# academicJournals

Vol. 7(31), pp. 4065-4072, 2 August, 2013 DOI: 10.5897/AJMR2013.5356 ISSN 1996-0808 ©2013 Academic Journals http://www.academicjournals.org/AJMR

Full Length Research Paper

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

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

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

Accepted 30 April, 2013

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

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

# INTRODUCTION

The biosphere is dominated by microorganisms, which produce numerous secondary metabolites with various biological activities. Efforts to discover new bioactive molecules from microbes have lasted one century. However, the number of novel compounds discovered in recent years has not increased in proportion to the progress of culture-based screening methods (Lu et al., 2008; Firn and Jones, 2000), because only a small fraction of all microbes can be cultured by traditional methods. The knowledge about the "underexplored majority" is still poor (Sogin et al., 2006).

To avoid culture limit, modern biological technology approaches have provided direct access to explore genes or gene clusters that are responsible for the synthesis of microbial secondary metabolites (Cowan et al., 2005; Carlson et al., 2010). Polyketide synthase (PKS) gene clusters are suitable to be the screening target. PKS synthesize polyketides, a large family of secondary metabolites that include many clinically important drugs erythromycin (antibacterial), epothilone such as (antitumor), soraphen (antifungal), rapamycin (immunosuppressant) and lovastatin (antihypercholesterolemic) (Staunton and Weissman, 2001; Fisch et al., 2009). Among the three types of PKS that have been reported to date, only PKS I shows a modular organization, allowing the comparative study of modules within or between

\*Corresponding authors. Email: jiaobh@uninet.com.cn; wsh928@hotmail.com.. Tel: 86-21-65493936; 86-21-81870967. Fax: 86-21-65334344.

enzymes. A minimal module is composed of a ketoacyl synthase (KS) domain; an acyltransferase (AT) domain and an acyl carrier protein (ACP), each module catalyzes the condensation of the acyl groups to the growing acyl chain. Frequently ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains are also embedded in the module to modify the growing acyl chain (Trindade-Silva et al., 2013; Shen et al., 2003). Using standard genetic engineering, enzymes were deleted or added, or their substrate specificity was turned in a particular manner, the numbers of novel polyketides potentially accessible by this route could be enormous (Mondol et al., 2011). Therefore recognition of the availability and diversity of PKS domains in the environment is very important for future drug discovery and combinatorial biosynthesis efforts (Woo et al., 2010). This study focus on the conserved KS modules belonging to the PKS I gene clusters.

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

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

#### MATERIALS AND METHODS

#### Sampling sites

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

#### **DNA extractions**

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

#### PCRs

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

#### **Cloning and Sequencing**

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



**Figure 1.** Agarose gel electrophoresis of total DNA extracted from soil and seawater by MO BIO PowerSoil DNA Isolation Kit. Lane 1,  $\lambda$ DNA/ Hind III marker (1µg); lane 2, DNA extracted from soil; lane 3, DNA extracted from seawater(contents that cannot pass through 0.8 µm cellulose acetate membrane); lane 4, DNA extracted from seawater(contents that pass through 0.8 µm membrane but cannot pass through 0.15 µm membrane).

into pMD18-T Vector according to the manufacturer's instructions (TaKaRa). Positive recombinants were then submitted for sequencing using an ABI3730 DNA Sequencer (USA) with M13 primer at the Invitrogen Biotechnology Company, China.

#### Sequence alignment and phylogenetic analysis

DNA sequences obtained from the PCR product had been translated into amino acid sequences using the Primer Premier 5 software after cutting off the vector and primer regions. Sequence analyses were performed using the BLAST programs provided by the National Center for Biotechnology Information (NCBI). Their closest one or two matches were retrieved from reported gene clusters at NCBI. Finally 42 KS fragments were obtained. All the KS amino acid sequences used in this study were completely aligned using ClustalV 2.0 and adjusted manually using the BioEdit (version 7.01). Mega5 program had been employed to construct the NJ/MP/UPGMA phylogenetic trees. The interior branch length supports were from 1000 replicates. All the sequences used in the phylogenetic analysis could be found in the Supplementary Materials.

Accession numbers. 19 of all the 23 KS sequence data obtained in this study had been submitted to the GenBank databases under accession number DQ640993, DQ640997, DQ641926, DQ641927, and DQ673137 $\sim$ DQ673152.

#### **RESULTS AND DISCUSSION**

#### **DNA** extraction

There were mainly two different extraction techniques

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

## PCRs

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

About twenty 700 bp fragments amplified from the soil sample and three from the seawater sample were purified from agarose gels and then cloned into a pMD18-T Vector. Approximately sixty positive recombinants were selected for the following sequencing. After excluding the identical clones, 23 unique nucleotide fragments (ranged from 630 bp to 690 bp after cutting off the primer and vector sequences) were obtained. Among them 19 clones were amplified from soil (DQ640993, DQ640997, DQ641926, DQ641927, and DQ673137~DQ673151)and 4 from the seawater(01529-2, 0848-1, 0848-2 and DQ673151). The PCR results are shown in Figure 2 and 3.

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

Table 1 . Sequence of primers used in this study.

Primer name	sequences(5'-3')
KSFA	cgcTccATggAcccccAAcA
KSFB	cgcTccATggAcccccAgcA
KSFC	cgcTccATggAcccgcAAcA
KSFD	cgcTccATggAcccgcAgcA
KSFE	cgcTccATggATccccAAcA
KSFF	cgcTccATggATccccAgcA
KSFG	cgcTccATggATccgcAAcA
KSFH	cgcTccATggATccgcAgcA
KSRA	gTcccggTcccgTgcgccTccA
KSRB	gTcccggTcccgTgcgccTcgA
KSRC	gTcccggTcccgTgcgcTTccA
KSRD	gTcccggTcccgTgcgcTTcgA
KSRE	gTcccggTcccgTgccAcTccA
KSRF	gTcccggTcccgTgccATTcgA
KSRG	gTcccggTcccgTggcAcTccA
KSRH	gTcccggTgccATgcgccTccA
KSRI	gTcccggTgccATgcgccTcgA
KSRJ	gTcccggTgccATgcgcTTccA
KSRK	gTcccggTgccATgcgcTTcgA
KSRL	gTcccggTgccATgccAcTccA
KSRM	gTcccggTgccATgccATTcgA
KSRN	gTcccggTgccATggcAcTccA

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

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

#### Sequence analysis

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

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



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



**Figure 3.** Agarose gel electrophoresis of 700 bp KS fragments amplification products from DNA of sea water. Lane 1: Tiangen DNA marker III; lanes 2-4: blank control; lane 5: KSFD/KSRH, 0.15 µm membrane sample; lane 6: KSFB/KSRI, 0.15 µm membrane sample; lane7: KSFD/KSRH, 0.8 µm membrane sample.

	5	15	25	35	45	55	65	75	\$5	95	105	115
DQ673139	<b>GFLALEVTWE</b>	ALEHAGIAPA	SLAGIATONE	IGVCS	SCHPORVLOR	CD-AAI	DAYLA	SO-WHENS	GRVAYFLNLR	<b>GPALSIDIAC</b>	SSSLAALHVA	LOSLESGEVE
DQ67314S	HELLEVIWE	ALESACQAPD	RLGOSRI GVY	VG	SDYACFMLAT	TDSGRL	-DAYYA	SO-NUHSMAA	GRLSVILGLO	GPSLTVDTAC	SSSLVA VALA	VOSLESGECE
DQ673145	<b>GELLLEVCKE</b>	ALEHAGLAPD	FLEGSRIGHT	LO-AAS	SDYCKLVFAE	GP-GAL	DVHSA	SO-NAPS IAA	GRLSYFLGLR	<b>GPSLAVDIAC</b>	SSSLVAVHLA	CECLRSGOCT
DQ640993	CRVLLEVANE	ALDDAGL VLE	FLSOSSTOW	IG-VCS	STALSAGLAA	GLGV	DAYTT	IC-NANS IVA	NEL SYHLDLR	APSMIIDIAC	SSSLVAVILA	COSLRACECD
DQ641927	<b>CFMILLEVVIE</b>	ALADACMIPE	GLAGSET GVF	IG ISN	LEVIFLOAGE	D4AT	-DAYSG	TO-TALS LAA	NEL SYNFOLR	<b>GPSWAVDIAC</b>	SSSLVA VHCA	CESLERKETD
DQ673150	<b>GFLLLEVARE</b>	ALEDANDAID	<b>QLAGIKIGUF</b>	IGIST	NEWGRACIPSH	PSLI	-DAYAG	TO-NALS LAA	NEISYQFDFL	<b>GPSVIIDIAC</b>	SSSI VA VHIA	CRSLWNGEAT
DQ673147	HELLIETSER	AVEHSGIAPT	SLANIQUONE	VO-LST	HDYLOWASSE	LTYPEI	EAYLA	IG-TSGAAGA	GRISYRLGLH	GPA VTVD IAC	SSSLVA INCA	COALCECED
0849-3	HELMLESAWH	ALEHACL PPA	DLVGIRI GVF	MCLST	HEWLSLL THE	MSYRSI	-DAYPG	IG-ISPAAGV	GRISYRLE	GPA VAID TAC	SSSLVAVIDA	CEALRSCECD
DQ541925	OFL ILEVANE	ALEHACQAPD	KLACSRI GVY	VO	IGY ITLVINA.	RGL SEI	DVYLA	IGA VSHA VSS	GRISVILGLO	GPALSLDIAC	STSLVSVILA	COALFLOED
DQ673143	HEPLLQAAWH	ALEN4GVVPS	ELHDISTOW	VG-AGT	GEYARARLDG	40	-DTYAL	TO-SLASENA	GRIAVHLELQ	GPTLADDIAC	SSSLVAVILA	CHELRAGECE
01 529-2	CEMILEVINE	ALE RACOAPG	WLVGIDI GVF	VO-IST	GOVGIEL ISC	DPT-MI	DPYTG	TO-OGEC LAS	GRUSHFLOLE	<b>GENEAIDIAC</b>	SSSLVAVELA	CENTRAGER
00640997	OFL FLECASE	ALEHACCOPA	FROC-RIGHT	AG-AGP	RHWLGPVMDA	LRAGOS	-EHESYRALM	1-NAPDELA	TRVAFELGLE	GPA ITVOTAC	STSLSAVELA	COS111 GECD
DQ673151	<b>CRUFLECARE</b>	AFEHAOCINS	FLPG-AVOVE	AG-LALN	-DYFOFNLAT	EPELTA	-LLGTYAITL	G-NENDFAA	TEVAYKLOLE	GPA IGVOTAC	STSLVALHLA	COSLESCESD
00573144	HENEL BC AWE	ALENA HPPE	FEAC-PUCK	10	-SYFYENICS	NPELVD	-SVG FILEH	TG-NEEDFLA	TRUST IN R	CPS TMOTIC	STSI VA THI A	COSTI & RECD
00673145	HEHL LEC VIE	ALEDACHPTE	KFAG-AIGLE	40-COG	-AYFMENIAS	IPELVR-	-SSGLFLLRH	IG-NEEDFLT	TRISYLUNE	CPSVC/DTAC	STSLVAVELA	COSILIVORCD
00573140	HEVELECAWE	ALFEACYAPD	KHDO-SLOVE	AG-TAVN	-SY-SLSY	ASRIGT	-LECOLLASE	LC-NAADFLS	TRUSYKLDLR	CPSLTICTAC	STSLVA MINA	COSLISCECD
00673138	HELELET CHAN	ALTHALTAPE	SEPO-STON	INTRO 200	-KYELSNI HA	REGOLE	-PETMICAN	LC-NENDVLT	SEVSYKINE	CPSUNIOTAC	STST 12 WOA	CONTI TWOCD
00673149	CEVEL FTCMS	ALE CHOYDES	BAKO-AVOVE	00-VSAN	-TYYLHALHO	RPFITE	-I VISNITUM	HO-NEKDYLA	TEVAVELOUT	CPL UNUSTLC	STSI VE VICE	COSILITYOCD
00673137	CEL EL ET AND	ALENACE CO	FTHE-ATCHE	AG-CYLG	-TYLLANI CA	NEL VIEGILS	FREVER YOTE	1G-NEEDVLT	SEVAYKLOLK	CPL VTICTAC	STSLVAVICA	COST BOGECD
00573142	CRITE PC ANT	AVETACYTER	CCKC-DUCKY	10	-TVCWO-UEA	CEGNI O	-A TADAPET	IC-NEEDVIT	TRUSTICET	CPENETSTAC	STAL VA WITA	INALIDACIO
10673152	CEL 11 FMARE	INF NUCKOPS	ALAGTIC AVY	VG	I TWGTEGI DD	154174		TC-NTI STAA	NEL SYVEDI H	CEST AVDTAC	SCCM21HH	CTCULEGOSS
00573141	CRITITEL VAR	CLEDIGLEED	DCAGERVOVE	2222 EV	-DYWSWIDS	ANESDIDG	-WIN	10-0515115	SET SHA FOR R	CPSFTTDT4C	SSST VA ANT A	KEST SECECS
00/0-1	CHCLI FROM	ALLICACCOVA	STI CAMON	Va-CEA	CERCOURAUT	PACALO-		TO-RACENTC	CRUSEN CAR	CREARING	CACITATUSC	1541 CELECE
Clustel Co	-* *-		SELORITORS		de ton an	· Added	V Look	10 1900 110	*	* ###	*****	*
cruster co												
	- I I	. I. I								. I.I.I		- I
00673139	 125 Val 405/NM	 135 	 145 LTKAHM APD	 155 GRCKTFDAAA		175	 155 RADCORVLAV	 195 IRGSASNODG	205	215 GPACEAVIEA	225	235 DIDYV
DQ673139 DQ673148	 125 VAL AGGVINM MAVAOGVINI	 135 CSFQITIA LSFENTIA	145 LTKAHALAPD PSKARALAAD	 155 GRCKIFDAAA GRCKIFDAFA	 165 DGF ARGE GCG DGF SEGE GCG		 155 RADGERVLAV LEDGERVLAV	195 IRGSASNQDG VRGSAVNQDG	-HSIQE TVPS -HSIQE TVPS -PSAGE TAPS	 215 GPAQEAVIRA GP	 225 ALTIDAQLIOS	235 DIDYV
DQ673139 DQ67314S DQ673145	125 VAL AOGVINM MAVAOGVINI LAL AOGVINI		145 LIKAHALAPD PSKARALAAD LSFMGLLAPD	IS5 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF ARGE GCG DGF SEGE GCG NGF VRGE GCG	175 MLVLKFLADA WVLKFLADA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV	195 IPGSASNQDG VRGSAVNQDG IPG IAVNQDG	-HSIQL TVPS -FSAQL TVPS -PSAQL TAPS -RSNQL TAPN	215 GPAQEAVIRA GP	ALTINGVER	235 DIDYV QVQYV
DQ673139 DQ673148 DQ673146 DQ673146	125 VALAGGVMM MAVAGGVNLI LALAGGVNLI LIVAGGVNLI	135 CSPQI HA LSPENHA LIPOGSLL LIPOGSLL	145 LIKAHALAPD PSKARALAAD LSRMGLLAPD PSRWGMAPD	155 GPC KIFDAAA GPC KIFDAAA GPC KIFDAAA GPC KIFDAAA	165 DOF ARCE GOG DOF SEGE GOG NOF VRCE GOG	175 MLVIKFLADA WVLKFLADA LWIKFLADA WVLKFLADA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV	195 IRGSASNQDG VRGSAVNQDG IRGTAVNQDG VRGSAVNQDG	-HSOGLIVPS -HSOGLIVPS -PSAGLIAPS -RSNGLIAPN -RSIVLIAPN	215 GPAQEAVIRA GP GQAQQALLFD ALAQQAA IFA	ALTDAQLIGS	235 DIDYV QVQYV ELSFV
DQ673139 DQ673148 DQ673146 DQ673146 DQ640993 DQ641927	 125 VALAGGVMM MAVAOGVHLT LALAGGVNLI LIVAGGVNLI RAICOGVNLI	135 -CSPQI HA -LSPENHA -LIPGGSLL -LMPIGAVS -LSPELIEV	145 LIKAHMLAPD PSKAPMLAAD LSFMGLLAPD PSRWBMAPD FTRAGMISAT	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA HRC RIFDAGA	165 DOF ARCE CCG DOF SECE CCG NOF VRCE CCG DOF VRCE CCG DOF VRCE CCG	175 MIVIERIADA WVIERIADA LWIERIADA WVIERIADA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV LSDREPIWAL VADERPVHAV	195 IRGSASNQDG VRGSAVNQDG IRGTAVNQDG VRGSAVNQDG IRGSAVNQDG IRGSAVNQDG	205 -HSOGL IVPS -PSAGL IAPS -RSNGL IAPN -RSIVL IAPN -RSNGL IAPN	215 GPAQEAVIRA GP	ALT DAQLIGS ALT DAQLIGS ALA DAGVRAD ALSNARVSPR ALHAAGVRPS	235 DIDYV QVQYV ELSFV ELGYV
DQ573139 DQ573143 DQ673145 DQ540993 DQ541927 DQ673150	 125 VALAGOVMM MAVAOGVHI LALAOGVNLI LIVAGGVNLI RAICOGVNLI LALAGGVNLI		145 LIKAHAAPD PSKARMLAAD LSRMELLAPD PSRWGMAAPD PTRAGMISAT PSKAGMAPD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KAFDAAA HRC RIFDAGA GRC KAFDAFA	165 DOF ARCE GOG DOF SECE GOG NOF VRCE GOG DOF VRCE GOG DOF VRCE GAG DOF VRCE GAG	175 MLVLKFLADA WVLKFLADA LWLKFLADA WVLKFLADA WVLKFLADA WVLKFLEDA WVLKFLSTA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV LADGERVLAV LADGERVLAV	195 IRGSASNQDG VRGSAVNQDG IRGTAVNQDG VRGSAVNQDG IRGSAVNQDG ILGSAVNDG	205 -KSOGL IVPS -PSAGL IAPS -RSNGL IAPN -RSNGL IAPN -RSNGL IAPN -RSNGL IAPN	215 GPAQEAVIRA GP	ALT DAQLIGS ALT DAQLIGS ALADAGVRAD ALSNARVSPR ALHAAGVRPS AVE LAGVSPG	235 DIDYV QVQYV EISFV EIGYV LVQYV
DQ573139 DQ573145 DQ573145 DQ540993 DQ541927 DQ573150 DQ573150	125 VALAGGVNM MAVAOGVHLT LALAGGVNLI LTVAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGNM	135 CSPQITIA LSPENTIA LSPENTIA LMPIGAVS LSPELTRV LSPELTRV LSPELTRV	145 LIKAMLAPD PSKARMLAAD LSRMCLLAPD FSRWGMAPD FIRAGMSAT FSKAGMAPD FSCSM SPD	ISS GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KAFDAFA HPC RIFDAGA GRC KAFDAFA GRC KAFDAFA	165 DGF ARGE GGG DGF SEGE GGG NGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG	ITS MLVLRFLADA WULRFLADA WULRFLADA WULRFLADA WULRFLADA WULRFLADA WULRFLADA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV LADGERVLAV LADGERVLAV VNDGERLAV	195 IPGSASNQDG VPGSAVNQDG IPG TAVNQDG VPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG	205 -KSQZ IVPS -PSAZ IAPS -RSNZ IAPN -RSNZ IAPN -RSNZ IAPN -RSNZ IAPN -RSNZ IAPN -RSNZ IAPN	CPA QEAVIRA CP- CQA QQALLED ALA QQAA LEA CPA QQEVVEE REA QEAVIEE CA KORVITE	ALI DAQUIOS ALI DAQUIOS ALA DAGVRAD ALANARVSPR ALIHAAGVRPS AYELAGVSPG ALE RAGLAG	235 DIDYV QAQYV EISFV EIGYV LVQYV DIDYL
DQ673139 DQ673145 DQ673146 DQ640993 DQ641927 DQ673150 DQ673147 OS49-3	125 VALACGVINM MAVACGVINI LALACGVINI RAICCGVINI LALACGVINI LALACGVINI		145 145 LTKAHALAPD FSKARMLAAD LSRWGLIAPD FSRWGMAPD FTRAGMSAT FSKARMLSPD FSRARMLSPD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KAFDAFA HPC RIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE COG DGF SEGE COG DGF VFGE COG	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILSTA	ISS RADGERVLAV LEDGERVLAV LADGERVLAV LSDREPVHAV LADREPVHAV VEDGERIRAV VEDGERIRAV	IPS AVNQDG IPS SAVNQDG IPS SAVNQDG IPS SAVNQDG IPS SAVNQDG IPS SAVNQDG IPS SAVNQDG IPS SAVNQDG	205 -ISOG IVPS -PSAG IAPS -PSAG IAPS -RSNG IAPN -RSNG IAPN -RSNG IAPN -ASOG IVPN -ASOG IVPN	215 GPA QEAVIRA GP	ALTDAQLTOS ALTDAQLTOS ALADAGVRAD ALADAGVRAD ALSNARVSPR ALENAGVRPS ALENAGLAG ALERAGLAG ALENAGLAG	235 DIDW QAQW EISFV EIGW LVQW DIDYL FVDYL
DQ673139 DQ673145 DQ673146 DQ640993 DQ640993 DQ641927 DQ673150 DQ673147 OS49-3 DD641926	125 VALAGUNM MAVAGOVELI LALAGUNLI LALAGUNLI LALAGOVELI LALAGGUNLI LALAGGUNLI LALAGGUNLI LALAGGUNKI	ISS -CSPQITIA -LSPENTIA -LSPENTIA -LSPELTRV -LSPDIAIN -LSPDIAIN -LSPAIMIT -LSPAIMIT -LSPAIMAN -LSPAIMAN	145 LIKAMA APD PSKAMA AAD LSFMCLLAPD PSRWOMAPD FTRAGMSAT PSRWOMAPD PSRAMA APD LAKGMA APD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KAFDAFA GRC KAFDAFA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE COG DGF SEGE COG DGF VFGE COG DGF SFALF COG	175 MULIFILADA WULFILADA UVLIFILADA UVLIFILADA WULFILADA WULFILADA VULFILADA UVLIFILADA	ISS RADGEVLAV LADGDQVLAV LADGDQVLAV LADGDQVLAV LADGEVLAV LADGEVLAV FROGERI RAV ERDGERI RAV	195 IFGSASNQDG VRGSAVNQDG IFGSAVNQDG IFGSAVNQDG IFGSAVNQDG IFGSAVNQDG IFGSAVNQDG	205 -ISOG TVPS -PSAG TVPS -RSNG TAPN -RSNG TAPN -RSNG TAPN -RSNG TVPN -ASOG TVPN -RSOG TVPN -RSOG TVPN	215 GPAQEAVIRA GP	225 ALT DAQL TOS ALA DAGVRAD ALA DAGVRAD ALA DAGVRAD ALA DAGVRAS AVELAGVSPS ALQ AGLEPS ALQ AGLEPS ALQ AGLEPS	235 DIDYV QVQYV EISFV EIGYV LVQYV DIDYL AVGYV
DQ673139 DQ673145 DQ64993 DQ64993 DQ64997 DQ673150 DQ673147 0549-3 DQ641926 DQ673143	125 VALACOVIM MAVACOVILI LALACOVINLI LALACOVINLI LALACOVINLI LALACOVINLI LALACOVINLI LALACOVINLI LALACOVINU LALACOVINU	135 -CSPQITIA -LSPENTIA -LSPENTIA -LSPELTRV -LSPELTRV -LSPENTA -LSPATATI -LSPATATI -LSPATATI -LSPATATI	145 145 LIKAMLAPD FSKAMLAD LSRMELAPD FSKAMAPD FSKAMAPD FSKAMAPD FSKAMAPD FSKAMAPD LAKGMLAPD LAKGMLAPD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KAFDAFA GRC KAFDAFA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF ARGE CGG DGF SBCE CGG DGF SBCE CGG DGF VRGE CGG DGF VRGE CGG DGF VRGE CGG DGF VRGE CGG DGF SRAE GAG DGF SRAE GAG	175 MLVLKRLADA WVLKRLADA LWLKRLADA UVLKRLADA WVLKRLADA WVLKRLEDA VLVLKRLEDA MALLKRLEDA	ISS RADGERVLAV LEDGERVLAV LADGQVLAV LADGQVLAV LADREPVHAV LADREPVHAV VEDGERLAV EFDGERLAV EFDGERVHAV	195 IPGSASNQDG IPGSASNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG	205 -KSQL TVPS -PSACL TAPN -RSNCL TAPN -RSNCL TAPN -RSNCL MAPS -ASQL TVPN -RSQCL TAPN -RSQCL TAPN -RSQCL TAPN	215 GPAQEAVIRA GPAQEAVIRA GPAQQALLED ALAQQAA IXA GPAQQEVVRE RRAQEAVIRE GRAQQEVIRA GPAQEAVIRA GPAQEAVIRA GPAQEAVIRA	225 ALTDAQLTOS ALADAGVRAD ALADAGVRAD ALADAGVRPS AYELAGVSPG ALERAGLAG ALQAGLEPG ALERAAVEPA ALDDAGLEPG	235 DIDYV QXQYV EISFV EIGYV LVQYV DIDYL EVDYL AVGYV
DQ873139 DQ873145 DQ873145 DQ84993 DQ841927 DQ873147 0849-3 DQ873147 0849-3 DQ841926 DQ873143 01830-2	125 VAL AGGVNM MAVAGGVHLT LAL AGGVNLT LTVAGGVNLT LAL AGGVNLT LAL AGGVNLT LAL AGGVNLT LAL AGGVNLT LAL AGGVNLT LAL AGGVNLT		145 145 17KAMLAPD FSKAPMLAAD ISRACLAPD FSRAMAAPD FSRAMAPD ISRAPMLAPD LAXGPMLAPD LCSTLALAPD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE GGG DGF SEGE GGG DGF VFGE GGG DGF VFGE GGG DGF VFGE GGG DGF VFGE GGG DGF SFAE GGG DGF SFAE GGG DGF SFAE GGG DGF GFGE GFG	175 MLVLRI ADA WULRI RDA LWULRI RDA LWULRI RDA IIVLRI RDA WULRI STA WULRI STA VLVLRI GDA MALKRI SDA VLVLRI STA	ISS RADGEVLAV LEDGEVLAV LSDEDVLAV LSDEDVLAV LSDEPVHAV VEDGERI FAV EFDGERI FAV EFDGERI FAV LFDGERI LAV LFDGERI LAV	195 IPGSASNQDG VPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG IPGSAVNQDG	205 -NSIGE TVPS -PSAGE TAPS -PSAGE TAPS -RSIGE TAPN -RSIGE TAPN -RSIGE TAPN -ASIGE TVPN -ASIGE TAPN -RSIGE TAPN -RSIGE TAPN -RSIGE TAPN	215 GPAQEAVIFA GPAQEAVIFA GPAQQALLFD ALAQQALIFD GPAQQEVIFE GPAQGEVIFE GPAQGEVIFE GPAQGEVIFE GPAQGEVIFE GPAQGEVIFE GPAQGEVIFE	225 ALTDAQLIGS ALADAGVRAD ALSNARVSFR ALHAAGVRFS ALHAGVSFG ALERAGUSFG ALERAGUSFG ALERAAVEPA ALDAGLAPA ALADAGUAPA	235 DIDIW QQW EISFV EIGW LVQW DIDN EVDN AVGW EVSW
DQ673139 DQ673145 DQ673145 DQ640993 DQ640993 DQ641927 DQ673147 OS49-5 DQ641926 DQ673143 OS49-5 DQ641926 DQ673143 OS29-2 DO640997	125 VAL AGGVNM MAVAGGVHLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT		145 145 LTKAMLAPD FSKARLAAD LSRWELLAPD FSRWEMAPD FSRAMLAPD LSRWELAPD LSRWELAPD LCSSQLAPD LCKTLAIAPD LCKTLAIAPD	155 GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE CCG DGF SEGE CCG DGF VFGE CCG DGF VFGE CCG DGF VFGE CCG DGF VFGE CCG DGF VFGE CCG DGF VFGE CCG DGF SFAE C4G DGF CFGE CCG DGF CFGE CCG	175 MLVLRELADA WVLRELADA WVLRELADA WVLRELADA WVLRELADA WVLRELADA WVLRELADA MALIRELADA MIVLRELADA	ISS FADGERVLAV LEDGERVLAV LEDGERVLAV LSDEPVHAV VADGERLFAV EFDGERLFAV EFDGERLFAV EFDGERLFAV LEDGESLAL LSDGELFAV	195 IRGSASNQDG VRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNDG IRGSAVNDG	205 -KSQL TVPS -PSACL TAPS -PSACL TAPS -RSNCL TAPN -RSNCL TAPN -RSNCL TAPN -RSNCL TAPN -RSNCL TVPN -RSNCL TVPN -RSNCL TVPN -RSNCL TAPN -RSNCL TAPN -RSNCL TAPN -RSNCL TAPN -RSNCL TAPN	215 GPAQEAVIFA GPAQEAVIFA GPAQQALLFD ALAQQAA IFA GPAQQEVIFE GRAQQEVIFE GRAQQEVIFE GFAQQEVIFE GFAQQEVIFE GFAQQEVIFA GLAQAVIFK GLAQAVIFK	ALTDAQLTOS ALTDAQLTOS ALTDAQLTOS ALTDAQLTOS ALTDAQVRAD ALTAGVSPG ALTAGVSPG ALTERAGLAG ALQAGLEPG ALTRAAVEPA ALDDAGLAPA ALDDAGVDPY AVEVAGVDPR	235 DIDW QQYV EISFV EISFV EIGYV LVQYV DIDYL EVDYL AVGYV EVSYV EVSYV SXDIV
DQ673139 DQ673145 DQ673146 DQ640993 DQ640993 DQ641927 DQ673147 OS49-3 DQ641926 DQ673143 O1529-2 DQ640997 DQ673151	125 VAL AGGVNM MAVAGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI VAL AGGVSLI VAL AGGVSLI VAL AGGVSLI		145 LIKAHALAPD FSKAMLAAD LSFMGLIAPD FSR/MMAPD FSR/MMAPD FSR/MLAPD LCSSQLIAPD LCSSQLIAPD LCKTLAIAPD RAEGIKSTD	155 GRC KLIFDAAA GRC KLIFDAAA	165 DGF ARGE GGG DGF SRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF SRAE GGG DGF SRAE GGG DGF GRGE GGG NGF VTSSGLE DGF GRGE GGG	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA MULIFILADA MIVLIFILADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDREPTHAL VADDERVYAV VEDGERTRAV EFDGERTRAV EFDGERTRAV LEDGESTLAL LSDGETTAV	IPS	205 HS OGL TVPS PS AGL TAPS PS AGL TAPS PS TAL TAPS PS TAL TAPS PS TAL TAPS PS TAL TAPS PS TAPS PS OGL TAPS PS O	215 GPA QEAVIFA GP- GPA QEAVIFA GPA QEAVIFA ALA QQAA IFA GPA QEVVFE FRA QEAVIFE GPA QEVVFE GPA QEAVIFA GPS GPS GPS GPS GPS GPS GPS GPS GPS GPS	ALTDAQUIGS ALTDAQUIGS ALADAGVRAD ALSNARVSPR ALENAGVRPS AYELAGVSPG ALENAGVRPS ALENAGVRPS ALENAGVRPA ALDDAGLAPA ALDDAGLAPA ALDDAGUPA DAGUGVRD	235 DIDYV QQQV EISFV EIGYV LVQVV DIDYL AVGVV EVSVV EVSVV EVSVV EVSVV
DQ5731 39 DQ5731 45 DQ5731 45 DQ541927 DQ5731 50 DQ5731 50 DQ5731 47 OS 49-5 DQ6731 47 OS 49-2 DQ6731 43 O1 529-2 DQ6731 51 DQ5731 51 DQ5731 51	125 VAL AGGINMM MAVAGGINEI LAL AGGINEI LAL AGGINEI LAL AGGINEI LAL AGGINEI LAL AGGINEI LAL AGGINEI LAL AGGINEI VAL AGGINEI VAL AGGINEI MAL AGGINEI		145 LIKAHALAPD FSKARALAAD LSRAGLAPD FSRAGMAPD FSRAGMAPD FSRAGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD SAEGGIKSTD VDGGIPSTD VDGGISSD	155 GPC KLIFDAAA GPC KLIFDAAA	165 DGF ARGE COG DGF SEGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF SRAE GAG DGF GRGE COG DGF GRGE COG DGF GRGE COG DGF GRGE COG DGF GRGE COG DGF GRGE COG CGT VRGS CAG	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA MALLFILSA MULIFILSA IVULFILSA WULFILADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDEDPINAL LADFERVIAV VEDGERI RAV EFDGERI RAV EFDGERVRAV LEDGERI LSD LEDGERI LSV LEDGERI LSV LEDGERI LSV LEDGERI LSV	IPS 195 IPCSALSNQDG VPCSAVNQDG IPCSAVNQDG IPCSAVNQDG IPCSAVNQDG IPCSAVNQDG IPCSAVNQDG IPCSAVNQDG IPCSALNAED IPCSALNAED IPCSALNAED IPCSALNAED	205 -NSOC TVPS -PSAC TAPS -PSAC TAPS -RSNC TAPN -RSNC TAPN -RSNC TAPN -RSNC TAPN -RSOC TVPN -RSOC TVPN -RSOC TAPN -RSOC TAPN -RSOC TAPS CREAT APS -RSOC TAPS -RSOC TAPS -R	215 GPAQEAVIRA GP	ALT DAQLTOS ALT DAQLTOS ALA DAGVRAD ALSNARVSPR AYE LAGVSPG ALE RAGLAAG ALQAGLEPG ALD DAGLAPA ALDDAGLAPA ALDDAGLAPA ALDDAGLAPA ALDDAGLAPA ALDDAGLAPA ALDDAGLAPA	235 DIDYV EISFV EISFV EISFV DIDYL EVDYL EVDYL EVSTV EVSTV EVSTV SNQLV IIGVV
DQ5731 39 DQ5731 45 DQ5731 45 DQ540993 DQ5731 47 DQ5731 50 DQ5731 47 OS 49~3 DQ5731 47 OS 49~3 DQ5731 43 OS 529~2 DQ5731 43 DQ5731 51 DQ5731 44 DQ5731 44 DQ5731 44	125 VAL AGGVNM MAVAGGVALI LAL AGGVNLI LIVAGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI VAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVIID		145 145 17KAPMLAPD PSKAPMLAPD PSKAPMLAPD PSRVMAPD PSRVMAPD PSRVMAPD PSRVMAPD LAKGMLAPD LAKTLATAPD RAEGGIKSTD VIDGGIPSPD VIEGEVLSPD	155 GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE COG DGF SEGE COG DGF VEGE COG DGF VEGE COG DGF VEGE COG DGF VEGE COG DGF VEGE COG DGF VEGE COG DGF COG COG DGF COG COG NGT VEGE COG NGT VESS COG QGT VEGE COG NGT VESS COG DGF VEGE COG DGF VEGE COG NGT VESS COG DGF VEGE COG DGF	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILSNA WULFILSNA UVUFILADA WULFILSNA WULFILSNA WULFILSNA WULFILSNA WULFILADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDEDEVLAV LSDEDEVLAV VEDGERI RAV EFDGERI RAV EFDGERI RAV EFDGERI RAV LEDGERI RAV LEDGERI RAV LEDGERI RAV LEDGERI RAV LEDGERI RAV	IPS INSTANCE IPS INSTANCE IPS INNOT IPS INNOT IPS INNOT IPS INNOT IPS INNOT IPS INNOT IPS INNOT IPS INNOT IPS INNOT	205 -NSOC TVPS -PSACLIAPS -PSACLIAPS -RSNCLIAPN -RSNCLIAPN -RSNCLIAPN -RSNCLIAPS -ASOCLIVPN -RSNCLIAPS -RSNCLIAPN -RSNCLIAPS ASKACFIAPS OKLACFIAPS OKLACFIAPS OKLACFIAPS	215 GPA QEAVIRA GP	ALT DAQLIGS ALT DAQLIGS ALA DAGVRAD ALSNARVSPR AYELAGVSPG ALE RAGLAAG ALQAGLEPG ALE RAGVEPS ALDDAGLAPA ALDDAGLAPA ALADORVDPV AYE VAGVDPR ANAVAGVTAD	235 DIDW EISFV EIGW EIGW EIGW DIDN EVSW EVSW SVQLV FVSW SVQLV IIGW IIDW
DQ673139 DQ673145 DQ673145 DQ640993 DQ640993 DQ641927 DQ673147 O649-5 DQ673147 O649-5 DQ641926 DQ673143 DQ673143 DQ673151 DQ673144 DQ673145	125 VALAGOVNIM MAVAGOVNII LALAGOVNII RAICOGVNII LALAGOVNII LALAGOVNII LALAGOVNII LALAGOVNII LALAGOVNIV IALAGOVNIV MALAGOVSII MALAGOVSII MALAGOVSII		145 145 1TKAMLAPD FSKAPMLAAD ISRMELLAPD FSKAPMLAPD FSKAPMLAPD ISSAPMLSPD ISSAPMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKTALLAPD YEGEVLSPD YEGEVLSPD YEGEVLSPD	155 GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE COG DGF SEGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF SFAE CAG DGF GFGE COG DGF GFGE COG CVFAS COA QGT VFGS CAG SGS VL TSCAG	ITS MULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA MULERIADA WULERIADA WULERIADA WULERIADA	ISS RADGERVLAV LEDGERVLAV LSDEDRVLAV LSDEDRVLAV LSDEDRVHAV VEDGERI RAV ERDGERI RAV ERDGERI RAV LEDGERI RAV LEDGERI VSV LEDGERI VSV LEDGERI VSV LEDGERI VSV	IPS INSTANCE IPS INSTANCE IPS INNO IPS INNO	205 -NSIGE TVPS -PSAGE TAPS -PSAGE TAPS -PSNGE TAPN -RSIGE TAPN -RSIGE TAPN -RSIGE TAPN -RSIGE TAPN -RSIGE TAPN AGK LIGF TAPS ASKLOVE TAPS ASKLOVE TAPS ASKLOVE TAPS	215 GPA QEAVIRA GPA QEAVIRA GPA QQAL LFD ALA QQAL LFD ALA QQAL LFD ALA QQAL LFD GPA QQEVIRA GPA QQEVIRA GPA QQEVIRA GRAQGRVIRA GIS QQAVIRA GLAQAAVIRA ELOGIADI IQR VEGQAEA LAR VEGQAEMAE VEGQAEMAE	225 ALTDAQLIGS ALADAGVRAD ALSNARVSFR ALRAAGVRFS ALRAGVRFS ALRAGVRFS ALRAGVRFS ALADGGVRFV ALADGGVDFV AYEVAGVDFR ALADGGVDFY ANALAGVIAD ASALAGVIAD	235 DIDW QQW EISFV EISFV EIGW LVQW DIDU EVSW EVSW EVSW EVSW SVQLV IIGW IIGW IIDW SVARM
DQ673139 DQ673145 DQ673145 DQ640993 DQ640993 DQ641927 DQ673147 OS49-5 DQ651926 DQ651926 DQ653143 O1529-2 DQ65097 DQ673145 DQ673145 DQ673145 DQ673145	125 VAL AGGUNMM MAVAGOVALI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI LAL AGGUNLI VAL AGGUNLI VAL AGGUNLI MAL VGGGRIN MAL AGGUNLI LAL AGGUNLI MAL AGGUNLI LAL AGGUNLI MAL AGGUNLI		145 LIKAMA APD FSKAMA AD LSRMCLIAPD FSRMMAPD FTRAMAPD FSRAM APD LCSSQALAPD LCSSQALAPD LCKTLAIAPD RAEGOIKSTD VIDGEIPSPD VIDGEIPSPD VIEGELISPD HEPGGIASPD	1155 GRC KIFDAAA GRC KIFDAAA	165 DGF ARGE GGG DGF SEGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG SGS VLTS GGG SGS VLTS GGG GGT VRS GAG SGS VLTS GGG GGT VRS GAG	ITS MULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA MULERISA WULERIADA WULERIADA WULERIADA WULERIADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV VADDERVLAV LADRERVLAV VEDGERLRAV EFDGERVAAV QAQEFR ILGV LEDGESTLAL LSDGDT IFAV RADGEHLVAV RADGEHLVAV	IPS	205 -ISOG IVPS -PSAG IAPS -PSAG IAPS -PSAG IAPS -RSTULIAPN -RSTULIAPN -RSOG IVPN -ASOG IVPN -ASOG IVPN -ASOG IAPN -RSOG IAPN -	215 GPA QEAVIRA GP- GQA QQALIFO GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE COLORNIC AND COLORNIC AND COLORNAL VICOLEVIAN VICOLEVIAN		235 DIDW QAQW EISFV EISFV EIGW LVQW LVQW EVDIL AVGW EVDIL AVGW EVSW SVQLV IIGW IIGW IIDW SVARM DIGII IIDW
DQ5731 39 DQ5731 45 DQ5731 45 DQ540993 DQ540993 DQ541927 DQ5731 47 DQ5731 47 DQ5731 43 O1 529-2 DQ5731 51 DQ5731 51 DQ5731 45 DQ5731 45 DQ5731 40	125 VAL AGGVNM MAVAGGVALI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI VALAGGVSIV MALAGGVSIV MALAGGVTID LALAGGVSIX AALAGGVSIX AALAGGVSIX		145 LIKAMA APD FSKAMA AD LSFMCLAPD FSKMMAPD FSKAMAPD FSKAMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD VICOLISTD YEEQELSTD HEPGGIASTD FVEQGIASTD	155 GPC KLIFDAAA GPC KLIFDAAA	165 DGF ARGE COG DGF SEGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF SRAE COG DGF SRAE COG DGF SRAE COG DGF SRAE COG DGF SRAE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG CGT VRGE COG C	ITS MLVLKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELSDA WULKELSDA WULKELSDA WULKELADA WULKELADA WULKELADA WULKELADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDREPVRAV LSDREPVRAV VEDGERIRAV FEDGERVRAV GAQKER ILGV LEDGELIRAV LEDGELIRAV LEDGELIRAV RADGEHIVAV LADGENVALI LADGENVALI	195 IRGSAISNQDG VRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG	205 -NSOGLIVPS -PSAGLIAPS -PSAGLIAPS -RSNGLIAPN -RSNGLIAPN -RSNGLIAPN -RSNGLIAPN -RSOGLIVPN -RSOGLIVPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPS -RSNGFIAPS -	215 GPA QEAVIRA GP- GQA QQALIRO GPA QEAVIRA GPA QEAVIRA GPA QEAVIRA GPA QEAVIRA GRA QQAVIRA GRA QQAVIRA GIS QQAVIRA GIS QQAVIRA GIS QQAVIRA UDG QARCMAE VDG QARCMAE VDG QAEVIAA VDG QAEVIAA	ALT DAQUES ALT DAQUES ALA DAGVRAD ALSNARVSPR ALSNARVSPR ALE RAGLAGG ALQAGLEPG ALE RAGLAGG ALQAGLEPG ALE RAGVRAD ALDAGUSPR ALA NAGVRAD ALA NAGVRAD ALA VAGVRAD ALA VAG IDAG AQSAADVAPD ALA VAG IDAG	235 DIDYV EISFV EISFV EIGW LVQW DIDYL EVDYL AVGW EVSW EVSW EVSW EVSW SQLV SQLV IIGW IIDW SVAPM DIGLI IIDW
DQ5731 39 DQ5731 45 DQ5731 45 DQ541927 DQ5731 50 DQ5731 50 DQ5731 50 DQ5731 47 OS 49-3 DQ6731 43 O1 529-2 DQ6731 51 DQ5731 51 DQ5731 51 DQ5731 40 DQ6731 45 DQ5731 49 DQ5731 49 DQ5731 49	125 VAL AGGVNM MAVAGGVELT LALAGGVNLT LIVAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT VALAGGVNLT VALAGGVNLT VALAGGVTTD MALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD		145 LIKAHALAPD FSKARALAAD LSRAGLAPD FSRWOMAPD FSRWOMAPD FSRWOMAPD FSRWOMAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD LAKGMLAPD TREGEILSPD FVEGIGISPD FVEGIGISPD FVEGIGISPD FVEGIGISPD FVEGIGISPD	155 GPC KLIFDAAA GPC RAFDAFA GPC RAFDAFA GPC RAFDAFA GPC RAFDAFA GPC RAFDAFA GPC RAFDAFA GPC RAFDAFA	165 DGF ARGE COG DGF SEGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF CRGE COG DGF CRGE COG DGF CRGE COG DGF CRGE COG DGF CRGE COG DGF CRGE COG CGT VRGE COG AGT LRGE CAG CGT VRGE COG CGT VRGE COG	ITS MLVLKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA WULKELADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDEDPINAL LSDEDPINAL LSDEDPINAL LSDEDVIAV VEDGERIAV LEDGERIAV LEDGERIAL LSDGDIBAV LEDGERIAV LEDGERIAV LEDGERIAV LADGERIAV LADGERIAV	195 IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGTAINADA IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG IRGSAINADG	205 -NSOC TVPS -PSAC TAPS -PSAC TAPS -PSAC TAPS -RSOC TAPS -RSOC TVPS -RSOC TVPS -RSOC TVPS -RSOC TVPS -RSOC TAPS ASKACYTAPS ASKACYTAPS STRVSFAAPS STRVSFAAPS STRVSFAAPS	215 GPA QEAVIRA GPA QEAVIRA GPA QEAVIRA GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRA GLAQAVIRE CIS QGVIEA GLAQAAVIRE DIG JALANA VEG QALENIAE VEG QALENIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE VEG QALEVIAE	LI DAQLIGS ALI DAQLIGS ALA DAQVRAD ALSNARVSPR AYE LAGVSPG ALE RAGLAAG ALQAGLEPG ALQAGLEPG ALQAGLEPG ALQAGLEPG ALDDGLAPA ALDDGLAPA ALDDGLAPA ALADGRVDPV AYE VAGVDPV ALAVGGVDPP ALAVAG IDAG AQSAADVAPD ALAVGGIDAG AQSAADVAPD	235 DIDYV EISFV EISFV EIGVV LVQVV DIDYL EVSVV EVSVV EVSVV EVSVV EVSVV SVQLV IIGVV SVQLV IIGVV SVQLV IIDVV SVAFM DIGLI IIDVV
DQ673139 DQ673145 DQ640993 DQ640993 DQ641927 DQ673147 OS49-3 DQ641926 DQ673147 OS49-3 DQ641926 DQ673147 DQ673145 DQ673145 DQ673145 DQ673145 DQ673149 DQ673149 DQ673149 DQ673149	125 VAL AGGVNM MAVAGGVALI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI LAL AGGVNLI MAL AGGVNLI MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII MAL AGGVSII		145 145 17KAPMLAPD PSKAPMLAPD PSKAPMLAPD PSRVMAPD PSRVMAPD PSRVMAPD PSRVMAPD PSRVMAPD PSRVMAPD LAKTLAIAPD LASSQLAPD LASSQLAPD LASSQLAPD VDGGIPSPD VPEGEILSPD VPEGEILSPD VPEGGIGSPD VPEGGISPD VPEGGISPD VPEGGISPD VPEGGISPD	155 GRC KIFDAAA GRC KIFDAAA	165 DGF AFGE COG DGF SEGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF VFGE COG DGF CFGE COG DGF CFGE COG NGT VFGE COG NGT VFGE COG SGS VLTSCAG SGS VLTSCAG QGT VFSS COG QGT VFSS COG CGT CGT CGT CGT	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDEDEVLAV LSDEDEVLAV VEDGERVRAV VEDGERIRAV EFDGERIRAV EFDGERIRAV EFDGERVRAV LEDGERIRAV LEDGERIRAV LEDGERIRAV RADGERIYAV RADGERIYAV MEDGERIYAV MEDGERIYAV	195 IRGSASNQDG VRGSAVNQDG IRG TAVNQDG IRG TAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRG TAUNNDG IRG TAUNNDG IRG TAUNNDG IRG SAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAUNNDG IRGSAUNNDG IRGSAUNNDG IRGSAUNNDG IRGSAUNNDG IRGSAUNNDG	205 -NSOCLTVPS -PSACLTAPS -PSACLTAPS -PSACLTAPS -RSNCLTAPN -RSNCLTAPN -RSNCLTAPN -RSNCLTAPN -RSNCLTAPN -RSNCLTAPS OKLOF TAPS OKLOF TAPS OKLOF TAPS OKLOF TAPS SCRUSP AAPC SCRUSP AAPC SCRUSP TAPS SCRUSP AAPC SCRUSP TAPS	215 GPA QEAVIRA GP	ALT DAQLIGS ALT DAQLIGS ALA DAGVRAD ALSNARVSPR AYELAGVSPG ALE RAGLAAG ALQAGLEPG ALE RAGLAAG ALDAGLAPA ALDDAGLAPA ALDDAGLAPA ALADGRVDPV AYE VAGVDPR ANAVAGVTAD ALAVAG TDAG AQSAADVAPD ARALACVRPE AQALGSVARE	235 DIDW EISFV EISFV EISFV EIGW LVQW DIDL EVDU EVSW SVQLV IIGW IIGW IIDW SVARM DIGLI IIDW IISVI IISVI IISVI
DQ6731.39 DQ6731.45 DQ640993 DQ640993 DQ640993 DQ641927 DQ6751.60 DQ6731.47 OS49-5 DQ6731.43 OS29-2 DQ641926 DQ6731.43 DQ6731.45 DQ6731.45 DQ6731.45 DQ6731.49 DQ6731.37 DQ6731.37 DQ6731.37 DQ6731.20	125 VAL ADGVNM MAVAGOVALI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI LAL ADGVNLI VAL ADGVSNL MAL VAGVSNL MAL VAGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL MAL ADGVSNL		145 145 LIKAMA APD FSKAMA AD LSRMCLIAPD FSKAMAPD FTRAMAPD FSRAMAPD LOSSALAPD LOSSALAPD LOSSALAPD LOSSALAPD LOSSALAPD LOSSALAPD LOSSALAPD VECGIOSPD	1155 GPC KLIFDAAA GPC KLIFTAAAA GPC KLIFTAAAA GPC KLIFTAAAA GPC KLIFTAAAA GPC KLIFTAAAA GPC KLIFTAAAAA GPC KLIFTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	165 DGF ARGE GGG DGF SEGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG QGT VRGE GGG Q	ITS MULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA MULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA WULERIADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV LADGERVLAV LEDGERVLAV GAGERIRAV EFDGERVAAV GAGERIRAV LEDGESILAL LSDGEIIRAV LEDGESILAL LSDGEIIRAV LADGEHVAAI LADGEHVAAI LADGEHVAAI LADGEHVAAI LADGEHVAAI LADGEHVAAI	IPS	205 -ISOG IVPS -PSAG IAPS -PSAG IAPS -PSAG IAPS -RSTILIAPN -RSTILIAPN -RSTILIAPN -ASOG IVPN -ASOG IVPN -ASOG IVPN -ASOG IAPN -RSQC IVPN -RSQC -RSQC -RSQC -RSQC -RSQC -RSQC -RSQC -RSQC -RSQC -	215 GPA QEAVIFA GP- GQA QQALIFO GQA QQALIFO GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE GPA QEAVIFE HIG QEAVIFE VEG QUEAL TAR VEG QUEAVIAN VEG VEG VEG VEG VEG VEG VEG VEG	ALL DAGUAR ALL DAGUAR ALA DAGUAR ALSNRYSPR ALSNRYSPR ALE RAGLAGU ALE RAGLAGU ALE RAGLAGU ALE RAGUAR ALQORVEPY ALA DORVEPY ANY UGVER ANY UGVER ALA VGGVEP ALA VIG DAG AGS ADVAPD ASALLCVISPE ALA VIG TIAG	235 DIDW QQW EISFV EISFV EIGW LQW LQW EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL EVDIL IIGW IIGW IIDW SARM DIGLI IIDW SLSW
DQ673139 DQ673145 DQ673146 DQ640993 DQ640993 DQ641927 DQ673147 O849-3 DQ673147 O849-3 DQ673143 O1529-2 DQ673143 DQ673145 DQ673145 DQ673145 DQ673135 DQ673137 DQ673142 DQ673142 DQ673142	125 VAL AGGVNM MAVAGGVALI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI LALAGGVNLI VALAGGVNLI MALAGGVNLI MALAGGVSIK AALAGGVSIK MALAGGVSIK MALAGGVSIK MALAGGVSIK AALAGGVSIK MALAGGVSIK AALAGGVSIK AALAGGVSIK		145 145 LIKAMA APD FSKAMA AD LSFAML APD FSKAMA PD FSKAMA PD FSKAMA PD FSKAMA PD LSSQAL APD LCSTAL APD LCSTAL APD LCSTAL APD LCSTAL APD LCSTAL APD LCSTAL APD LCSTAL APD PYE CELL SPD YPE CELL SPD YPE CELL SPD YPE CELL SPD YPE COLL SPD	1155 GRC KLIFDAAA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA GRC RAFDAGA	165 DGF ARGE GGG DGF SRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF VRGE GGG DGF SRAE GGG DGF SRAE GGG DGF SRAE GGG CGT VRGE GGG GGT VRGE GGG GGT VRGE GGG GGT VRGE GGG SGT VRGE GGG SGT VRGE GGG SGT VRGE GGG SGT VRGE GGG SGT VRGE GGG SGT VRGE GGG	ITS MULICIANA WULCIAN	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LADGERVLAV LADGERVLAV VEDGERVLAV VEDGERVLAV VEDGERI RAV EFDGERVLAV LEDGERI RAV LEDGESI LAL LSDGEI GAV RADGENVALI LADGENVALI LADGENVALI LADGEN VAV IEDGERVI VAV IEDGERVI VAV IEDGERVI VAV IEDGERVI VAV IEDGERVI VAV	III III III IIII IIII IIII IIIII IIIIII	205 -NSOGLIVPS -PSAGLIAPS -PSAGLIAPS -RSINGLIAPN -RSINGLIAPN -RSINGLIAPN -RSINGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN -RSIGLIAPN SIKVSYLAPS SIKVSYLAPS SIKVSYLAPS SIKVSYLAPS GRAAYIAPS GRAAYIAPS GRAAYIAPS	215 GPA QEAVIFA GP- GQA QQAL LFD GPA QEAVIFA GPA QEAVIFE GPA QEAVIFE GPA QEAVIFA GPA QEAVIFA GPA QEAVIFA GPA QEAVIFA GPA QEAVIFA GPA QEAVIFA CLAQAAVIFA VEGAALAMAE VEGAALAM	225 ALTDAQLIGS ALADAGVRAD ALSNARVSPR ALSNARVSPR ALENAGVRPS AYELAGVSPG ALENAGVRPS ALENAGVRPS ALENAGVRPA ALDDAGLAPA ALDDAGLAPA ALADGRVDPV AXEVAGVRAD ANALAGVRAD AGALOGVPAE ALANGGVAE ALASOGVASE VLINAGIASE	235 DIDYV EISFV EISFV EISFV DIDYL EVDYL AVGYV EVSYV EVSYV EVSYV SVQLV IIGYV IIGYV IIGYV IIGYV IIDVV SVQFM DIGLI IIDYV SLSYV EVDFM
DQ5731 39 DQ5731 45 DQ5731 45 DQ540993 DQ540993 DQ541927 DQ5731 45 DQ5731 43 O1 529-2 DQ5731 43 O1 529-2 DQ5731 51 DQ5731 45 DQ5731 45 DQ5731 49 DQ5731 49 DQ5731 42 DQ5731 42 DQ5731 42 DQ5731 42 DQ5731 42 DQ5731 42	125 VAL AGGVNM MAVAGGVELT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT LALAGGVNLT VALAGGVSTT MALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD LALAGGVTTD MALAGGVTTS AALVGGVML LALVGGVML LALVGGVML		145 145 LTKAHALAPD FSKAMLAPD FSKAMLAPD FSKAMLAPD FSKAMAPD FSKAMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD LAKGMAPD FSKAMAPD VECGINSPD VFECELISPD FVECGINSPD VECGINSPD VECGINSPD VECGINSPD VECGINSPD VECGINSPD FIKAMASN SKAMASN	155 GPC KLIFDAAA GPC RAFDAFA GPC RAFDAFA GPS LIFFDAFA	165 DGF ARGE COG DGF SECE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF SRAE GAG DGF SRAE GAG DGF SRAE GAG CGT VRGE COG GGT VRGE COG GGT VRGE COG SGT VRGE COG SGT VRGE COG SGT VRGE COG SGT VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG	ITS MULIFILADA WULFILADA WULFILADA WULFILADA WULFILFILA WULFILFILA WULFILFILA MALIFILGA WULFILADA WULFILADA WULFILADA WULFILEDA ULLIFILDA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LSDREPVBAV SDREPVBAV VEDGERI RAV FEDGERI RAV FEDGERI RAV FEDGERI RAV LEDGERI RAV LEDGERI RAV LEDGERI VAV LEDGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV LADGERI VAV	195 IRGSAISNQDG VRGSAVNQDG IRGSAVNDG IRGSAVNDG IRGSAVNDG IRGSAVNQD	205 -NSOGLIVPS -PSAGLIAPS -PSAGLIAPS -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN -RSOGLIAPN SCHOFIAPS SC	215 GPA QEAVIRA GP- GQA QQALIRO GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE USGA DATA VIEGAALAIRE VIEGAALAIR	ALT DAQLIGS ALT DAQLIGS ALSNARVSPR ALSNARVSPR ALSNARVSPR ALE RAGLAGG ALQAGLEPG ALE RAGLAGG ALQAGLEPG ALDDAGLAPA ALDDAGLAPA ALDDAGVAP ALAVAG IDAG AQALGGVAE ALAVAG IDAG AQALGGVAE ALAVGGVAE ALAVGGVAE ALAVGGVAE ALAVGGVAE ALAVGGVAE ALAVGGVAE ALAVGGVAE	235 DIDYV EISFV EISFV EIGW LVQW DIDYL EVDYL EVSW EVSW EVSW EVSW EVSW EVSW SQLV IIGW IIDW SVRM DIGLI IIDW SVRM DIGLI IIDW SVRM EVDRM SVRW EVDRM
DQ5731 39 DQ5731 45 DQ5731 45 DQ541927 DQ5731 50 DQ5731 50 DQ5731 50 DQ5731 50 DQ5731 47 OS 49-3 DQ5731 43 OI 529-2 DQ5731 51 DQ5731 51 DQ5731 40 DQ5731 45 DQ5731 49 DQ5731 49 DQ5731 42 DQ5731 52 DQ5731 42 DQ5731 52 DQ5731 42 DQ5731 52 DQ5731 42 DQ5731 52 DQ5731 42 DQ5731 52 DQ5731 42 DQ5731 43 DQ5731 45 DQ5731 42 DQ5731 42	125 VAL AGGINMM MAVAGGINLI LALAGGINLI LIVAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI VALAGGINLI VALAGGINLI VALAGGINLI VALAGGINLI VALAGGINLI VALAGGINLI LALAGGINLI VALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI LALAGGINLI VALAGGINLI LALAGGINLI VALAGGINLI LALVASANLM RALAGGINMI LALVASANLM RALAGGINMI		145 145 LTKAHMLAPD FSKAPMLAAD FSKAPMLAPD FSKAPMLAPD FSKAPMLAPD FSKAPMLAPD LAKGPMLAPD LAKGPMLAPD LAKGPMLAPD LAKGPMLAPD LAKGPMLAPD LAKGPMLAPD VEGGISSPD VEGGISSPD VEGGISSPD VEGGISSPD VEGGISSPD FXEGGISSPD F	155 GPC NIFDAAA GPC NIFDAAA	165 DGF ARGE COG DGF SEGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF VRGE COG DGF CRGE COG CGT VRGE COG QGT VRGE COG QGT VRGE COG DGF VRGE COG D	ITS MULIFILADA WULFILADA WULFILADA WULFILFIL WULFILFIL WULFILFIL WULFILSA WULFILSA IVULFILSA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA WULFILADA	ISS RADGERVLAV LEDGERVLAV LEDGERVLAV LEDGERVLAV SDERPVLAV SDERPVLAV SDERVVLAV SEDGERVKAV SEDGERVKAV LEDGERVKAV LEDGESILAL LSDGEI DAV LEDGESILAL LSDGEI SAV LEDGERI YAV LADGERVVAI LADGERVVAI LADGERVVAI LADGERVVAI LADGERVVAI SSDERI YAV SSDERI YAL -FGGEDAAQ	195 IRGSAISNQDG VRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNQDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSAVNNDG IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR IRGSALNNDR	205 -NSOCI TVPS -PSACI TAPS -PSACI TAPS -RSOCI TAPN -RSOCI TAPN SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS SIGNAFIAPS -RIFICI TVPS -RIFICI TVPS -RIFICI TVPS -RIFICI TAPS -RIFICI TAPS -RIFI	215 GPA QEAVIRA GPA QEAVIRA GPA QEAVIRA GPA QEAVIRE GPA QEAVIRE GPA QEAVIRE GPA QEAVIRA GLAQAVIRA GLAQAVIRA CISQ QXVIRA GLAQAAVIRA DIG QALANARE VIEG QALENARE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALEVIAE VIEG QALIRES LQA CROMINE	LI DAQLIGS ALI DAQLIGS ALI DAQLIGS ALSNARVSPR ALBAAGVRPS AYE LAGVSPG ALE RAGLAAG ALQAGLEPG ALDAGUSPG ALDAGUSPA ALDAGUSPA ALDAGUSP ALAVAG DAG AGSAADVAPD ALAVAG DAG AQSAADVAPD ALAVAG DAG AQSAADVAPD ALAVAG DAG AQSAADVAPD ALASGGVASE VLNKAG IAPS AYADAGVAG	235 DIDYV EISFV EISFV EIGW LVQW DIDYL EVSW EVSW EVSW EVSW EVSW EVSW EVSW EVSW

Figure 4. Clustal V alignment of predicted amino acid sequences of soil and seawater PKS I gene fragments. (\*, identity; : , strong similarity; . , weak similarity). Soil KS genes obtained in present study and their identities in the GenBank were as follows: dz-ks-1, 63% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (215amino-acid[aa] alignment);dz-ks-2, 53% identity to Nostoc sp. PCC 7120 PKS gene (223aa alignment);dz-ks-A1, 64% identity to Nostoc punctiforme PCC 73102 PKS gene (218aa alignment); dz-ks- P1, 66% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene(215aa alignment); dz-ks-810-3, 61% identity to Anabaena variabilis ATCC 29413 beta- ketosynthase gene(226aa alignment); dz-ks-HH-1, 85% identity to uncultured bacterium beta ketosynthase gene (224aa alignment); dz-ks-HH-2, 61% identity to Crocosphaera watsonii WH 8501 betaketosynthase gene (216aa alignment); dz-ks-HK-1, 61% identity to Lyngbya majuscula JamP gene 218aa alignment); dz-ks-HK-3, 48% identity to Nodularia spumigen NdaFgene (230 aa alignment); dz-ks-HN-1, 51% identity to Lyngbya majuscula BarE gene (218aa alignment); dz-ks-HN-2, 65% identity to Polyangium cellulosum soraphen polyketide synthase A(213aa alignment); dz-ks-DG-1, 76% identity to Myxococcus xanthus DK 1622 PKS/NRPS synthase gene (221aa alignment); dz-ks-DG-3, 66% identity to Myxococcus xanthus DK 1622 PKS/NRPS synthase gene (221aa alignment); dz-ks-DH-1, 63% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene(216aa alignment); dz-ks-DH-2, 86% identity to Mycobacterium sp. JLS beta-ketoacyl synthase gene (217aa alignment); dz-ks-48-3, 66% identity to Crocosphaera watsonii WH 8501 beta-ketoacyl synthase gene(193aa alignment); dz-ks-Dl-3, 58% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (220aa alignment); dz-ks-BH-2, 67% identity to Anabaena sp. WH 0404 KS gene (215aa alignment); dz-ks-BK-1, 58% identity to Nostoc punctiforme PCC 73102 PKS gene (220aa alignment). Seawater KS genes obtained in present study and their identities in the GenBank were as follows: 015-2, 75% identity to Nitrosomonas europaea ATCC 19718 WcbR gene (216aa alignment); 01529-2, 64% identity to Nostoc sp. PCC 7120 all1648 gene (216aa alignment); 0848-1, 45% identity to Anabaena variabilis ATCC 29413 beta-ketosynthase gene (210aa alignment); 0848-3, 76% identity to Mycobacterium vanbaalenii PYR-1 beta-ketosynthase gene (217aa alignment).

unit found that aspartate (Pavlidou, 2011).

Among the 23 KS sequences obtained in this study, 9 sequences displayed a unique pattern N(DE)KD, 22 amino acids upstream from the cysteine active site in the KS domain and the conserved pattern VDTACSSS was replaced by VQTACSTS (these amino acid residues were underlined in Figure 4). These two patterns were shown to identify KS domains belonging to hybrid NRPS/PKS systems, the remaining 15 KS sequences (the 4 KS sequences derived from the sea water included) showed typical conserved patterns of KS domains. No KS<sup>Q</sup> like fragment was found from all the 23 KS fragments.

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

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

# Phylogenetic analysis

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

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

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

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

Data presented in this study showed that the PCR method using degenerate primer to isolate the secondary metabolites biosynthesis gene fragments from the environmental samples were practically effective. This



Figure 5. Phylogenetic analysis of the KS fragments amplified by PCR from soil and seawater DNA and the closest sequences derived from the NCBI. The reconstructiong was computed for all 64 KS amino acid sequences by the distance method (NJ, Posson correction distance model) with interior branch length supports 1000 replicates. E. coli FabF was used as outgroup. Only bootstrap value  $\geq$ 50% are shown ( $\blacktriangle$  KS fragments that belong to Hybrid or Mixed PKS/NRPS enzymes complexes; •KS fragments that belong to the Type I ketosynthase).

study will lay the foundation for the biosynthesis and heterogeneous expression of polyketides and contribute to the exploitation of microbial genetic resources belong to the "underexplored majority".

## ACKNOWLEDGEMENTS

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

#### REFERENCES

- Bisang C, Long PF, Cortes J, Westcott J, Crosby J, Matharu AL, Cox RJ, Simpson TJ, Staunton J,Leadlay PF (1999) A chain initiation factor common to both modular and aromatic polyketide synthases. Nature 401:502-505.
- Carlson JC, Fortman JL, Anzai Y, Li S, Burr DA, Sherman DH(2010): Identification of the tirandamycin biosynthetic gene cluster from Streptomyces sp. 307-9.Chembiochemistry 11(4):564-72
- Cheng YQ, Tang GL, Shen B (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. USA 100:3149-3154.
- Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR, Nalin R, Guerineau M, Jeannin P, Simonet P,Pernodet JL (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl. Environ. Microbiol. 69:49-55.
- Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R,Wittwer P (2005) Metagenomic gene discovery: past, present and future. Trends Biotechnol. 23:321-329.
- Edwards DJ, Gerwick WH (2004). Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. J. Am. Chem. Soc. 126:11432-11423
- Ehrenreich IM, Waterbury JB, Webb EA (2005) Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. Appl. Environ. Microbiol. 71:7401-7413.
- Firn RD, Jones CG (2000) The evolution of secondary metabolism a unifying model. Mol. Microbiol. 37:989-994
- Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, Taudien S, Platzer M, Rubio BK, Robinson SJ, Crews P, Piel J(2009). Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting.Nat. Chem. Biol. 5(7):494-501
- Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K, Bertrand H, Vogel TM, Perriere G, Simonet P,Nalin R (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl. Environ. Microbiol. 70:5522-5527
- Jenke-Kodama H, Sandmann A, Muller R,Dittmann E (2005). Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22:2027-2039.
- Kim TK, Hewavitharana AK, Shaw PN, Fuerst JA (2006) Discovery of a new source of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction. Appl. Environ. Microbiol. 72:2118-2125
- Lim GE, Haygood MG (2004) "*Candidatus Endobugula glebosa*," a specific bacterial symbiont of the marine bryozoan *Bugula* simplex. Appl. Environ. Microbiol. 70:4921-4929.

- Lopanik NB, Targett NM,Lindquist N (2006) Isolation of two polyketide synthase gene fragments from the uncultured microbial symbiont of the marine bryozoan *Bugula neritina*. Appl. Environ. Microbiol. 72:7941-7944.
- Lu XL, Xu QZ, Liu XY, Cao X, Ni KY, Jiao BH (2008). Marine Druga-Macrolactins. Chem. Biodiver. 5:1669-1674.
- Moffitt MC, Neilan BA (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. J. Mol. Evol. 56:446-457.
- Mondol MA, Kim JH, Lee HS, Lee YJ, Shin HJ (2011). Macrolactin W, a new antibacterial macrolide from a marine *Bacillus* sp. Bioorg. Med. Chem. Lett. 21(12):3832-3855.
- Pavlidou M, Pross EK, Musiol EM, Kulik A, Wohlleben W, Weber T (2011). The phosphopantetheinyl transferase KirP activates the ACP and PCP domains of the kirromycin NRPS/PKS of Streptomyces collinus Tü 365.FEMS Microbiol. Lett. 319(1):26-33.
- Piel J (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. Proc. Natl. Acad. Sci. USA 99:14002-14007.
- Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N,Matsunaga S (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. USA 101:16222-16227.
- Rose TM, Henikoff JG,Henikoff S (2003) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. Nucleic Acids Res. 31:3763-3766.
- Schirmer A, Gadkari R, Reeves CD, Ibrahim F, DeLong EF, Hutchinson CR (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl. Environ. Microbiol. JT 71:4840-4849.
- Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr. Opin. Chem. Biol. 7:285-295.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. USA 103:12115-20.
- Staunton J, Weissman KJ (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18(4):380-416.
- Trindade-Silva AE, Rua CP, Andrade BG, Vicente AC, Silva GG, Berlinck RG, Thompson FL(2013) Polyketide synthase gene diversity within the microbiome of the sponge Arenosclera brasiliensis, endemic to the Southern Atlantic Ocean. Appl. Environ. Microbiol. Mar; 79(5):1598-605.
- Woo PC, Tam EW, Chong KT, Cai JJ, Tung ET, Ngan AH, Lau SK, Yuen KY(2010). High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in Penicillium marneffei. FEBS J. 277(18):3750-3758.
- Zhou J, Bruns MA,Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62:316-322.