Mining Metagenomes Reveals Diverse Antibiotic Biosynthetic Genes in Uncultured Microbial Communities

Dina H. Amin (dina.hatem@sci.asu.edu.eg)
Ain Shams University Faculty of Science https://orcid.org/0000-0002-3673-8889
Wedad M. Nageeb
Suez Canal University
Amr Elkelish
Suez Canal University
Rabab R. Makharita
Suez Canal University

Research Article

Keywords: Bioinformatic tools, Metagenomic library, NRPS, Phylogenetic analysis, antimicrobials.

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1214433/v1

Abstract

Background: Pathogens resistant to antimicrobials form a significant threat to public health worldwide. Tackling multidrug-resistant pathogens via screening metagenomic libraries has become a common approach for the discovery of new antibiotics from uncultured microorganisms. This study focuses on capturing non-ribosomal peptide synthase (NRPS) gene clusters implicated in the synthesis of many natural compounds of industrial relevance. A NRPS PCR assay was used to screen 2976 *Escherichia coli* clones in a soil metagenomic library to target NRPS genes. Bioinformatics analysis were conducted to detect predict NRPS domains and their substrate specificity.

Results: Successfully, 17 NRPS positive hits with a biosynthetic potential were identified. DNA sequencing and BLAST analysis confirmed that NRPS protein sequences shared similarities with members of genus *Delftia* in the *Proteobacteria* taxonomic position. Multiple alignment and phylogenetic analysis demonstrated that clones no. 15cd35 and 15cd37 shared low bootstrap values (54%) and were distantly far from close phylogenetic neighbors. Additionally, NRPS domain substrates specificity has no hits with the known ones hence they are more likely to use different substrates to produce new diverse antimicrobials.

Conclusions: We confirmed that the analyses of the soil metagenomic library revealed a diverse set of NRPS related to the genus *Delftia.* An in-depth understanding of those positive NRPS hits is a crucial step for genetic manipulation of NRPS, shedding light on alternative novel antimicrobial compounds that can be used in drug discovery and hence supports the pharmaceutical sector.

Background

According to a report released by the World Health Organization (WHO), the number of new antibiotics in development is insufficient to relieve the growing threat of antimicrobial resistance. Few new antibiotics threaten worldwide efforts to contain drug-resistant infections [1]. WHO warns against the urgent need for new antibiotics to tackle antimicrobial resistance and it advised more investments needed in basic science, drug discovery and clinical development. As most agents currently in the pipeline are modifications of existing antibiotic classes and would not provide long-term solutions to antimicrobial resistance [2–4]. This problem needs urgent solutions to beat the multidrug-resistant pathogens.

Usually, research focuses on bacterial isolates that can grow in vitro and thus lack the 99% of bacteria that remain uncultivated [5, 6] or even harbor cryptic gene clusters [6, 7]. Metagenomics has developed as an alternative approach to conventional microbial screening that allows comprehensive screening of microbial genomes in their natural environments. A metagenome is a culture-independent, molecular method to freshly analyze the microorganism inhabiting in a certain environment [8]. It signifies a snapshot of a microbial population at a particular time when its DNA is extracted [9]. Metagenomics enables us to discover in less time and with higher precision new genes and proteins or even the full genomes of non-cultivable organisms than traditional microbiology or molecular methods [9]. Several

studies have recently adopted the screening approach of metagenomic libraries to assess the diversity of biosynthetic gene clusters and thus, as a tool for drug discovery [10].

The screening of large biosynthetic gene libraries has been shown to detect numerous potentially interesting clones [11, 12]. Metagenomics is supported by the integration of computational methods including BLAST algorithms to obtain a list of related hits with a certain annotation which can also be used to exploit taxonomic information and metabolic potential [13] [9].

The synthesis of bioactive peptides produced by micro-organisms through the ribosomal and non ribosomal mechanisms is well known. The synthetic route of non-ribosomal peptides is an alternative means of producing highly specialized polypeptides [14]. Numerous non-ribosomal peptides create various secondary metabolites such as antibiotics, antifungals, toxins, anticancer drugs, siderophores, and immunosuppressants [15, 16]. Non-ribosomal peptide synthetases (NRPSs) are modular mega-synthases that catalyze non-ribosomal peptide assembly from protein and non-protein amino acids [17, 18]. NRPSs consisting of an adenylation (A) domain, condensation (C) domain, and a peptidyl carrier protein (PCP) as a minimum core structure [19, 20]. Criteria such as composition, number, and arrangement of elongation modules in NRPS system, commands the mass and chemical structure of the produced natural compounds.

Several studies confirmed that mutations in NRPS genes and phylogenetically far domains in the modular organization of NRPS lead to a functional complex that can generate new bioactive molecules [21–23]. Additionally, the DNA sequence of the NRPS gene cluster could be used to predict the chemical nature of peptide, consequently, it correlates to the chemical nature and the biological function of the compounds produced via the NRPS biological system [18, 24]. Therefore, it is crucial to study the phylogenic insight of NRPS in the potential actinomycete that would provide new opportunities for drug discovery [21, 25, 26].

The present study aims to screen the soil fosmid metagenomic library to target biosynthetic pathways. Based on the evidence, fosmid libraries evaluate the diversity of biosynthetic gene clusters and help in finding several new bioactive compounds [27, 28]. NRPS PCR assay and DNA sequencing were used to capture NRPS biosynthetic gene clusters from the environmental metagenomic DNA library recovered from different sites in Cuba [28, 29]. Further, molecular and bioinformatics analysis of the positive NRPS hits retrieved from the metagenomic library is considered as a preliminary step for the manipulation of these genes in heterologous hosts [11, 30].

Materials And Methods

The metagenomics DNA source and routine cultivation

Thirty-one plate (96-well) of metagenomic fosmid library constructed from Cuban soil were kindly provided by Prof. Elizabeth Wellington, The University of Warwick. The Cuban library included 2976 clones, with an average insert size of approximately 35 kb, for a total library size of 104 Mb. All plates

were recovered in replica using fresh Luria-Bertani (LB) broth media containing chloramphenicol (35 mg/ml). An aliquot of 150 µl LB broth media containing chloramphenicol was placed in each well of the plates. An amount of 15 µl metagenomic DNA was transferred from each well individually of old plates into new plates with fresh LB broth media using an electronic micropipette. Each plate was covered with a permeable sheet and incubated at 37°C for 24 hours. All plates were held at 4^0C

Extraction of environmental DNA from E. coli **clones in metagenomics libraries**

Using a multichannel pipette, a volume of 150 µl LB broth containing 12.5 mg/ml of chloramphenicol was added to each well of (96 well) culture plates (Becton Dickinson Labware). Every well was inoculated with 15 µl of *E. coli* clones. Plates were then tapped and incubated at 37°C for 24 hrs. An amount of 10 µl from every 96 wells was transferred into a 500 µl Eppendorf tube using a multichannel pipette. Centrifugation of each sample was carried at 13,000 xg for 5 min. The supernatants were discarded, and the fosmid-bearing *E. coli* pellets were collected using the GeneJET plasmid miniprep package(QIAGEN). All steps are performed following the manufacturer's instructions and fosmid DNA was collected for PCR assay.

NRPS PCR assay and screening of the fosmid metagenomics library

The environmental DNA metagenomic library, of about 104 Mb in size, was screened using Polymerase chain reaction (PCR) for identification of NRPS clusters with ADEdom5 and ADEdom3 primer pairs obtained from (Sigma Company, Egypt) as shown in Table 1. PCR mixture included 12.5µl PCR Master Mix (Promega, Madison, WI, USA), 6.25 µl of Distilled water, 1.25µl DMSO, 2µl of (0.8µM) of each primer in 25µl total volume, and 1 µl of (0.1 µM) extracted fosmid DNA Template. The PCR program was set as follows: (5 min at 95°C followed by 40 cycles of 1 min at 95°C, 1 min at 60°C and 1.5 min at 72°C, followed by a final extension of 10 min at 72°C) [45–46]. Five microliters of every PCR product and 1Kb ladder (Fermentus) were detected using agarose gel electrophoresis for 30 min at 90 V. *Streptomyces coelicolor* was positive control in all PCR re-actions as it contains NRPS genes. The gel was stained with 50mg/ml of ethidium bromide and images of DNA bands of the predictable size was recorded using UV transilluminator (Hercules, CA).

Table 1
Primer pair used for amplifying NRPS genes obtained from fosmid metagenomic library.

Primers	Sequence (5'_'3)	Fragment (bp)	Reference
ADEdom5	5'-ACS GGC NNN CCS AAG GGC GT-3'	450	[45–46]
ADEdom3	5'-CTC SGT SGG SCC GTA-3'	450	[45–46]

Purification And Sequencing Of Pcr Products

Bands of NRPS gene fragments recovered from four fosmids *E. coli* clones were purified from 1% agarose gel using the QIAquick Gel Extraction Kit (QIAGEN; Venlo, Netherlands). An amount of fifty microliters of each PCR sample were added to the 125 µl PB buffer according to the manufacturer's instructions. Then, ten microliters of 3 M sodium acetate (pH 5.0) were included until the mixture became yellow. Samples were transferred to a QIAquick column, and centrifuged at 17, 900 xg for 60 sec, then the flow was removed. QIAquick columns were washed with 0.75 ml BE Buffer and centrifugated for 60 sec at 17, 900 xg. Each QIAquick column was set in a clean 1.5 ml microcentrifuge tube to elute the DNA, then 50 µl of BE Buffer (10 mM Tris·Cl, pH 8.5) was carefully added to the center of the QIAquick membrane. The column was then centrifuged at 17, 900 xg for 60 sec. Storage of the purified samples remained in a deep freezer at -20^{0}C and ready for sequencing. The sequencing of purified PCR products of NRPS gene fragments was conducted at GATC Biotech AG, Cologne, Germany [52]. All sequenced amplicons were placed under GenBank accession numbers.

Comparative Analyses Of The Obtained Gene Sequences Via Blast

Sequenced amplicons of selected strains were transformed into amino acid sequences to define open reading frames (ORF) using the ORF finder server (https://www.ncbi.nlm.nih.gov/orffinder/). Comparative analyses of the deduced amino acid sequences were conducted against corresponding non-ribosomal peptide domain sequences in the GenBank database using the BLASTP algorithm (https://blast.ncbi.nlm.nih.gov/Blast). Multiple alignments of NRPS amino acid sequences against similar sequences were constructed using Bioedit software to represent the patterns of amino acids with similar descent or shared functional constraints [33–38].

Neighbor-joining Phylogenetic Tree Construction

Neighbor-joining Phylogenetic tree were built using translated nucleotides of NRPS clones versus the related amino acid sequences accessible in the NCBI database using BLASTP. Multiple sequence alignment of amino acid sequences was performed using the CLUSTAL W program [31] and Neighbour-Joining method within the Molecular Evolutionary Genetics Analysis (MEGA) software version 6. 0 [32–53]. The numbers at each branch's node correspond to a percentage bootstrap value based on 1000 replicates. The scale bar showed the nucleotide sequence dissimilarity.

Downstream Analysis Of The Defined Amino Acid Sequences

A comprehensive overview of NRPS sequence domains was predicted via the NRPS Predictive BLAST web server available at (http://nrps.igs.umaryland.edu/blast.html). This web server uses Hidden Markov

Model (HMM) to predict the identity of every domain in NRPS gene sequences based on a statistical representation of protein groups with similar sequences and hence the functional similarity. Additionally, substrates specificity of each NRPS clone was expected using Non-Ribosomal Peptide Synthase Substrate Predictor based on the NRPSsp database which is available at (http://www.nrpssp.com/execute.php) [54]. Additionally, the PDBsum pictorial database in Protein Data Bank (PDB) was also used to analyze the NRPS sequence against the sequences of all proteins in the PDB. The top hits are listed with links to their PDBsum pages showing molecules that make up the structure including protein chains, DNA, ligands, and metal ions (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) [55].

Results

PCR assay detection of NRPS gene cluster in fosmid metagenomic library

NRPS PCR assay screening of 2976 clones of the soil metagenomic library retrieved a total of 17 positive clones thus putatively harboring NRPS biosynthetic pathway as in Figure 1. Consequently, the hit rate can be calculated as the total DNA captured divided by the number of positive hits (17 hits), therefore it is 1 in 6 Mb for the soil metagenomic library. This expects that the NRPS assay can detect an average greater than one gene cluster per 1.4 genomes, based on that the average *E. coli* genome is 4.6 Mb in size. The sequenced clones primarily had similarity to sequences found in Gram-negative bacteria, *Proteobacteria* including the genus *Delftia*. Selected sequenced clone amplicons can be viewed under GenBank accession numbers (MT538186, MT538187, MT538188, MT538189)

Bioinformatic analysis DNA sequences obtained from E. coli **fosmid clones**

Bioinformatics investigations of ORF amino acid sequences of *E. coli* clones derived from metagenomic libraries revealed significant homology to the NRPS gene available in NCBI data base. Clones no. 15cd30 and 15cd34 showed (100%) blast identity with NRPS genes of *Delftia tsuruhatensis* and *Delftia acidovorans*, respectively. While NRPS gene sequences of clone 15cd35 and clone 15cd37 exhibited recognizable blast identity (98.82%) and (95.24%) with corresponding NRPS gene fragments of *Delftia acidovorans*, and all sequences were not identical to each other as shown in Table 2.

Table 2
Results of BLAST identity of the positive clones retrieved from soil metagenomic library using the BLASTP algorithm.

Primer pairs	ORF	Size of product (aa)	Characteristics of homolog (s)				
			Protein ID	Deduced role	BLAST Identity	Origin	Accession no. of the sequenced amplicon
15cd30	ORF5	127 aa	2347175	non-ribosomal peptide synthetase	100.00%	*Delftia tsuruhatensis*	MT538186
15cd34	ORF6	127 aa	2347190	AMP-binding protein in NRPS cluster	100%	*Delftia acidovorans*	MT538187
15cd35	ORF5	175 aa	2347562	AMP-binding protein	98.82%	*Delftia acidovorans*	MT538188
15cd37	ORF4	90 aa	2347577	AMP-binding protein	95.24%	*Delftia acidovorans*	MT538189

Phylogenetic Analysis Of The Detected Nrps Gene Fragments

Phylogenetic analyses were carried to locate metagenomic NRPS gene fragments inside the phylogenetic tree built with the NRPS published sequences on Genbank. Clone 15cd30 was clustered with *Delftia tsuruhatensis* (WP 082254717), *Delftia tsuruhatensis* (WP 070080972), *Delftia* sp. BR1 (WP 151019177), and *Delftia* sp. GW456-R20 (WP 063325845) with (bootstrap value, 100) as shown in Figure 2. Additionally, amino acid sequence multiple alignment pattern of NRPS clone no. 15cd30 was identical to corresponding amino acid sequences of *Delftia tsuruhatensis* (WP 082254717), *Delftia tsuruhatensis* (WP 070080972), *Delftia* sp. GW4 (WP063325845.1) at the core region of the NRPS domain, while at the terminal several mismatches were detected as presented in Figure 3.

The closest match to the NRPS gene of clone no. 15cd34 is AMP-binding protein (within the NRPS gene cluster) was *Delftia tsuruhatensis* (WP 154834667.1). It is clustered with members of *Pandoraea* spp. and *Achromobacter* spp. with bootstrap value (100%) as shown in Figure 4. Amino acid sequence multiple alignment pattern of NRPS clone no. 15cd34 was identical to corresponding amino acid sequences of *Delftia tsuruhatensis* (WP 154834667.1) except for the terminal region as presented in Figure 5.

Phylogenetic analysis confirmed that clone 15cd35 and clone 15cd37 were clustered in a separate clade from AMP binding protein sequences (within NRPS) that belonged to members of genus *Delftia* spp. within the b-*Proteobacteria*. *Delftia acidovorans* (WP 099752190.1), and *Delftia* sp. RIT313

(WP043825786.1) was identified as the nearest phylogenetic relative to both clones with a low bootstrap value (54%) as shown in Figure 4. Amino acid sequence multiple alignment pattern of NRPS clone no. 15cd35 shows several amino acid substitutions to corresponding conserved regions of amino acid sequences of all corresponding *Delftia* spp. in the middle core region as presented in Figure 5.

Prediction of NRPS clones was conducted using PKS/NRPS analysis web site. HMMs hits of NRPS positive clones retrieved from the soil metagenomic library resolve diversity in NRPS domains as shown in Table 3. The substrate of NRPS clones retrieved from soil metagenomic library was predicted by NRPSpredictor2 as presented in Table 4. The substrate for the NRPS adenylation domains of clones no. 15cd30 and 15cd34 appear to be phenylalanine and alanine, respectively. No substrates were recorded to other NRPS domains of clones 15cd35 and 15cd37 as presented in Table 4. PDBsum Bioinformatics tool demonstrated hits of all clones against all proteins sequences and related 3D structures deposited in the Protein Data Bank as shown in Table 5. The bioinformatic analysis confirmed the presence of diverse NRPS domains in all sequences retrieved from the metagenomic library.

Table 3
List of top HMMs hits of NRPS positive clones retrieved from soil metagenomic library using NRPS analysis web site.

Clones	Protein Identifier	Domain	Coordinates of the hit		Hit probability score
15cd30	ORF5	ER Domain	1	127	7.2
15cd34	ORF6	DH Domain	1	92	7.1
15cd35	ORF5	AT Domain	1	131	7.4
15cd37	ORF4	KR Domain	17	91	9.3

*Several domains of NRPS gene cluster with defined functions (ER): Enoylreductase (DH): Dehydratase (AT): Acyltransferase (KR) Ketoreductase.

Table 4
Results of substrates that bind to a given NRPS positive clones retrieved from soil metagenomic library Using Non-Ribosomal Peptide Synthase Substrate Predictor based on HMM predictor.

Clones	Protein Identifier	Adenylation domain start position	Adenylation domain end position	Substrate Name	Score
15cd30	ORF5	0	124	Phenylalanine	82.7
15cd34	ORF6	0	120	Alanine	128.2
15cd35	ORF5	-	-	ND	-
15cd37	ORF4	-	-	ND	-
*(ND): Not detected.					

Table 5
Results of Top hits of positive clones retrieved from soil metagenomic library against all protein sequences and related 3D structures deposited in the Protein Data Bank using the PDBsum database.

Clones	ORF	PDB code	Amino acid overlap	z-score	Ligands	Protein name	Organism
15cd30	ORF5	5n9x	129	215.2	THR, ATP, 8QN	adenylation domain thr1	*Streptomyces* sp. Oh-5093
15cd34	ORF6	3fce	118	382.8	ATP	d-alanyl carrier protein ligase	*Bacillus cereus*
15cd35	ORF5	4oae	66	118.7	SO4, CLM, EDO	gnat superfamily acetyltransferase	*Pseudomonas aeruginosa*
15cd37	ORF4	4oae	66	130.9	SO4, CLM, EDO	gnat superfamily acetyltransferase	*Pseudomonas aeruginosa*

Discussion

Our environment still contains lots of undiscovered micro-organisms that can support potential drug discovery [5]. In this study, the soil metagenomic library was screened to discover the biosynthetic pathways in uncultured microorganisms. Similarly, several studies have involved in the mining of metagenomic libraries to determine the variety of biosynthetic gene clusters [10]. Screening of large biosynthetic gene libraries detected abundant remarkable clones with biosynthetic potential [11, 12, 28]. In this study, the NRPS PCR assay captured successfully several NRPS gene fragments in *E. coli* clones that matched corresponding NRPS genes on the GenBank. Those positive hits indicate the existence of NRPS biosynthetic genes within the soil metagenome. A similar PCR screening approach was used by other research groups to capture NRPS clones from soil metagenomes [28]. These findings are in agreement with other study that confirmed capturing NRPS genes in Cuban soil metagenome using NRPS PCR using different primer pair [28].

In the herein study, NRPS PCR assay recovered clones that had similarity to sequences belonging to the *Delftia* genus, *Proteobacteria* phyla, suggesting that they are the dominant phyla and the tested soil is abundant in metabolite-producing bacteria. Several reports confirmed the biotechnological potential of the *Delftia* genus. For instance, a recent study has identified Delftibactin A (NRP) isolated from novel environmental *Delftia* spp. with potent antimicrobial activity against methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus*, *Acinetobacter baumannii*, and *Klebsiella pneumonia* [34]. Furthermore, the biotechnological perspective of *Delftia* sp. JD2 was previously confirmed using a genomic approach [35]. In the same context, functional NRPS genes were identified in *Delftia tsuruhatensis* MTQ3 for bacteriocins and siderophore production [36].

The level of diversity in the metagenomic library is low since all positive hits related only to the genus of *Delftia*. Reversely, another study declared high bacterial biodiversity in Cuban soil with abundant enzymatic activity [28]. Due to the small number of clones selected for the sequencing phase, and a huge sequencing work is needed to identify all the gene clusters in the metagenomic library. Additionally, higher biodiversity would be achieved if a range of primer sets were used. Similarly, Amos et al suggested that using different primers focusing on different groups of microorganisms would increase the biodiversity of the positive hits [28]. Several related NRPS sequences were isolated from different sites suggesting a widespread environmental distribution of bacteria harboring NRPS genes. Similar results were recorded by a research study on the metagenomic library extracted from the soil in Cuba [28].

In the case of natural product science, phylogenetic relationships are highly informative in the design and function of the genes involved in secondary metabolite biosynthesis. Non ribosomal peptide synthetases provide a model in which individual domain phylogenies exhibit various predictive capabilities, determining features of substrate specificity correlated to the final metabolic product [37]. In this study, sequencing of selected NRPS clones showed that they belonged to different regions in NRPS gene clusters, thus they were separated in different phylogenetic trees. It also may evidence for the evolutionary process of NRPS genes [38]. Phylogenetic relationships showed that clones no. 15cd30 and 15cd34 were clustered together as a sister group to the genus *Delftia*, supported by high blast identity values and the tree was in agreement with a previous phylogenetic analysis of the group [28]. Clone no 15cd30 and 15cd34 shared amino acid similarity with strains of *Delftia tsuruhatensis* and *Delftia acidovorans*, suggesting similar secondary metabolites production. However, they even have some mismatches at the terminal region indicating a distinct scope of bioactive molecules production. In case of clone 15cd35 and clone 15cd37, NRPS gene sequences of exhibited noticeable similarity values to AMP (A) domain sequences of belonged to members of genus *Delftia* spp. Though they still showed several differences in amino acid sequences and low bootstrap value (54%) with the closest phylogenetic neighbors. Thus, the metagenomic NRPS genes are predicted to produce distinct non-ribosomal peptides or suggesting many of the clones recovered came from undiscovered NRP pathways. These changes in amino acid sequences can be due to conservative mutation or radical substitution [39, 40]. Especially, the mutations are detected in the protein interior which may be a sign of structural constraints [41]. A similar study confirmed that multiple alignments and the phylogenic tree of amino acid sequences of NRPS genes of *Streptomyces* sp. BDUSMP 02 reveals the potential to produce a new type of antibacterial compounds belonging to the NRPS type [42]. Other studies supported using multiple alignments and phylogenetic trees approach in identifying conserved sequence regions and establishing evolutionary relationships [38]. Similarly, a study emphasized that sequence analysis of the 21 most active improved variants revealed that each contained between one and three changes to their primary amino acid sequence, suggesting that minimal sequence variation was able effect dramatic improvements in NRPS domain function [43]. Fischbach et al. (2007) declared that NRPS is considered as enzymes that assemble the key skeletons of natural products, so any mutation of these genes is expected to give the clearest impact on the metabolite pattern and their functions. In agreement with our results, a

substitutional mutation in NRPS genes can lead to ≈10-fold improvements in enzyme activity and can trigger the creation of new derivatives of NRP antibiotics [23].

In this study, bioinformatics tools such as NRPS predictor tool, NRPS predictive blast webserver, and PDBsum database were used to identify NRPS domains at both genome/proteome level and their substrate suggesting different structures with relevant biological activities. Similarly, other studies support the use of the NRPS predictor tool for reliable prediction of adenylation domain (NRPS) specificities [44, 45]. None of the predictive methods could infer any substrate specificity for 15cd35 and 15cd37 NRPS sequences, suggesting completely new types of specificity and PDBsum database showed their resemblance to acetyltransferase domains still with low z- score. This may be due to mutations from corresponding sequences recorded in the multiple alignments. Different substrate specificity using bioinformatic tools refers to variation in the NRPS clones, thus different non-ribosomal peptide biosynthesized. Similar results recorded that substrate specificity is important in determining the final bioactive product such as the A domain TycB_m3 activates l-tryptophan and phenylalanine for tyrocidine biosynthesis [46, 47]. Additionally, A domain of the barbamide biosynthetic gene cluster activates 100% specificity for leucine and valine, and 80% for trichloroleucine [48].

Finally, this study emphasizes the presence of biosynthetic NRPS genes in the soil metagenomic library. The molecular, phylogenetic, and bioinformatic analysis confirmed that mainly 15cd35 and 15cd37 are distinct clones harboring biosynthetic potential with undetected substrate specificity and thus can produce improved yield or even new antibiotics. Similar results confirmed that directed mutation in substrate specificity code within the A-domain of NRPS genes can produce a high yield of certain antibiotic or even a novel antimicrobial compound [49–52]. This study supports the methodology of using PCR assay for screening soil metagenomes as a tool for drug discovery. Further genetic manipulation of NRPS clones will provide a positive impact on the pharmaceutical sector and consequently the health sector.

Conclusion

In conclusion, PCR assays for screening soil metagenomic libraries are a useful approach to explore the biosynthetic potential in uncultivated bacteria. Molecular and bioinformatic methods confirmed that all contained clones belong to *Proteobacteria*, and code for an NRPS biosynthetic pathway. This research highlights the richness of Cuban soil with diverse biosynthetic gene clusters. These diverse biosynthetic clusters can be exploited in genetic engineering, they are more likely to produce new antibiotics. This is expected to have a positive effect on the pharmaceutical industry soon.

Abbreviations

NRPS
non ribosomal peptide synthase
PKS

polyketide synthase
PDB
Protein Data Bank
HMM
Hidden Markov Model
AMP
Adenosine monophosphate.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors agreed to publish this manuscript.

Availability of Data and Materials

All data is available in this study.

Competing interests

All authors declare that they have no conflicts of interest to report regarding the present study.

Funding

Not applicable.

Author's Contribution

Data curation, DHA, WMN; Funding acquisition, DHA; Investigation, DHA; Methodology, DHA; Resources, DHA, WMN; Software, DHA; Supervision, WMN, RRM, AE and Writing – original draft, DHA; Writing – review & editing, WMN, RRM, and AE.

Acknowledgment

Thanks to the members of lab C123 in the University of Warwick.

References

1. World Health Organization (2020) Lack of new antibiotics threatens global efforts to contain drug-resistant infections. New release Geneva
2. Kmietowicz Z (2017) Few novel antibiotics in the pipeline, WHO warns, BMJ: British Medical Journal (Online). 358. https://doi.org/ 10.1136/bmj.j4339
3. Falagas ME, Lourida P, Poulikakos P, Rafailidis PI, Tansarli GS (2013) Antibiotic treatment of infections due to carbapenem-resistant Enterobacteriaceae: systematic evaluation of the available evidence", Antimicrobil agen chem. AAC 01222–01213. https://doi.org/10.1128/AAC.01222-13
4. Khanna M, Solanki R, Lal R (2011) Selective isolation of rare actinomycetes producing novel antimicrobial compounds. Int J Adv Biotechnol Res 2(3):357–375
5. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68(9):4301–4306. https://doi.org/10.1128/AEM.68.9.4301-4306.2002
6. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Mol Biol Rev 59(1):143–169. https://doi.org/10.1128/mr.59.1.143-169.1995
7. Bentley S, Chater K, Cerdeno-Tarraga A, Challis G, Thomson N, James K, Harris D, Quail M, Kieser H, Harper D (2002) O= Neil, S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, and Hopwood, DA 141-147
8. Hugenholtz P, Tyson GW (2008) Metagenomics, Nature 455(7212):481–483. https://doi.org/10.1038/455481a
9. Escobar-Zepeda A, Vera-Ponce de Leon A, Sanchez-Flores A (2015) The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics". Front Genet. 6 348 https://doi.org/10.3389/fgene.2015.00348
10. PD S, Handelsman J (2003) Biotechnological prospects from metagenomics",Curr Opin Biotechnol. 14303–310
11. Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K, Bertrand H, Vogel TM, Perriere G, Simonet P (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol 70(9):5522–5527. https://doi.org/10.1128/AEM.70.9.5522-5527.2004
12. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69(1):49–55. https://doi.org/10.1128/AEM.69.1.49-55.2003
13. Carr R, Borenstein E (2014) Comparative analysis of functional metagenomic annotation and the mappability of short reads, PloS one. 9(8). https://doi.org/10.1371/journal.pone.0105776

14. Marahiel MA, Stachelhaus T, Mootz HD (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem rev 97(7):2651–2674. https://doi.org/10.1021/cr960029e
15. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proceed Nat Acad Sci 111(25):9259–9264. https://doi.org/10.1073/pnas.1401734111
16. Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster AL, Magarvey NA (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat chem biol 12(12):1007–1014. https://doi.org/10.1038/nchembio.2188
17. Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. Annu Rev Microbiol 58:453–488. https://doi.org/10.1146/annurev.micro.58.030603.123615
18. Sieber SA, Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem rev 105(2):715–738. https://doi.org/10.1021/cr0301191
19. Cane DE, Walsh CT, Khosla C (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations, Science. 282(5386), 63-68. https://doi.org/.282.5386.63
20. Cane DE, Walsh CT (1999) The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem & biol 6(12):R319–R325. https://doi.org/10.1016/S1074-5521(00)80001-0
21. Amin DH, Tolba S, Abolmaaty A, Abdallah NA, Wellington EM (2017) Phylogenetic and Antimicrobial Characteristics of a Novel Streptomyces sp. Ru87 Isolated from Egyptian Soil. Int J Curr Microbiol App Sci 6(8):2524–2541. http://dx.doi.org/10.20546/ijcmas
22. Amin DH, Abolmaaty A, Tolba S, Abdallah NA, Wellington EM (2017) Phylogenic Characteristics of a Unique Antagonistic Micromonospora Sp. Rc5 to S. aureus Isolated from Sinai Desert of Egypt. Cur Res Microbiol and Biotech 5(6):1295–1306. https:://doi.org/10.9734/ARRB/2018/38318
23. Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR (2007) Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proceed Nat Acad Sci 104(29):11951–11956. https://doi.org/10.1073/pnas.0705348104
24. Komaki H, Harayama S (2006) Sequence diversity of type-II polyketide synthase genes in Streptomyces. Actinomycetologica 20(2):42–48. https://doi.org/10.1136/bmj.j4339
25. Amin DH, Borsetto C, Tolba S, Abolmaaty A, Abdallah NA, Wellington EM (2017) Phylogenic Analysis of NRPS and PKS Genes Associated with Antagonistic Micromonospora Rc5 and Streptomyces Ru87 Isolates., J Adv Biology & Biotechnology. 1–22. . doi: 10.9734/JABB/2017/37592
26. Amin DH, Abolmaaty A, Tolba S, Abdallah NA, Wellington EM (2017) Phylogenic Characteristics of a Unique Antagonistic Micromonospora Sp. Rc5 to S. aureus Isolated from Sinai Desert of Egypt. Cur Res Microbiol and Biotech 5(6):1295–1306. https:://doi.org/10.9734/ARRB/2018/38318
27. Kallifidas D, Kang HS, Brady SF (2012) Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. J Ameri Chem Soc 134(48):19552–19555. https://doi.org/10.1021/ja3093828

28. Amos GC, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA, Vallin C, Wellington EM (2015) Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. PLoS ONE 10(9):e0138327. https://doi.org/10.1371/journal.pone.0138327
29. Ayuso-Sacido A, 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. Microbiol ecol 49(1):10–24. https://doi.org/10.1007/s00248-004-0249-6
30. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69(1):49–55. https://doi.org/10.1128/AEM.69.1.49-55.2003
31. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl acids res 22(22):4673–4680. https://doi.org/10.1093/nar/22.22.4673
32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol biol evol 30(12):2725–2729. https://doi.org/10.1093/molbev/mst197
33. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series
34. Tejman-Yarden N, Robinson A, Davidov Y, Shulman A, Varvak A, Reyes F, Rahav G, Nissan I (2019) Delftibactin-A, a Non-ribosomal Peptide With Broad Antimicrobial Activity. Front in Microbiol 10(2377). https://doi.org/10.3389/fmicb.2019.02377
35. Morel MA, Iriarte A, Jara E, Musto H, Castro-Sowinski S (2016) Revealing the biotechnological potential of Delftia sp. JD2 by a genomic approach. AIMS Bioeng 3(2):156–175. https://doi.org/10.1128/jb.179.21.6843-6850.1997
36. Guo H, Yang Y, Liu K, Xu W, Gao J, Duan H, Du B, Ding Y, Wang C (2016) Comparative Genomic Analysis of Delftia tsuruhatensis MTQ3 and the Identification of Functional NRPS Genes for Siderophore Production, BioMed Res Intern. 2016 3687619. https://doi.org/10.1155/2016/3687619
37. Ziemert N, Jensen PR (2012) Phylogenetic approaches to natural product structure prediction, Meth in enzym. 517:161–182. https://doi.org/10.1016/B978-0-12-404634-4.00008-5
38. Wang L, Jiang T (1994) On the complexity of multiple sequence alignment. Journal of comput bio 1(4):337–348. https://doi.org/10.1089/cmb.1994.1.337
39. Zhang J (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes". J mol evol 50(1):56–68. https://doi.org/10.1007/s002399910007
40. Dagan TY, Graur D (2002) Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol biol evol 19(7):1022–1025. https://doi.org/10.1093/oxfordjournals.molbev.a004161
41. Sivalingam P, Muthuselvam M, Pote J, Prabakar K (2019) Phylogenetic insight of Nonribosomal peptide synthetases (NRPS) Adenylate domain in Antibacterial potential Streptomyces BDUSMP 02

isolated from Pitchavaram Mangrove. Bioinformation 15(6):412. https://dx.doi.org/10.6026%2F97320630015412

42. Owen JG, Calcott MJ, Robins KJ, Ackerley DF (2016) Generating functional recombinant NRPS enzymes in the laboratory setting via peptidyl carrier protein engineering. Cell chem biol 23(11):1395–1406. https://doi.org/10.1016/j.chembiol.2016.09.014
43. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucl acids res 33(18):5799–5808. https://doi.org/10.1093/nar/gki88
44. Agüero-Chapin G, Pérez-Machado G, Sánchez-Rodríguez A, Santos MM, Antunes A (2016) Alignment-free methods for the detection and specificity prediction of adenylation domains. Nonribosomal Peptide and Polyketide Biosynthesis. Humana Press, New York, NY, pp 253–272
45. Miller, Lipman M (1973) Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. Proceed Nat Acad Sci 70(1):190–194. https://doi.org/10.1073/pnas.70.1.190
46. Mootz HD, Marahiel MA (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. https://doi.org/10.1128/jb.179.21.6843-6850.199
47. Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH (2002) The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. Gene 296(1–2):235–247. https://doi.org/10.1016/S0378-1119(02)00860-0
48. Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem & biol 6(8):493–505. https://doi.org/10.1016/S1074-5521(99)80082-9
49. Mootz HD, Kessler N, Linne U, Eppelmann K, Schwarzer D, Marahiel MA (2002) Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes. J Amer Chem Soc 124(37):10980–10981. https://doi.org/10.1021/ja027276m
50. Symmank H, Franke P, Saenger W, Bernhard F (2002) Modification of biologically active peptides: production of a novel lipohexapeptide after engineering of Bacillus subtilis surfactin synthetase. Protein Eng 15(11):913–921. https://doi.org/10.1093/protein/15.11.913
51. Roongsawang N, Washio K, Morikawa M (2011) Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J mol sci 12(1):141–172
52. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceed nat acad sci 74(12):5463–5467. https://doi.org/10.1073/pnas.74.12.5463
53. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol biol evol 4(4):406–425. https://doi.org/10.3390/ijms12010141
54. Prieto C, García-Estrada C, Lorenzana D, Martín JF (2012) NRPSsp: non-ribosomal peptide synthase substrate predictor. Bioinformatics 28(3):426–427. https://doi.org/10.1093/bioinformatics/btr659

55. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. Tren biochem sci 22(12):488–490. https://doi.org/10.1016/S0968-0004(97)01140-7

Figures

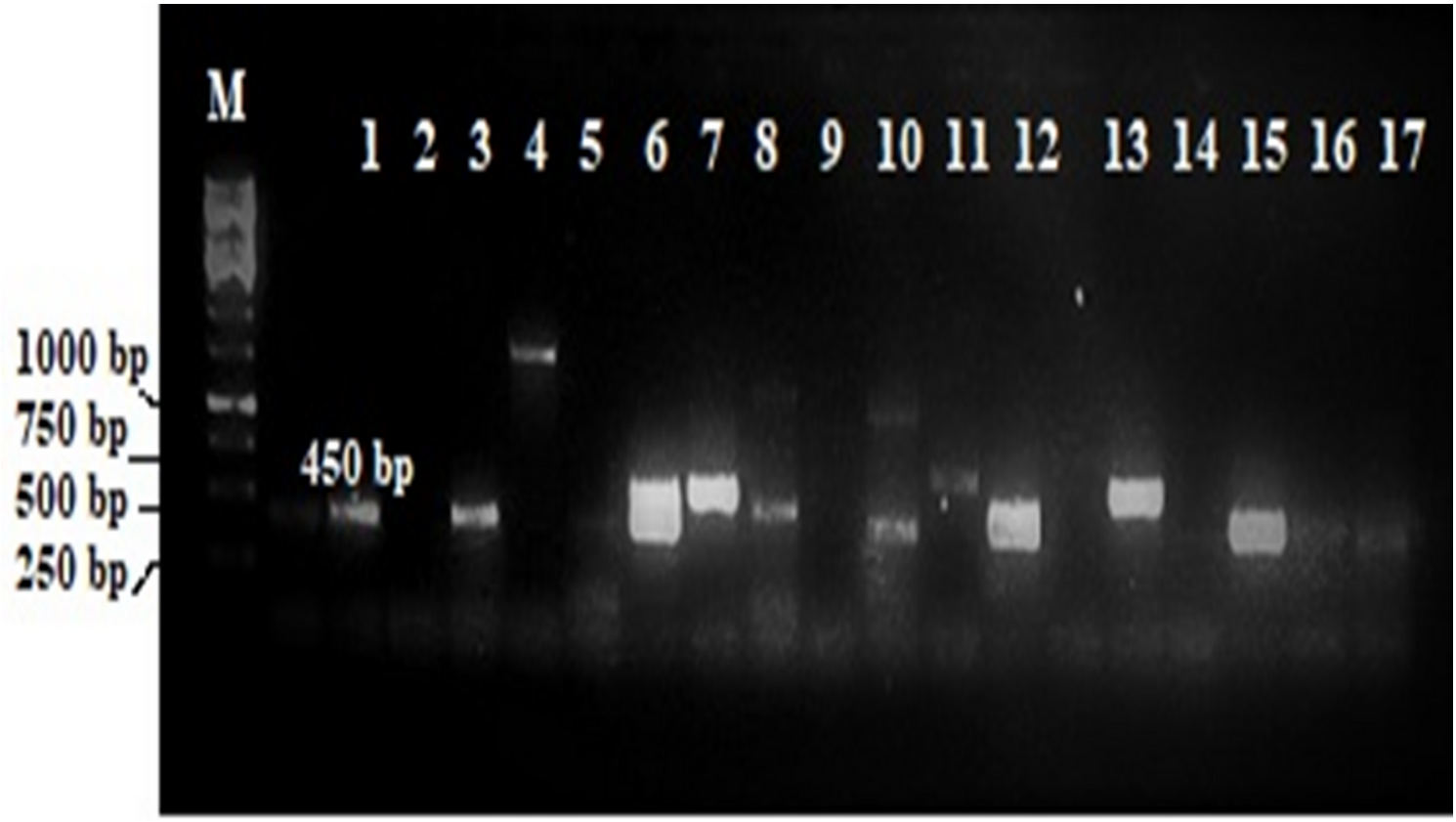

Figure 1

NRPS PCR screening of soil metagenomic fosmid library. PCR amplicons were subjected to electrophoresis in 1% agarose gels as previously described in the Materials and Methods section. Lane M: 1 kb ladder, Lane 1: control (*S. coelicolor*), Lane (3,6,8,12,15): Positive hits of 450 bp of NRPS genes isolated from *E. coli* clones of the soil metagenomic library by ADEdom3/5 primers.

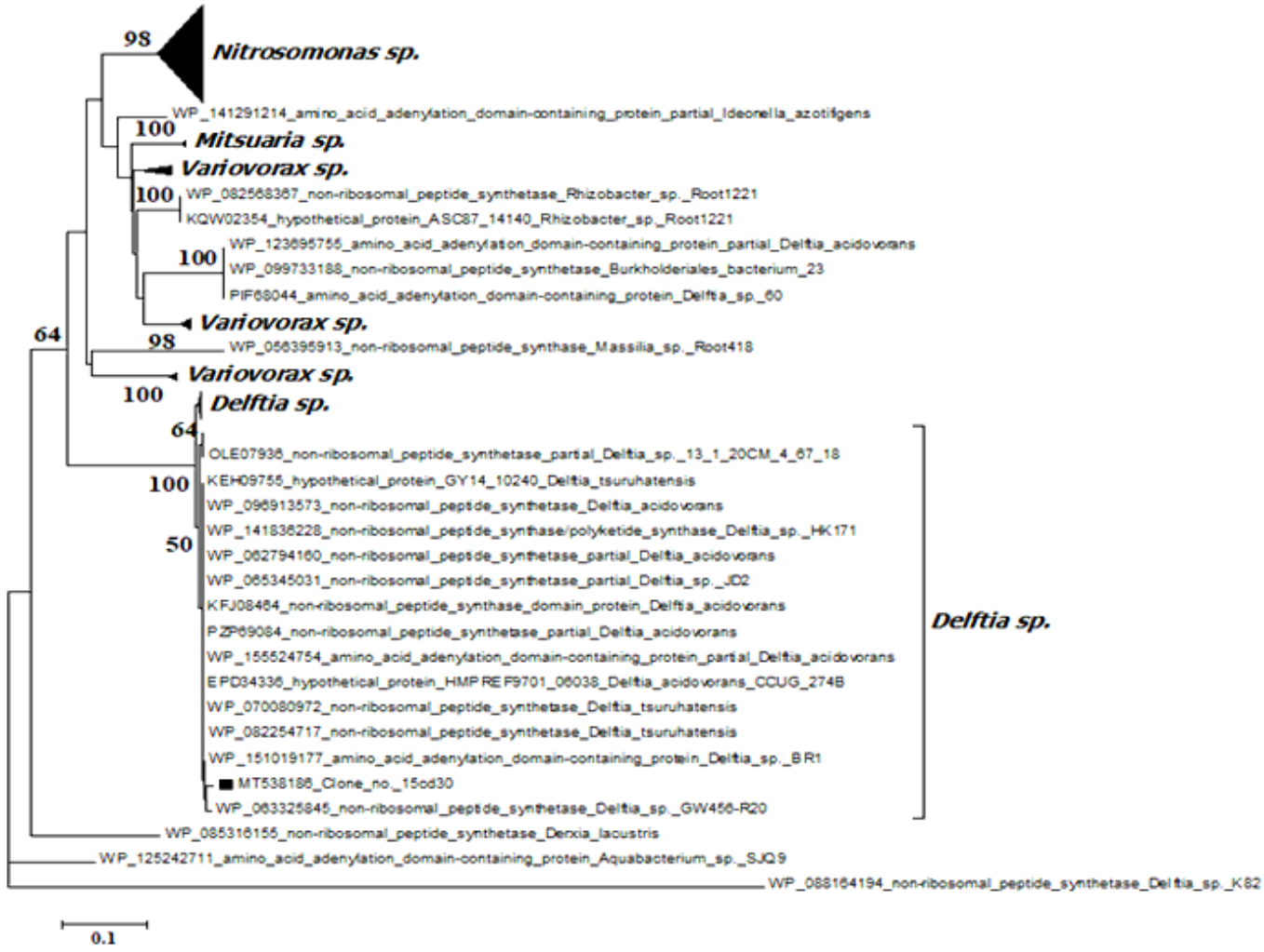

Figure 2

Phylogenetic tree based on the amino acid sequence of NRPS gene fragment of positive clone retrieved from soil metagenomic library. Multiple sequences were aligned using the CLUSTAL W program [31] against corresponding amino acid sequences. The tree was constructed using the neighbor-joining method using MEGA software version 6.0 [32]. The numbers beside the branches indicate the percentage bootstrap value of 1000 replicates. Bootstrap values greater than 50% are at the node. The scale bar indicates nucleotide sequence dissimilarity.

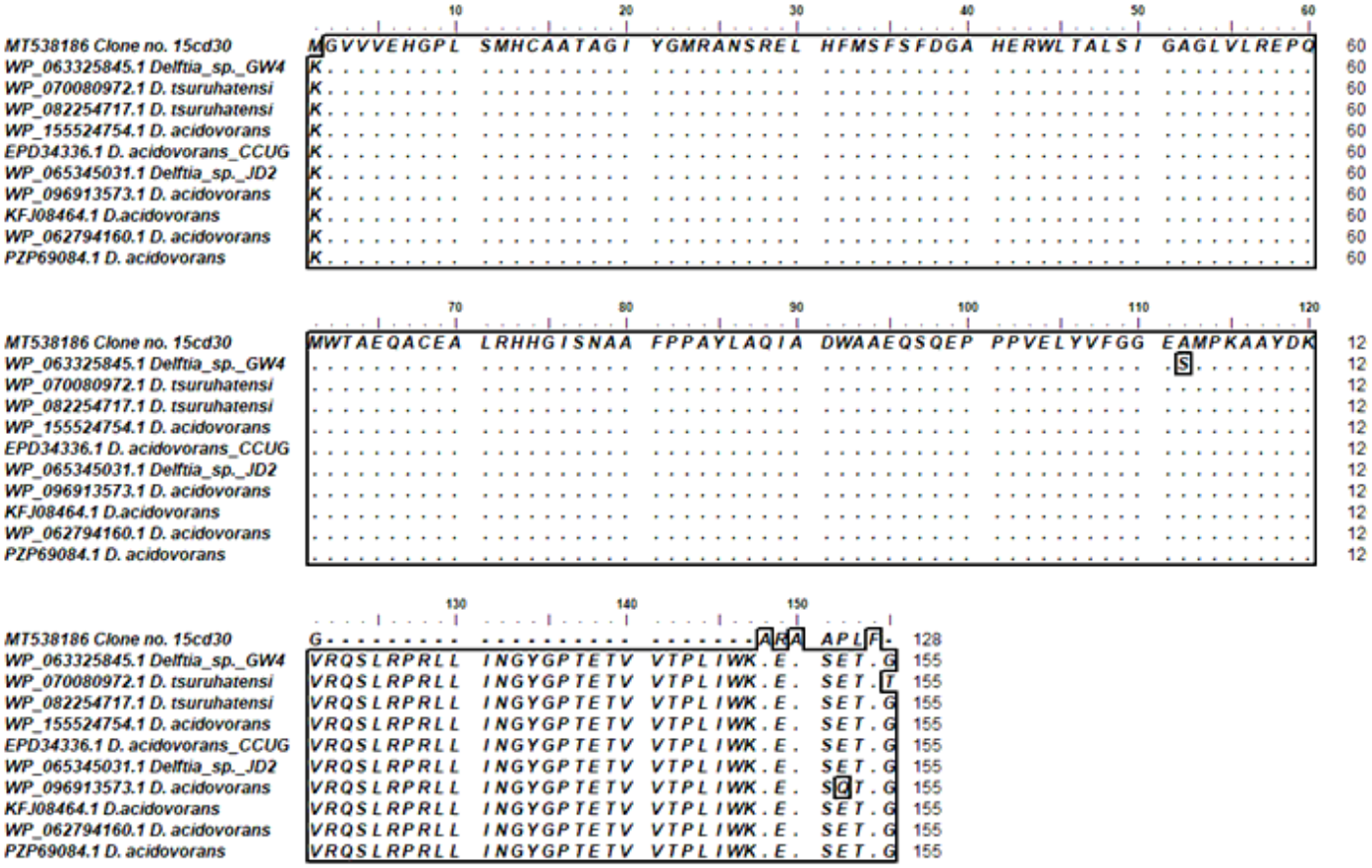

Figure 3

Amino acid sequence multiple alignment pattern for the identification of conserved motifs in NRPS positive clone retrieved from soil metagenomic library. Multiple sequences were aligned using the CLUSTAL W program[31] against corresponding amino acid sequences using Bioedit program [33]. Conservation is viewed by plotting identities to the first sequence as dots with outlining.

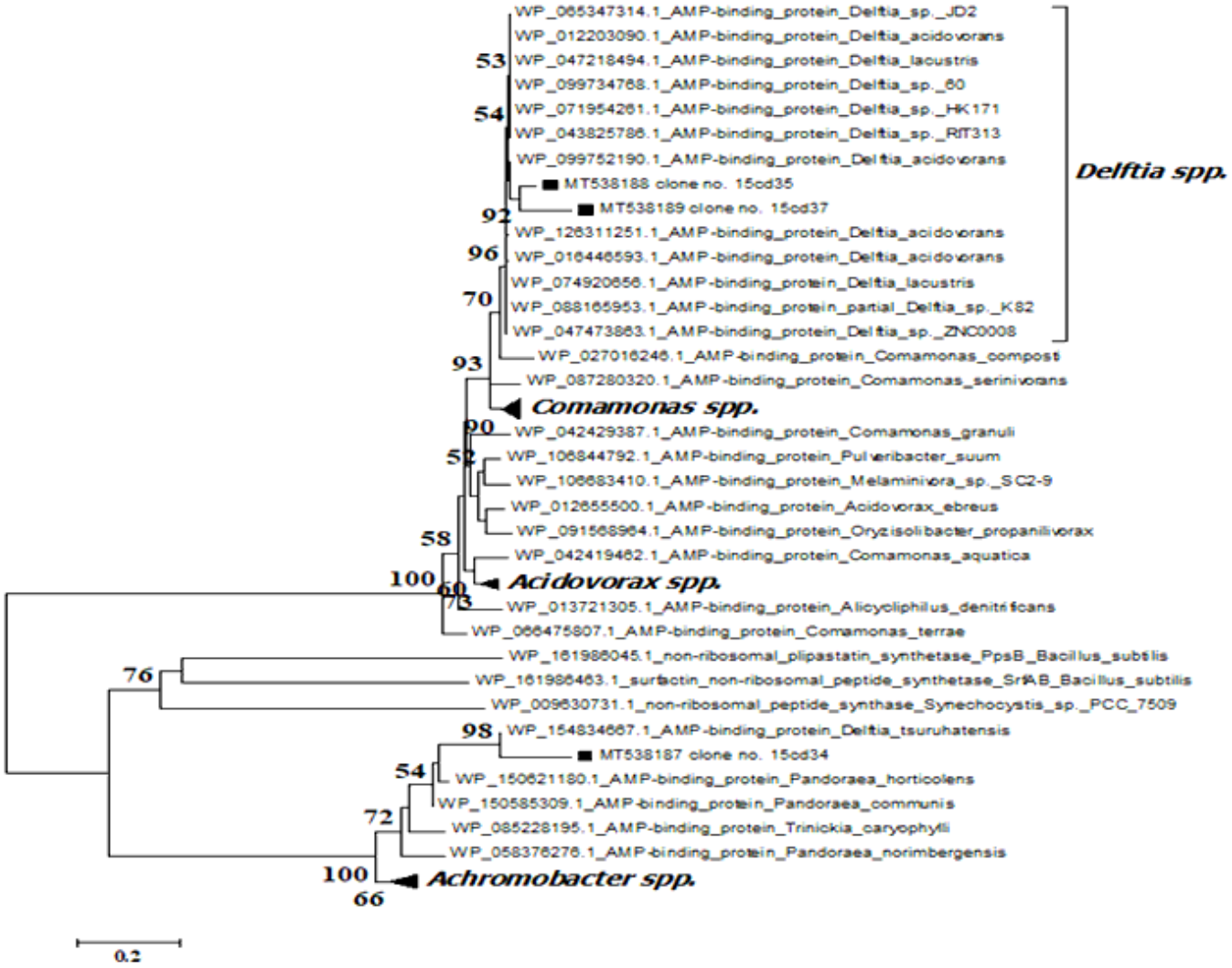

Figure 4

Phylogenetic tree based on the amino acid sequence of NRPS gene fragment of positive clones retrieved from soil metagenomic library. Multiple sequences were aligned using the CLUSTAL W program [31] against corresponding amino acid sequences. The tree was constructed using the neighbor-joining method using MEGA software version 6.0 [32] .The numbers beside the branches indicate the percentage bootstrap value of 1000 replicates. Bootstrap values greater than 50% are at the node. Scale bars indicate nucleotide substitutions per site.

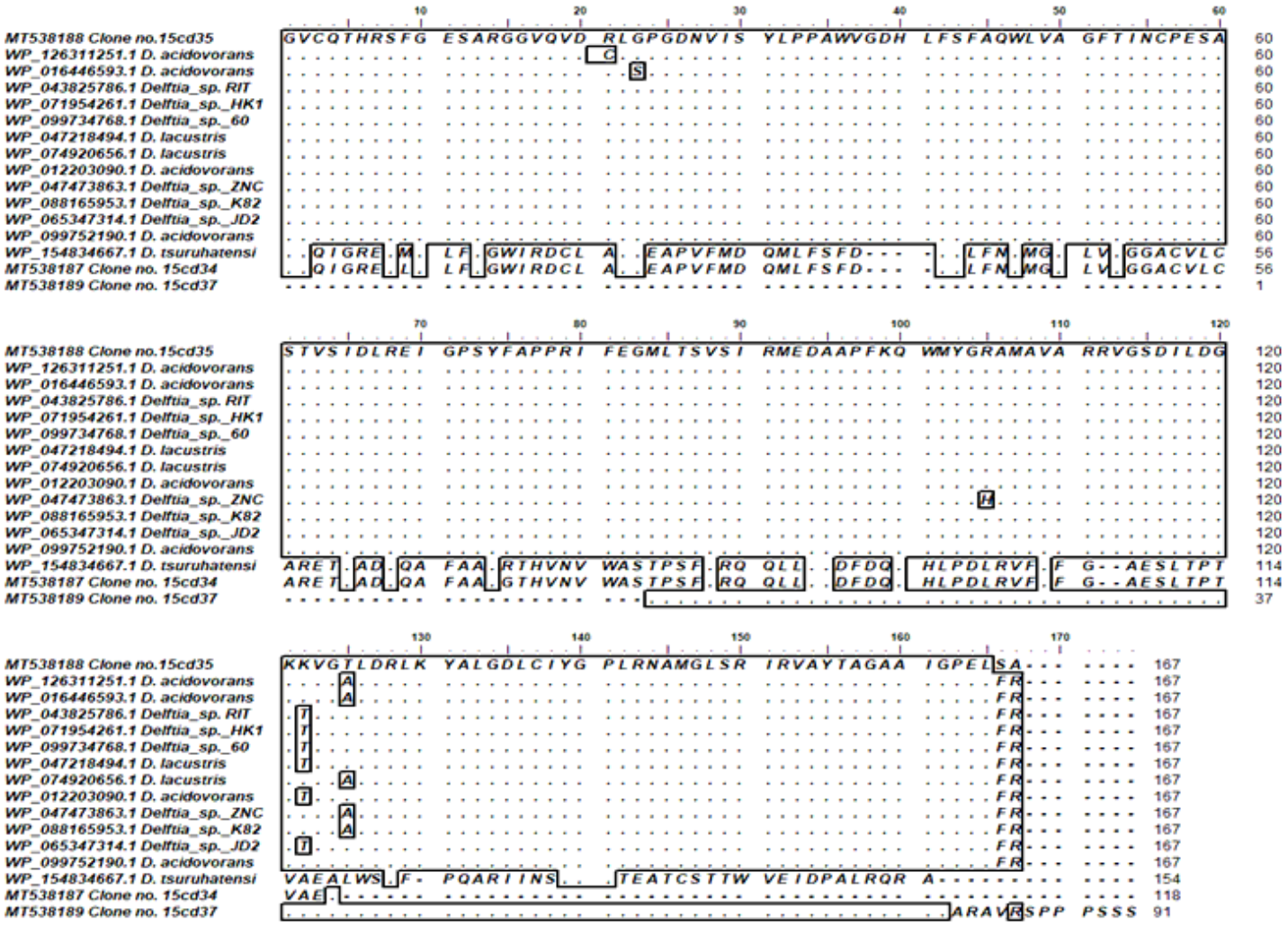

Figure 5

Amino acid sequence multiple alignment pattern for the identification of conserved motifs in NRPS positive clone retrieved from soil metagenomic library. Multiple sequences were aligned using the CLUSTAL W program [31] against corresponding amino acid sequences using Bioedit program [33]. Conservation is viewed by plotting identities to the first sequence as dots with outlining.