

Supplementary Information

Integrated Omics Strategy Reveals Cyclic Lipopeptides Empedopeptins from *Massilia* sp. YMA4 and their Biosynthetic Pathway

Shang-Tse Ho^{1, 2#}, Ying-Ning Ho^{3, 4#}, Chih Lin¹, Wei-Chen Hsu¹, Han-Jung Lee¹, Chia-Chi Peng¹, Han-Tan Cheng¹, Yu-Liang Yang^{1, *}

¹ Agricultural Biotechnology Research Center, Academia Sinica, Taipei 11529, Taiwan

² Department of Wood Based Materials and Design, National Chiayi University, Chiayi 60004, Taiwan

³ Institute of Marine Biology and Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung 20224, Taiwan

⁴ Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung 20224, Taiwan

Tables

Table S1. Secondary metabolite gene clusters identified in <i>Massilia</i> sp. YMA4	2
Table S2. Structure information of cyclic lipopeptides	3
Table S3. Structure information of linear lipopeptides	4
Table S4. Genetic context and features of the BGC 6 in <i>Massilia</i> sp. YMA4 and its related BGCs in various microbes.	6
Table S5. Oligonucleotides primer sets used in this study	9
Table S6. Primer sets for empedopeptin biosynthetic genes of <i>Massilia</i> sp. YMA4	10

Figures

Figure S1. Phylogenetic tree of genus <i>Massilia</i> .	11
Figure S2. Tandem mass spectrum and fragment annotation of empedopeptin	12
Figure S3. Schematic diagram of insertion mutant for <i>Massilia</i> sp. YMA4	13
Figure S4. PCR check of mutants.	14

Table S1. Secondary metabolite gene clusters identified in *Massilia* sp. YMA4 by genome mining. Identified secondary metabolite clusters were predicted by antiSMASH 5.1.

Clusters	Type	Most similar known cluster	Similarity ^a
1	Siderophore	Unknown	
2	Hserlactone	Obafluorin	14%
3	Terpene	Unknown	
4	Acyl amino acids	Unknown	
5	NRPS	Turnerbactin	76%
6	NRPS	Entolysin	8%
7	Bacteriocin	Unknown	
8	NRPS, arylpolyene	APE Cf	10%
9	Thiopeptides	Unknown	
10	Hserlactone	Unknown	
11	Terpene	Carotenoid	100%
12	NRPS	Unknown	

^a The percentage of genes within the closest known compound that have a significant BLAST hit to genes within the current gene cluster.

Table S2. Structure information of cyclic lipopeptides (core structure: 3-OH-Fatty acid-AA1-AA2-AA3-AA4-AA5-AA6-AA7-AA8)

Producing strains	Formula	[M+H] ⁺	[M+H] ⁺ calcd ^a	Error ^b	Sequences (AA8 to AA1)	3-OH-fatty acid
<i>ΔempB</i>	C ₄₉ H ₇₉ N ₁₁ O ₁₆	1078.5750	1078.5785	-3.2	D/P/S/D/R/P/S/P	C14
wt ^c	C ₄₇ H ₇₅ N ₁₁ O ₁₈	1082.5368	1082.5370	-0.2	3-OH-D/P/S/3-OH-D/R/P/S/P	C12
wt	C ₄₆ H ₇₃ N ₁₁ O ₁₉	1084.5156	1084.5162	-0.6	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C11
<i>ΔempA</i>	C ₄₉ H ₇₉ N ₁₁ O ₁₇	1094.5679	1094.5734	-5.0	D/P/S/3-OH-D/R/P/S/P	C14
<i>ΔempB</i>	C ₄₉ H ₇₉ N ₁₁ O ₁₇	1094.5699	1094.5734	-3.2	3-OH-D/P/S/D/R/P/S/P	C14
wt	C ₄₇ H ₇₅ N ₁₁ O ₁₉	1098.5315	1098.5319	-0.4	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C12
<i>ΔempB</i>	C ₅₁ H ₈₁ N ₁₁ O ₁₆	1104.5927	1104.5941	-1.3	D/P/S/D/R/P/S/P	C16:1
<i>ΔempB</i>	C ₅₁ H ₈₃ N ₁₁ O ₁₆	1106.6094	1106.6098	-0.4	D/P/S/D/R/P/S/P	C16
wt	C ₄₈ H ₇₇ N ₁₁ O ₁₉	1112.5481	1112.5475	0.5	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C13
<i>ΔempB</i>	C ₄₉ H ₇₉ N ₁₁ O ₁₇	1120.5876	1120.5890	-1.2	3-OH-D/P/S/D/R/P/S/P	C16:1
<i>ΔempA</i>	C ₅₁ H ₈₃ N ₁₁ O ₁₇	1122.6049	1122.6047	0.2	D/P/S/3-OH-D/R/P/S/P	C16
<i>ΔempB</i>	C ₅₁ H ₈₃ N ₁₁ O ₁₇	1122.6036	1122.6047	-1.0	3-OH-D/P/S/D/R/P/S/P	C16
all strains	C ₄₉ H ₇₉ N ₁₁ O ₁₉	1126.5630	1126.5632	-0.2	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C14
<i>ΔempB</i>	C ₅₁ H ₈₁ N ₁₁ O ₁₈	1136.5819	1136.5839	-1.8	3-OH-D/3-OH-P/S/D/R/P/S/P	C16
wt	C ₅₀ H ₈₁ N ₁₁ O ₁₉	1140.5783	1140.5788	-0.4	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C15
wt	C ₅₁ H ₈₁ N ₁₁ O ₁₉	1152.5775	1152.5788	-1.1	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C16:1
wt	C ₅₁ H ₈₃ N ₁₁ O ₁₉	1154.5932	1154.5945	-1.1	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C16

^a Calculated [M + H]⁺, ^b ppm, ^c Wild-type.

Table S3. Structure information of linear lipopeptides (core structure: 3-OH-Fatty acid-AA1-AA2-AA3-AA4-AA5-AA6-AA7-AA8)

Producing strains	Formula	[M+H] ⁺	[M+H] ⁺ calcd ^a	Error ^b	Sequences	3-OH-fatty acid
	6 AAs				AA6 to AA1	
wt, $\Delta empA$	C ₃₈ H ₆₅ N ₉ O ₁₄	872.4707	872.4729	-2.5	S/3-OH-D/R/P/S/P	C12
wt, $\Delta empA$	C ₄₀ H ₆₉ N ₉ O ₁₄	900.5041	900.5042	-0.1	S/3-OH-D/R/P/S/P	C14
$\Delta empA$	C ₄₁ H ₇₁ N ₉ O ₁₄	914.5175	914.5199	-2.6	S/3-OH-D/R/P/S/P	C15
wt, $\Delta empA$	C ₄₀ H ₆₉ N ₉ O ₁₅	916.4994	916.4991	0.3	S/3-OH-D/R/3-OH-P/S/P	C14
wt, $\Delta empA$	C ₄₂ H ₇₁ N ₉ O ₁₄	926.5195	926.5199	-0.4	S/3-OH-D/R/P/S/P	C16:1
wt, $\Delta empA$	C ₄₂ H ₇₃ N ₉ O ₁₄	928.5344	928.5355	-1.2	S/3-OH-D/R/P/S/P	C16
	7 AAs				AA7 to AA1	
wt, $\Delta empA$	C ₄₃ H ₇₂ N ₁₀ O ₁₅	969.5242	969.5257	-1.5	P/S/3-OH-D/R/P/S/P	C12
$\Delta empB$	C ₄₅ H ₇₆ N ₁₀ O ₁₄	981.5599	981.5621	-2.2	P/S/D/R/P/S/P	C14
wt, $\Delta empA$	C ₄₃ H ₇₂ N ₁₀ O ₁₆	985.5182	985.5206	-2.4	3-OH-P/S/3-OH-D/R/P/S/P	C12
$\Delta empB$	C ₄₆ H ₇₈ N ₁₀ O ₁₄	995.5744	995.5777	-3.3	P/S/D/R/P/S/P	C15
wt, $\Delta empA$	C ₄₅ H ₇₆ N ₁₀ O ₁₅	997.5558	997.5570	-1.2	P/S/3-OH-D/R/P/S/P	C14
$\Delta empB$	C ₄₇ H ₇₈ N ₁₀ O ₁₄	1007.5753	1007.5777	-2.4	P/S/D/R/P/S/P	C16:1
$\Delta empB$	C ₄₇ H ₈₀ N ₁₀ O ₁₄	1009.5840	1009.5834	0.6	P/S/D/R/P/S/P	C16
$\Delta empA$	C ₄₆ H ₇₈ N ₁₀ O ₁₅	1011.5728	1011.5726	0.2	P/S/3-OH-D/R/P/S/P	C15
wt, $\Delta empA$	C ₄₅ H ₇₆ N ₁₀ O ₁₆	1013.5517	1013.5519	-0.2	3-OH-P/S/3-OH-D/R/P/S/P	C14
wt, $\Delta empA$	C ₄₇ H ₇₈ N ₁₀ O ₁₅	1023.5724	1023.5726	-0.2	P/S/3-OH-D/R/P/S/P	C16:1
wt, $\Delta empA$	C ₄₇ H ₈₀ N ₁₀ O ₁₅	1025.5853	1025.5883	-2.9	P/S/3-OH-D/R/P/S/P	C16
wt, $\Delta empA$	C ₄₅ H ₇₆ N ₁₀ O ₁₇	1029.5483	1029.5468	1.5	3-OH-P/S/3-OH-D/R/3-OH-P/S/P	C14
wt, $\Delta empA$	C ₄₇ H ₇₈ N ₁₀ O ₁₆	1039.5680	1039.5679	0.1	3-OH-P/S/3-OH-D/R/P/S/P	C16:1
$\Delta empA$	C ₄₈ H ₈₂ N ₁₀ O ₁₅	1039.6040	1039.6041	-0.1	P/S/3-OH-D/R/P/S/P	C17

	8 AAs				AA8 to AA1	
wt	C ₄₇ H ₇₇ N ₁₁ O ₂₀	1114.5269	1114.5268	0.1	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C12:1
wt	C ₄₇ H ₇₇ N ₁₁ O ₂₀	1116.5414	1116.5425	-1.0	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C12
wt	C ₄₉ H ₈₁ N ₁₁ O ₂₀	1144.5723	1144.5738	-1.3	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C14
wt	C ₅₁ H ₈₃ N ₁₁ O ₂₀	1170.5879	1170.5894	-1.3	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C16:1
wt	C ₅₁ H ₈₅ N ₁₁ O ₂₀	1172.6031	1172.6051	-1.7	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C16
wt	C ₅₁ H ₈₅ N ₁₁ O ₂₁	1188.6003	1188.6000	0.3	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C12
wt	C ₅₃ H ₈₇ N ₁₁ O ₂₀	1198.6215	1198.6207	0.7	3-OH-D/3-OH-P/S/3-OH-D/R/P/S/P	C18:1

^a Calculated [M + H]⁺, ^b ppm, ^c Wild-type.

Table S4. Genetic context and features of the BGC 6 in *Massilia* sp. YMA4 and its related BGCs in various microbes

Microbes	Accession number	Gene name (Locus tag)	Proposed function	AA length
<i>Massilia</i> sp. YMA4	CP030092.1	<i>empA</i> (DPH57_09130)	TauD/TfdA family dioxygenase	305
		<i>empB</i> (DPH57_09135)	TauD/TfdA family dioxygenase	326
		<i>empC</i> (DPH57_09140)	non-ribosomal peptide synthetase	2751
		<i>empD</i> (DPH57_09145)	amino acid adenylation domain-containing protein	1122
		<i>empE</i> (DPH57_09150)	amino acid adenylation domain-containing protein	5322
<i>C. fungivorans</i> ESAIA1	RAIL01000001.1	<i>geneA</i> (DFH22_2667)	TfdA family taurine catabolism dioxygenase TauD	308
		<i>geneB</i> (DFH22_2666)	TauD/TfdA family dioxygenase	317
		<i>geneC</i> (DFH22_2665)	amino acid adenylation domain-containing protein	2756
		<i>geneD</i> (DFH22_2664)	amino acid adenylation domain-containing protein	1171
		<i>geneE</i> (DFH22_2663)	amino acid adenylation domain-containing protein	5358
<i>C. fungivorans</i> Ter331	CP002745.1	<i>geneA</i> (CFU_2180)	TauD/TfdA family dioxygenase	274
		<i>geneB</i> (CFU_2181)	TauD/TfdA family dioxygenase	317
		<i>geneC</i> (CFU_2182)	putative non-ribosomal peptide synthetase	2815
		<i>geneD</i> (CFU_2184)	non-ribosomal peptide synthetase	1168
		<i>geneE</i> (CFU_2185)	amino acid adenylation	5375
<i>C. fungivorans</i> Ter6	CP013232.1	<i>geneA</i> (CFter6_2734)	TauD/TfdA family dioxygenase	305
		<i>geneB</i> (CFter6_2735)	TauD/TfdA family dioxygenase	317
		<i>geneC</i> (CFter6_2736)	amino acid adenylation domain-containing protein	2763
		<i>geneD</i> (CFter6_2737)	non-ribosomal peptide synthetase	1170
		<i>geneE</i> (CFter6_2738)	non-ribosomal peptide synthetase	5361

<i>Duganella sacchari</i>	FRCX01000009.1	<i>geneA</i> (SAMN05192549_102249)	TauD/TfdA family dioxygenase	305
		<i>geneB</i> (SAMN05192549_102250)	TauD/TfdA family dioxygenase	316
		<i>geneC</i> (SAMN05192549_102251)	non-ribosomal peptide synthetase	2747
		<i>geneD</i> (SAMN05192549_102252)	AMP-binding protein	871
		<i>geneE</i> (SAMN05192549_102253)	non-ribosomal peptide synthetase	5334
<i>V. guangxiensis</i>	RXFT01000001.1	<i>geneA</i> (EJP67_06565)	TauD/TfdA family dioxygenase	304
		<i>geneB</i> (EJP67_06570)	TauD/TfdA family dioxygenase	328
		<i>geneC</i> (EJP67_06575)	amino acid adenylation domain-containing protein	2794
		<i>geneD</i> (EJP67_06580)	amino acid adenylation domain-containing protein	1172
		<i>geneE</i> (EJP67_06585)	amino acid adenylation domain-containing protein	5529
<i>Variovorax</i> sp. OV084	FOII01000021.1	<i>geneA</i> (SAMN05443580_1317)	TauD/TfdA family dioxygenase	318
		<i>geneB</i> (SAMN05443580_1316)	Taurine dioxygenase, alpha-ketoglutarate-dependent	314
		<i>geneC</i> (SAMN05443580_1315)	amino acid adenylation domain-containing protein	2767
		<i>geneD</i> (SAMN05443580_1314)	non-ribosomal peptide synthetase	1166
		<i>geneE</i> (SAMN05443580_1313)	non-ribosomal peptide synthetase	5350

<i>Variovorax</i> sp. YR752	OCMW01000002.1	<i>geneA</i> (SAMN05518800_2816)	TauD/TfdA family dioxygenase	314
		<i>geneB</i> (SAMN05518800_2815)	Taurine dioxygenase, alpha-ketoglutarate-dependent	318
		<i>geneC</i> (SAMN05518800_2814)	amino acid adenylation domain-containing protein	2765
		<i>geneD</i> (SAMN05518800_2813)	non-ribosomal peptide synthetase	1170
		<i>geneE</i> (SAMN05518800_2812)	amino acid adenylation domain-containing protein	5355

^a The proposed function of each gene was predicted by NCBI BlastP.

Table S5. Oligonucleotides primer sets used in this study

Primer	Sequence (5'→3')
qRT-PCR	
16S-F	CCTGAAGAATAAGCACCGGCTAACTA
16S-R	GGGGATTTACGACAGACTTACAAAA
<i>empA</i> -F	CTTGACAGGGTGTTGCACG
<i>empA</i> -R	ATCAACCATGAACGACGGGA
<i>empB</i> -F	GCCTCGATCAGGTAGGGGAA
<i>empB</i> -R	CACTACTGGACGAACACACCA
<i>empC</i> -F	TTGTTTGCGTCGAACAGCTC
<i>empC</i> -R	CCCAATGCGATACCGGAGAG
<i>empD</i> -F	CAGGTCCTCCTTGTCGAA
<i>empD</i> -R	TTGAGGAGGTCGAGCTGTGA
<i>empE</i> -F	CAACGACGCCGATATTGCAG
<i>empE</i> -R	CCCCCTGCGTCTGTTTTCAA
PCR	
<i>empA</i> -ck1-F	ATCAAAGGTACGCAAGGCGA
<i>empA</i> -ck2-R	GATCGAGCAATCGGTCAGCA
<i>empB</i> -ck1-F	CTTTCGAGGTGACGGCCA
<i>empB</i> -ck2-R	TGTTTCGACGCATGCCACA
<i>empC</i> -ck1-F	GACGCAAACAAGGTCGAGGT
<i>empC</i> -ck2-R	CTCTCGATGTGGCGGATACC
<i>empD</i> -ck1-F	GCACTGGAATTGCTGGACAA
<i>empD</i> -ck2-R	CGTGTTGAGGAACATGCCCA
<i>empE</i> -ck1-F	TCAGTTGAGTTCACCCCAGC
<i>empE</i> -ck2-R	GTAGCAGCGTTTCAGTTCCG
pCM184-ck3-R	CGAACGACATGGAGCGGCAC
pCM184-ck4-2-F	AAAGCGGTTAGCTCCTTCGG

Table S6. Primer sets for empedopeptin biosynthetic genes of *Massilia* sp. YMA4

Target gene	Primer set for positive PCR (left arm; primer 1 and 3)	Primer set for positive PCR (right arm; primer 2 and 4)	Primer set for negative PCR (primer 1 and 2)	PCR product size for negative PCR (wild type)	PCR product size for positive PCR (left arm)	PCR product size for positive PCR (right arm)
<i>empA</i>	<i>empA</i> -ck1-F/pCM184-ck3-R	pCM184-ck4-2-F/ <i>empA</i> -ck2-R	<i>empA</i> -ck1-F/ <i>empA</i> -ck2-R	780 bp	769 bp	1497 bp
<i>empB</i>	<i>empB</i> -ck1-F/pCM184-ck3-R	pCM184-ck4-2-F/ <i>empB</i> -ck2-R	<i>empB</i> -ck1-F/ <i>empB</i> -ck2-R	690 bp	905 bp	1477 bp
<i>empC</i>	<i>empC</i> -ck1-F/pCM184-ck3-R	pCM184-ck4-2-F/ <i>empC</i> -ck2-R	<i>empC</i> -ck1-F/ <i>empC</i> -ck2-R	1355 bp	972 bp	1612 bp
<i>empD</i>	<i>empD</i> -ck1-F/pCM184-ck3-R	pCM184-ck4-2-F/ <i>empD</i> -ck2-R	<i>empD</i> -ck1-F/ <i>empD</i> -ck2-R	1161 bp	644 bp	1280 bp
<i>empE</i>	<i>empE</i> -ck1-F/pCM184-ck3-R	pCM184-ck4-2-F/ <i>empE</i> -ck2-R	<i>empE</i> -ck1-F/ <i>empE</i> -ck2-R	949 bp	699 bp	1358 bp

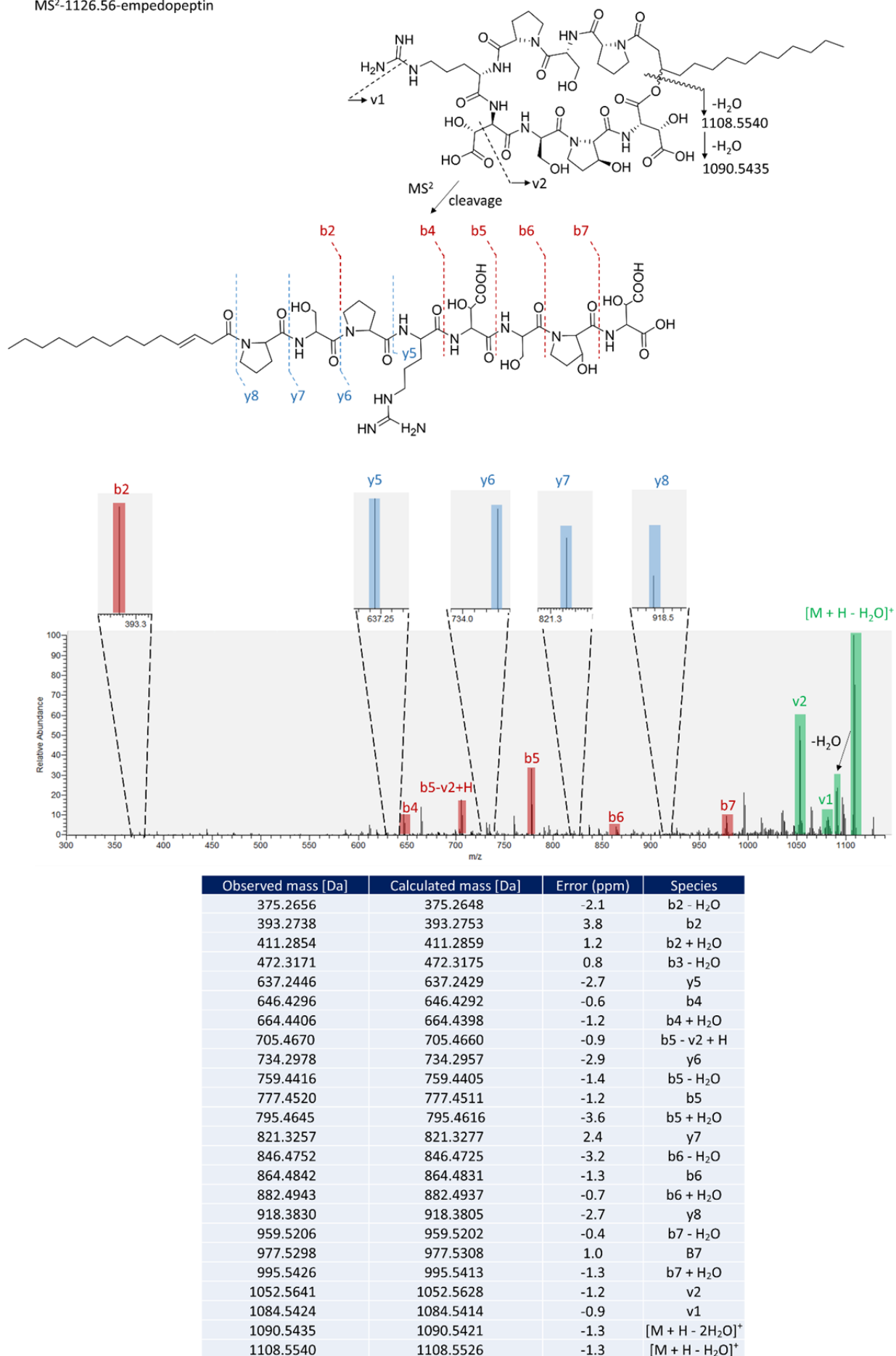


Figure S2. Tandem mass spectrum and fragment annotation of empedopeptin.

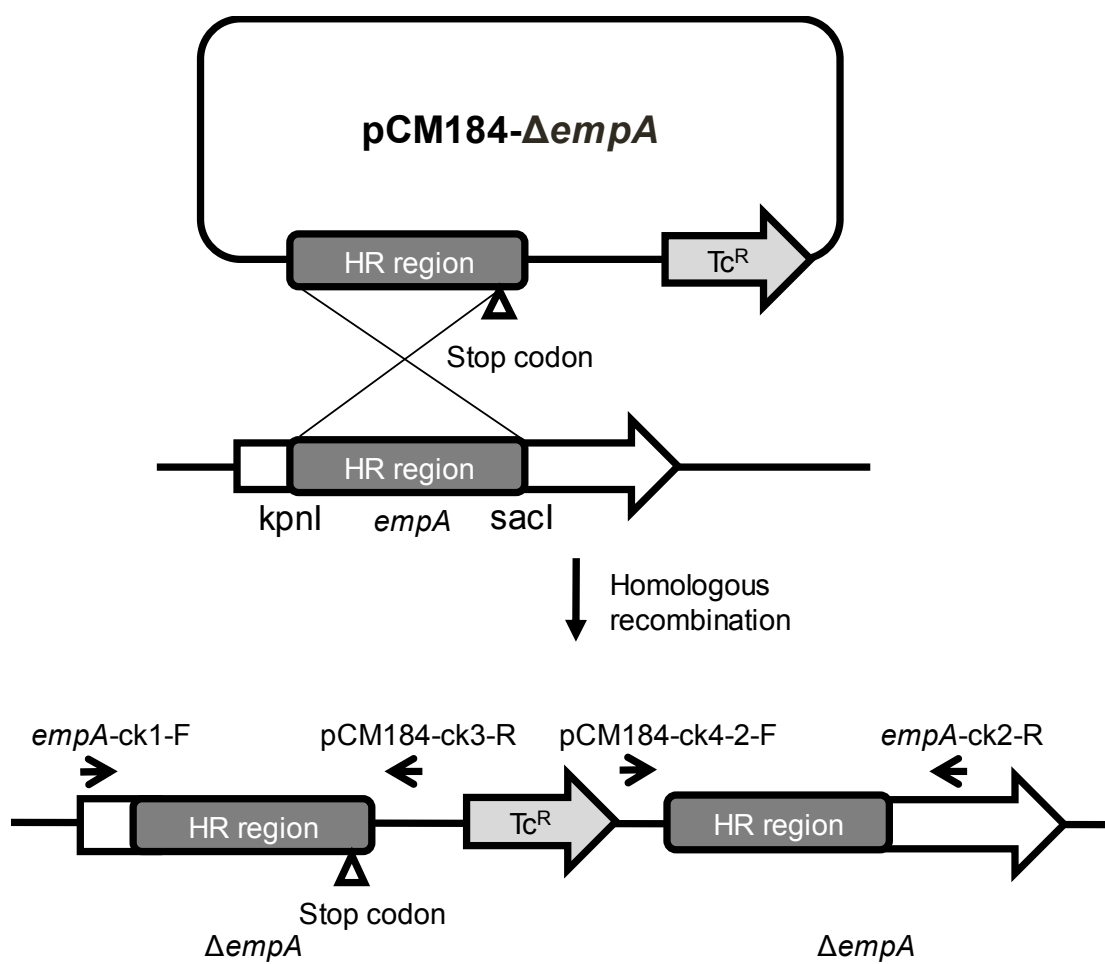


Figure S3. Schematic diagram of insertion mutant for *Massilia* sp. YMA4.

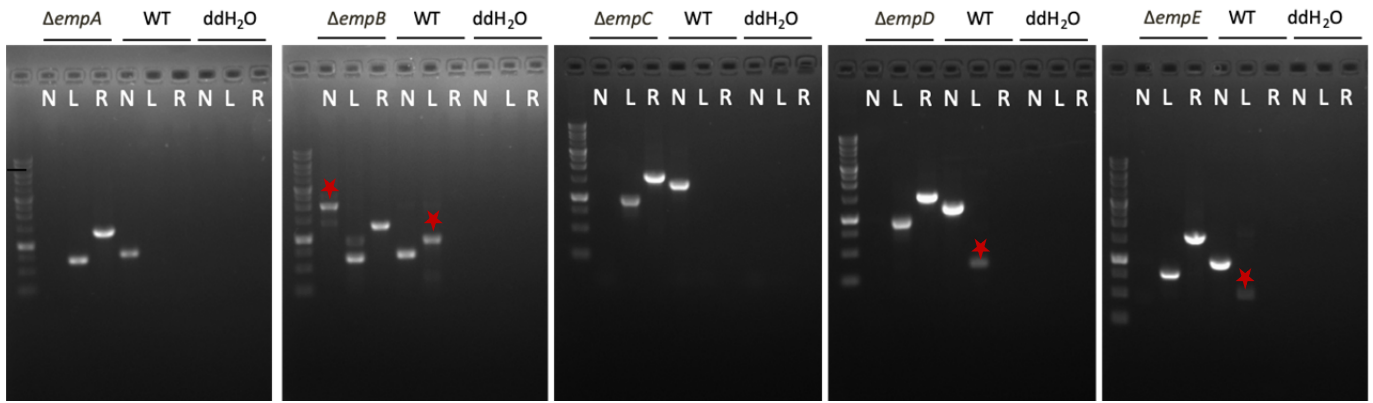


Figure S4. PCR check of mutants. The represented primer set of PCR check for $\Delta empA$: N: *empA*-ck1-F + *empA*-ck2-R; L: *empA*-ck1-F + pCM184-ck3-R; R: pCM184-ck4-2-F + *empA*-ck2-R. The positions of primers were shown in **Figure S3**. The sequences of all the bands were confirmed and ★ represented non-specific bands.