The University of South Bohemia in České Budějovice

Faculty of Science

Searching for the Polyenes in *Streptomyces* Associated with Bark Beetles

Master´s Thesis

Bc. Sara Khaled Youssef Sayed

Supervisor: RNDr. Alica Chroňáková, PhD.

Co-supervisor: Ana Catalina Lara Rodriguez, PhD.

České Budějovice

2023

Genome Screening of antifungal polyenic secondary metabolites

1. Research Objective:

Data mining from polyene databases was carried out to search for a specific gene marker that is essential for the synthesis of polyenes. This was accomplished by a deep understanding of the biosynthetic gene clusters (BGCs) of the polyenes produced by *Streptomyces sp*. and evolutionary related actinomycetes. The first hypothesis was that Cytochrome P450 could be used as a specific gene marker for polyene production. The main aim of this work was to evaluate the matching of the primer pair that was designed in a previous study with CYP 450 sequences of a more robust dataset of polyenes. Furthermore, an alternative hypothesis was formulated after conducting an extensive data mining to search for a specific gene marker. The hypothesis was that thioesterase type I is a good gene marker for the identification of polyene-producing actinomycetes, specifically streptomycetes.

2. Methodology:

2.1 Data mining from polyene databases

Since polyenes have a common structure essential for their functions, they have common genes which may act as gene markers (Figure 1). Data mining from polyene databases such as the National Center for Biotechnology Information (NCBI), Minimum Information about a Biosynthetic Gene cluster (MIBiG) and Database of Biosynthesis clusters Curated and Integrated (DOBISCUIT), was carried out to identify a specific gene marker required for the production of Polyenes.

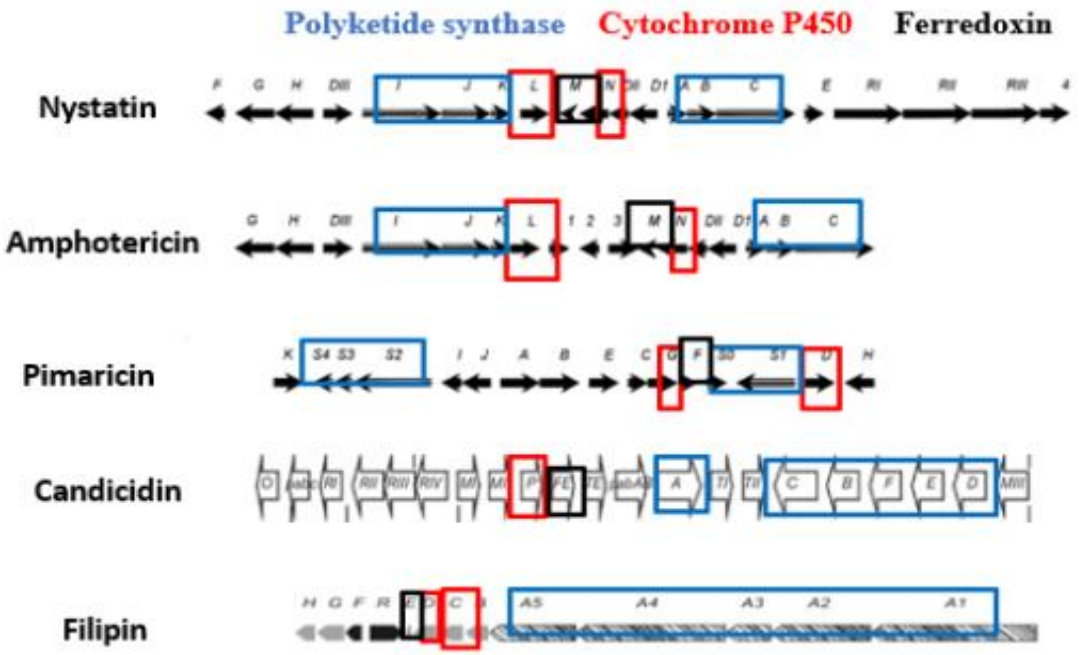


Figure 1: Illustration for the biosynthetic gene cluster of the polyenes with common genes highlighted.

- **Cytochrome P450 as a gene marker**

Based on a previous study, where a primer pair was specifically designed for cytochrome P450

and utilized for PCR-based genome screening to isolate potential polyene-producing actinomycetes strains.

Firstly, nucleotide and amino acid sequences of polyene BGCs were downloaded from publicly available databases such as DOBISCUIT (Accessed in October 2022) and MIBiG (Accessed in October 2022). Then a FASTA file was created for the amino acid sequence of CYP 450 of 6 polyenes. Multiple sequence alignment of CYP 450 using the graphical interface of Geneious version 2022.0.2 (Biomatters Ltd., Auckland, New Zealand) was done using Muscle (Version 3.5). The aligned sequences were edited manually by removing the gaps and used for phylogenetic tree construction using Molecular Evolutionary Genetics Analysis (MEGA) software version 11.0.13. Maximum likelihood statistical method and 100 bootstrap replications were selected.

- **Thioesterase type I as a gene marker**

After conducting an extensive data mining process to search for another gene marker, the alternative hypothesis was that thioesterase type I could serve as an effective gene marker for identifying polyene-producing actinomycetes, particularly streptomycetes. It was based on a previous literature that highlighted the presence of a distinct group of PKS-TE domains responsible for producing polyene macrolides. These domains exhibited relatively low sequence similarity compared to TEs of other non-polyene macrolides such as Erythromycin and Pikromycin.

Firstly, nucleotide sequences of polyene biosynthetic gene clusters were downloaded from JDB, NCBI and MIBiG. Then FASTA file was created for nucleotide sequences of the last module of type I polyketide synthase that associated with thioesterase type I, which included 26 polyenes and 2 non-polyenes (Pikromycin and Erythromycin). The sequences were checked that they are starting with start codon and have stop codon at the end. Then the sequences of the last module were uploaded on AliView software (Alignment Viewer and Editor). The sequences were translated into amino acid sequences using bacterial genetic code and translation frame 1. Then multiple sequence alignment of the last module of polyketide synthase using AliView was performed using Muscle and selecting the option "Realign everything as Translated Amino Acids". The aligned amino acid sequences were checked, edited manually and the gaps were removed. Then the amino acid sequences of ACP-TE were annotated using antiSMASH bacterial version and extracted from the whole module. The results of the alignment of amino acid sequences of the thioesterase (TE) gene were compared to the results provided by Zhou, Yucong, et al.. The Conserved region of TE for the polyenes and non-polyenes was extracted and the large gaps that occupied only by non-polyenes were removed. Then the Amino acid sequences were translated back to nucleotide sequences and used for designing the forward and reverse Primers.

2.2 Designing of primer pairs

Ten primer pairs were designed on the consensus sequences using Geneious software version R8.1.9 (Biomatters Ltd., Auckland, New Zealand) by Clicking the Primers button and selecting Design New Primers. The parameters were adjusted as shown in Figure 2 for optimization of the primer design. The optimum product size was adjusted to 500 bp. Then the best option for the primer pairs was selected based on the % GC content, primer melting temperature (Tm), Hairpin Tm and Self-Dimer Tm and used for designing degenerate primer pairs. Moreover, it is important that the sequence of the primers have guanine (G) or cytosine (C) within the last 5 nucleotides at the 3' end as they form 3 hydrogen bonds thus better stability. This is known as GC clamp.

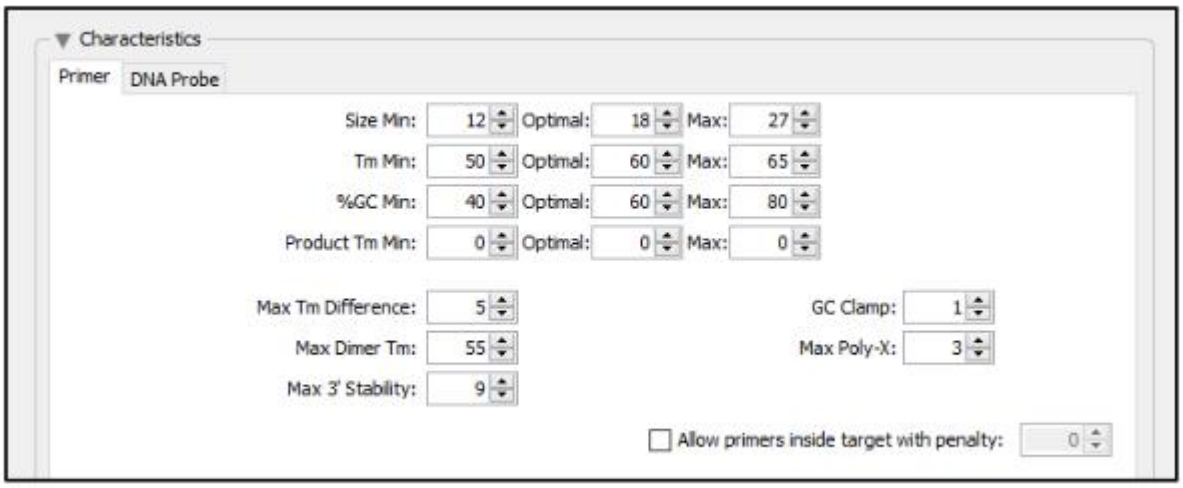


Figure 2: Optimization of the primer's properties.

Moreover, the degenerate primer pair was designed manually using Geneious software version 8.0.4 (Biomatters Ltd., Auckland, New Zealand) by following the procedure that mentioned on Geneious website. Firstly, the file with the aligned sequences was selected. Then the consensus bottom was selected, and the threshold was set to 50 % at which the bases matching was at least 50% of the sequences. The highlighting of the sequences was adjusted as depicted in Figure 3 for easy observation of the differences.

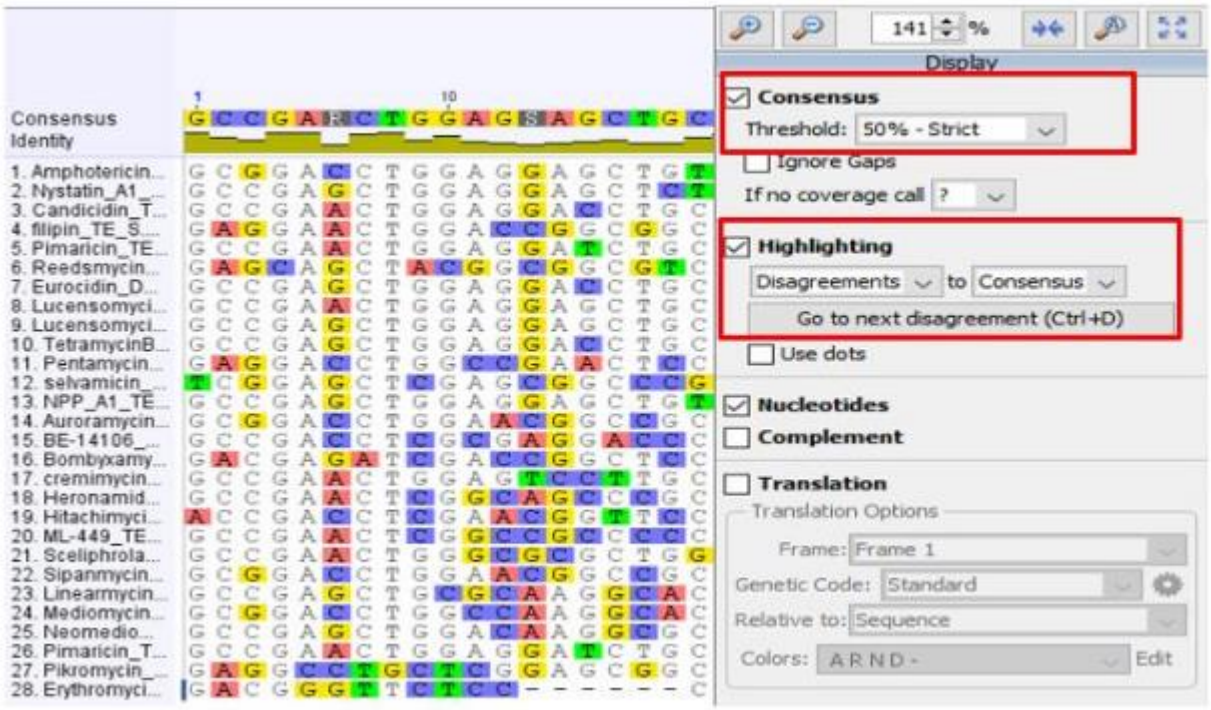


Figure 3: Screenshot illustrating the settings for designing degenerate primers.

The regions of the best option out of the 10 primer pairs that were designed before were selected on the consensus sequence in the alignment. For the forward primer, the region was selected from left to right while the region of the reverse primer was selected from right to left. The Add Annotation button was selected and the calculated characteristics of the primer, including the Tm range based on the degeneracy were checked. Then the name of the primer was entered, and the primer was added to the consensus as an annotation by clicking Ok.

2.3 In silico PCR

In silico PCR, also known as virtual PCR, is a computational method used to simulate the polymerase chain reaction (PCR) process. An online tool called “In silico PCR amplification” (http://insilico.ehu.es/PCR/index.php?mo=Streptomyces, accessed in October 2023) was used for testing the efficiency and specificity of the designed primer pair. The primer pair was uploaded and tested against all available Streptomyces strains allowing up to 2 mismatches, but in 1 nucleotide in 3' end. The maximum lengths of the bands were adjusted to 600 nucleotides (Figure 4).

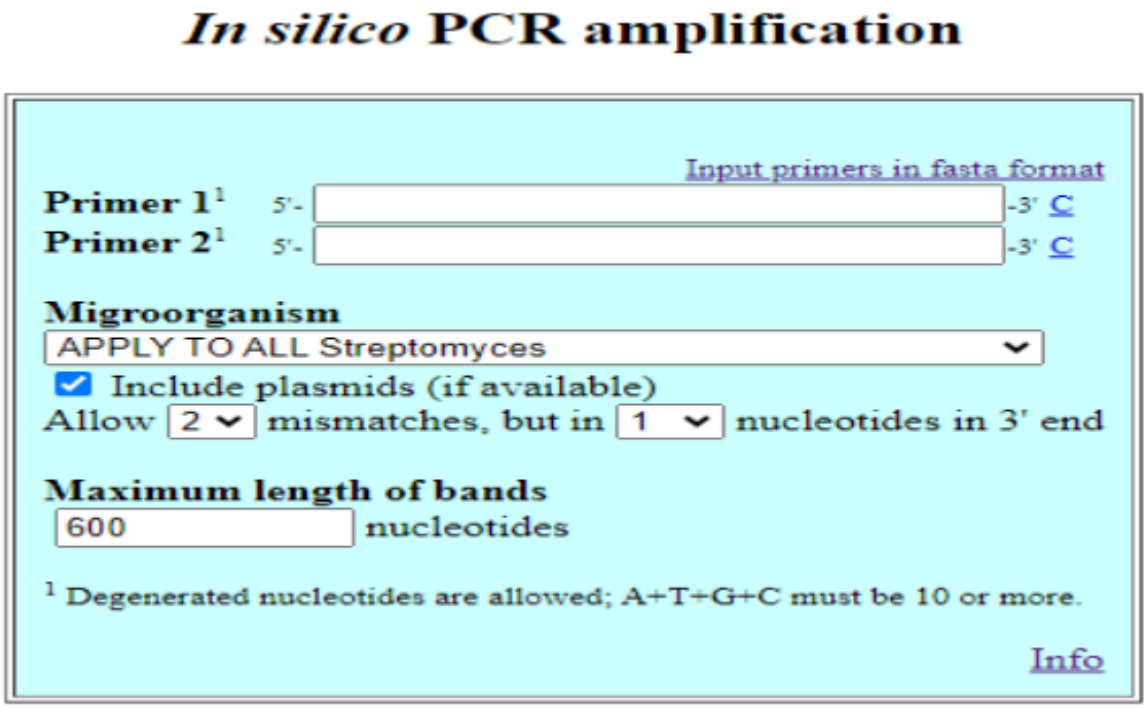


Figure 4: Screenshot illustrating the settings for testing the primer pair.

3. Results:

3.1 Data mining from polyene databases

Robust dataset of 26 linear and cyclic polyenes (macrolactone and macrolactam compounds) was created and used for the analysis of cytochrome P450 (CYP 450) and thioesterase type I (TE I) genes. They are listed in groups based on their structures together with bacterial strain, classification, database, and biological activity (Table 1).

Table 1: List of polyene secondary metabolites produced by Streptomyces sp. and phylogenetically close actinomycetes that were used in this study.

Compound	Bacterial strain	Classification	Biological activity	Database	Used for analysis of CYP 450	Used for analysis of TE
Macrolactone						
Amphotericin	*Streptomyces nodosus*-ATCC 14899	Heptaenes	Antifungal	JDB	✓	✓
Candicidin	*Streptomyces sp.* FR-008	Heptaenes	Antifungal	Mibig	✓	✓
Filipin	*Streptomyces avermitilis* MA-4680	Pentaenes	Antifungal and low anti-	Mibig	✓	✓

			bacterial activity			
Nystatin	*Streptomyces noursei* ATCC 11455	Tetraenes	Antifungal	JDB	✓	✓
Pimaricin	*Streptomyces Natalensis ATCC-27448*	Tetraenes	Antifungal and anti-protozoal	JDB	✓	✓
Rimocidin	*Streptomyces diastaticus 108*	Tetraenes	Antifungal	JDB	✓	
Reedsmycins	*Streptomyces* Sp. OUC6819	Pentaenes	Antifungal	Mibig		✓
Eurocidin-D	*Streptomyces eurocidicus* ATCC 27428	Pentaenes	Antifungal	Mibig		✓
Lucensomycin	*Streptomyces achromogenes_ NBRC*-14001 *Streptomyces cyanogenus*	Tetraenes	Antifungal and antiviral activity.	Mibig		✓
Tetramycin	*Streptomyces hygrospinosus* CGMCC 4.1123	Tetraenes	Antifungal	Mibig		✓
NPP	*Pseudonocardia-sp.* AL041005-10	Tetraenes	Antifungal	NCBI		✓
pentamycin	*Streptomyces sp*. S816	Pentaenes	Antifungal	NCBI		✓
Macrolactam						
Auroramycin	*Streptomyces filamentosus strain* NRRL 15998	Pentaenes	Antifungal, Anti-MRSA	Mibig, NCBI		✓
ML-449	*Streptomyces sp*. MP39-85	Tetraenes	Anti-bacterial, Cytotoxicity	Mibig		✓
BE-14106	*Streptomyces sp.* DSM-21069	Tetraenes	Anti-bacterial, Cytotoxicity	Mibig		✓
Sipanmycin	*Streptomyces sp*. CS149	Pentaenes	Cytotoxicity	Mibig		✓

Hitachimycin	*Streptomyces scabrisporus* CMB-0406	Triene	Anti-protozoal antifungal	Mibig		✓
Heronamide_C	*Streptomyces sp*. CMB-0406	Tetraenes	No anti-bacterial activity and cytotoxicity	Mibig		✓
Sceliphrolactam	*Streptomyces sp*.SD85	Pentaenes	Antifungal	NCBI		✓
Bombyxamycin A	*Streptomyces sp*.SD53	Pentaenes	Anti-bacterial, cytotoxicity	NCBI		✓
cremimycin	*Streptomyces sp.* MJ635-86F5	Triene	Anti-bacterial (G+), cytotoxicity	NCBI		✓
Linear						
Linearmycin A	*Streptomyces sp. Mg1*		Antifungal	Mibig		✓
Mediomycin A	*Streptomyces blastmyceticus*		Antifungal	Mibig		✓
Neomediomycin	*Streptomyces sp. RK95-74*		Antifungal	Mibig		✓

3.2 Analysis of Cytochrome P450

Amino acid sequences of cytochrome P450 from 6 polyene biosynthetic gene clusters (BGCs) were used for multiple sequence alignment as depicted in Figure 5. The majority of those polyene BGCs contain two cytochrome P450 genes, with the exception of Candicidin and Rimocidin BGCs.

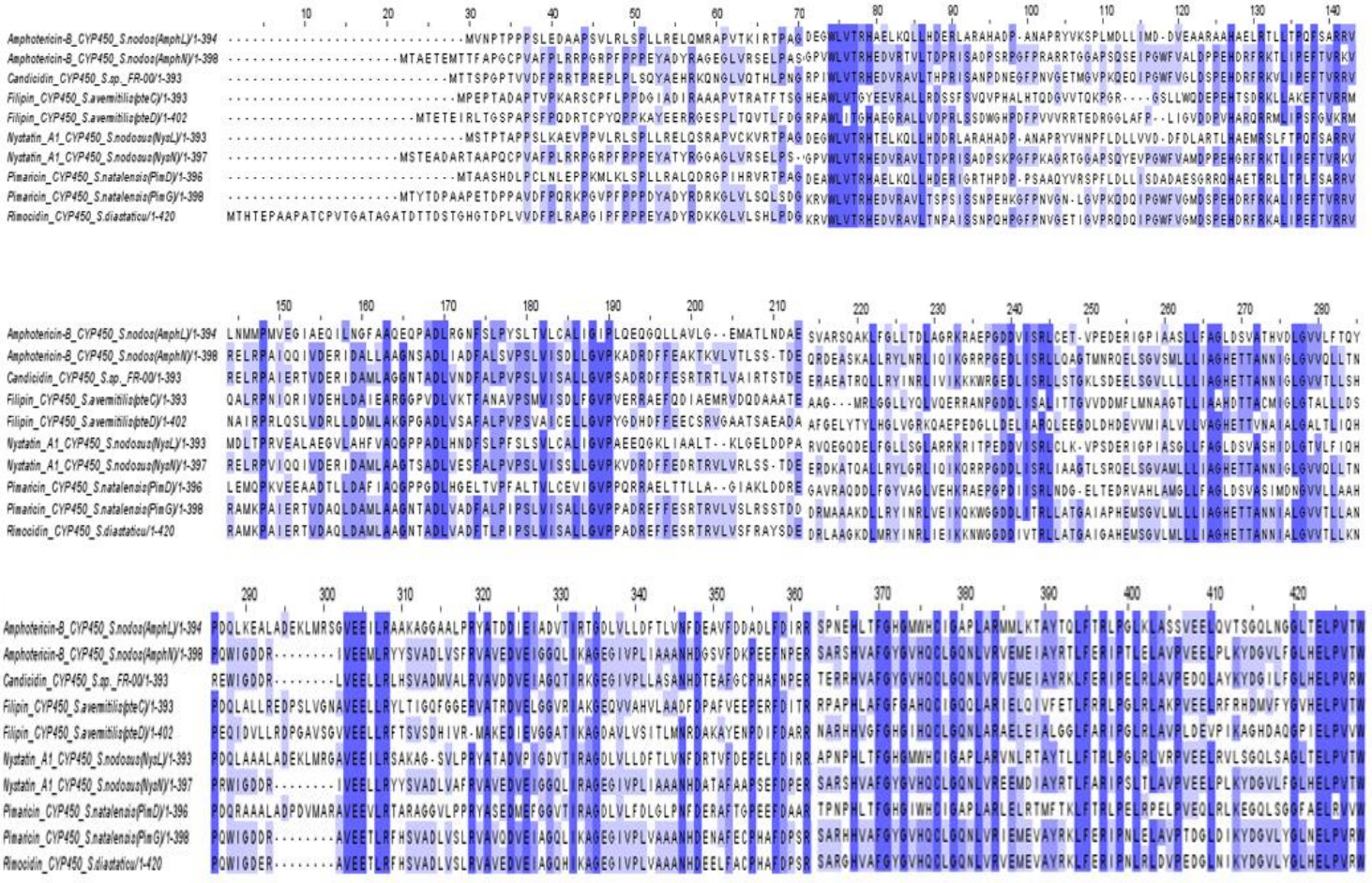


Figure 5: Multiple sequence alignment of the amino acid sequences of cytochrome P450. The highly conserved amino acid residues were highlighted in dark blue.

The results demonstrate the presence of the regions where the amino acid residues are highly conserved across the CYP sequences of the six-polyene BGCs. Based on the alignment, a phylogenetic tree was constructed to determine the similarity of the sequences among the CYPs from different BGCs (Figure 6).

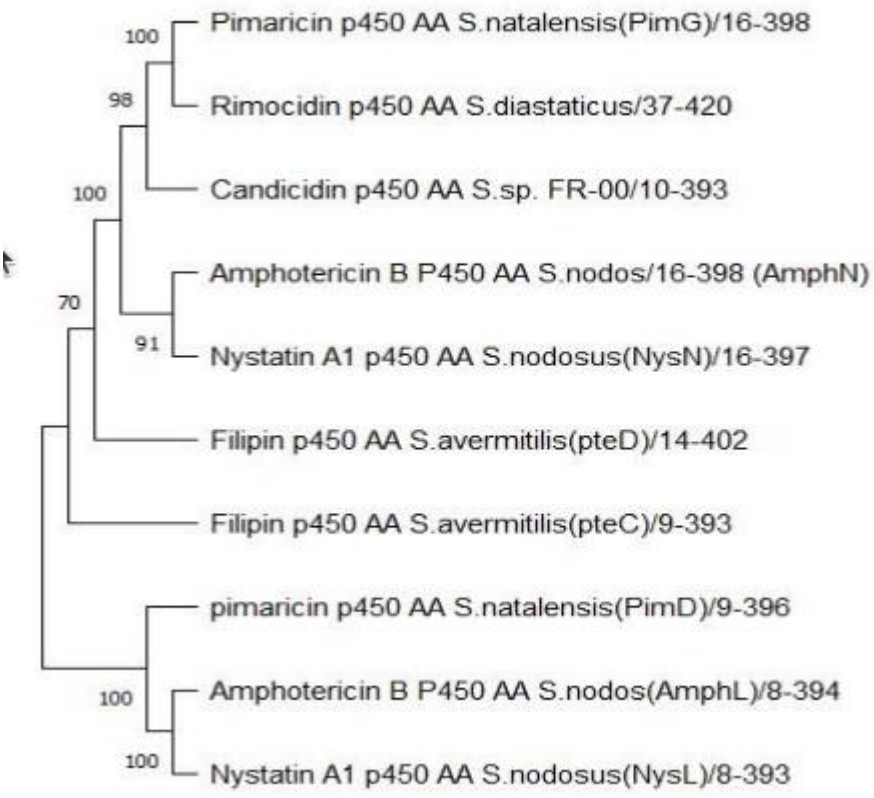


Figure 6: Phylogenetic tree of CYP450 gene extracted from BGCs coding for the 6 different Polyenes.

This analysis showed that cytochrome P450 genes encoded by the polyene clusters can be classified into two distinct similarity groups within the constructed tree. As shown in Figure 6, the first group consists of PimG, AmphN, NysN, PteD, RimG and fscP whereas the second group includes PimD, AmphL, PteC and NysL.

Upon searching for more polyene biosynthetic gene clusters (BGCs) for further analysis of cytochrome P450 (CYP) sequences, it has been discovered that certain polyenes, such as Reedsmycin (a cyclic polyene), and linear polyenes like Linearmycin-A, Mediomycin-A, and Neomediomycin-B, do not possess the genes encoding for cytochrome P450 as well as ferredoxin.

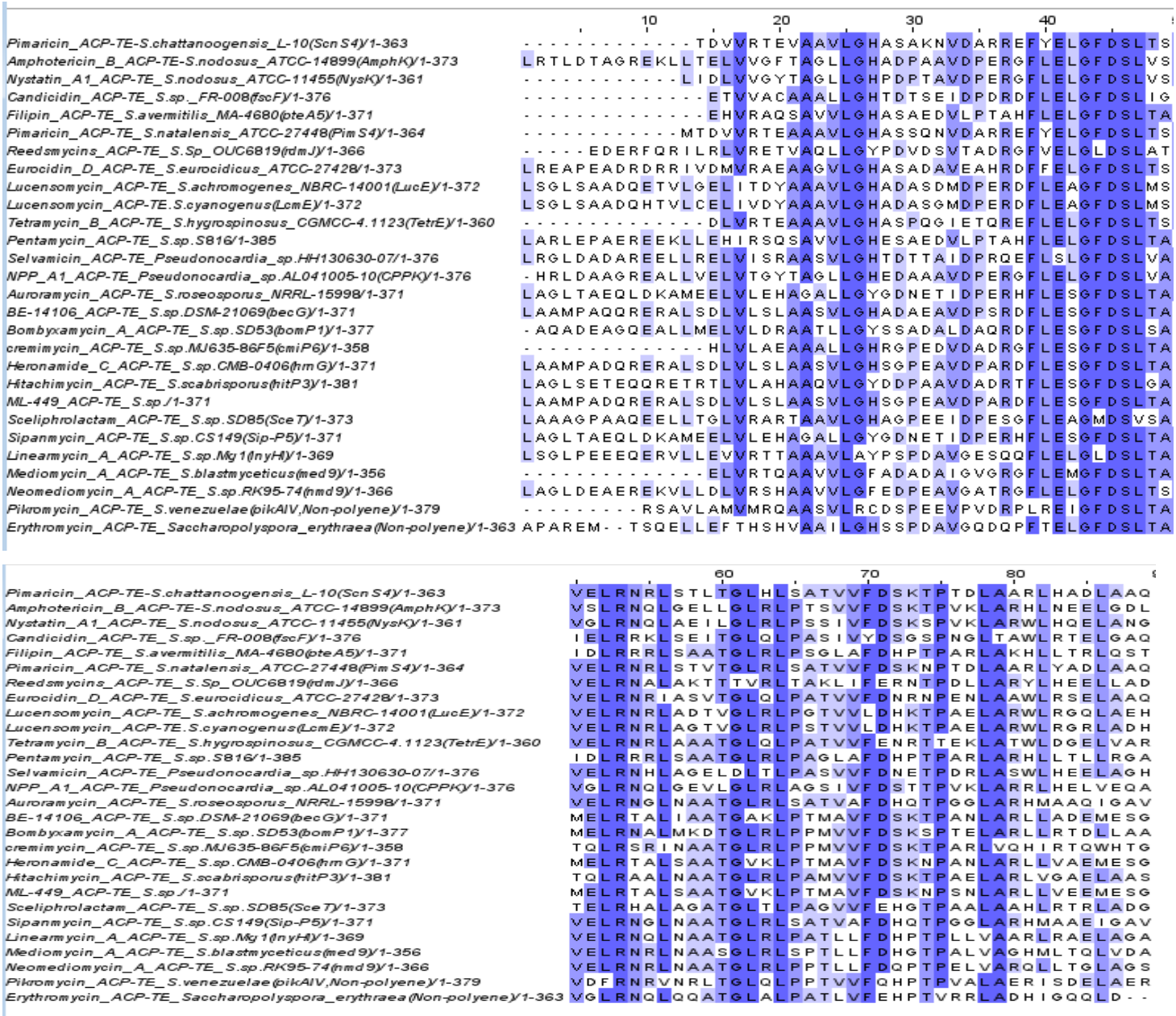
Therefore, the alternative hypothesis was formulated, and the dataset was expanded by including an additional 20 polyene and 2 non-polyene biosynthetic gene clusters (BGCs).

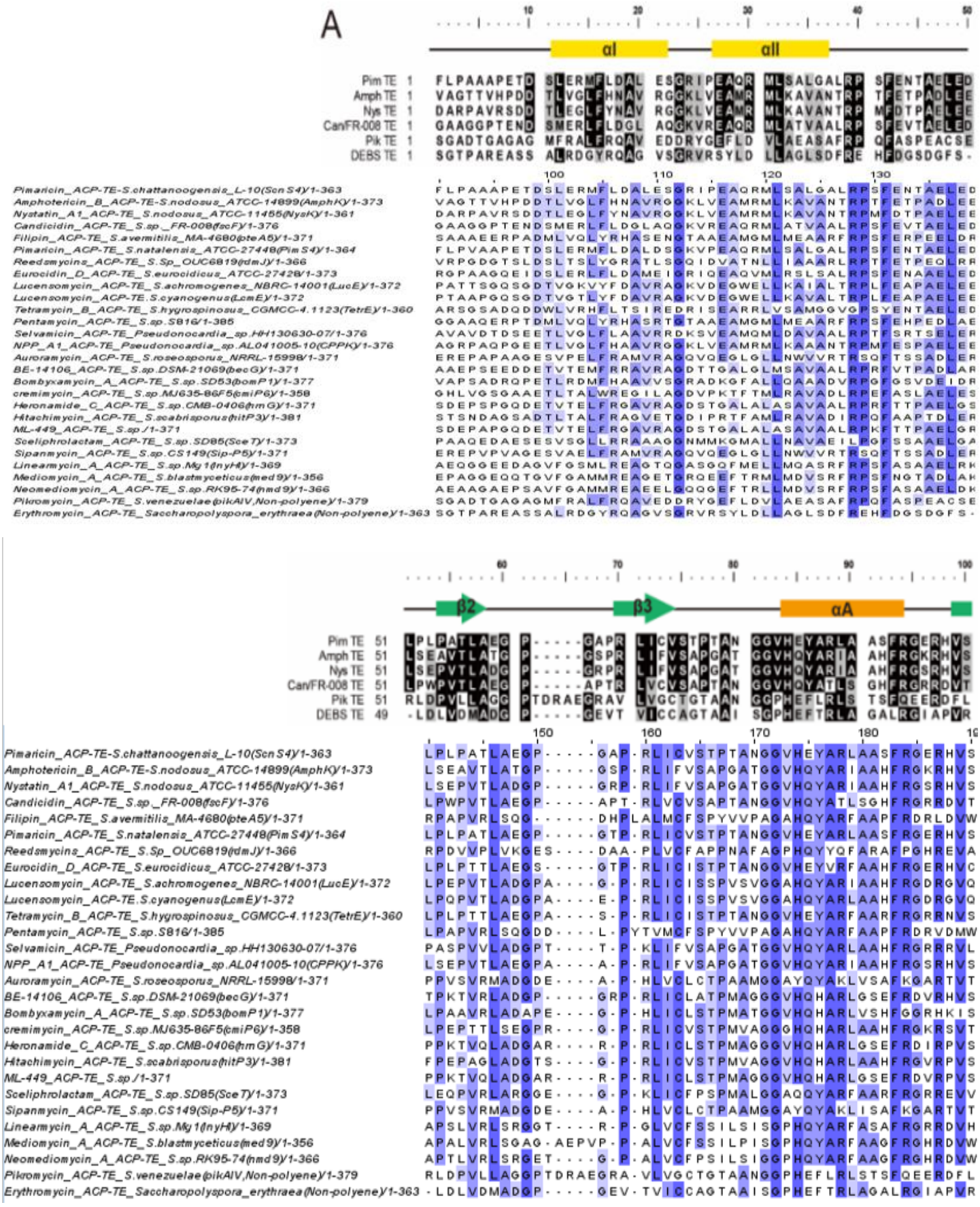
3.3 Analysis of Thioesterase type I

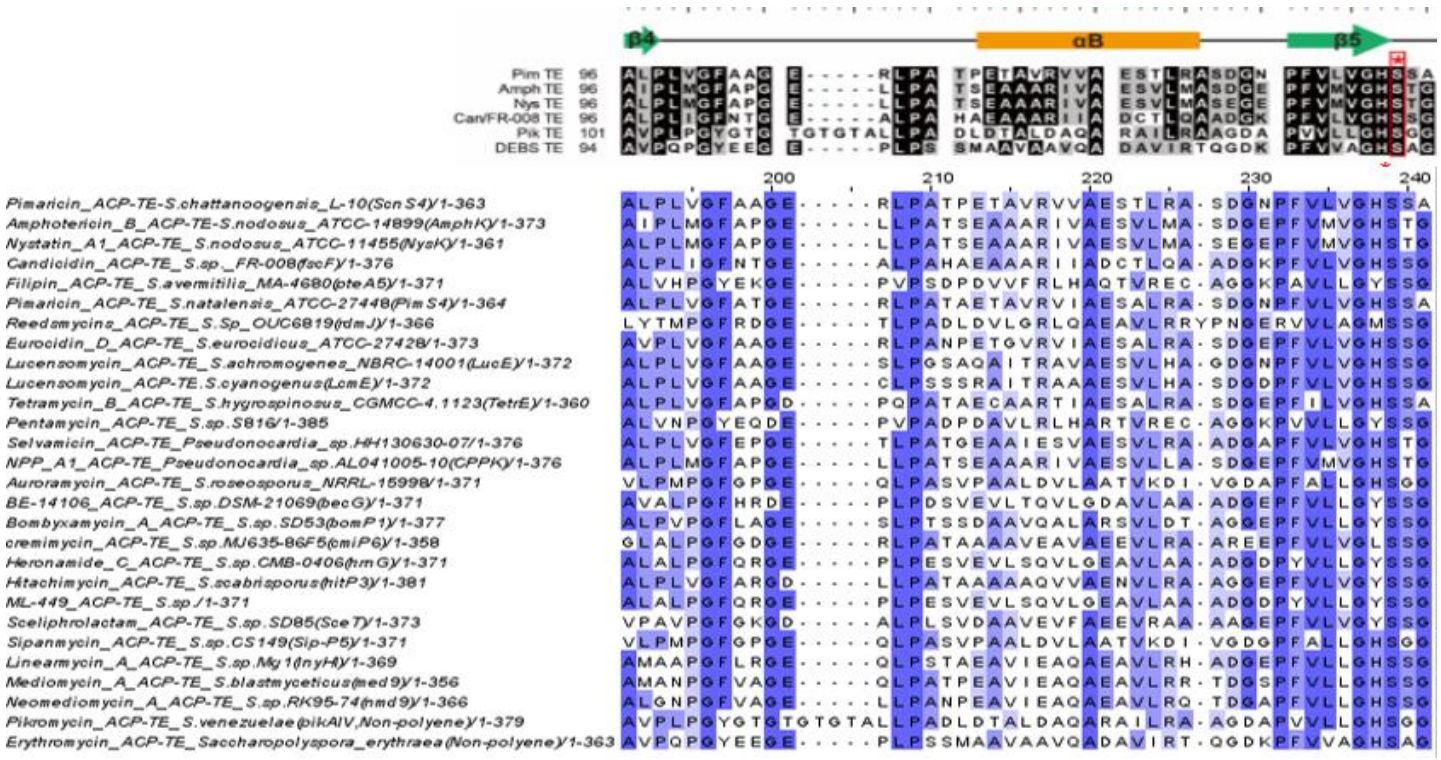
Since thioesterase type I is associated with the acyl carrier protein (ACP) of the last module of type I polyketide synthase, the nucleotide sequence of the whole module was downloaded. The analysis of the nucleotide sequences of the last module of type I polyketide synthase that associated with thioesterase type I revealed that there is only one starting codon and one stop codon, which means that the genes within the module cannot be synthesized individually but

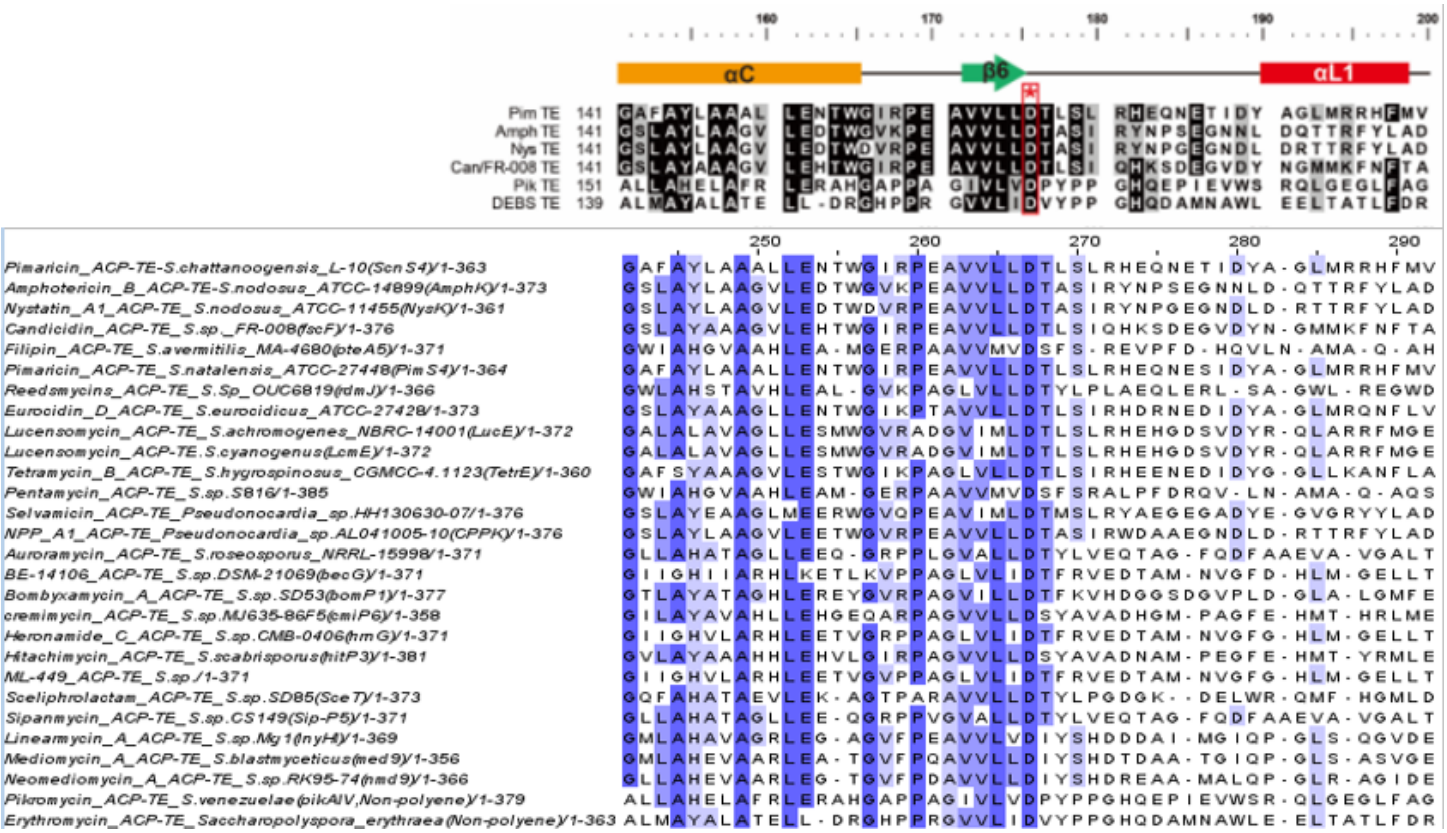
rather as a whole module. The nucleotide sequences of the last module of 26 polyenes and 2 non-polyenes were translated and the amino acid sequences were used for multiple sequence alignment. Then the aligned amino acid sequences of thioesterase type I (TE I) and acyl carrier protein (ACP) were extracted.

Then the hypothesis was tested by analyzing the sequence similarity of the gene responsible for encoding TE I within the dataset. The results of the multiple sequence alignment of thioesterase gene were compared to the results provided by Zhou, Yucong, et al (Figure 7).




A
αI
αII
β2
β3
αA
Pim TE
Amph TE
Nys TE
CanFR-008 TE
Pik TE
DEBS TE
Pimaricin_ACP-TE-S.chattanoogensis_L-10(ScnS4)/1-363
Amphotericin_B_ACP-TE-S.nodosus_ATCC-14899(AmphK)/1-373
Nystatin_A1_ACP-TE_S.nodosus_ATCC-11455(NysK)/1-361
Candicidin_ACP-TE_S.sp._FR-008(fscF)/1-376
Filipin_ACP-TE_S.avermitilis_MA-4680(pteA5)/1-371
Pimaricin_ACP-TE_S.natalensis_ATCC-27448(PimS4)/1-364
Reedsmycins_ACP-TE_S.Sp_OUC6819(rdmJ)/1-366
Eurocidin_D_ACP-TE_S.eurocidicus_ATCC-27428/1-373
Lucensomycin_ACP-TE_S.achromogenes_NBRC-14001(LucE)/1-372
Lucensomycin_ACP-TE.S.cyanogenus(LcmE)/1-372
Tetramycin_B_ACP-TE_S.hygrospinosus_CGMCC-4.1123(TetrE)/1-360
Pentamycin_ACP-TE_S.sp.S816/1-385
Selvamicin_ACP-TE_Pseudonocardia_sp.HH130630-07/1-376
NPP_A1_ACP-TE_Pseudonocardia_sp.AL041005-10(CPPK)/1-376
Auroramycin_ACP-TE_S.roseosporus_NRRL-15998/1-371
BE-14106_ACP-TE_S.sp.DSM-21069(becG)/1-371
Bombyxamycin_A_ACP-TE_S.sp.SD53(bomP1)/1-377
cremimycin_ACP-TE_S.sp.MJ635-86F5(cmiP6)/1-358
Heronamide_C_ACP-TE_S.sp.CMB-0406(hmG)/1-371
Hitachimycin_ACP-TE_S.scabrisporus(hitP3)/1-381
ML-449_ACP-TE_S.sp./1-371
Sceliphrolactam_ACP-TE_S.sp.SD85(SceT)/1-373
Sipanmycin_ACP-TE_S.sp.CS149(Sip-P5)/1-371
Linearmycin_A_ACP-TE_S.sp.Mj1(nyH)/1-369
Mediomycin_A_ACP-TE_S.blastmyceticus(med9)/1-356
Neomediomycin_A_ACP-TE_S.sp.RK95-74(nmd9)/1-366
Pikromycin_ACP-TE_S.venezuelae(pikAIV,Non-polyene)/1-379
Erythromycin_ACP-TE_Saccharopolyspora_erythraea(Non-polyene)/1-363


β4
αB
β5
Pim TE 96
Amph TE 96
Nys TE 96
Can/FR-008 TE 96
Pik TE 101
DEBS TE 94
200
210
220
230
240
Pimaricin_ACP-TE-S.chattanoogensis_L-10(ScnS4)/1-363
Amphotericin_B_ACP-TE-S.nodosus_ATCC-14899(AmphK)/1-373
Nystatin_A1_ACP-TE_S.nodosus_ATCC-11455(NysK)/1-361
Candicidin_ACP-TE_S.sp._FR-008(fscF)/1-376
Filipin_ACP-TE_S.avermitilis_MA-4680(pteA5)/1-371
Pimaricin_ACP-TE_S.natalensis_ATCC-27448(PimS4)/1-364
Reedsmycins_ACP-TE_S.Sp._OUC6819(dmJ)/1-366
Eurocidin_D_ACP-TE_S.eurocidicus_ATCC-27428/1-373
Lucensomycin_ACP-TE_S.achromogenes_NBRC-14001(LucE)/1-372
Lucensomycin_ACP-TE.S.cyanogenus(LcmE)/1-372
Tetramycin_B_ACP-TE_S.hygroscopinosus_CGMCC-4.1123(TetrE)/1-360
Pentamycin_ACP-TE_S.sp.S816/1-385
Selvamicin_ACP-TE_Pseudonocardia_sp.HH130630-07/1-376
NPP_A1_ACP-TE_Pseudonocardia_sp.AL041005-10(CPPK)/1-376
Auroramycin_ACP-TE_S.roseosporus_NRRL-15998/1-371
BE-14106_ACP-TE_S.sp.DSM-21069(becG)/1-371
Bombyxamycin_A_ACP-TE_S.sp.SD53(bomP1)/1-377
cremimycin_ACP-TE_S.sp.MJ635-86F5(cmiP6)/1-358
Heronamide_C_ACP-TE_S.sp.CMB-0406(hrnG)/1-371
Hitachimycin_ACP-TE_S.scabrisporus(hitP3)/1-381
ML-449_ACP-TE_S.sp./1-371
Sceliphrolactam_ACP-TE_S.sp.SD85(SceT)/1-373
Sipanmycin_ACP-TE_S.sp.CS149(Sip-P5)/1-371
Lineamycin_A_ACP-TE_S.sp.Mg1(nyH)/1-369
Mediomycin_A_ACP-TE_S.blastmyceticus(med9)/1-356
Neomediomycin_A_ACP-TE_S.sp.RK95-74(nmd9)/1-366
Pikromycin_ACP-TE_S.venezuelae(pikAIV,Non-polyene)/1-379
Erythromycin_ACP-TE_Saccharopolyspora_erythraea(Non-polyene)/1-363


αC
β6
αL1
Pim TE 141 GAFAYLAAAL LENTWGIRPE AVVLLDTLSL RHEQNETIDY AGLMRRHFMV
Amph TE 141 GSLAYLAAGV LEDTWGVKPE AVVLLDTASI RYNPSEGNNL DQTTRFYLAD
Nys TE 141 GSLAYLAAGV LEDTWDVRPE AVVLLDTASI RYNPGEGNDL DRTTRFYLAD
Can/FR-008 TE 141 GSLAYAAAGV LEHTWGIRPE AVVLLDTLSI GHKSDEGVDY NGMMKFNFTA
Pik TE 151 ALLAHELAFR LERAHGAPPA GIVLVDPYPP GMQEPIEVWS RQLGEGLFAG
DEBS TE 139 ALMAYALATE LL-DRGHPPR GVVLIDVYPP GHQDAMNAWL EELTATLFDR
250
260
270
280
290
Pimaricin_ACP-TE-S.chattanoogensis_L-10(ScnS4)Y1-363 GAFAYLAAALLENTWGIRPEAVVLLDTLSLRHEQNETIDYA-GLMRRHFMV
Amphotericin_B_ACP-TE-S.nodosus_ATCC-14899(AmphK)Y1-373 GSLAYLAAGVLEDTWGVKPEAVVLLDTASIRYNPSEGNNLD-QTTRFYLAD
Nystatin_A1_ACP-TE_S.nodosus_ATCC-11455(NysK)Y1-361 GSLAYLAAGVLEDTWDVRPEAVVLLDTASIRYNPGEGNDLD-RTTRFYLAD
Candicidin_ACP-TE_S.sp._FR-008(fscF)Y1-376 GSLAYAAAGVLEHTWGIRPEAVVLLDTLSIQHKSDEGVDYN-GMMKFNFTA
Filipin_ACP-TE_S.avermitilis_MA-4680(pteA5)Y1-371 GWIAHGVAAHLEA-MGERPAAVVMVDSFS-REVPFD-HQVLN-AMA-Q-AH
Pimaricin_ACP-TE_S.natalensis_ATCC-27448(PimS4)Y1-364 GAFAYLAAALLENTWGIRPEAVVLLDTLSLRHEQNESIDYA-GLMRRHFMV
Reedsmycins_ACP-TE_S.Sp_OUC6819(rdmJ)Y1-366 GWLAHSTAVHLEAL-GVKPAGLVLLDTYLPLAEQLERL-SA-GWL-REGWD
Eurocidin_D_ACP-TE_S.eurocidicus_ATCC-27428/1-373 GSLAYAAAGLLENTWGIKPTAVVLLDTLSIRHDRNEDIDYA-GLMRQNFLV
Lucensomycin_ACP-TE_S.achromogenes_NBRC-14001(LucE)Y1-372 GALALAVAGLLESMWGVRADGVIMLDTLSLRHEHGDSVDYR-QLARRFMGE
Lucensomycin_ACP-TE.S.cyanogenus(LcmE)Y1-372 GALALAVAGLLESMWGVRADGVIMLDTLSLRHEHGDSVDYR-QLARRFMGE
Tetramycin_B_ACP-TE_S.hygrospinosus_CGMCC-4.1123(TetrE)Y1-360 GAFSYAAAGVLESTWGIKPAGLVLLDTLSIRHEENEDIDYG-GLLKANFLA
Pentamycin_ACP-TE_S.sp.S816/1-385 GWIAHGVAAHLEAM-GERPAAVVMVDSFSRALPFDRQV-LN-AMA-Q-AQS
Selvamicin_ACP-TE_Pseudonocardia_sp.HH130630-07/1-376 GSLAYEAAGLMEERWGVQPEAVIMLDTMSLRYAEGEGADYE-GVGRYYLAD
NPP_A1_ACP-TE_Pseudonocardia_sp.AL041005-10(CPPK)Y1-376 GSLAYLAAGVLEETWGVRPEAVVLLDTASIRWDAAEGNDLD-RTTRFYLAD
Auroramycin_ACP-TE_S.roseosporus_NRRL-15998/1-371 GLLAHATAGLLEEQ-GRPPLGVALLDTYLVEQTAG-FQDFAAEVA-VGALT
BE-14106_ACP-TE_S.sp.DSM-21069(becG)Y1-371 GIIGHIIARHLKETLKVPPAGLVLIDTFRVEDTAM-NVGFD-HLM-GELLT
Bombyxamycin_A_ACP-TE_S.sp.SD53(bomP1)Y1-377 GTLAYATAGHLEREYGVRPAGVILLDTFKVHDGGSDGVPLD-GLA-LGMFE
cremimycin_ACP-TE_S.sp.MJ635-86F5(cmiP6)Y1-358 GILAYAVAHLLEHGEQARPAGVVLLDSYAVADHGM-PAGFE-HMT-HRLME
Heronamide_C_ACP-TE_S.sp.CMB-0406(hmG)Y1-371 GIIGHVLARHLEETVGRPPAGLVLIDTFRVEDTAM-NVGFG-HLM-GELLT
Hitachimycin_ACP-TE_S.scabrisporus(hitP3)Y1-381 GVLAYAAAHHLEHVLGIRPAGVVLLDSYAVADNAM-PEGFE-HMT-YRMLE
ML-449_ACP-TE_S.sp./1-371 GIIGHVLARHLEETVGVPPAGLVLIDTFRVEDTAM-NVGFG-HLM-GELLT
Sceliphrolactam_ACP-TE_S.sp.SD85(SceT)Y1-373 GQFAHATAEVLEK-AGTPARAVVLLDTYLPGDGK--DELWR-QMF-HGMLD
Sipanmycin_ACP-TE_S.sp.CS149(Sip-P5)Y1-371 GLLAHATAGLLEE-QGRPPVGVALLDTYLVEQTAG-FQDFAAEVA-VGALT
Linearmycin_A_ACP-TE_S.sp.Mg1(nyH)Y1-369 GMLAHAVAGRLEG-AGVFPEAVVLVDIYSHDDDAI-MGIQP-GLS-QGVDE
Mediomycin_A_ACP-TE_S.blastmyceticus(med9)Y1-356 GMLAHEVAARLEA-TGVFPQAVVLLDIYSHDTDAA-TGIQP-GLS-ASVGE
Neomediomycin_A_ACP-TE_S.sp.RK95-74(nmd9)Y1-366 GLLAHEVAARLEG-TGVFPDAVVLLDIYSHDREAA-MALQP-GLR-AGIDE
Pikromycin_ACP-TE_S.venezuelae(pikAIV,Non-polyene)Y1-379 ALLAHELAFRLERAHGAPPAGIVLVDPYPPGHQEPIEVWSR-QLGEGLFAG
Erythromycin_ACP-TE_Saccharopolyspora_erythraea(Non-polyene)Y1-363 ALMAYALATELL-DRGHPPRGVVLIDVYPPGHQDAMNAWLE-ELTATLFDR
αL2
β7
Pim TE 191 DEVSPVRMTN SRLSAMARWM GMLNQLEVRH LTVPVLIIRA AKETFG--IG
Amph TE 191 IDSPSVTLNS ARMSAMAHWF MAMTDIDAPA LTAPTLLLRA TQANNG----
Nys TE 191 IDSPSVTLNS ARMSAMAHWF MAMTDIQAPA PTAPTLLVRA ARALDG----
Can/FR-008 TE 191 VDDSPVRLTN SRLSAMGRWM VLLNALDVHP LTVPVLEIKC TRALIEGVPA
Pik TE 201 ELEP---MSD ARLLAMGRYA RFLAGPRPGR SSAPVLLVRA SEPLGD--WQ
DEBS TE 188 ETVR---MDD TRLTALGAYD RLTGQWRPRE TGLPTLLVSA GEPMGP--WP
300
310
320
330
340
Pimaricin_ACP-TE-S.chattanoogensis_L-10(ScnS4)Y1-363 DEVSPVRMT-NSRLSAMARWMGMLNQL-E-VRHTTVPVLIIRAAKETFG--IG
Amphotericin_B_ACP-TE-S.nodosus_ATCC-14899(AmphK)Y1-373 IDSPSVTLN-SARMSAMAHWFMAMTDI-D-APATTAPTLLLRATQANNG----
Nystatin_A1_ACP-TE_S.nodosus_ATCC-11455(NysK)Y1-361 IDSPSVTLN-SARMSAMAHWFMAMTDI-Q-APAPTAPTLLVRAARALDG----
Candicidin_ACP-TE_S.sp._FR-008(fscF)Y1-376 VDDSPVRLT-NSRLSAMGRWMVLLNAL-D-VHPTTVPVLEIKCTRALIEGVPA
Filipin_ACP-TE_S.avermitilis_MA-4680(pteA5)Y1-371 SQRLDFMRSGGEQLTAMGRYMRL-FDDW-AAPQIAAPTLLVRAS-EPMLAGAG
Pimaricin_ACP-TE_S.natalensis_ATCC-27448(PimS4)Y1-364 DEVSPVRMT-NSRLSAMARWMGMLNQL-E-VRHTTVPVLIIRATKETFG--IG
Reedsmycins_ACP-TE_S.Sp_OUC6819(rdmJ)Y1-366 RAENYNYLD-GRALTATGWYMRLFGQLWK-PSEIATPTLLVRATELVDG--EG
Eurocidin_D_ACP-TE_S.eurocidicus_ATCC-27428/1-373 DEVSPVRMT-NSRLSAMGRWMGLLNRV-D-VQHTTAPVLNLQALRQHAGV-ES
Lucensomycin_ACP-TE_S.achromogenes_NBRC-14001(LucE)Y1-372 TDSATVTVD-SNRLSAMAHYLNRMSAL-E-VPPTTAPTLLVRCSVPLLGDPDA
Lucensomycin_ACP-TE.S.cyanogenus(LcmE)Y1-372 TDSATVTVD-SNRLSAMAHYLNRMSAL-E-VPPTTAPTLLVRCSVPLLGDPDT
Tetramycin_B_ACP-TE_S.hygrospinosus_CGMCC-4.1123(TetrE)Y1-360 SEATQVRIL-NANLSAMGRWMHLLNSI-D-VPQMNTPVLSIQATRLYPGIAAG
Pentamycin_ACP-TE_S.sp.S816/1-385 ERLDFMRSG-GEQLTAMGRYMRLFDDW-A-APEITAPTLLVRASEPVPAAAAD
Selvamicin_ACP-TE_Pseudonocardia_sp.HH130630-07/1-376 IDSPAVALT-STRLTAMVHWYNRAAAL-RPVGETTAPTLLVRASIPLPGGKGP
NPP_A1_ACP-TE_Pseudonocardia_sp.AL041005-10(CPPK)Y1-376 IDSPSVTLN-SARMSAMAHWFMTMTEI-D-SPPTTAPTLLVRAARDVEGFILD
Auroramycin_ACP-TE_S.roseosporus_NRRL-15998/1-371 REISYGQLGNATSLSAMAQYVALMPEI-E-LKELATPTLLVKASERFATGPED
BE-14106_ACP-TE_S.sp.DSM-21069(becG)Y1-371 VETTLGNYD-AARLSAMPHYFQVLAGF-D-PVRLDTPTLFVQASEPFVQPPEG
Bombyxamycin_A_ACP-TE_S.sp.SD53(bomP1)Y1-377 KEAVFGRFD-SSRLSAMGRWVELVPQL-P-LNPVESPVLFVQCTQSFVPDGDD
cremimycin_ACP-TE_S.sp.MJ635-86F5(cmiP6)Y1-358 LETRFGPYR-SAELTAMSRYFHFLPEF-T-RASLKAPVLFVGAEKTFIPETGS
Heronamide_C_ACP-TE_S.sp.CMB-0406(hmG)Y1-371 VETTVGSFD-TARLSAMPYYFQVLAGF-D-PVRLAAPTLFVQASEPFVQPPEG
Hitachimycin_ACP-TE_S.scabrisporus(hitP3)Y1-381 MEATFGPYD-SAQLCAMARYFHLLPDF-T-RESVGAPVLFVGAGKSFKPETGD
ML-449_ACP-TE_S.sp./1-371 VETTVGSFD-TARLSAMPYYFQVLAGF-D-PVRLAAPTLFVQASEPFVQPPEG
Sceliphrolactam_ACP-TE_S.sp.SD85(SceT)Y1-373 RESSFGRFS-AARLAAMSRYSDLITGC-L-PGALSAPVLFVRPAESFATGAG-
Sipanmycin_ACP-TE_S.sp.CS149(Sip-P5)Y1-371 REISYGQLGNATSLSAMAQYVALMPEI-E-LKELATPTLLVKASERFATGPEG
Linearmycin_A_ACP-TE_S.sp.Mg1(nyH)Y1-369 RQGSYVPVD-DSRLLAMGAYFRLFGGW-K-PEAVKAPTLLVRAQEQFFDWTRA
Mediomycin_A_ACP-TE_S.blastmyceticus(med9)Y1-356 RSESHVPLD-DARLLAMGAYFRLFGEW-R-PREIKTPTLLVRAGERLFDWTRG
Neomediomycin_A_ACP-TE_S.sp.RK95-74(nmd9)Y1-366 RSTGPVPVD-DARLLAMGAYFRLFGQW-E-PREVKTPTLLVRAEEQFFDWSR-
Pikromycin_ACP-TE_S.venezuelae(pikAIV,Non-polyene)Y1-379 ELEP---MS-DARLLAMGRYARFLAGP-R-PGRSSAPVLLVRASEPLGD--WQ
Erythromycin_ACP-TE_Saccharopolyspora_erythraea(Non-polyene)Y1-363 ET---VRMD-DTRLTALGAYDRLTGQW-R-PRETGLPTLLVSAGEPMGP--WP

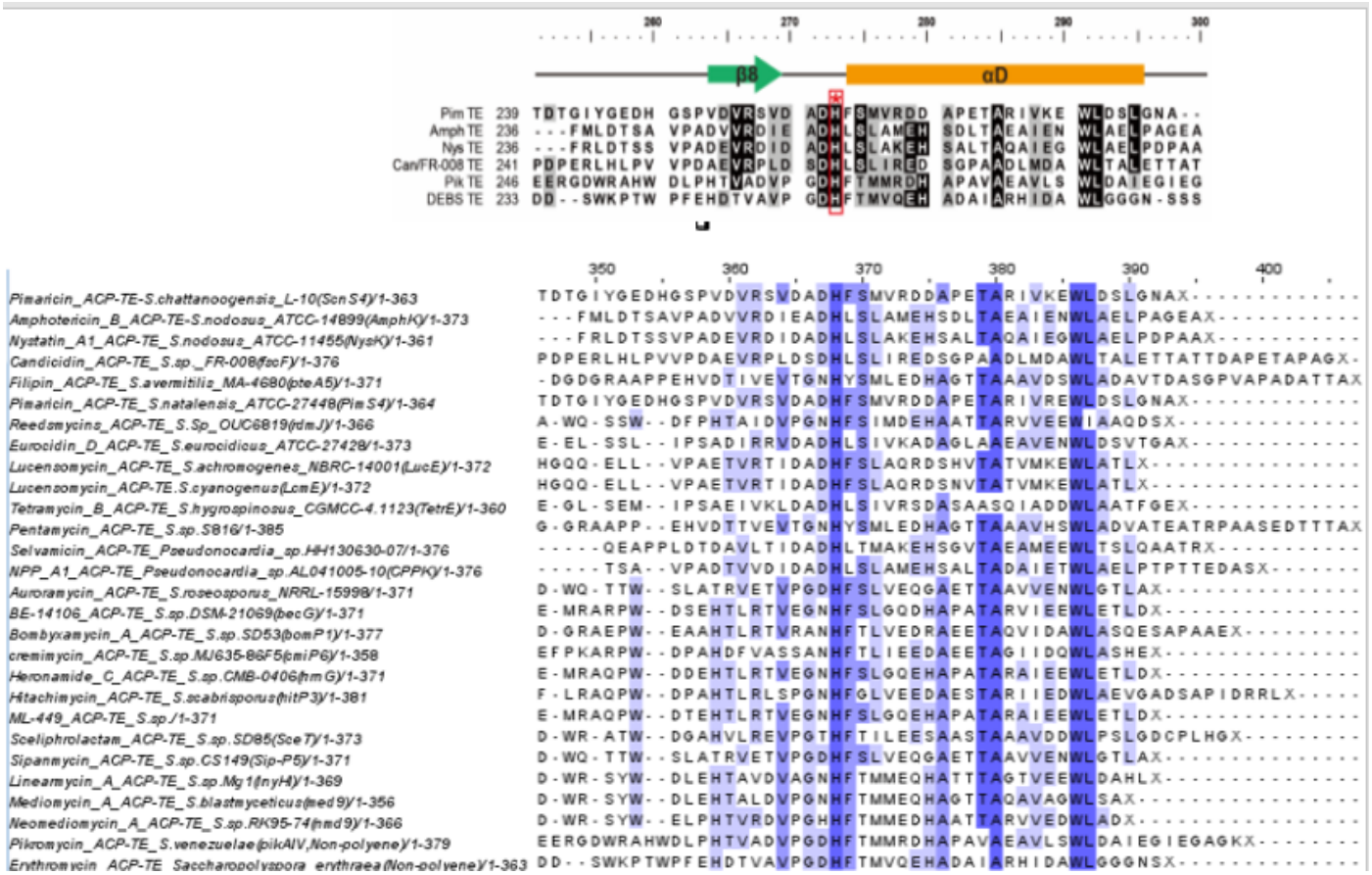
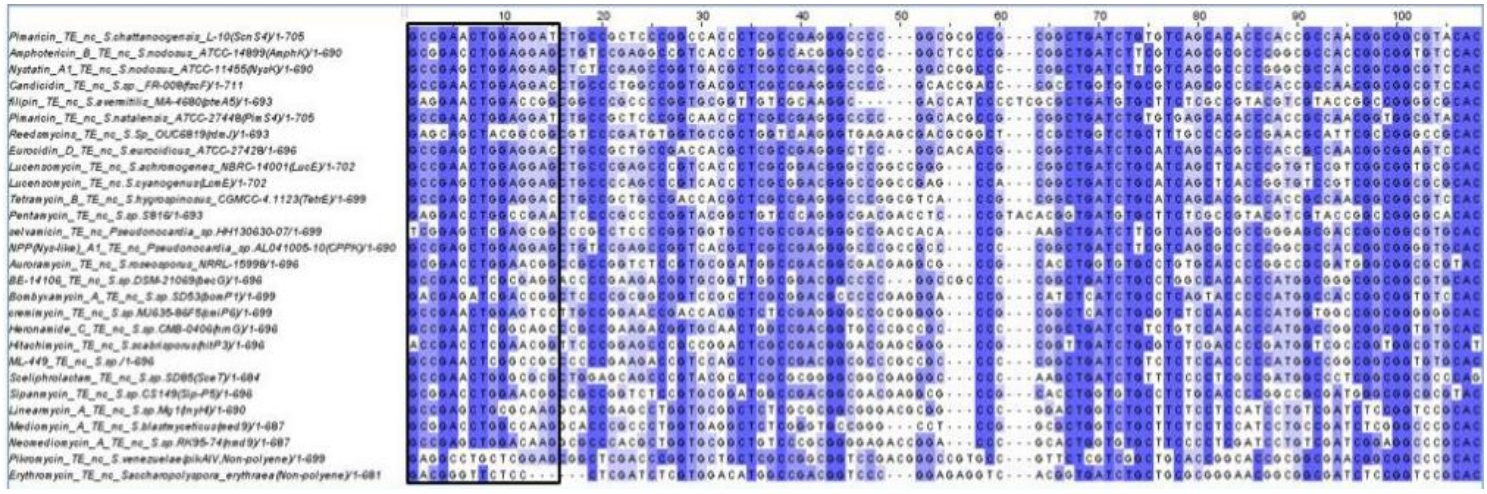
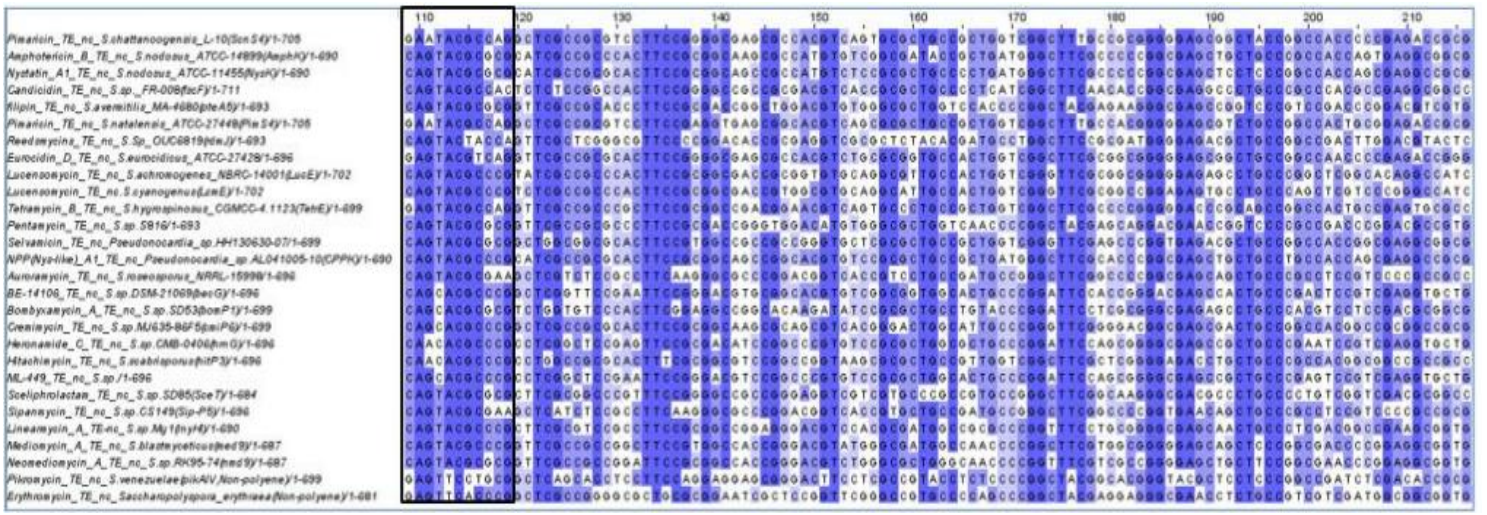


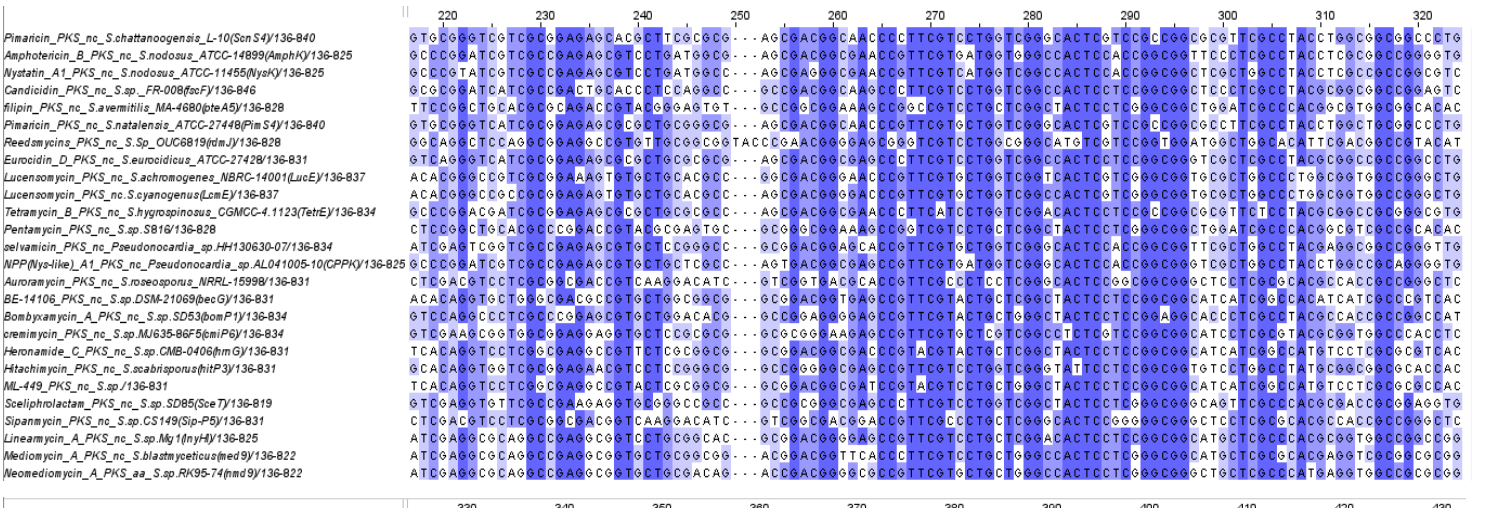
Figure 7: Multiple sequence alignment of amino acid residues of TE type I gene associated with ACP of the last module of polyketide synthase (PKS). The result of the alignment of thioesterase gene was compared to the results provided by Zhou, Yucong, et al.

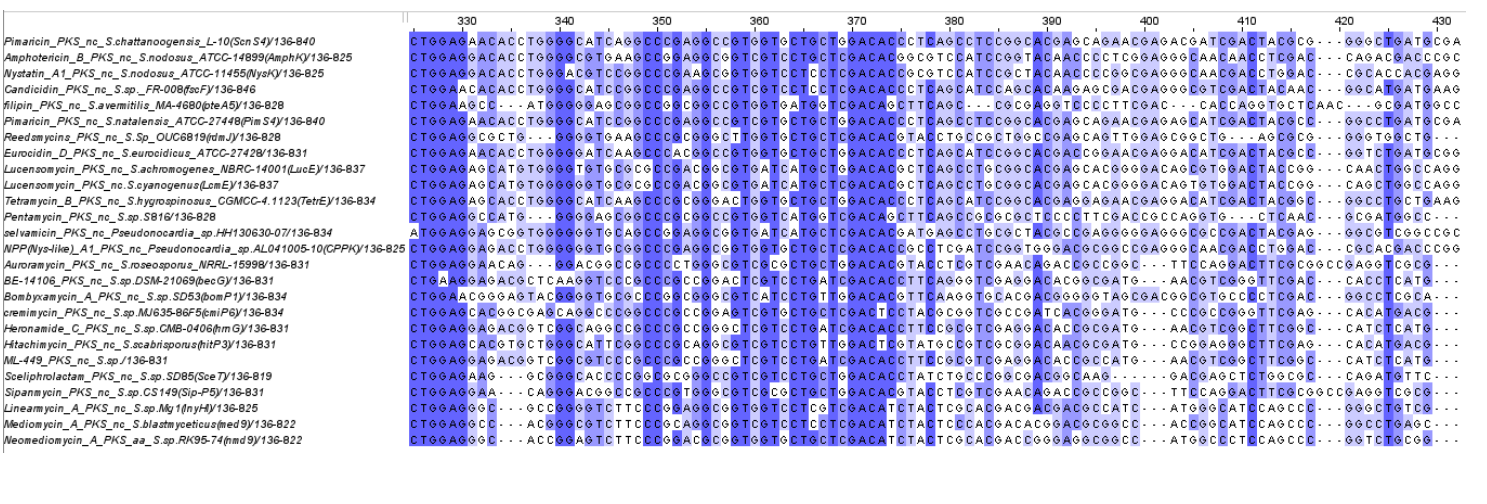
The results demonstrated that the ACP gene is highly conserved among both polyene and non-polyene polyketide synthases, since it codes for important functional domains that are crucial for its activity. Moreover, the analysis of the multiple sequence alignment of the comprehensive dataset and its comparison to the aligned sequence in previous literature indicated that the cyclic (macrolactone and macrolactam) and linear polyenes share conserved domains with some variability in this region. Although different amino acids may be present within the conserved regions, these variations do not alter the electric charges and the overall geometry of the conserved domain.

Then the highly conserved region of the thioesterase for the polyenes from amino acid 45 to 387 was extracted and translated back into nucleotide sequences for further analysis at the nucleotide level and for the purpose of designing primers.


Pimaricin_TE_nc_S.chattanoogensis_L-10(ScnS4)/1-705
Amphotericin_B_TE_nc_S.nodosus_ATCC-14899(AmphK)/1-690
Nystatin_A1_TE_nc_S.nodosus_ATCC-11455(NysK)/1-690
Candicidin_TE_nc_S.sp._FR-008(fscF)/1-711
filipin_TE_nc_S.avermitilis_MA-4680(pteA5)/1-693
Pimaricin_TE_nc_S.natalensis_ATCC-27448(PimS4)/1-705
Reedsmycins_TE_nc_S.Sp_OUC6819(rdmJ)/1-693
Eurocidin_D_TE_nc_S.eurocidicus_ATCC-27428/1-696
Lucensomycin_TE_nc_S.achromogenes_NBRC-14001(LucE)/1-702
Lucensomycin_TE_nc_S.cyanogenus(LcmE)/1-702
Tetramycin_B_TE_nc_S.hygroscopinosus_CGMCC-4.1123(TetE)/1-699
Pentamycin_TE_nc_S.sp.SB16/1-693
selvamicin_TE_nc_Pseudonocardia_sp.HH130630-07/1-699
NPP(Nys-like)_A1_TE_nc_Pseudonocardia_sp.AL041005-10(CPPK)/1-690
Auroramycin_TE_nc_S.roseosporus_NRRL-15998/1-696
BE-14106_TE_nc_S.sp.DSM-21069(becG)/1-696
Bombyxamycin_A_TE_nc_S.sp.SD53(bomP1)/1-699
cremimycin_TE_nc_S.sp.MJ635-86F5(cmiP6)/1-699
Heronamide_C_TE_nc_S.sp.CMB-0406(hmG)/1-696
Hitachimycin_TE_nc_S.scabrisporus(hitP3)/1-696
ML-449_TE_nc_S.sp./1-696
Sceliphrolactam_TE_nc_S.sp.SD85(SceT)/1-684
Sipanmycin_TE_nc_S.sp.CS149(Sip-P5)/1-696
Linearmycin_A_TE_nc_S.sp.My1(nyH)/1-690
Mediomycin_A_TE_nc_S.blastmyceticus(med9)/1-687
Neomediomycin_A_TE_nc_S.sp.RK95-74(hmd9)/1-687
Pikromycin_TE_nc_S.venezuelae(pikAIV,Non-polyene)/1-699
Erythromycin_TE_nc_Saccharopolyspora_erythraea(Non-polyene)/1-681


Pimaricin_TE_nc_S.chattanoogensis_L-10(ScnS4)/1-705
Amphotericin_B_TE_nc_S.nodosus_ATCC-14899(AmphK)/1-690
Nystatin_A1_TE_nc_S.nodosus_ATCC-11455(NysK)/1-690
Candicidin_TE_nc_S.sp._FR-008(fscF)/1-711
filipin_TE_nc_S.avermitilis_MA-4680(pteA5)/1-693
Pimaricin_TE_nc_S.natalensis_ATCC-27448(PimS4)/1-705
Reedsmycins_TE_nc_S.Sp_OUC6819(rdmJ)/1-693
Eurocidin_D_TE_nc_S.eurocidicus_ATCC-27428/1-696
Lucensomycin_TE_nc_S.achromogenes_NBRC-14001(LucE)/1-702
Lucensomycin_TE_nc_S.cyanogenus(LcmE)/1-702
Tetramycin_B_TE_nc_S.hygroscopinosus_CGMCC-4.1123(TetE)/1-699
Pentamycin_TE_nc_S.sp.SB16/1-693
Selvamicin_TE_nc_Pseudonocardia_sp.HH130630-07/1-699
NPP(Nys-like)_A1_TE_nc_Pseudonocardia_sp.AL041005-10(CPPK)/1-690
Auroramycin_TE_nc_S.roseosporus_NRRL-15998/1-696
BE-14106_TE_nc_S.sp.DSM-21069(becG)/1-696
Bombyxamycin_A_TE_nc_S.sp.SD53(bomP1)/1-699
cremimycin_TE_nc_S.sp.MJ635-86F5(cmiP6)/1-699
Heronamide_C_TE_nc_S.sp.CMB-0406(hmG)/1-696
Hitachimycin_TE_nc_S.scabrisporus(hitP3)/1-696
ML-449_TE_nc_S.sp./1-696
Sceliphrolactam_TE_nc_S.sp.SD85(SceT)/1-684
Sipanmycin_TE_nc_S.sp.CS149(Sip-P5)/1-696
Linearmycin_A_TE_nc_S.sp.My1(nyH)/1-690
Mediomycin_A_TE_nc_S.blastmyceticus(med9)/1-687
Neomediomycin_A_TE_nc_S.sp.RK95-74(hmd9)/1-687
Pikromycin_TE_nc_S.venezuelae(pikAIV,Non-polyene)/1-699
Erythromycin_TE_nc_Saccharopolyspora_erythraea(Non-polyene)/1-681


Pimaricin_PKS_nc_S.chattanoogensis_L-10(ScnS4)/136-840
Amphotericin_B_PKS_nc_S.nodosus_ATCC-14899(AmphK)/136-825
Nystatin_A1_PKS_nc_S.nodosus_ATCC-11455(NysK)/136-825
Candicidin_PKS_nc_S.sp._FR-008(fscF)/136-846
filipin_PKS_nc_S.avermitilis_MA-4680(pteA5)/136-828
Pimaricin_PKS_nc_S.natalensis_ATCC-27448(PimS4)/136-840
Reedsmycins_PKS_nc_S.Sp_OUC6819(rdmJ)/136-828
Eurocidin_D_PKS_nc_S.eurocidicus_ATCC-27428/136-831
Lucensomycin_PKS_nc_S.achromogenes_NBRC-14001(LucE)/136-837
Lucensomycin_PKS_nc_S.cyanogenus(LcmE)/136-837
Tetramycin_B_PKS_nc_S.hygroscopinosus_CGMCC-4.1123(TetE)/136-834
Pentamycin_PKS_nc_S.sp.SB16/136-828
selvamicin_PKS_nc_Pseudonocardia_sp.HH130630-07/136-834
NPP(Nys-like)_A1_PKS_nc_Pseudonocardia_sp.AL041005-10(CPPK)/136-825
Auroramycin_PKS_nc_S.roseosporus_NRRL-15998/136-831
BE-14106_PKS_nc_S.sp.DSM-21069(becG)/136-831
Bombyxamycin_A_PKS_nc_S.sp.SD53(bomP1)/136-834
cremimycin_PKS_nc_S.sp.MJ635-86F5(cmiP6)/136-834
Heronamide_C_PKS_nc_S.sp.CMB-0406(hmG)/136-831
Hitachimycin_PKS_nc_S.scabrisporus(hitP3)/136-831
ML-449_PKS_nc_S.sp./136-831
Sceliphrolactam_PKS_nc_S.sp.SD85(SceT)/136-819
Sipanmycin_PKS_nc_S.sp.CS149(Sip-P5)/136-831
Linearmycin_A_PKS_nc_S.sp.My1(nyH)/136-825
Mediomycin_A_PKS_nc_S.blastmyceticus(med9)/136-822
Neomediomycin_A_PKS_aa_S.sp.RK95-74(hmd9)/136-822


Pimaricin_PKS_nc_S.chattanoogensis_L-10(ScnS4)/136-840
Amphotericin_B_PKS_nc_S.nodosus_ATCC-14899(AmphK)/136-825
Nystatin_A1_PKS_nc_S.nodosus_ATCC-11455(NysK)/136-825
Candicidin_PKS_nc_S.sp._FR-008(fscF)/136-846
filipin_PKS_nc_S.avermitilis_MA-4680(pteA5)/136-828
Pimaricin_PKS_nc_S.natalensis_ATCC-27448(PimS4)/136-840
Reedsmycins_PKS_nc_S.Sp_OUC6819(rdmJ)/136-828
Eurocidin_D_PKS_nc_S.eurocidicus_ATCC-27428/136-831
Lucensomycin_PKS_nc_S.achromogenes_NBRC-14001(LucE)/136-837
Lucensomycin_PKS_nc_S.cyanogenus(LcmE)/136-837
Tetramycin_B_PKS_nc_S.hygroscopinosus_CGMCC-4.1123(TetE)/136-834
Pentamycin_PKS_nc_S.sp.SB16/136-828
selvamicin_PKS_nc_Pseudonocardia_sp.HH130630-07/136-834
NPP(Nys-like)_A1_PKS_nc_Pseudonocardia_sp.AL041005-10(CPPK)/136-825
Auroramycin_PKS_nc_S.roseosporus_NRRL-15998/136-831
BE-14106_PKS_nc_S.sp.DSM-21069(becG)/136-831
Bombyxamycin_A_PKS_nc_S.sp.SD53(bomP1)/136-834
cremimycin_PKS_nc_S.sp.MJ635-86F5(cmiP6)/136-834
Heronamide_C_PKS_nc_S.sp.CMB-0406(hmG)/136-831
Hitachimycin_PKS_nc_S.scabrisporus(hitP3)/136-831
ML-449_PKS_nc_S.sp./136-831
Sceliphrolactam_PKS_nc_S.sp.SD85(SceT)/136-819
Sipanmycin_PKS_nc_S.sp.CS149(Sip-P5)/136-831
Linearmycin_A_PKS_nc_S.sp.My1(nyH)/136-825
Mediomycin_A_PKS_nc_S.blastmyceticus(med9)/136-822
Neomediomycin_A_PKS_aa_S.sp.RK95-74(hmd9)/136-822

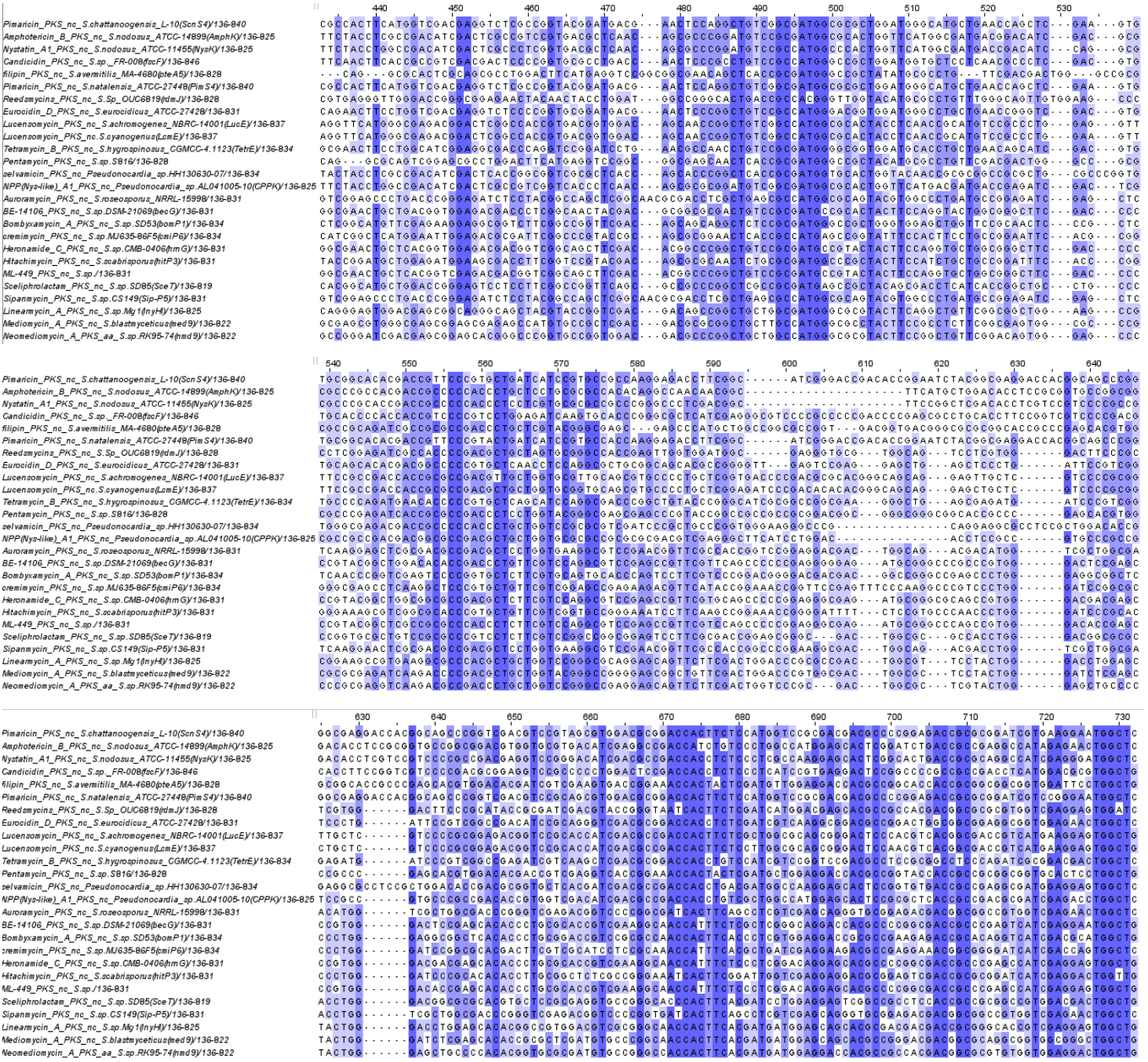


Figure 8: Multiple sequence alignment of the conserved region of TE type I nucleotide sequences. The regions of interest that could be used for designing the primers are indicated with black boxes.

The analysis of the coding region of the thioesterase gene showed that there is a sequence variability at the level of nucleotides so that the degenerative positions needed to be used.

3.4 Designing of primers

Factors such as G+C content, length, melting temperature (Tm) and degeneracy should be considered during designing of primers as they provide optimal specificity and efficiency in PCR amplification.

The results of the designed primers are listed in table 2, analyzed and compared. Then the best option for the primer pairs was selected based on the factors mentioned above.

Table 2: List of the ten primers that were designed by Geneious software version R8.1.9. The best primer pair is highlighted in green.

Primer pairs	Sequences	length	Interval	%GC	Tm	Hairpin Tm	Dimer Tm	Product size
1	TE-Forward GCCGAGCTGGAGGAGC TE-Reverse GCCCATCGCGGACAG	16 15	1-16 501-484	75.0 72.0	59.5 62.9	49.6 None	6.4 None	501
2	TE-Forward GCCGAGCTGGAGGAGC TE-Reverse CAGCAGGTGGAAGTAG CGC	16 19	1-16 499-481	75.0 63.2	59.5 61.1	49.6 None	6.4 None	499
3	TE-Forward GCCGAGCTGGAGGAGC TE-Reverse CAGCAGGTGGAAGTAG CG	16 18	1-16 499-482	75.0 61.1	59.5 57.8	49.6 None	6.4 None	499
4	TE-Forward GGCCCGCGGCTGATC TE-Reverse GGCCCGGACGAGCAG	15 15	51-65 550-536	80.0 80.0	60.3 59.4	49.6 None	19.3 2.8	500
5	TE-Forward GGCCCGCGGCTGATC TE-Reverse GCCCGGACGAGCAGC	15 15	51-65 549-535	80.0 80.0	60.3 60.2	49.6 None	19.3 None	499
6	TE-Forward CCCGCGGCTGATCTGC TE-Reverse GTGGCCCGGACGAGC	16 15	53-68 552-538	75.0 80.0	60.7 59.7	43.7 None	22.9 None	500
7	TE-Forward TGATCTGCGTCAGCACC C TE-Reverse GAACGGCTCGGTGGCC	18 16	61-78 562-547	61.1 75.0	59.7 60.5	50.4 47.9	2.7 5.1	502
8	TE-Forward TGATCTGCGTCAGCACC C TE-Reverse ACGGCTCGGTGGCCC	18 15	61-78 560-546	61.1 80.0	59.7 61.6	50.4 48.8	2.7 8.6	501
9	TE-Forward TGATCTGCGTCAGCACC C TE-Reverse	18	61-78 561-547	61.1 73.3	59.7 59.2	50.4	2.7	500

	AACGGCTCGGTGGCC	15				47.9	5.1	
10	TE-Forward TGATCTGCGTCAGCACCC	18	61-78	61.1 75.0	59.7 62.2	50.4	2.7	501
	TE-Reverse AACGGCTCGGTGGCCC	16	561-546			48.8	8.6	

The analysis of the results indicated that primer pair 1 is the optimal choice. This is because not only does it fulfill the required criteria, but upon examining the binding regions for this primer pair, it was observed that these regions were highly conserved. Consequently, designing a less degenerate pair was possible.

Then the sequences of the best primer pair were utilized to design degenerate primers with the objective of targeting a minimum of 50% of the polyenes present in the dataset.

Table 3: Degenerate PCR primer pair properties.

Sequence	Length	Tm	%GC	Hairpin Tm	Self-Dimer Tm	Degeneracy
TE-Forward 5`GCCGARCTGGAGSAGC3`	16	52.5 - 56.6	66.7-73.3	45.2	6.4	4
TE-Reverse 5`SCCCATCGCGGACAG3`	15	56.3 - 57.1	73.3	36.6	None	2

3.5. *In silico* PCR

The specificity of the primer pair against the *Streptomyces* genomes was assessed by conducting *in silico* PCR amplification.

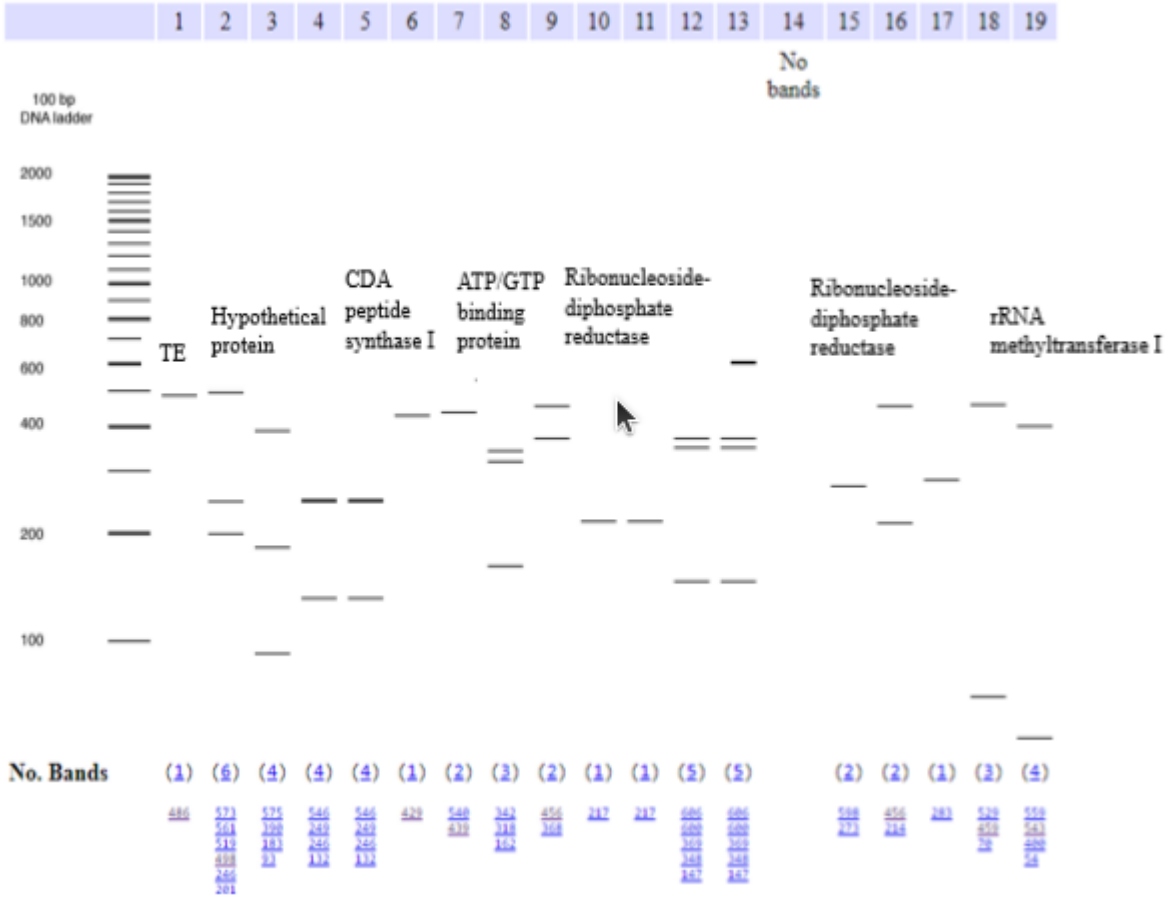


Figure 9: *In-silico* PCR gel illustrating the amplified genes.

The results of the analysis showed that the primer pair, which was designed to target the thioesterase gene, exhibited some degree of non-specific binding. This non-specific binding resulted in the amplification of the intended target as well as other genes with the same band size (around 500 bp).

5. Discussion:

Two hypotheses were formulated to identify a specific gene marker crucial for the biosynthesis of polyenes in Streptomyces and evolutionarily related actinomycetes. The first hypothesis was that Cytochrome P450 could serve as a reliable gene marker for polyene production. The multiple sequence alignment results revealed that the CYP sequences of the six-polyene BGCs exhibited shared regions where the amino acid residues are highly conserved. This suggests that these conserved regions may correspond to important catalytic sites or binding domains involved in the biosynthesis of polyene antifungal compounds. Furthermore, apart from the Candicidin and Rimocidin BGCs, the other 4 polyene BGCs possess two cytochrome P450 genes. The phylogenetic tree analysis revealed the presence of two distinct groups within the constructed tree. The possible explanation is that each group is involved in regiospecific oxidation. The first group in figure 6, including PimG, AmphN, NysN, PteD, RimG and

fscP are catalyzing the oxidation of the exocyclic methyl branch to the carboxyl group, whereas the second group including PimD, AmphL, PteC and NysL, are involved in oxidative modifications of the polyol segment. Example for the regiospecific oxidation by PimG and Pim D is illustrated in Figure 10.

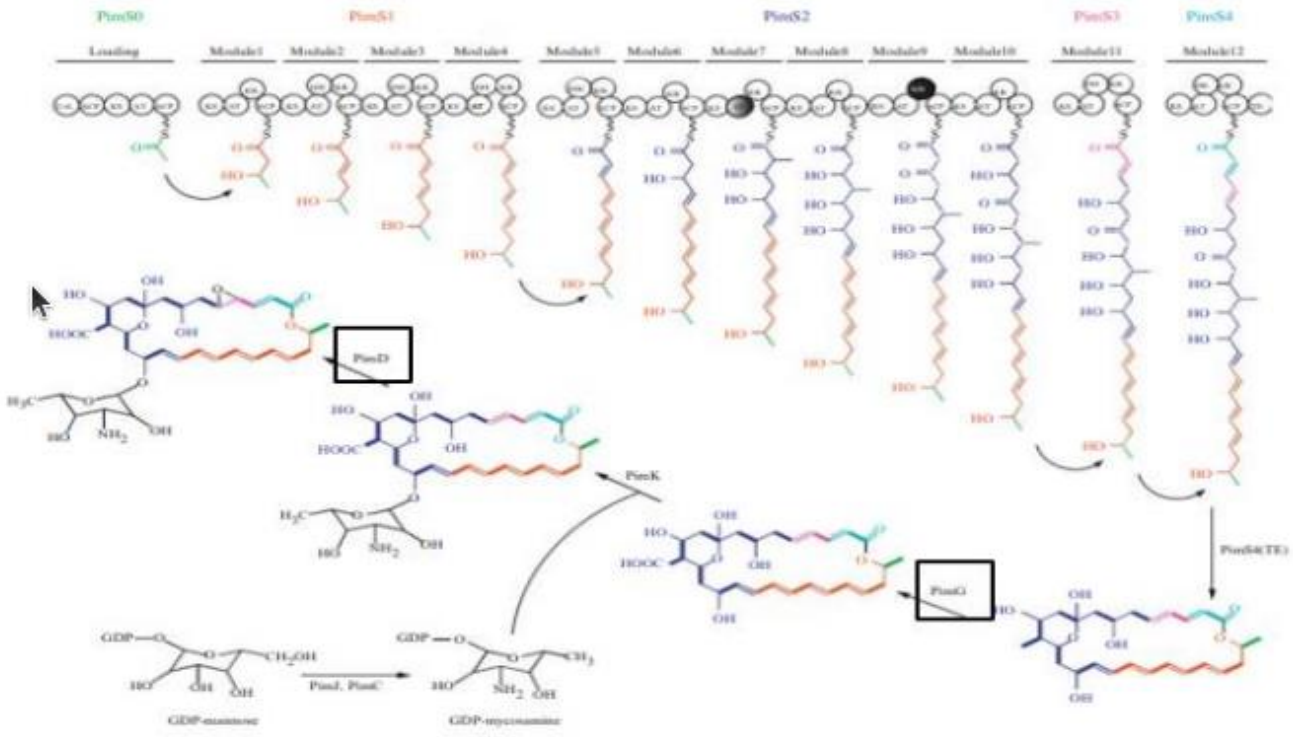


Figure 10: The biosynthesis of Pimaricin. The genes involved in regiospecific oxidation are marked with black boxes.

However, the results of searching for more polyene BGCs indicated that certain polyenes such as Reedsmycin, Linearmycin-A, Mediomycin-A, and Neomediomycin-B are devoid of cytochrome P450 genes. The biosynthetic gene clusters of those polyenes are listed in the appendix. This means that cytochrome P450 cannot serve as a reliable gene marker for polyene production as it is not a gene marker for all polyene biosynthetic gene clusters. Therefore, the hypothesis was rejected, and we have focused on alternative gene. The second hypothesis was that thioesterase type I can be used as an effective gene marker for identifying polyene-producing actinomycetes, particularly streptomycetes. Then the BGCs of 20 polyenes and 2 non-polyenes were included to the dataset.

The results of multiple sequence alignment of thioesterase type I within the biosynthetic gene clusters (BGCs) of 26 polyenes and 2 non-polyenes were compared to the aligned sequences of TE type I of the cyclic (macrolactone) polyenes in previous literature. The analysis revealed that the results in our study are similar to the results in the literature. Despite including additional

sequences of cyclic (macrolactone and macrolactam) and linear polyenes, they still exhibited conserved domains with some degree of variability in these regions. However, these variations in the amino acids that present within the conserved regions do not significantly impact the electric charges or overall geometry of the conserved domain for example, at position 47 (Figure 7), both aspartate (D) and glutamate (E) amino acids have negative charge. These domains share low sequence similarity with non-polyenes at the amino acid level, suggesting that these domains are specific for cyclization and offloading of the polyhydroxylated intermediates with continuous conjugated double bonds. As a result, the amino acid sequences of these domains are translated back to nucleotide sequences and used for designing the primers. The analysis of the domains at the nucleotide level indicated that those different amino acids, that have the same properties, are represented by different codons, and even identical amino acids have multiple possible codons. This poses a challenge when designing primers for thioesterase type I.

Based on the circumstances, 10 primer pairs could be designed. The results showed that primer pair 1 is the best option as both forward and reverse primers have the optimum GC %. This indicates the stability and specificity of primer binding to the target DNA sequence. Furthermore, the melting temperatures (Tm) of the primers are close to each other. Having closely matched Tm values for the forward and reverse primers is important so that they are likely to anneal to the target DNA sequence at a similar temperature during PCR amplification.

Moreover, formation of dimers and hairpins are possible only for the forward primer, unlike primer pairs 7-10. This difference in behaviour lowers the probability of the primers annealing effectively. The most important advantage of this primer pair over the others is that the analysis of the binding regions for this primer pair revealed a high degree of conservation for polyenes thus, designing less degenerate primer pair.

Then the sequences of the best primer pair were used to design the degenerate primers. Degenerate primers provide flexibility in primer design, enabling them to bind to target regions with slight variations. However, degeneracy of the primers was kept low to ensure specific binding. In silico PCR was used to evaluate the specificity of the degenerate primer pairs by testing these primers against 19 Streptomyces strains. The results show that not only the thioesterase type I gene was amplified, but also other genes. The lack of specificity observed with this primer pair suggests that the sequences of the primer pair may have similarity with other genes within the genome. Moreover, the relatively short length of the primer pair (15 bp) may increase the probability of non-specific binding to other regions of the genome as shorter primers have a higher likelihood of encountering sequences with partial complementarity elsewhere in the genome, leading to unintended amplification during PCR. Therefore, the hypothesis was rejected. To enhance primer specificity and minimize non-specific amplification, it is recommended to design longer primers (18-30 bases) that are more unique to the target gene sequence or to use nested PCR to ensure specific binding. Due to time constraints, the primer pair was only tested using *in silico* PCR. However, it is preferable to conduct experimental testing in order to facilitate further analysis, such as sequencing the amplicons.

6. Conclusion:

The analysis of cytochrome P450 sequences in the polyene biosynthetic gene clusters revealed two different groups. Each group leads to regiospecific oxidation. However, some polyene biosynthetic gene clusters did not have cytochrome P450 enzymes, making it an unreliable gene marker. Furthermore, the thioesterase type I gene was found in all polyene BGCs, but was not effective as a gene marker due to the lack of specificity of the designed primer pair. Therefore, testing the primer pair in the lab and sequencing the products can provide additional information about the amplified regions.