
371002	3710	987	HQ286568	Figure 27
374601	3746	997	HQ286569	Figure 28
374605	3746	988	HQ286570	Figure 29

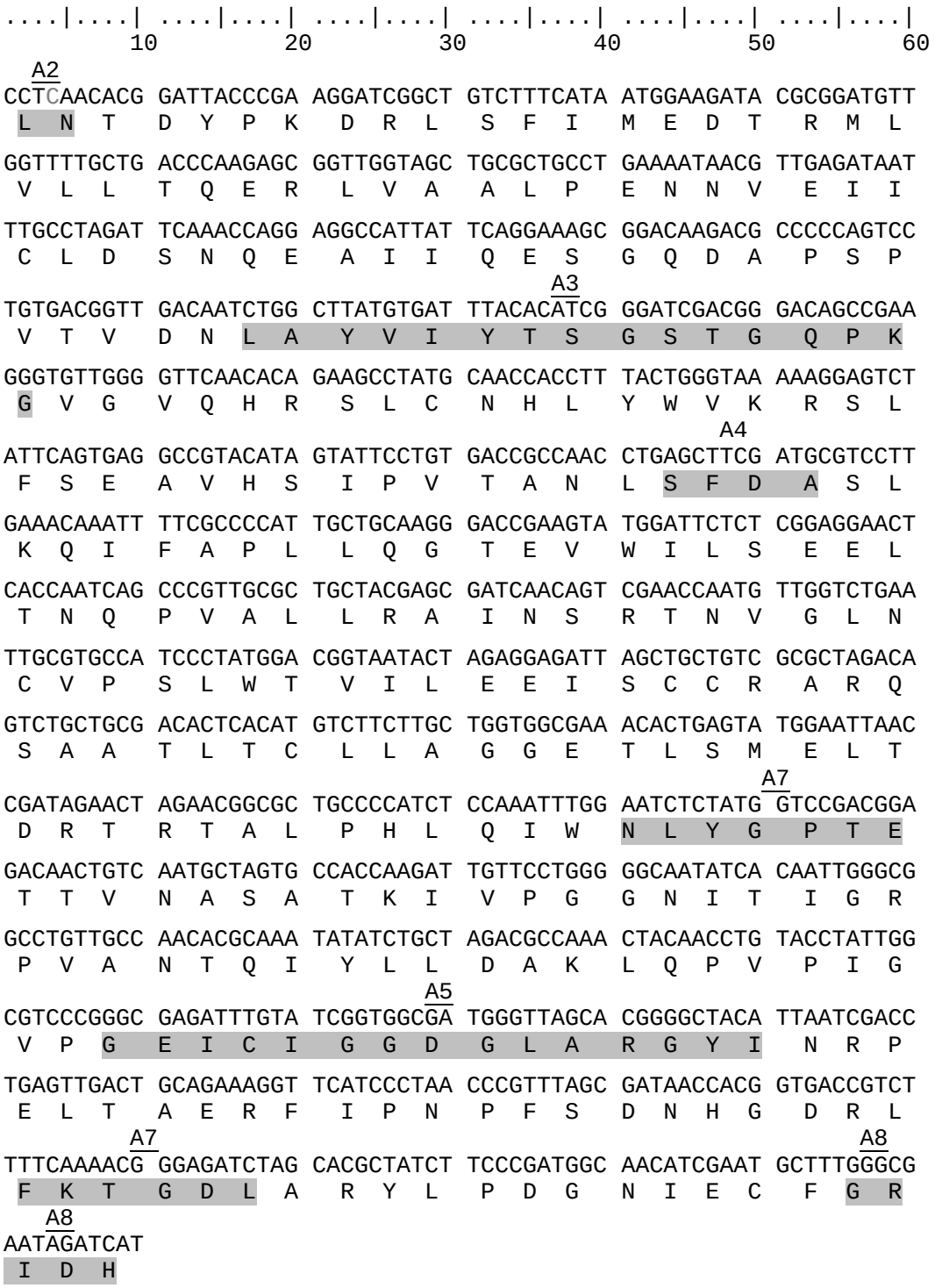

Figure 19 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 360305 - conserved core motifs of A domain of NRPS are highlighted in gray.

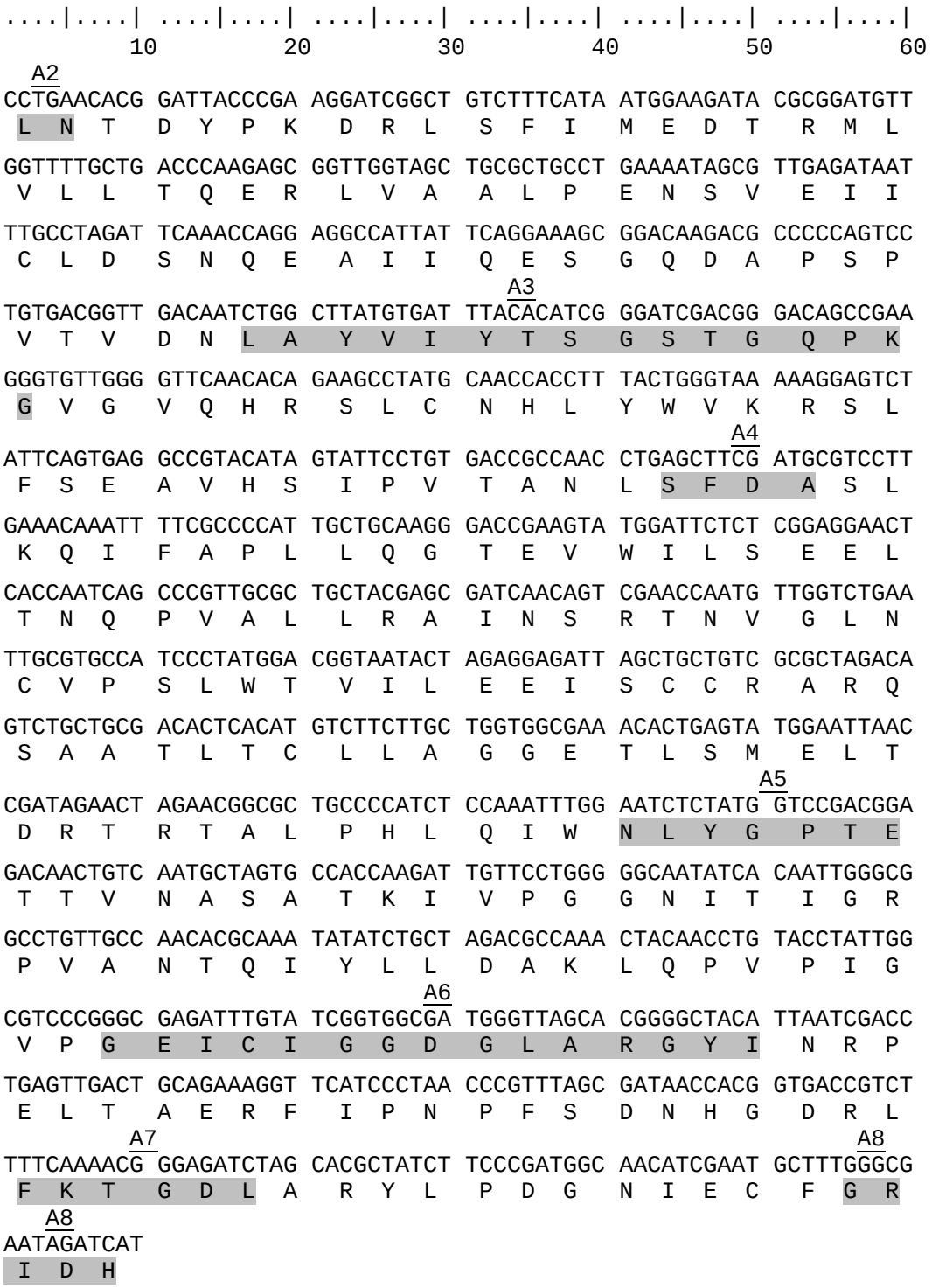

Figure 20 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 360310 - conserved core motifs of A domain of NRPS are highlighted in gray.

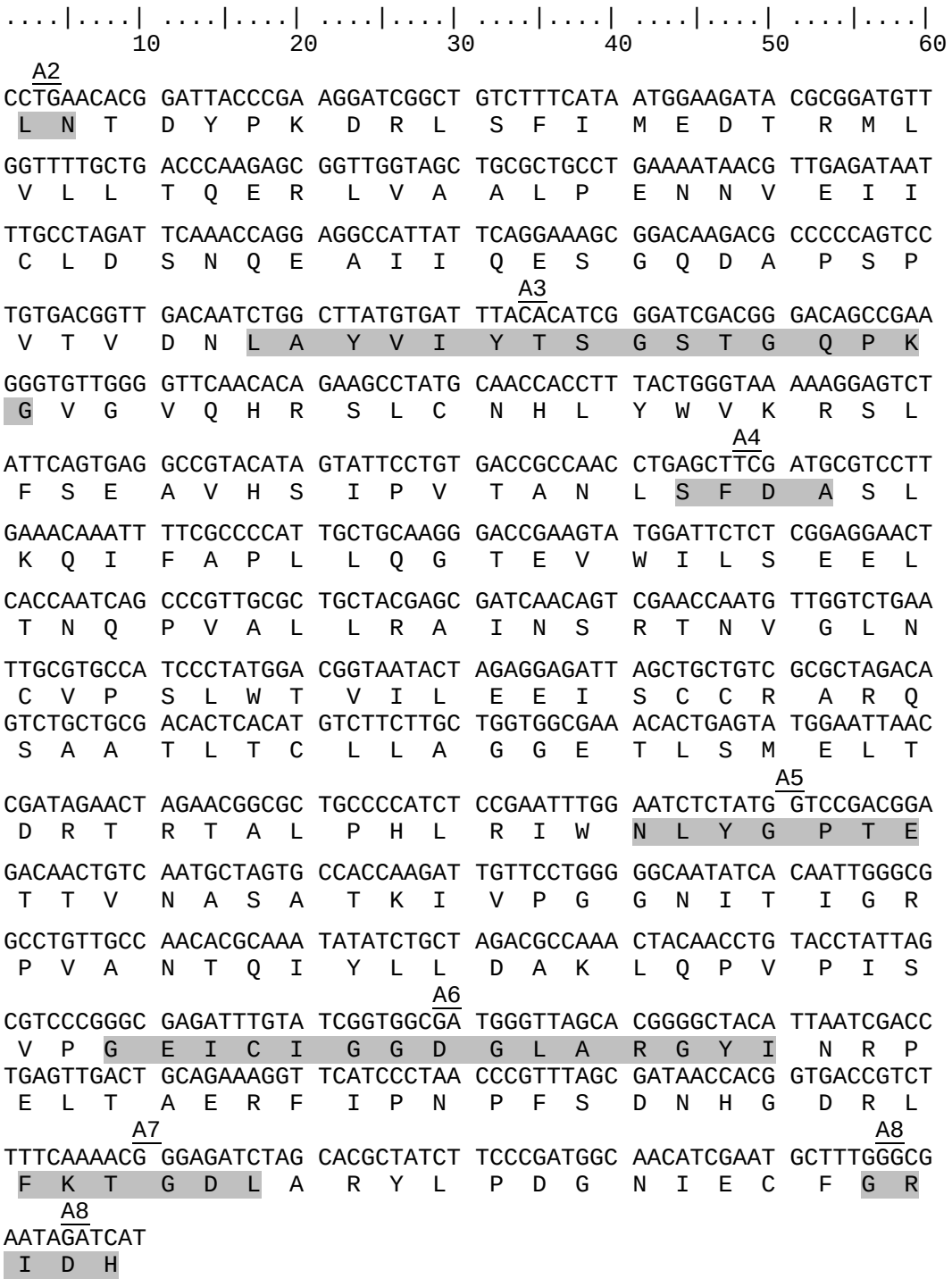

Figure 21 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 360312 - conserved core motifs of A domain of NRPS are highlighted in gray.

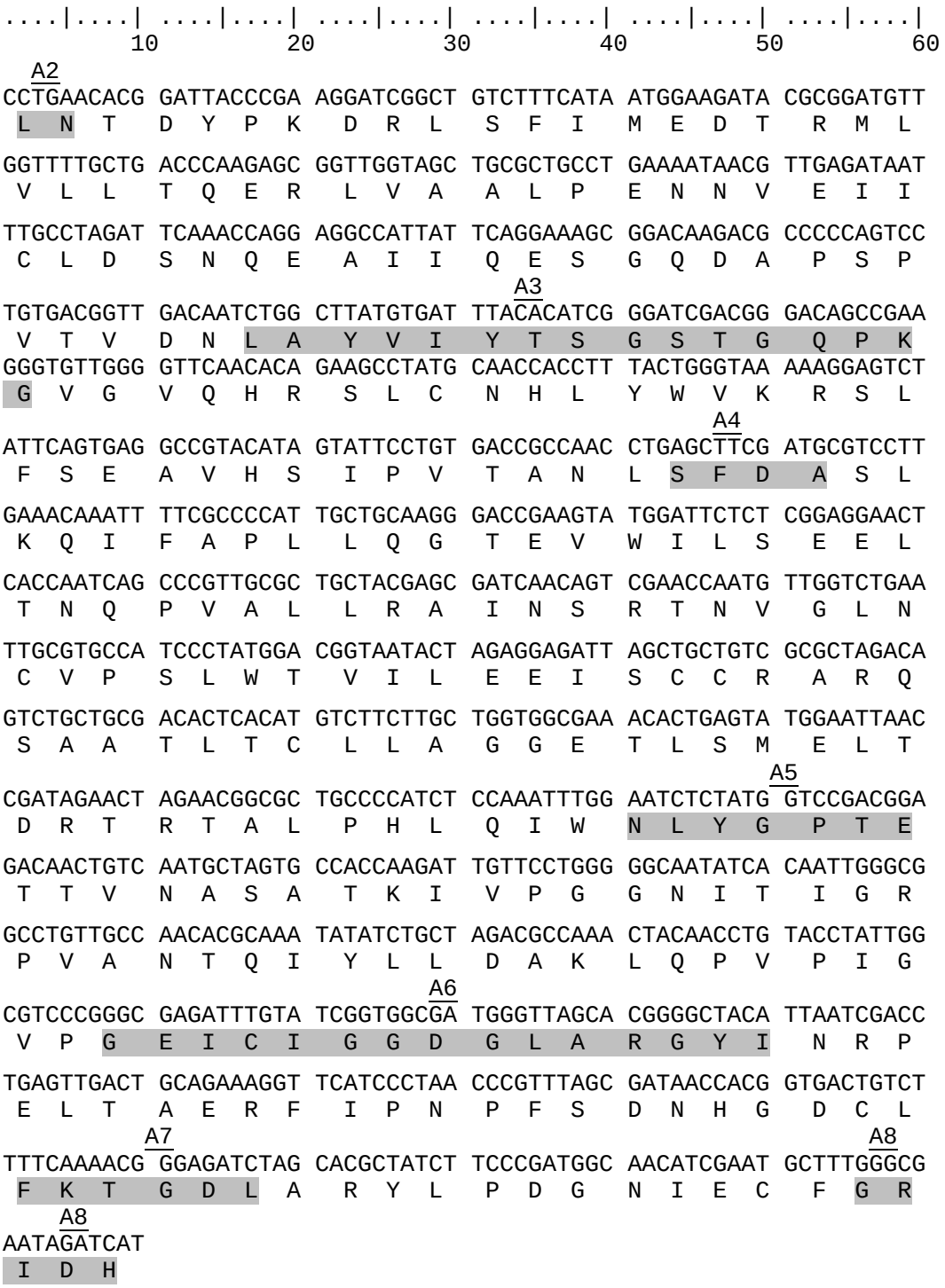

Figure 22 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 360802 - conserved core motifs of A domain of NRPS are highlighted in gray.

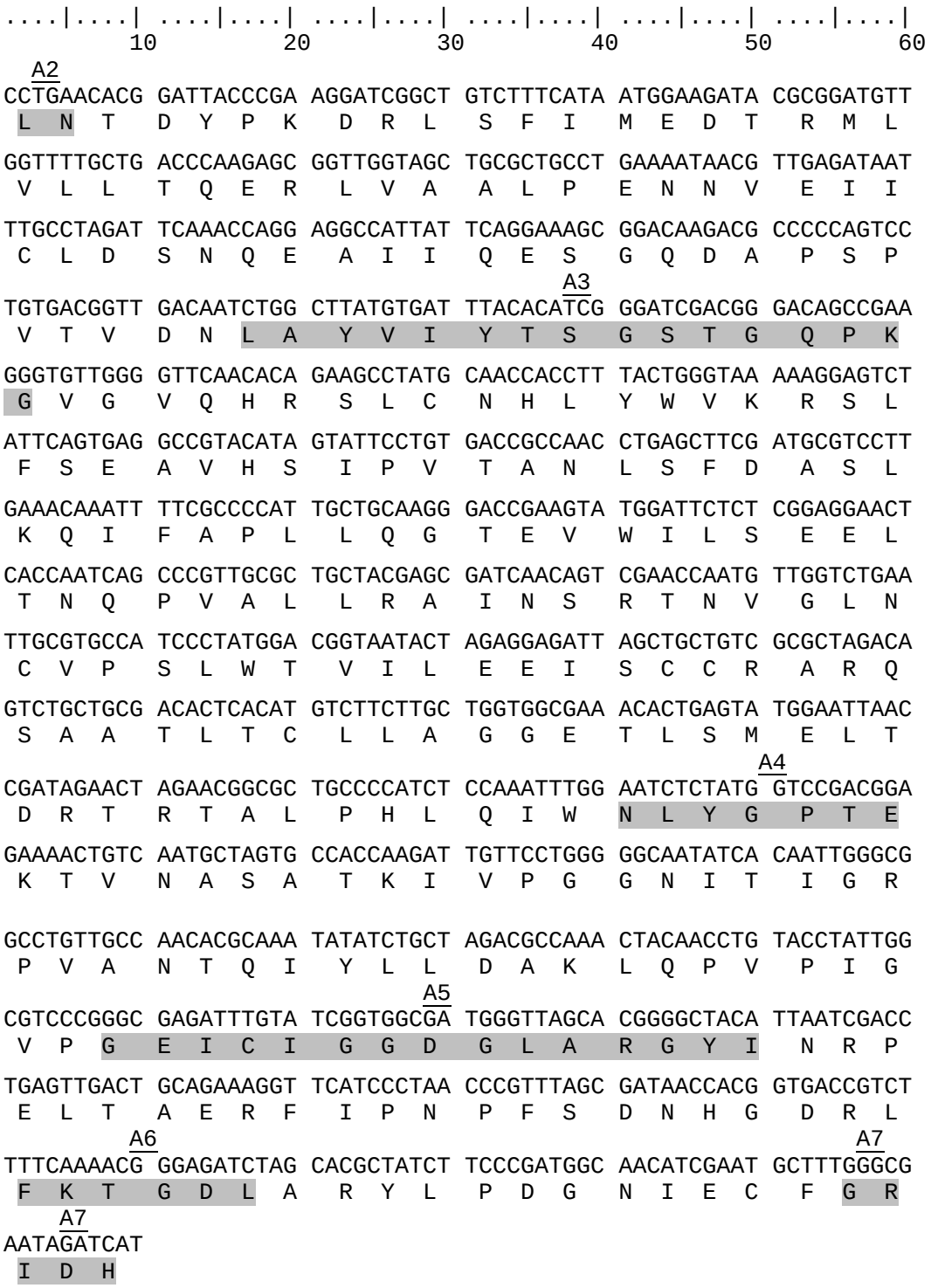

Figure 23 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 360804 - conserved core motifs of A domain of NRPS are highlighted in gray.

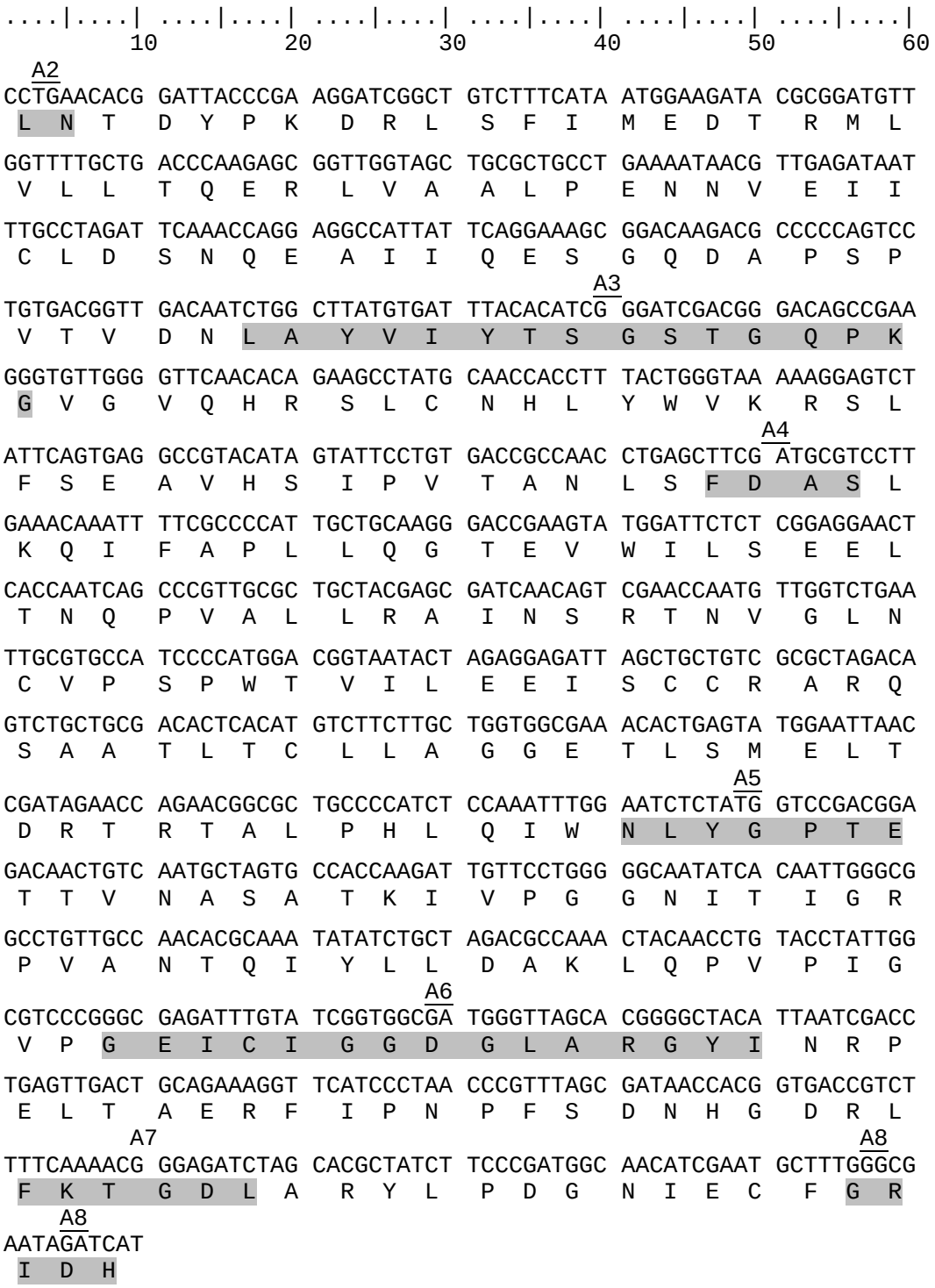

Figure 24 Nucleic acid and deduced amino acid sequences of partial A domain of NRPS of clone 361101 - conserved core motifs of A domain of NRPS are highlighted in gray.

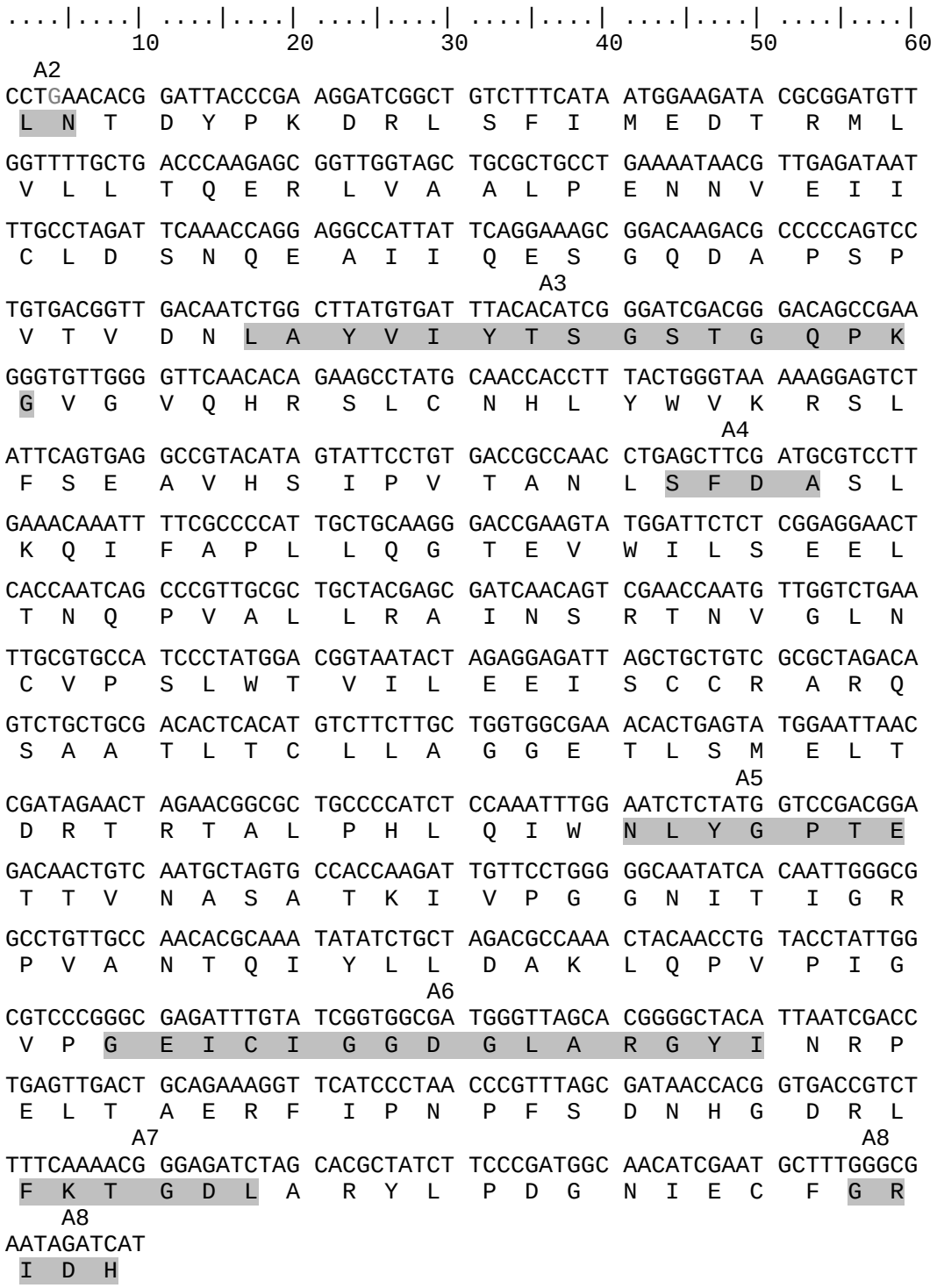

Figure 25 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 361102 - conserved core motifs of A domain of NRPS are highlighted in gray.

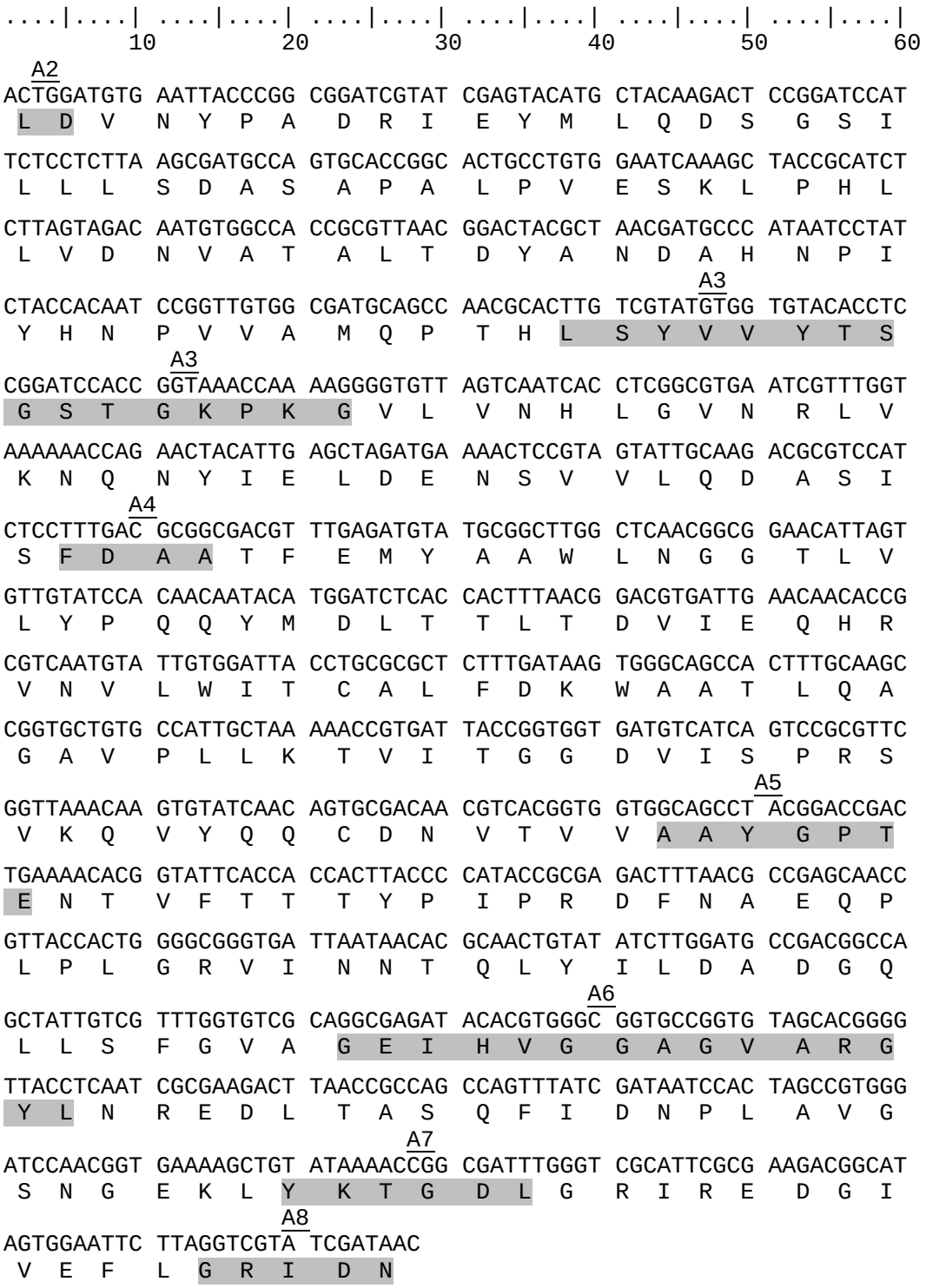

Figure 26 Nucleotide Nucleic acid and deduced amino acid sequences of partial A domain of NRPS of clone 371001 - conserved core motifs of A domain of NRPS are highlighted in gray.

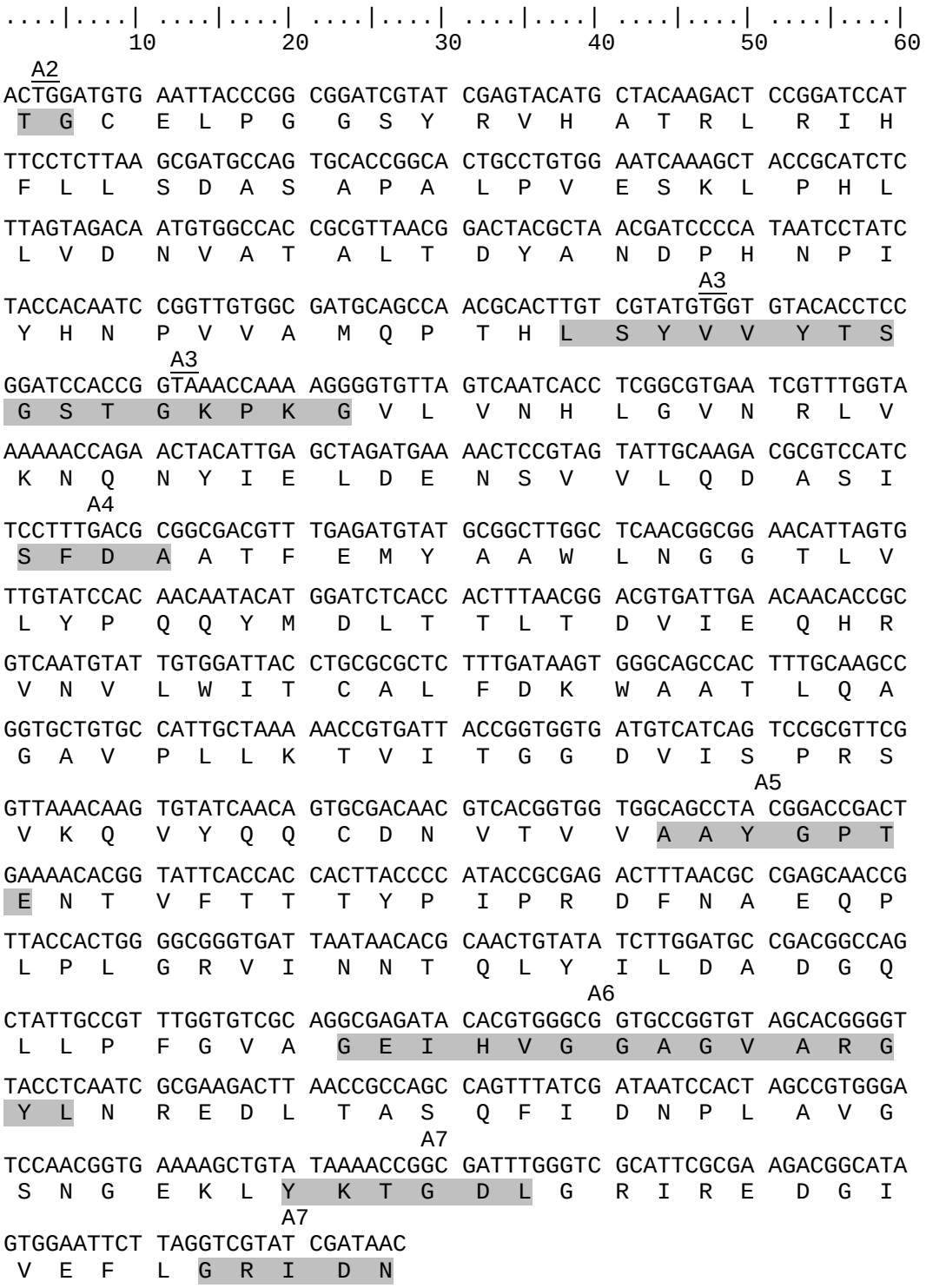

Figure 27 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 371002 - conserved core motifs of A domain of NRPS are highlighted in gray.

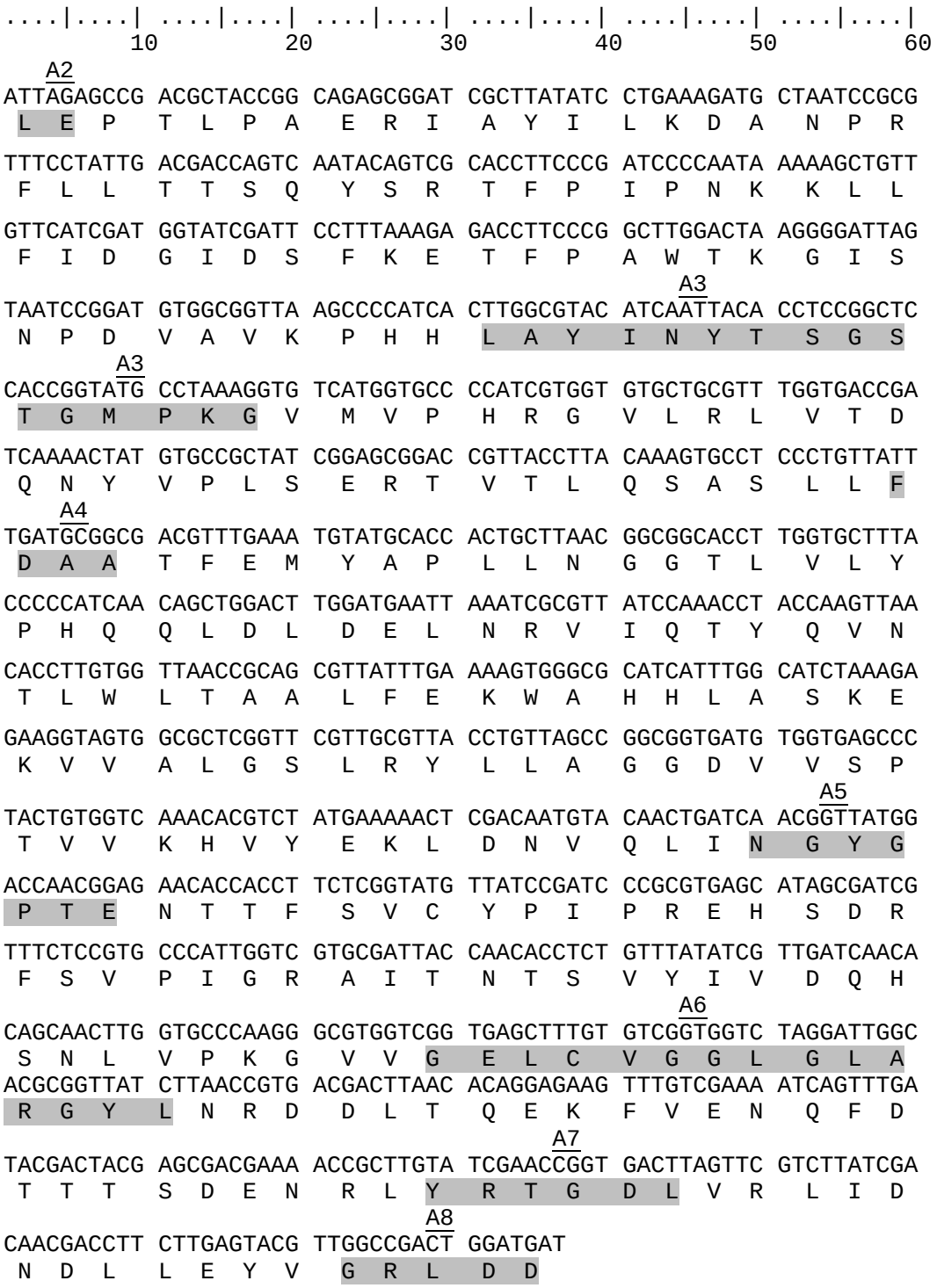

Figure 28 Nucleotide and deduced amino acid sequences of partial A domain of NRPS of clone 374601 - conserved core motifs of A domain of NRPS are highlighted in gray.

```
....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
         10         20         30         40         50         60
  A2
ACTGGATGTG AATTACCCGG CGGATCGTAT CGAGTACATG CTACAAGACT CCGGATCCAT
 L  D  V   N  Y  P  A   D  R  I   E  Y  M   L  Q  D  S   G  S  I

TCTCCTCTTA AGCGATGCCA GTGCACCGGC ACTGCCTGTG GAATCAAAGC TACCGCATCT
 L  L  L   S  D  A  S   A  P  A   L  P  V   E  S  K  L   P  H  L

CTTAGTAGAC AATGTGGCCA CCGCGTTAAC GGACTACGCT AACGATGCCC ATAATCCTAT
 L  V  D   N  V  A  T   A  L  T   D  Y  A   N  D  A  H   N  P  I
                                                  A3
CTACCACAAT CCGGTTGTGG CGATGCAGCC AACGCACTTG TCGTATGTGG TGTACACCTC
 Y  H  N   P  V  V  A   M  Q  P   T  H  L   S  Y  V  V   Y  T  S
          A3
CGGATCCACC GGTAAACCAA AAGGGGTGTT AGTCAATCAC CTCGGCGTGA ATCGTTTGGT
 G  S  T   G  K  P  K   G  V  L   V  N  H   L  G  V  N   R  L  V

AAAAAACCAG AACTACATTG AGCTAGATGA AAACTCCGTA GTATTGCAAG ACGCGTCCAT
 K  N  Q   N  Y  I  E   L  D  E   N  S  V   V  L  Q  D   A  S  I
        A4
CTCCTTTGAC GCGGCGACGT TTGAGATGTA TGCGGCTTGG CTCAACGGCG GAACATTAGT
 S  F  D   A  A  T  F   E  M  Y   A  A  W   L  N  G  G   T  L  V

GCTGTATCCA CAACAATACA TGGATCTCAC CACTTTAACG GACGTGATTG AACAACACCG
 L  Y  P   Q  Q  Y  M   D  L  T   T  L  T   D  V  I  E   Q  H  R

CGTCAATGTA TTGTGGATTA CCTGCGCGCT CTTTGATAAG TGGGCAGCCA CTTTGCAAGC
 V  N  V   L  W  I  T   C  A  L   F  D  K   W  A  A  T   L  Q  A

CGGTGCTGTG CCATTGCTAA AAACCGTGAT TACCGGTGGT GATGTCATCA GTCCGCGTTC
 G  A  V   P  L  L  K   T  V  I   T  G  G   D  V  I  S   P  R  S
                                                     A5
GGTTAAACAA GTGTATCAAC AGTGCGACAA CGTCACGGTG GTGGCAGCCT ACGGACCGAC
 V  K  Q   V  Y  Q  Q   C  D  N   V  T  V   V  A  A  Y   G  P  T

TGAAAACACG GTATTCACCA CCACTTACCC CATACCGCGA GACTTTAACG CCGAGCAACC
 E  N  T   V  F  T  T   T  Y  P   I  P  R   D  F  N  A   E  Q  P

GTTACCACTG GGGCGGGTGA TTAATAACAC GCAACTGTAT ATCTTGGATG CCGACGGCCA
 L  P  L   G  R  V  I   N  N  T   Q  L  Y   I  L  D  A   D  G  Q
                                              A6
GCTATTGCCG TTTGGTGTCG CAGGCGAGAT ACACGTGGGC GGTGCCGGTG TAGCACGGGG
 L  L  P   F  G  V  A   G  E  I   H  V  G   G  A  G  V   A  R  G

TTACCTCAAT CGCGAAGACT TAACCGCCAG CCAGTTTATC GATAATCCAC TAGCCGTGGG
 Y  L  N   R  E  D  L   T  A  S   Q  F  I   D  N  P  L   A  V  G
                                     A7
ATCCAACGGT GAAAAGCTGT ATAAAACCGG CGATTTGGGT CGCATTCGCG AAGACGGCAT
 S  N  G   E  K  L  Y   K  T  G   D  L  G   R  I  R  E   D  G  I
                        A8
AGTGGAATTC TTAGGTCGTA TCGATAAC
 V  E  F   L  G  R  I   D  N
```

Figure 29 Nucleic acid and deduced amino acid sequences of partial A domain of NRPS of clone 374605 - conserved core motifs of A domain of NRPS are highlighted in gray.

The alignment scores of deduced amino acid sequences of partial A domains from metagenomic library were shown in Table 9. Theses sequences could be divided into three groups [clone form pool 36; clone 37a (371001, 371002 and 374605) and clone 374601], according to their identity. Among the sequences from pool 36, most of them had 99% similarity to one another. Amino acid sequences of clone 360305 and 361102 had 100% identity. Although their amino acid sequences were identical, alignment of their nucleotide sequence showed one base difference. These high similarities among clones from pool 36 indicated that these sequences derived from the same A domain. Streak for isolation for a single colony was

Table 9 Percent identity of deduced amino acid sequences of partial A domain of NRPSs from clone 36 and 37 of metagenomic library. Total number of amino acid in each sequence (323-332) was shown in Figure 2B in Appendix B.

360305	100										
360310	99	100									
360312	99	99	100								
360802	99	99	99	100							
360804	99	99	99	99	100						
361101	99	99	99	99	99	100					
361102	100	99	99	99	99	99	100				
371001	30	30	30	30	30	31	30	100			
371002	29	30	29	29	29	30	29	93	100		
374601	32	32	31	32	32	32	32	47	44	100	
374605	31	31	30	31	31	31	31	99	93	48	100
Clone	**360305**	**360310**	**360312**	**360802**	**360804**	**361101**	**361102**	**371001**	**371002**	**374601**	**374605**

necessary because of contamination of PCR-positive clone during the isolation of transformants. This resulted in collection of redundant clone. The minor difference in nucleotide sequences might arise from PCR or cloning process. According to their identity scores, sequences from pool 37 were separated into two groups: 37a (371001, 371002 and 374605 which showed 93-99% identity) and 374601 which showed <50% identity to 37a.

All of the NRPSs from clone from pool 36 and 37 were found to consist of partial A2, complete A3-A7 and partial A8 conserved core sequences based on highly conserved core motifs (A1-A10) of adenylation domain of NRPS described by Marahiel *et al*, (1997) (Figure 19-29). Severn core motifs were rather conserved in the NRPS A domain. Core motifs data of the sequences in the same group (as divided by sequences alignment data) were identical. Amino acids sequences in each group were searched for the similar sequences by blastx (Table 10).

Phylogenetic analysis of A domains of clones from pool 36 placed them as sister clade of *Bacillus* in Phylum Firmicutes, as shown in Figure 30. This suggested that novel *nrps* gene encoding A domains in clone from pool 36 was evolutionary related to NRPS of *Bacillus* with boostrap value of 64.4%. Phylogenetic analysis of A domain of clone 37a placed them as sister clade of Cyanobacteria, Chloroflexi and Proteobacteria with boostrap support of 47.9%, as shown in Figure 31. This suggested that their NRPSs were more evolutionary related than those from Firmicutes. Phylogenetic analysis of A domain of clone 374601 placed it in the clade of Proteobacteria with boostrap value of 44.1%, as shown in Figure 32. These data suggested that novel *nrps* gene encoding A domain in clones from pool 36 and 37 were diverse. Summary data of phylogenetic analysis were summarized in Table 11. It was noticeable that NRPSs having evolutionary related to these clones and NRPSs having maximum % identity (Cyanobacteria, Table 10) were from bacteria in different phylum.

Table 10 Summary data of species with maximum score retrieved from similarity search of NRPS sequences from Klongkone mangrove soil

Clone	Known species with maximum score	% identity	% positive	Name
360305	Cyanobacteria (*Nostoc* sp. MV6)	47 (155/329)	61 (202/329)	NRPS
360310	Cyanobacteria (*Nostoc* sp. MV6)	47 (155/329)	61 (202/329)	NRPS
360312	Cyanobacteria (*Nostoc* sp. MV6)	46 (150/326)	61 (206/326)	NRPS
360802	Cyanobacteria (*Nostoc* sp. MV6)	46 (150/326)	61 (206/326)	NRPS
360804	Cyanobacteria (*Nostoc* sp. MV6)	47 (155/329)	61 (202/329)	NRPS
361101	Cyanobacteria (*Nostoc* sp. MV6)	47 (155/329)	61 (202/329)	NRPS
361102	Cyanobacteria (*Nostoc* sp. MV6)	47 (155/329)	61 (202/329)	NRPS
371001	Cyanobacteria (*Cyanothece* sp.)	43 (145/333)	59 (199/333)	NRPS
371002	Cyanobacteria (*Cyanothece* sp.)	43 (145/333)	59 (199/333)	NRPS
374601	Cyanobacteria (*Nostoc punctiform*)	48 (161/333)	63 (213/333)	NRPS
374605	Cyanobacteria (*Cyanothece* sp.)	43 (145/333)	59 (199/333)	NRPS

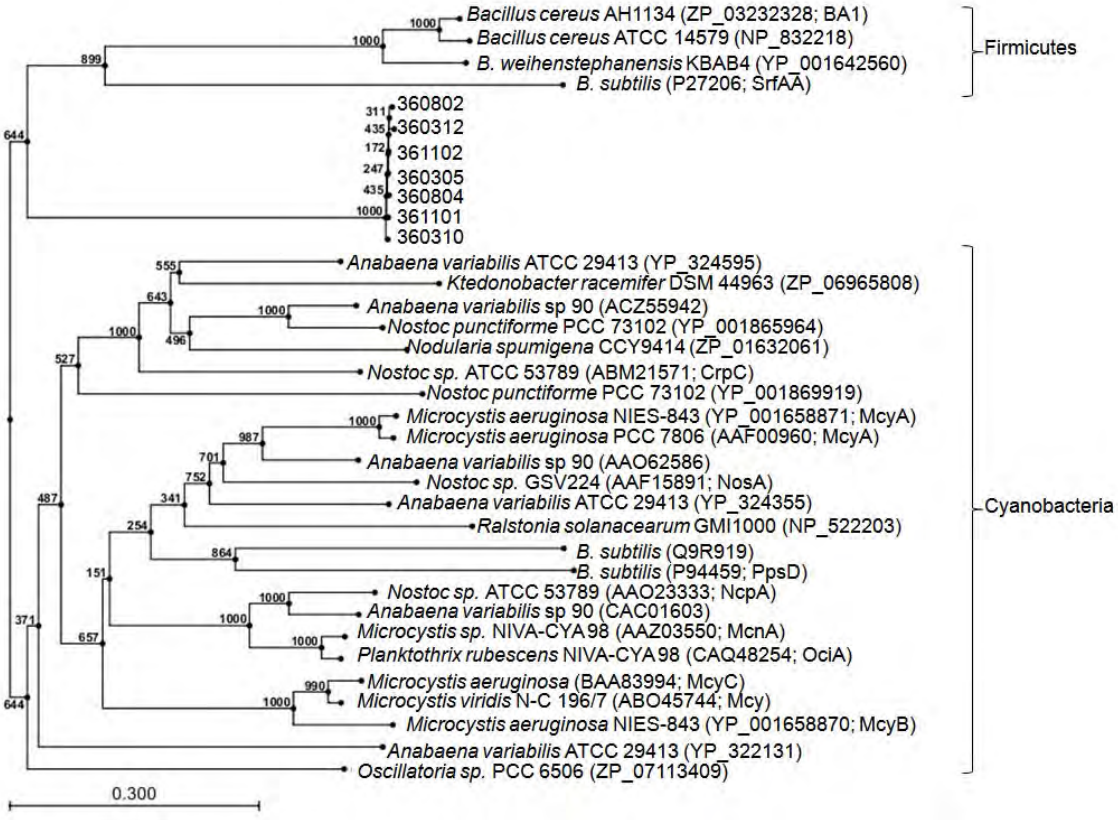

Figure 30 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPSs of clones from pool 36 and 28 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.30 amino acid substitution per position. Abbreviations; BA1: bacitracin synthetase 1; SrfAA: surfactin synthase subunit 1; CrpC: cryptophycins synthetase; McyA: microcystin synthetase subunit A; NosA: nostopeptolides synthesis enzyme subunit A; PpsD: plipastatin synthase subunit D; NcpA: nostocyclopeptide synthetase A; McnA: *Microcystis* cyanopeptolin synthesis enzyme; OciA: *Planktothrix* cyanopeptolin synthesis enzyme subunit A; McyC: microcystin synthetase subunit C; ZP_07113409: D-alanine-poly(phosphoribitol) ligase

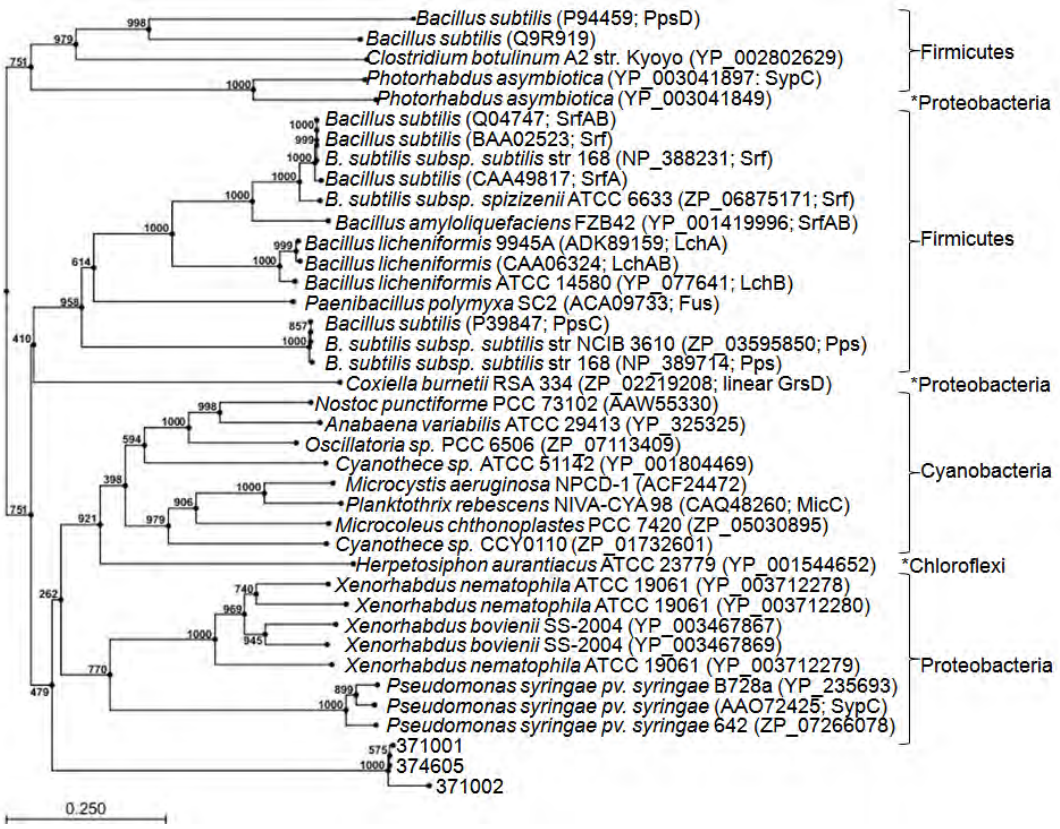

Figure 31 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of clone 37a (371001, 371002 and 374605) from pool 37 and 36 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.25 amino acid substitution per position. Abbreviations; PpsD: plipastatin synthase subunit D; SypC: syringopeptin synthetase subunit C; SrfAB: surfactin synthase subunit 2; SrfA: surfactin synthetase; LchA: lichenysin A synthetase; LchAB: lichenysin A synthetase in *Bacillus*; LchB: lichenysin B synthetase; Fus: fusaricidin synthetase; PpsC: plipastatin synthase subunit C; Pps: plipastatin synthase; GrsD: gramicidin synthetase subunit D; MicC: microginin synthetase; YP_003712280, YP_003467867: phenylalanine racemase; YP_003467869: ornithine racemase

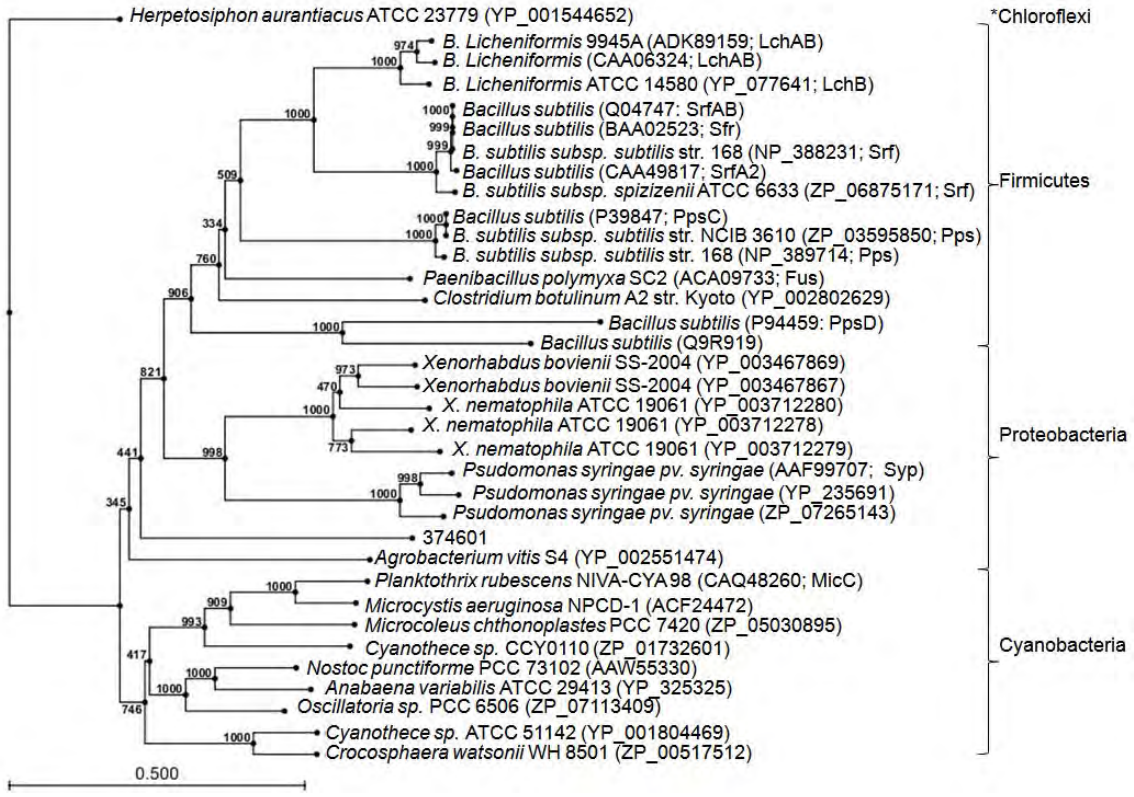

Figure 32 Neighbour-joining phylogenetic tree generated from the deduced amino acid sequences of partial A domain of NRPS of clone 374601 from pool 37 and 34 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.50 amino acid substitution per position. Abbreviations, LchA: lichenysin A synthetase; LchAB: lichenysin A synthetase in *Bacillus*; LchB: lichenysin B synthetase; SrfAB: surfactin synthase subunit 2; Srf: surfactin synthase; PpsC: plipastatin synthase subunit C; Fus: fusaricidin synthetase; PpsD: plipastatin synthase subunit D; Syp: syringopeptin synthetase; MicC: microginin synthetase; YP_003467869: ornithine racemase; YP_003467867, YP_003712280: phenylalanine racemase

Table 11 Summary data of phylogenetic analysis of A domain of NRPS amplified from soil metagenomic library.

Transformant	Maximum % identity to known species			Related Phylum (% Boostrap value)	Figure of Neighbour-joining tree
	Nonredundant proteins	Reference proteins	Swissprot proteins		
360305	47	46	44	Firmicutes (64.4)	Figure 30
360310	47	46	44		Figure 30
360312	47	46	43		Figure 30
360802	47	46	43		Figure 30
360804	43	43	43		Figure 30
361101	47	46	44		Figure 30
361102	47	46	44		Figure 30
371001	44	44	43	Cyanobacteria, Proteobacteria and Firmicutes (47.9)	Figure 31
371002	38	38	38		Figure 31
374605	44	44	44		Figure 31
374601	49	49	47	Proteobacteria (44.1)	Figure 32

All sequences used in phylogenetic analysis in this study were used for constructing a new phylogenetic tree in order to view the overall image of sequence grouping. Based on the rectangular cladogram shown in Figures 33A and 33B, NRPSs from Klongkone mangrove soil metagenome were divided into two large groups: A, SM23, SM27 and clones 37; B, SM2_2, SM3, SM48 and clones 36. Summary of Phyla of bacteria having NRPS evolutionary related to NRPSs obtained from Klongkone mangrove soils analyzed separately and totally as shown in Table 12. Most of them were correlated with one another.

Most of them were correlated. The abbreviation of protein products used in Figure 32 was listed in the following. PpsD: plipastatin synthase subunit D; Srf: surfactin synthetase; SrfAB, SrfA2: surfactin synthetase subunit 2; LchAB: lichenysin A synthetase in *Bacillus*; LchA: lichenysin A synthetase; LchB: lichenysin B synthetase; Fus: fusaridin synthetase; MassB: massetolide A synthetase; Syp: syringopeptin synthetase; SypC: syringopeptin synthetase subunit C; YP_003467869, YP_003467869: putative Ornithine racemase; YP_003467867, YP_003712280: Phenylalanine racemase; ZP_07113409: D-alanine--poly(phosphoribitol) ligase; MicC: microginin synthetase; OciC: *Planktothrix* cyanopeptolin synthesis enzyme subunit C; McnE: *Microcystis* cyanopeptolin synthesis enzyme subunit E; CrpC: cryptophycins synthetase; NosA: nostopeptolides synthetase subunit A; McyA: microcystin synthetase subunit A; McyC: microcystin synthetase subunit C; McyB: microcystin synthetase subunit B; NcpA: nostocyclopeptide synthetase; McnA: *Microcystis* cyanopeptolin synthesis enzyme; OciA: *Planktothrix* cyanopeptolin synthesis enzyme subunit A; GrsB: gramicidin S synthase subunit B; SrfA: surfactin synthetase subunit 1; PpsA: plipastatin synthase subunit A; PchD: pyochelin synthetase subunit D; PchF: pyochelin synthetase subunit F, HctE: hectochlorin synthetase; Ybt; yersiniabactin synthetase; GrsC: linear gramicidin synthetase subunit C; PvdJ(2): pyoverdine synthase; SypB: syringopeptin synthetase subunit B; ArfB: arthrofactin synthetase subunit B; BA1: bacitracin synthetase 1

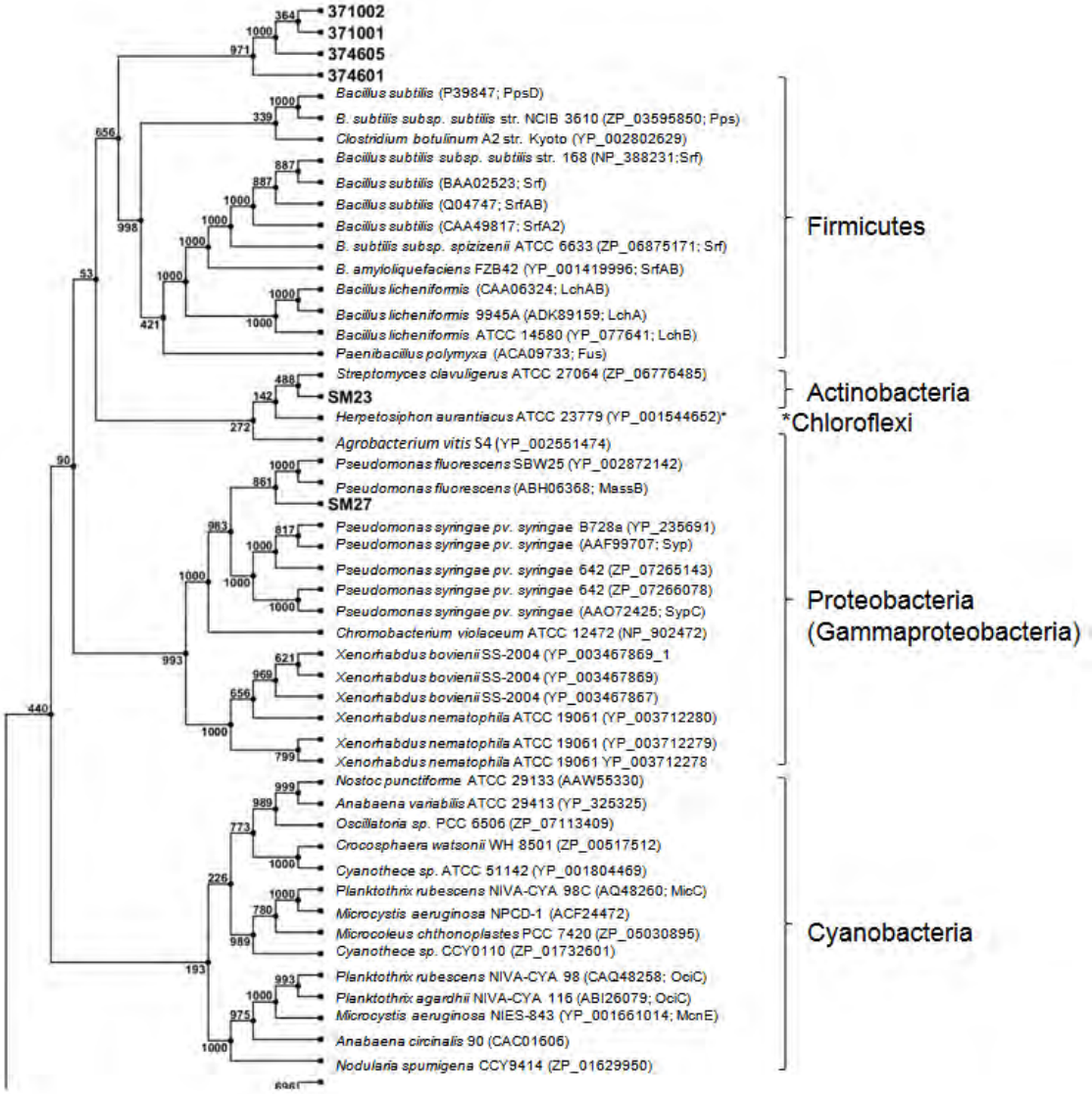

Figure 33A Rectangular cladogram from neighbour-joining algorithm generated from the deduced amino acid sequences of all NRPS sequences in this study (Part1) and 114 known taxa having similar conserved core motifs. The numbers at internal node indicate the bootstrap values from 1,000 bootstrap replications. The number in brackets is the GenBank accession numbers of the reference amino acid sequences. The scale bar represents 0.30 amino acid substitution per position.

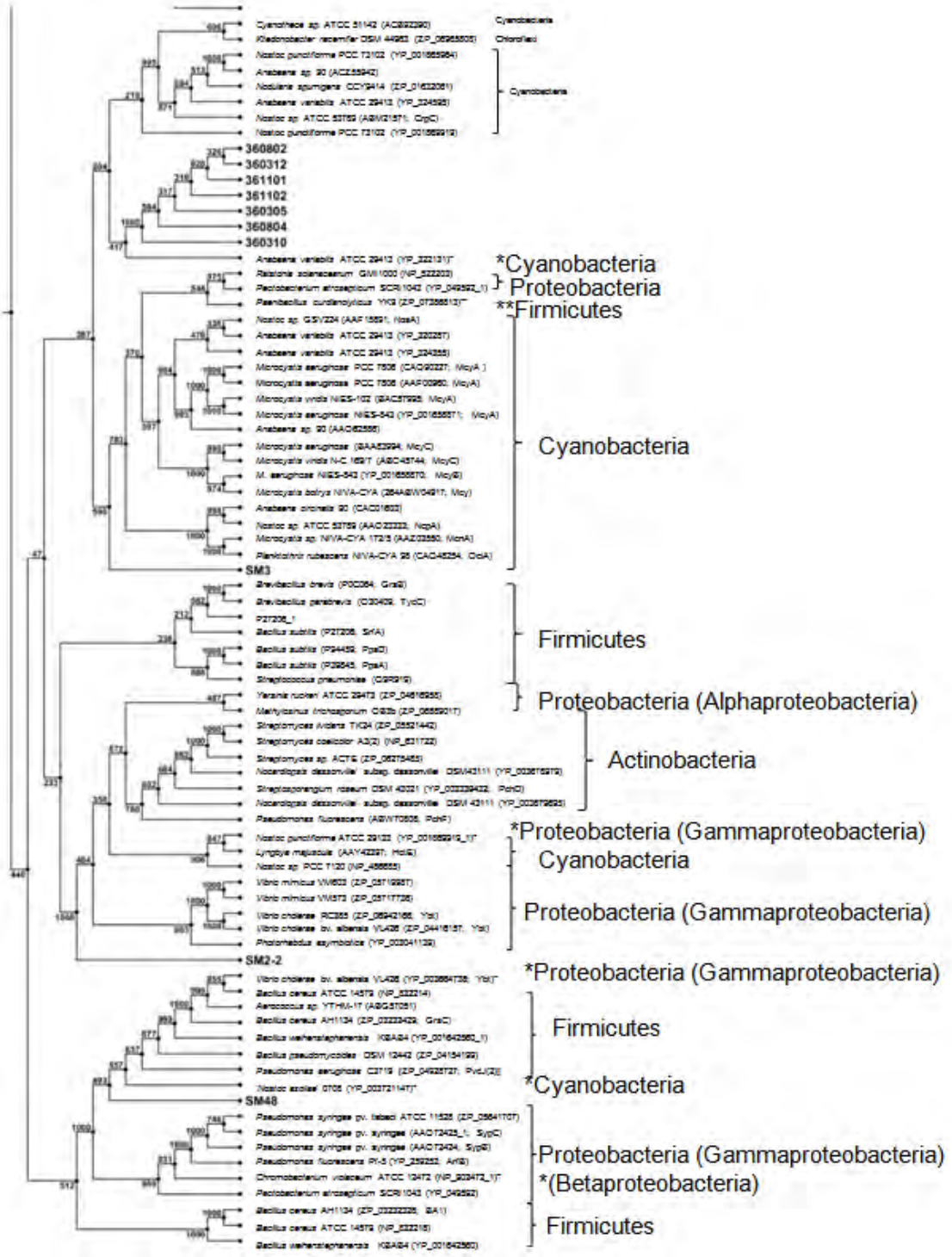

Figure 33B Rectangular cladogram from neighbour-joining algorithm generated from the deduced amino acid sequences of all NRPS sequences in this study (Part2).

Table 12 Phyla of bacteria having NRPS evolutionary related to NRPSs obtained from Klongkone mangrove soil metagenome.

Clone	**BLAST (% Identity)**	**Separate phylogenetic analysis (Bootstrap value)**	**Phylogenetic analysis of all sequences (Bootstrap value)**
SM2_2	Proteobacteria (52)	Cyanobacteria (62.7)	Proteobacteria, Cyanobacteria and Actinobacteria (100)
SM3	Cyanobacteria (52)	Cyanobacteria (57.5)	Cyanobacteria (59.0)
SM23	Proteobacteria (57)	Actinobacteria (85.6)	Actinobacteria (48.8)
SM27	Proteobacteria (57)	Proteobacteria (83.2)	Proteobacteria (99.3)
SM48	Cyanobacteria (62)	Cyanobacteria (43.4)	Cyanobacteria, Firmicutes and Proteobacteria (100)
36*	Cyanobacteria (47)	Firmicutes (64.4)	Cyanobacteria (41.7)
37a**	Cyanobacteria (43)	Cyanobacteria Proteobacteria and Firmicutes (79.1)	Firmicutes (65.6)
374601	Cyanobacteria (48)	Proteobacteria (44.1)	Firmicutes (65.6)

*Clone 36 refers to 360305, 360310, 360312, 360802, 360804, 361101 and 361102.;

**Clone 37a refers to 371001, 371002 and 374605. % identity (for BLAST analysis) and Bootstrap value (for phylogenetic analysis) were showed in parenthesis.

4.4.2.2 Predicted amino acid which is activated by A domain of NRPS from metagenomic library

The prediction of amino acid activated by the aminoacyl domain of NRPS was processed as indicated in section 3.5.3 in methodology. Summary data of the prediction of amino acid activated by partial A domain of NRPS from metagenomic library were shown in Table 13. Data of predicted amino acids divided NRPS sequences into three groups correlated with data from phylogenetic analysis. A domain in each group recognized and activated different amino acid substrates. As shown in Table 13, A domains in clones 371001, 371002 and 374605 were predicted to activate the same amino acid, tryptophan. In addition, amino acid sequence identities of clones 371001 and 371002, clones 371001 and 374605, and clones 371002 and 374605, were 93%, 99%, and 93%, respectively, as shown in Table 9.

Table 13 Summary data of the prediction of amino acid activated by A domain of NRPS from clone 36 and 37 of metagenomic library.

NRPS	Residues in the Binding pocket	Predicted amino acid
360305, 360310, 360312, 360802, 360804, 361101, 361102	DAKCLGLV	SrfAA-Ml-Glu/Asp
371001, 371002, 374605	DAFXIGAV	Cdal-M3-Trp
374601	DAFXLGGT	TycC-M4-Val

SrfAA: Surfactin synthetase; CdaI: CDA peptide synthetase; TycC: Tyrocidine synthase subunit C; Glu: Glutamic acid; Asp: Aspartic acid; Thr: Threonine; Val: Valine

To verify (a) The possibility that A domains of clones 371001, 371002 and 374605, would be in the same peptide synthetase and (b) The possibility that A domains of clones 371001 and 374605 would be the same A domain, analyses of five known peptide synthetases in NCBI database were attempted to determine variation in the identities of A domain activating the same and different amino acids. As shown in Table 14, the amino acid sequence identities of two A domains activating tryptophan and three A domains activating asparagine of CDA peptide synthetase from *Streptomyces coelicolor* A3(2) were 66% and 88-92%, respectively. Sequence identities of A domains activating different amino acids were found to be 33-52%. For cyclosporine synthetase from *Tolypocladium inflatum* ATCC 34921, the sequence identities of two A domains activating alanine, three A domains activating leucine and two A domains activating valine were 54%, 61-62%, and 71%, respectively, as shown in Table 15. Sequence identities of A domains activating different amino acids were found to be 49-64%. For gramicidin S synthetase from *Paenibacillus polymyxa* E681, the sequence identities of three A domains activating glutamic acid and two A domains activating valine were 38-61% and 41%, respectively, as shown in Table 16. Sequence identities of A domains activating different amino acids were found to be 35-76%. It was found that in plipastatin synthase from *Bacillus subtilis* three A domains activating glutamic acid and two A domains activating tyrosine showed identities of 56-95% and 53%, respectively, as shown in Table 17. Sequence identities of A domains activating different amino acids were found to be 30-49%. Two A domains activating leucine in surfactin synthetase from *Bacillus subtilis* subsp. subtilis str. 168 showed identity of 98%, as shown in Table 18. Sequence identities of A domains activating different amino acids were found to be 33-41%. Moreover, sequence identities of A domains activating valine from different peptide synthetases were analyzed. It was found that they were 30-67%, as shown in Table 19.

Table 14 Percent identity of A domains of CDA peptide synthetase from *Streptomyces coelicolor* A3(2)

1. Cda1-A1-Ser	100										
2. Cda1-A2-Thr	44	100									
3. Cda1-A3-Trp	48	42	100								
4. Cda1-A4-Asp	45	41	42	100							
5. Cda1-A5-Asp	45	42	43	92	100						
6. Cda1-A6-HPG	42	41	39	34	34	100					
7. Cda2-A1-Asp	44	41	44	88	90	34	100				
8. Cda2-A2-Gly	50	41	44	41	41	37	41	100			
9. Cda2-A2-No Hit	38	36	43	48	49	36	48	38	100		
10. Cda3-A1-3-Me-Glu	41	38	42	52	51	33	50	37	42	100	
11. Cda3-A2-Trp	45	46	66	38	39	40	40	42	43	38	100
	1	2	3	4	5	6	7	8	9	10	11

Table 15 Percent identity of A domains of cyclosporine synthetase from *Tolypocladium inflatum* ATCC 34921 **(**Bmt: (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine**)**

1. CssA-A1-Ala	100										
2. CssA-A2-No Hit	59	100									
3. CssA-A3-Leu	56	64	100								
4. CssA-A4-Val	54	56	54	100							
5 CssA-A5-Bmt	59	59	58	57	100						
6. CssA-A6-No Hit	54	54	51	50	52	100					
7. CssA-A7-No Hit	57	56	59	57	54	51	100				
8. CssA-A8-Leu	57	64	61	56	57	53	56	100			
9. CssA-A9-Val	58	57	57	71	57	50	57	55	100		
10. CssA-A10-Leu	55	61	62	55	57	50	54	61	56	100	
11. CssA-A11-Ala	54	54	55	59	55	49	57	57	58	53	100
	1	2	3	4	5	6	7	8	9	10	11

Table 16 Percent identity of A domains of gramicidin S synthetase I and II from *Paenibacillus polymyxa* E681

1. GrsA-A1-Glu	100								
2. GrsA-A2-Val	41	100							
3. GrsA-A3-Val	76	41	100						
4. GrsA-A4-Tyr	74	40	73	100					
5. GrsA-A5-Glu	39	38	37	38	100				
6. GrsB-A1-Glu	61	42	61	62	38	100			
7. GrsB-A2-Thr	46	41	45	45	39	48	100		
8. GrsB-A3-No Hit	39	42	40	39	35	43	39	100	
9. GrsB-A4-No Hit	39	42	40	40	35	43	40	96	100
	1	2	3	4	5	6	7	8	9

Table 17 Percent identity of A domains of plipastatin synthase from *Bacillus subtilis*

1. PpsA-A1-Glu	100									
2. PpsA-A2-No Hit	32	100								
3. PpsB-A1-Tyr	38	33	100							
4. PpsB-A2-Thr	38	36	36	100						
5. PpsC-A1-Glu	95	32	37	39	100					
6. PpsC-A2-Val	37	33	36	39	37	100				
7. PpsD-A1-Pro	31	30	32	34	31	35	100			
8. PpsD-A2-Glu	56	32	37	34	56	39	30	100		
9. PpsD-A3-Tyr	38	34	53	40	38	36	33	38	100	
10. PpsE-A1-Ile	35	34	35	42	35	49	33	35	34	100
	1	2	3	4	5	6	7	8	9	10

Table 18 Percent identity of A domains of surfactin synthetase from *Bacillus subtilis* subsp. *subtilis* str. 168

1. SrfAA-A1-Glu/Asp	100					
2. SrfAA-A2-Leu/Ile/Val	36	100				
3. SrfAA-A3-Leu	41	36	100			
4. SrfAB-A1-Val	38	40	36	100		
5. SrfAB-A2-Asp	34	34	36	33	100	
6. SrfAB-A3-Leu	40	36	98	36	35	100
	1	2	3	4	5	6

Table 19 Percent identity of A domains which activate Valine

1. LchB-A1-Val (*B. licheniformis*)	100						
2. PpsC-A2-Val (*B. subtilis*)	51	100					
3. SrfA-A1-Val (*B. subtilis*)	67	50	100				
4. TycC-A4-Val (*Br. brevis*)	60	52	58	100			
5. GrsB-A2-Val (*P. asymbiotica*)	36	32	36	34	100		
6. SypC-A1-Val (*P. asymbiotica*)	44	39	45	43	35	100	
7. CssA-A4-Val (*T. inflatum*)	34	33	33	33	30	35	100
	1	2	3	4	5	6	7

B. licheniformis: Bacillus licheniformis, Br. brevis: Brevibacillus brevis, P. asymbiotica: Photorhabdus asymbiotica, T. inflatum: Tolypocladium inflatum

In summary, (a) A domains activating the same amino acid in each peptide synthetase could have sequence identity as low as 38% or as high as 98%, (b) A domains activating the different amino acid in each peptide synthetase could have sequence identity in the range of 30-76%, (c) A domains activating the same amino acid from different peptide synthetases were only 30-67%. These results suggested that A domains activating tryptophan in clones 371001, 371002, 374605 were tentatively different from A domains in the same peptide synthetase. A domain activating valine in clone 374601 and three A domains activating tryptophan (44-48% identity) might be in the same peptide synthetase. A domain activating Glu/Asp in clone 36 and four A domains in clones 37 (29-31% identity) might or might not be in the same peptide synthetase.

Sequence-based screening of mangrove soil metagenomic for novel *nrps* genes by MTF2/MTR PCR primer pair revealed the PCR-positive clones from pool 36 and 37. PCR products amplified from those clones were cloned into pGEM®-T Easy vector, giving several clones concluded in Table 8. Sequencing results from pool 36 and 37 clones were used for further analysis. Data from sequence alignment of amino acids divided these sequences into three different groups, clone from pool 36, clone 37a and clone 374601. The similarity search by BLAST and phylogenetic analysis data suggested the novelty of A domains of *nrps* in mangrove soil metagenomic library. These data could be used for further research of these *nrps* containing clones. Primer walking or other DNA sequencing method of clone 367310, 3710 and 3746 should be performed in order to investigate the rest sequences of inserted DNA. Data of entire *nrps* fragment can be used for prediction of NRPS protein architecture, function or structure of peptide product. Sequence-based screening of the rest of library was recommended because this may discover more bioactive compound producing genes.

CHAPTER V

CONCLUSION

Due to several problems from antimicrobial resistance microorganisms, many strategies are developed and applied for novel antibiotics discovery. Metagenomics is among the potential tools for screening new genes that could produce bioactive compounds from unculturable microorganisms from natural samples including soil. Mangrove soil is one of the extreme habitats. It harbors several medically important microorganisms including cyanobacteria and actinobacteria. In this research, mangrove soil was collected from Klongkone mangrove, Samutsongkhram province in September 2008. Mangrove soil metagenomic DNA was extracted by direct extraction method and purified by agarose gel electrophoresis and dialysis yielding DNA with a suitable size for fosmid cloning. Soil metagenomic DNA was then screened for *nrps* gene by PCR using MTF2/MTR primer pair which is specific for adenylation domain (A domain) of *nrps* gene. Cloning of PCR products from soil DNA amplification resulted in five partial A domains designated as SM2_2, SM3, SM23, SM27 and SM48. The sequence alignment of these deduced amino acid sequences demonstrated that they were different from one another. Each sequence was blasted against non-redundant protein sequences, reference proteins and swissprot protein sequences databases of NCBI using blastx program. Blasting results suggested the novelty of those five NRPSs (Table 4, Chapter 3). Hit species from blasting which showed the similar conserved core sequences with each of A domain were used for phylogenetic analysis. According to phylogenetic analysis, partial A domains derived from PCR products of mangrove soil were evolutionary related to NRPSs of cyanobacteria, actinobacteria and proteobacteria. The prediction of amino acid which is activated by A domain of SM2_2, SM3 and SM27 NRPSs showed no hit amino acid (Table 6, Chapter 3). A domain of SM23 NRPS was predicted to activate valine

similar to tyrocidine synthetase 3 of *Brevibacillus brevis*. A domain of SM48 NRPS was predicted to activate threonine similar to coelichelin synthetase of *Streptomyces coelicolor*. These results suggested that Klongkone mangrove soil had novel and diverse *nrps*.

The DNA from this soil sample was used to construct the metagenomic library using CopyControl™ Fosmid Library Production Kit with pCC2FOS™ vector and *E. coli* strain DH5α host cell. The construction of library resulted in approximately 14,000 clones of metagenomic library divided into 95 pools. Approximately 4,300 clones from 29 pools (1-20, 41-44, 91-95) were screened for the activity against *Candida albicans* ATCC 90028 and chloramphenicol resistance *Bacillus subtilis*. No clone with activity was retrieved from the functional screening. Therefore, the library was screened by sequence-based approach. At least 1,500 clones from ten pools (30-39) were screened by PCR using MTF2/MTR primer pair. Clones 3671310-03, 3671310-08, 3671310-11, 3671314-25, 3710 and 3746 were discovered from the screening. PCR products of these six clones were cloned into pGEM®-T Easy vector, resulted in eleven clones designated as 360305, 360310, 360312, 360802, 360804, 361101, 361102, 371001, 371002, 374601 and 374605. DNAs from these clones were sequenced for further analysis. The results from blast searching against protein sequences databases in NBCI suggested the novelty of these NRPSs. These eleven sequences were separated into three groups based on sequence similarity from sequence alignment data. The data from activated amino acid prediction also divided these eleven NRPSs into three groups which activate glutamine/aspartic acid, tryptophan and valine (like SM23). The data revealed that clone 3671310, 3710 and 3746 contained different novel *nrps*. All sequences of NRPSs obtained in this study were used for constructing new phylogenetic tree. Summary of Phyla of bacteria having NRPS evolutionary related to NRPSs obtained from Klongkone mangrove soil analyzed separately and totally exhibited that most of them were correlated.

In conclusion, Klongkone mangrove soil metagenome is a potential source of novel and diverse *nrps* genes. Further study of clone 3671310, 3710 and 3746 should be conducted to get

complete *nrps* gene that tentatively encodes NRPS for synthesis of novel bioactive peptide. It is necessary to explore the rest of metagenomic library to discover the hidden *nrps* genes which may lead to the discovery of novel *nrps* genes for novel bioactive compounds.

REFERENCES

Agoramoorthy, G., Chen, F.A., and, Hsu, M.J. 2007. Threat of heavy metal pollution in halophytic and mangrove plants of Tamil Nadu, India. Environment. Pollut. 115(2): 320-326.

Aislabie, J.M., Jordan, S., and Barker, G.M. 2008. Relation between soil classification and bacterial diversity in soils of the Ross Sea region, Antarctica. Geoderma. 144: 9–20.

Alexander, M. 1964. Biochemical ecology of soil microorganism. Annu. Rev. Microbiol. 18: 217-250.

Alvarez, A. M., *et al*. 1990. A strain of *Nocardia mediterranei* that produces a mixture of Rifamycin B and W. Biotechnol. Lett. 12(4): 283-288.

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(1): 143–169.

Ananda, K., and Sridhar, K.R. 2002. Diversity of endophytic fungi in the roots of mangrove species on the west coast of India. Can. J. Microbiol. 48(10): 871-878.

Arbeli, Z., and Fuentes, C.L. 2007. Improved purification and PCR amplification of DNA from environmental samples. FEMS Microbiol. Lett. 272: 269–275.

Ayuso-Sacido, A., and Genilloud, O. 2005. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: Detection and distribution of these biosynthetic gene sequences in major taxonomic groups. Microb. Ecol. 49(1): 10–24.

Bachmann, B.O., and Ravel, J. 2009. In silico prediction of microbial secondary metabolic pathways from DNA sequence data. Methods. Enzymol. 458: 181-217.

Bakken, L.R. 1985. Separation and purification of bacteria from soil. Appl. Environ. Microbiol. 49(6): 1482-1487.

Baldock, J.A., and Skjemstad, J.O. 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. Org. Geochem. 31: 697-710.

Balestra, G. M., and Misaghi, I.J. 1997. Increasing the efficiency of the plate counting method for estimating bacterial diversity. J. Microbiol. Methods. 30(2): 111-117.

Baltz, R. H. 2008. Renaissance in antibacterial discovery from actinomycetes. Curr. Opin. Pharmacol. 8(5): 557-563.

Barrios-Llerena, M.E, Burja, A.M., and Wright, P.C. 2007. Genetic analysis of polyketide synthase and peptide synthetase genes in cyanobacteria as a mining tool for secondary metabolites. J. Ind. Microbiol. Biotechnol. 34:443–456.

Beja, O., *et al*. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ. Microbiol. 2(5): 516-529.

Beja, O., *et al*. 2002. Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. Appl. Environ. Microbiol. 68(1): 335-345.

Berenguer, B., *et al*. 2006. Protective and antioxidant effects of *Rhizophora mangle* L. against NSAID-induced gastric ulcers. J. Ethnopharmacol. 103(2): 194-200.

Bergmann, W., and Burke, D.C. 1955. Contributions to the study of marine products. XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. J. Org. Chem. 20: 1501 – 1507.

Bergmann, W., and Feeney, R.F. 1950. The Isolation of a new thymine pentoside from sponges. J. Am. Chem. Soc. 72: 2809 – 2810.

Brady, S.F., *et al*. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org. Lett. 3: 1981-1984.

Bredholt, H., *et al*. 2008. Actinomycetes from sediments in the trondheim fjord,Norway: Diversity and biological activity. Mar. Drugs. 6(1): 12–24.

Bushley, K.E., and Turgeon, B.E. 2010. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC. Evol. Biol. 10: 26-49.

Bushley, K.E., Ripoll, D.R., and Turgeon, G. 2008. Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. BMC. Evol. Biol. 8: 328-352.

Caboche, S., *et al*. 2008. NORINE: a database of nonribosomal peptides. Nucl. Acids. Res. 36: D326-D331.

Cadel-Six, S., *et al*. 2008. Halogenase genes in nonribosomal peptide synthetase gene clusters of *Microcystis* (Cyanobacteria): Sporadic distribution and evolution. Mol. Biol. Evol. 25(9): 2031–2041.

Canestri, V., and Ruiz, O. 1973. The *destruction of mangroves.* Mar. Poll. Bull. 4(12): 183–185.

Castle, J. W., and Rodgers, J. H. Jr. 2009. Hypothesis for the role of toxin-producing algae in Phanerozoic mass extinctions based on evidence from the geologic record and modern environments. Environ. Geosci. 16: 1-23.

Chakrabarti, R. and Schutt, C. E. 2001. The enhancement of PCR amplification by low molecular-weight sulfones. Gene. 274(1-2): 293–298.

Challis, G.L., and Naismith, J.H. 2004. Structural aspects of non-ribosomal peptide biosynthesis. Curr. Opin. Struct. Biol. 14:748–756.

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.

Chen, K., and Pachter, L. 2005. Bioinformatics for whole-genome shotgun sequencing of microbial communities. PloS. Comp. Biol. 1: 24.

Chen, W.L., *et al*. 2009. A novel marine compound xyloketal B protects against oxidized LDL-induced cell injury in vitro. Biochem. Pharmacol. 78(8): 941-950.

Cheplick, G. P., and Clay, K. 1988. Acquired chemical defences in grasses: the role of fungal endophytes. – Oikos. 52: 309-318.

Chrisoffersen, R.E. 2006. Antibiotics – an investment worth making? Nat. Biotechnol. 24: 1512-1514.

Chronakova, A., *et al*., 2010. Biodiversity of streptomycetes isolated from a succession sequence at a post-mining site and their evidence in Miocene lacustrine sediment. Microbiol. Res. 165(7): 594-608.

Committee on Metagenomics: Challenges and functional applications, National Research Council. 2007. The New science of metagenomics: Revealing the secrets of our microbial planet. Washington: The national academics press.

Couto, G.H. *et al*. 2010**.** Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. Genet. Mol. Res**.** 9(1): 514-523.

Dadenko, E. V., *et al*. 2009. Changes in the enzymatic activity of soil samples upon their storage. Eurasian. Soil. Sci. 42(12): 1380–1385.

Daniel, R. 2004. The soil metagenome – a rich resource for the discovery of novel natural products. Curr. Opin. Biotechnol. 15(3): 199-204.

Daniel, R. 2005. The metagenomics of soil. Nat. Rev. Microbiol. 3: 470-478.

Davidson, S.K., *et al*. 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "*Candidatus* Endobugula sertula" of the bryozoan *Bugula neritina*. Appl. Environ. Microbiol. 67: 4531 - 4537.

Davies, J. 2006. Where have all the antibiotics gone? Can. J. Infect. Dis. Med. Microbiol. 17(5): 287-290.

Davis, K. E. R., Joseph, S. J., and Janssen, P. H. 2005. Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Appl. Environ. Microbiol. 71(2): 826-834.

De Rosa, S., *et al*. 1995. A new bioactive eunicellin-type diterpene from the gorgonian *Eunicella cavolini*. Nat. Prod. Lett. 7: 259-265.

Den Burg, B.V., 2003. Extremophiles as a source for novel enzymes. Curr. Opin. Microbiol. 6: 213–218.

Doekel, S., *et al*. 2008. Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. Microbiology. 154: 2872–2880.

Donadio, S., Monciardini, P., and Sosio, M. 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat. Prod. Rep. 24: 1073–1109.

Dunlap, W.C. *et al*. 2007. Biomedicinals from the phytosymbionts of marine invertebrates: A molecular approach. Methods. 42: 358–376.

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

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids. Res. 32(5): 1792-1797.

Edwards, D.J., *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.

Ehrenreich, I.M., Waterbury, J.B., and Webb, E.A. 2005. Distribution and diversity of natural product genes in marine and freshwater Cyanobacterial cultures and genomes. Appl. Environ. Microbiol. 71(11): 7401-7413.

Ellison, J.C. 1998. Impacts of sediment burial on mangroves. Mar. Pollut. Bulletin. 37(8): 420-426.

Entcheva, E., *et al*. 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl. Environ. Microbiol. 67(1): 89–99.

Faber, M. 2006. Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? J. Ind. Microbiol. Biotechnol. 33: 507–513.

Fang, X., Jiang, B., and Wang, X. 2006. Purification and partial characterization of an acidic polysaccharide with complement fixing ability from the stems of *Avicennia marina*. J. Biochem. Mol. Biol. 39(5): 546-555.

Felnagele, E.A., *et al*. 2007. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol. Pharmaceutics. 5(2): 191-211.

Fenical, W., and Jensen, P.R. 2006. Developing a new resource for drug discovery: Marine actinomycete bacteria. Nat. Chem. Biol. 2(12): 666-673.

Ferreira, T.O., *et al*. 2007. Are mangrove forest substrates sediments or soils? A case study in southeastern Brazil. Catena. 70: 79-91.

Ferrer, M., *et al*. 2007. Mining enzymes from extreme environments. Curr. Opin. Microbiol. 10: 207–214.

Food and Agriculture Organization of the United Nations. 2007. The World's Mangroves 1980-2005. Rome: (n.p.).

Foster, R.C. 1988. Microenvironments of soil microorganisms. Biol. Fertil. Soils. 6: 189-203.

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(2): 153– 163.

Gall, Y. M., and Konashev, M. B. 2001. The discovery of Gramicidin S: The intellectual transformation of G.F. Gause from biologist to researcher of antibiotics and on its meaning for the fate of Russian genetics. Hist. Philos. Life. Sci. 23(1): 137-150.

Ganesan, A. 2008. The impact of natural products upon modern drug discovery. Curr. Opin. Chem. Biol. 12(3): 306-317.

Gans, J., Wolinsky, M., and Dunbar, J. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science. 309(5739): 1387–1390.

Gao, H., *et al*. 2010. New oxidized sterols from *Aspergillus awamori* and the endo-boat conformation adopted by the cyclohexene oxide system. Magn. Reson. Chem. 48(1): 38-43.

Gao, P., and Huang, Y. 2009. Detection, distribution, and organohalogen compound discovery implications of the reduced flavin adenine dinucleotide-dependent halogenase gene in major filamentous Actinomycete taxonomic groups**.** Appl. Environ. Microbiol. 75(14): 4813-4820

Ghosh, A. *et al*. 2010. Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. Saline. Syst. 6(1): 4-11.

Gillespie, D.E., *et al*. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl. Environ. Microbiol. 68: 4306-4310.

Girvan, M.S., *et al*. 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Appl. Environ. Microbiol. 69(3): 1800–1809.

Gonzalez, I., *et al*. 2005. Actinomycetes isolated from lichens: Evaluation of their diversity and detection of biosynthetic gene sequences. FEMS. Microbiol. Ecol. 54(3): 401-415.

Gulick, A.M. 2009. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of nonribosomal peptide synthetases, and firefly luciferase. ACS. Chem. Biol. 4(10): 811-827.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95-98.

Hallam, S. J., *et al*. 2003. Identification of methyl coenzyme M reductase A (*mcr* A) genes associated with methane-oxidizing archaea. Appl. Environ. Microbiol. 69(9): 5483–91.

Han, L., *et al*. 2004. New diterpenoids from the marine mangrove *Bruguiera gymnorrhiza*. J. Nat. Prod. 67(9): 1620-1623.

Han, L., Huang, X., Sattler, I., Dahse, H. Hongzheng, F., Grabley, S., Wenhan, and Lin, M. 2005a. Three new pimaren diterpenoids from marine mangrove plant, *Bruguiera gymnorrhiza*. Pharmazie. 60(9): 705-707.

Han, L., Huang, X., Sattler, I., Moellmann, U., Fu, H., Lin, W., and Grabley, S. 2005b. New aromatic compounds from the marine mangrove *Bruguiera gymnorrhiza*. Planta. Med. 71(2): 160-164.

Han. L., *et al*. 2007. Unusual naphthoquinone derivatives from the twigs of *Avicennia marina*. J. Nat. Prod. 70(6): 923-927.

Handelsman, J. 2004. Metagenomics: Application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68(4): 669–685.

Handelsman, J., *et al*. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5(10): R245–249.

Harvey, A.L. 2008. Natural products in drug discovery. Drug. Discov. Today. 13(19/20): 894-901.

Haygood, M.G., and Davidson, S.K. 1997. Small-subunit rRNA genes and in situ hybridization with oligonucleotides specific for the bacterial symbionts in the larvae of the bryozoan *Bugula neritina* and proposal of "*Candidatus* endobugula sertula". Appl. Environ. Microbiol. 63: 4612–4616.

Henne, A., *et al*. 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. 65(9): 3901- 3907.

Hildebrand, M., *et al*. 2004. Approaches to identify, clone, and express symbiont bioactive metabolite genes. Nat. Prod. Rep. 21: 122 – 142.

Hong, K., *et al*. 2009. Actinomycetes for marine drug discovery isolated from mangrove soils and plants in China. Mar. Drugs. 7(1): 24–44.

Hopwood, D. A., *et al*. 1985. Genetic manipulation of Streptomyces: A laboratory manual. Norwich: John Innes Foundation.

Hori, K., *et al*. 1989. Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. J. Biochem. 106(4): 639-645.

Howard, D., *et al*. 2001. Measuring the economic costs of antimicrobial resistance in hospital settings: Summary of the centers for disease control and prevention–Emory Workshop. Clin. Infect. Dis. 33:1573–1578.

Howells, J.D., and Albert, R. 1971. Fermentation Process for 9-(b-D-arabinofuranosyl) adenine, US Patent 3616208.

Huang, H., *et al*. 2007. Cyclic peptides from an endophytic fungus obtained from a mangrove leaf (*Kandelia candel*). J. Nat. Prod. 70(11): 1696-1969.

Hung, T., Mak, K., and Fong, K. 1999. A specificity enhancer for polymerase chain reaction. Nucleic. Acids. Res. 18(16): 4953.

Jensen, P.R., *et al*. 2005 Marine actinomycete diversity and natural product discovery. Antonie. van. Leeuwenhoek. 87: 43–48.

Jia, G.M., *et al*. 2005. Microbial biomass and nutrients in soil at the different stages of secondary forest succession in Ziwulin, northwest China. Forest. Ecol. Manag. 217: 117–125.

Jirakkakul, J., *et al*. 2008. Identification of the nonribosomal peptide synthetase gene responsible for bassianolide synthesis in wood-decaying fungus *Xylaria* sp. BCC1067. Microbiol. 154: 995–1006.

Jones, A.C., *et al*. 2009. New tricks from ancient algae: Natural products biosynthesis in marine cyanobacteria. Curr. Opin. Chem. Biol. 13: 216–223.

Jones, W.P., *et al*. 2005. Antitumour activity of 3-chlorodeoxylapachol, a naphthoquinone from *Avicennia germinans* collected from an experimental plot in southern Florida. J. Pharm. Pharmacol. 57(9): 1101-1108.

Kathiresan, K. 2003. Polythene and Plastics-degrading microbes from the mangrove soil. Rev. Biol. Trop. 51: 3-4.

Kathiresan, K., and Bingham, B. L. 2001. Biology of mangroves and mangrove ecosystems. Adv. Mar. Biol. 40: 81-251.

Kehriga, H.A., *et al*. 2003. Heavy metals and methylmercury in a tropical coastal estuary and a mangrove in Brazil. Org. Geochem. 34: 661–669.

Kennedy, J., Marchesi, J. R., and Dobson, A. D. W. 2007. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. Appl. Microbiol. Biotechnol. 75(1): 11-20.

Kessler, N., *et al*. 2004. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. J. Biol. Chem. 279: 7413-7419.

Khaldi, N., *et al*. 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. Genome. Biol. 9: R18-R28.

Kirk, J.L., *et al*. 2004. Methods of studying soil microbial diversity. J. Microbiol. Methods. 58: 169– 188.

Kjer, J., *et al*. 2009. Xanalteric Acids I and II and Related Phenolic Compounds from an Endophytic *Alternaria* sp. Isolated from the Mangrove Plant *Sonneratia alba*. J. Nat. Prod. 72(11): 2053–2057.

Kleinkauf, H. and Dohren, H.V. 1990. Nonribosomal biosynthesis of peptide antibiotics. Eur. J. Biochem. 192: 1-15.

Komiyama, A. Ong, J.E., and Poungparn, S. 2008. Allometry, biomass, and productivity of mangrove forests: A review. Aquat. Bot. 89: 128–137.

Kratzschmar, J., Krause, M., and Marahiel, M. A. 1989. Gramicidin S biosynthesis operon containing the structural genes grsA and grsB has an open reading frame encoding a protein homologous to fatty acid thioesterases. J. Bacteriol. 171(10): 5422–5429.

Kristensen, E., *et al*. 2008. Organic carbon dynamics in mangrove ecosystems: A review. Aquat. Bot. 89: 201–219.

Kumar, S., Hatha, A.A., and Christi, K.S. 2007. Diversity and effectiveness of tropical mangrove soil microflora on the degradation of polythene carry bags. Rev. Biol. Trop. 55(3-4): 777-786.

Labud, V., Garcia, C., and Hernandez, T. 2007. Effect of hydrocarbon pollution on the microbial properties of a sandy and a clay soil. Chemosphere. 66: 1863–1871.

Lal, R. 2005. Forest soils and carbon sequestration. Forest. Ecol. Manag. 220: 242–258.

Lam, K.L. 2006. Discovery of novel metabolites from marine actinomycetes. Curr. Opin. Microbiol. 9: 245–251.

Larkin M.A., *et al*. 2007. ClustalW and ClustalX version 2. Bioinformatics. 23(21): 2947-2948.

Lauber, C.L., *et al*. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human associated samples. FEMS. Microbiol. Lett. 307: 80–86.

Laurent-Martin, F., *et al*. 2001. DNA Extraction from Soils: Old bias for new microbial diversity analysis methods. Appl. Environ. Microbiol. 67(5): 2354–2359.

Lautru, S., and Challis, G.L. 2004. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. Microbiology. 150: 1629–1636.

Lavahun, M. F. E., Joergensen, R.G. and Meyer, B. 1996. Activity and biomass of soil microorganisms at different depths. Biol. Fert. Soils. 23(1): 38-42.

Le, S.Y. 2008. Mangrove macrobenthos: Assemblages, services, and linkages. J. Sea. Res. 59: 16–29.

Li, L., *et al*. 2006. Pongamone A-E, five flavonoids from the stems of a mangrove plant, *Pongamia pinnata*. Phytochemistry. 67(13):1347-1352.

Li, X., and Qin, L. 2005. Metagenomics-based drug discovery and marine microbial diversity. Trends Biotechnol. 23(11): 539-543.

Liaw, R.B., *et al*. 2010. Use of metagenomic approaches to isolate lipolytic genes from activated sludge. Bioresour. Biotechnol. 101(21): 8323-8329.

Lim, H.K., *et al*. 2005. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Appl. Environ. Microbiol. 71(12): 7768-7777.

Lin, W., *et al.* 2005. New cyclopentenone derivatives from an endophytic *Streptomyces* sp. isolated from the mangrove plant *Aegiceras comiculatum*. J. Antibiot. 58(9): 594-598.

Lin, Z., Zhu, T., Fang, Y., and Gu, Q. 2008a. 1H and 13C NMR assignments of two new indolic enamide diastereomers from a mangrove endophytic fungus *Aspergillus* sp. Magn. Reson. Chem. 46(12): 1212-1216.

Lin, Z., Zhu, T., Fang, Y., Gu, Q. and Zhu, W. 2008b. Polyketides from *Penicillium* sp. JP-1, an endophytic fungus associated with the mangrove plant *Aegiceras corniculatum*. Phytochemistry. 69(5):1273-1278.

Lubertozzi, D., and Keasling, J. D. 2009. Developing *Aspergillus* as a host for heterologous expression. Biotechnol. Adv. 27(1): 53-75.

Lugo, A.E. 1998. Mangrove forests: a tough system to invade but an easy one to rehabilitate. Mar. Pollut. Bull. 37(8): 427-430.

Malborg, A. S. 1986. The renaissance of erythromycin. J. Antimicrob. Chemother. 18(3): 293-296.

Maragakis, L.L., Perencevich, E.N., and Cosgrove, S.E. 2008. Clinical and economic burden of antimicrobial resistance. Expert. Rev. Anti. Infect. Ther. 6(5): 751-763.

Marahiel, M.A. 1997. Protein templates for the biosynthesis of peptide antibiotics. Chem. Biol. 4: 561–567.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97(7): 2651-2673.

Marchand, C., Lallier-Verges, E., and Baltzer, F. 2003. The composition of sedimentary organic matter in relation to the dynamic features of a mangrove-fringed coast in French Guiana. Estuar. Coast. Shelf. Sci. 56(1): 119–130.

Marrero, E., *et al*. COX-2 and sPLA2 inhibitory activity of aqueous extract and polyphenols of *Rhizophora mangle* (red mangrove). Fitoterapia. 77(4): 313-315.

Martens, T., *et al*. 2007. Bacteria of the Roseobacter clade show potential for secondary metabolite production. Microbe. Ecol. 54: 31-42.

Matsuno, Y., *et al*. 1992. High-efficiency transformation of *Bacillus subtilis* NB22, an antifungal antibiotic iturin producer, by electroporation. Ferment. Bioeng. 73(4): 261-264.

McCormick, M. H., *et al*. 1955–1956. Vancomycin, a new antibiotic. I. Chemical and biologic properties. Antibiot. Annu. 3: 606-611.

Mitra, A., Santra, S.C., and Mukherjee, J. 2008. Distribution of actinomycetes, their antagonistic behavior and the physico-chemical characteristics of the world's largest tidal mangrove forest. Appl. Microbiol. Biotechnol. 80: 685–695.

Mitsui, H., *et al*. 1997. Incubation time and media requirements of culturable bacteria from different phylogenetic groups. J. Microbiol. Methods. 30(2): 103-110.

Mo, X.C., *et al*. 2010. Identification and characterization of a novel xylanase derived from a rice straw degrading enrichment culture. Appl. Microbiol. Biotechnol. 87(6): 2137-2146.

Mootz, H.D. and Marahiel, M.A. 1997. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. J. Bacteriol. 179(21): 6843-6850.

Mootz, H.D., Schwarzer, D. and Marahiel, M.A. 2002. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chem. Biochem. 3: 490–504.

Nakashima, H., *et al*. 1996. In vitro anti human immunodeficiency virus activity of mangrove plants. Indian. J. Med. Res. 103: 278-281.

Nambiar, G.R., and Raveendran, K. 2009. Manglicolous Marine Fungi on *Avicennia* and *Rhizophora* along Kerala Coast (India). Middle-East J. Sci. Res. 4(1): 48-51.

National Employee Development staff. 1987. USDA textural classification study guide. Washington: United States Department of Agriculture–Soil Conservation Service, U.S. Gov. Print Office.

Neilan, B.A. *et al*. 1999. Nonribosomal peptide synthesis and toxigenicity of Cyanobacteria. J. Bact. 181(13): 4089–4097.

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.

Nicolau, D.P. 2009. Containing Costs and Containing Bugs: Are They Mutually Exclusive? J. Manag. Care. Pharm. 15(2): S12-S17.

Odokuma, L.O., and Dickson, A.A. 2003. Bioremediation of a crude oil polluted tropical mangrove environment. J. Appl. Sci. Environ. Manag. 7(2): 23-29.

Olano, C., Mendez, C. and Salas, J.A. 2009. Antitumor compounds from marine actinomycetes. Mar. Drugs. *7*: 210-248.

Oliveira, A., and Pampulha, M.E. 2006. Effects of long-term heavy metal contamination on soil microbial characteristics. J. Biosci. Bioeng. 102(3): 157-161.

Osburne, M.S., *et al*. 2000. Tapping into microbial diversity for natural products drug discovery. Am. Soc. Microbiol. News. 66: 411-417.

Owen, P.L., *et al*. 2007. Endothelial cytoprotection from oxidized LDL by some crude Melanesian plant extracts is not related to their antioxidant capacity. J. Biochem. Mol. Toxicol. 21(5): 231-242.

Paphavasit, N., *et al*. 2002. Ecological assessment of mangrove reforestation/afforestation in Samut Songkharm Province on zooplankton and benthic communities. 1st ed. Bangkok: National Research Council of Thailand.

Pearson, L.A., and Neilan, B.A. 2008. The molecular genetics of cyanobacterial toxicity as a basis for monitoring water quality and public health risk. Curr. Opin. Biotechnol. 19(3): 281-288.

Perera, L.M., Ruedas, D., and Gómez, B.C. 2001. Gastric antiulcer effect of *Rhizophora mangle* L. J. Ethnopharmacol. 77(1): 1-3.

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

Picard, C., *et al*. 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. 58(9): 2717– 2722.

Piel, J., *et al*. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. *U S A*. 101: 16222-16227.

Premanathan, M., *et al*. 1999. Antiviral properties of a mangrove plant, *Rhizophora apiculata* Blume, against human immunodeficiency virus. Antiviral. Res. 44(2): 113-122.

Qin, S., *et al*. 2009. Isolation, diversity, and antimicrobial activity of rare actinobacteria from medicinal plants of tropical rain forests in Xishuangbanna, China. Appl. Environ. Microbiol. 75(9): 6176–6186.

Rainey, F.A., and Oren, A. 2006. Extremophile microorganisms and the methods to handle them. Meth. Microbiol. 35: 1-25.

Rajendhran, J., and Gunasekaran, P. 2008. Strategies for accessing soil metagenome for desired applications. Biotech. Adv. 26: 576–590.

Rausch, C., *et al*. 2007. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC. Evol. Biol. 7: 78-93.

Rausch, C., *et al*., 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucl. Acids Res. 33(18): 5799–5808.

Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. 2004. Metagenomics: Genomic analysis of microbial communities. Annu. Rev. Genet. 38: 525–52.

Rojas Hernández, N.M., and Coto Pérez, O. 1987. Antimicrobial properties of extracts from *Rhizophora mangle* L. Rev. Cubana. Med. Trop. 30(3): 181-187.

Roongsawang, N., *et al*. 2005. Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. FEMS Microbiol. Lett. 252: 143–151.

Roose-Amsaleg, C.L., Garnier-Sillam, E., and Harry, M. 2001. Extraction and purification of microbial DNA from soil and sediment samples. Appl. Soil. Ecol. 18: 47–60.

Rounge, T.B., *et al*. 2009. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. BMC. Genomics. 10: 396-407.

Saikkonen, K., *et al*. 1998. Fungal endophytes: A continuum of interactions with host plants. Annu. Rev. Ecol. Syst. 29: 319–343.

Saikkonen, K., *et al*. 2004. Evolution of endophyte-plant symbioses. Trends. Plant. Sci. 9(6): 275-280.

Saito, F., *et al*. 1994. Entire nucleotide sequence for *Bacillus brevis* Nagano Grs2 gene encoding gramicidin S synthetase 2: a multifunctional peptide synthetase. J. Biochem. 116(2): 357-367.

Sambrook, J., and Russell, D. W. 2001. Molecular Cloning: a laboratory manual, 3rd ed. New York: Cold Spring Harbor Laboratory Press.

Sardessa, Y.N., and Bhosle, S.J. 2002. Organic solvent-tolerant bacteria in mangrove ecosystem. Curr. Sci. 82(6): 622-623.

Schiraldi, C., and Rosa, M.D. 2002. The production of biocatalysts and biomolecules from extremophiles. Trends. Biotechnol. 20(12): 515-521.

Schloss, P.D., and Handelsman, J. 2003. Biotechnological prospects from metagenomics. Curr .Opin. Biotechnol. 14: 303–310.

Schloter, M., Dilly, O., and Munch, J.C. 2003. Indicators for evaluating soil quality. Agric. Ecosyst. Environ. 98(1-3): 255-262.

Schmeisser, C., Steele, H., and Streit, W.R. 2007. Metagenomics, biotechnology with non-culturable microbes. Appl. Microbiol. Biotechnol. 75:955–962.

Schmidt, E.W., *et al*. 2005**.** Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. Proc. Natl. Acad. Sci. U S A.. 102: 7315–7320.

Schwarzer, D., Finking, R., and Marahiel, M.A. 2003. Nonribosomal peptides: from genes to products. Nat. Prod. Rep. 20: 275–287.

Shen, L.R., *et al*. 2009. Xylomexicanins A and B, new Delta 14,15-mexicanolides from seeds of the Chinese mangrove *Xylocarpus granatum*. Z. Naturforsch. C. 64(1-2): 37-42.

Simu, K., and Hagstrom, A. 2004. Oligotrophic bacterioplankton with a novel single-cell life strategy. Appl. Environ. Microbiol. 70(4): 2445–2451.

Sipahi, O.R. 2008. Economics of antibiotic resistance. Expert. Rev. Anti. Infect. Ther. 6(4): 523- 539.

Sipari, H., *et al*. 2010. Development of a chip assay and quantitative PCR for detecting microcystin synthetase E gene expression. Appl. Environ. Microbiol. 76: 3797-3805.

Smith, C. G. 1956. Fermentation Studies with *Streptomyces niveus*. Appl. Microbiol. 4(5): 232–236.

Soil Survey staff. 2006. Keys to soil taxonomy. 10th ed. United States Department of Agriculture–Natural Resources Conservation Service. Washington: U.S. Gov. Print Office.

Sousa, O.V. 2006. The impact of shrimp farming effluent on bacterial communities in mangrove waters, Ceara, Brazil. Mar. Poll. Bull. 52: 1725–1734.

Stachelhaus, T., and Marahiel, M. A. 1995. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J. Biol. Chem. 270(11): 6163–6169.

Stachelhaus, T., Schneider, A., and Marahiel, M. A. 1996. Engineered biosynthesis of peptide antibiotics. Biochem. Pharmacol. 52(2): 177-186.

Stack, D., Neville, C., and Doyle, S. 2007. Nonribosomal peptide synthesis in *Aspergillus fumigatus* and other fungi. Microbiology. 153: 1297–1306.

Staley, J. T., and Konopka, A. 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. 39:321–346.

Steffan, R. J., *et al*. 1988. Recovery of DNA from soils and sediments. Appl. Environ. Microbiol. 54(12): 2908-2915.

Stein, J. L., *et al*. 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. 178(3): 591-599.

Stein, T. and Vater, J. 1996. Amino acid activation and polymerization at modular multienzymes in nonribosomal peptide biosynthesis. Amino. Acids. 10:201-227.

Stephen F. A., *et al*. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic. Acids. Res. 25: 3389-3402.

Streit, W. R., and Schmitz, R. A. 2004. Metagenomics—the key to uncultured microbes. Curr. Opin. Microbiol. 7(5): 492–498.

Strobel, G., *et al*. 2004. Natural products from endophytic microorganisms. J. Nat. Prod. 67: 257-268.

Suganthy, N., *et al*. 2009. Mangrove plant extracts: Radical scavenging activity and the battle against food-borne pathogens. Forsch. Komplementmed. 16(1): 41-48.

Suganthy, N., Pandian, S.K., and Devi, K.P. 2009. Cholinesterase inhibitory effects of *Rhizophora lamarckii*, *Avicennia officinalis*, *Sesuvium portulacastrum* and *Suaeda monica*: Mangroves inhabiting an Indian coastal area (Vellar Estuary). J. Enzyme. Inhib. Med. Chem. 24(3): 702-707.

Takahashi, Y., and Omura, S. 2003. Isolation of new actinomycete strains for the screening of new bioactive compounds. J. Gen. Appl. Microbiol. 49(3): 141-154.

Tamrakar, A.K., *et al*. 2008. Stimulatory effect of Ceriopstagal on hexose uptake in L6 muscle cells in culture. Nat. Prod. Res. 10;22(7): 592-599.

Tan, L.T. 2007. Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry. 68: 954–979.

Tan, N.H. 2010. Isolation and characterization of the thrombin-like enzyme from *Cryptelytrops purpureomaculatus* venom. Comp. Biochem. Physiol. C. Toxicol. Pharmacol. 151(1): 131-136.

Tapi, A., *et al*. 2010. New approach for the detection of non-ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction. Appl. Microbiol. Biotechnol. 85: 1521–1531.

Tomita, F. 2003. Endophytes in Southeast Asia and Japan: Their taxonomic diversity and potential applications. Fungal. Divers. 14: 187-204.

Tooming-Klunderud, A., *et al*. 2007. Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from *Microcystis*: implications for evolution of cyanopeptolins. Microbiol. 153: 1382–1393.

Torsvik, V., Goksøyr, J., and Daae, F. L. 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56(3): 782–787.

Turgay, K., Krause, M., and Marahiel, M.A. 1992. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. Mol. Microbiol. 6(4): 529-546.

Ulrich, A., and Becker, R. 2006. Soil parent material is a key determinant of the bacterial community structure in arable soils. FEMS Microbiol. Ecol. 56: 430–443.

Vane, C.H., *et al*. 2009. Organic and metal contamination in surface mangrove sediments of South China. Mar. Pollut. Bull. 50: 134-144.

Velkov, T., and Lawen, A. 2009. Non-ribosomal peptide synthetases as technological platforms for the synthesis of highly modified peptide bioeffectors – Cyclosporin synthetase as a complex example. Biotechnol. Annu. Rev. 9: 151-197.

Vijayavel, K., Anbuselvam, C., and Balasubramanian, M.P. 2006. Free radical scavenging activity of the marine mangrove *Rhizophora apiculata* bark extract with reference to naphthalene induced mitochondrial dysfunction. Chem. Biol. Interact. 163(1-2): 170-175.

Vizcaino, J.A., *et al*. 2005. Detection of putative peptide synthetase genes in *Trichoderma* species: Application of this method to the cloning of a gene from *T. harzianum* CECT 2413. FEMS. Microbiol. Lett. 244: 139–148.

Voget, S., *et al*. 2003. Prospecting for novel biocatalysts in a soil metagenome. Appl. Environ. Microbiol. 69(10): 6235-6242.

Volkman, J.K., *et al*. 1994. Oil spills. In Swan, J.M., Neff, J.M., and Young, P.C. (eds.), Environmental implications of offshore oil and gas development in Australia, the findings of an independent scientic review, pp 509-695. Sydney: Australian Petroleum Exploration Association (APEA) & Energy Research and Development Corporation (ERDC).

Wallenius, K., *et al*. 2010. Sample storage for soil enzyme activity and bacterial community profiles. J. Microbiol. Methods. 81(1): 48-55.

Walsh, C.T., *et al*. 2001. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr. Opin. Chem. Biol. 5: 525-534.

Wang, G. Y., *et al*. 2000. Novel natural products from soil DNA libraries in a streptomycete host. Org. Lett. 2(16): 2401-2404.

Wang, K., *et al*. 2010. A novel metagenome-derived β-galactosidase: gene cloning, overexpression, purification and characterization. Appl. Microbiol. Biotechnol. 88(1): 155-165.

Wangensteen, H., *et al*. 2009. Antioxidant and antimicrobial effects of the mangrove tree *Heritiera fomes*. J. Nat. Prod. 4(3): 371-376.

Weber, T., and Marahiel, M.A. 2001. Exploring the domain structure of modular nonribosomal peptide synthetases. Structure. 9: R3–R9.

Weiss, A., Jerome, V., and Freitag, R. 2007. Comparison of strategies for the isolation of PCR-compatible, genomic DNA from a municipal biogas plants. J. Chromatogra. B. 853(1-2): 190-197.

Wilfinger, W. W., Mackey, K., and Chomczynski, P. 1997. Effect of pH and ionic strength on the spectroscopic assessment of nucleic acid purity. Biotechniques. 22(3): 474-481.

William, K. 2004. Streptomycin, Schatz v. Waksman, and the balance of credit for discovery. J. Hist. Med. Allied Sci. 59(3): 441-462.

Williamson, L.L., *et al*. 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl. Environ. Microbiol. 71(10): 6335-6344.

World Health Organization. (2001). WHO global strategy for containment of antimicrobial resistance. Geneva: World Health Organization.

World Health Organization. (2005). WHO policy perspectives on medicine - Containing antimicrobial resistance. Geneva: World Health Organization.

Wu, Y., Tam, N.F., and Wong, M.H. 2008. Effects of salinity on treatment of municipal wastewater by constructed mangrove wetland microcosms. Mar. Pollut. Bull. 57(6-12): 727-734.

Xia, J., *et al*. 2006. Comparisons of extraction and purification methods of soil microorganism DNA from rhizosphere soil. J. For. Res. 17(1): 31–34.

Xu, J., *et al*. 2009. Chromones from the endophytic fungus *Pestalotiopsis* sp. isolated from the chinese mangrove plant *Rhizophora mucronata*. J. Nat. Prod. 72(4): 662-665.

Xu, M., *et al*. 2004. Chemical constituents from the mangrove plant, *Aegiceras corniculatum*. J. Nat. Prod. 67(5): 762-766.

Ye, M., *et al*. 2010. Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. Appl. Microbiol. Biotechnol. 87:1023–1031.

Ye, Y., Tam, N.F., and Wong, Y.S. 2001. Livestock wastewater treatment by a mangrove pot-cultivation system and the effect of salinity on the nutrient removal efficiency. Mar. Pollut. Bull. 42(6): 513-521.

Yin, S., *et al*. 2007. Limonoids from the seeds of the marine mangrove *Xylocarpus granatum*. J. Nat. Prod. 70(4): 682-685.

Young, C.C., *et al*. 1993. Polyvinylpyrrolidone-agarose gel electrophoresis purification of polymerase chain reaction-amplifiable DNA from soils. Appl. Environ. Microbiol. 59(6): 1972-1974.

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.

Zhihong, Y.E., and Yang, J. 2009. Metal accumulation and tolerance in wetland plants. Biol. China. 4(3): 282–288.

Zhou, H.W., *et al*. 2009. Polycyclic aromatic hydrocarbon-induced structural shift of bacterial communities in mangrove sediment. Microb. Ecol. 58(1): 153-160.

APPENDICES

APPENDIX A

Table A1 Natural products from marine actinomycetes and marine cyanobacteria. Data shown in this table were collected from Edwards *et al*., 2004; Lam, 2006; Dunlap *et al*., 2007; Tan, 2007; Olano *et al*., 2009.

Sources	Species	Chemical or Fraction	Activity
Actinomycetes	*Actinomadura* sp.	Chandrananimycins	Antialagl, antibacterial, anticancer and antifungal
Actinomycetes	*Actinomadura* sp.	IB-00208	Anticancer
Actinomycetes	*Janibacter limosus*	Helquinoline	Antibacterial
Actinomycetes	*Marinispora* sp.	Marinomycins	Antibacterial and anticancer
Actinomycetes	*Micromonosproa* sp.	Diazepinomicin (ECO-4601)	Antibacterial, anticancer and anti-inflammatory
Actinomycetes	*Nocardiopsis lucentensis* CNR-712	Lucentamycins	Antitumor
Actinomycetes	*Salinispora arenicola* CNR-005	Saliniketals	Antitumor
Actinomycetes	*Salinispora arenicola* CNT-088	Arenamides	Antitumor
Actinomycetes	*Salinispora pacifica* CNS-237	Salinipyrones	Antitumor
Actinomycetes	*Salinispora tropica*	Salinosporamide A (NPI-0052)	Anticancer
Actinomycetes	*Salinispora tropica*	Salinosporamides, sporolides	Antitumor

Sources	Species	Chemical or Fraction	Activity
Actinomycetes	*Streptomyces albogriseolus* A2002	Echinosporins	Antitumor
Actinomycetes	*Streptomyces aureoverticillatus*	Aureoverticillactam	Anticancer
Actinomycetes	*Streptomyces chinaensis* AUBN1/7	1-hydroxy-1-norresistomycin	Antitumor
Actinomycetes	*Streptomyces chinaensis* AUBN1/7	Resitoflavine	Antitumor
Actinomycetes	*Streptomyces griseus*	Frigocyclinone	Antibacterial
Actinomycetes	*Streptomyces nodosus*	Lajollamycin	Antibacterial
Actinomycetes	*Streptomyces nodosus* NPS007994	Lajollamycin	Antitumor
Actinomycetes	*Streptomyces sioyaensis* SA-1758	Altemicidin	Antitumor
Actinomycetes	*Streptomyces sp.*	Bonactin	Antibacterial and antifungal
Actinomycetes	*Streptomyces* sp.	Caprolactones	Anticancer
Actinomycetes	*Streptomyces* sp.	Chinikomycins	Anticancer
Actinomycetes	*Streptomyces* sp.	3,6-disubstituted indoles	Anticancer
Actinomycetes	*Streptomyces* sp.	Glaciapyrroles	Antibacterial
Actinomycetes	*Streptomyces* sp.	Gutingimycin	Antibacterial
Actinomycetes	*Streptomyces* sp.	Himalomycins	Antibacterial
Actinomycetes	*Streptomyces* sp.	Komodoquinone A	Neuritogenic activity
Actinomycetes	*Streptomyces* sp.	Trioxacarcins	Antibacterial, anticancer and antimalarial
Actinomycetes	*Streptoverticillium luteoverticillatum* 11014	Butenolides	Antitumor

Sources	Species	Chemical or Fraction	Activity
Actinomycetes	*Thermoactinomyces* sp.	Mechercharmycins	Anticancer
	Verrucosispora maris AB-18-032	Proximicins	Antitumor
Actinomycetes	*Verrucosispora* sp.	Abyssomicins	Antibacterial
Cyanobacteria	*Bursatella leachii*	Malyngamide	Antiinflammatory
Cyanobacteria	*Bursatella leachii*	Hectochlorin and deacetylhectochlorin	Cytotoxicity (KB and NCI-H187 cells)
Cyanobacteria	*Cyanobacterial* diets of sea hare *Dolabella auricularia*	Dolastatins	Anti-cancer mitotic inhibitors
Cyanobacteria	*Lyngbya confervoides*	Obyanamide	Cytotoxicity (KB cells)
	Lyngbya confervoides	Lobocyclamides	Antifungal (Candida sp)
Cyanobacteria	*Lyngbya majuscula*	Apratoxin A, aurilides B and C, dolabellin, homodolastatin 16, jamaicamides A-C, lyngbyabellins E, pitipeptolides A and B,	Cytotoxicity
Cyanobacteria	*Lyngbya majuscula*	Antillatoxins	Sodium channel-activating and ichthyotoxic
Cyanobacteria	*Lyngbya majuscule*	Curacin A	Cytotoxicity (Colon, renal, and breast cancer derived cell lines)
Cyanobacteria	*Lyngbya semiplena*	Wewakpeptins A and B	Cytotoxicity (H420 Human lung tumor)

Sources	Species	Chemical or Fraction	Activity
Cyanobacteria	*Nostoc ellipsosporum*	Cyanovirin	Anti HIV-1, HIV-2, influenza A and B viruses. (Human immunodeficiency virus; HIV)
Cyanobacteria	*Nostoc linkia* and *N. spongaeforme*	Borophycin	Cytotoxicity (Human epidermoid carcinoma and colorectal adenocarcinoma cell lines)
Cyanobacteria	*Nostoc* sp.	Cryptophycin 1	Cytotoxicity
Cyanobacteria	*Philinopsis speciosa*	Kulokekahilides	Cytotoxicity (P388, SK-OV-3, MDA-MB-435, and A-10 cells)
Cyanobacteria	*Symploca hydnoides*	Malevamide D	Cytotoxicity (P-388, A-549, HT-29, and MEL-28 cells)
Cyanobacteria	*Symploca* sp.	Belamide, guamamide, micromide, symplostatin, tasiamide and tasipeptins	Cytotoxicity

Table A2 Natural products from mangrove environment

Sources	Species	Chemical or Fraction	Activity	Reference
Plant	*Aegiceras corniculatum*	5-O-ethylembelin, 5-O-methylembelin	Cytotoxic activitiy (HL-60, Bel(7402), U937, and Hela cell lines)	Xu *et al*., 2004
Plant	*Avicennia germinans*	3-chlorodeoxylapachol	Cytotoxic activitiy (KB human cancer cells)	Jones *et al*., 2005
Plant	*Avicennia marina*	Avicequinone A, stenocarpoquinone B, avicequinone C	Cytotoxic activitiy	Han *et al*., 2007
Plant	*Avicennia marina*	Polysaccharides	Anticomplementary activity	Fang *et al*., 2006
Plant	*Avicennia officinalis*	Methanolic extract from leaf	Acetyl and ***butyryl cholinesterase inhibitor***	Suganthy *et al*., 2008
Plant	*Bruguiera gymnorrhiza*	Bruguierols C	Antimycobacteria activity	Han, Huang, Sattler, Moellmann *et al*., 2005

Sources	Species	Chemical or Fraction	Activity	Reference
Plant	*Bruguiera gymnorrhiza*	Pimaren diterpenoids	Cytotoxic activitiy (L-929 and K562 cells)	Han, L., Huang, X., Sattler, I., Dahse, H.M., *et al*. 2005.
Plant	*Bruguiera gymnorrhiza*	Polyphenols	Antiatherogenic	Owen *et al*., 2007
Plant	*Ceriops tagal*	Ethanolic extract	Glucose uptake stimulant	Tamrakar *et al*., 2008
Plant	*Heritiera fomes*	Ethanolic extract from bark	Antibacterial activities (*Bacillus subtilis, Kocuria rhizophilia, Pseudomonas aeruginosa, Staphylococcus aureus* and mycobacteria) and antifungal activity	Rojas and Coto, 1987; Han *et al*., 2005; Wangensteen *et al*., 2009
Plant	Mangroves species e.g. *Rhizophora apiculata*	Pyrethrin-like compound	Mosquitoes repellant	Kathiresan and Bingham, 2001
Plant	*Rhizophora apiculata*	Methanolic extract from bark	Antioxidant	Vijayavel *et al*., 2006

Sources	Species	Chemical or Fraction	Activity	Reference
Plant	*Rhizophora lamarckii*	Methanolic extract from leaf	Acetyl and ***butyryl cholinesterase inhibitor***	Suganthy *et al*., 2008
Plant	*Rhizophora mangle*	Aqueous extract from bark	Gastroprotective effect via antioxidant and prostaglandin dependent pathway	Perera *et al*., 2001; Berenguer *et al*., 2006
Plant	*Rhizophora mangle*	Aqueous extract from bark	Anti-inflammation	Marrero *et al*., 2006
Plant	*Rhizophora mucronata*	Methanolic extract from bark	Antioxidant	Suganthy *et al*., 2009
Plant	*Rhizophoraceae* e.g. *Rhizophora apiculata*	No information	Anti-HIV	Nakashima *et al*., 1996 ; Premanathan *et al*., 1999; Li *et al*., 2006
Plant	*Sesuvium portulacastrum*	Methanolic extract from leaf	Acetyl and butyryl cholinesterase inhibitor	Suganthy *et al*., 2008
Plant	*Suaeda monica*	Methanolic extract from leaf	Acetyl and butyryl cholinesterase inhibitor	Suganthy *et al*., 2008

Sources	Species	Chemical or Fraction	Activity	Reference
Plant	*Xylocarpus granatum*	Xylomexicanins A and B	Cytotoxic activitiy (KT breast carcinoma)	Shen *et al*., 2009
Endophytic fungi from *Sonneratia alba*	*Alternaria* sp.	Altenusin, xanalteric acids	Antibacterial and antifungal activitiy	Kjer *et al*., 2009
Endophytic fungi from *Kandelia candel*	Endophytic fungus No. 1962	Cyclic depsipeptides 1962A	Cytotoxic activitiy (MCF-7 breast carcinoma)	Huang *et al*., 2007
Endophytic fungi from *Aegiceras corniculatum*	*Penicillium* sp. JP-1	Polyketides (Leptosphaerone C and penicillenone)	Cytotoxic activitiy (A-549 and P388 cells)	Lin *et al*., 2008b
Endophytic fungi from *Rhizophora mucronata*	*Pestalotiopsis* sp.	Pestalotiopsones	Cytotoxic activitiy (Murine L5178Y cells)	Xu *et al*., 2009
Endophytic fungi from *Acanthus ilicifolius*	*Aspergillus* sp. w-6	Terpeptin A and B	Cytotoxic activitiy (A-549 cells)	Lin *et al*., 2008a

Sources	Species	Chemical or Fraction	Activity	Reference
Soil	*Aspergillus awamori* (fungus)	Oxidized sterols	Cytotoxic activitiy (A-549 cells)	Gao *et al*., 2009
Soil	*Xylaria* spp. no. 2508	Xyloketal B	Endothelial oxidative injury protection	Chen *et al*., 2009
Animals	*Cryptelytrops purpureomaculatus* (pit viper)	Purpurase (Thrombin-like enzyme)	Arginine ester hydrolase and amidase inhibitor, Anticoagulant	Tan, 2009

Table A3 IUPAC nucleotide symbol

Nucleotide symbol	Full Name	Nucleotide symbol	Full Name
A	Adenine	S	G/C
C	Cytosine	W	A/T
G	Guanine	B	G/T/C
T	Thymine	D	G/A/T
U	Uracil	H	A/C/T
R	G/A (purine)	V	G/C/A
Y	C/T (pyrimidine)	N	A/G/C/T
K	G/T	I	Inosine
M	A/C		

Table A4. Table of standard amino acid abbreviations

Amino acid	Three letters	One letter	Amino acid	Three letters	One letter
Alanine	Ala	A	Lysine	Lys	K
Arginine	Arg	R	Methionine	Met	M
Asparagine	Asn	N	Phenylalanine	Phe	F
Aspartic acid	Asp	D	Proline	Pro	P
Cysteine	Cys	C	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glycine	Gly	G	Tyrosine	Tyr	Y
Histidine	His	H	Valine	Val	V
Isoleucine	Ile	I			
Leucine	Leu	L			

APPENDIX B

```
SM3         LEPEYPIERLALMLEDARPLVVLT--SESLQKTLPLHGG--ITLCLDSDWRSLSKESRDN 56
SM48        LDPSYPEERLAFMLDDIRATVLIS--QTGLQGKIPSKNKNIRTIFMDGDREVISGQNLEN 58
SM2-2       VDPDYPAKRVEYLLIHSEARMILS--QPQLITELPATDT----PILDVTGGEVASMPAKP 54
SM23        LDADYPPRRLDFMLRDTDAAVLLATRDT--AEAVADFDGT--LVLLDSPWEEIADQAVDN 56
SM27        LDEVMPGERQSLLAKDAGAKWIVSNRGQGLAPELSDLG----RVDVDS--EEVLTQSTDN 54
            ::   * .*   :  .  .  :::         :.  .       :*     :     .

SM3         PVPAGGPNNTAYVIYTSGSTGKPKGVLIRRSALQNFALSLRDNCNLAPNDRVLQIASSCF 116
SM48        PLNSASPDNLAYIIYTSGSTGKPKGVMITRYNVVRLFQSTRKWFHFNGEDVWTLFHSFAF 118
SM2-2       PGRQCSPNDLAYVIYTSGSTGTPKGVMISHGAAVNTIVDINQRFRVTKFDRILGFSSLSF 114
SM23        LPAQAGPDSLAYVMYTSGSTGRPKGVEVVHRGVVRLVCGT-DYVELGPGEAILQFAPLSF 115
SM27        PEIVANGESLAYVMYTSGSTGQPKGVLVQQRGVSRLVLNN-GYAAFSAEDRVAFAANPAF 113
                 . :. **::******* **** : :    .   .      .   :        .*

SM3         DMSVAEIFPTLLAGAALALPQPGEQRDPARLARFISRLQVTVLFSVPSLLDVLLE-EPG- 174
SM48        DFSVWELWGALLHGGRLVVVPFWVSRSPDRFLDLLICQRVTVLNITPSAFRQLIQ-EEGN 177
SM2-2       DLSVWDIFGTLGAGGTLVILPRESLKSPSRWFDLIEREGITIWNSVPTAMKMLLDFCEGR 174
SM23        DASTFEIWAALLHGGRLAVFP-PGLPSIDELGRFIHDRRITTLWLTAGLFQQMVD----- 169
SM27        DASTMEVWGALLNGGEVIVIEKTALLDTGLFKSALEENGVTVLFLTTALFNQYAH----- 168
            * *. ::: :*  *. : :       .       :    :*    ..  :    .

SM3         ----FHRCSALRLVIAAGDVLPPQLCERFFKQFKAD----LHNLYGPTEATVQTT--IWR 224
SM48        ASGAAGREMALRLVIFGGEALQMRTLKPWYERHEERCPL-LVNMYGITETTVHVTYQPLK 236
SM2-2       R---VCESTTLRLAMLSGDWIPLDLPGRIKAYFEDCK---VVSLGGATEASIWSIYYPIE 228
SM23        --FGLEHLSGVRQLLAGGDVVPPAHAARALAAL-PECC--LINGYGPTENTTFTCCHRMA 224
SM27        --SIPKTLAGLRYLLCGGERGDPSCFRRVLEYNGPEH---LIHCYGPTETTTYASTHEVR 223
                      :*  : .*:                     :    * ** :

SM3         CQRGIQPVRIPIGRPIDNYQVYVLDRNLQLLPVGVPGELCIGGAGLARGYLNSPELTSQK 284
SM48        AADARENSASLIGRPIPDLQVYILDQNLHPVPVGVFGEIYVGGAGLARGYLNRPQLTSER 296
SM2-2       TVDTQWN-SIPYGKPLGRQRFYIFDDQLQPVSDGEVGELCIGGRGVAMGYYREPERTARS 287
SM23        TPKDVGP-TVSIGRPIANTRVYVLDRQGRPVPWGVPGELYAASDGLARGYLARPELTAER 283
SM27        SVAADAK-TISIGRPIGNTTIYILDTNGQPVAPGVAGEIHIGGDGVAKGYLNQAQLSAES 282
                       *:*:    .*::* : : :. *  **:  .. *:* **   .: ::.

SM3         FVPNPFG-DTGDRLYRTGDLAKYLPDGSIDFLGRVD 319
SM48        FIPNSYCEKNGSRLYKTGDLARYLPDGSIEFLGRTD 332
SM2-2       FISDPET---GQTLYRTGDLGRIMNDGNIEIIGRID 320
SM23        FLPDPFSEEPGARMYRTGDLVRWRPDGTLEFLGRMD 319
SM27        FLPDPFSDKPEAKMYKTGDLAYWSANGTIEYLGRND 318
            *:.:.        :*:****     :*.:: :** *
```

Figure 1B Alignment of amino acid sequence of partial A domain of NRPSs amplified from Klongkone mangrove soil

```
360305      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
361101      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
360312      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
360310      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENSVEIICLDSNQEAIIQESGQDAPSP 60
360804      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
360802      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
361102      LNTDYPKDRLSFIMEDTRMLVLLTQERLVAALPENNVEIICLDSNQEAIIQESGQDAPSP 60
371001      LDVNYPADRIEYMLQDSGSILLLSDASAP-ALPVESKLPHLLVDNVATALTDYANDAHNP 59
374605      LDVNYPADRIEYMLQDSGSILLLSDASAP-ALPVESKLPHLLVDNVATALTDYANDAHNP 59
371002      TGCELPGGSYRVHATRLRIHFLLSDASAP-ALPVESKLPHLLVDNVATALTDYANDPHNP 59
374601      LEPTLPAERIAYILKDANPRFLLTTSQYSRTFPIPNKK-LLFIDGIDSFKETFP--AWTK 57
                 *              **:      ::*  .     : ..  :        . .

360305      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
361101      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
360312      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
360310      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
360804      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
360802      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
361102      VTVDN--------LAYVIYTSGSTGQPKGVGVQHRSLCNHLYWVKRSLFSEAVHSIPVTA 112
371001      IYHNPVVAMQPTHLSYVVYTSGSTGKPKGVLVNHLGVN-RLVKNQNYIELDENSVVLQDA 118
374605      IYHNPVVAMQPTHLSYVVYTSGSTGKPKGVLVNHLGVN-RLVKNQNYIELDENSVVLQDA 118
371002      IYHNPVVAMQPTHLSYVVYTSGSTGKPKGVLVNHLGVN-RLVKNQNYIELDENSVVLQDA 118
374601      GISNPDVAVKPHHLAYINYTSGSTGMPKGVMVPHRGVL-RLVTDQNYVPLSERTVTLQSA 116
               :         *:*: ******* **** * * .:  :*   :. :  .        *

360305      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
361101      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSPWTVILEE 172
360312      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
360310      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
360804      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
360802      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
361102      NLSFDASLKQIFAPLLQGTEVWILSEELTNQPVALLRAINSRTNVGLNCVPSLWTVILEE 172
371001      SISFDAATFEMYAAWLNG-GTLVLYPQQYMDLTTLTDVIEQHRVNVLWITCALFDKWAAT 177
374605      SISFDAATFEMYAAWLNG-GTLVLYPQQYMDLTTLTDVIEQHRVNVLWITCALFDKWAAT 177
371002      SISFDAATFEMYAAWLNG-GTLVLYPQQYMDLTTLTDVIEQHRVNVLWITCALFDKWAAT 177
374601      SLLFDAATFEMYAPLLNG-GTLVLYPHQQLDLDELNRVIQTYQVNTLWLTAALFEKWAHH 175
            .: ***:  :::*. *:*  . :*  .   :   *  .*:      *  . : :

360305      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTETTVNASATKIVP 232
361101      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTETTVNASATKIVP 232
360312      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLRIWNLYGPTETTVNASATKIVP 232
360310      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTETTVNASATKIVP 232
360804      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTEKTVNASATKIVP 232
360802      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTETTVNASATKIVP 232
361102      ISCCRARQSAATLTCLLAGGETLSMELTDRTRTALPHLQIWNLYGPTETTVNASATKIVP 232
371001      LQAG----AVPLLKTVITGGDVISPRSVKQVYQQCDNVTVVAAYGPTENTVFTTTYPIPR 233
374605      LQAG----AVPLLKTVITGGDVISPRSVKQVYQQCDNVTVVAAYGPTENTVFTTTYPIPR 233
371002      LQAG----AVPLLKTVITGGDVISPRSVKQVYQQCDNVTVVAAYGPTENTVFTTTYPIPR 233
374601      LASKEKVVALGSLRYLLAGGDVVSPTVVKHVYEKLDNVQLINGYGPTENTTFSVCYPIPR 235
            : .     :   *  :::**:.:*   ..:.     :: :   *****.*. :    *
```

```
360305      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
361101      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
360312      GG----NITIGRPVANTQIYLLDAKLQPVPISVPGEICIGGDGLARGYINRPELTAERFI 288
360310      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
360804      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
360802      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
361102      GG----NITIGRPVANTQIYLLDAKLQPVPIGVPGEICIGGDGLARGYINRPELTAERFI 288
371001      DFNAEQPLPLGRVINNTQLYILDADGQLLSFGVAGEIHVGGAGVARGYLNREDLTASQFI 293
374605      DFNAEQPLPLGRVINNTQLYILDADGQLLPFGVAGEIHVGGAGVARGYLNREDLTASQFI 293
371002      DFNAEQPLPLGRVINNTQLYILDADGQLLPFGVAGEIHVGGAGVARGYLNREDLTASQFI 293
374601      EHSDRFSVPIGRAITNTSVYIVDQHSNLVPKGVVGELCVGGLGLARGYLNRDDLTQEKFV 295
                   :.:** : **.:*::* . : :. .* **: :** *:****:** :** .:*:

360305      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
361101      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
360312      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
360310      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
360804      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
360802      PNPFSDNHGDC--LFKTGDLARYLPDGNIECFGRIDH 323
361102      PNPFSDNHGDR--LFKTGDLARYLPDGNIECFGRIDH 323
371001      DNPLAVGSNGE-KLYKTGDLGRIREDGIVEFLGRIDN 329
374605      DNPLAVGSNGE-KLYKTGDLGRIREDGIVEFLGRIDN 329
371002      DNPLAVGSNGE-KLYKTGDLGRIREDGIVEFLGRIDN 329
374601      ENQFDTTTSDENRLYRTGDLVRLIDNDLLEYVGRLDD 332
             * :    ..   *::**** *   :. :* .**:*.
```

Figure 2B. Sequence alignment of partial A domain of NRPSs of clone 36 and 37 of metagenomic library

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
SM2_2           VDPDYPAKRV EYLLIHSEAR MILSQ---PQ L--ITE-LPA TDTPI-LDVT
AAY42397        IDPELPKERR EFLLTQGEVQ LVLTQ---ES LLEQLA-IPE GI-E-CLSVD
YP_0018699      IDPQLPSQRQ QQLLEQSQAR VIVTD---DP LVATSA-WVG LIPV-ML-ID
NP_486685       IDPQLPAQRR LHLLQETQAQ IILTQ---SW LDTTLE-WAD HLTRICVDLS
ABW70808        LDTNQPEARR QLILDNAEVA RVLSQ---SW LSDSLC-WPA RVTQV-IAVD
ZP_0461695      VDPQLPEQRQ HRLIERCAAK AVLTF----- -DGNVA-VQG MV----IVVT
YP_0030411      IDASYPQQRI HQLLASGEVD TVLTQ---PK FAQQMS-WPD NV-QV-ISLD
NP_929573       IDASYPQQRI HQLLASGEVD TVLTQ---PK FAQQMS-WPD SV-QV-ISLD
ZP_0694216      IDGAYPEPRI QALLKQGAVS TIISDSSEPC RTDDYR---- ------VLIP
ZP_0441615      IDGAYPEPRI QALLKQGAVS TIISDSSEPC RTNDYR---- ------VLIP
ZP_0571998      VDATYPAQRI QALLQQGGVN TVIAQ----- -TSDLAILPH YR----IIVP
ZP_0571773      VDATYPAQRI QALLQQGGVN TVIAQ----- -TSDLEILPH YR----IIVP
ZP_0688901      VDPALPEDRR RRLLAAGEVD TILTT---AA LAAR---WPR DI-RA-IAVD
YP_0033394      VDTGQPPARR ARILADAGVR HVLTQ---SW IRDRPE-GA- DL----VDVD
YP_0036796      VDARQPAARR GLLLGDAGAG LVLAQ---PW TADGAA-AGT GA-RV-LTVD
YP_0036789      VDTTQPAARR RAILRDAGVR HVLTQ---SW LAEIGD-WEQ DV-EP-VEVD
ZP_0627548      VDTAQPPARR DTIIADSGVR TVLTQ---SW LAEIDD-LPA DV-TA-LAVD
ZP_0552144      VDTAQPAARR DTIIGDAGVR TVLTQ---SW LAELED-LPS TV-SP-VAVD
NP_631722       VDTAQPAARR DTIIGDAGVR TVLTQ---SW LAELED-LPS TV-SP-VAVD
Clustal Co      :*   *  *    ::    .   :::                      : :

SM2-2           GGEVASMPAK PPG------- RQCSPNDLAY VIYTSGSTGT PKGVMISHGA
AAY42397        TFESIKNDSI SFV------- PVHNPEDLAY VIYTSGSTGL PKGVIIKHQA
YP_0018699      ERMQ-SQEPT LPLAWV---- -Q-TPEDLAY VIYTSGSTGI PKGVAIDHCS
NP_486685       PVEPILNSPP SLVGKGAGGL GQPT--DLAY VIYTSGSTGT PKGVMIDHQG
ABW70808        QGPRAAQRAL P--------S LDIDPQQLAY VIYTSGSTGV PKGVMINHQA
ZP_0461695      TIVSGTLRPL PPT------P RKQSPDDLAY VIFTSGSTGE PKGVMISHTN
YP_0030411      ETLLNRLPVN PGV-----QS LSARPEDLAY VIFTSGSTGK PKGVMIDHQG
NP_929573       ELLLNRLPKN TGG-----QR LSAHPEDLAY VIFTSGSTGK PKGVMIDHQG
ZP_0694216      ALMTEAQAHF IPV------- -ANQPTDLAY VIFTSGSTGQ PKGVMMEHGA
ZP_0441615      ALMTEAQAHF IPV------- A-NQPTDLAY VIFTSGSTGQ PKGVMMEHGA
ZP_0571998      ELSNDTIGDF TPV------- -PIRATDLAY VIFTSGSTGQ PKGVMMEHAA
ZP_0571773      ELSNDTIGDF TPV------- -PIRATDLAY VIFTSGSTGQ PKGVMMEHAA
ZP_0688901      GL-----APA PRPGS---IV GRARPDHLAY VIFTSGSTGE PKGVMIEHRA
YP_0033394      LL-LP-SA-L PAPQP----A A--DPGDLAY VIYTSGSTGD PKGVMISHRA
YP_0036796      DDGA-DSVPE PSDVPDPT-G AG--PDDLAY VIYTSGSTGR PKGVMVSHRA
YP_0036789      AVPAADAVPA AWT-PDAL-P PPVDPDALAY VIYTSGSTGT PKGVMVSHRA
ZP_0627548      LLPEDAVATP GET------A ARRDPDDLAY IIYTSGSTGT PKGVMISHRA
ZP_0552144      LVGEA-TADL PP-------A ARRDPDDLAY VIYTSGSTGT PKGVMISHRA
NP_631722       LVGEA-TADL PP-------A ARRDPDDLAY VIYTSGSTGT PKGVMISHRA
Clustal Co                              *** :*:******  **** :.*
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
SM2-2           AVNTIVDINQ RFRVTKFDRI LGFSSLSFDL SVWDIFGTLG AGGTLVILPR
AAY42397        VVNTILDINQ RFNVTANDRI LAVSALNFDL SVYDIFGILA VGGTVVIPSA
YP_0018699      AVNTLLDINS RFCVSPADRV LALSALNFDL SVYDIFGVLA AGGTIVMPGV
NP_486685       AVNTILDINQ RFGVTENDRV LAVSSLSFDL SVYDIFGILA AGGTIIIPKS
ABW70808        ALNTIVDINQ RFAIEAQDRV LALASLGFDL SVYDIFGLLA VGGALVLPDP
ZP_0461695      AVNTVADINR RFSVNSQDRV YSIAPAGFDL SVYDYFGVLG AGGSILFAAE
YP_0030411      AVNTILDINQ RIALNEHDSV LAISELTFDL SVYDLFGTLS CGAKLVIPSP
NP_929573       AVNTILDINQ RIALNEHDSV LAISELTFDL SVYDLFGTLS CGAKLVIPSP
ZP_0694216      VVNTLLDINQ RIALDHRDRV LAISSLNFDL SVFDIFSTLS CGARLVIPQT
ZP_0441615      VVNTLLDINQ RIALDHRDRV LAISSLNFDL SVFDIFSTLS CGARLVIPQT
ZP_0571998      VVNTLLALNQ RIALNSHDRV LAISALNFDL SVFDIFSTLS RGARIVIPSI
ZP_0571773      VVNTLLDLNQ RIALNSHDRV LAISALNFDL SVFDIFSTLS RGARIVIPSI
ZP_0688901      ALNTICDVNE RFGVGPHDRM LALSELGFDL SVYDIFGVLG AGGALVLPEG
YP_0033394      ALNTVADVNR RFAVTAEDRV LGLAGLGFDL SVYDIFGPLS LGGALVLPDA
YP_0036796      ALNTLHDVRG RFGVTADDRG IALASLGFDL SVFDVFGLLG AGACLVLPDA
YP_0036789      ALNTIADVNR RFSVGSGDRV LGLAALGFDL SVYDLFGPLG AGGAIVLPHA
ZP_0627548      ALNTVEDINR RFTVTGRDRV LGIAGLGFDL SVYDLFGPLA VGATLVLPHA
ZP_0552144      ALNTVEDINR RFAVDERDRV LGIAGLGFDL SVYDLFGPLA VGATLVLPRS
NP_631722       ALNTVEDINR RFAVDERDRV LGIAGLGFDL SVYDLFGPLA VGATLVLPRS
Clustal Co      .:**:  :.  *: :   *    ..:   *** **:* *. *.  *. :::

SM2-2           ESLKSPSRWF DLIEREGITI WNSVPTAMKM LLDFCEGRRV CE-STTLRLA
AAY42397        IDAKDPARWY ELIVKHQVTL WNSVPALMQM LVEYLSGQLN QS-HGPLRLA
YP_0018699      TEVKEPAHWV ELMRQHHVTL WNTVPALGQM LADYLSRERM TP-PQGLRLA
NP_486685       G--NDPTHWM QLINQHQVTI WNTVPALMQL LLDTSPT--- --QNQTLRLI
ABW70808        QRRADPSHWA ECVREHGVTL WNSVPAQLQM LTHYLQAVPS MA-PGSLRLA
ZP_0461695      NEPADPGLWA EQIVKQGVTL WNTVPAPVKA LFERAGEQLR ---DSSLRLV
YP_0030411      GDSRQPDKLL TWLQQESVTV WNSVPAFVQL LEEYAR-SYP HS-LDSLRWI
NP_929573       GDSRHPDKLL TWLQQESVTV WNSVPAFVQL LEEYAR-DYP HS-LNSLRWV
ZP_0694216      SPSQDPEALL HLAQQSAITV WNSVPAFAQL LVDLLE-NRS NP-LPHLRQI
ZP_0441615      SPSQDPEALL HLAQQSAITV WNSVPAFAQL LVDLLE-NRS NP-LPHLRQI
ZP_0571998      SSSQDPEALV RLAQQSGITI WNSVPAFAQL LADLLE-LSS SP-LQSLRHI
ZP_0571773      SSSQDPEALV RLAQQSGITI WNSVPAFAQL LADLLE-LNS SP-LQSLRHI
ZP_0688901      RSARDPDHLA RLVLDHGVTL WNAVPSFMQL FVASAEAQAA ---LRRLRLV
YP_0033394      EGRGDPAHWA RLIAGHGVTV WNSVPAQLQM LDHYLGSQPE PD-LPSLRLA
YP_0036796      DRRGDPSHWA ELVERHGVTV WNSVPAQMQM LEDHLASGGG RD-VGSLRLA
YP_0036789      DRRGDPSHWA ELAREHGVTL WNSVPAQMGM LADYLSTAPA QA-PTTLRVA
ZP_0627548      DRRGDPSHWA ELVRDFGVTV WNSVPGQLHM LCDWLRSVPP TD-DASLRLA
ZP_0552144      DRRGDPSHWA ELVRDFGVTV WNSVPGQLHM LCDWLRSEPP TD-DGSLRLA
NP_631722       DLRGDPSHWA ELVRDFGVTV WNSVPGQLHM LCDWLRSEPP TD-DGSLRLA
Clustal Co           *           :*: **:**      :               **
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         210        220        230        240        250
SM2-2            MLSGDWIPLD LPGRIKAYFE DCKVVSLGGA TEASIWSIYY PIETVDTQWN
AAY42397         LLSGDWIPLT LPEQIKDLWS QIQIVSLGGA TEASIWSIYH PIEQITPVTK
YP_0018699       LLSGDWLPLS LPAQLRQLWS QMEIVSLGGA TEASIWSICY PIDEVDPSWS
NP_486685        LLSGDWIPLT LPPRIRSQFN HPQIISLGGA TEASIWSIFY PIETIDPNWK
ABW70808         LLSGDWIPLN LPAEAAQLLP GLRLISLGGA TEAAIWSIYY PITQVNPQWR
ZP_0461695       LMSGDWIPVD LPDQIRHVSE KIDVISLGGA TEGSIWSIVY PVQTVDTTWK
YP_0030411       LMSGDWIPTH LPAKLYALHP ELNLLSLGGA TEASIWSIAY PIAHIDPNWR
NP_929573        LMSGDWIPTS LPERLSALHP ALNLLSLGGA TEASIWSIAY PIAQVDPNWR
ZP_0694216       MMSGDWIPVN LPDRLNTLMP QAKLLSLGGA TEAAIWSICY PIEKSYATHT
ZP_0441615       MMSGDWIPVN LPDRLNTLMP QAKLLSLGGA TEAAIWSICY PIEKSYATHT
ZP_0571998       MMSGDWIPVN LPDRLTKVAP NAQLLSLGGA TEAAIWSICH PIEKSYADQT
ZP_0571773       MMSGDWIPVN LPDRLTKVAP NAQLLSLGGA TEAAIWSICH PIEKSYADQT
ZP_0688901       LMSGDWIPLD LPPRLVAANP GLEVVSLGGA TEASIWSILH PIGPLDPSWS
YP_0033394       MLSGDWIPIA LPGRVRGRLP GLELISLGGA TEAAIWSIYH PIGEVRDGLR
YP_0036796       MLSGDWIPVA LPDRIRRRAP GLRVVSLGGA TEAAVWSIAF PVDEVDPAWA
YP_0036789       LLSGDWIPLT LADRVRAHLP GTALYSFGGA TEGSIWSIHH PIGEADPSWP
ZP_0627548       LISGDWIPVA LPDQARELLP GLEVISLGGA TEGSIWSIVH PIDKVDTARP
ZP_0552144       LISGDWIPVA LPDQARELLP GLEIVSLGGA TEGSIWSIAH PIGEVDTARP
NP_631722        LISGDWIPVA LPDQARELLP GLEIVSLGGA TEGSIWSIAH PIGEVDTARP
Clustal Co       ::****:*   *. .          : *:*** **.::*** . *:
SM2-2            SIPYGKPLGR QRFYIFDDQL QPVSDGEVGE LCIGGRGVAM GYYREPERTA
AAY42397         SIPYGKSLGN QTVSVLNDLM QPTPVWVCGD LYIGGVGLAS GYLLDEKKTN
YP_0018699       SIPYGKPLVN QTFVVFDDRL NARPVWVPGE LYIGGIGLAR GYWQDEERTA
NP_486685        SIPYGYSLTN QQVYVLNHSL EPCPTWAIGE IYISGLGVAK GYWQNPELTA
ABW70808         SIPYGMPLAN QRFMVLDEQG RDRPQGVAGE LYIAGSGLAL GYLGDAEKTA
ZP_0461695       SIPYGKPLAN QRFHVLNEWL EPSPKWVTGE IFIAGDGVAQ GYLGDDEKTQ
YP_0030411       SIPYGKPLAN QTFYVLNSAL SPCPVWVTGE LYIGGQGLAL GYWADSEKTE
NP_929573        SIPYGKPLAN QTFYVLNATL SPCPVWVTGE LYIGGQGLAL GYWADLEKTA
ZP_0694216       SIPYGKPLTH QQFYVLDEQL NPCADWVTGE LYIGGFGLAR GYWHDQEKTD
ZP_0441615       SIPYGKPLTH QQFYVLDEQL NPCADWVTGE LYIGGRGLAR GYWHDQEKTD
ZP_0571998       SIPYGKPLSN QHFYILDQQL EPCPEWVTGE LYIGGHGLAR GYWQDQARTD
ZP_0571773       SIPYGKPLSN QHFYILDQQL EPCPDWVTGE LYIGGHGLAR GYWQDQERTD
ZP_0688901       SVPYGSPMRN QRFHVLGVDL EDCPDHVAGE LYIAGEGLAR GYWRDDARTA
YP_0033394       SIPYGTPLAN QAFHVLDEAL RPCPDWVPGE LYISGAGLAL GYLGDEDRTA
YP_0036796       SIPYGRPLAN QTFHVLDHAL RDRPDHVPGE LYIGGAGLAS GYLGDPERTA
YP_0036789       SVPYGRPLTN QTFHVLDGRM GDRPDWVAGE LHIGGAGVAT GYLGDEARTA
ZP_0627548       SVPYGTPLTN QTFAVLDRHL RPRPEWVPGE LYIGGAGVAL GYFGDEGRTA
ZP_0552144       SIPYGKPLTN QTFAVLDRHL RPRPEWVPGE LYIGGAGVAL GYLGDGERTA
NP_631722        SIPYGKPLTN QTFAVLDRHL RPRPEWVPGE LYIGGAGVAL GYLGDGERTA
Clustal Co       *:*** .: . * . ::.       .    *: : *.* *:*  **  :   *
```

```
                ....|....| ....|....| ....|....| ....|
                         310        320        330
SM2-2           RSFISDPETG QTLYRTGDLG RIMNDGNIEI IGRID
AAY42397        ASFITHPVTH ERLYKTGDLG RYLPDGNIEF LGRCD
YP_0018699      TSFITHPVTG ERLYKTGDLG RYLPNGNIEF LGRLD
NP_486685       EKFIQHPYTP TPLYKTGDLG RYLSDGTIEF LGRED
ABW70808        ERFVDHPRSG ERLYRTGDLG RYLDDGLIEF LGRED
ZP_0461695      ARFFQHPRTG ERLYRTGDLG RYINEGLIEI LGRED
YP_0030411      HAFITHPQTG ERLYRTGDLG RWRPDGNIEF LGRND
NP_929573       QAFITHPQTG ERLYRTGDLG RWRPDGNIEF LGRND
ZP_0694216      LAFIEHPTLG QRLYKTGDLG RYLPDGNIEF LGRND
ZP_0441615      FAFIEHPTLG QRLYKTGDLG RYLPDGNIEF LGRND
ZP_0571998      HAFIIHPMSG ERLYKTGDLG RYLPDGNIEF LGRND
ZP_0571773      HAFIIHPMSG ERLYKTGDLG RYLPDGNIEF LGRND
ZP_0688901      ERFLLHPRSG ERLYRTGDLG RYRDEGLIEF LGRAD
YP_0033394      QRFIRRPGTG ERLYRTGDLG RYHDDGTIEF LGRED
YP_0036796      DRFVTRPGSG ERLYRTGDLG RYRGDGTIEF LGRSD
YP_0036789      QRFVTHPRTG ERLYRTGDLG RYRPGGDIEF LGRED
ZP_0627548      ERFLTDPATG ERLYRTGDLG RYLPDGTIEF LGRED
ZP_0552144      QRFLTDLATG ERLYRTGDLG RYLPDGTIEF LGRED
NP_631722       QRFLTDLATG ERLYRTGDLG RYLPDGTIEF LGRED
Clustal Co        *.         **:***** *    * **: :** *
```

Figure 3B. Sequence alignment of SM2_2 NRPS with protein sequences with similar conserved core sequences from database

```
           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   10         20         30         40         50         60
P27206     IDPDYPDQRI EYILQDSGAK LLLKQEGIS- -VPDSYTGDV ILLDGSRTIL SLPLDENDEE
ACB52390   LDPAYPQERL AFMVSDSQIS VLLTTETLAP TIPQ-AQAQV ICLDRDWKTI R----QKSQD
ZP_0696580 LDPRYPSERL AFMLEDAQVS IILTRQDIVK KLPS-HNAHF VRMDEDWKTL A----QQNGN
P94459     IDPDYPEERI SFLLEDSGTN ILLLQS-AGL HVPE-FTGEI VYLNQTNSGL A----HRL-S
Q9R919     IDPTYPEERI RYILEDSDTK LLLVQHHLRE KVP--FTGKV --LDMEDPQT F----SEDGS
SM3        LEPEYPIERL ALMLEDARPL VVLTSESLQK TLPL-HGGIT LCLDSDWRSL S----KESRD
ZP_0738681 LDPEHPADRV QMILEDTSLR LVLTQSQLKE RLADRTDLQL ICMDKPFASS G----ARPNT
CAQ48254   LDPEYPLERL SFMLEDAAVN VLLTQQKLIN KLPE-HQAQL ICLDADWELI F----QFSRD
AAZ03550   LDPDYPIERI IFMLEDAAVK VLLTQQKLIN KLPE-HQAQL ICLDADWELI S----QFSQD
AAO23333   LDPEYPQERL TFMLADAQVS VLLTQQHLVE KLPR-HQARV VHLDKDWVAI A----KSSQE
CAC01603   LDPEYPTERL TFMLADAQVS VLLTQQHLVE KLPE-NQEPV VCLDTDWLVI C----ESSQE
ABW04917   LDPDYPTERL GDILSDSGVS LVLTQESLGD FLPQ-TGAEL LCLDRDWEKI A----TYSPE
CAO90227   LDPNYPQERL SYLLEDTGVK VIITGESLRG LLDE-YRGIV VALDTDWSAI S----QESQN
YP_0016588 LDPNYPQERL SYLLEDTGVK VIITAESLRG LLGE-YRGIV VALDTDWPAI S----QESQN
BAC57995   LDPNYPQERL SYLLEDTGVK VIITAESLRG LLGE-YRGIV VALDTDWPAI S----QESQN
YP_320287  LDPSYPRERL AFMLQDAQVA VLLTQEKFLP SLPE-HQATV VCLDKDNEVW A----SETIV
NP_522203  LDPSYPQDRL TYMLEDSAPV AVLTQGLVRE QLGM-LSVPV LDLDGPQ--- -----EDAEH
YP_049592  LDPGYPAERL AYMLDDARPV ALLTQANQRA LLT--GDIPV VMLDTADFS- -----HLSED
Clustal Co ::* :* :*:   :: *:     ::          :           ::
P27206     NP-ETAVTAE NLAYMIYTSG TTGQPKGVMV EHHALVNLCF WHHDAFSMTA EDRSAKYAGF
ACB52390   NP-IGGVTPQ NLAYLIYTSG STGTPKGVLV SHGGLVNLTE DKIRVCQVSP DSCVLQFFSF
ZP_0696580 NP-RSETIAH NLAYIIYTSG STGTPKGVLV SHQSLCNLAT AQIQVFHVSP QSRVLQFASL
P94459     NP-NVDVLPQ SLAYVIYTSG STGMPKGVEI EHRSAVNFLN SLQSRYQLKH SDMIMHKTSY
Q9R919     NL-ESISGPN QLAYVIYTSG STGKPKGVMV EHRSVINRLV WMQENYPLDE RDAILQKTAI
SM3        NP-VPAGGPN NTAYVIYTSG STGKPKGVLI RRSALQNFAL SLRDNCNLAP NDRVLQIASS
ZP_0738681 AA-RTSATPD DLAYVLYTSG STGKPKGVMI PHAGLMNRLM WMQEEYRLSY QDRVLQKTPY
CAQ48254   NL-ITDIQAT NLAYVIYTSG STGQPKGVML SHSNLSNHMF WMQETFPLTR ADRVLQKTSF
AAZ03550   NP-ITDVQAT NLAYVIYTSG STGQPKGVML SHSNLSNHMF WMQETFPLTK TDRVLQKTPF
AAO23333   NP-IAQVQAS NVAYVIYTSG STGQPKGVIL SHSNLCNHMF WMQATFPLTK EDKVLQKTPF
CAC01603   SP-ITEVQPG NLAYVIYTSG STGTPKGVML SHSNLCNHMS WMQATFPLTE KDKVLQKTPF
ABW04917   NP-FNLTTPE NLAYVIYTSG STGKPKGVMN IHQGICNTLK YNIDNYNLNS EERILQITPF
CAO90227   NC-DSGVTGE NLAYVIYTSG STGKPKGVMN NHKGIRNRLL WMQDTYQLTK SDCILQKTPF
YP_0016588 NC-DSGVTGE NLAYVIYTSG STGKPKGVMN NHKGIRNRLL WMQDTYQLTK SDGILQKTPF
BAC57995   NC-DSGVTGE NLAYVIYTSG STGKPKGVMN NHKGIRNRLL WMQDTYQLTK SDGILQKTPF
YP_320287  NP-VNEVTTH NLAYVIYTSG STGRPKGVMN THRGICNRLA WMQETYQLTI VDRVLQKTPF
NP_522203  DPQVEALKPH HLAYVIYTSG STGRPKGVMN EHRGVVNRLW WAQQTYRLDA SDRVLQKTPF
YP_049592  NPHVVGLDAH HLAYVIYTSG STGKPKGVMN SHRGLCNRLV WMQNTYRLTP DDRVLQKTPF
Clustal Co             **::**** :** ****    :    *          :    .   :
P27206     GFDASIWEMF PTWTIGAELH VIEEAIRLDI VRLNDYFETN GVTITFLPTQ LAEQFME---
ACB52390   SFDASIPEII MALGCGAKLC LAKLESLLPG PNLLKLLKDE KITHITITPS ALSNLAV---
ZP_0696580 NFDVSISEIL MALLAGATLY LGSQEAILPG TVLLHFLQQN AITIATFPPA VLKALPD---
P94459     SFDASIWELF WWPYAGASVY LLPQGGEKEP EVIAKAIEEQ KITAMHFVPS MLHAFLEHIK
Q9R919     TFDVSVWELF WWSIVGSKVV LLPNGGEKNP ELILDTIEQK GVSTLHFVPA MLHAFLESME
SM3        CFDMSVAEIF PTLLAGAALA LPQPGEQRDP ARLARFISRL QVTVLFSVPS LLDVLLE---
ZP_0738681 SFDVSVWEFL WPLMYGAGLI ILEPGEHRNP AYLVEIIKRH EVSIIHFVPS MLQLFVD---
CAQ48254   SFDASVWEFY APLLVGGQLL IAQPGGHTDS DYLLKTIAQQ QVTTVQLVPS LLQMLLE---
AAZ03550   SFDASVWEFY APLLVGGQLL IAQPGGHTDS DYLLKTIAQQ QVTTVQLVPS LLQMLLE---
AAO23333   GFDASVWEFY APLLAGGQLL IAEPRGHTDS AYLLRLIAQQ QVTTIQLVPS LLQMLLE---
CAC01603   GFDASVWEFY APLLAGGQLL IAKPGGHTDS AYLLRLIAQQ QVTIVQLVPS LLQMLLE---
ABW04917   SFDVSVWEVF SSLTSGATLV VAKPDGYKDI DYLIDLIVQE QVTYFTCVPS ILRVFLQ---
CAO90227   SFDVSVWEFF WPLLAGATLV VAEPEGHKDS TYLIQLIQKQ QITTLHFVPS MLRVFLQ---
YP_0016588 SFDVSVWEFF WPLLAGATLV VAKPEGHKDS TYLIQLIQKQ QITTLHFVPS MLRVFLQ---
BAC57995   SFDVSVWEFF WPLLAGATLV VAKPEGHKDS TYLIQLIQKQ QITTLHFVPS MLRVFLQ---
YP_320287  SFDVSIWEFF WPLTTGACLV MARPGGHQDS AYLVKLIQEQ QITTIHFVPS MLQVFLA---
NP_522203  GFDVSVWELF WPLLAGARLV MARPEGHKAP AYLAATIEQA GITTLHFVPS MLQLFLD---
YP_049592  SFDVSVWEFF WPLLYGARLV MARPDGHKDA AYLAQLIERT GITTLHFVPS MLQQFVQ---
Clustal Co  ** *: *.       *. :  :            :   :     ::     .      :
```

```
            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   190        200        210        220        230        240
P27206      --------LE NTSLRVLLTG GDKLKRAVK- --KPY----- --TLVNNYGP TENTVVATSA
ACB52390    --------TD LPDLEMVLVG GEAPSPELI- --DNWSG--- DRLFINAYGP TEVTVNASMV
ZP_0696580  --------AL LPSLQTIISA GEACSPDIV- --ARWGH--- NRQFFNAYGP TETTVYATID
P94459      ---YRSVPIK TNRLKRVFSG GEQLGTHLV- --SRFYELLP NVSITNSYGP TEATVEAAFF
Q9R919      QTPSGKLKRK LASLRYVFAS GEALTPKHVD GFQRIITPVS HAQIINLYGP TEATIDVSYF
SM3         ---EPGFH-R CSALRLVIAA GDVLPPQLC- --ERFFKQF- KADLHNLYGP TEATVQTTIW
ZP_0738681  ---QPGSE-S CISLRDVICS GEALPYQLK- --EKFSEKL- CANLHNLYGP TEASIDVTYW
CAQ48254    ---QGGIE-N CQLLKRVFCG GEILPVALQ- --EKLFSQL- NVNLCNLYGP TECCIDVTFW
AAZ03550    ---QGGIE-N CQLLKRVFCG GEILPVALQ- --EKLLSQL- NVNLCNLYGP TECCIDVTFW
AAO23333    ---QGGIE-T CHSLKHVFCG GEVLPVTLQ- --ESLLSKL- DVNLHNLYGP TEACIDATFW
CAC01603    ---QGGIE-T CHSLKHVFCG GEVLPVALL- --EGLLSKL- DVNLHNLYGP TETCIDATFC
ABW04917    ---HPKSK-D CHCLKRVIVG GEALSYELN- --QRFFQQL- NCELYNAYGP TEAAVDATVW
CAO90227    ---EPELK-G CSSLKRVFCS GEALSLELT- --QRFFEHF- DCELHNLYGP TEAAIDVTYW
YP_0016588  ---EPELK-E CSSLKRVFCS GEALSLDLT- --QRFFEHF- DCELHNLYGP TEAAIDVTYW
BAC57995    ---EPELK-E CSSLKRVFCS GEALSLDLT- --RRFFEHF- DCELHNLYGP TEAAIDVTYW
YP_320287   ---EPSVE-A CKCLRRVICS GEILPVQLQ- --EHFFTRL- DAELHNLYGP TEAAIDVTFW
NP_522203   ---QVEAG-R CQGLRRMLCS GEALPHALQ- --QRSLARFP HSELHNLYGP TEAAIDVTAW
YP_049592   ---WADADCA CDSLRRVICS GEALPAELQ- --QRFFARF- NAQLHNLYGP TEAAIDVTFW
Clustal Co              *. :: . *:                      : * *** **  : .:
P27206      EIHPEEGS-- LSIGRAIANT RVYILGEGNQ VQPEGVAGEL CVAGRGLARG YLNREDETAK
ACB52390    PC--GNGHPT LPTLLPSANK QLYILDRHLQ PVPVGVLGEL HIGGVGLARG YLNRPDLTAE
ZP_0696580  ECTSRQEK-- ISIGRPIANT QVYILDQEMQ IVPVGILGEL YIGGAGVARG YLNRPELTKD
P94459      DCPPHEKLER IPIGKPVHHV RLYLLNQNQR MLPVGCIGEL YIAGAGVARG YLNRPALTEE
Q9R919      ECEADKRYNS VPIGKPISNI QLYILQAGY- MQPVGVAGEL CIAGDGLARG YLNRPELTAE
SM3         RCQRGIQPVR IPIGRPIDNY QVYVLDRNLQ LLPVGVPGEL CIGGAGLARG YLNSPELTSQ
ZP_0738681  NCSEPIGKRI VPIGRPIWNT QIYVLNTKLQ SVPVGVIGEL YIGGVGLAKG YLNRDDLTLE
CAQ48254    NCQREMYGQR IPIGRPIFNT QIYILDSNLQ PVPVGIPGEL HIGGAGLARG YLNRPELTQE
AAZ03550    NCQREMYGQR IPIGRPISNT QIYILDSNLQ SLPVGIPGEL HIGGAGLARG YLNRPELTQE
AAO23333    NCQREIYPQL IPIGRPIDNT QIYILDQNLQ PVPVGVPGEL HIGGAGLAKG YLNLPELTQE
CAC01603    NCQREIYAQI VPIGRPISNT QIYILDQNLQ ALPVGVPGEL HISGAGLARG YLNRPELTQE
ABW04917    CCQPNSQ--L IPIGTPIANA QVYILDSYLQ PVPIGVAGEL HIGGMGLARG YLNQPELTAE
CAO90227    PCLPESQKAI VSIGRPIANT QIYILNPHLQ PVPIGIVGEL HIGGIGLARG YLNRPELTAE
YP_0016588  PCLPENQKAL VSIGQPIANT QIYILNPHLQ PVPIGIVGEL HIGGIGLARG YLNRPELTAE
BAC57995    PCLPENQKAL VSIGQPIANT QIYILNPDLQ PVPIGIVGEL HIGGIGLARG YLNRPELTAE
YP_320287   ACNRHSDKNI VPIGRAIANT QIYILDKHLQ PVPIGVPGEL HIGGVGVARG YLNQPQLTAE
NP_522203   RCNAEIHPGV VPIGRPIANT QIYVLDAYRQ PVPLGVTGEI YIGGAGVARG YLNRPELTAE
YP_049592   ACQPDDHRSF VPIGRPIANT QLYILDTLGQ PVPLGVAGEL HIGGVGVARG YLNRPDLTAE
Clustal Co              :.   .  :  ::*:*        * *  **:  :.* *:*:* ***    * .
```

```
                ....|....| ....|....| ....|....| ....|....| ....|..
                       310        320        330        340
P27206          RFVADPFV-- ---------- PGERMYRTGD LVKW-TGGGI EYIGRID
ACB52390        RFIPSPFDPP LPLLRKGGKK QGARLYKTGD LACYLPDGRI KLLGRLD
ZP_0696580      RFVPHPFS-- --------DT IGDQLYKTGD LARYLPDGRI ELIGRAD
P94459          RFLEDPFY-- ---------- PGERMYKTGD VARWLPDGNV EFLGRTD
Q9R919          KFVKNPFS-- ---------- AGERMYRTGD LARWLPDGNI EYLGRID
SM3             KFVPNPFG-- ---------D TGDRLYRTGD LAKYLPDGSI DFLGRVD
ZP_0738681      KFIANPFR-- ---------- SGEKMYRTGD LVRFLSEGAI DYVGRAD
CAQ48254        KFIPNPFS-- --------NY PDSRLYKTGD LARYLPDGNI EYLGRID
AAZ03550        KFIPNPFS-- --------NY PDSRLYKTGD LARYLPDGNI EYLGRID
AAO23333        KFIPNPFQ-G SRGAGEQGSR GRERLYKTGD LARYLPDGNI EYLGRID
CAC01603        KFIANPFS-- --------TY PGSRLYKTGD LARYLPNGNI EYLGRID
ABW04917        KFIPHPFA-- ---------- -EGKLYKTGD LARYLPDGNI EYLGRID
CAO90227        KFIPNPFA-- -KVEAGIGGE IRAKLYKTGD LARYLPDGNI EFLGRID
YP_0016588      KFIPNPFA-- -KVEGEIGGE IRAKLYKTGD LARYLPDGNI EFLGRID
BAC57995        KFIPNPFA-- -KVEGEIGGE IRAKLYKTGD LARYLPDGNI EFLGRID
YP_320287       KFIVNPFS-- --------NN SNNRLYKTGD LARYHTDGSI EYLGRLD
NP_522203       RFVVNPFH-- --------GE GRERMYRTGD LGRWLPDGSL EYQGRAD
YP_049592       RFIPDPFS-- --------NQ HGARLYKTGD LARWLPDGSL EYLGRND
Clustal Co      :*:  **                ::*:*** :  : . * : .  ** *
```

Figure 4B. Sequence alignment of SM3 NRPS with protein sequences with similar conserved core sequences from database

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
SM23             LDADYPPRRL DFMLRDTDAA VLLATRDTAE AVAD-F-D-G TLVLLDS---
ZP_0677648       LDPEDPPARH ELLLGDAGVG MVITEEALRE RVPD---G-V AAVGEEG---
YP_0015446       VDPSYPVERL AWMLSDLQPT VVIAQHGVLD RLPSV---AC SVVVLET---
ZP_0051751       LDPNIPPERL TILLEDTQIN LLLTQNDINL PWPN----TL TVIDLQQ---
ZP_0711340       LDPTYPKERL AFMLEDASVP VLLTQTRLVE SLPH----QA RVVCLDA---
AAW55330         LDPGYPRERL AFMLLDTQVS ILLTQKDLVA KLPT-H--TA FVICLDA---
P2706            IDPDYPDQRI EYILQDSGAK LLLKQEGISV PDSY----TG DVILLDG--S
P94459           IDPDYPEERI SFLLEDSGTN ILLLQSAGLH VPEF----TG EIVYLNQ---
Q9R919           IDPTYPEERI RYILEDSDTK LLLVQHHLRE KVPF----TG KVLDMED---
YP_0037122       LDPNYPAERL TYILDDSAPV ALLTQEAHLN KLSA----TL PTVLLDN---
YP_0025514       VDPVYPKDRI DFVARDARPA VVITMSRHAE LFVGLH-PSV PVISIDADHN
YP_0037122       LDPDYPTERL AYMLEDAAPV VLLTQTSQLD KLSG----TM PVVILDT---
YP_0034678       LDPAYPAERL AYMLDDAAPV VLLTQTAWVD TLVSPVTTSV PIIVLDA---
YP_0034678       LDPAYPTERL AYMLNDAAPV ALLTQAAQVG TLAS----TV PTVVLDG---
Clustal Co       :*.  *  *    :  *      ::                    :  :
SM23             PWE-EIADQ- AVDNLPAQA- GPDSLAYVMY TSGSTGRPKG VEVVHRGVVR
ZP_0677648       ----PVPVVR PVRASPAPGP GPDRLAYVSY TSGSTGEPKG VAVPHRAVDR
YP_0015446       IAA-HLAAY- PTTA-PTVDI SPENLAYVMY TSGSTGRPKG IMINQRNIVR
ZP_0051751       -----QEIYQ ESQNTLPTDT TAEHLAYVMY TSGSTGIPKG ICIPHRGVTR
ZP_0711340       DWE-VIERQ- -SEENPSPQV IHDNLAYVMY TSGSTGIPKG VSVIHQGVVR
AAW55330         DWH-TIAQN- -KKENLSTNV TAENLAYVMY TSGSTGTPKG VSVIHRGVVR
P2706            RTILSLPLDE NDEENPETAV TAENLAYMIY TSGTTGQPKG VMVEHHALVN
P94459           ----TNSGLA HRLSNPNVDV LPQSLAYVIY TSGSTGMPKG VEIEHRSAVN
Q9R919           -----PQTFS EDGSNLESIS GPNQLAYVIY TSGSTGKPKG VMVEHRSVIN
YP_0037122       DETLLATQPI DNPDIQALGL TSHHLAYVLY TSGSTGQPKG VMTEHRNVLR
YP_0025514       EWST-M---- -SGAPPEMGG NDSRLAYICY TSGSTGTPKG VMIDHAAVVR
YP_0037122       QNALLESQSI HNPETQMQGL TSRHLAYVIY TSGSTGQPKG VMVEHRNVLR
YP_0034678       QEPAVAAQPT HNPEPQTLGL TSRHLAYVIY TSGSTGLSKG VMVEHRNVLR
YP_0034678       QDASLMAQPT HNPDTQALGL TSRHLAYVIY TSGSTGLPKG VMVEHRNVLR
Clustal Co                                ***: * ***:** .** :   :    .
150
SM23             LVCG-TDYVE LGPGEAILQF APLSFDASTF EIWAALLHGG RLAVFPPG-L
ZP_0677648       LVRG-ADWME VRPGDVFFHI APVAFDASTL EIWAPLVNGC RLAVFPPG-T
YP_0015446       LVRN-TTYAA FGPDQVGLLL ATVAFDASTF ELWGCLLNGG RLVIAPPQ-Q
ZP_0051751       LVKN-SNYVA LGEDDIFLQA APYTFDASTF EIWGALLNGG RLVILPSQ-T
ZP_0711340       LVKD-TNYVN LSAEEVFLQL APISFDASTL EIWGSLLNGG RLVIMPPH-T
AAW55330         LVKE-TNYAH LTAEEIILQL APISFDASTF EIWGCLLNGG QLVICPPH-T
P2706            LCFWHHDAFS MTAEDRSAKY AGFGFDASIW EMFPTWTIGA ELHVIEEAIR
P94459           FLNSLQSRYQ LKHSDMIMHK TSYSFDASIW ELFWWPYAGA SVYLLPQGGE
Q9R919           RLVWMQENYP LDERDAILQK TAITFDVSVW ELFWWSIVGS KVVLLPNGGE
YP_0037122       LIIN-SGFAD IGPDDCIAHC ANMAFDASTW EIWSALLNGA RLHVVSPSVL
YP_0025514       TVMA-TDYAN FGVRETFLQF APLAFDASTF EIWGALLNGG RLVFAPPG-K
YP_0037122       LIIN-NGFAD IGPDDCIAHC ANMAFDASTW EIWSALLNGG CLHVVSQPVL
YP_0034678       LIIN-NGFAD IGSDDCIAHC ANIAFDASTW EIWSALLNGG RLYVVPPSVL
YP_0034678       LIIN-NGFAD IGSDDCIAHC ANIAFDASTW EIWSALLNGG RLHVVSPSVL
Clustal Co                .   :      :   **.*   *::     *   : .
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
SM23             PSIDELGRFI HDRRITTLWL TAGLFQQMVD FGLE------ ---HLSGVRQ
ZP_0677648       IALAEVARTV RAEGVTVLLL TTGLFHRMAG SHPE------ ---AFAGVRH
YP_0015446       LSLAELGHLV EREQITTLWL TAGLFHQMVD HALD------ ---RLGSLRQ
ZP_0051751       PSLEEIGETL ENYGVTTLWL TAGLFQVMVE EKLE------ ---SFKNVRY
ZP_0711340       PSLQELGEAI WGYQITTLWL TAGLFHIMVD EHLE------ ---DLKQVRQ
AAW55330         PSLEELGQII QQYQVTTLWL TAGLFHLIVD EKID------ ---ALKSLRQ
P2706            LDIVRLNDYF ETNGVTITFL PTQLAEQFME L--------- ---ENTSLRV
P94459           KEPEVIAKAI EEQKITAMHF VPSMLHAFLE HIKYRSVP-- IK-TNRLKRV
Q9R919           KNPELILDTI EQKGVSTLHF VPAMLHAFLE S-MEQTPSGK LKRKLASLRY
YP_0037122       LDPVRFCDSL MQGQVTALWL TAGLFHEYLD TLKP------ ---LYGQLRY
YP_0025514       VGLDEVCDLV QKFNVTTLWL TAGIFQLLSE EHLQ------ ---CLFSLRQ
YP_0037122       LDPVRFCDSL IRGKVTGLWL TAGLFNEYLD TLKP------ ---VFRQLRY
YP_0034678       LDPVRFCDSL IKGQVTALWL TAGLFNEYLS DLNP------ ---LLGRLRY
YP_0034678       FDPVRFRDSL IKGQVTALWL TVGLFNEYLS DIQP------ ---LFGQLRY
Clustal Co             .   .     ::   :    : .                       *
SM23             LLAGGDVVPP AHAA---RA- LAALPECCLI NGYGPTENTT FTCCHRMATP
ZP_0677648       VLTGGDVASP SHVE---RL- LTLHPGLVYT NGYGPTENTT YTTCWTSDTL
YP_0015446       LLAGGDRLSP VHVH---KV- LERWPQCRLI NGYGPTENTT FSCCQQLSAT
ZP_0051751       LLAGGDVLSP THVK---TV- LQTYPHCSVI NGYGPTENTT FTCCSVLTDV
ZP_0711340       LLAGGDILSV PHVQ---KV- IQELKGCQLI NGYGPTENTT FTCCYRITEV
AAW55330         LLAGGDVLSV LHVQ---KF- LQTVENCRLI NGYGPTENTT FTCCHLITAP
P2706            LLTGGDKLKR AVKK---PY- -------TLV NNYGPTENTV VATSAEIHPE
P94459           -FSGGEQLGT HLVS---RF- YELLPNVSIT NSYGPTEATV EAAFFDCPPH
Q9R919           VFASGEALTP KHVDGFQRI- ITPVSHAQII NLYGPTEATI DVSYFECEAD
YP_0037122       LLVGGDILDP GKIQ---QVK LAESQPAHLI NGYGPTETTT FATTYDIASP
YP_0025514       LLAGGDVLSL DTIN---RV- NKALPNCQVI NGYGPTEATT FSVCHAF--P
YP_0037122       LLIGGDVLDP NKIQ---QVQ LAESKPTYLI NGYGPTETTT FAATYTIPSS
YP_0034678       LLIGGDVLDP RKIQ---RAQ LAESQPAHLI NGYGPTETTT FATTYRIASP
YP_0034678       LLIGGDVLDP QKIR---RTQ LSEFQPAHLI NGYGPTETTT FAVTYTIASP
Clustal Co        : .*:                          * ***** *
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         260        270        280        290        300
SM23             KDVGPTVSIG RPIANTRVYV LDRQGRPVPW GVPGELYAAS DGLARGYLAR
ZP_0677648       TNRE-RVPIG GPISGTRIAV LDSELRPVPA GECGELYAAG AGLARGYLNR
YP_0015446       TDLAQGVPIG QPIANSTAYI LDRLLQLVPI GVVGELYLGG AGLARGYLAR
ZP_0051751       EQIGYSVPIG QPISQTQVYI LDNYLQPVPF GVPGELYIGG DGLARGYLNR
ZP_0711340       NLIENSIPIG RSISNTQVYL LDTHLQLVPI GVPGELYIGG DGLARGYLNR
AAW55330         VQPGVSIPIG RPIANTQVYI LDNNFQTVAI GEIGELHIAG DGLARGYLNR
P2706            E---GSLSIG RAIANTRVYI LGEGNQVQPE GVAGELCVAG RGLARGYLNR
P94459           EKLE-RIPIG KPVHHVRLYL LNQNQRMLPV GCIGELYIAG AGVARGYLNR
Q9R919           KRYN-SVPIG KPISNIQLYI L-QAGYMQPV GVAGELCIAG DGLARGYLNR
YP_0037122       VDVTRSIPIG RPIGNTRIYI LDSRGQPVPL GIVGEIHIAG AGVARGYLNR
YP_0025514       KGIATEIPIG KPIANTKVYV LDKCLAPVPI GVVGELYIAG RGVGRGYLNH
YP_0037122       VDVARSIPIG RPIANTQIYI LDSQGRPVPV GVAGEIYIAG NGVARGYLNR
YP_0034678       VDVAHSIPIG RPIANTRIYI LDCHNQPVPL GVAGEIYIAG AGVARGYLNR
YP_0034678       VDVTRSIPIG RPIANTRIYI LDSLGQPVPF GVAGEIHIAG AGVARGYLNR
Clustal Co             :.**  .:      : *       .  *  **:  ..  *:.**** :
```

```
                ....|....| ....|....| ....|....| ....|....| ...
                       310        320        330        340
SM23            PELTAERFLP DPFSEEPGAR MYRTGDLVRW RPDGTLEFLG RMD
ZP_0677648      PGATAERFLP DPSGTEPGAR MYRTGDLVRW TPDGTLEFVG RAD
YP_0015446      PDQTAAAFIP NPMSQTAGER LYRSGDLARY RDDGTIEFIG RRD
ZP_0051751      PQLTAERFIA SPFAT--GER LYKTGDLVRY DRQRNIEFLG RKD
ZP_0711340      PELTAERFIL NPFSDKPSDR LYKTGDLARY LPDGNIEFLG RID
AAW55330        PELTAEKFIS HSFDSNLATR LYKTGDLARY LPDGNIEFLG RID
P2706           EDETAKRFVA DPFVP--GER MYRTGDLVKW TGGG-IEYIG RID
P94459          PALTEERFLE DPFYP--GER MYKTGDVARW LPDGNVEFLG RTD
Q9R919          PELTAEKFVK NPFSA--GER MYRTGDLARW LPDGNIEYLG RID
YP_0037122      PELTTERFLL DPFSQGTHAR MYKTGDLGRW LPEGNIEYLG RYD
YP_0025514      PSLTCEKFIS SPFGD-SGDR LYRTGDLVRW GRDGLLRFLG RAD
YP_0037122      PELTAERFLA DPFSQDTDAH MYKTGDLGRW LADGNIEYLG RND
YP_0034678      PELTAERFVP DTFSADPDER MYKTGDLGRW LFDGNIDYLG RND
YP_0034678      PELTAERFLA DPFSSDPDAR MYRTGDLGRW RPDGNIDYLG RND
Clustal Co         *   *:   .       : :*::**: ::      : ::* * *
```

Figure 5B. Sequence alignment of SM23 NRPS with protein sequences with similar conserved core sequences from database

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
SM27             LDEVMPGER- QSLLAKDAGA KWIVSNRGQG ----LAPE-L SDL-----GR
YP_0028721       LDINAPAER- QGFMLQDSGA AWLLTRS--- ------DASV DYP----AQR
ABH06368         LDVNAPPER- QAFMVQDSGA R--------- ---------- --QRLDNLAE
NP_902472        LDEGLTAERR QWLQADSGAK L---SVDAA- ---W-L---- ---------A
YP_0037122       LDPDYPTERL -AYMLEDAAP VVLLTQTSQL ----DKLSG- TMPVVI-L-D
AAO72425         LDINAPAER- QAFMLQDCGA RQVLTL-S-- --RHDLPDGI --QRID----
YP_235693        LDINAPAER- QAFMLQDCGA RQVLTL-S-- --RHDLPDGI --QRID----
ZP_0726607       LDINAPAER- QAFMLHDCGA RQVLTL-A-- --RHDLPEGI --QRID----
P39845           LDPAYPKERL SYMLKDSGAS LLLTQP---- --GCSAP-NF S---VD--MT
Q9R919           IDPTYPEERI RYILEDSDTK LLLVQH---- HLREKVP--F TGKVLD--ME
Clustal Co       :*   . **       ..
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
SM27             VD-VDS-EEV LT--QSTDNP EI-VANGES- -LAYVMYTSG STGQPKGVLV
YP_0028721       LD-LDT---L VLDPQPSHNP DL-SQSSDS- -VAYIMYTSG STGTPKGVLV
ABH06368         L-NLDVMPAT --------NP AV-AQSSDS- -VAYIMYTSG STGTPKGVLV
NP_902472        TPG-TWPEE- --------NP AV-AGDAES- -VAYLMYTSG STGEPKGVLA
YP_0037122       TQNALL--ES QSIHNPETQM ----QGLTSR HLAYVIYTSG STGQPKGVMV
AAO72425         LDLLELQSDA -------PNP -VHSASAES- -VAYIMYTSG STGMPKGVLV
YP_235693        LDLLELQSDA -------PNP -VHSASAES- -VAYIMYTSG STGMPKGVLV
ZP_0726607       LDLLQLPGDT -------PNP -VPSASAES- -VAYIMYTSG STGMPKGVLV
P39845           SLA-SEKAE- --------NH EFTPADGGS- -LAYVIYTSG STGQPKGVAV
Q9R919           DPQ-TFSEDG -------SNL ESISGP-NQ- -LAYVIYTSG STGKPKGVMV
Clustal Co                          :          .   :**::**** *** **** .
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
SM27             QQRGV-SRLV LNNGYAAFSA EDRVAFAANP AFDASTMEVW GALLNGGEVI
YP_0028721       PHRGI-TRLV LNNGYADFNA SDRVAFASNP AFDASTMDVW GPLLNGGQVQ
ABH06368         THRGI-SRLV INNGYADFNP HDRIAFASNP AFDASTMDVW GALLNGGQVQ
NP_902472        PHRGI-TRLV CGNRYAAFQA DDRIAWAANP AFDASTLEIW GALAHGASLV
YP_0037122       EHRNVL-RLI INNGFADIGP DDCIAHCANM AFDASTWEIW SALLNGGCLH
AAO72425         PHRAV-SRLV LNNGYADFNA GDRVAFASNP AFDASTLDVW APLLNGGCVV
YP_235693        PHRAV-SRLV LNNGYADFNA GDRVAFASNP AFDASTLDVW APLLNGGCVV
ZP_0726607       PHRAV-SRLV LNNGYADFNA QDRVAFASNP AFDASTLDVW APLLNGGCVV
P39845           EHRQAVSFLT GMQHQFPLSE DDIVMVKTSF SFDASVWQLF WWSLSGASAY
Q9R919           EHRSVINRLV WMQENYPLDE RDAILQKTAI TFDVSVWELF WWSIVGSKVV
Clustal Co        :*     *    :    :    * :   :   :**.*. :::      *.
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                        160        170        180        190        200
SM27             VIEKTALLDT GLFKSALEEN GVTVLFLTTA LFNQYAHSIP KT----LA--
YP_0028721       VIDHATLLDP TVFGAALAD- -VTVLFVTTA LFNQYVQLIP QA----LA--
ABH06368         VIDHATLLDP LAFGAELK-- GATVLFVTTA LFNQYVQLIP QA----LA--
NP_902472        AIDKDTLLSA EALGARLQRD RITILWLTAG LFQRHARGLG AA----LS--
YP_0037122       VVSQPVLLDP VRFCDSLIRG KVTGLWLTAG LFNEYLDTLK PV----FR--
AAO72425         VVEQSVLLSL DEFRALLLSQ SVSVLWMTAG LFHQYASGLM EA----LA--
YP_235693        VVEQSVLLSL DEFRALLLSQ SVSVLWMTAG LFHQYASGLM EA----LA--
ZP_0726607       VVAQSVLLSL DEFRALLLSQ SVSVLWMTAG LFHQYADGLM EA----FA--
P39845           LLPPGWEKDS ALIVQAIHQE NVTTAHFIPA MLNSFLDQAE IE---RLSDR
Q9R919           LLPNGGEKNP ELILDTIEQK GVSTLHFVPA MLHAFLESME QTPSGKLKRK
Clustal Co        :      .    :   :      :   . .. ::: .            :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
SM27            --GLRYLLCG GERGDPSCFR RVL-EYNGPE H---LIHCYG PTETTTYAST
YP_0028721      --GLRILLCG GERADPAAFR SLL-ARAPAL R---LVHCYG PTETTTYATA
ABH06368        --GLRILLCG GERADPAAFR SLL-AQAPAL R---LVHCYG PTETTTYATT
NP_902472       --GLRYLMVG GDVVDPRVAA QVR-RDNPPA H---LLNCYG PTETTTFATT
YP_0037122      --QLRYLLIG GDVLDPNKIQ QVQLAESKPT Y---LINGYG PTETTTFAAT
AAO72425        --RLRYLIVG GDVLDPAVIA RVL-AEGAPQ H---LLNGYG PTEATTFSTT
YP_235693       --RLRYLIVG GDVLDPAVIA RVL-AEGAPQ H---LLNGYG PTEATTFSTT
ZP_0726607      --RLRYLIVG GDVLDPAVIG RVL-KEGAPR H---LLNGYG PTEATTFSTT
P39845          -TSLKRVFAG GEPLAPRTAA RFA-SVLPQV S---LIHGYG PTEATVDAAF
Q9R919          LASLRYVFAS GEALTPKHVD GFQ-RIITPV SHAQIINLYG PTEATIDVSY
Clustal Co         *: :: . *:   *      .             ::: ** ***:*   :
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        260        270        280        290        300
SM27            HEVRSVAADA K-TISIGRPI GNTTIYILDT NGQPVAPGVA GEIHIGGDGV
YP_0028721      YEVRSLAEDA D-SVPIGRPI SNTQIHVLDA QLQPVPLGVT GEICIGGDGV
ABH06368        YEVRALASDA D-SVPVGRPI SNTQIYVLDA QLQPVPLGIT GEICIGGEGV
NP_902472       HEIGAEAETA A-SLPIGKPI GNTRIYILDG DGQLAPLGVA GELYIGGAGV
YP_0037122      YTIPSSVDVA R-SIPIGRPI ANTQIYILDS QGRPVPVGVA GEIYIAGNGV
AAO72425        HEITSVGSG- --GIPIGRPI GNSQVYVLDT LRQPVAVGVA GELYIGGQGV
YP_235693       HEITSVGSG- --GIPIGRPI GNSQVYVLDT LRQPVAVGVA GELYIGGQGV
ZP_0726607      HEITSVGNG- --GIPVGRPI GNSQVYVLDT LRQPVAVGVV GELYIGGQGV
P39845          YVLDPERDRD RLRIPIGKPV PGARLYVLDP HLAVQPSGVA GELYIAGAGV
Q9R919          FECEA-DKRY N-SVPIGKPI SNIQLYILQA -GYMQPVGVA GELCIAGDGL
Clustal Co      .   .         :.:*:*:  .  :::*:       . *:. **: *.* *:
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
SM27            AKGYLNQAQL SAESFLPDPF SDKPEAKMYK TGDLAYWSAN GTIEYLGRND
YP_0028721      AKGYLNRPAL TAEKFVRDPF D--ADALMYR TGDLGRWTAG GLLECIGRND
ABH06368        AKGYLNRAQL TAEKFVNNPF VDQPGALMYR TGDLGRWSEE GLLECLGRND
NP_902472       ARGYLNRPEL TAERFIADPY SADPQARLYK TGDLGRWLPD GSIEYLGRND
YP_0037122      ARGYLNRPEL TAERFLADPF SQDTDAHMYK TGDLGRWLAD GNIEYLGRND
AAO72425        AKGYLNRPEL NATQFVANPF SDDAGALLYR TGDLGRWNAD GIVEYLGRND
YP_235693       AKGYLNRPEL NATQFVANPF SDDAGALLYR TGDLGRWNAD GIVEYLGRND
ZP_0726607      AKGYLNRPEL NATQFVANPF SDDAGALLYR TGDLGRWNGD GVIEYLGRND
P39845          ARGYLNRPAL TEERFLEDPF Y--PGERMYK TGDVARWLPD GNVEFLGRTD
Q9R919          ARGYLNRPEL TAEKFVKNPF S--AGERMYR TGDLARWLPD GNIEYLGRID
Clustal Co      *:****:. * .   *: :*:    .   :*: ***:. *    * :* :** *
```

Figure 6B. Sequence alignment of SM27 NRPS with protein sequences with similar conserved core sequences from database

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
YP_0037211       LDIAYPKERL AFILSDSQLS IILTQQHLVE RLP-RN-QAR VVCLDNDWED
SM48             LDPSYPEERL AFMLDDIRAT VLISQTGLQG KIPSKNKNIR TIFMDGDREV
YP_049592        LDPSYPAERL TYMLDDATPV ALLTQSALTA TLP--DTALP TVLLDAHDV-
NP_902472        LDPGYPPDRL SYMLADSSPK AMLTQTSLLP SLH--DWIGA QVVLDDVEEV
YP_259253        IDPAYPRERI AYTLQDSDPV ALLVQAGTQS LVA--DLRVP LIDLDSRT--
ZP_0492872       LDPRYPSDRL GYMIEDSGIR LLLTQRAARE RLPL-GEGLP CLLLDAEHE-
AAO72424         LDPAYPPERL AYTLGDSTPV ALLSQQSVQQ ALP--VSQVP VIYLDDAG--
AAO72425         LDPAYPLERL AYTLGDSAPV ALLSQRSVQS TLP--ASEVP VISLDDD---
ZP_0564170       LDPAYPLERL AYTLGDSAPL ALLSQRSVQH ALP--VSDVP VISLDDAD--
ZP_0415419       IDPTIPKARL DYFIQDSGIN LLLTQDDLSN --EYCTQNID KILLDKDWLK
YP_0016425       LDPSYPESRL RYILEDTGIQ VLVTNEALEG --WI-TEEIK TVCLDRDKAM
ZP_0323242       LDPTYPEQRL QYILEDASIQ LFVTQESLKE LNWL-PENVE SICLDRDQDE
YP_0036647       LDPTYPEQRL QYILEDASIQ LFVTQESLKE LNWL-PENVE SICLDCDQDE
NP_832214        LDPTYPEQRL QYILEDASIQ LFVTQESLKE LNWL-PENVE SICLDCDQDE
ABG57051         LDPTYPEQRL QYILEDASIQ LFVTQESLKE LNWL-PENVE SICLDRDQDE
O30409           IDPDYPLERQ AFMLEDSEAK LLLTLQKMNS QV---AFPYE TFYLDTET--
P0C064           IDIDYPQERI SYMMEDSGAA LLLTQQKLTQ QI---AFSGD ILYLDQEE--
Clustal Co       :*   *  *   : : *      ::                    . :*

YP_0037211       --IVKIPIQH TEIT---VEP DNLAYIIYTS GSTGKPKGVI ILHHNVVCLF
SM48             --ISGQNLEN PLNSA---SP DNLAYIIYTS GSTGKPKGVM ITRYNVVRLF
YP_049592        --FDAQPDHN PDAHALGVTP DHLAYVIYTS GSTGKPKGVM VEHASVTRLL
NP_902472        DRLSRLPDHN PDAARRGLTS SHLAYIIYTS GSTGAPKGVM VEHRQVVRLF
YP_259253        --LAHEAQDD PEVP--GLTP AHLAYVIYTS GSTGLPKGVM VEHRNVARLF
ZP_0492872       --WAGYPESD PQSA---VGV DNLAYVIYTS GSTGKPKGTL LPHGNVLRLF
AAO72424         --LQDESVDN PQIS---VKP DNLAYVIYTS GSTGLPKGVM VEHRNVARLF
AAO72425         --LQGESVCN PQVP---VKP TNLAYVIYTS GSTGLPKGVM VEHRNVARLF
ZP_0564170       --LQDESASN PQVP---VKP TSLAYIIYTS GSTGQPKGVM IEHRNVARLF
ZP_0415419       --ISKESKEN LNSD---VHP GNLAYVIYTS GSTGDPKGTL IPHENITRLF
YP_0016425       --ISRESTLS PICE---VTG ENLAYVIYTS GSTGNPKGVM VEHHNVIRLF
ZP_0323242       --IGKESKTL PVSS---VGP QNLAYVIYTS GSTGNPKGVM IEHHNVIRLF
YP_0036647       --IGKESKTL PVSS---VGP QNLAYVIYTS GSTGNPKGVM IEHHNVIRLF
NP_832214        --IGKESKTL PVSS---VGP QNLAYVIYTS GSTGNPKGVM IEHHNVIRLF
ABG57051         --IGKESKTL PVSS---VGP QNLAYVIYTS GSTGNPKGVM IEHHNVIRLF
O30409           --VDQEETGN LEHV---AQP ENVAYIIYTS GTTGKPKGVV IEHRSYANVA
P0C064           --WLHEEASN LEPI---ARP QDIAYIIYTS GTTGKPKGVM IEHQSYVNVA
Clustal Co                           :**:**** *:** ***.: : : .   :
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
YP_0037211       AATQPWFQFN NNDVWSC-FH SYAFDFSVWE IWGALLYGGR LVIIPYYVSR
SM48             QSTRKWFHFN GEDVWTL-FH SFAFDFSVWE LWGALLHGGR LVVVPFWVSR
YP_049592        DATQDYFHFD SNDVWT-QFH SFAFDFSVWE IWGALAYGGK LVVVPTLCAR
NP_902472        GATDHWFHFG EQDVWSL-FH SFAFDFSVWE IWGALAHGGK LLIVPKDIAR
YP_259253        SATRDWFDFN WRDVWAL-FH SFAFDFSVWE IWGALVHGGQ LLVVPQAVSR
ZP_0492872       DATRHWFGFS ADDAWSL-FH SYAFDFSVWE IFGALLHGGR LVIVPYETSR
AAO72424         SATEDWFGFN EQDVWAL-FH SFAFDFSVWE IWGALLHGGR LLIVPQLVSR
AAO72425         SATEEWFGFN QQDVWAL-FH SFAFDFSVWE IWGALLHGGR LLIVPQLVSR
ZP_0564170       SATDDWFGFN EKDVWAL-FH SFAFDFSVWE IWGALLHGGR LLIVPQLVSR
ZP_0415419       ASTSKWFQFN ENDTWTL-FH SYAFDFSVWE IWGALLYGGK LVIVPYWVSR
YP_0016425       KSTECWYQFD EKDTWTL-FH SYAFDFSVWE IWGALLHGGR LIVVPYWISR
ZP_0323242       KSTDCWYQFN EKDTWTL-FH SYAFDFSVWE IWGALLYGGK LVVVPYWISR
YP_0036647       KSTDCLYQFN EKDTWTL-FH SYAFDFSVWE IWGALLYGGK LVVVPYWISR
NP_832214        KSTDCLYQFN EKDTWTL-FH SYAFDFSVWE IWGALLYGGK LVVVPYWISR
ABG57051         KSTDCLYQFN EKDTWTL-FH SYAFDFSVWE IWGALLYGGK LVVVPYWISR
O30409           FAWKDEYHLD SFPVRLLQMA SFAFDVSTGD FARALLTGGQ LVICPNGVKM
P0C064           MAWKDAYRLD TFPVRLLQMA SFAFDVSAGD FARALLTGGQ LIVCPNEVKM
Clustal Co        :    : :.    .    :  *:***.*. : :  **  **: *:: *

YP_0037211       SPELFYKLLS QEGITILNQT PSAFKQLIQL ET--SLNNHS DLSLRFVIFG
SM48             SPDRFLDLLI CQRVTVLNIT PSAFRQLIQE EGNASGAAGR EMALRLVIFG
YP_049592        SPQEFYSLLC RERVTVLNQT PGAFRQLIA- -----ARDDT DHSLRCIIFG
NP_902472        SPDQFYQLLC EQKVTVLNQT PSAFRQLIGA --Q--ARSSQ AHHLRYVVFG
YP_259253        SPDDCYRLLC EARVSILNQT PSAFRSLIAA --Q--DQSPL KHSLRQVIFG
ZP_0492872       SPEDFLRLLC RERVTVLNQT PSAFKQLMQV ACA--GQEVP PLALRHVVFG
AAO72424         SPEDFYTLLC STAVTVLNQT PSAFRQLITA --Q--GENQQ AHSLRQVIFG
AAO72425         SPEDFYNLLC SAGVTVLNQT PSAFRQLIAA --Q--AENTQ AHSLRQVIFG
ZP_0564170       SPEDFYNLLC SAGVTVLNQT PSAFRQLIAA --Q--GEQAQ AHSLRQVIFG
ZP_0415419       DTEKFYDLLI KEKVTILNQT PSAFYQLIKI DEKR-LLSPT QLSLRKVVFG
YP_0016425       SPKDFYQLLV KEKVTVLNQT PSAFRQLIQV CEQ--EDEKK DLHLRYVIFG
ZP_0323242       SPKDFYQLLV EEEVTVLNQT PSAFRQLIQV CEQ--EDKNK NLQLRYVIFG
YP_0036647       SPKDFYQVLV EEEVTVLNQT PSAFRQLIRV CEQ--EDKNK NLQLRYVIFG
NP_832214        SPKDFYQVLV EEEVTVLNQT PSAFRQLIRV CEQ--EDKNK NLQLRYVIFG
ABG57051         SPKDFYQLLV EEEVTVLNQT PSAFRQLIRV CEQ--EDKNK NLQLRYVIFG
O30409           DPASLYETIR RHEITIFEAT PALIMPLMHY VY---ENELD MSQMKLLILG
P0C064           DPASLYAIIK KYDITIFEAT PALVIPLMEY IY---EQKLD ISQLQILIVG
Clustal Co       ..      :     ::::: * *. .  *:                :: ::.*
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
YP_0037211      GEALDIQSLK PWVDKHRDKF TQLVNMYGIT ETTVHVTYRP INIDDLN-SS
SM48            GEALQMRTLK PWYERHEERC PLLVNMYGIT ETTVHVTYQP LKAADARENS
YP_049592       GEALELHMLA PWIADNPLER TRLINMYGIT EITVHATFRE LSAADITAGR
NP_902472       GEALETSMLA PWYARHIDHG PLLINMYGIT ETTVHVTYRP LSAEDVNRRG
YP_259253       GEALEPGMLK PWYAHLENVG TQLVNMYGIT ETTVHVTYRP LQAADAQLVG
ZP_0492872      GEALEVQALR PWFERFGDRA PRLVNMYGIT ETTVHVTYRP LSLADLDGGA
AAO72424        GEALETAMLK PWYARNVNAA TQLVNMYGIT ETTVHVTYYP LQPEDAQRVG
AAO72425        GEALETAMLK PWYARQANAG TQLVNMYGIT ETTVHVTYYP LQPEDAQRLG
ZP_0564170      GEALETAMLK PWYARNVNAA TQLVNMYGIT ETTVHVTYYP LQPEDAMRVG
ZP_0415419      GEALEYRLLR PWIQKYGDKV PQLVNMYGIT ETTVHVTYRP ITYEDIEKNI
YP_0016425      GEALDPTSLV PWFQRYGGQE PQLINMYGIT ETTVHVTYYP ITQDDVQHAS
ZP_0323242      GEALEPTSLL PWFQRYGEKN PQLINMYGIT ETTVHVTYYP ITLDDVQQAS
YP_0036647      GEALEPIGLL PWFQRYGEKK PQLINMYGIT ETTVHVTYYP ITLDDVQHAS
NP_832214       GEALEPIGLL PWFQRYGEKK PQLINMYGIT ETTVHVTYYP ITLDDVQHAS
ABG57051        GEALEPIGLL PWFQRYGEKK PQLINMYGIT ETTVHVTYYP ITLDDVQHAS
O30409          ADSCPAEDFK TLLARFGQKM -RIINSYGVT EACIDTSYYE ETDVTAIRSG
P0C064          SDSCSMEDFK TLVSRFGSTI -RIVNSYGVT EACIDSSYYE QPLSSLHVTG
Clustal Co      .::     :  .            ::* **:* *  :. ::

YP_0037211      PKVIGCAIPN LQLYILNSHL QPVPVGVAGE IYVGGAGLAR GYLNNLELTA
SM48            ASLIGRPIPD LQVYILDQNL HPVPVGVFGE IYVGGAGLAR GYLNRPQLTS
YP_049592       GSLIGRPLPD LRAYLLDPHG QPVPVGVAGE LYIGGAGVAR GYLNRPDLTA
NP_902472       ASPIGVKIPD LSVYILDANR QLAPLGVAGE LYIGGAGVAR GYLNRPELTA
YP_259253       SSPIGRRIPD LQLYVLDAHR EPLPSGVVGE LYVGGAGVAR GYLNRDQLTA
ZP_0492872      ASPIGEPIPD LSWYLLDAGL NPVPRGCIGE LYVGGAGLAR GYLNRPELSC
AAO72424        ASPIGTRIPD LQLYLLDTCG EPVPVGVVGE LYVGGAGVAR GYLNREALTA
AAO72425        ASPIGRRIPD LQLYVLDARG EPVPVGVVGE LYVGGAGVAR GYLNREALTA
ZP_0564170      ASPIGKRIPD LQMYVLDARG EPVPVGVVGE LYVGGAGVAR GYLNREALTA
ZP_0415419      KSMIGITIPD LYVLVLDAYM QPVPVGVQGE LFVGGAGLAR GYLNKPELTA
YP_0016425      RSNIGKQIPD LEVYVLDACQ QPVPIGVAGE LFIGGAGLAR GYLNRSELTA
ZP_0323242      RSNIGKRIPD LEVYILDAYQ QPVPIGVDGE LYIGGAGLAR GYLNRPELTA
YP_0036647      RSNIGKRIPD LEVYILDAYQ QPVPIGVDGE LYIGGAGLAR GYLNRPELTA
NP_832214       RSNIGKRIPD LEVYILDAYQ QPVPIGVDGE LYIGGAGLAR GYLNRPELTA
ABG57051        RSNIGKRIPD LEVYILDAYQ QPVPIGVDGE LYIGGAGLAR GYLNRPELTA
O30409          TVPIGKPLPN MTMYVVDAHL NLQPVGVVGE LCIGGAGVAR GYLNRPELTE
P0C064          TVPIGKPYAN MKMYIMNQYL QIQPVGVIGE LCIGGAGVAR GYLNRPDLTA
Clustal Co         **   .: :   :::    .  * *  ** : :****:** ****.  *:
```

```
                ....|....| ....|....| ....|....| ....|...
                         310        320        330
YP_0037211      ERFIPHPFNN QAKARLYKTG DLARYLPSGD IEYLGRID
SM48            ERFIPNSYCE KNGSRLYKTG DLARYLPDGS IEFLGRTD
YP_049592       ERFIVDPFSD SPATRLYKTG DLARWLPDGT LDYLGRND
NP_902472       ERFIADPYSA DPQARLYKTG DLGRWLPDGS IEYLGRND
YP_259253       ERFIADPFSH EPGARLYKTG DLARWRSDGS LEYLGRND
ZP_0492872      TRFVADPFST T-GGRLYRTG DLARYRCDGV VEYVGRID
AAO72424        ERFIDNPFNT APGARLYRTG DLGRWLADGT LEYLGRND
AAO72425        ERFLDNPFSH TADARMYRTG DLGRWLADGS LEYLGRND
ZP_0564170      ARFLDNPFST APGARMYRTG DLGRWMADGS LEYMGRND
ZP_0415419      TRFIDNPFSQ EPE-KLYRTG DLGKILLNGE IEYCGRID
YP_0016425      ERFIPHPFSS DPGARLYRTG DLARYLPDGN LDYLGRID
ZP_0323242      ERFISHPFSS DLKARLYRTG DLARYLPDGN LDYRGRID
YP_0036647      ERFISHPFSS NPKARLYRTG DLARYLPDGN LDYRGRID
NP_832214       ERFISHPFSS NPKARLYRTG DLARYLPDGN LDYRGRID
ABG57051        ERFISHPFSS NPKARLYRTG DLARYLPDGN LDYRGRID
O30409          EKFVPNPFA- -PGERLYRTG DLAKWRADGN VEFLGRND
P0C064          EKFVPNPFV- -PGEKLYRTG DLARWMPDGN VEFLGRND
Clustal Co       :*: ..:       ::*:** **.:   .*  ::: ** *
```

Figure 7B. Sequence alignment of SM48 NRPS with protein sequences with similar conserved core sequences from database

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
360305           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
360310           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENSVEII CLDSNQEAI-
360312           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
360802           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
360804           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
361101           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
361102           LNTDYPKDRL SFIMEDTRML VLLTQERLVA ALPENNVEII CLDSNQEAI-
P27206           IDPGFPAERI QYILEDCGAD FILTESKV-A A-PEADAELI DLD---QAI-
YP_0016425       LDPSYPESHL RYILEDTGIQ ILVTNEVSQG WMPE-EVETV CLDRDQAMI-
NP_832218        LDPSYPESRL RYILEDTGIQ VLVTNESLQD WIPK-EIKIV CLDRDQAMI-
ZP_0323232       LDPSYPESRL RYILEDTGIQ VLVTNESLED WIPK-EIKIV CLDRDQIMI-
YP_322131        LDPNYPVERL SYMLADSQLP ILLTQKHLLK QLPNNQTQTI CLDEDWQKL-
ZP_0711340       LDPTYPKERL AFMLEDASVP VLLTQTRLVE SLP-HQARVV CLDADWEVI-
PPSD_BACSU       IDPDYPEERI SFLLEDSGTN ILLLQSAGLH -VPEFTGEIV YLNQTNSGL-
ZP_0696580       LDPRYPSERL AFMLEDAQVS IILTRQDIVK KLPSHNAHFV RMDEDWKTL-
ABM21571         LDPNYPQERL SYMLADSGVE VLLAQKSLLE SLPSHTAQVV CLDSDWGVI-
ZP_0163206       LDPAYPQERL NFILQDAQLP IILTQQHFIT KLLPTSAKII CTDID---I-
YP_324595        LDPNYPSDRL AFMLNDAQLP VLLTQQQLVE KLPEHQAIAI CLDADWNEI-
YP_0018659       LDPTYPKERL SFMLSDSQVQ VVLTQEKFVD DLAASGAKLV CLD-DKKSF-
ACZ55942         LDPTYPKERL SFMLSDSQVQ VLLTQQKFVE SFADSGAKTV CLDQDWELI-
Q9R919           IDPTYPEERI RYILEDSDTK LLLVQHHLRE KVP-FTGKVL DMEDPQT---
YP_0018699       LDPGYPSERL GYALSDAQIS VLLTQQHLVE KLPEHQAQVV YLDQNWDAIL
NP_522203        LDPSYPQDRL TYMLEDSAPV AVLTQGLVRE QLGMLSVPVL DLDGPQE---
CAQ48254         LDPEYPLERL SFMLEDAAVN VLLTQQKLIN KLPEHQAQLI CLDADWELI-
AAZ03550         LDPDYPIERI IFMLEDAAVK VLLTQQKLIN KLPEHQAQLI CLDADWELI-
CAC01603         LDPEYPTERL TFMLADAQVS VLLTQQHLVE KLPENQEPVV CLDTDWLVI-
AAO23333         LDPEYPQERL TFMLADAQVS VLLTQQHLVE KLPRHQARVV HLDKDWVAI-
YP_0016588       LAPDYPTERL GDILSDSGVS LVLTQESLGD FLPQTEAELL CLDRDWEKI-
ABO45744         LDPNYPPERL DYMISDSAIS LLLTQQSLVQ FLPENQAEIL CLDTDWLKI-
BAA83994         LDPNYPPERL DYMISDSAIS LLLTQQSLVQ FLPENQAEIL CLDTDWSRI-
YP_324355        IDPEYPQERI AYMLEDSQVK VLLTQEKLLN QIPHHQAQTI CVDREWEKI-
AAF15891         LDPDYPQERL SFMLEDAQLR VLLTQHQLKE KLPQHQGQVV CLDTDWQFI-
AAO62586         LDPSYPKERL SYMLEDTGVK VLLTQRSLTE LLPENQAIVV SLDGDWQVI-
AAF00960         LDPNYPQERL SYLLEDTGVK VIITGESLRG LLDEYRGIVV ALDTDWSAI-
YP_0016588       LDPNYPQERL SYLLEDTGVK VIITAESLRG LLGEYRGIVV ALDTDWPAI-
Clustal Co       : . :* .::    : *      ::                :   :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90        100
360305          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
360310          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
360312          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
360802          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
360804          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
361101          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
361102          ------IQES GQDAPSPVTV DNLAYVIYTS GSTGQPKGVG VQHRSLCNHL
P27206          ------EEGA EESLNADVNA RNLAYIIYTS GTTGRPKGVM IEHRQVHHLV
YP_0016425      ------SQEN TLSPICKVTG ENLAYVIYTS GSTGNPKGVM VQHHSVLNLS
NP_832218       ------SQES ILSPKCEVTG ENLAYVIYTS GSTGNPKGVL IQHHSVLNLS
ZP_0323232      ------SQES ILSPKCEVTG EDLAYVIYTS GSTGNPKGVS IQHHSVLNLS
YP_322131       ------ANYS DENPCSQVKS DNLAYIIYTS GSTGKPKGTM IVHRGVVNYL
ZP_0711340      ------ERQS EENPSPQVIH DNLAYVMYTS GSTGIPKGVS VIHQGVVRLV
PPSD_BACSU      -------AHR LSNPNVDVLP QSLAYVIYTS GSTGMPKGVE IEHRSAVNFL
ZP_0696580      ------AQQN GNNPRSETIA HNLAYIIYTS GSTGTPKGVL VSHQSLCNLA
ABM21571        ------EQHS QENLDVGVCS DNLAYVIYTS GSTGVPKGVG IEHFSLCNLI
ZP_0163206      ------HSQP SDNPSSSVKS DNLAYVIYTS GSTGKPKGVM VAHRGLCNLA
YP_324595       ------AKNN SFNPTSTVTT ANLAYVIYTS GSTGKPKGVM VEHTGLCNLA
YP_0018659      ------HQES NENPSSGVAP ENLAYVIYTS GSTGTPKGVL IQHQGVCNLA
ACZ55942        ------TRQN QENPTSDVTA ENLAYVIYTS GSTGTPKGVM IQHRGVCNLA
Q9R919          ------FSED GSNLESISGP NQLAYVIYTS GSTGKPKGVM VEHRSVINRL
YP_0018699      TADYAYAQFP KDNVHSQVQP TNLAYVLYTS GSTGKPKGVA IEHHSPVALV
NP_522203       -------DAE HDPQVEALKP HHLAYVIYTS GSTGRPKGVM NEHRGVVNRL
CAQ48254        ------FQFS RDNLITDIQA TNLAYVIYTS GSTGQPKGVM LSHSNLSNHM
AAZ03550        ------SQFS QDNPITDVQA TNLAYVIYTS GSTGQPKGVM LSHSNLSNHM
CAC01603        ------CESS QESPITEVQP GNLAYVIYTS GSTGTPKGVM LSHSNLCNHM
AAO23333        ------AKSS QENPIAQVQA SNVAYVIYTS GSTGQPKGVI LSHSNLCNHM
YP_0016588      ------ATYS PENPFNLTTP ENLAYVIYTS GSTGKPKGVM NIHRGICNTL
ABO45744        ------ANYS QENLTSPVKP ENLAYVIYTS GSTGKPKGVM NIHRGICNTL
BAA83994        ------ANYS QENLTSPVKP ENLAYVIYTS GSTGKPKGVM NIHQGICNTL
YP_324355       ------STQA NTNPKSNIKT DNLAYVIYTS GSTGKPKGAM NTHKGICNRL
AAF15891        ------SQSS QENLITTVQA SNLAYVIYTS GSTGKPKGAM NTHLGICNRL
AAO62586        ------AQEN QNNLNSGVKG ENLAYVIYTS GSTGKPKGAM NTHKGISNRL
AAF00960        ------SQES QNNCDSGVTG ENLAYVIYTS GSTGKPKGVM NNHKGIRNRL
YP_0016588      ------SQES QNNCDSGVTG ENLAYVIYTS GSTGKPKGVM NNHKGIRNRL
Clustal Co                            :**::*** *:** ***.    *
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
360305           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
360310           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
360312           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
360802           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
360804           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
361101           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
361102           YWVKRSLFSE AVH---SIPV TANLS--FDA SLKQIFAPLL QGTEVWILSE
P27206           ESLQQTIYQS GSQ-TLRMAL LAPFH--FDA SVKQIFASLL LGQTLYIVPK
YP_0016425       YGLQKEVFSH RAHDNMRVGL --NASIAFDS SVKQLQM-LL YGSSLYIIST
NP_832218        HGLQKEVFEH EIPSNMHVGL --NASIAFDA SIQQLQM-LL YGSSLYIIPN
ZP_0323232       YGLQKEVFEH EIPSNMHVGL --NASIAFDA SIQQLQM-LL YGSSLYIIPS
YP_322131        SWCTKAYDVA AGV---GSTV --NSSLSFDA TITSLFSPLL VGAKVLLLPE
ZP_0711340       K-DTNYVNLS AEE---VFLQ LAPIS--FDA STLEIWGSLL NGGRLVIMPP
PPSD_BACSU       NSLQSRYQLK HSD---MIMH KTSYS--FDA SIWELFWWPY AGASVYLLPQ
ZP_0696580       TAQIQVFHVS PQS---RVLQ FASLN--FDV SISEILMALL AGATLYLGSQ
ABM21571         QAQKNLFYLE PNS---RVLQ FASIS--FDA SVSEIFIALT SGAMLILAIA
ZP_0163206       TAQIKLFEVR PDS---SVLQ FASIS--FDA SISEIVMAIC AGAKLCLATR
YP_324595        KAQIQTFDVQ TSS---RILQ FASFS--FDA SIFEVVMALG TGARLYLGTK
YP_0018659       QAQVKLFNVQ QNS---RVLQ FASFS--FDA SVWEIFMALC SGASLYIGTQ
ACZ55942         QAQVKLFGVN QNS---RVLQ FASFS--FDA SVSEIVMALC SGASLYLGNQ
Q9R919           VWMQENYPLD ERD---AILQ KTAIT--FDV SVWELFWWSI VGSKVVLLPN
YP_0018699       AWAKEVFTPE QLA---GVLA CTSIC--FDL SVFELFVPLS WGRKVILAEN
NP_522203        WWAQQTYRLD ASD---RVLQ KTPFG--FDV SVWELFWPLL AGARLVMARP
CAQ48254         FWMQETFPLT RAD---RVLQ KTSFS--FDA SVWEFYAPLL VGGQLLIAQP
AAZ03550         FWMQETFPLT KTD---RVLQ KTPFS--FDA SVWEFYAPLL VGGQLLIAQP
CAC01603         SWMQATFPLT EKD---KVLQ KTPFG--FDA SVWEFYAPLL AGGQLLIAKP
AAO23333         FWMQATFPLT KED---KVLQ KTPFG--FDA SVWEFYAPLL AGGQLLIAEP
YP_0016588       KYTIGHYNIT SED---RILQ IISLS--FDG SVWEIFSSLI SGASLVVAKP
ABO45744         KYNIDNYNLN SED---RILQ ITPFS--FDV SVWEVFSSLT SGATLVVAKP
BAA83994         KYNIDNYNLN SEE---RILQ ITPFS--FDV SVWEIFLSLT SGATLVVAKP
YP_324355        LWMQEAYQID STD---SILQ KTPFS--FDV SVWEFFWTLL TGARLVIAKP
AAF15891         LWMQQAYQLT ALD---CILQ KTPFS--FDV SVWEFFWPLI TGARLVVAKP
AAO62586         VWMQNTYQLT SSD---RILQ KTPFS--FDV SVWEFFWPLL AGATLVVVKP
AAF00960         LWMQDTYQLT KSD---CILQ KTPFS--FDV SVWEFFWPLL AGATLVVAEP
YP_0016588       LWMQDTYQLT KSD---GILQ KTPFS--FDV SVWEFFWPLL AGATLVVAKP
Clustal Co                                     **  :  ..       *  : :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
360305          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
360310          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
360312          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
360802          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
360804          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
361101          ELTNQP---V ALLRAINSRT NVGLNCVPSP WTVILEEISC CRARQSAATL
361102          ELTNQP---V ALLRAINSRT NVGLNCVPSL WTVILEEISC CRARQSAATL
P27206          KTVTNGAALT AYYRK-NSIE ATDGTPAHLQ --MLAAAGDF EGLK-----L
YP_0016425      EVRSDPQQFI SYIRE-NKLE MFDITPSLLQ --LLIDEGLL ETND-SVHVP
NP_832218       EVRSDPEQFV AYIRE-NKLE IFDITPSLLQ --LLIDAGLL ETCD-GVHAP
ZP_0323232      EVRSDPEQFV AYIRE-NKLE IFDITPSLLQ --LLIDVGLL ETCD-GVHVP
YP_322131       EEEIEALK-T ALCSGTKFSL VKITPAHLEI LS---HLFTS EAVN---IQA
ZP_0711340      H-TPSLQELG EAIWGYQITT LWLTAGLFHI MV---DE-HL EDLK----QV
PPSD_BACSU      GGEKEPEVIA KAIEEQKITA MHFVPSMLHA FL---EHIKY RSVPIKTNRL
ZP_0696580      EAILPGTVLL HFLQQNAITI ATFPPAVLKA LP---DA--- -LLP----SL
ABM21571        SELIPGSDLK QILQERCVTH VTLPPSALAV LA---TD--- -EFP----AL
ZP_0163206      DSLQPGQPLQ KLLQIQNISH VTLVPSALAA LS---PQ--- -DLP----NL
YP_324595       ESLLPGSSLI QLLQKYGITH ITLPPSALAV LP---AD--- -ELP----AL
YP_0018659      DSLRPGIDLM RLLQEQSITH VTLPPTALAA LP---KE--- -ELP----NL
ACZ55942        DSLRPGIDLI RFLRQQSITH ATLPPTALAA LP---KE--- -ELP----NL
Q9R919          GGEKNPELIL DTIEQKGVST LHFVPAMLHA FL---ESMEQ TPSGKLKRKL
YP_0018699      ALHLPTLP-- ---AAEQVTL INTVPSVITE LI------RI NGLPG---GV
NP_522203       EGHKAPAYLA ATIEQAGITT LHFVPSMLQL FL------DQ VEAGRCQ-GL
CAQ48254        GGHTDSDYLL KTIAQQQVTT VQLVPSLLQM LL------EQ GGIENCQ-LL
AAZ03550        GGHTDSDYLL KTIAQQQVTT VQLVPSLLQM LL------EQ GGIENCQ-LL
CAC01603        GGHTDSAYLL RLIAQQQVTI VQLVPSLLQM LL------EQ GGIETCH-SL
AAO23333        RGHTDSAYLL RLIAQQQVTT IQLVPSLLQM LL------EQ GGIETCH-SL
YP_0016588      DGYKDIDYLI DLIVQEQVTY FTCVPSILRV FL------QH PKSKYCH-YL
ABO45744        DGYKDIDYLI DLIVQEQVTY FTCVPSILRV FL------QH PKSKDCH-CL
BAA83994        DGYKDIDYLI DLIVQEQVTC FTCVPSILRV FL------QH SKSKDCH-CL
YP_324355       GGHKDSAYLI DLITQEQITT LHFVPSMLQV FL------QN RHVSKCS-SL
AAF15891        GGHKDSAYLV NLILEQQVTH VHFVPSMLQV FL------EE QNLENCR-SL
AAO62586        QGHKDNTYLI KLIQQQQITT IHFVPSMLRV FL------QE PSLENCS-CL
AAF00960        EGHKDSTYLI QLIQKQQITT LHFVPSMLRV FL------QE PELKGCS-SL
YP_0016588      EGHKDSTYLI QLIQKQQITT LHFVPSMLRV FL------QE PELKECS-SL
Clustal Co
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
360305          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TETTVNASAT
360310          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TETTVNASAT
360312          TCLLAGGETL SMELTD---R TRTALPHL-- --RIWNLYGP TETTVNASAT
360802          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TETTVNASAT
360804          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TEKTVNASAT
361101          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TETTVNASAT
361102          TCLLAGGETL SMELTD---R TRTALPHL-- --QIWNLYGP TETTVNASAT
P27206          KHMLIGGEGL SSVVAD---K LLKLFKEAGT APRLTNVYGP TETCVDASVH
YP_0016425      SKVLVGGEAI MPSLWE---Q LVEND-HI-- --HFYNVYGP TECTVDATCY
NP_832218       SKVLVGGEAI MPSLWE---Q LVETD-KI-- --QFYNVYGP TECTVDATCY
ZP_0323232      SKVLVGGEAI MPSLWE---Q LVETD-KI-- --QFYNVYGP TECTVDATCY
YP_322131       QAFIIGGEAL SEKIAS---F WKKRAPET-- --KLINEYGP TETVVGCCIY
ZP_0711340      RQLLAGGDIL SVPHVQ---K VIQELKGC-- --QLINGYGP TENTTFTCCY
PPSD_BACSU      KRVFSGGEQL GTHLVS---R FYELLPNV-- --SITNSYGP TEATVEAAFF
ZP_0696580      QTIISAGEAC SPDIVA---R WGH---NR-- --QFFNAYGP TETTVYATID
ABM21571        GQIIVAGEAC NLELAN---Q WSV---GR-- --RLFNGYGP TESTIGAAVA
ZP_0163206      KNLIVAGEPC PGDLAA---S WAV---GR-- --QFFNAYGP TEATVCATVL
YP_324595       QTIIVAGEAC PPDLVE---R WSR---GR-- --RFFNAYGP TEATVWSTVA
YP_0018659      QTLIVAGEAC NPKLIA---E WSK---GR-- --RFFNAYGP TESTICATVA
ACZ55942        QTLIVAGEAC NPKLIA---Q WSK---ER-- --RFFNAYGP TESTVCATVA
Q9R919          ASLRYVFASG EALTPKHVDG FQRIITPVSH A-QIINLYGP TEATIDVSYF
YP_0018699      STVNLAGEPL QNQLVQ---Q IYQQQIVK-- --YIFNLYGP SEDTTYSTFA
NP_522203       RRMLCSGEAL PHALQQ---R SLARFPHS-- --ELHNLYGP TEAAIDVTAW
CAQ48254        KRVFCGGEIL PVALQE---K LFSQL-NV-- --NLCNLYGP TECCIDVTFW
AAZ03550        KRVFCGGEIL PVALQE---K LLSQL-NV-- --NLCNLYGP TECCIDVTFW
CAC01603        KHVFCGGEVL PVALLE---G LLSKL-DV-- --NLHNLYGP TETCIDATFC
AAO23333        KHVFCGGEVL PVTLQE---S LLSKL-DV-- --NLHNLYGP TEACIDATFW
YP_0016588      KRVIVGGEAL SYELNQ---R FFQQL-NC-- --ELYNAYGP TEVAVETTIW
ABO45744        KRVIVGGEAL SYELNQ---R FFQQL-NY-- --QLYNAYGP TEAAVDATIW
BAA83994        KRVIVGGEAL SYELNQ---R FFQQL-NC-- --ELYNAYGP TEVAVETTIW
YP_324355       KRVICSGEAL SIDLQN---R FFQHL-QC-- --ELHNLYGP TEAAIDVTFW
AAF15891        KRVICSGEAL PVELQE---R FFARL-EC-- --ELHNLYGP TEAAIDVTYW
AAO62586        KRVICSGEAL PYELTQ---R FFERL-NC-- --ELHNLYGP TEAAIDVTFW
AAF00960        KRVFCSGEAL SLELTQ---R FFEHF-DC-- --ELHNLYGP TEAAIDVTYW
YP_0016588      KRVFCSGEAL SLDLTQ---R FFEHF-DC-- --ELHNLYGP TEAAIDVTYW
Clustal Co        .                                 : * *** :*
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
360305          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
360310          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
360312          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPISVP GEICIGGDGL
360802          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
360804          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
361101          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
361102          KIVP--GGN- --ITIGRPVA NTQIYLLDA- KLQPVPIGVP GEICIGGDGL
P27206          PVIPENAVQS AYVPIGKALG NNRLYILDQ- KGRLQPEGVA GELYIAGDGV
YP_0016425      RIKK----DS KRVTIGRPLP NVQAYVLDE- KLLPVPVGVT GELYIGGAGL
NP_832218       HIKK----DS KRVTIGRPLP NVQTYVLDS- NRLLVPVGVM GELYIGGVGL
ZP_0323232      HIKK--GSK- -RVTIGRPLP NIQTYVLDR- NRLPVPVGVM GELYIGGAGL
YP_322131       EVEK-LGYPG SNIPIGRPIA NTQLYILDS- HLQPVPIGVP GELYIGGDGV
ZP_0711340      RITE-VNLIE NSIPIGRSIS NTQVYLLDT- HLQLVPIGVP GELYIGGDGL
PPSD_BACSU      DCPP--HEKL ERIPIGKPVH HVRLYLLNQ- NQRMLPVGCI GELYIAGAGV
ZP_0696580      ECTS--RQE- K-ISIGRPIA NTQVYILDQ- EMQIVPVGIL GELYIGGAGV
ABM21571        QISH--GSEK --VTIGRPIA NTQIYILDK- HLEPVPISVS GELYIGGYGL
ZP_0163206      LYQP--GMK- --ISIGQAIA HTQIYILDH- YLQPVPIGVP GELHIAGVGL
YP_324595       ECSS--NSTN K-PPIGRPIT NTQIYLLDQ- DLQPVPVGVP GELHIGGIGL
YP_0018659      EYTG--DTQ- --LTIGRAIA NTQIYILAQ- DRQPVPIGTP GELYIGGDGL
ACZ55942        ECTF--GETQ --PTIGRAIA NIQIYILDH- NLQPVPIGVP GELYIGGDGL
Q9R919          ECEA--DKRY NSVPIGKPIS NIQLYILQAG YMQPV--GVA GELCIAGDGL
YP_0018699      LIEK--GTTF A-PPIGRPIA NTQIYILDE- YLQPVPVGVA GELHIAGAGL
NP_522203       RCNA--EIHP GVVPIGRPIA NTQIYVLDA- YRQPVPLGVT GEIYIGGAGV
CAQ48254        NCQR--EMYG QRIPIGRPIF NTQIYILDS- NLQPVPVGIP GELHIGGAGL
AAZ03550        NCQR--EMYG QRIPIGRPIS NTQIYILDS- NLQSLPVGIP GELHIGGAGL
CAC01603        NCQR--EIYA QIVPIGRPIS NTQIYILDQ- NLQALPVGVP GELHISGAGL
AAO23333        NCQR--EIYP QLIPIGRPID NTQIYILDQ- NLQPVPVGVP GELHIGGAGL
YP_0016588      CCQP--NSQ- --ISIGTPIA NAQVYILDS- YLQPVPIGVA GELHIGGMGL
ABO45744        CCQP--NSQL --IPIGRPIA NAQVYILDS- YLQPVPIGVA GELHIGGMGL
BAA83994        CCQP--NSQ- --ISIGTPIA NAQVYILDS- YLQPVPIGVA GELHIGGMGL
YP_324355       QCRK--DSNL KSVPIGRPIA NTQIYILDA- DLQPVNIGVT GEIYIGGVGV
AAF15891        QCFP--NGHL RTVPIGRAIA NTQIYILDE- HLQPVPVGVA GELHIAGVGL
AAO62586        HCLP--QIQQ QIVPIGRPIA NTQIYILDQ- YLQPVPVGIA GELHIGGVGL
AAF00960        PCLP--ESQK AIVSIGRPIA NTQIYILNP- HLQPVPIGIV GELHIGGIGL
YP_0016588      PCLP--ENQK ALVSIGQPIA NTQIYILNP- HLQPVPIGIV GELHIGGIGL
Clustal Co                 .** .:  : : *:*          .   **: *.* *:
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        310        320        330        340        350
360305          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
360310          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
360312          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
360802          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DCLFKT
360804          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
361101          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
361102          ARGYINRPEL TAERFIPNPF SDNHG----- ---------- ----DRLFKT
P27206          GRGYLHLPEL TEEKFLQDPF VPG------- ---------- ----DRMYRT
YP_0016425      ARGYLNRPEL TLERFIPHPF NEG------- ---------- ----ERLYRT
NP_832218       AKGYLNRPEL TSERFISHPF KEG------- ---------- ----ERLYRT
ZP_0323232      ARGYLNRPEL TSERFISHPF KEG------- ---------- ----ERLYRT
YP_322131       ARGYLNRPEL TQQKFIPNPF EKSQG----- ---------- ----SRLYKT
ZP_0711340      ARGYLNRPEL TAERFILNPF SDKPS----- ---------- ----DRLYKT
PPSD_BACSU      ARGYLNRPAL TEERFLEDPF --YPG----- ---------- ----ERMYKT
ZP_0696580      ARGYLNRPEL TKDRFVPHPF SDTIG----- ---------- ----DQLYKT
ABM21571        ARGYLNRPEL TLEKFIPNPF NSR------- ---------- ----SKLYKT
ZP_0163206      ARGYLNQPDL TAQKFIPNPF SNDTN----- ---------- ----SRLYKT
YP_324595       ARGYLNRPEL TQQKFIPHPF SNEPE----- ---------- ----ARLYKT
YP_0018659      ARGYLNRPEL TKEKFIPHPF EKAEG----- ---------- ----SRLYKT
ACZ55942        ARGYLNRPEL TKEKFISNPF KKTEG----- ---------- ----SRLYKT
Q9R919          ARGYLNRPEL TAEKFVKNPF SAG------- ---------- ----ERMYRT
YP_0018699      ARGYLNRPQL TIEKFIPNPF STDPH----- ---------- ----SRLYKT
NP_522203       ARGYLNRPEL TAERFVVNPF HGEGR----- ---------- ----ERMYRT
CAQ48254        ARGYLNRPEL TQEKFIPNPF SNYPD----- ---------- ----SRLYKT
AAZ03550        ARGYLNRPEL TQEKFIPNPF SNYPD----- ---------- ----SRLYKT
CAC01603        ARGYLNRPEL TQEKFIANPF STYPG----- ---------- ----SRLYKT
AAO23333        AKGYLNLPEL TQEKFIPNPF QGSRGAGE-- --------QG SRGRERLYKT
YP_0016588      ARGYLNRLEL TQEKFISNPF AE-------- ---------- ----GKLYKT
ABO45744        ARGYLNRPEL TAEKFIPHPF AQ-------- ---------- ----GKLYKT
BAA83994        ARGYLNQPEL TAEKFIPHPF AQ-------- ---------- ----GKLYKT
YP_324355       ARGYLNKEEL TKEKFIINPF PNSEF----- ---------- ----KRLYKT
AAF15891        AKGYLNRPDL TTDKFIPNPF SREVGEQGSK GAKILPNSQS LVPNPQLYKT
AAO62586        ARGYLNRPEL TSHKFISHSF GD-------- ---------- ----GKLYKT
AAF00960        ARGYLNRPEL TAEKFIPNPF AKVEAGIG-- --------GE IR--AKLYKT
YP_0016588      ARGYLNRPEL TAEKFIPNPF AKVEGEIG-- --------GE IR--AKLYKT
Clustal Co      .:**::   * * .:*: ..*                          :::*
```

```
                ....|....| ....|....
                       360
360305          GDLARYLPDG NIECFGRID
360310          GDLARYLPDG NIECFGRID
360312          GDLARYLPDG NIECFGRID
360802          GDLARYLPDG NIECFGRID
360804          GDLARYLPDG NIECFGRID
361101          GDLARYLPDG NIECFGRID
361102          GDLARYLPDG NIECFGRID
P27206          GDVVRWLPDG TIEYLGRED
YP_0016425      GDLVRYLADG HLDYLGRID
NP_832218       GDLVRYLPDG NIDYLGRMD
ZP_0323232      GDLVRYLPDG NIAYLGRMD
YP_322131       GDLARYLSDG NIEYLGRID
ZP_0711340      GDLARYLPDG NIEFLGRID
PPSD_BACSU      GDVARWLPDG NVEFLGRTD
ZP_0696580      GDLARYLPDG RIELIGRAD
ABM21571        GDLARYLPDG NIEFLGRLD
ZP_0163206      GDLGRYLPDG NIEFLGRID
YP_324595       GDLARYLSDG NIEYLGRID
YP_0018659      GDLARFLPDG NIEFLGRVD
ACZ55942        GDLARYLPDG NIEFLGRVD
Q9R919          GDLARWLPDG NIEYLGRID
YP_0018699      GDLARYLPDG NIEYLGRID
NP_522203       GDLGRWLPDG SLEYQGRAD
CAQ48254        GDLARYLPDG NIEYLGRID
AAZ03550        GDLARYLPDG NIEYLGRID
CAC01603        GDLARYLPNG NIEYLGRID
AAO23333        GDLARYLPDG NIEYLGRID
YP_0016588      GDLARYLPEG NIEYLGRID
ABO45744        GDLARYLPDG NIEYLGRID
BAA83994        GDLARYLPDG NIEYLGRID
YP_324355       GDLARYLPDG NIEYLGRTD
AAF15891        GDLARYLPDG TIEYIGRID
AAO62586        GDLARYLPDG NIEFLGRID
AAF00960        GDLARYLPDG NIEFLGRID
YP_0016588      GDLARYLPDG NIEFLGRID
Clustal Co      **: *:*.:*  :   ** *
```

Figure 8B. Sequence alignment of NRPS from clone from pool 36 with protein sequences with similar conserved core sequences from database**.**

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        10         20         30         40         50
371001          LDVNYPADRI EYMLQDSGSI LLLS--DASA PA--LP---V ESKLPHLLVD
371002          TGCELPGGSY RVHATRLRIH FLLS--DASA PA--LP---V ESKLPHLLVD
374605          LDVNYPADRI EYMLQDSGSI LLLS--DASA PA--LP---V ESKLPHLLVD
AAO72425        LDINAPAERQ AFMLQDCGAR QVLTLSRHDL PD-------- ----GIQRID
AAW55330        LDPGYPRERL AFMLLDTQVS ILLT-QKDLV AK--LP---T HTA-FVICLD
ACA09733        IDPDYPEDRV RYMLEDSNAK LLLVQK--GE LIN--V---- DYGL-PIVDL
ACF24472        IDPDYPEERI SFMIQDTQVK IILT-CESLQ TS--LP---N HQA-IVVCLD
ADK89159        IDPGYPEERI RFLLEDSGSK IVLT-KD-ST QIS--L---E GYE---VLAV
BAA02523        LDPALPGDRL RFMAEDSSVR MVLI-GN-SY TGQ-AH---Q LQV-PVLTLD
CAA06324        IDPGYPEERI RFLLEDSGSK IVLT-KD-GT QIS--L---E GYE---VLAV
```

```
CAA49817     LDPALPGDRL RFMAEDSSVR MVLI-GN-SY TGQ-AH---Q LQV-PVLTLD
CAQ48260     IDSDYPQERI SFMFQDTQVK ILLT-QESLL AS--LP---N HEA-IVVCLD
NP_388231    LDPALPGDRL RFMAEDSSVR MVLI-GN-SY TGQ-AH---Q LQV-PVLTLD
NP_389714    IDSNLPVERI AYMLSDSRAA LLLQ-SEKTE KR--L---LG IEC-EQIIIE
P39847       IDSNLPVERI AYMLSDSRAA LLLQ-SEKTE KR--LL---G IEC-EQIIIE
P94459       IDPDYPEERI SFLLEDSGTN ILLL-QSAGL HV------PE FTG-EIVYLN
Q04747       LDPALPGDRL RFMAEDSSVR MVLI-GN-SY TGQ-AH---Q LQV-PVLTLD
Q9R919       IDPTYPEERI RYILEDSDTK LLLV-QHHLR EK------VP FTG-KVLDME
YP_0014199   IDPAFPEDRL RFMAEDSSIR LVLT-VQ-DY QEQ-A---GA LQV-PVVMLD
YP_0015446   VDPSYPVERL AWMLSDLQPT VVIA-QHGVL DR--LP---S VAC-SVVVLE
YP_0018044   IDPNAPSERI DFLLEDTQIN LLLT-QRNID HQ--WP---- --N-TVTVID
YP_0028026   IDPSYPVERI EYMIEDAKID ILIT-SEEFI NK------VK FTK-SIVNIK
YP_0030418   LDPAYPGERL IHILTDAAPA ILLA--D-SA -GCDALG--- EKVLTRLTLL
YP_0030418   LDPTHPGERL TYMLTDAAPA ILLA--D-NA -GQTALS--- EEVMATLTVL
YP_0034678   LDPAYPAERL AYMLDDAAPV VLLTQTAWVD T--LVSPVTT SV-PIIVLDA
YP_0034678   LDPTYPAERL AYMLDDAAPV ALLTQAAWVD T------LDS PV-PTVVLDA
YP_0037122   LDPDYPTERL AYMLEDAAPV VLLTQTSQLD K------LSG TM-PVVILDT
YP_0037122   LDPAYPTERL AYMLKDAAPV VLLT-ETAQF DR--LSGTLP AMM-SVVMLD
YP_0037122   LDPNYPAERL TYILDDSAPV ALLTQEAHLN K------LSA TL-PTVLLDN
YP_077641    IDPGYPEERI RFLLEDSGAK IVLT-KD-SP QIS--L---E GYE---VLAA
YP_235693    LDINAPAERQ AFMLQDCGAR QVLTLSRHDL PD-------- ----GIQRID
YP_325325    LDAGYPQERL AFMLVDTQIP VLLT-QKELV KK--LP---N HEA-RVICLD
ZP_0173260   IDPNYPQERI EYMLEDSGIR ILVT-QESFR PL--YS---E FST-QLISLD
ZP_0221920   LDPDYPKNRL EFMIQDSHEG LIVTQKNIVS ENHFLK---Q LHTH-ELLIL
ZP_0359585   IDSNLPVERI AYMLSDSRAA LLLQ-SEKTE KR--LL---G IEC-EQIIIE
ZP_0503089   IDPTYPTERI TYMLEDAQVQ VLLT-QESLT QE--LP---V NHT-QLICLD
ZP_0687517   LDPVLPEDRL RFMAEDSSIQ MVLA-GK-SY TEQ-AH---Q LQV-PVITLD
ZP_0711340   LDPTYPKERL AFMLEDASVP VLLT-QTRLV ES--LP---- HQA-RVVCLD
ZP_0726607   LDINAPAERQ AFMLHDCGAR QVLTLARHDL PE-------- ----GIQRID
Clustal Co    .   *                ::
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        60         70         80         90        100
371001          NVATALTDYA NDAHNPIYHN PVVAM---QP THLSYVVYTS GSTGKPKGVL
371002          NVATALTDYA NDPHNPIYHN PVVAM---QP THLSYVVYTS GSTGKPKGVL
374605          NVATALTDYA NDAHNPIYHN PVVAM---QP THLSYVVYTS GSTGKPKGVL
AAO72425        LDLL-----E LQSDAP--NP VHSA----SA ESVAYIMYTS GSTGMPKGVL
AAW55330        ADWH-----T IAQNKK--EN LSTN---VTA ENLAYVMYTS GSTGTPKGVS
ACA09733        SSAEA----- -YASEPV-QA EVVQ----GP EGLAYVIYTS GTTGRPKGVM
ACF24472        NDWQ-----Q IKQASQ--EN LNNA---VSA DNLAYIIYTS GSTGTPKGVE
ADK89159        NAMDA----- -EKEDAA-NL EHVN----KP EDLAYYIYTS GSTGRPKGVM
BAA02523        IGFE------ -ESE-AADNL NLPS----AP SDLAYIMYTS GSTGKPKGVM
CAA06324        NAMDA----- -EKEDAA-NL EHVN----KP EDLAYYIYTS GSTGRPKGVM
CAA49817        IGFE------ -ESE-AADNL NLPS----AP SDLAYIMYTS GSTGKPKGVM
CAQ48260        KDWE-----Q INQASQ--EN LNSA---VSA ENLAYVIYTS GSTGTPKGVE
NP_388231       IGFE------ -ESE-AADNL NLPS----AP SDLAYIMYTS GSTGKPKGVM
NP_389714       DIQK------ -QGEAK--NV ESSA----GP HSLAYIIYTS GSTGKPKGVM
P39847          DIQK------ -QGEAK--NV ESSA----GP HSLAYIIYTS GSTGKPKGVM
P94459          QTNS------ GLAHRL--SN PNVD-V--LP QSLAYVIYTS GSTGMPKGVE
Q04747          IGFE------ -ESE-AADNL NLPS----AP SDLAYIMYTS GSTGKPKGVM
Q9R919          DPQT------ FSEDGS--NL ESIS----GP NQLAYVIYTS GSTGKPKGVM
YP_0014199      ESAD------ -ETVSGTD-L NLPA----GG NDLAYIMYTS GSTGKPKGVM
YP_0015446      TIAA-----H LAAYPT--TA PTVD---ISP ENLAYVMYTS GSTGRPKGIM
YP_0018044      LDEK-----A IAQESP--TL PVTD---TTS EHLAYVMYTS GSTGIPKGVC
YP_0028026      EKSL------ --LKKD--NL DIIN---K-S SDLAYVIYTS GSTGRPKGVM
YP_0030418      DPNS------ -LLDQPDSNP LVSSL---TS RHLAYVIYTS GSTGTPKGVM
YP_0030418      DPNIQ----- --PDQPDSNP QVPEL---TS RHLAYVIYTS GSTGRPKGVM
YP_0034678      QEP-----AV AAQPT--HNP EPQTL-GLTS RHLAYVIYTS GSTGLSKGVM
YP_0034678      QES-----AM AAQPT--HNP DAQAL-GLTS RHLAYVIYTS GSTGLPKGVM
YP_0037122      QNA-----LL ESQSI--HNP ETQMQ-GLTS RHLAYVIYTS GSTGQPKGVM
YP_0037122      EQNTPPATQL LATQSD-HNP TAQAS-GLTS RHLAYVIYTS GSTGQPKGVM
YP_0037122      DET-----LL ATQPI--DNP DIQAL-GLTS HHLAYVLYTS GSTGQPKGVM
YP_077641       NAVDA----- -EKEDAA-NL VHAN----KP GDLAYYIYTS GSTGKPKGVM
YP_235693       LDLL------ ELQSDA--PN PVHSA---SA ESVAYIMYTS GSTGMPKGVL
YP_325325       TDWE-----I INQHTP--EN QNIS---ITP DNLAYVMYTS GSTGQPKGVS
ZP_0173260      TDQQ-----K WERENQ--TN PIHQ---THS HHLAYINYTS GSTGQPKGVM
ZP_0221920      DSDEVKTD-L SN--QTTENL ALIS----GP RNLAYVIYTS GTTGKPKGVM
ZP_0359585      DIQK------ -QGEAK--NV ESSA----GP HSLAYIIYTS GSTGKPKGVM
ZP_0503089      SQWQ-----I IAQQSP--DN PLTD---VTS DNLAYINYTS GSTGKPKGVE
ZP_0687517      SGFE------ -ESG-AADNL NLPS----AP SDLAYIMYTS GSTGKPKGVM
ZP_0711340      ADWE-----V IERQSE--EN PSPQ---VIH DNLAYVMYTS GSTGIPKGVS
ZP_0726607      LDLL-----Q LPGDTP--NP VPSA----SA ESVAYIMYTS GSTGMPKGVL
Clustal Co                                        ::*  *** *:** .**:
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        110        120        130        140        150
371001          VNHLGVNRLV KN-QNYIELD ENSVVLQDAS ISFDAATFEM YAAWLNGGTL
371002          VNHLGVNRLV KN-QNYIELD ENSVVLQDAS ISFDAATFEM YAAWLNGGTL
374605          VNHLGVNRLV KN-QNYIELD ENSVVLQDAS ISFDAATFEM YAAWLNGGTL
AAO72425        VPHRAVSRLV LN-NGYADFN AGDRVAFASN PAFDASTLDV WAPLLNGGCV
AAW55330        VIHRGVVRLV KE-TNYAHLT AEEIILQLAP ISFDASTFEI WGCLLNGGQL
ACA09733        VEHRNVVRLV KE-TNYVELN ESTRILQTGA VAFDASTFEI WGALLNGGQL
ACF24472        ITHRSVNRLL FG-VNYVHLD ATQRLLQMAP IAFDASTFEI WGALLHGGRC
ADK89159        VEHRNIVRLV KN-AGCIPLK SGVKMAQTGA VSFDASTFEV FGALLNGGTL
BAA02523        IEHKSILRLV KN-AGYVPVT EEDRMAQTGA VSFDAGTFEV FGALLNGAAL
CAA06324        VEHRNIVRLV KN-AGCIPLK SGVKMAQTGA VSFDASTFEV FGALLNGGTL
CAA49817        IEHKSILRLV KN-AGYVPVT EEDAMAQTGA VSFDAGTFEV FGALLNGAAL
CAQ48260        VIHRSVNRLL FG-INYVDLD ANETFLQMAP IAFDASTFEI WGALLHGARC
NP_388231       IEHKSILRLV KN-AGYVPVT EEDRMAQTGA VSFDAGTFEV FGALLNGAAL
NP_389714       IEQRSVIRLV KN-SNYITFT PEDRLLMTSS IGFDVGSFEI FGPLLNGAAL
P39847          IEQRSVIRLV KN-SNYITFT PEDRLLMTSS IGFDVGSFEI FGPLLNGAAL
P94459          IEHRSAVNFL NSLQSRYQLK HSDMIMHKTS YSFDASIWEL FWWPYAGASV
Q04747          IEHKSILRLV KN-AGYVPVT EEDRMAQTGA VSFDAGTFEV FGALLNGAAL
Q9R919          VEHRSVINRL VWMQENYPLD ERDAILQKTA ITFDVSVWEL FWWSIVGSKV
YP_0014199      IEHRNIIRLV KH-SNYVPVH EEDRMAQTGA VSFDAGTFEV FGALLNGASL
YP_0015446      INQRNIVRLV RN-TTYAAFG PDQVGLLLAT VAFDASTFEL WGCLLNGGRL
YP_0018044      IPHRGVIRLV KN-SNYVDIR EDDVFLQAAP YTFDASTFEI WGALLNGGRL
YP_0028026      VEHKSLINLC NYHNKKFNIK EEDKSTSYAE FSFDASVWEV FPYLIIGATI
YP_0030418      IEHRGLVNLI QEKIVQFDIH SGSRMLQFAS FGFDASVWEV MMALCGGATL
YP_0030418      VEHHGVVNLT LTQNAQFNVD AASRMLQFAS FGFDASVWEI MMALSSGAIL
YP_0034678      VEHRNVLRLI IN-NGFADIG SDDCIAHCAN IAFDASTWEI WSALLNGGRL
YP_0034678      VEHRSINRLV IN-NPYADIS SDDCVAHCAN IAFDASTWEI WSALLNGGRL
YP_0037122      VEHRNVLRLI IN-NGFADIG PDDCIAHCAN MAFDASTWEI WSALLNGGCL
YP_0037122      VEHHNVNRLI IN-NGYADIT AEDCVAHCAN IAFDASTWEI WSALLNGGRL
YP_0037122      TEHRNVLRLI IN-SGFADIG PDDCIAHCAN MAFDASTWEI WSALLNGARL
YP_077641       VEHRNIVRLV KN-AGYIPLK SDVKMAQTGA VSFDASTFEV FGALLNGGTL
YP_235693       VPHRAVSRLV LN-NGYADFN AGDRVAFASN PAFDASTLDV WAPLLNGGCV
YP_325325       VVHRGVVRLV KQ-TNYANFT NTEIFLQFAP ISFDASTFEI WGCLLNGGKL
ZP_0173260      IPHRGVIRLL IN-SDYVELD EAKTFLHLSP IAFDASTFEI WGALLYGGKC
ZP_0221920      VEHKSINSLV VN-NAYLHMS ERDALLSLSS LVFDASTFEI WMPLLNGSKL
ZP_0359585      IEQRSVIRLV KN-SNYITFT PEDRLLMTSS IGFDVGSFEI FGPLLNGAAL
ZP_0503089      VLHRGVIRLL FG-IDYVHLD GKQRLLQMAP ISFDAATFEI WGALLHGARC
ZP_0687517      IEHKSILRLV KN-AGYVPIH EEDRMAQTGA VSFDAGTFEV FGALLNGAAL
ZP_0711340      VIHQGVVRLV KD-TNYVNLS AEEVFLQLAP ISFDASTLEI WGSLLNGGRL
ZP_0726607      VPHRAVSRLV LN-NGYADFN AQDRVAFASN PAFDASTLDV WAPLLNGGCV
Clustal Co        :               .               **..  ::       *.
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
371001          VLY--PQQYM -DLTTLTDVI EQHRVNVLWI TCALFDKWAA TLQAG-----
371002          VLY--PQQYM -DLTTLTDVI EQHRVNVLWI TCALFDKWAA TLQAG-----
374605          VLY--PQQYM -DLTTLTDVI EQHRVNVLWI TCALFDKWAA TLQAG-----
AAO72425        VVV--EQSVL LSLDEFRALL LSQSVSVLWM TAGLFHQYAS GL-ME-----
AAW55330        VIC--PPHTP -SLEELGQII QQYQVTTLWL TAGLFHLIVD EK-ID-----
ACA09733        YFV--ENDDI LIADRLKAAI AKYGITTMWL TSPLFNQLSL QD-EY-----
ACF24472        VIF--AEDIP -TATSLKNAI DKNGITVLWL TAALFNRIID DN-SQ-----
ADK89159        YPV--PKETL LDGKRFNMFL KETGITTMWL TSPLFNQLAQ QD-PG-----
BAA02523        YPVK--KETL LDAKQFAAFL REQSITTMWL TSPLFNQLAA KD-AG-----
CAA06324        YPV--PKETL LDGKRFNMFL KETGITTMWL TSPLFNQIAQ QD-PG-----
CAA49817        YPVK--KRHV LDAKQFAAFL REQSITTMWL TSPLFNQLAA KD-AG-----
CAQ48260        VLF--PGNIP -TAKSLRDAI DKHGITILWL TTALFNAIID DD-SQ-----
NP_388231       YPVK--KETL LDAKQFAAFL REQSITTMWL TSPLFNQLAA KD-AG-----
NP_389714       HLS--DQQTF LDSHQLKRYI EHQGITTIWL TSSLFNHLTE QN-EQ-----
P39847          HLS--DQQTF LDSHQLKRYI EHQGITTIWL TSSLFNHLTE QN-EQ-----
P94459          YLL--PQGGE KEPEVIAKAI EEQKITAMHF VPSMLHAFLE HIKYR---SV
Q04747          YPVK--KETL LDAKQFAAFL REQSITTMWL TSPLFNQLAA KD-AG-----
Q9R919          VLL--PNGGE KNPELILDTI EQKGVSTLHF VPAMLHAFLE SMEQTPSGKL
YP_0014199      HPVK--KETL LDAGQFAQFL KEQRITTMWL TSPLFNQLAQ KD-AG-----
YP_0015446      VIA--PPQQL -SLAELGHLV EREQITTLWL TAGLFHQMVD HA-LD-----
YP_0018044      VIL--PSPTP -SLEELGEAI ENYGVTTLWL TAGLFHLMVE EK-LE-----
YP_0028026      YII--NENIK LDIIKLNKYY EKNNITISFL PTPICQQFME VD-NT-----
YP_0030418      DIP--VDIVR QEPHHLWHYL EEHTVTHACL TPTMLREGAG LP--------
YP_0030418      VIP--TETVR QDPGRLWHYL EEQTVTHACL TPAMFHD--G TG-LP-----
YP_0034678      YVV--PPSVL LDPVRFCDSL IKGQVTALWL TAGLFNEYLS DL-NP-----
YP_0034678      HVV--SSSAL LDPVRFRDSL VKGQVTALWL TAGLFNEYLG DL-EP-----
YP_0037122      HVV--SQPVL LDPVRFCDSL IRGKVTGLWL TAGLFNEYLD TL-KP-----
YP_0037122      HIV--SQSVL LDPAQFRDSL IKGKVTALWL TAGLFNEYLD TL-KP-----
YP_0037122      HVV--SPSVL LDPVRFCDSL MQGQVTALWL TAGLFHEYLD TL-KP-----
YP_077641       YPV--PKETL LDGKRFRVFL EETGITTMWL TSPLFNQLAQ QD-PG-----
YP_235693       VVV--EQSVL LSLDEFRALL LSQSVSVLWM TAGLFHQYAS GL-ME-----
YP_325325       VLY--PSNTP -SIDELGQVI QKYQITTIWL TAGLFHLMVD EN-IH-----
ZP_0173260      IIF--PEKIP -TALTLKEAI NQYQVTTLWL TAALFNLVID EL-PE-----
ZP_0221920      VLAKDTKELT SHLEQFKKVI IQHQITTLWL TKTLFDSLYI QD-KY-----
ZP_0359585      HLS--DQQTF LDSHQLKRYI EHQGITTIWL TSSLFNHLTE QN-EQ-----
ZP_0503089      VLF--PETVP -TAQTLKQVI QTHNITTLWL TSALFNGIVA ED-AE-----
ZP_0687517      YPAK--KETL LDAKQFAAFL REQRITTMWL TSPLFNQLAA KD-AG-----
ZP_0711340      VIM--PPHTP -SLQELGEAI WGYQITTLWL TAGLFHIMVD EH-LE-----
ZP_0726607      VVV--AQSVL LSLDEFRALL LSQSVSVLWM TAGLFHQYAD GL-ME-----
Clustal Co                     :         :.   :    :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
371001          --AVPLLKTV ITGGDVISPR SVKQV-Y--Q QCD-NVTVVA AYGPTENTVF
371002          --AVPLLKTV ITGGDVISPR SVKQV-Y--Q QCD-NVTVVA AYGPTENTVF
374605          --AVPLLKTV ITGGDVISPR SVKQV-Y--Q QCD-NVTVVA AYGPTENTVF
AAO72425        --ALARLRYL IVGGDVLDPA VIARV-L--A EGA-PQHLLN GYGPTEATTF
AAW55330        --ALKSLRQL LAGGDVLSVL HVQKF-L--Q TVE-NCRLIN GYGPTENTTF
ACA09733        --LFRGLKAL LVGGDVLSIS HINRV-I--E ANP-DLVPIN GYGPTENTTF
ACF24472        --ALSGIKQL LIGGEALSVA HVHKA-L--A ALP-LTQITN GYGPTESTTF
ADK89159        --MFATLNDL IIGGDALVPG IVNRV-K--R ESP-ELSLWN GYGPTENTTF
BAA02523        --MFGTLRHL IIGGDALVPH IVSKV-K--Q ASP-SLSLWN GYGPTENTTF
CAA06324        --MFATLNDL IIGGDALVPG IVNRV-K--R ESP-ELSLWN GYGPTENTTF
CAA49817        --MFGTLRHL IIGGDALVPH IVSKV-K--Q ASP-SLSLWN GYGPTENTTF
CAQ48260        --ALSGIKQL LIGGEALSIA HVQKA-L--F TLP-FTQIIN GYGPTESTTF
NP_388231       --MFGTLRHL IIGGDALVPH IVSKV-K--Q ASP-SLSLWN GYGPTENTTF
NP_389714       --TFSQLKHL IIGGEALSPS HVNRI-R--N VCP-EVSIWN GYGPTENTTF
P39847          --TFSQLKHL IIGGEALSPS HVNRI-R--N VCP-EVSIWN GYGPTENTTF
P94459          PIKTNRLKRV FSGGEQLGTH LVSRF-Y--E LLP-NVSITN SYGPTEATVE
Q04747          --MFGTLRHL IIGGDALVPH IVSKV-K--Q ASP-SLSLWN GYGPTENTTF
Q9R919          KRKLASLRYV FASGEALTPK HVDGFQRIIT PVS-HAQIIN LYGPTEATID
YP_0014199      --MFNTLRHL IIGGDALVPH IVSKV-R--K ASP-ELSLWN GYGPTENTTF
YP_0015446      --RLGSLRQL LAGGDRLSPV HVHKV-L--E RWP-QCRLIN GYGPTENTTF
YP_0018044      --SFKNVRYL LAGGDVLFPD HVKTV-L--R TYP-HCCVIN GYGPTENTTF
YP_0028026      -----SLRVI LTGGDKLNNY KEKQI-S--- -------IIN NYGPTEATVL
YP_0030418      --VMTIKPTL ILGGEAPSAA LLGAL-S--G RVN----LFN AYGPTEITVC
YP_0030418      --AITIRPTI IFAGEAPGLT LFQAL-C--N QAD----LFN AYGPTETTVC
YP_0034678      --LLGRLRYL LIGGDVLDPR KIQRAQL--A ESQ-PAHLIN GYGPTETTTF
YP_0034678      --LFGQLRYL LVGGDVLDPR KIRRTQL--A ECQ-PAHLIN GYGPTETTTF
YP_0037122      --VFRQLRYL LIGGDVLDPN KIQQVQL--A ESK-PTYLIN GYGPTETTTF
YP_0037122      --LLGQLRYL LIGGDVLDPN KIQQVQS--A VLK-PTYLLN GYGPTETTTF
YP_0037122      --LYGQLRYL LVGGDILDPG KIQQVKL--A ESQ-PAHLIN GYGPTETTTF
YP_077641       --MFATLNDL IIGGDALVPG IVNRV-K--R ESP-ELSLWN GYGPTENTTF
YP_235693       --ALARLRYL IVGGDVLDPA VIARV-L--A EGA-PQHLLN GYGPTEATTF
YP_325325       --ALKPLRQL LAGGDVLSVS HVQKF-L--K TVE-NCQLIN GYGPTENTTF
ZP_0173260      --AFIRVKEL LTGGEALSVH HVKKA-L--Q ALP-STQLIN GYGPTENTTF
ZP_0221920      --LFSGLKTL LVGGEALNID LINQL-I--S QNQRPCRVLN GYGPTEGTTF
ZP_0359585      --TFSQLKHL IIGGEALSPS HVNRI-R--N VCP-EVSIWN GYGPTENTTF
ZP_0503089      --ALSGVPQL LTGGEALSVN PVKKA-L--A ALP-STQIIN GYGPTENTTF
ZP_0687517      --MFGTLRHV IIGGDALVPH IVSKV-K--Q ASP-SLSLWN GYGPTENTTF
ZP_0711340      --DLKQVRQL LAGGDILSVP HVQKV-I--Q ELK-GCQLIN GYGPTENTTF
ZP_0726607      --AFARLRYL IVGGDVLDPA VIGRV-L--K EGA-PRHLLN GYGPTEATTF
Clustal Co              : : .*:                              ***** *
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         260        270        280        290        300
371001           TTTYPI-PRD FNAEQPL-PL GRVINNTQLY ILDADGQLLS FGVAGEIHVG
371002           TTTYPI-PRD FNAEQPL-PL GRVINNTQLY ILDADGQLLP FGVAGEIHVG
374605           TTTYPI-PRD FNAEQPL-PL GRVINNTQLY ILDADGQLLP FGVAGEIHVG
AAO72425         STTHEI--TS VG-SGGI-PI GRPIGNSQVY VLDTLRQPVA VGVAGELYIG
AAW55330         TCCHLI-TAP VQPGVSI-PI GRPIANTQVY ILDNNFQTVA IGEIGELHIA
ACA09733         STTYKI---P GRVEGGV-PI GRPISNSTAY VVNESLQLQP IGAWGELIVG
ACF24472         TCCYPIPKQL EATIKSI-PI GRPISNTQVY ILDNYLQPVP IGVVGELHIS
ADK89159         STCFLI---D QAYERTI-PI GKPIGNSTAY IVDEYGSLQP IGVPGELCVG
BAA02523         STSFLI---D REYGGSI-PI GKPIGNSTAY IMDEQQCLQP IGAPGELCVG
CAA06324         STCFLI---D QAYERTI-PI GKPIGNSTAY IVDEHGSLQP IGVPGELCVG
CAA49817         STSFLI---D REYGGSI-PI GKPIGNSTAY IMDEQQCLQP IGAPGELCVG
CAQ48260         TCCYPIPKQL ETKIKSI-PL GKPIANTQVY ILDKYLQPVP VGVSGELHIG
NP_388231        STSFLI---D REYGGSI-PI GKPIGNSTAY IMDEQQCLQP IGAPGELCVG
NP_389714        STCLHI---Q KTYELSI-PI GRPVGNSTAF ILNQWGVLQP VGAVGELCVG
P39847           STCLHI---Q KTYELSI-PI GRPVGNSTAF ILNQWGVLQP VGAVGELCVG
P94459           AAFFDC--PP HEKLERI-PI GKPVHHVRLY LLNQNQRMLP VGCIGELYIA
Q04747           STSFLI---D REYGGSI-PI GKPIGNSTAY IMDEQQCLQP IGAPGELCVG
Q9R919           VSYFEC--EA DKRYNSV-PI GKPISNIQLY ILQA-GYMQP VGVAGELCIA
YP_0014199       STSFLI---D QDYDGSV-PI GKPIGNSTAY IMDENRNLQP IGAPGELCVG
YP_0015446       SCCQQL-SAT TDLAQGV-PI GQPIANSTAY ILDRLLQLVP IGVVGELYLG
YP_0018044       TCCAVL-TDV EQIGHSV-PI GRPISQTQVY ILDPYLHPVP FGVPGELYIG
YP_0028026       TTSYNV---- KSKVNNI-PI GKPMYNQRVY ILNNK-KVAP IGVSGELCIS
YP_0030418       ATVWYC---P PDYTDELISI GRPTANTRIY LLDTYGQPVP LGAVGELYIG
YP_0030418       ATTWDC---P PDYMGRLTPI GRPTANKRLY LLDKHGQPVP LGAVGELYIG
YP_0034678       ATTYRI-ASP VDVAHSI-PI GRPIANTRIY ILDCHNQPVP LGVAGEIYIA
YP_0034678       AATYRI-SSP VDVNRPI-PI GCPIANTQIY LLDPYGQPVP LGVAGEIHIA
YP_0037122       AATYTI-PSS VDVARSI-PI GRPIANTQIY ILDSQGRPVP VGVAGEIYIA
YP_0037122       AATYTI-PLS IDVTRSI-PI GRPISNTQIY ILDSYGQPVP LGVTGEIYIA
YP_0037122       ATTYDI-ASP VDVTRSI-PI GRPIGNTRIY ILDSRGQPVP LGIVGEIHIA
YP_077641        STCFFI---D QAYERTI-PI GKPIGNSTAY IVDEHGALQP IGVPGELCVG
YP_235693        STTHEI---T SVGSGGI-PI GRPIGNSQVY VLDTLRQPVA VGVAGELYIG
YP_325325        TCCYHI-KDP VRPDSSI-PI GRPIAHTQVY ILDENLQPVA MGATGELYIG
ZP_0173260       TCCYSIPSFL ESKVSSI-LI GRPINNTQIY ILDPNLQPVP VGVPGELHIG
ZP_0221920       TTIYEC-QKN IE-GNSV-PI GRPISQRKVF ILDANLNPVP VGVTGELYIG
ZP_0359585       STCLHI---Q KTYELSI-PI GRPVGNSTAF ILNQWGVLQP VGAVGELCVG
ZP_0503089       TCCYSLPKQL PGTELSI-SI GRPISNTQVY LLDAYWQPVP IGVIGELYIG
ZP_0687517       STSFLI---D REYSGSI-PI GKPIGNSTAY IMDEQQRLQP IGAPGELCVG
ZP_0711340       TCCYRI-TEV NLIENSI-PI GRSISNTQVY LLDTHLQLVP IGVPGELYIG
ZP_0726607       STTHEI---T SVGNGGI-PV GRPIGNSQVY VLDTLRQPVA VGVVGELYIG
Clustal Co                    :  : *    :   : :::      . .*  **: :.
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         310        320        330        340        350
371001           GAGVARGYLN REDLTASQFI DNPLA----- VGSNGEKLYK TGDLGRIRED
371002           GAGVARGYLN REDLTASQFI DNPLA----- VGSNGEKLYK TGDLGRIRED
374605           GAGVARGYLN REDLTASQFI DNPLA----- VGSNGEKLYK TGDLGRIRED
AAO72425         GQGVAKGYLN RPELNATQFV ANPFS----- -DDAGALLYR TGDLGRWNAD
AAW55330         GDGLARGYLN RPELTAEKFI SHSFD----- -SNLATRLYK TGDLARYLPD
ACA09733         GEGVARGYLN RPDLTAEKFV PSPVK----- ---EGERCYR TGDLVRWLPD
ACF24472         GVGVARGYLN RLELTQEKFI ANPFS----- -TDSQSRLYK TGDLARYLPD
ADK89159         GDGVARGYLN QPELTDEKFV GDPFA----- ---EGKRMYR TGDLAKWLPD
BAA02523         GIGVARGYVN LPELTEKQFL EDPFR----- ---PGERIYR TGDLARWLPD
CAA06324         GDGVARGYLN QPELTDEKFV GDPFA----- ---EGKRMYR TGDLAKWLPD
CAA49817         GIGVARGYVN LPELTEKQFL EDPFR----- ---PGERIYR TGDLARWLPD
CAQ48260         GAGLARGYLN RLELTAEKFI PNPFEPLSKV SNQ-QSKLYK TGDLARYLPD
NP_388231        GIGVARGYVN LPELTEKQFL EDPFR----- ---PGERIYR TGDLARWLPD
NP_389714        GDGVARGYLG RPDLTKEKFV PHPFA----- ---PGDRLYR TGDLARWLSD
P39847           GDGVARGYLG RPDLTKEKFV PHPFA----- ---PGDRLYR TGDLARWLSD
P94459           GAGVARGYLN RPALTEERFL EDPFY----- ---PGERMYK TGDVARWLPD
Q04747           GIGVARGYVN LPELTEKQFL EDPFR----- ---PGERIYR TGDLARWLPD
Q9R919           GDGLARGYLN RPELTAEKFV KNPFS----- ---AGERMYR TGDLARWLPD
YP_0014199       GSGVARGYVN LPELTEKQFV RDPFR----- ---PDEMIYR TGDLAKWLPD
YP_0015446       GAGLARGYLA RPDQTAAAFI PNPMS----- -QTAGERLYR SGDLARYRDD
YP_0018044       GGGLARGYLN RPELTAERFI PIPPTPITKG GGKQGERLYK TGDLGRYDRK
YP_0028026       GDGLARGYLN NPELTSEKFV DNPFE----- ---PGERMYR TGDLARWLPD
YP_0030418       GIGVARGYLN HPELTVEHFL TDPFS----- -DDPNARIYR TGDLARYLPD
YP_0030418       GAGVARGYLN RPELTAERFL TDPFS----- -NKTGAQMYR TGDLARYLPD
YP_0034678       GAGVARGYLN RPELTAERFV PDTFS----- -ADPDERMYK TGDLGRWLFD
YP_0034678       GAGVARGYLN RPELTAERFL TDPFS----- -SDPDARMYK TGDLGRWLPD
YP_0037122       GNGVARGYLN RPELTAERFL ADPFS----- -QDTDAHMYK TGDLGRWLAD
YP_0037122       GFGVARGYLN HAELTAERFL TDPFAS-PF- VSNLNARMYK TGDLGRWLPD
YP_0037122       GAGVARGYLN RPELTTERFL LDPFS----- -QGTHARMYK TGDLGRWLPE
YP_077641        GDGVARGYLN QPELTDEKFV GDPFA----- ---EGKRMYR TGDLAKWLPD
YP_235693        GQGVAKGYLN RPELNATQFV ANPFS----- -DDAGALLYR TGDLGRWNAD
YP_325325        GDGLARGYLH RPELTKERFI ELNNS----- -NFQSLTLYK TGDLARYLPD
ZP_0173260       GDGLARGYLN RPDLTAEKFI PNPFG----- ---TG-KLYK TGDLCRYRRD
ZP_0221920       GAGVARGYLN RPELTKEHFI PNPFA--KEL DLPSSDRIYK TGDLASWLPD
ZP_0359585       GDGVARGYLG RPDLTKEKFV PHPFA----- ---PGDRLYR TGDLARWLSD
ZP_0503089       GDGLARGYLN RPELTGEKFI ANPFS----- -NQPNARLYK TGDLARYRAD
ZP_0687517       GIGVARGYVN LPELTEKQFL EDPFR----- ---PGERIYR TGDLARWLPD
ZP_0711340       GDGLARGYLN RPELTAERFI LNPFS----- -DKPSDRLYK TGDLARYLPD
ZP_0726607       GQGVAKGYLN RPELNATQFV ANPFS----- -DDAGALLYR TGDLGRWNGD
Clustal Co       * *:*:**:      .   *:                  *: :**:     .
```

```
                ....|....|
                        360
371001          GIVEFLGRID
371002          GIVEFLGRID
374605          GIVEFLGRID
AAO72425        GIVEYLGRND
AAW55330        GNIEFLGRID
ACA09733        GNLEFKGRID
ACF24472        GNIEYLGRID
ADK89159        GNIEFLGRID
BAA02523        GNIEFLGRID
CAA06324        GNIEFLGRID
CAA49817        GNIEFLGRID
CAQ48260        GKIEYLGRID
NP_388231       GNIEFLGRID
NP_389714       GTIEYVGRID
P39847          GTIEYVGRID
P94459          GNVEFLGRTD
Q04747          GNIEFLGRID
Q9R919          GNIEYLGRID
YP_0014199      GTIEFLGRID
YP_0015446      GTIEFIGRRD
YP_0018044      GNIEFLGRKD
YP_0028026      GNIEFLGRID
YP_0030418      GNLMFVGRND
YP_0030418      GNLVFVGRND
YP_0034678      GNIDYLGRND
YP_0034678      GNLEYLGRND
YP_0037122      GNIEYLGRND
YP_0037122      GNIEFRGRND
YP_0037122      GNIEYLGRYD
YP_077641       GNIEFLGRID
YP_235693       GIVEYLGRND
YP_325325       GNIEFLGRID
ZP_0173260      GNIEYIGRID
ZP_0221920      GNLEYLGRMD
ZP_0359585      GTIEYVGRID
ZP_0503089      GTIEFVGRVD
ZP_0687517      GNIEFLGRID
ZP_0711340      GNIEFLGRID
ZP_0726607      GVIEYLGRND
Clustal Co      * : : ** *
```

Figure 9B. Sequence alignment of NRPS from clone 37a (371001, 371002 and 374605) with protein sequences with similar conserved core sequences from database**.**

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40         50
374601          LEPTLPAERI AYILKDANPR FLLTTSQYSR TFPIP--NKK LLFIDGIDSF
Q9R919          IDPTYPEERI RYILEDSDTK LLLVQHH--L REKVP----- -FTGKVLDME
P94459          IDPDYPEERI SFLLEDSGTN ILLLQSAGLH VPEFT----- -GEIVYLNQT
YP_0034678      LDPAYPAERL AYMLDDAAPV VLLTQTA--W VDTLVSPVTT SVPIIVLDAQ
YP_0034678      LDPTYPAERL AYMLDDAAPV ALLTQAA--W VDTLDSP--- -VPTVVLDAQ
YP_0037122      LDPNYPAERL TYILDDSAPV ALLTQEA--H LNKLSAT--- -LPTVLLDND
YP_0037122      LDPAYPTERL AYMLKDAAPV VLLTETA--Q FDRLSGTLPA MMSVVMLDEQ
YP_0037122      LDPDYPTERL AYMLEDAAPV VLLTQTS--Q LDKLSGT--- -MPVVILDTQ
ZP_0726514      LDINAPVERQ AFMIEDSQAH VLLTHIHVSL TTT------- -AQRVDLDVL
YP_235691       LDVNAPVERQ SFMIEDSRAR VLLTHSQ--V SLT------- -TGAQRVDLD
AAF99707        LDVNAPVERQ TFMIEDSQAH VLLTLSR--M SLT------- -ASTQRIDLD
YP_0025514      VDPVYPKDRI DFVARDARPA VVITMSR--H AELFVGLHPS -VPVISIDAD
ZP_0173260      IDPNYPQERI EYMLEDSGIR ILVTQES--F RPLYS--EFS -TQLISLDTD
ZP_0503089      IDPTYPTERI TYMLEDAQVQ VLLTQES--L TQELP--VNH -TQLICLDSQ
ACF24472        IDPDYPEERI SFMIQDTQVK IILTCES--L QTSLP--NHQ -AIVVCLDND
CAQ48260        IDSDYPQERI SFMFQDTQVK ILLTQES--L LASLP--NHE -AIVVCLDKD
ZP_0051751      LDPNIPPERL TILLEDTQIN LLLTQND--I NLPWP----- -NTLTVIDLQ
YP_0018044      IDPNAPSERI DFLLEDTQIN LLLTQRN--I DHQWP----- -NTVTVIDLD
ZP_0711340      LDPTYPKERL AFMLEDASVP VLLTQTR--L VESLP---HQ -ARVVCLDAD
YP_325325       LDAGYPQERL AFMLVDTQIP VLLTQKE--L VKKLP--NHE -ARVICLDTD
AAW55330        LDPGYPRERL AFMLLDTQVS ILLTQKD--L VAKLP--THT -AFVICLDAD
YP_0015446      VDPSYPVERL AWMLSDLQPT VVIAQHG--V LDRLPSVA-- -CSVVVLETI
ZP_0359585      IDSNLPVERI AYMLSDSRAA LLLQSEK--T EKRL------ -LGIECEQII
NP_389714       IDSNLPVERI AYMLSDSRAA LLL-QSE--K TEKRL----- -LGIECEQII
P39847          IDSNLPVERI AYMLSDSRAA LLLQSEK--T EKRL------ -LGIECEQII
YP_0028026      IDPNLPKKRI DFILNDCKVN IVLGKGL--K EID------- -SNIRLLDIE
ACA09733        IDPDYPEDRV RYMLEDSNAK LLLVQKGELI NVD------- -YGLPIVDLS
YP_077641       IDPGYPEERI RFLLEDSGAK IVLTKDS--P QIS------- -LEGYEVLAA
CAA06324        IDPGYPEERI RFLLEDSGSK IVLTKDG--T QISLEG---- -YEVLAVNAM
ADK89159        IDPGYPEERI RFLLEDSGSK IVLTKDS--T QISLEG---- -YEVLAVNA-
ZP_0687517      LDPVLPEDRL RFMAEDSSIQ MVLAGKS--Y TEQAHQLQ-- -VPVITLDSG
CAA49817        LDPALPGDRL RFMAEDSSVR MVLIGNS--Y TGQAHQLQ-- -VPVLTLDIG
NP_388231       LDPALPGDRL RFMAEDSSVR MVLIGNS--Y TGQAHQLQ-- -VPVLTLDIG
BAA02523        LDPALPGDRL RFMAEDSSVR MVLIGNS--Y TGQAHQLQ-- -VPVLTLDIG
Q04747          LDPALPGDRL RFMAEDSSVR MVLIGNS--Y TGQAHQLQ-- -VPVLTLDIG
Clustal Co      ::   * .*    :  *      ::
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                          60         70         80         90        100
374601           KE-----TFP AWTKGISN-P DVAVKPHHLA YINYTSGSTG MPKGVMVPHR
Q9R919           DP-----QTF SEDG--SN-L ESISGPNQLA YVIYTSGSTG KPKGVMVEHR
P94459           NS-------- GLAHRLSN-P NVDVLPQSLA YVIYTSGSTG MPKGVEIEHR
YP_0034678       EP-----AVA AQPTHNPEPQ TLGLTSRHLA YVIYTSGSTG LSKGVMVEHR
YP_0034678       ES-----AMA AQPTHNPDAQ ALGLTSRHLA YVIYTSGSTG LPKGVMVEHR
YP_0037122       ET-----LLA TQPIDNPDIQ ALGLTSHHLA YVLYTSGSTG QPKGVMTEHR
YP_0037122       NTPPATQLLA TQSDHNPTAQ ASGLTSRHLA YVIYTSGSTG QPKGVMVEHH
YP_0037122       NA-----LLE SQSIHNPETQ MQGLTSRHLA YVIYTSGSTG QPKGVMVEHR
ZP_0726514       DL-----DGL KDTD-----L ALPQSSESVA YIMYTSGSTG IPKGVLVPHR
YP_235691        GL------TL ERLKGTDL-A LPPQSSESVA YIMYTSGSTG TPKGVLVPHR
AAF99707         GL-----TLD GLKD--TD-L TLPQSSESVA YIMYTSGSTG VPKGVLVPHR
YP_0025514       HN-----EWS TMSG--AP-P EMGGNDSRLA YICYTSGSTG TPKGVMIDHA
ZP_0173260       QQ-----KWE RENQ--TN-P IHQTHSHHLA YINYTSGSTG QPKGVMIPHR
ZP_0503089       WQ-----IIA QQSP--DN-P LTDVTSDNLA YINYTSGSTG KPKGVEVLHR
ACF24472         WQ-----QIK QASQ--EN-L NNAVSADNLA YIIYTSGSTG TPKGVEITHR
CAQ48260         WE-----QIN QASQ--EN-L NSAVSAENLA YVIYTSGSTG TPKGVEVIHR
ZP_0051751       QQ-----EIY QESQ--NT-L PTDTTAEHLA YVMYTSGSTG IPKGICIPHR
YP_0018044       EK-----AIA QESP--TL-P VTDTTSEHLA YVMYTSGSTG IPKGVCIPHR
ZP_0711340       WE-----VIE RQSE--EN-P SPQVIHDNLA YVMYTSGSTG IPKGVSVIHQ
YP_325325        WE-----IIN QHTP--EN-Q NISITPDNLA YVMYTSGSTG QPKGVSVVHR
AAW55330         WH-----TIA QNKK--EN-L STNVTAENLA YVMYTSGSTG TPKGVSVIHR
YP_0015446       AA-----HLA AYPT--TA-P TVDISPENLA YVMYTSGSTG RPKGIMINQR
ZP_0359585       IE-----DIQ KQGE-AKN-V ESSAGPHSLA YIIYTSGSTG KPKGVMIEQR
NP_389714        IE-----DIQ KQGE-AKN-V ESSAGPHSLA YIIYTSGSTG KPKGVMIEQR
P39847           IE-----DIQ KQGE-AKN-V ESSAGPHSLA YIIYTSGSTG KPKGVMIEQR
YP_0028026       ER-----IEY EECT--AN-P DIFKKTNLLA YIMYTSGSTG VPKGVLVEQK
ACA09733         SA-----EAY ASEP--VQ-A EVVQGPEGLA YVIYTSGTTG RPKGVMVEHR
YP_077641        NA------VD AEKEDAAN-L VHANKPGDLA YYIYTSGSTG KPKGVMVEHR
CAA06324         DA-----EKE DAAN-----L EHVNKPEDLA YYIYTSGSTG RPKGVMVEHR
ADK89159         --------MD AEKEDAAN-L EHVNKPEDLA YYIYTSGSTG RPKGVMVEHR
ZP_0687517       FE-------- -ESGAADN-L NLPSAPSDLA YIMYTSGSTG KPKGVMIEHK
CAA49817         FE-------- -ESEAADN-L NLPSAPSDLA YIMYTSGSTG KPKGVMIEHK
NP_388231        FE-------- -ESEAADN-L NLPSAPSDLA YIMYTSGSTG KPKGVMIEHK
BAA02523         FE-------- -ESEAADN-L NLPSAPSDLA YIMYTSGSTG KPKGVMIEHK
Q04747           FE-------- -ESEAADN-L NLPSAPSDLA YIMYTSGSTG KPKGVMIEHK
Clustal Co                                 :* *  ****:**  .**:   :
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
374601          GVL-RLVTDQ NYVPLSERTV TLQSASLLFD AATFEMYAPL LNGGTLVLYP
Q9R919          SVINRLVWMQ ENYPLDERDA ILQKTAITFD VSVWELFWWS IVGSKVVLLP
P94459          SAVNFLNSLQ SRYQLKHSDM IMHKTSYSFD ASIWELFWWP YAGASVYLLP
YP_0034678      NVL-RLIINN GFADIGSDDC IAHCANIAFD ASTWEIWSAL LNGGRLYVVP
YP_0034678      SIN-RLVINN PYADISSDDC VAHCANIAFD ASTWEIWSAL LNGGRLHVVS
YP_0037122      NVL-RLIINS GFADIGPDDC IAHCANMAFD ASTWEIWSAL LNGARLHVVS
YP_0037122      NVN-RLIINN GYADITAEDC VAHCANIAFD ASTWEIWSAL LNGGRLHIVS
YP_0037122      NVL-RLIINN GFADIGPDDC IAHCANMAFD ASTWEIWSAL LNGGCLHVVS
ZP_0726514      AIS-RLVINN GYADFNAQDR VAFASNPAFD ASTLDVWAPL LNGGCVVVIG
YP_235691       AIS-RLAINN GYADFNAQDR VAFASNPAFD ASTLDVWAPL LNGGCVVVIG
AAF99707        AIS-RLVINN GYADFNAQDR VAFASNPAFD ASTLDVWAPL LNGGCVVVIG
YP_0025514      AVV-RTVMAT DYANFGVRET FLQFAPLAFD ASTFEIWGAL LNGGRLVFAP
ZP_0173260      GVI-RLLINS DYVELDEAKT FLHLSPIAFD ASTFEIWGAL LYGGKCIIFP
ZP_0503089      GVI-RLLFGI DYVHLDGKQR LLQMAPISFD AATFEIWGAL LHGARCVLFP
ACF24472        SVN-RLLFGV NYVHLDATQR LLQMAPIAFD ASTFEIWGAL LHGGRCVIFA
CAQ48260        SVN-RLLFGI NYVDLDANET FLQMAPIAFD ASTFEIWGAL LHGARCVLFP
ZP_0051751      GVT-RLVKNS NYVALGEDDI FLQAAPYTFD ASTFEIWGAL LNGGRLVILP
YP_0018044      GVI-RLVKNS NYVDIREDDV FLQAAPYTFD ASTFEIWGAL LNGGRLVILP
ZP_0711340      GVV-RLVKDT NYVNLSAEEV FLQLAPISFD ASTLEIWGSL LNGGRLVIMP
YP_325325       GVV-RLVKQT NYANFTNTEI FLQFAPISFD ASTFEIWGCL LNGGKLVLYP
AAW55330        GVV-RLVKET NYAHLTAEEI ILQLAPISFD ASTFEIWGCL LNGGQLVICP
YP_0015446      NIV-RLVRNT TYAAFGPDQV GLLLATVAFD ASTFELWGCL LNGGRLVIAP
ZP_0359585      SVI-RLVKNS NYITFTPEDR LLMTSSIGFD VGSFEIFGPL LNGAALHLSD
NP_389714       SVI-RLVKNS NYITFTPEDR LLMTSSIGFD VGSFEIFGPL LNGAALHLSD
P39847          SVI-RLVKNS NYITFTPEDR LLMTSSIGFD VGSFEIFGPL LNGAALHLSD
YP_0028026      NVI-RLVKNT NYMDF-KNIR ILQTGSIAFD ASTFEIWGSL LNGGMLCLVN
ACA09733        NVV-RLVKET NYVELNESTR ILQTGAVAFD ASTFEIWGAL LNGGQLYFVE
YP_077641       NIV-RLVKNA GYIPLKSDVK MAQTGAVSFD ASTFEVFGAL LNGGTLYPVP
CAA06324        NIV-RLVKNA GCIPLKSGVK MAQTGAVSFD ASTFEVFGAL LNGGTLYPVP
ADK89159        NIV-RLVKNA GCIPLKSGVK MAQTGAVSFD ASTFEVFGAL LNGGTLYPVP
ZP_0687517      SIL-RLVKNA GYVPIHEEDR MAQTGAVSFD AGTFEVFGAL LNGAALYPAK
CAA49817        SIL-RLVKNA GYVPVTEEDA MAQTGAVSFD AGTFEVFGAL LNGAALYPVK
NP_388231       SIL-RLVKNA GYVPVTEEDR MAQTGAVSFD AGTFEVFGAL LNGAALYPVK
BAA02523        SIL-RLVKNA GYVPVTEEDR MAQTGAVSFD AGTFEVFGAL LNGAALYPVK
Q04747          SIL-RLVKNA GYVPVTEEDR MAQTGAVSFD AGTFEVFGAL LNGAALYPVK
Clustal Co                      .             ** ..  :::      *.
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
374601           HQQ-LDLDEL NRVIQTYQVN TLWLTAALFE KWAHHLAS-K EKVVAL---G
Q9R919           NGGEKNPELI LDTIEQKGVS TLHFVPAML- HAFLESMEQT PSGKLKRKLA
P94459           QGGEKEPEVI AKAIEEQKIT AMHFVPSML- HAFLEHIKYR SVPIKT---N
YP_0034678       PSVLLDPVRF CDSLIKGQVT ALWLTAGLF- ----NEYL-S DLNPLL---G
YP_0034678       SSALLDPVRF RDSLVKGQVT ALWLTAGLF- ----NEYL-G DLEPLF---G
YP_0037122       PSVLLDPVRF CDSLMQGQVT ALWLTAGLF- ----HEYL-D TLKPLY---G
YP_0037122       QSVLLDPAQF RDSLIKGKVT ALWLTAGLF- ----NEYL-D TLKPLL---G
YP_0037122       QPVLLDPVRF CDSLIRGKVT GLWLTAGLF- ----NEYL-D TLKPVF---R
ZP_0726514       QSDLLSPMNF QHLLLEQAVT VLWMTAGLF- ----HQYA-S GLGEAF---S
YP_235691        QHDLLSPLNF QRLLLEQSVS VLWMTAGLF- ----HQYA-S GLGEAF---S
AAF99707         QHDLLSPLNF QRLLLEQSVS VLWMTAGLF- ----HQYA-T GLGEAF---S
YP_0025514       PGK-VGLDEV CDLVQKFNVT TLWLTAGIF- ----QLLS-E EHLQCL---F
ZP_0173260       EKI-PTALTL KEAINQYQVT TLWLTAALF- ----NLVI-D ELPEAF---I
ZP_0503089       ETV-PTAQTL KQVIQTHNIT TLWLTSALF- ----NGIV-A EDAEAL---S
ACF24472         EDI-PTATSL KNAIDKNGIT VLWLTAALF- ----NRII-D DNSQAL---S
CAQ48260         GNI-PTAKSL RDAIDKHGIT ILWLTTALF- ----NAII-D DDSQAL---S
ZP_0051751       SQT-PSLEEI GETLENYGVT TLWLTAGLF- ----QVMV-E EKLESF---K
YP_0018044       SPT-PSLEEL GEAIENYGVT TLWLTAGLF- ----HLMV-E EKLESF---K
ZP_0711340       PHT-PSLQEL GEAIWGYQIT TLWLTAGLF- ----HIMV-D EHLEDL---K
YP_325325        SNT-PSIDEL GQVIQKYQIT TIWLTAGLF- ----HLMV-D ENIHAL---K
AAW55330         PHT-PSLEEL GQIIQQYQVT TLWLTAGLF- ----HLIV-D EKIDAL---K
YP_0015446       PQQ-LSLAEL GHLVEREQIT TLWLTAGLF- ----HQMV-D HALDRL---G
ZP_0359585       QQTFLDSHQL KRYIEHQGIT TIWLTSSLF- ----NHLT-E QNEQTF---S
NP_389714        QQTFLDSHQL KRYIEHQGIT TIWLTSSLF- ----NHLT-E QNEQTF---S
P39847           QQTFLDSHQL KRYIEHQGIT TIWLTSSLF- ----NHLT-E QNEQTF---S
YP_0028026       DKEILNTESI KSNITKNCIN TIWITSAFF- ----NKLA-E EKTSIF---S
ACA09733         NDDILIADRL KAAIAKYGIT TMWLTSPLF- ----NQLS-L QDEYLF---R
YP_077641        KETLLDGKRF RVFLEETGIT TMWLTSPLF- ----NQLA-Q QDPGMF---A
CAA06324         KETLLDGKRF NMFLKETGIT TMWLTSPLF- ----NQIA-Q QDPGMF---A
ADK89159         KETLLDGKRF NMFLKETGIT TMWLTSPLF- ----NQLA-Q QDPGMF---A
ZP_0687517       KETLLDAKQF AAFLREQRIT TMWLTSPLF- ----NQLA-A KDAGMF---G
CAA49817         KRHVLDAKQF AAFLREQSIT TMWLTSPLF- ----NQLA-A KDAGMF---G
NP_388231        KETLLDAKQF AAFLREQSIT TMWLTSPLF- ----NQLA-A KDAGMF---G
BAA02523         KETLLDAKQF AAFLREQSIT TMWLTSPLF- ----NQLA-A KDAGMF---G
Q04747           KETLLDAKQF AAFLREQSIT TMWLTSPLF- ----NQLA-A KDAGMF---G
Clustal Co                .    :    :.  : :.. ::      .
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         210        220        230        240        250
374601           SLRYLLAGGD VVSPTVVKHV -YEKLDNV-- QLINGYGPTE NTTFSVCYPI
Q9R919           SLRYVFASGE ALTPKHVDGF QRIITPVSHA QIINLYGPTE ATIDVSYFEC
P94459           RLKRVFSGGE QLGTHLVSRF -YELLPNV-- SITNSYGPTE ATVEAAFFDC
YP_0034678       RLRYLLIGGD VLDPRKIQRA QLAESQPA-- HLINGYGPTE TTTFATTYRI
YP_0034678       QLRYLLVGGD VLDPRKIRRT QLAECQPA-- HLINGYGPTE TTTFAATYRI
YP_0037122       QLRYLLVGGD ILDPGKIQQV KLAESQPA-- HLINGYGPTE TTTFATTYDI
YP_0037122       QLRYLLIGGD VLDPNKIQQV QSAVLKPT-- YLLNGYGPTE TTTFAATYTI
YP_0037122       QLRYLLIGGD VLDPNKIQQV QLAESKPT-- YLINGYGPTE TTTFAATYTI
ZP_0726514       RLRYLIVGGD VLDPAVIGRV -LANSPPQ-- HLLNGYGPTE ATTFSATYEI
YP_235691        RLRYLIVGGD VLDPAVIGRV -LANNPPQ-- HLLNGYGPTE ATTFSATYEI
AAF99707         RLRYLIVGGD VLDPAVIARV -LANNAPQ-- HLLNGYGPTE ATTFSATYEI
YP_0025514       SLRQLLAGGD VLSLDTINRV -NKALPNC-- QVINGYGPTE ATTFSVCHAF
ZP_0173260       RVKELLTGGE ALSVHHVKKA -LQALPST-- QLINGYGPTE NTTFTCCYSI
ZP_0503089       GVPQLLTGGE ALSVNPVKKA -LAALPST-- QIINGYGPTE NTTFTCCYSL
ACF24472         GIKQLLIGGE ALSVAHVHKA -LAALPLT-- QITNGYGPTE STTFTCCYPI
CAQ48260         GIKQLLIGGE ALSIAHVQKA -LFTLPFT-- QIINGYGPTE STTFTCCYPI
ZP_0051751       NVRYLLAGGD VLSPTHVKTV -LQTYPHC-- SVINGYGPTE NTTFTCCSVL
YP_0018044       NVRYLLAGGD VLFPDHVKTV -LRTYPHC-- CVINGYGPTE NTTFTCCAVL
ZP_0711340       QVRQLLAGGD ILSVPHVQKV -IQELKGC-- QLINGYGPTE NTTFTCCYRI
YP_325325        PLRQLLAGGD VLSVSHVQKF -LKTVENC-- QLINGYGPTE NTTFTCCYHI
AAW55330         SLRQLLAGGD VLSVLHVQKF -LQTVENC-- RLINGYGPTE NTTFTCCHLI
YP_0015446       SLRQLLAGGD RLSPVHVHKV -LERWPQC-- RLINGYGPTE NTTFSCCQQL
ZP_0359585       QLKHLIIGGE ALSPSHVNRI -RNVCPEV-- SIWNGYGPTE NTTFSTCLHI
NP_389714        QLKHLIIGGE ALSPSHVNRI -RNVCPEV-- SIWNGYGPTE NTTFSTCLHI
P39847           QLKHLIIGGE ALSPSHVNRI -RNVCPEV-- SIWNGYGPTE NTTFSTCLHI
YP_0028026       GLRHLLIGGD TLSPKHINMV -LDKCNGL-- KIINGYGPTE NTTFSTAYLI
ACA09733         GLKALLVGGD VLSISHINRV -IEANPDL-- VPINGYGPTE NTTFSTTYKI
YP_077641        TLNDLIIGGD ALVPGIVNRV -KRESPEL-- SLWNGYGPTE NTTFSTCFFI
CAA06324         TLNDLIIGGD ALVPGIVNRV -KRESPEL-- SLWNGYGPTE NTTFSTCFLI
ADK89159         TLNDLIIGGD ALVPGIVNRV -KRESPEL-- SLWNGYGPTE NTTFSTCFLI
ZP_0687517       TLRHVIIGGD ALVPHIVSKV -KQASPSL-- SLWNGYGPTE NTTFSTSFLI
CAA49817         TLRHLIIGGD ALVPHIVSKV -KQASPSL-- SLWNGYGPTE NTTFSTSFLI
NP_388231        TLRHLIIGGD ALVPHIVSKV -KQASPSL-- SLWNGYGPTE NTTFSTSFLI
BAA02523         TLRHLIIGGD ALVPHIVSKV -KQASPSL-- SLWNGYGPTE NTTFSTSFLI
Q04747           TLRHLIIGGD ALVPHIVSKV -KQASPSL-- SLWNGYGPTE NTTFSTSFLI
Clustal Co        :  :: .*:  :    :                  * *****  *

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         260        270        280        290        300
```

```
374601       -PREHSDRFS VPIGRAITNT SVYIVDQHSN LVPKGVVGEL CVGGLGLARG
Q9R919       --EADKRYNS VPIGKPISNI QLYILQ-AGY MQPVGVAGEL CIAGDGLARG
P94459       --PPHEKLER IPIGKPVHHV RLYLLNQNQR MLPVGCIGEL YIAGAGVARG
YP_0034678   -ASPVDVAHS IPIGRPIANT RIYILDCHNQ PVPLGVAGEI YIAGAGVARG
YP_0034678   -SSPVDVNRP IPIGCPIANT QIYLLDPYGQ PVPLGVAGEI HIAGAGVARG
YP_0037122   -ASPVDVTRS IPIGRPIGNT RIYILDSRGQ PVPLGIVGEI HIAGAGVARG
YP_0037122   -PLSIDVTRS IPIGRPISNT QIYILDSYGQ PVPLGVTGEI YIAGFGVARG
YP_0037122   -PSSVDVARS IPIGRPIANT QIYILDSQGR PVPVGVAGEI YIAGNGVARG
ZP_0726514   ---TSAGNGS VPIGKPVGNS RLYVLDSQGQ PVPLGVPGEL YIGGQGVARG
YP_235691    ---VSVGNGS IPIGKPVGNS RLYVLDNQGQ PAPLGVGGEL YIGGQGVARG
AAF99707     ---TSVDNGS IPIGKPVGNT RLYVLDSQGQ PAPLGVAGEL YIGGQGVARG
YP_0025514   ---PKGIATE IPIGKPIANT KVYVLDKCLA PVPIGVVGEL YIAGRGVGRG
ZP_0173260   PSFLESKVSS ILIGRPINNT QIYILDPNLQ PVPVGVPGEL HIGGDGLARG
ZP_0503089   PKQLPGTELS ISIGRPISNT QVYLLDAYWQ PVPIGVIGEL YIGGDGLARG
ACF24472     PKQLEATIKS IPIGRPISNT QVYILDNYLQ PVPIGVVGEL HISGVGVARG
CAQ48260     PKQLETKIKS IPLGKPIANT QVYILDKYLQ PVPVGVSGEL HIGGAGLARG
ZP_0051751   -TDVEQIGYS VPIGQPISQT QVYILDNYLQ PVPFGVPGEL YIGGDGLARG
YP_0018044   -TDVEQIGHS VPIGRPISQT QVYILDPYLH PVPFGVPGEL YIGGGGLARG
ZP_0711340   -TEVNLIENS IPIGRSISNT QVYLLDTHLQ LVPIGVPGEL YIGGDGLARG
YP_325325    -KDPVRPDSS IPIGRPIAHT QVYILDENLQ PVAMGATGEL YIGGDGLARG
AAW55330     -TAPVQPGVS IPIGRPIANT QVYILDNNFQ TVAIGEIGEL HIAGDGLARG
YP_0015446   -SATTDLAQG VPIGQPIANS TAYILDRLLQ LVPIGVVGEL YLGGAGLARG
ZP_0359585   ---QKTYELS IPIGRPVGNS TAFILNQWGV LQPVGAVGEL CVGGDGVARG
NP_389714    ---QKTYELS IPIGRPVGNS TAFILNQWGV LQPVGAVGEL CVGGDGVARG
P39847       ---QKTYELS IPIGRPVGNS TAFILNQWGV LQPVGAVGEL CVGGDGVARG
YP_0028026   ---EKRYYSS IPIGKPISNS NVYIVDKNCN LTPIGISGEL CVGGEGLAKG
ACA09733     ---PGRVEGG VPIGRPISNS TAYVVNESLQ LQPIGAWGEL IVGGEGVARG
YP_077641    ---DQAYERT IPIGKPIGNS TAYIVDEHGA LQPIGVPGEL CVGGDGVARG
CAA06324     ---DQAYERT IPIGKPIGNS TAYIVDEHGS LQPIGVPGEL CVGGDGVARG
ADK89159     ---DQAYERT IPIGKPIGNS TAYIVDEYGS LQPIGVPGEL CVGGDGVARG
ZP_0687517   ---DREYSGS IPIGKPIGNS TAYIMDEQQR LQPIGAPGEL CVGGIGVARG
CAA49817     ---DREYGGS IPIGKPIGNS TAYIMDEQQC LQPIGAPGEL CVGGIGVARG
NP_388231    ---DREYGGS IPIGKPIGNS TAYIMDEQQC LQPIGAPGEL CVGGIGVARG
BAA02523     ---DREYGGS IPIGKPIGNS TAYIMDEQQC LQPIGAPGEL CVGGIGVARG
Q04747       ---DREYGGS IPIGKPIGNS TAYIMDEQQC LQPIGAPGEL CVGGIGVARG
Clustal Co               : :* .: :    ::::       . *  **:  :.* *:.:*
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         310        320        330        340        350
374601           YLNRDDLTQE KFVENQFD-- --TTTSDENR LYRTGDLVRL IDNDLLEYVG
Q9R919           YLNRPELTAE KFVKNPFS-- ------AGER MYRTGDLARW LPDGNIEYLG
P94459           YLNRPALTEE RFLEDPFY-- ------PGER MYKTGDVARW LPDGNVEFLG
YP_0034678       YLNRPELTAE RFVPDTFS-- ----ADPDER MYKTGDLGRW LFDGNIDYLG
YP_0034678       YLNRPELTAE RFLTDPFS-- ----SDPDAR MYKTGDLGRW LPDGNLEYLG
YP_0037122       YLNRPELTTE RFLLDPFS-- ----QGTHAR MYKTGDLGRW LPEGNIEYLG
YP_0037122       YLNHAELTAE RFLTDPFA-- SPFVSNLNAR MYKTGDLGRW LPDGNIEFRG
YP_0037122       YLNRPELTAE RFLADPFS-- ----QDTDAH MYKTGDLGRW LADGNIEYLG
ZP_0726514       YLNRDELTLE KFVADPFD-- ----SDPEAR LYRTGDLVRW RADGNLDYLG
YP_235691        YLHRDELTLE KFVADPFD-- ----SDPQAR LYRTGDLVRW RADGNLEYLG
AAF99707         YLHRDELTLE KFLADPFD-- ----SDPQAR LYRTGDLVRW RADGNLEYLG
YP_0025514       YLNHPSLTCE KFISSPFG-- -----DSGDR LYRTGDLVRW GRDGLLRFLG
ZP_0173260       YLNRPDLTAE KFIPNPFG-- ------TGK- LYKTGDLCRY RRDGNIEYIG
ZP_0503089       YLNRPELTGE KFIANPFS-- ----NQPNAR LYKTGDLARY RADGTIEFVG
ACF24472         YLNRLELTQE KFIANPFS-- ----TDSQSR LYKTGDLARY LPDGNIEYLG
CAQ48260         YLNRLELTAE KFIPNPFE-P LSKVSNQQSK LYKTGDLARY LPDGKIEYLG
ZP_0051751       YLNRPQLTAE RFIASPFA-- ------TGER LYKTGDLVRY DRQRNIEFLG
YP_0018044       YLNRPELTAE RFIPIPPTPI TKGGGKQGER LYKTGDLGRY DRKGNIEFLG
ZP_0711340       YLNRPELTAE RFILNPFS-- ----DKPSDR LYKTGDLARY LPDGNIEFLG
YP_325325        YLHRPELTKE RFIELNNS-- ----NFQSLT LYKTGDLARY LPDGNIEFLG
AAW55330         YLNRPELTAE KFISHSFD-- ----SNLATR LYKTGDLARY LPDGNIEFLG
YP_0015446       YLARPDQTAA AFIPNPMS-- ----QTAGER LYRSGDLARY RDDGTIEFIG
ZP_0359585       YLGRPDLTKE KFVPHPFA-- ------PGDR LYRTGDLARW LSDGTIEYVG
NP_389714        YLGRPDLTKE KFVPHPFA-- ------PGDR LYRTGDLARW LSDGTIEYVG
P39847           YLGRPDLTKE KFVPHPFA-- ------PGDR LYRTGDLARW LSDGTIEYVG
YP_0028026       YLNREDLTAE KFIENPFE-- ------PGKR MYRTGDLARW LPDGNIEFLG
ACA09733         YLNRPDLTAE KFVPSPVK-- ------EGER CYRTGDLVRW LPDGNLEFKG
YP_077641        YLNQPELTDE KFVGDPFA-- ------EGKR MYRTGDLAKW LPDGNIEFLG
CAA06324         YLNQPELTDE KFVGDPFA-- ------EGKR MYRTGDLAKW LPDGNIEFLG
ADK89159         YLNQPELTDE KFVGDPFA-- ------EGKR MYRTGDLAKW LPDGNIEFLG
ZP_0687517       YVNLPELTEK QFLEDPFR-- ------PGER IYRTGDLARW LPDGNIEFLG
CAA49817         YVNLPELTEK QFLEDPFR-- ------PGER IYRTGDLARW LPDGNIEFLG
NP_388231        YVNLPELTEK QFLEDPFR-- ------PGER IYRTGDLARW LPDGNIEFLG
BAA02523         YVNLPELTEK QFLEDPFR-- ------PGER IYRTGDLARW LPDGNIEFLG
Q04747           YVNLPELTEK QFLEDPFR-- ------PGER IYRTGDLARW LPDGNIEFLG
Clustal Co       *:     *    *:                   *::**: :    .  : : *
```

```
                ...

374601          RLD
Q9R919          RID
P94459          RTD
YP_0034678      RND
YP_0034678      RND
YP_0037122      RYD
YP_0037122      RND
YP_0037122      RND
ZP_0726514      RND
YP_235691       RND
AAF99707        RND
YP_0025514      RAD
ZP_0173260      RID
ZP_0503089      RVD
ACF24472        RID
CAQ48260        RID
ZP_0051751      RKD
YP_0018044      RKD
ZP_0711340      RID
YP_325325       RID
AAW55330        RID
YP_0015446      RRD
ZP_0359585      RID
NP_389714       RID
P39847          RID
YP_0028026      RID
ACA09733        RID
YP_077641       RID
CAA06324        RID
ADK89159        RID
ZP_0687517      RID
CAA49817        RID
NP_388231       RID
BAA02523        RID
Q04747          RID
Clustal Co      * *
```

Figure 10B. Sequence alignment of NRPS from clone 374601 with protein sequences with similar conserved core sequences from database**.**

APPENDIX C

Buffer

1. Lysis buffer

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

Adjust pH to 8

2. TBE buffer (5X)

Tris base	54 g
Boric acid	27.5 g
EDTA 0.5 M (pH 8)	20 ml

Water to. 1000 ml

3. PCR reaction for A3f/A7r primer

DNA template	2 µl
Forward primer	0.4 µM
Reverse primer	0.4 µM
dNTPs (each)	0.2 mM
Taq polymerase	1 U
DMSO	2 µl

BIOGRAPHY

Mr. Nattawut Leelakanok has been graduated with a Bachelor of Science in Pharmacy (Second Class Honor) from the Faculty of Pharmaceutical Sciences, Chulalongkorn University, Thailand in 2007. He has been studying for a Master of Science in Pharmacy Program in Microbiology, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Thailand since 2007. During studying in Master's Degree, he was supported by Chulalongkorn University Graduate Scholarship to Commemorate the 72nd Anniversary of His Majesty King Bhumibol Adulyadej, Graduate School, Chulalongkorn University.