

SUPPLEMENTAL MATERIAL

Discovery and biosynthesis of the antibiotic bicyclomycin in distant bacterial classes

Natalia M. Vior, Rodney Lacret, Govind Chandra, Siobhán Dorai-Raj, Martin Trick,
Andrew W. Truman

TABLE OF CONTENTS

Table S1. Statistics of the *S. cinnamoneus* DSM 41675 genome sequence assemblies.

Table S2. *bcm* gene clusters in *S. cinnamoneus* DSM 41675 and *P. aeruginosa* SCV20265.

Figure S1. Alignment of a *P. aeruginosa* CDPS with BcmA from *S. cinnamoneus*.

Figure S2. MRM analysis of bicyclomycin from *S. cinnamoneus*.

Figure S3. MS and MS² spectra for bicyclomycin from *Pseudomonas*.

Figure S4. Comparison of the LC-MS spectra of *P. fluorescens* SBW25-pJH-BCMclp-PA and an empty vector control.

Table S3. Exact masses of the BCM-like compounds produced by *Pseudomonas*.

Figure S5. Main ¹H-¹³C HMBC correlations identified for bicyclomycin from *Pseudomonas*.

Table S4. ¹H and ¹³C NMR data for BCM produced by *Pseudomonas* in CD₃SOCD₃.

Figure S6. 400 MHz ¹H NMR spectrum of BCM produced by *Pseudomonas* in CD₃SOCD₃.

Figure S7. 100 MHz ¹³C NMR spectrum of BCM produced by *Pseudomonas* in CD₃SOCD₃.

Figure S8. 400 MHz ¹H-¹³C HSQC NMR spectrum of BCM produced by *Pseudomonas* in CD₃SOCD₃.

Figure S9. 400 MHz ¹H-¹H COSY NMR spectrum of BCM produced by *Pseudomonas* in CD₃SOCD₃.

Figure S10. 400 MHz ^1H - ^{13}C HMBC NMR spectrum of BCM produced by *Pseudomonas* in CD_3SOCD_3 .

Table S5. Mobile genetic elements and tRNA genes surrounding the *bcm* cluster in different bacteria.

Figure S11. Genomic context of the *bcm* cluster in *P. aeruginosa*.

Figure S12. Genomic context of the *bcm* cluster in *Mycobacterium*.

Figure S13. Alignment and percentage identity matrix of the *bcm* 2-OG/Fe dioxygenases from *S. cinnamomeus*.

Figure S14. Unrooted version of the phylogenetic tree shown in Fig. 6.

| | Illumina only assembly | Nanopore only assembly | Hybrid assembly |
|---------------------|-----------------------------------|-----------------------------------|----------------------------|
| Sequence size | 6,802,277 | 6,597,724 | 6,662,690 |
| Number of contigs | 415 | 4 | 2 |
| GC content (%) | 72.4 | 72.3 | 72.4 |
| Longest contig size | 896,920 | 4,033,611 | 6,463,510 |
| N50 value | 403,337 | 4,033,611 | 6,463,510 |
| L50 value | 6 | 1 | 1 |

Table S1. Statistics of the different *S. cinnamoneus* DSM 41675 genome sequence assemblies described in this work.

| Gene | Locus tag in <i>S. cinnamomeus</i> DSM 41675 | Locus tag in <i>P. aeruginosa</i> SCV20265 | Gene product | Size (aa) in Scin/Paer |
|--------------------|---|---|---|-----------------------------------|
| <i>bcmA</i> | CYQ11_26550 | SCV20265_RS31155 | tRNA-dependent cyclodipeptide synthase | 241/244 |
| <i>bcmB</i> | CYQ11_26545 | SCV20265_RS31160 | 2OG-Fe(II) oxygenase | 325/322 |
| <i>bcmC</i> | CYQ11_26540 | SCV20265_RS31165 | 2OG-Fe(II) oxygenase | 305/313 |
| <i>bcmD</i> | CYQ11_26535 | SCV20265_RS31170 | Cytochrome P450 | 488/476 |
| <i>bcmE</i> | CYQ11_26530 | SCV20265_RS31175 | 2OG-Fe(II) oxygenase | 314/293 |
| <i>bcmF</i> | CYQ11_26525 | SCV20265_RS31180 | 2OG-Fe(II) oxygenase | 296/315 |
| <i>bcmG</i> | CYQ11_26520 | SCV20265_RS31185 | 2OG-Fe(II) oxygenase | 300/296 |
| <i>bcmT</i> | CYQ11_26555 | SCV20265_RS31190 | MFS transporter | 473/401 |

Table S2. *bcm* gene clusters in *S. cinnamomeus* DSM 41675 and *P. aeruginosa* SCV20265.

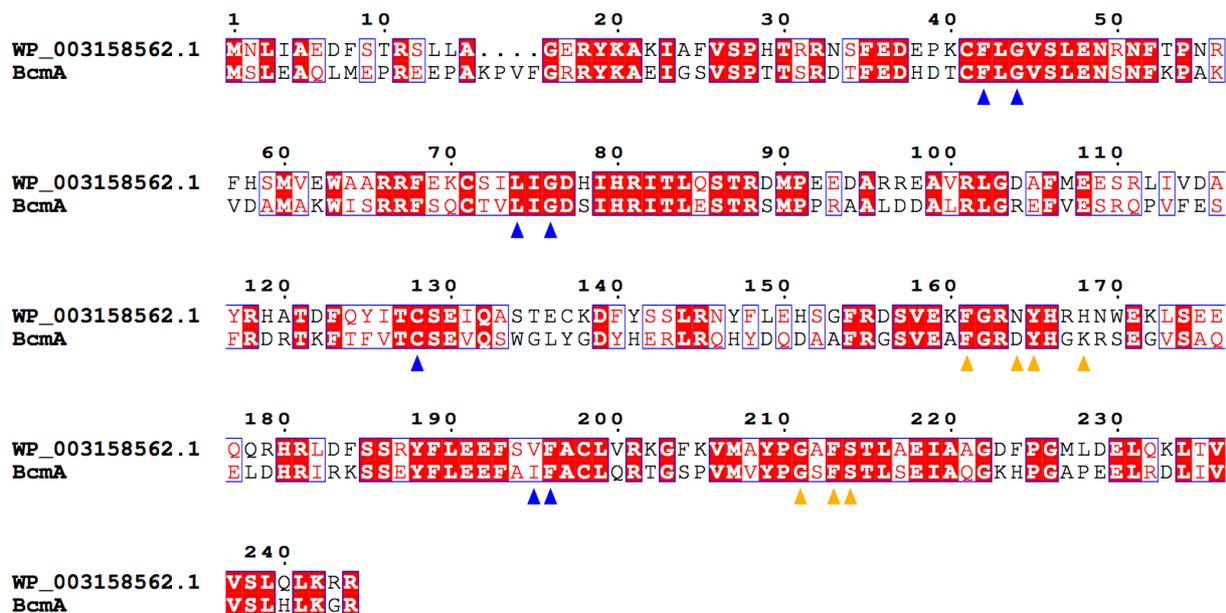


Figure S1. Alignment of a *P. aeruginosa* CDPS (WP_003158562.1) with selectivity for cyclo(L-Ile-L-Leu) and BcmA from *S. cinnamoneus*. Specificity-determining residues are highlight for P1 (blue triangles) and P2 (orange triangles) binding pockets, as determined by Jacques *et al.* (1). Identical residues have a red background while similar residues have red lettering. Figure generated using Esript 3.0 (2).

M1146 + pJ-BCM

10 µg/mL BCM standard

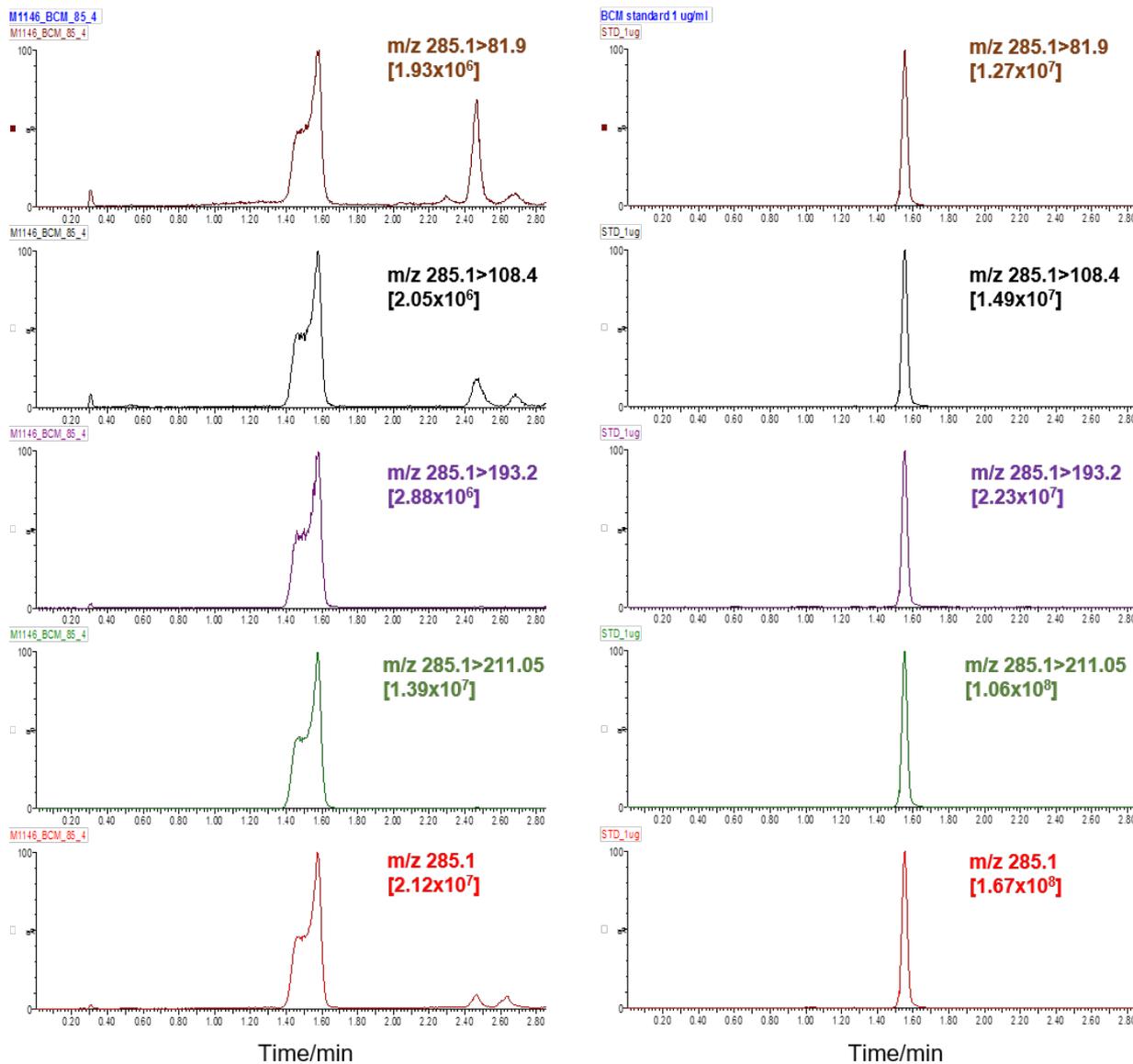


Figure S2. Multiple Reaction Monitoring (MRM) analysis of bicyclomycin (parent ion m/z 285.1, $[M-H_2O+H]^+$) produced by the heterologous expression strain M1146 carrying pJ-BCM (left) compared with a pure BCM standard (right). The measured intensity for each monitored transition is indicated in brackets.

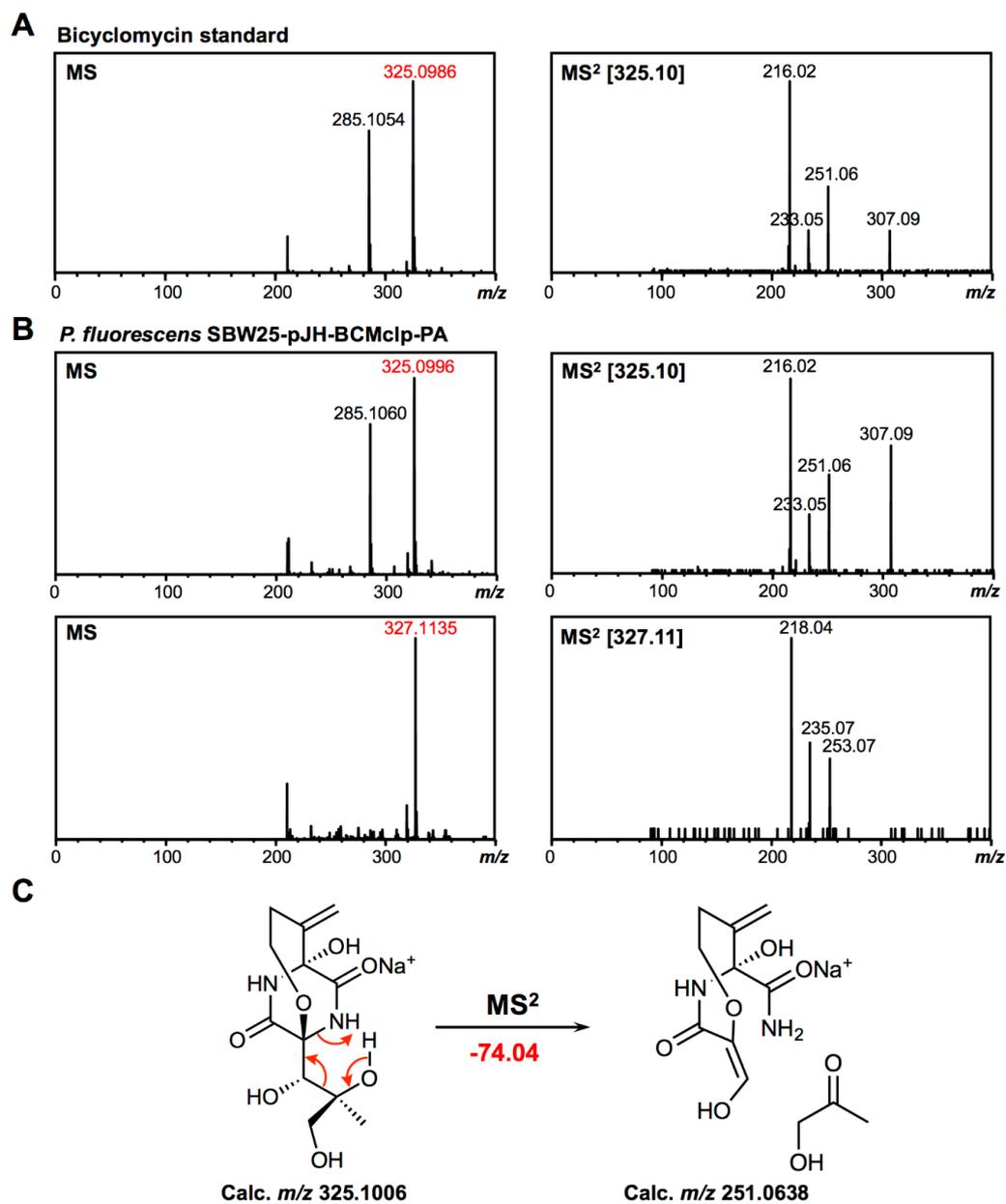


Figure S3. MS and MS² spectra for bicyclomycin. (A) MS spectrum for a commercial standard of BCM and MS² spectra for [BCM+Na]⁺. (B) MS and MS² spectra for compounds with m/z 325.10 and m/z 327.11 produced by *P. fluorescens* SBW25-pJH-BCMclp-PA. (C) Proposed fragmentation mechanism leading to the loss of 74.04 Da. An alternative mechanism would involve opening of the ether instead of breaking the C-N bond.

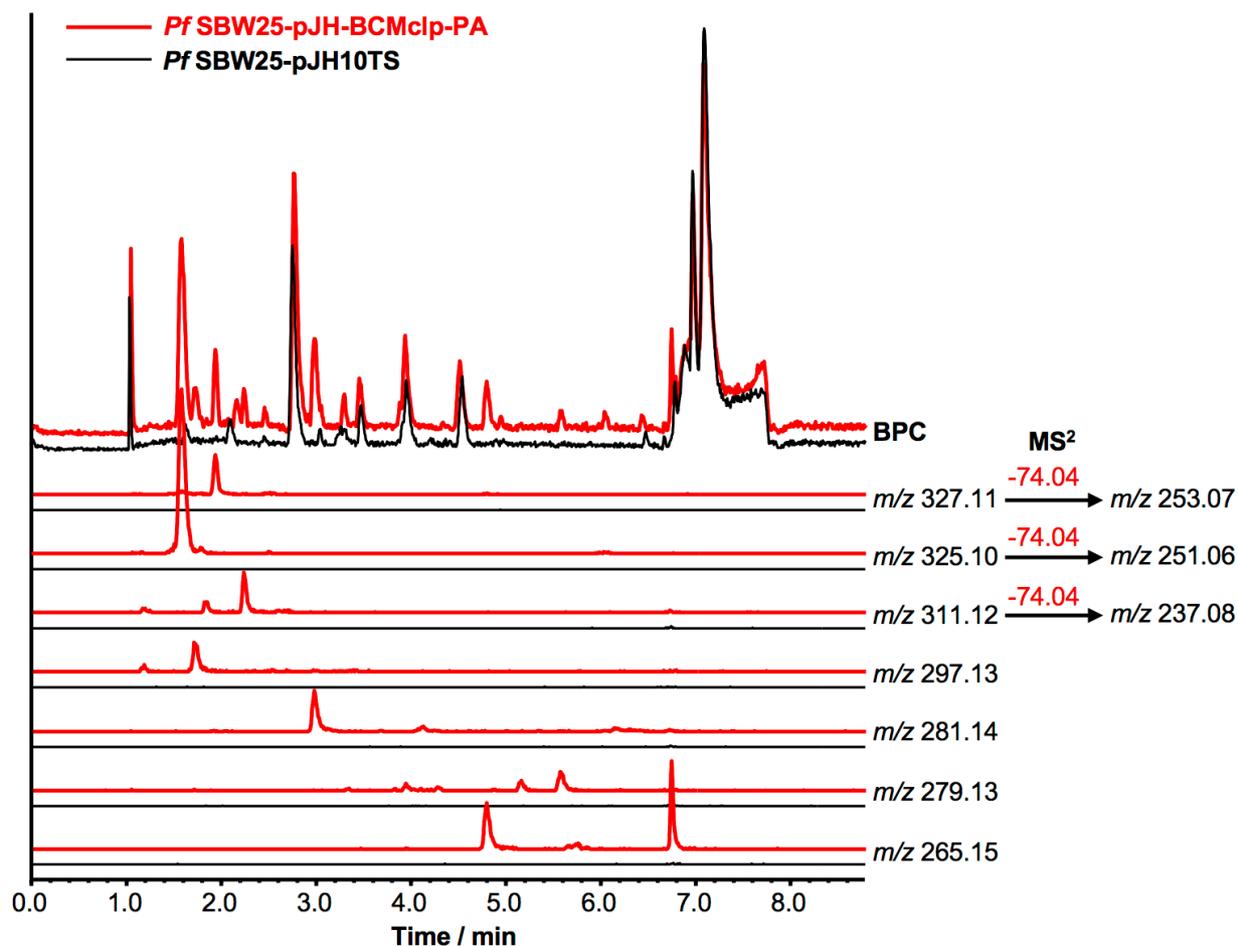


Figure S4. Comparison of the LC-MS spectra of *P. fluorescens* SBW25-pJH-BCMclp-PA (red) and an empty vector control (black). Extracted ion chromatograms are shown for putative $[M+Na]^+$ signals for the major metabolites that differed between these two fermentations (BPC = base peak chromatogram). Compounds with a BCM-like loss of 74.04 Da are highlighted.

| Predicted compound | [M+Na] ⁺ chemical formula | Pred. m/z | Obs. m/z | Error (ppm) |
|--------------------|--|-----------|----------|-------------|
| BCM | C ₁₂ H ₁₈ N ₂ NaO ₇ ⁺ | 325.1006 | 325.1008 | -0.62 |
| BCM+2H | C ₁₂ H ₂₀ N ₂ NaO ₇ ⁺ | 327.1163 | 327.1163 | 0.00 |
| BCM+2H-O | C ₁₂ H ₂₂ N ₂ NaO ₅ ⁺ | 311.1214 | 311.1208 | 1.93 |
| BCM+4H-2O | C ₁₂ H ₂₂ N ₂ NaO ₅ ⁺ | 297.1421 | 297.1420 | 0.34 |
| BCM+4H-3O | C ₁₂ H ₂₂ N ₂ NaO ₄ ⁺ | 281.1472 | 281.1467 | 1.78 |
| BCM+2H-3O | C ₁₂ H ₂₀ N ₂ NaO ₄ ⁺ | 279.1315 | 279.1310 | 1.79 |
| BCM+4H-4O | C ₁₂ H ₂₂ N ₂ NaO ₃ ⁺ | 265.1523 | 265.1520 | 1.13 |

Table S3. Exact masses of the BCM-like compounds produced by *P. fluorescens* SBW25-pJH-BCMclp-PA.

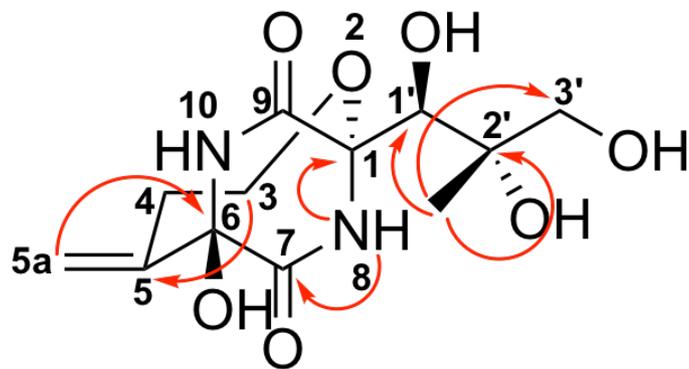


Figure S5. Main ¹H-¹³C HMBC correlations identified for bicyclomycin produced by SBW25-pJH-BCMclp-PA. Atom numbering relates to Table S3 and is identical to compound numbering in Kohn *et al.* (3).

| Position | δ_{H} , mult. (J in Hz) | δ_{C} |
|--------------------|--|------------------------|
| 1 | - | 87.8, C |
| 3 | 3.60, dd, 12.6, 8.5 3.79, dd, 12.6, 6.2 | 63.3, CH ₂ |
| 4 | 2.48, ddd, 14.3, 8.5, 6.2 | 35.3, CH ₂ |
| 5 | - | 149.1, C |
| 5a | 5.04, br s; 5.35, d, 1.8 | 115.2, CH ₂ |
| 6-OH | 6.88, s | - |
| 6 | - | 81.5, C |
| 7 | - | 169.6, C |
| 8-NH | 8.72, s | - |
| 9 | - | 166.3, C |
| 10-NH | 8.96, s | - |
| 1' | 3.89, d, 7.1 | 70.3, C |
| 1'-OH | 5.32, d, 7.5 | - |
| 2' | - | 77.1, C |
| 2'-CH ₃ | 1.16, s | 23.9, CH ₃ |
| 2'-OH | 4.53, s | - |
| 3' | 1.29, m | 66.3, CH ₂ |
| 3'-OH | 5.27, br s | - |

Table S4. ¹H and ¹³C NMR data for bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD₃SOCD₃.

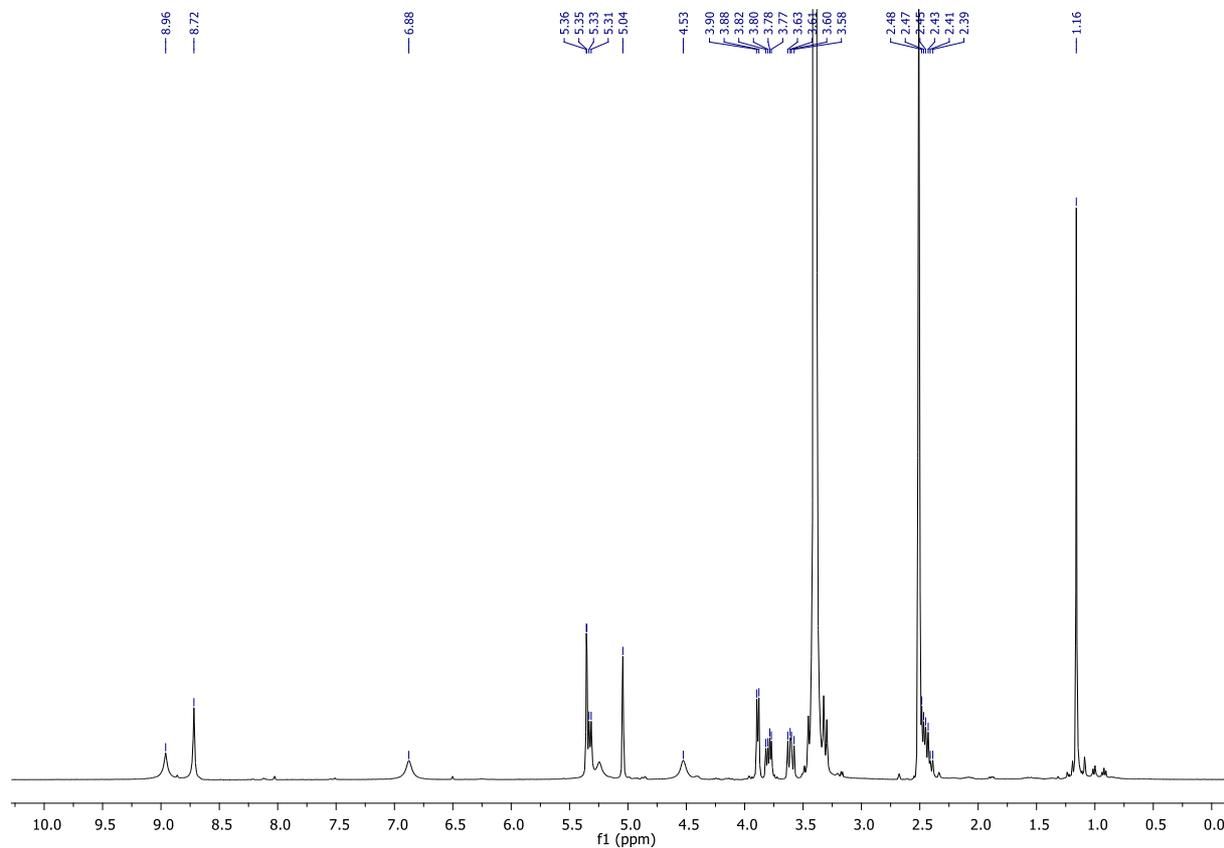


Figure S6. 400 MHz ¹H NMR spectrum of bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD₃SOCD₃.

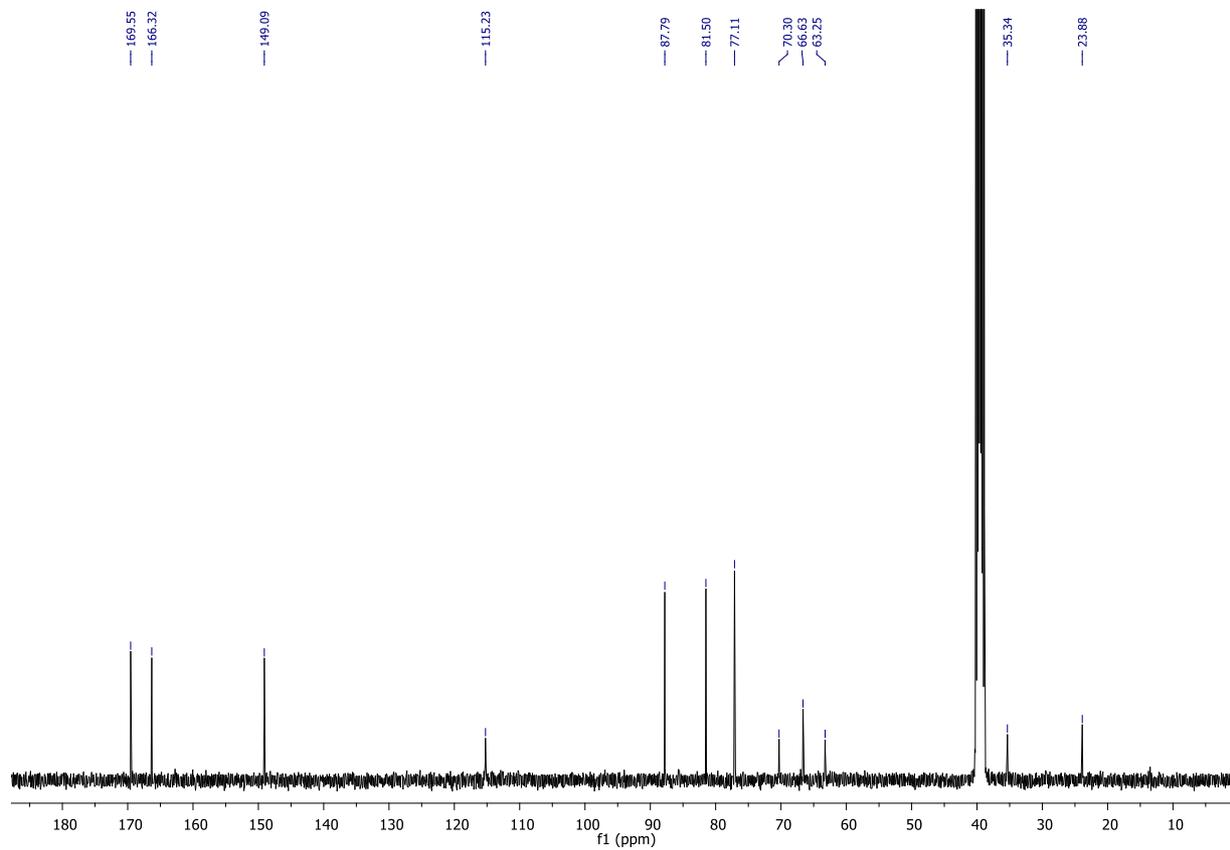


Figure S7. 100 MHz ^{13}C NMR spectrum of bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD_3SOCD_3 .

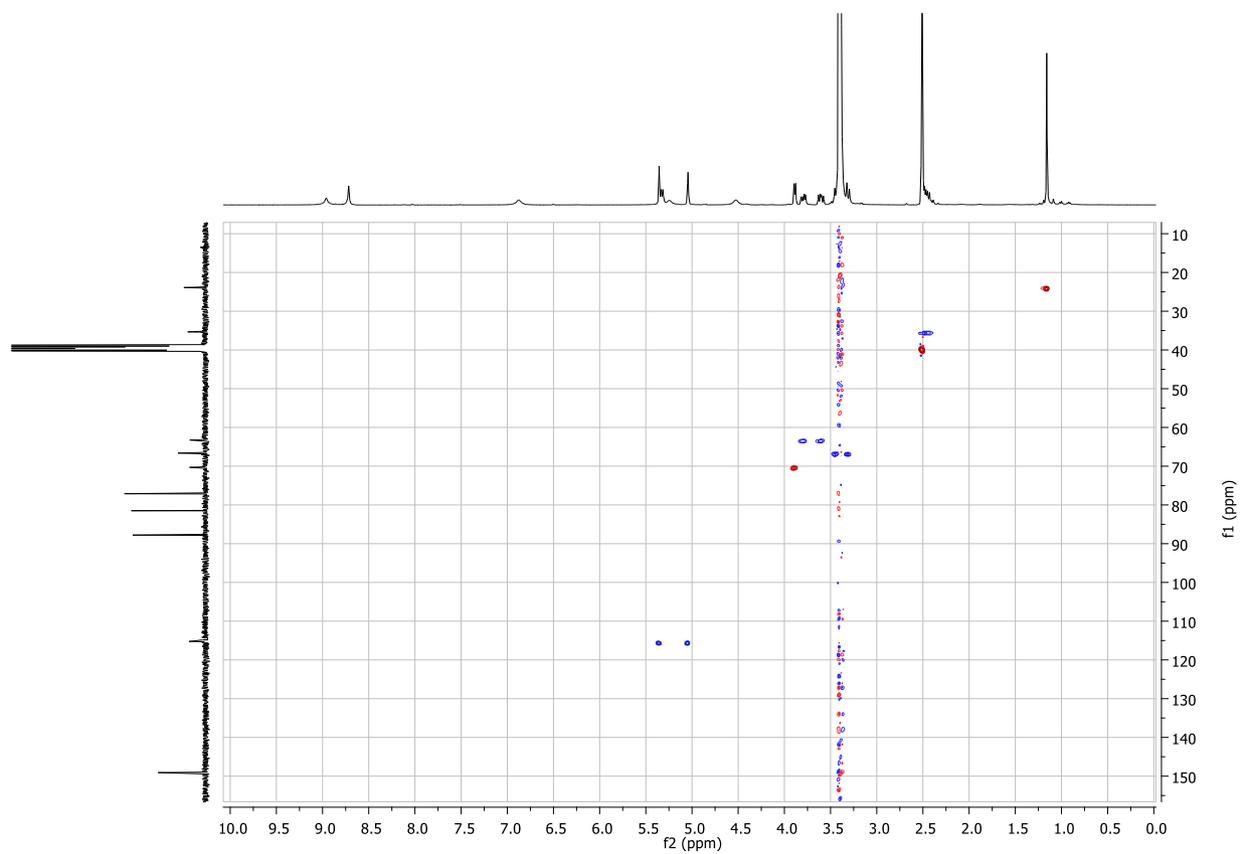


Figure S8. 400 MHz ^1H - ^{13}C HSQC NMR spectrum of bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD_3SOCD_3 .

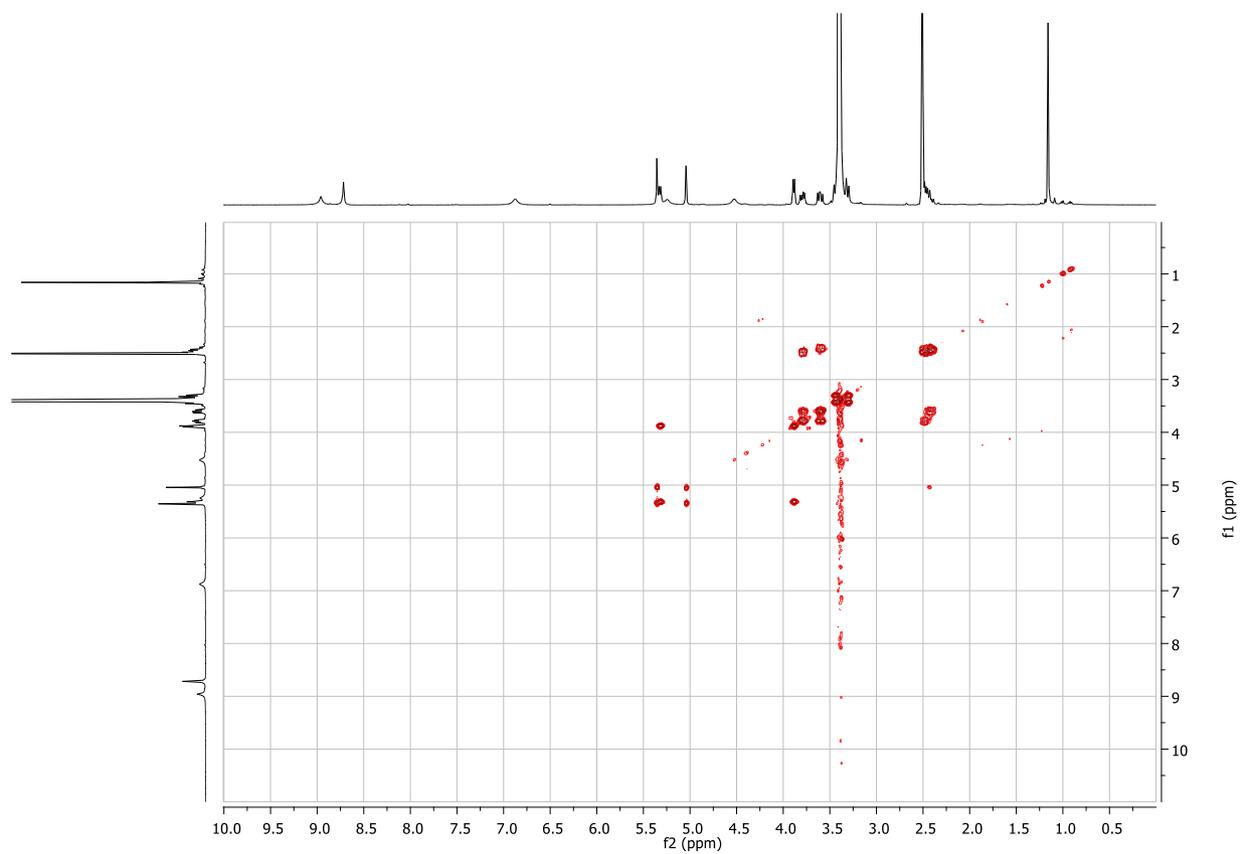


Figure S9. 400 MHz ^1H - ^1H COSY NMR spectrum of bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD_3SOCD_3 .

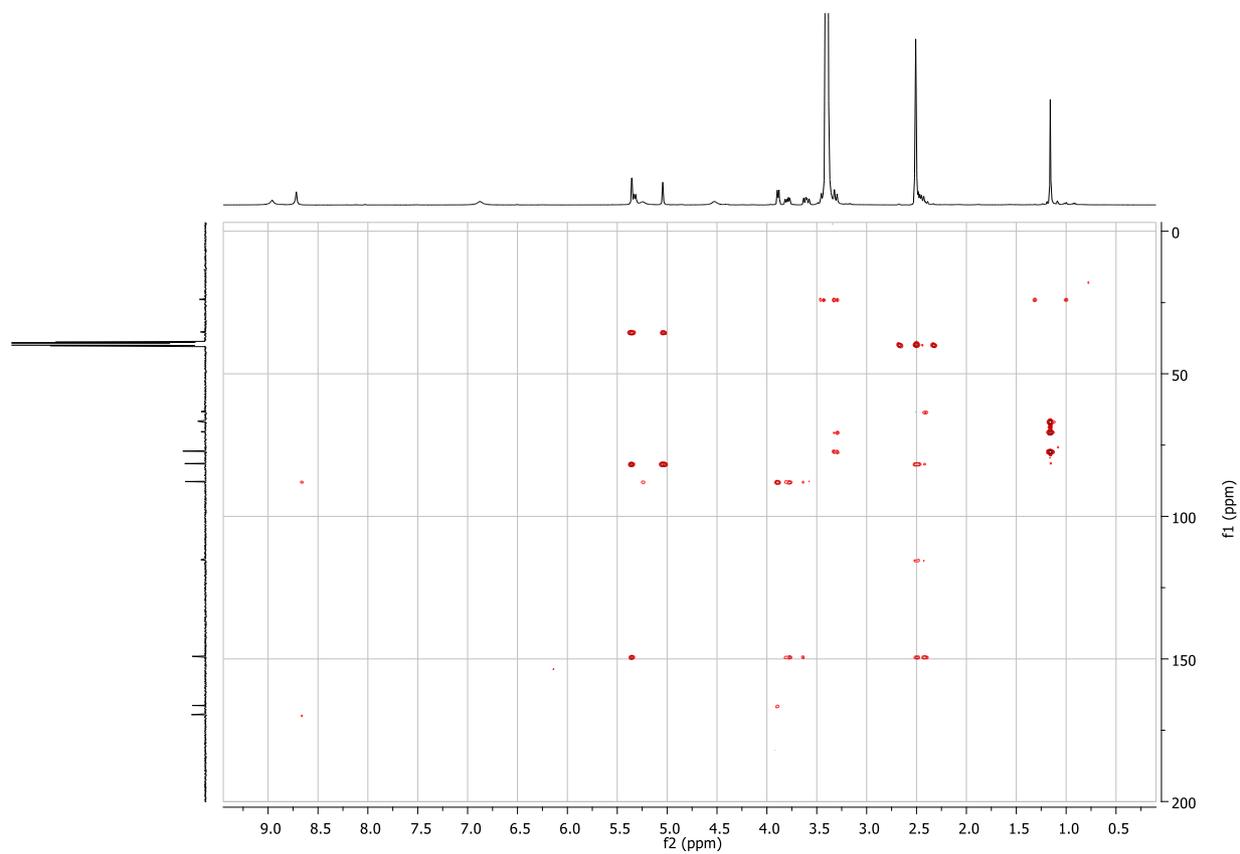


Figure S10. 400 MHz ^1H - ^{13}C HMBC NMR spectrum of bicyclomycin produced by SBW25-pJH-BCMclp-PA in CD_3SOCD_3 .

| Strain | Locus tag | Mobile genetic element | Conserved domain |
|---|---------------|---|----------------------|
| <i>M. chelonae</i> CCUG 47445 | BB28_RS01095 | tRNA-Ser | |
| | BB28_RS01100 | tRNA-Arg | |
| <i>M. chelonae</i> D16R20 | B4407_RS03550 | tRNA-Ser | |
| | B4407_RS03545 | tRNA-Arg | |
| <i>M. chelonae</i> D16R2 | B4391_RS18230 | tRNA-Ser | |
| | B4391_RS18235 | tRNA-Arg | |
| <i>M. chelonae</i> 15513 | BKG80_RS00985 | tRNA-Ser | |
| | BKG80_RS00990 | tRNA-Arg | |
| <i>M. chelonae</i> 15514 | BKG81_RS06600 | Resolvase | pfam00239 |
| | BKG81_RS06660 | Site-specific phage integrase | pfam00589 |
| | BKG81_RS06665 | tRNA-Phe | |
| | BKG81_RS06670 | tRNA-Asp | |
| | BKG81_RS06675 | tRNA-Glu | |
| <i>W. herbipolensis</i> ARP1 | TU34_RS19365 | IS110 family transposase | pfam02371/pfam01548 |
| <i>S. hygroscopicus</i> XM201 | --- | --- | --- |
| <i>S. violaceusniger</i> NRRL F-8817 | ADL28_RS46050 | Transposase DDE_Tnp_1_4 domain-containing protein | pfam13701 |
| | ADL28_RS46040 | Tn3 transposase DDE domain-containing protein | Pfam01526 |
| | ADL28_RS16000 | Tn3 transposase DDE domain-containing protein | Pfam01526 |
| <i>S. castelarensis</i> NRRL B-24289 | BZY55_RS12845 | Serine recombinase (resolvase domain) | cd03768/ pfam00239 |
| | BZY55_RS12840 | Transposase DDE_Tnp_1_4 domain-containing protein | pfam13701 |
| | BZY55_RS12795 | Tn3 transposase DDE domain- containing protein | pfam01526 |
| <i>S. kanamyceticus</i> NRRL B-2535 | --- | --- | --- |
| <i>S. formicae</i> KY5 | --- | --- | --- |
| <i>S. cinnamoneus</i> DSM 41675 | --- | --- | --- |
| <i>S. platensis</i> DSM 40041 | BG653_RS05930 | Transposase DDE_Tnp_4 superfamily endonuclease | pfam13359 |
| <i>S. ossamyceticus</i> NRRL B-3822 | --- | --- | --- |
| <i>A. spheciospongae</i> EG49 | UO65_RS00270 | Integrase | pfam00665/ pfam13683 |
| | UO65_RS00275 | Transposase DDE domain-containing protein | pfam01609 |
| <i>T. mobilis</i> MCCC 1A02139 | AUP44_RS28005 | Putative IS4/5 family transposase | pfam13340 |
| | bpln_RS33885 | Transposase DDE_Tnp_1_5 superfamily endonuclease | pfam13737 |
| <i>B. plantarii</i> ATCC 43733 | bpln_RS33900 | Putative IS4/5 family transposase | pfam13340 |
| | bpln_RS33905 | Transposase | pfam01527 |
| | bpln_RS10520 | IS66 Orf2 like protein | pfam05717 |
| <i>P. aeruginosa</i> SCV20265 | --- | --- | --- |

Table S5. Mobile genetic elements and tRNA genes surrounding the *bcm* cluster in different bacteria.

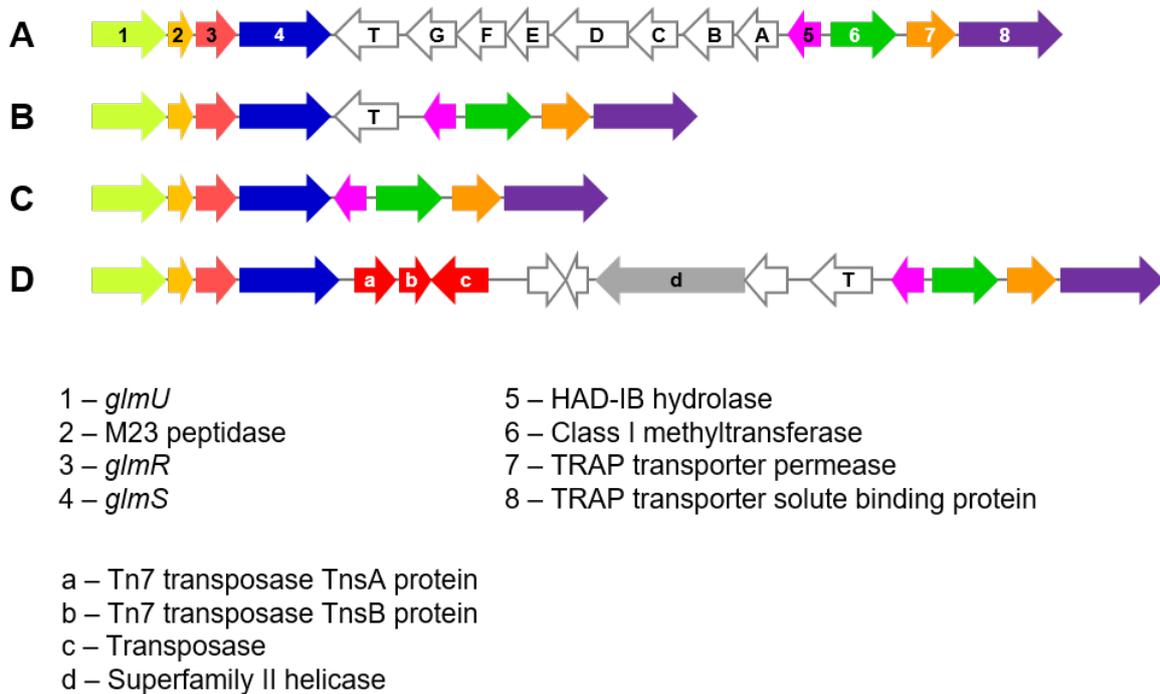
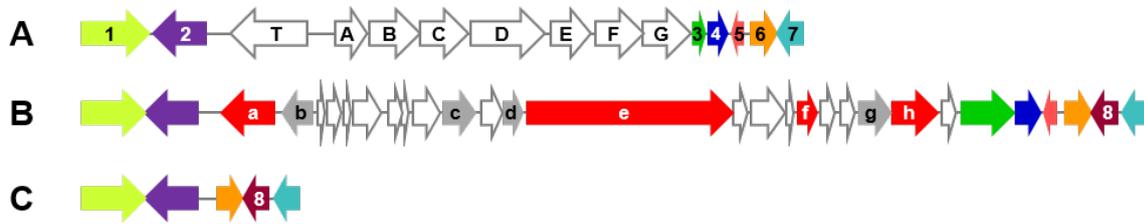


Figure S11. Genomic context of the *bcm* cluster (genes A to T) in *P. aeruginosa* and synteny of its flanking genes. (A) Schematic representation of the genes surrounding the *bcm* cluster in all *bcm*-positive *P. aeruginosa* strains, including SCV20265, M18 and ATCC 14886. (B) Organization of the flanking genes in *bcm*-minus strains that feature a *bcmT* homolog (99% identity, 100% coverage) adjacent to *glmS*, including *P. aeruginosa* PAO1 (AE004091), PA14 (ASWV0100021), ATCC 700888 (AKZF01000541) and LESB58 (FM209186). (C) Example of genetic organization where the flanking genes are contiguous to each other, observed in *P. aeruginosa* PA7 (CP000744) and VRFPA01 (AOBK01000101). (D) Genetic organization in *P. aeruginosa* BL08 (NZ_KI518902), where elements of transposon Tn7 along with a *bcmT* homolog are present instead of the full *bcm* cluster. *bcm* flanking genes are identified by numbers while insertion genes are labelled with lowercase letters. Products encoded by these genes are listed in the legend.



- | | |
|---------------------------------------|--|
| 1 – TQXA domain-containing protein | 5 – Hypothetical protein |
| 2 – Histidinol-phosphate transaminase | 6 – SRPBCC domain-containing protein |
| 3 – HMGL-like protein | 7 – TetR transcriptional regulator |
| 4 – Hypothetical protein | 8 – N-acetyltransferase |
| a – Integrase | e – Phage tail tape measure protein TP901 family |
| b – DUF1537 protein | f – HNH endonuclease |
| c – HTH domain-containing protein | g – SAM-dependent methyltransferase |
| d – HTH domain-containing protein | h – uncharacterized phage associated protein |

Figure S12. Genomic context of the *bcm* cluster (genes T to G) in *Mycobacterium* and synteny of its flanking genes. (A) Schematic representation of the genes surrounding the *bcm* cluster in *M. chelonae* CCUG 47445. (B) Organization of the same genomic area in *M. abscessus* ATCC 19977 (CU458896) and 6G-0125-R (AKUE01000005) where the *bcm* genes are substituted by a cluster of uncharacterized and phage-associated genes. (C) Gene organization observed in several isolates of *M. abscessus* subsp. *bolletii* (including accessions AKUO01000002, AKUL01000005 and AGQU01000004), where the *bcm* cluster and some of its flanking genes are absent, and no phage is integrated. *bcm* flanking genes are identified by numbers while insertion genes are labelled with lowercase letters. Products encoded by these genes are listed in the legend.

```

                                10      20      30
BcmF_Scin_ MTTVVDNE.....GHLHLPTARVTAGRLLFDAAEGADQAL
BcmC_Scin_ VSTETLR.....LQKARATEEGLAFETPGGLTRAL
bcmE_Scin_ .MASPDSATLREPVVLPMPGGEHEARAAAYPPIG...LERSRVTGGRLVFDRDEGFDRAL
BcmB_Scin_ MSRAPGNATAAPEIRRGRIYRDLYEKRASGPAVQGDALHLEARIQGDRLVFAFSRARETAL
BcmG_Scin_ MSTAQGYG.....WQTAALRGGELVFSTPGGIEQAL

                                40      50      60      70      80      90
BcmF_Scin_ ALGAFCLAVPEDDLVEPGLRFCSRFEYEP..AEPGTADR YRGHREDGHA...DSKLG YEDR
BcmC_Scin_ RDGCFLLAVPFGDTPGVTLCRERFRPVEQGGESTRAYRGFRDLGVD...YFDR
bcmE_Scin_ AQGFELVRIPEGTDPAAGDRFAAHFHEE..RAGGDP LDAYRGYRHVRVP...GDYQGYFDR
BcmB_Scin_ ADCVLELLEIPADIDVAAGDAFSRQFHL...GPDSP.PYGRFRDLGSEHFDPDLLGFHQR
BcmG_Scin_ RDGFHVEQEEGLDLTAGDRFARGFYLP..GEPDSTDPFRGFQHWTSERL.GPRQGYFCR

                                100     110     120     130     140     150
BcmF_Scin_ .PDQVEQLQLESHLWSRYLPEEVTALERMKDLTLDALYGVFVDVAGIPFHEHDRETVTGGAR
BcmC_Scin_ EHFQTEHVLIDGPGREERHFPPELRRMAEHMHELARHVLRTVLELGVARELWSEVTGGAV
bcmE_Scin_ EHDQWENFYVE RDNWD.VLPSVAVRVRGMAGLGVTILRGVLEHLR LPREHWARVTGGLT
BcmB_Scin_ .VNOIEFLLERRFWASDYPEEIAIRLGEQLTRLSQKVLCAVLSHVGVPERDRRRRATGGCS
BcmG_Scin_ DDDQTEQFFLESAHWDSVYPAQALRQAEMRS LALDVLRAVLAHLELPPELWDEATGRCL

                                160     170     180     190     200     210
BcmF_Scin_ QDTGLCYTTVNHYRADLSDRAGIVEHSDSGFITLICTDQPGYEILHEGRWRP VREEPGHF
BcmC_Scin_ DGRGTEWFAANHYRSE RDRLL.GCAPHKDGTGFVTVLYIEEGGLEAATGGSWTPVDPVPGCF
bcmE_Scin_ EDRGHQMLAENHRSRSHKGV.R.GSKFHRDSGWVTVLRSVDPGLLALVDGRLWAVDPEPGHF
BcmB_Scin_ RAA.GSYHLTFNHYRPEHRDV.GLSSHKDDGFLTILRITTPGLEVNRKDRWERVPVDPDCF
BcmG_Scin_ SAR.GTYNLTFNHRRPEVPRR.GLNVHKDSGWVTVLRSSTDPLEVERDGAWHPIIDPRPGTF

                                220     230     240     250     260
BcmF_Scin_ VVNIIGDAFRVLT RKLPRPVTA VYHRVPELRPDGA.AHHRSSFIT IY...MGPRYDMM LHQ
BcmC_Scin_ VVNFEGGAF ELLTSGLDRPVRALLHRV RQCAPRPE.SADRFSFAAF...VNPPPTGD LYR
bcmE_Scin_ IVNFGSSLEVLTERLDRPVRANVHG VVSTERAPG.QPDRTSYVTF...LDSDLTGT VYR
BcmB_Scin_ VINFGLSMEILTAPTKAPVAATMHRVARQ...GGDRSSEGHFSSSGCAPGMDEG VFR
BcmG_Scin_ IVNFGCAIEILLTRDTRTPVAAVAHRV VQQPRTDERKPD RFSYALFV DSSLDEDICPG LFR

                                270     280     290
BcmF_Scin_ YAADGTLHEYQGFRDFSV EKS KKLGYEF.....HSRI.....
BcmC_Scin_ VGADGTA TVARSTEDFL RDNFNERT.WGDGYADFGIAPPEPAGVAEDGVRA
bcmE_Scin_ FE.NGTPRP LQSVAEFA GQEVGRT.YDD.....SGAL.....
BcmB_Scin_ YLPGSGLDRVCGSR ELI DENDHEI.YAG.....TDAPGDKRREH.....
BcmG_Scin_ YEPGTGLRLETNFGTFL D TILHNT.YQK.....DTAGLY.....

```

| | BcmF_Scin | BcmC_Scin | BcmE_Scin | BcmB_Scin | BcmG_Scin |
|-----------|-----------|-----------|-----------|-----------|-----------|
| BcmF_Scin | 100 | 33.1 | 34.2 | 34.3 | 32.4 |
| BcmC_Scin | 33.1 | 100 | 37.8 | 31.1 | 38.7 |
| BcmE_Scin | 34.2 | 37.8 | 100 | 31.9 | 42.0 |
| BcmB_Scin | 34.3 | 31.1 | 31.9 | 100 | 42.1 |
| BcmG_Scin | 32.4 | 38.7 | 42.0 | 42.1 | 100 |

Figure S13. Alignment and percentage identity matrix of the *bcm* 2-OG/Fe-dioxygenases from *S. cinamomeus*. Figure generated using Esript 3.0 (2).

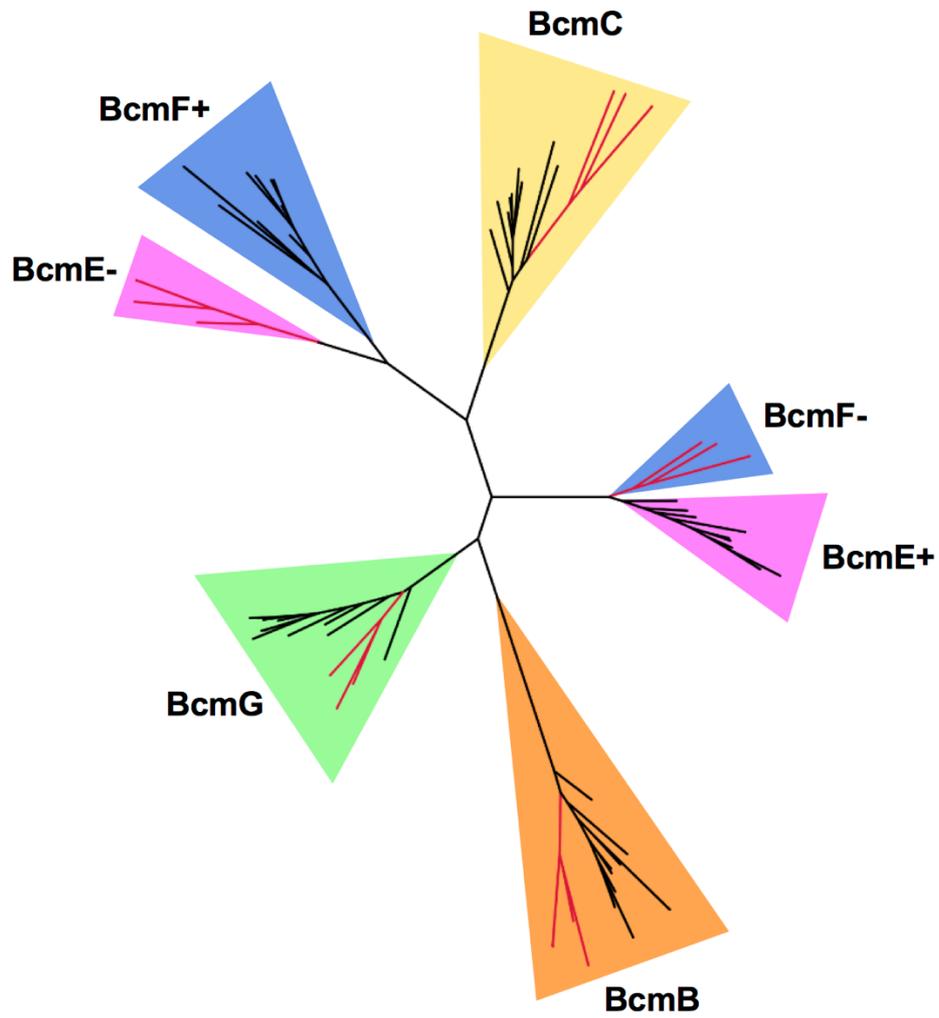


Figure S14. Unrooted version of the phylogenetic tree shown in Fig. 6 without the outgroup. Clade and branch colors are the same as those in Fig. 6.

SUPPLEMENTARY REFERENCES

1. Jacques IB, Moutiez M, Witwinowski J, Darbon E, Martel C, Seguin J, Favry E, Thai R, Lecoq A, Dubois S, Pernodet J-L, Gondry M, Belin P. 2015. Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat Chem Biol* 11:721–727.
2. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324.
3. Kohn H, Abuzar S, Korp JD, Zektzer AS, Martin GE. 1988. Structural studies of bicyclomycin. *J Heterocyclic Chem* 25:1511–1517.