Strain Prioritization for Natural Product Discovery by a High-Throughput Real-Time PCR Method

Hindra,[1‡] Tingting Huang,[1‡] Dong Yang,[1‡] Jeffrey D. Rudolf,[1‡] Pengfei Xie,[1] Guangbo Xie,[1] Qihui Teng,[1] Jeremy R. Lohman,[1] Xiangcheng Zhu,[2,3] Yong Huang,[2] Li-Xing Zhao,[4] Yi Jiang,[4] Yanwen Duan,[2,3] and Ben Shen[1,5,6*]*

[1]Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458, USA; [2]Xiangya International Academy of Translational Medicine, Central South University, Changsha, Hunan 410013, China; [3]Hunan Engineering Research Center of Combinatorial Biosynthesis and Natural Product Drug Discovery, Changsha, Hunan 410329, China; [4]Yunnan Institute of Microbiology, Yunnan University, Kunming, Yunnan 650091, China; [5]Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL 33458, USA; and [6]Natural Products Library Initiative, The Scripps Research Institute, Jupiter, FL 33458, USA.

[‡]These authors contributed equally

*****To whom correspondence should be addressed: Ben Shen, The Scripps Research Institute, 130 Scripps Way, #3A1, Jupiter, FL 33458; Tel: (561) 228-2456, Fax: (561) 228-2472; E-mail: shenb@scripps.edu or Yanwen Duan, E-mail: ywduan66@yahoo.com

Supporting Information (SI)

Table S1. Primers used in this study.

Primer	Nucleotide Sequence (5'-3')	Function (Reference)
ptmn1-S	CCGGGCTGGACATCCGGGCGGAC	PCR targeting *ptm/ptnT1*
ptmn1-AS	GGATGGCGCAGAGCAGGAGGTC	PCR targeting *ptm/ptnT1*
ptmn2-S	CTGCTCCCCCGCCGCCACC	PCR targeting *ptm/ptnT2*
ptmn2-AS	CGTAGTACGGCGAGGCGTGC	PCR targeting *ptm/ptnT2*
ptm3-S	TATCTGCTCGACGGCAGGCTCGAC	PCR targeting *ptmT3*
ptm3-AS	TTGGCCCAGGTCCGCAGATCGTT	PCR targeting *ptmT3*
ptmnT4mix-S	CTGBTSCACGACGAYVTSATGGAC	PCR targeting *ptm/ptnT4*
ptmnT4mix-AS	GCCSAKBABGTCGTCSRYVWDYTGGAA	PCR targeting *ptm/ptnT4*
16SrRNA_for	AGAGTTTGATCCTGGCTCAG	Phylogenetic Analysis
16SrRNA_rev	ACGGCTACCTTGTTACGACTT	Phylogenetic Analysis
recAfor	TAATACGACTCACTATAGGGCCGCRCTCGCACARATTGAACG	Phylogenetic Analysis
recArev	GCTAGTTATTGCTCAGCGGCGTCGGGGTTGTCCTTSAGGAAG	Phylogenetic Analysis
rpoB-2	CATCGACCACTTCGGCAAC	Phylogenetic Analysis
ActRpoB3303R	GAANCGCTGDCCRCCGAACTG	Phylogenetic Analysis[1]
trpBfor	TAATACGACTCACTATAGGGGCGCGAGGACCTGAACCACAC	Phylogenetic Analysis
trpBrev	GCTAGTTATTGCTCAGCGGCATGGCCGGGATGATGCCC	Phylogenetic Analysis
ptmT1for	GACATCGAGGGGCATGGGAAGG	Sequencing
ptmT1rev	GTCAATCCGGAGACCGGGGTAC	Sequencing
ptmT2for	CTTGTGGGCGTCCAGGAAGGAG	Sequencing
ptmT2rev	CCACGAACGACAACAACAGTTGCG	Sequencing
ptmT2mid	GATGTGTGCGTTGGTGCTGG	Sequencing
ptmT3for	CATCTGGCGCGACAACCGCATTG	Sequencing
ptmT3rev	CATCGCGTCCTTTGATCGGGAGG	Sequencing
ptmT4for2	CGGAACACCGCCGCGTAG	Sequencing
ptmT4rev	GAGCACCACGTCGCCGTG	Sequencing
ptmR1for	GTCTTCGGCAGCCGGCTCTC	Sequencing
ptmR1rev	CGACTCAACAGGGCGTAAAGGTGC	Sequencing
ptmRtgtF	CCCGCCGGAATCGGCCCTGATGGAGCAGTTCGGCATTTCCATTCCGGGGATCCGTCGACC	λRED-mediated PCR targeting replacement of *ptmR1*
ptmRtgtR	TCGAGGAGTTCCAGACGGGTATTGGCGCCGCTCGCATTCAATGTAGGCTGGAGCTGCTTC	λRED-mediated PCR targeting replacement of *ptmR1*
ptmRidF	CCTGATGGAGCAGTTCGG	Δ*ptmR1* PCR confirmation
ptmRidR	GGAGTTCCAGACGGGTATTG	Δ*ptmR1* PCR confirmation

Table S2. Strains and plasmids used in this study.

Strain/Plasmid	Genotype, Description	Source (Reference)
E. coli DH5α	*E. coli* host for cloning	Life Technologies
E. coli XL1-Blue MRF′	*E. coli* host for library construction	Agilent
E. coli ET12567/pUZ8002	Methylation-deficient *E. coli* host for intergeneric conjugation; contains pUZ8002, a nontransmissible *oriT* mobilizing plasmid	2
E. coli BW25113/pIJ790	*E. coli* host for PCR targeting	3
S. platensis MA7327	Wildtype PTM/PTN producer	Merck[4,5]
S. platensis MA7339	Wildtype PTN producer	Merck[6,7]
S. platensis SB12001	PTM/PTN overproducing strain	8
S. platensis SB12002	PTM/PTN overproducing strain	8
S. platensis SB12600	PTN overproducing strain	9
S. platensis CB00739	PTM-PTN strain hit	This study
S. platensis CB00765	PTM-PTN strain hit	This study
S. platensis CB00775	PTM-PTN strain hit	This study
S. platensis CB00789	PTM-PTN strain hit	This study
S. platensis CB02289	PTM-PTN strain hit	This study
S. platensis CB02304	PTM-PTN strain hit	This study
S. platensis SB12026	CB00739 with *ptmR1* replaced with *aac(3)IV*+oriT cassette	This study
S. platensis SB12027	CB00765 with *ptmR1* replaced with *aac(3)IV*+oriT cassette	This study
S. platensis SB12028	CB00775 with *ptmR1* replaced with *aac(3)IV*+oriT cassette	This study
SuperCos1	Vector for the construction of cosmid libraries	Agilent
pIJ773	Plasmid containing the apramycin resistance cassette (*aac(3)IV*+oriT)	3
pBS12031	Cosmid 18H9 from CB00739 cosmid library, containing partial *ptm* gene cluster	This study
pBS12032	Cosmid 18H10 from CB00765 cosmid library, containing partial *ptm* gene cluster	This study
pBS12033	Cosmid 10F8 from CB00775 cosmid library, containing partial *ptm* gene cluster	This study
pBS12034	pBS12031 with *ptmR1* replaced with *aac(3)IV*+oriT by PCR targeting	This study
pBS12035	pBS12032 with *ptmR1* replaced with *aac(3)IV*+oriT by PCR targeting	This study
pBS12036	pBS12033 with *ptmR1* replaced with *aac(3)IV*+oriT by PCR targeting	This study

Figure S1. Genetic organization of the PTM-PTN dual biosynthetic gene cluster in *S. platensis* MA7327 and the PTN biosynthetic gene cluster in *S. platensis* MA7339, and primer design targeting the genes encoding the four diterpene-related synthases *ent*-atiserene synthase (T1), *ent*-copalyl diphosphate synthase (T2), *ent*-kaurene synthase (T3), and geranylgeranyl diphosphate synthase (T4).[10] Black arrows depicting the relative locations of the primers with the predicted sizes for each of the products indicated.

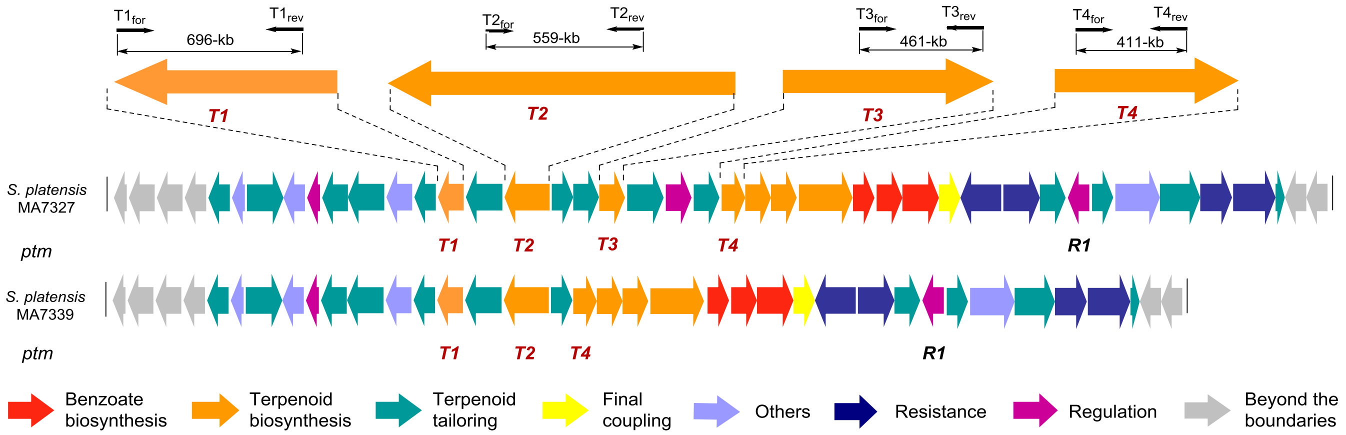


Figure S2. T1, T2, T3, and T4 sequence analysis for primer design. **(a)** T1, **(b)** T2, **(c)** T3, and **(d)** T4. For T1, T2, and T4, amino acid sequences were aligned using Clustal 2[11] and conserved sequences were highlighted using Boxshade. Consensus sequences highlighted in black were chosen for primer design, which was based on the associated nucleotide sequences. Accession codes of proteins used in the alignment: PtmT1 (ACO31274), PtnT1 (ADD83014), Sko3988_Orf2 (BAD86797), Swt1.2 (AEV45183), Swt2.2 (AEW22921), PtmT2 (ACO31276), PtnT2 (ADD83015), PtmT3 (ACO31279), PtmT4 (ACO31283), PtnT4 (ADD83016), Kgris_Orf1 (BAB07816), PlaT4 (ABB69754), Bra2 (BAG16276), and Sko3988_Orf4 (BAD86799).

a

```
PtmT1      61   IVAIISQNPLNAGLDIRADTHTSGK...(200)...RKHSLSLAMGGDLLLCAIPLLAL  308
PtnT1      61   IVAIISQNPLNAGLDIRADTHTSGK...(200)...RKHSLSLAMGGDLLLCAIPLLAL  308
consensus  61   *************************           ***********************
                     Sense primer region              Antisense primer region
```

```
Sense primer region:
Conserved residues                                   G  L  D  I  R  A  D
ptmT1      181  ATCGTCGCGATCATCTCCCAGAACCCGCTCAACGCCGGGCTGGACATCCGGGCGGACACC
ptnT1      181  ATCGTCGCGATCATCTCCCAGAACCCGCTCAACGCCGGGCTGGACATCCGGGCGGACACC
                                                  CCGGGCTGGACATCCGGGCGGAC
```

```
Antisense primer region:
Conserved residues                    D  L  L  L  C  A  I
ptmT1      868  CTCTCGCTGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCCCGCTGCTCGCGCTCTGA
ptnT1      868  CTCTCGCTGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCCCGCTGCTCGCGCTCTGA
                                     GACCTCCTGCTCTGCGCCATCC
```

b

```
Sko3988_Orf2   197  GGAVGASPAATSAW...(153)...GWLVEQQQPDGRWMDKWHASPYYATACGAA   393
Swt2.2         196  GGAVGASPAATAAW...(153)...GWLVEQQEPDGRWMDKWHASPYYATACGAA   392
Swt1.2         211  GGAVGASPAATAAW...(153)...DWLVDQQEPDGRWMDKWHASPYYATACGAA   407
PtmT2          219  NGSVGCSPAATAAW...(154)...DWLLDNQLPDGSWMDKWHASPYYATACCAL   416
PtnT2          219  NGSVGCSPAATAAW...(154)...DWLLDNQLPDGSWMDKWHASPYYATACCAL   416
consensus      231  .*.**.*****.**           .**...* ***.*****************.*.   430
                   Sense primer region              Antisense primer region
```

```
Sense primer region:
Conserved residues       S  P  A  A  T
sko3988_orf2   603  AGCCTCCCCCGCCGCCACCTCCGCCTGGCTGGGCGATCCGCCTCATACGGACGC
swt2_orf2      600  AGCCTCCCCCGCCGCCACCGCCGCCTGGCTGGGCGATCCGCCGCACACGGACGC
swt1_orf2      645  TGCCTCCCCCGCCGCCACCGCCGCCTGGCTGGGCGAACCGCCGCACACGGACGC
ptmT2          669  CTGCTCCCCCGCCGCCACCGCCGCCTGGCTGGGCGCCTCGCCCGATCCGGCCGC
ptnT2          669  CTGCTCCCCCGCCGCCACCGCCGCCTGGCTGGGCGCCTCGCCCGATCCGGCCGC
                     CGCTCCCCCGCCGCCACC
```

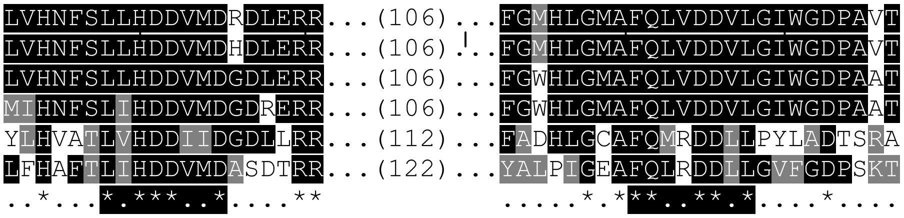
```
Antisense primer region:
Conserved residues                                     H  A  S  P  Y  Y
sko3988_orf2  1106  AGCAGCAGCCGGACGGCCGGTGGATGGACAAGTGGCACGCCTCTCCGTACTACG
swt2_orf2     1103  AGCAGGAGCCGGACGGCCGGTGGATGGACAAGTGGCACGCCTCTCCGTACTACG
swt1_orf2     1148  AGCAGGAGCCGGACGGCCGCTGGATGGACAAGTGGCACGCCTCCCCGTACTACG
ptmT2         1175  ACCAGCTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
ptnT2         1175  ACCAGCTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
                                                      GCACGCCTCGCCGTACTACG
```

Figure S2. (continued)

c

```
PtmT3  73   L  Y  L  L  D  G  R  L  D ...(137)... A  N  D  L  R  T  W  A  K   227
      217  ctctatctgctcgacggcaggctcgac...(411)...gccaacgatctgcggacctgggccaag
             Sense primer region                   Antisense primer region
```

d



```
PtmT4           96  LVHNFSLLHDDVMDRDLERR...(106)...FGMHLGMAFQLVDDVLGIWGDPAVT  246
PtnT4           96  LVHNFSLLHDDVMDHDLERR...(106)...FGMHLGMAFQLVDDVLGIWGDPAVT  246
Kgris_Orf1      96  LVHNFSLLHDDVMDGDLERR...(106)...FGWHLGMAFQLVDDVLGIWGDPAAT  246
PlaT4          106  MIHNFSLIHDDVMDGDRERR...(106)...FGWHLGMAFQLVDDVLGIWGDPAAT  256
Bra2            77  YLHVATLVHDDIIDGDLLRR...(112)...FADHLGCAFQMRDDLLPYLADTSRA  233
Sko3988_Orf4   120  LFHAFTLIHDDVMDASDTRR...(122)...YALPIGEAFQLRDDLLGVFGDPSKT  286
consensus      121  ..*...*.***..*....**           .....*.***..**.*....*....
                     Sense primer region          Antisense primer region
```

<u>Sense primer design:</u>

```
Conserved residues                                   L  X  H  D  D V/I M/I D
ptmT4         271  GCCGCGGCCGTCGAACTGGTCCACAACTTCTCGCTGTTGCACGACGACGTCATGGACCGC
ptnT4         271  GCCGCCGCCGTCGAACTCGTCCACAACTTCTCGCTGTTGCACGACGACGTCATGGACCAC
kgris_orf1    270  GCCGTCGCGGTCGAGTTGGTGCACAACTTCTCCCTGCTCCACGACGACGTCATGGACGGT
plaT4         279  GGCGCCGCCGTGGAGATGATCCATAACTTCTCTCTCATCCACGACGACGTCATGGACGGT
bra2          204  GCGGTGGCGCTGGAGTACCTGCACGTGGCCACCCTGGTGCACGACGACATCATCGACGGC
sko3988_orf4  299  GCGGCCTCGCTCGAACTCTTCCACGCCTTCACCCTCATCCATGACGACGTGATGGACGCC
                                                    CTGBTSCACGACGAYVTSATGGAC
```

<u>Antisense primer design:</u>

```
Conserved residues              F  Q  X  X  D  D  X  L
ptmT4         675  CCTGGGCATGGCGTTCCAGTTGGTCGACGACGTGCTCGGCATCTGGGGCGACCCGGCCGT
ptnT4         675  CCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGCATATGGGGCGACCCGGCCGT
kgris_orf1    674  TCTGGGGATGGCGTTCCAGCTGGTGGATGACGTCCTGGGCATCTGGGGCGACCCCGCGGC
plaT4         683  GCTCGGACTCGCCTTCCAGATCGCCGACGACATCCTGGGCATCTGGGGCGACCCCCGGCG
bra2          636  TCTCGGCTGCGCCTTCCAGATGCGCGACGATCTGCTGCCGTATCTGGCCGACACCAGCCG
sko3988_orf4  751  GATCGGAGAGGCCTTCCAACTGCGCGACGACCTGCTCGGCGTCTTCGGCGACCCGTCCAA
                                TTCCARHWBRYSGACGACVTVMTSGGC
```

Figure S3. Sequence comparisons of *T1*, *T2*, *T3*, and *T4* PCR hit fragments. **(a)** T4, **(b)** T2, **(c)** T3, and **(d)** T1. Relative to *ptmT1*, *ptmT2*, *ptmT3*, and *ptmT4* sequences, the percent identities of hit sequences was between 97-98, 63-98, 96-97, and 38-98, respectively.

a

```
ptmT4         1  CTGTTGCACGACGACGTCATGGACCGCGACCTGGAG-CGGCGCGGCCGGCCCAC------
CB00739       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGCGGCCGGCCCAC------
CB00765       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGCGGCCGGCCCAC------
CB00775       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGCGGCCGGCCCAC------
CB00789       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGCGGCCGGCCCAC------
CB02289       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGTGGCCGGCCCAC------
CB02304       1  CTGTTGCACGACGACGTCATGGACCGCGATCTGGAG-CGGCGCGGCCGGCCCAC------
CB00028       1  CTGGTGCACGACGATCTGATGGACCACGACACGACC-CGGCGCCACCGTGCCAC------
CB01059-1     1  CTGTTGCACGACGACCTGATGGACACCTCCGACACC-CGGCGCGGCCGTCCGACCGTGCA
CB01059-2     1  CTGGTGCACGACGACGTGATGGACGGGTC-GTCACTGCGTCGCGGGTCGCCGGCGCTCCA
CB00830       1  CTGGTGTACGACGACGTGATGGACGCCTCCGACACC-CGGCGCGGCCGTCCGACCGTCCA

ptmT4        54  --------------------------GGTCTGGAG---------------CAAGTTCGGC
CB00739      54  --------------------------GGTATGGAG---------------CAAGTTCGGC
CB00765      54  --------------------------GGTATGGAG---------------CAAGTTCGGC
CB00775      54  --------------------------GGTATGGAG---------------CAAGTTCGGC
CB00789      54  --------------------------GGTATGGAG---------------CAAGTTCGGC
CB02289      54  --------------------------GGTCTGGAG---------------CAAGTTCGGC
CB02304      54  --------------------------GGTCTGGAG---------------CAAGTTCGGC
CB00028      54  --------------------------GGCGTGGAC---------------GGTGTTCGGT
CB01059-1    60  CCGCG----CGCTGGAGACG------GGTGCGGACCCCACGGCCGCCG-AGCGGTTCGGC
CB01059-2    60  CGTCGAGTTCGCCGACAACCACGCACGGTCCGGTATGCGGGGAGACCACAACGCTCACGG
CB00830      60  CCGCG----CGCTGGAGACG------GGTGCGGACCCCACGACCGCCG-AGCGGTTCGGC

ptmT4        73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGAATGTTCGA
CB00739      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGCATGTTCGA
CB00765      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGCATGTTCGA
CB00775      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGCATGTTCGA
CB00789      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGCATGTTCGA
CB02289      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGGATGTTCGA
CB02304      73  ACCCCC-GCGGCGATCCTGGCGGGCGACATCCTGCTGGCGCGCGCCTGCGGGATGTTCGA
CB00028      73  ATCAAC-GCCGCGATCCTGGCGGGCGACGCGCTGCAGGCGTTGGCCGTGGACGTCCTGGC
CB01059-1   109  ACCAACAGCG-CGATCCTGCTCGGCGACCTGGTCATGGTGTGGTCGGACGAGCTGCTGTA
CB01059-2   120  GCGGGCAGCGGCGATCCTGGCCGGCGACCTGGCCCTGGCCTGGGCGGACGATCT--TGT-
CB00830     109  ACCAACAGCG-CGATCCTGCTCGGCGACCTGGTCATGGTGTGGTCGGACGAGCTGCTGTA

ptmT4       132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTCATCGACGCGAT
CB00739     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB00765     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB00775     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB00789     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB02289     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB02304     132  CGAGGCCTC---CGGCCAC-CAGG-G-CTGG--GC-GACCAAGGCCCTGATCGACGCGAT
CB00028     132  CGGCTCCGGA--CATCCACGTGCG---CTGGAGGC-GATCCGGACGCTGGGCGACGCGGT
CB01059-1   168  CTCCAGCGGCCTCGACCAG-CAGC-AACTGGTCGC-CGTTCGACCGCTGTTGGACGCCAT
CB01059-2   177  CGCCGACATCCTGAACGGATCAGCGAACCGCGAGCACGTTCAGAGGCTGTGGCGTGCAAT
CB00830     168  CTCCAGCGGCCTCGACCAG-CAGC-AACTGGTCGC-GGTCCGCCCGCTGCTGGACGCCAT
```

Figure S3a. (continued)

```
ptmT4       183  C---GCCGAGCTGGCCGCGGGCCAGATGGCGGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB00739     183  C---GCCGAGCTGGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB00765     183  C---GCCGAGCTGGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB00775     183  C---GCCGAGCTGGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB00789     183  C---GCCGAGCTGGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB02289     183  C---GCCGAGCTGGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB02304     183  C---GCCGAGCTTGCCGCGGGCCAGATGGCCGACCT-CGCGCTGGAGCGCCGCGC--CAC
CB00028     186  G---GTGGCGTTGCTCGACGGGCAGGCCTACGACCTCGACTTCGAGACGCGGGCCGACGT
CB01059-1   225  GCGCACCGAGGTGATGATCGGTCAGTACCGGGACCTGCTGGCC--ACCGGGGAGCC-CAA
CB01059-2   237  GCGCGGTGAAATGGTCGCAGGCCAGTATCTGGACATGCGCGCCCAAGCGGACGGCT-CAT
CB00830     225  GCGCACCGAGGTGATGATCGGTCAGTACCGGGACCTGCTGGCC--ACCGGGGAGCC-CAA

ptmT4       237  GGTCACCCTGGAAGAGGCCCTCACCGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB00739     237  GGTGACCCTGGAAGAGGCCCTCATGGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB00765     237  GGTGACCCTGGAAGAGGCCCTCATGGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB00775     237  GGTGACCCTGGAAGAGGCCCTCATGGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB00789     237  GGTGACCCTGGAAGAGGCCCTCATGGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB02289     237  GGTGACCCTGGAAGAGGCCCTCATGGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB02304     237  GGTGACCCTGGAAGAGGCCCTCACCGTCTCCGAGCAGAAGACCGCGGCGC-TGCTGCGCT
CB00028     243  CCAGGTGCCCGAGGTGGAGCGGATGGCGCGCGGT---AAGACCGGCGCTC-TGCT-CAGT
CB01059-1   282  CGCGGACGTCGAACGGTCACTGGCGGTGATCCGCTACAAGACGGCCAAGTACACCGTGGA
CB01059-2   296  CCA--CTACGGATCGGGCTTTGCGTATCGCCTGTCTCAAAACGGCTCTCTACACGGTCGA
CB00830     282  CGCGGACGTCGAGCGGTCGCTGGCGGTGATCCGCTACAAGACGGCCAAGTACACCGTGGA

ptmT4       296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGACGGGACCAGCCGCCGCTT
CB00739     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGGGCGCCCGAGGGGACCAGCCGCCGCTT
CB00765     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGAGGGGACCAGCCGCCGCTT
CB00775     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGAGGGGACCAGCCGCCGCTT
CB00789     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGAGGGGACCAGCCGCCGCTT
CB02289     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGACGGGACCAGCCGCCGCTT
CB02304     296  GCGCCTGCA-CGC-TGGGCGCGGACTCGTCGGCGCGCCCGACGCGACCAGCCGCCGCTT
CB00028     298  GCCTCAACGGCTC-TGGGCGCGTTGTACGGCGACG-GCACGGACGAGCAGATCGAGGCGA
CB01059-1   342  GCGGCCGCTGCACATCGGCGCGGCCCTCGCCGGAGCGGACCGGTCC-CTGCTGGACGTCT
CB01059-2   354  ACGACCCCTGGCTCTCGGCGCGGCACTCGCCGGAGCCGAC-GGCCGGACAACGTCCGCTC
CB00830     342  GCGGCCGCTGCACATCGGCGCGGCCCTGGCCGGAGCGGACCAGTCC-CTGCTGGACGTCT

ptmT4       354  CG-GCGCCTTCGGTATGCACCTGGGCATGGCGTTCCAGTTGGTCGACGACGTGCTCGGC
CB00739     354  CG-GTGCCTTCGGTATGCACCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGC
CB00765     354  CG-GTGCCTTCGGTATGCACCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGC
CB00775     354  CG-GTGCCTTCGGTATGCACCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGC
CB00789     354  CG-GTGCCTTCGGTATGCACCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGC
CB02289     354  CG-GCGCCTTCGGCATGCACCTGGGCATGGCGTTCCAACTGGTCGACGACGTACTCGGC
CB02304     354  CG-GCGCCTTCGGCATGCACCTGGGCATGGCGTTCCAGCTGGTCGACGACGTACTCGGC
CB00028     356  TGCGCGGGTACGGCGCGGAGCTCGGCCTGGCGTTCCAGCACACGGACGACCTCCTCGGC
CB01059-1   401  GCACCGCCTACGCCCTCCCGATCGGGAGGCCTTCCAATAGACCGACGACCTACTGGGC
CB01059-2   413  TCTGCTCCGCGGGTCGCAGCGCAGGTATCGCCTTCCAGTTGACCGACGACCTCATCGGC
CB00830     401  GCACCGCCTACGCCCTCCCGATCGGCGAGGCCTTCCAAATCATGGACGACATACTCGGC
```

Figure S3. (continued)

b

```
ptmT2       1   CGCCCGATCCGGCCGCACCGG------GCGTCGCCTACCTCCGTGACGTCCAGGCGCGGT
CB02304     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00789     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00765     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00739     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB02289     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00775     1   CCCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00028     1   CGCCCGATCCGGCCGCGCCGG------GCGTCGCCTATCTCCGTGACGTCCAGGCGCGGT
CB00830     1   CGCCGCACACGGACGCCGTGAAGGAGTGCCTGGAGTATCTGCGTCAGACCCAGGCACGTC
CB01059     1   CGCCGCACACGGACGCCGTGAAGGAGTGCCTGGAGTATCTGCGGCAGACCCAGGCACGTC

ptmT2      55   TCGGCGGTCCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB02304    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00789    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00765    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00739    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB02289    55   TCGGCGGGCCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00775    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00028    55   TCGGCGGACCGGTGCCCTCCATCACCCCGATCGTCTACTTCGAGCAGGCGTGGGTCCTCA
CB00830    61   ATGGCGGCCCGGTCTCGGGCATCACCTCGATCAACTACTTCGAACTGGCCTGGGTGGTGA
CB01059    61   ATGGCGGCCCGGTCTCGGGCATCACCTCGATCAACTACTTCGAACTGGCCTGGGTGGTGA

ptmT2     115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB02304   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GTCTC
CB00789   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB00765   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB00739   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB02289   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTGCTCGACA-GCCTC
CB00775   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB00028   115   ACTCGCTGGCCGCCTCCGGCCTGCGCTACGAGGCCCCGGCCGCGCTCCTCGACA-GCCTC
CB00830   121   CGGCGTTGTCCGGCTCCGGCCTGGACGTGGACTTCCCCGCCCAGATCGCGGACACGCTGC
CB01059   121   CGGCGTTGTCCGGCTCCGGCCTGGACGTGGACTTCCCCGCCCAGATCGCGGACACGCTGC

ptmT2     174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGTCTGCCGAGCGACTCCGAC
CB02304   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00789   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00765   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00739   174   GAAGCGGGTCTCACCGACGAGGGCATAGCCGCCGCCCCCGGTCTGCCGAGCGACTCCGAC
CB02289   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00775   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00028   174   GAAGCGGGTCTCACGGACGAGGGCATAGCCGCCGCCCCCGGGCTGCCGAGCGACTCCGAC
CB00830   181   GGACCGCGCTGGGCGCGAACGGAC-TGTCCGCTGGCCCGGGGCTCCCGGCCGACTCCGAC
CB01059   181   GGACCGCGCTGGGCGGGAACGGAC-TGTCCGCCGGCCCGGGGCTCCCGGCCGACTCCGAC
```

Figure S3b. (continued)

```
ptmT2    234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB02304  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB00789  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB00765  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB00739  234  GACACCGCCGCCGTTCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB02289  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCGGACAGC
CB00775  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB00028  234  GACACCGCCGCCGTCCTCTTCGCCCTGGCGCAGCACGGCAGGACGCACCGCCCCGACAGC
CB00830  240  GACACGTCGGCCGCCCTGCACGCCCTCGACCTCCTCGGGAAGCCGGAATCGGTCGACTGC
CB01059  240  GACACGTCGGCCGCCCTGCACGCCCTCGATCTCCTCGGGAAGCCGGAATCGGTCGACTGC

ptmT2    294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCTTCC
CB02304  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCATCGAGCGCACCCCGTCC
CB00789  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCGTCC
CB00765  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCTTCG
CB00739  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCTTCG
CB02289  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCCTCC
CB00775  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCTTCG
CB00028  294  CTGATGCACTTCCGCCGGGACGGCTACTTCTCCTGCTTCGGCGTCGAGCGCACCCCTTCG
CB00830  300  CTGTGGGATTACGACACCGGCCTGTACTTCACCTGCTTCCCCAAGGAACGCACCCCTTCC
CB01059  300  CTGTGGGATTACGACACCGGCCTGTACTTCACCTGCTTCCCGAAGGAACGCACCCCTTCC

ptmT2    354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB02304  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00789  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00765  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00739  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB02289  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00775  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00028  354  ACCAGCACCAACGCACACATCCTCGAAGCCCTCGGCCATCACGTCACGGTGCGCCCCGAC
CB00830  360  ACCTCGACGAACGCCCACATCCTCGTCGCCCTGGCGGATCGC--CGCGGA----CAGGGC
CB01059  360  ACCTCGACGAACGCCCACATCCTCGTCGCCCTGGCGGATCGC--CGCGGA----CAGGGC

ptmT2    414  GACGCGGGACGCTATGGCGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB02304  414  GACGCGGGACGCTACGGCGCGGAGATCCGGATGATCGGCGACTGGCTGCTGGACAACCAG
CB00789  414  GACGCGGGACGCTACGGGGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB00765  414  GACGCGGGACGCTACGGGGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB00739  414  GACGCGGGACGCTACGGGGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB02289  414  GACGCAGGACGCTATGGCGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB00775  414  GACGCGGGACGCTACGGGGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB00028  414  GACGCGGGACGCTACGGGGCGGAGATCCGGATGATCAGCGACTGGCTGCTGGACAACCAG
CB00830  414  GACACGCGCTGCGACGACGCCGCCGAGCGGGTG---GGCGGCTGGCTCGTCGAGCAGCAG
CB01059  414  GACACGCGCTGCGACGAGGCCGCCGAGCGAGTG---GGCGGCTGGCTCGTCGAGCAGCAG

ptmT2    474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB02304  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00789  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00765  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00739  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB02289  474  CTGCCGGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00775  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00028  474  CTGCCCGACGGCAGCTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB00830  471  GAGCCGGACGGCCGGTGGATGGACAAGTGGCACGCCTCGCCGTACTACG
CB01059  471  GAGCCGGACGGCCGGTGGACGGACAAGTGGCACGCCTCGCCGTACTACG
```

Figure S3. (continued)

c

```
ptmT3      1   TATCTGCTCGACGGCAGGCTCGACGACTTCGAGCACTACGGCACCCGGCCCGAGGACGTC
CB00739    1   TATCTCCTCGACGGCCGGCTCGACGACTCCGAGCACTACGGCACCCGGCCCGAGGACGTC
CB00765    1   TATCTCCTCGACGGCCGGCTCGACGACTCCGAGCACTACGGCACCCGGCCCGAGGACGTC
CB00775    1   TATCTCCTCGACGGCCGGCTCGACGACTCCGAGCACTACGGCACCCGGCCCGAGGACGTC
CB00789    1   TATCTCCTCGACGGCCGGCTCGACGACTCCGAGCACTACGGCACCCGGCCCGAGGACGTC
CB02289    1   TATCTGCTCGACGGCCGGCTCGACGACTGCGAGCGCTACGGCACCCGGCCCGAGGACGTC
CB02304    1   TACCTGCTCGACGGCAGGCTCGACGACTCCGAGCACTACGGCACCCGGCCCGAGGACGTC

ptmT3     61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGCGAGCCGACGACTTC
CB00739   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGAGAGCCGGCGACTTC
CB00765   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGAGAGCCGGCGACTTC
CB00775   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGAGAGCCGGCGACTTC
CB00789   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCCAGAGCCGGCGACTTC
CB02289   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGAGAGCCGGCGACTTC
CB02304   61   GGCCGGCGGGTGATCGCCGTACTCCGCGGCGGCCGTGCCGAGGCGAGAGCCGGCGACTTC

ptmT3    121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGACCCGGGACGGCTGCTGCGGG
CB00739  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG
CB00765  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG
CB00775  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG
CB00789  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG
CB02289  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG
CB02304  121   TTCGAGACCTCGCTCGCCGCACTGGTCGAGGAGCTGCGGAGGCGGGACGGCTGCTGCGGG

ptmT3    181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGTGCGGGGGTGCGCCAGGCG
CB00739  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGCGCGGGGGTGCGGCAGGCC
CB00765  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGCGCGGGGGTGCGGCAGGCC
CB00775  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGCGCGGGGGTGCGGCAGGCC
CB00789  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGCGCGGGGGTGCGGCAGGCC
CB02289  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGTGCGGGGGTGCGGCAGGCG
CB02304  181   CTGCTGGAGCGGTTCGTGCTGCGGCTGGTGGACGGGGTCCGTGCGGGGGTGCGGCAGGCG

ptmT3    241   GTGCTCAGTCGGCGGATCGCCGAGGGCGCGGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB00739  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACCGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB00765  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACCGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB00775  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACCGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB00789  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACCGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB02289  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACCGAGCCGCTGCCCACCATGGAGGACTTCCTC
CB02304  241   GTGCTCAGCCGGCGGATCGCCGAGGGCACGGAGCCGCTGCCCACCATGGAGGACTTCCTC

ptmT3    301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
CB00739  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
CB00765  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
CB00775  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
CB00789  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
CB02289  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCGCTGCTGATCACCGTG
CB02304  301   GAACTGGCCTACCGGCACGTCAACTACCGCAGTGTGGCCCTGGCACTGCTGATCACCGTG
```

Figure S3c. (continued)

```
ptmT3    361  GGCGAGCGTCCGGACAGCGCGGCGCAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB00739  361  GGCGAGCGCCCGGACAGCGCGGCACAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB00765  361  GGCGAGCGCCCGGACAGCGCGGCACAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB00775  361  GGCGAGCGCCCGGACAGCGCGGCACAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB00789  361  GGCGAGCGCCCGGACAGCGCGGCTCAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB02289  361  GGCGAGCACCCGGACAGCGCGGCCCAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG
CB02304  361  GGCGAGCGCCCGGACAGCGCGGCACAGGAGCGGCTCGACGCGGCCCTCGTCCCGGCGTCG

ptmT3    421  AGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB00739  421  CGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB00765  421  CGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB00775  421  CGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB00789  421  CGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB02289  421  AGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
CB02304  421  CGGGCGGTCCGGCTCGCCAACGATCTGCGGACCTGGGCCAA
```

d

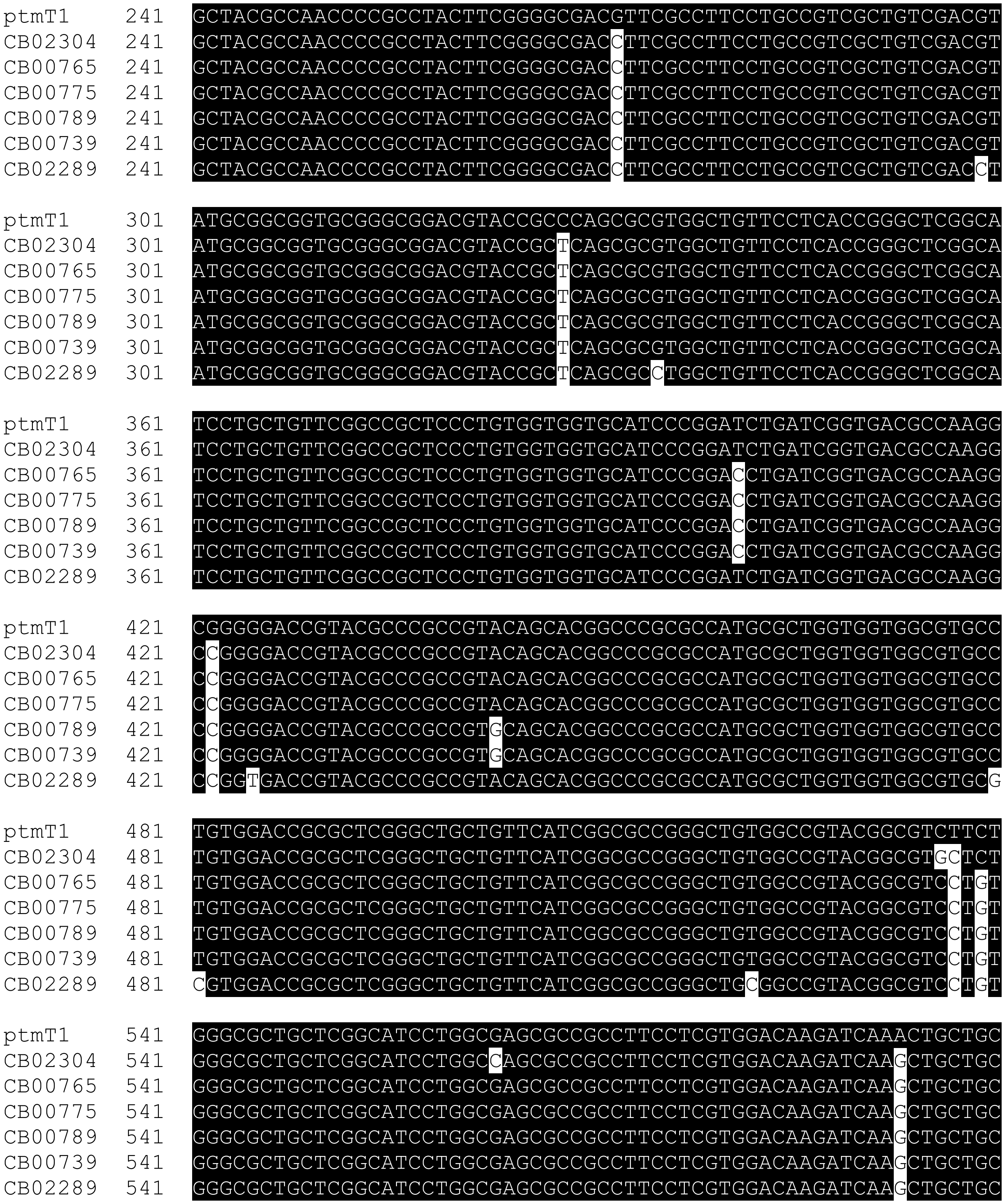
```
ptmT1      1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB02304    1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB00765    1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB00775    1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB00789    1  CCGGGCTGGACATCCGGGCGGACACACACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB00739    1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA
CB02289    1  CCGGGCTGGACATCCGGGCGGACACCCACACCAGCGGCAAGGAGAGCATCGCCCGCGCCA

ptmT1     61  CCCAGCACCTCAGCGTCCGCACCGCGTTCACCTGCGCGGCACTGGAGATGGCGCTCGCCC
CB02304   61  CCCAACATCTCAGCGTCCGCACCGCGTTCACCTGCGCGGCACTGGAGATGGCGCTCGCCC
CB00765   61  CCCAACGCCTCAGCGTCCGCACGGCGTTCACCTGCGCGGGACTGGAGATGGCGCTCGCGC
CB00775   61  CCCAACGCCTCAGCGTCCGCACGGCGTTCACCTGCGCGGGACTGGAGACGGCGCTCGCGC
CB00789   61  CCCAACGCCTCAGCGTCCGCACCGCGTTCACCTGCGCGGGACTGGAGATGGCGCTCGCGC
CB00739   61  CCCAACGCCTCAGCGTCCGCACCGCGTTCACCTGCGCGGGACTGGAGATGGCGCTCGCCC
CB02289   61  CCCAGCACCTCAGCGTCCGCACCGCGTTCACCTGCGCGGCACTGGAGATGGCGCTCGCCC

ptmT1    121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGCGGTGGGGGTGG
CB02304  121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGTCGTGGGTGTGG
CB00765  121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGTCGTGGGGGTGG
CB00775  121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGTCGTGGGGGTGG
CB00789  121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGTCGTGGGGGTGG
CB00739  121  TCGGGCTGTCGGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGTCGTGGGGGTGG
CB02289  121  TCGGGCTGTCCGTCTGGGTCGCGCTCTGGCTCGGCCGGCCGCTCGTCGCGGTGGGTGTGG

ptmT1    181  CGCTGTCGATCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB02304  181  CGCTGTCGATCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB00765  181  CGCTATCGGTCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB00775  181  CGCTATCGGTCGTGCTCCACTTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB00789  181  CGCTGTCGGTCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB00739  181  CGCTGTCGGTCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
CB02289  181  CACTGTCGATCGTGCTCCACCTCGCCTACAACCTGGAGCCGGTCCGGCTCAAGCGGCGCG
```

Figure S3d. (continued)



```
ptmT1    241   GCTACGCCAACCCCGCCTACTTCGGGGCGACGTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB02304  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB00765  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB00775  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB00789  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB00739  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACGT
CB02289  241   GCTACGCCAACCCCGCCTACTTCGGGGCGACCTTCGCCTTCCTGCCGTCGCTGTCGACCT

ptmT1    301   ATGCGGCGGTGCGGGCGGACGTACCGCCCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB02304  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB00765  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB00775  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB00789  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB00739  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCGTGGCTGTTCCTCACCGGGCTCGGCA
CB02289  301   ATGCGGCGGTGCGGGCGGACGTACCGCTCAGCGCCTGGCTGTTCCTCACCGGGCTCGGCA

ptmT1    361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGATCTGATCGGTGACGCCAAGG
CB02304  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGATCTGATCGGTGACGCCAAGG
CB00765  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGACCTGATCGGTGACGCCAAGG
CB00775  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGACCTGATCGGTGACGCCAAGG
CB00789  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGACCTGATCGGTGACGCCAAGG
CB00739  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGACCTGATCGGTGACGCCAAGG
CB02289  361   TCCTGCTGTTCGGCCGCTCCCTGTGGTGGTGCATCCCGGATCTGATCGGTGACGCCAAGG

ptmT1    421   CGGGGGACCGTACGCCCGCCGTACAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB02304  421   CCGGGGACCGTACGCCCGCCGTACAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB00765  421   CCGGGGACCGTACGCCCGCCGTACAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB00775  421   CCGGGGACCGTACGCCCGCCGTACAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB00789  421   CCGGGGACCGTACGCCCGCCGTGCAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB00739  421   CCGGGGACCGTACGCCCGCCGTGCAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCC
CB02289  421   CCGGTGACCGTACGCCCGCCGTACAGCACGGCCCGCGCCATGCGCTGGTGGTGGCGTGCG

ptmT1    481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTCTTCT
CB02304  481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTGCTCT
CB00765  481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTCCTGT
CB00775  481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTCCTGT
CB00789  481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTCCTGT
CB00739  481   TGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGTGGCCGTACGGCGTCCTGT
CB02289  481   CGTGGACCGCGCTCGGGCTGCTGTTCATCGGCGCCGGGCTGCGGCCGTACGGCGTCCTGT

ptmT1    541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAACTGCTGC
CB02304  541   GGGCGCTGCTCGGCATCCTGGCCAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
CB00765  541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
CB00775  541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
CB00789  541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
CB00739  541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
CB02289  541   GGGCGCTGCTCGGCATCCTGGCGAGCGCCGCCTTCCTCGTGGACAAGATCAAGCTGCTGC
```

Figure S3d. (continued)

```
ptmT1    601  GGAGCATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB02304  601  GGAGCATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB00765  601  GGCACATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB00775  601  GGCACATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB00789  601  GGCACATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB00739  601  GGCACATCTCGCGGGAGAACCTCCCACACGAGTCCACGATGCGCAAGCACAGCCTCTCGC
CB02289  601  GGCACATCTCGCGGGAGAACCTCCCGCACGAGTCCACGATGCGCAAGCACAGCCTCTCGC

ptmT1    661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB02304  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB00765  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB00775  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB00789  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB00739  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
CB02289  661  TGGCGATGGGCGGGGACCTCCTGCTCTGCGCCATCC
```

Figure S4. Genetic organization of the PTM-PTN biosynthetic gene clusters in *S. platensis* MA7327, *S. platensis* CB00739, and *S. platensis* CB00765. The PTM-PTN clusters in *S. platensis* CB00739 (KJ189771) and *S. platensis* CB00765 (KJ189772) share 99% identity in nucleotide sequence, and their nucleotide sequences are 97% identical to *S. platensis* MA7327[10] (FJ655920).

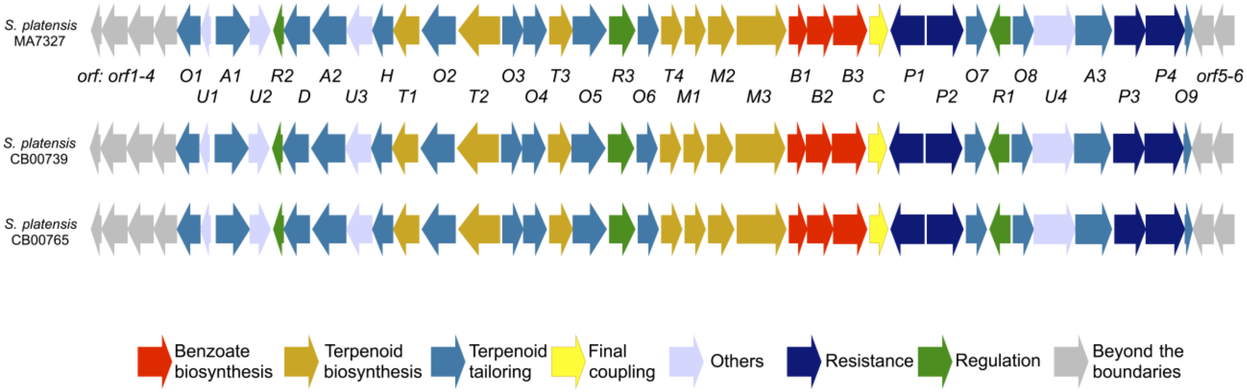


Figure S5. The new platensimycin and platencin dual-producers are distinct from *S. platensis* MA7327, showing superior genetic amenability. **(a)** Phylogenetic tree generated from the alignment of concatenated partial sequences of the four housekeeping genes 16S rRNA, *recA*, *rpoB*, and *trpB* (2975-bp total) using the Tamura-Nei evolutionary distance method and the neighbor-joining algorithm. Bootstrap values ≥70% (based on 100 resampled trials) are given at nodes. Bar, 0.02 substitutions per nucleotide position. Nucleotide sequences of the housekeeping genes for *S. coelicolor* A3(2) (AL645882), *S. avermitilis* MA-4680 (BA000030), *S. rimosus* ATCC 10970 (ANSJ00000000), *S. griseus* NBRC 13350 (AP009493), *S. scabiei* 87.22 (NC_013929), *S. venezuelae* ATCC 10712 (FR845719), and *S. clavuligerus* ATCC 27064 (ADWJ00000000) were obtained from the NCBI database and used as relevant *Streptomyces* spp. for phylogenetic analysis. *Streptosporangium roseum* DSM43021 (CP001814) and *Mycobacterium tuberculosis* H37Rv (AL123456) were used as outgroups. **(b)** Morphology of the six new platensimycin and platencin dual-producers of *S. platensis* CB00739, CB00765, CB00775, CB00789, CB02289, and CB02304 in comparison with *S. platensis* MA7327 and MA7339 on an ISP4 agar plate.

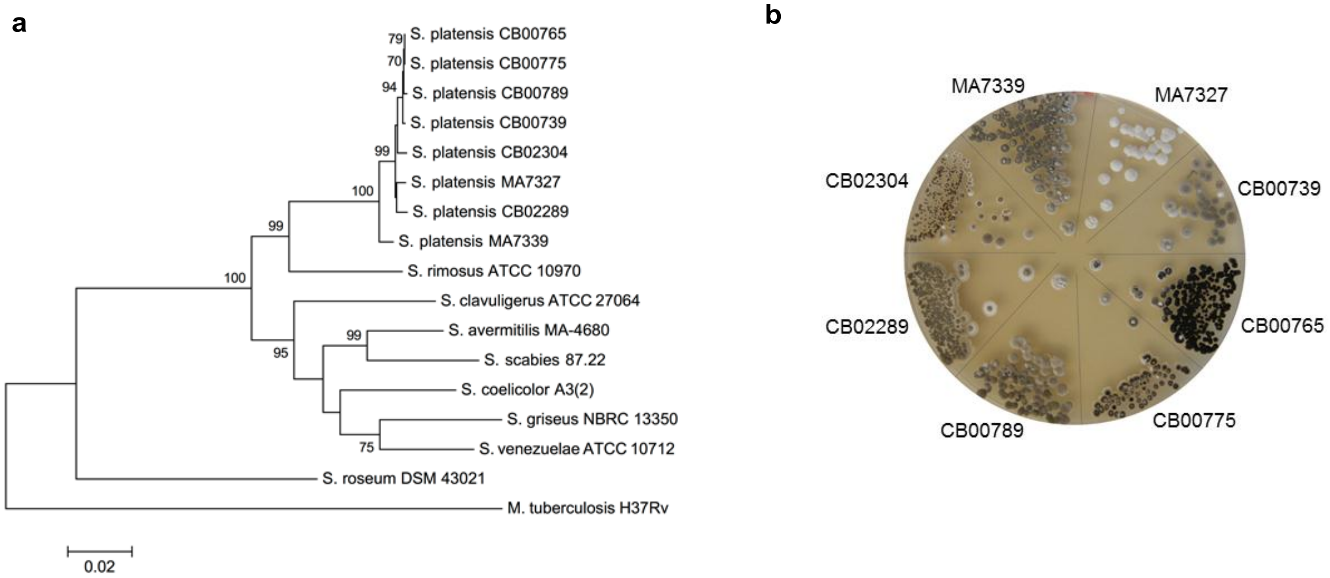
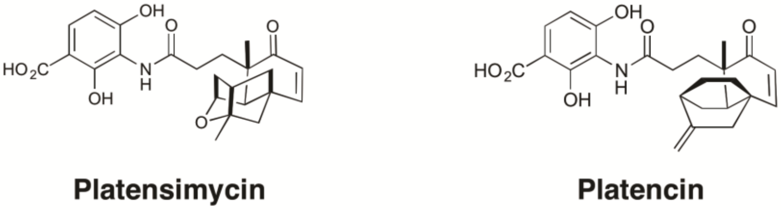
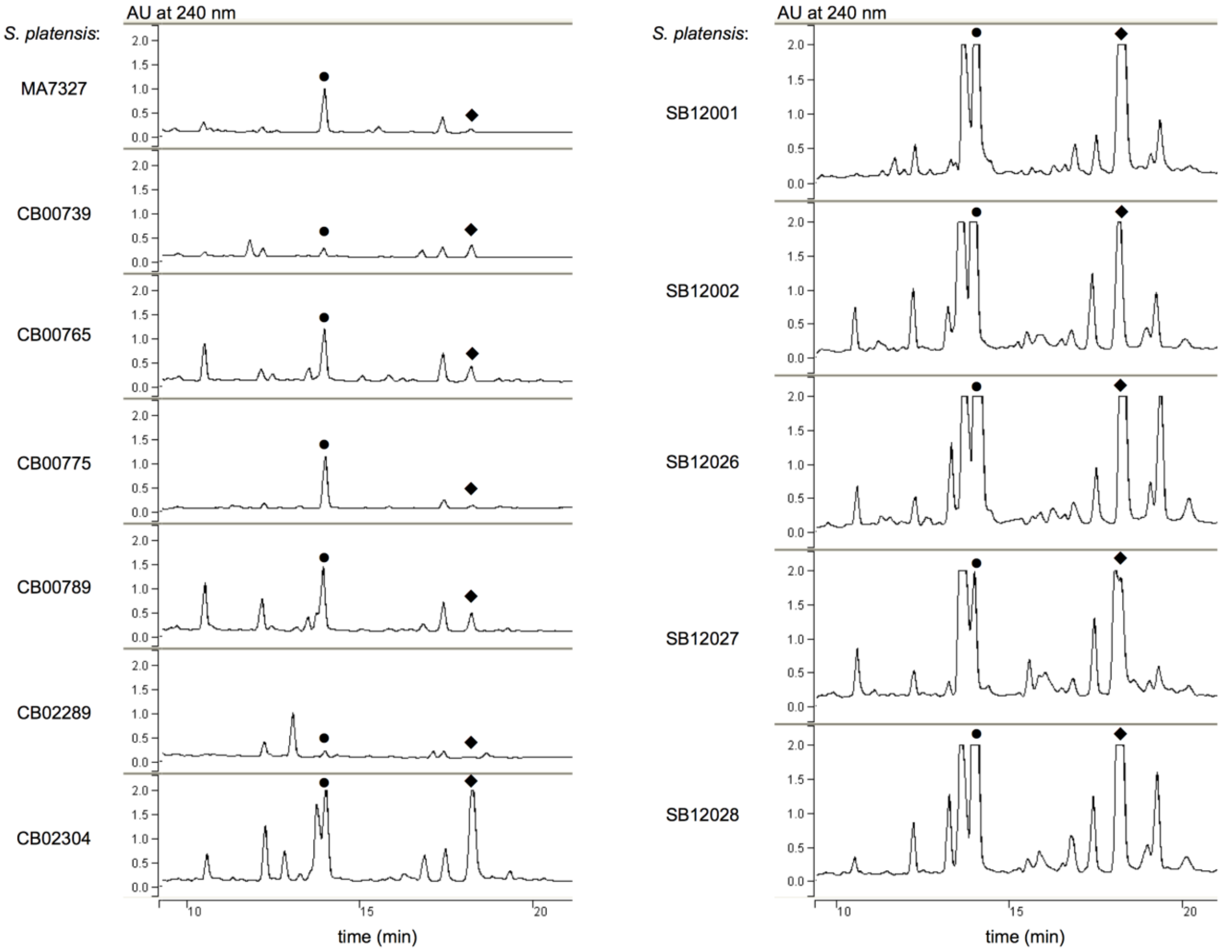


Figure S6. Production of platensimycin (●) and platencin (♦) by *Streptomyces platensis* spp. **(a)** Structures of platensimycin (PTM) and platencin (PTN), **(b)** HPLC chromatograms of crude extracts prepared from: left column, original and new PTM-PTN-producing *Streptomyces platensis* spp.; and right column, previously reported and new Δ*ptmR1* deletion mutants. The y-axes are kept constant between the two panels for visualization of the significant increase in titers. Crude extracts for SB12001, SB12002, SB12026, and SB12028 were diluted 4-fold for better representation of PTM and PTN production, **(c)** Extracted ion (*m/z* at 442.1863 for the $[PTM + H]^+$ ion and *m/z* at 426.1914 for the $[PTN + H]^+$ ion) chromatograms from LC-MS analyses of: left column, original and new PTM-PTN-producing *Streptomyces platensis* spp.; and right column, previously reported and new Δ*ptmR1* deletion mutants.

a



b



c

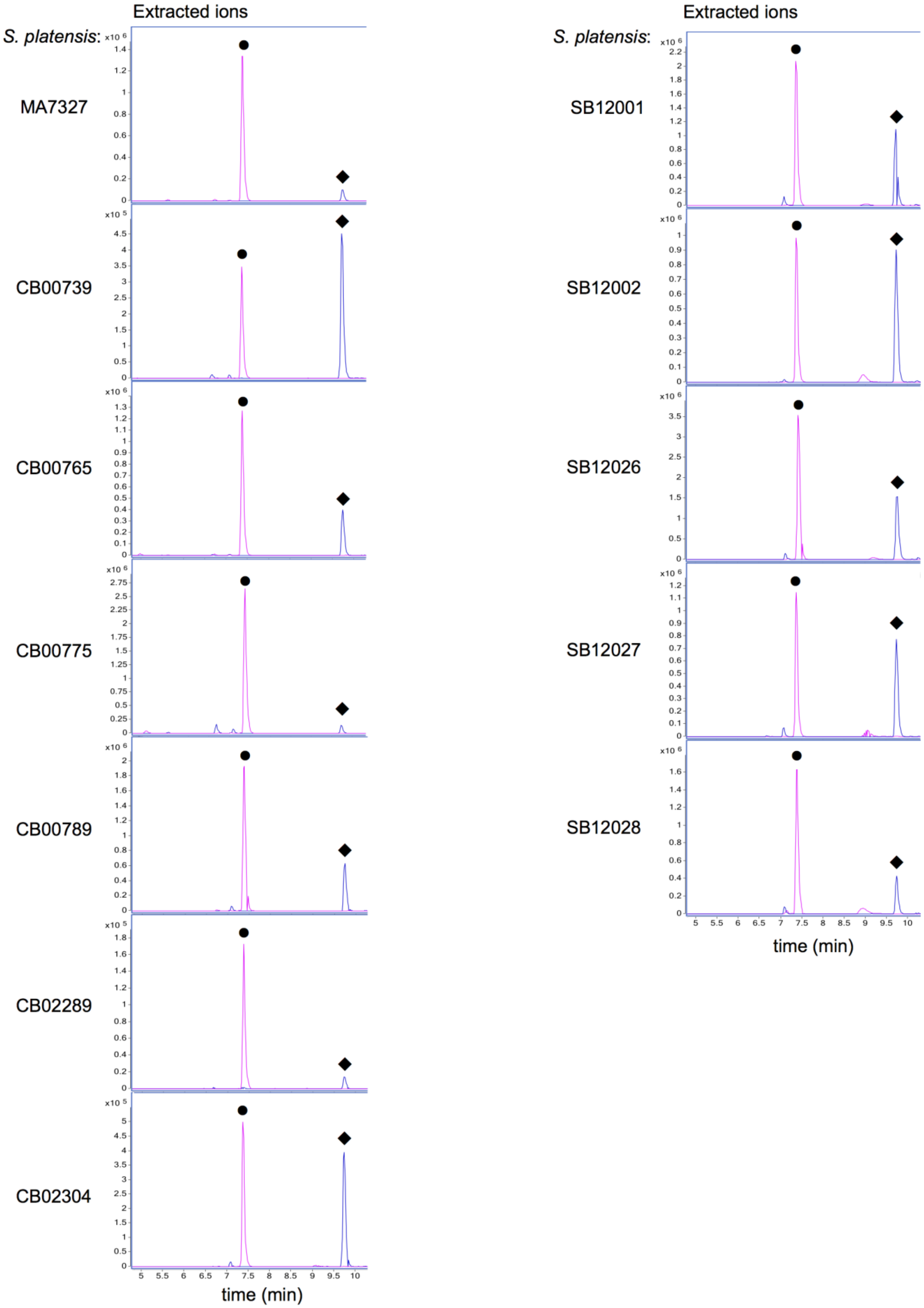
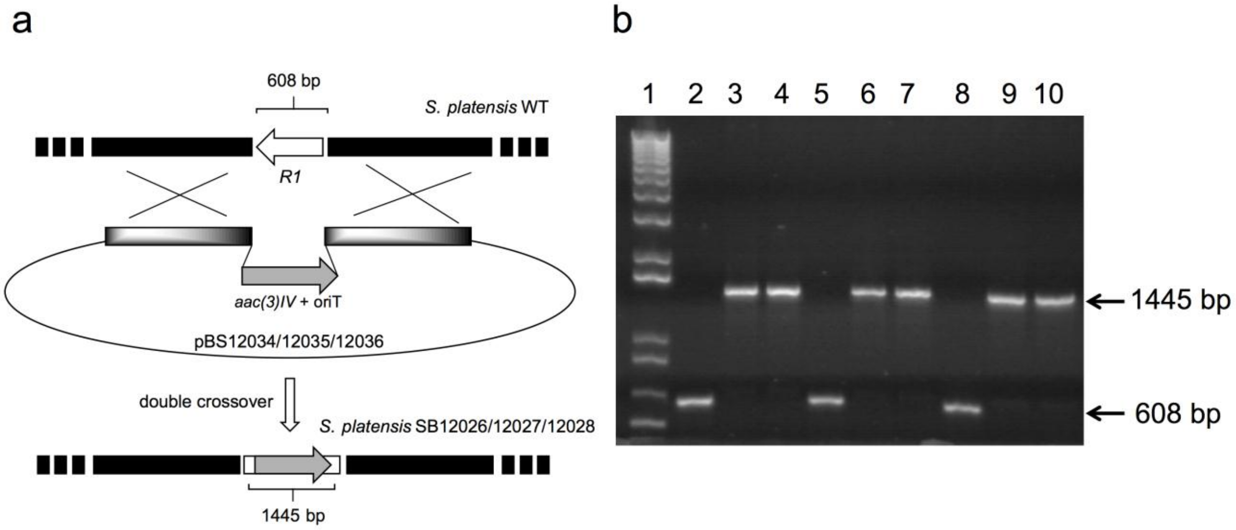

Extracted ions
S. platensis:
MA7327
x10 6
CB00739
x10 5
CB00765
x10 6
CB00775
x10 6
CB00789
x10 6
CB02289
x10 5
CB02304
x10 5
5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10
time (min)
Extracted ions
S. platensis:
SB12001
x10 6
SB12002
x10 6
SB12026
x10 6
SB12027
x10 6
SB12028
x10 6
5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10
time (min)

Figure S7. Inactivation of *ptmR1* in three new PTM-PTN producers affording *S. platensis* SB12026, SB12027, and SB12028. **(a)** Schematic representation of the deletion of *S. platensis ptmR1* by insertion of an apramycin resistance-oriT cassette *(aac(3)IV + oriT*), **(b)** PCR verification of wild-type and double crossover mutant genotypes, using the primers ptmRidF and ptmRidR. Lane 1, 1 Kb Plus DNA ladder (Invitrogen); lane 2, *S. platensis* CB00739; lanes 3-4, two isolates of *S. platensis* SB12026; lane 5, *S. platensis* CB00775; lanes 6-7, *S. platensis* SB12028; lane 8, *S. platensis* CB00765; lanes 9-10, *S. platensis* SB12027.



References

(1) Adekambi, T.; Butler, R. W.; Hanrahan, F.; Delcher, A. L.; Drancourt, M.; Shinnick, T. M. *PloS One* **2011**, *6*, e14792.

(2) MacNeil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; MacNeil, T. *Gene* **1992**, *111*, 61-68.

(3) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 1541-1546.

(4) Wang, J.; Soisson, S. M.; Young, K.; Shoop, W.; Kodali, S.; Galgoci, A.; Painter, R.; Parthasarathy, G.; Tang, Y. S.; Cummings, R.; Ha, S.; Dorso, K.; Motyl, M.; Jayasuriya, H.; Ondeyka, J.; Herath, K.; Zhang, C.; Hernandez, L.; Allocco, J.; Basilio, A.; Tormo, J. R.; Genilloud, O.; Vicente, F.; Pelaez, F.; Colwell, L.; Lee, S. H.; Michael, B.; Felcetto, T.; Gill, C.; Silver, L. L.; Hermes, J. D.; Bartizal, K.; Barrett, J.; Schmatz, D.; Becker, J. W.; Cully, D.; Singh, S. B. *Nature* **2006,** *441*, 358-361.

(5) Singh, S. B.; Jayasuriya, H.; Ondeyka, J. G.; Herath, K. B.; Zhang, C.; Zink, D. L.; Tsou, N. N.; Ball, R. G.; Basilio, A.; Genilloud, O.; Diez, M. T.; Vicente, F.; Pelaez, F.; Young, K.; Wang, J. *J. Am. Chem. Soc.* **2006**, *128*, 11916-11920.

(6) Jayasuriya, H.; Herath, K. B.; Zhang, C.; Zink, D. L.; Basilio, A.; Genilloud, O.; Diez, M. T.; Vicente, F.; Gonzalez, I.; Salazar, O.; Pelaez, F.; Cummings, R.; Ha, S.; Wang, J.; Singh, S. B. *Angew. Chem. Int. Ed.* **2007,** *46*, 4684-4688.

(7) Wang, J.; Kodali, S.; Lee, S. H.; Galgoci, A.; Painter, R.; Dorso, K.; Racine, F.; Motyl, M.; Hernandez, L.; Tinney, E.; Colletti, S. L.; Herath, K.; Cummings, R.; Salazar, O.; Gonzalez, I.; Basilio, A.; Vicente, F.; Genilloud, O.; Pelaez, F.; Jayasuriya, H.; Young, K.; Cully, D. F.; Singh, S. B. *Proc. Natl. Acad. Sci. USA* ***2007***, *104*, 7612-7616.

(8) Smanski, M. J.; Peterson, R. M.; Rajski, S. R.; Shen, B. *Antimicrob. Agents Chemother.* **2009**, *53*, 1299-1304.

(9) Yu, Z.; Smanski, M. J.; Peterson, R. M.; Marchillo, K.; Andes, D.; Rajski, S. R.; Shen, B. *Org. Lett.* **2010**, *12*, 1744-1747.

(10) Smanski, M. J.; Yu, Z.; Casper, J.; Lin, S.; Peterson, R. M.; Chen, Y.; Wendt-Pienkowski, E.; Rajski, S. R.; Shen, B. *Proc. Natl. Acad. Sci. USA* ***2011***, *108*, 13498-13503.

(11) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947-2948.